Weak Poincaré Inequalities, Simulated Annealing, and Sampling from Spherical Spin Glasses

Brice Huang[*]   Sidhanth Mohanty[†]   Amit Rajaraman[‡]   David X. Wu[§]

November 22, 2024

Abstract

There has been a recent surge of powerful tools to show rapid mixing of Markov chains, via functional inequalities such as *Poincaré inequalities*. In many situations, Markov chains fail to mix rapidly from a worst-case initialization, yet are expected to approximately sample from a random initialization. For example, this occurs if the target distribution has *metastable states*, small clusters accounting for a vanishing fraction of the mass that are essentially disconnected from the bulk of the measure. Under such conditions, a Poincaré inequality cannot hold, necessitating new tools to prove sampling guarantees.

We develop a framework to analyze simulated annealing, based on establishing so-called *weak Poincaré inequalities*. These inequalities imply mixing from a suitably warm start, and simulated annealing provides a way to chain such warm starts together into a sampling algorithm. We further identify a local-to-global principle to prove weak Poincaré inequalities, mirroring the spectral independence and localization schemes frameworks for analyzing mixing times of Markov chains.

As our main application, we prove that simulated annealing samples from the Gibbs measure of a spherical spin glass for inverse temperatures up to a natural threshold, matching recent algorithms based on algorithmic stochastic localization. This provides the first Markov chain sampling guarantee that holds beyond the *uniqueness threshold* for spherical spin glasses, where mixing from a worst-case initialization is provably slow due to the presence of metastable states. As an ingredient in our proof, we prove bounds on the operator norm of the covariance matrix of spherical spin glasses in the full replica-symmetric regime.

Additionally, we resolve a question related to sampling using data-based initializations.

Contents

1 Introduction

A common task of interest in computer science, probability, and physics is to efficiently sample from Gibbs distributions. For a Hamiltonian energy function $H : \Omega \to \mathbb{R}$ over state space $\Omega \subseteq \mathbb{R}^N$, the associated Gibbs distribution μ_H is defined by $\mathrm{d}\mu_H(x) \propto \exp(H(x)) \, \mathrm{d}x$.

The class of *Markov chain Monte Carlo* (MCMC) algorithms is arguably the most widely used tool for sampling from Gibbs distributions. In this paradigm, one sets up a Markov chain P_H whose stationary distribution is μ_H, and outputs the final state of a $\mathrm{poly}(N)$-time random walk according to P_H. Common choices include the Glauber dynamics, for discrete state spaces such as $\Omega = \{\pm 1\}^N$, and the Langevin diffusion, for continuous state spaces such as $\Omega = \mathbb{R}^N$ or $\sqrt{N} \cdot \mathbb{S}^{N-1}$.

To prove that such an algorithm indeed correctly samples from μ_H, one bounds the *mixing time* of the Markov chain. A common route to prove a bound on the mixing time is to establish functional inequalities, such as *Poincaré* inequalities. There are now powerful frameworks for proving such functional inequalities, such as *spectral independence* [ALOG21] and *localization schemes* [CE22]. The development of these frameworks has led to a flurry of activity in analyzing mixing times of Markov chains, including the resolution of several long-standing open problems in the algorithmic theory of counting and sampling [ALGV24, ALOG21, AJK⁺22, EKZ22, CE22].

The implications of these inequalities are quite strong. In particular, they imply that for *any* initial distribution ν, for an appropriate divergence function, a single step of the Markov chain shrinks the distance to the stationary distribution by a significant multiplicative factor:

$$\mathrm{Divergence}(P_H \nu \| \mu_H) \leqslant \left(1 - \frac{1}{\mathrm{poly}(N)}\right) \mathrm{Divergence}(P_H \nu \| \mu_H) \,.$$

The presence of such a functional inequality typically implies that a Markov chain mixes rapidly from a *worst-case initialization*.

Sampling from random initializations. Many natural Markov chains are expected to produce approximate samples from the Gibbs measure when started at a *random* initialization, but fail to mix rapidly from a worst-case initialization. Often, this is because the Gibbs measure contains pathological clusters (termed *metastable states* in the physics literature) that are essentially disconnected from most of the measure, and account for a vanishing fraction of the total mass. A Markov chain initialized in such a cluster will remain trapped inside it and fail to mix, and therefore methods that show mixing from worst-case initializations cannot give effective guarantees in such settings.

However, one may still hope to show that from a random initialization, the Markov chain samples from the non-pathological part of the Gibbs measure, which is statistically indistinguishable from the true Gibbs measure. In our work, we prove that under suitable conditions, the *simulated annealing* algorithm samples from a distribution close to the Gibbs measure.

Simulated annealing. In the simulated annealing algorithm, one defines a "schedule" of inverse temperatures, i.e. for $i = 0, \ldots, T$, let $\beta_i := i/T$. The algorithm initializes at a sample from the uniform distribution $\mu_{\beta_0 H}$. Then, for $i = 1, \ldots, T$, the i-th stage of the algorithm runs the Markov chain $P_{\beta_i H}$ corresponding to $\mu_{\beta_i H}$ for $\mathrm{poly}(N)$ time, initialized at the output of the previous stage.

The underlying idea of this algorithm is that, for T sufficiently large, the Gibbs distribution $\mu_{\beta_{i-1} H}$ is a "warm start" for $\mu_{\beta_i H}$, i.e. an initialization with suitably bounded likelihood ratio with $\mu_{\beta_i H}$. So, if one could show that each of the Markov chains $P_{\beta_i H}$ (approximately) mixes rapidly

1

from a warm start, one may inductively argue that the output of the i-th stage of the algorithm is an approximate sample from $\mu_{\beta_i H}$. In other words, simulated annealing chains a sequence of warm starts together into a sampling algorithm.

This algorithmic idea is widely used empirically, and has also been employed to obtain algorithms for approximating the volumes of convex bodies [DFK91, DF91, LS90, KLS97], approximating the number of perfect matchings in a bipartite graph [JSV04], and sampling from the random field Ising model at sufficiently high temperatures [AEGP23], among others. However, we lack a general theory for why simulated annealing achieves provable guarantees beyond the settings of sampling from log-concave distributions and convex bodies. Indeed, in contrast to the general recipes available to prove mixing from worst-case initialization, proofs of rapid mixing from warm starts often employ ad-hoc arguments.

One of the main contributions of this work is a framework for proving mixing from warm starts, which combined with the above discussion provides general sufficient conditions under which simulated annealing samples from the Gibbs measure. We achieve this by generalizing the frameworks of spectral independence and localization schemes, previously employed to prove mixing from worst-case initialization, to show mixing from warm starts (see Section 6 for details). As we discuss just below, our framework gives sampling guarantees for simulated annealing in regimes where mixing from worst-case initializations is provably false.

Main application: spherical spin glasses. In a *spherical mixed p-spin glass*, $H : \sqrt{N} \cdot \mathsf{S}^{N-1} \to \mathbb{R}$ is a random Hamiltonian parameterized by coefficients $\beta, \gamma_2, \ldots, \gamma_{p_*} \geqslant 0$ where:

$$H(\sigma) = \beta \sum_{p \geqslant 2} \frac{\gamma_p}{N^{(p-1)/2}} \sum_{i_1, \ldots, i_p} g_{i_1, \ldots, i_p} \sigma_{i_1} \cdots \sigma_{i_p}, \tag{1}$$

for i.i.d. $g_{i_1, \ldots, i_p} \sim \mathcal{N}(0, 1)$. The Gibbs distribution μ_H is very well-studied in probability, statistical physics, and average-case algorithm design, as it simultaneously exhibits rich behavior and is amenable to analytic tools. Notably, this model undergoes numerous sharp phase transitions as one increases β. For small β, the model satisfies a Poincaré inequality [GJ19]. Beyond a *uniqueness* transition β_{uniq}, small isolated clusters in μ_H known as *metastable states* start to appear [AJ24]. In particular, the natural Markov chain *Langevin diffusion* initialized from such states mixes slowly, thereby precluding a Poincaré inequality. However, these states account for a vanishing fraction of the measure under μ_H, and the Langevin diffusion with a random initialization is expected to still mix rapidly over a $1 - o_N(1)$ fraction of μ_H, thereby producing a sample with vanishing total variation distance from μ_H.

The threshold for efficient algorithmic sampling is believed to occur at the *shattering* transition β_{shatter} — beyond this transition, the Gibbs measure shatters into an exponential number of poorly-connected clusters with exponentially small mass, and mixing is provably slow [CHS93, AMS23b, GJK23]. It is expected that all efficient algorithms fail to sample from the Gibbs measure above β_{shatter}, and recently [AMS23b] gave rigorous evidence for this picture by showing that all *stable* algorithms fail.

We use our framework to prove that *annealed* Langevin diffusion, where one begins by running Langevin diffusion for $\beta_0 = 0$, and slowly increases the inverse temperature to the target β, samples from the spherical mixed p-spin glass. This leads to the first rigorous guarantee in this problem for a Markov chain beyond the uniqueness threshold.

2

Theorem 1.1 (Informal). *For any choice of $\gamma_2, \ldots, \gamma_{p_*}$, there is a threshold $\beta_{\mathrm{SL}} \leqslant \beta_{\mathrm{shatter}}$ such that for any $\beta < \beta_{\mathrm{SL}}$, with probability at least $1 - e^{-\Omega(N^{1/5})}$ over the randomness of H, annealed Langevin diffusion run for $\mathrm{poly}(N)$ time samples from a distribution whose total variation distance to μ_H is at most $e^{-\Omega(N^{1/5})}$.*

The thresholds β_{SL} and β_{shatter} are formally defined as the supremal β such that the inequalities (SL) and (Non-shattering) below hold. The recent work [HMP24] produces a different sampling algorithm based on algorithmic stochastic localization, which succeeds to the same threshold β_{SL}; see below for further discussion. This threshold is a fundamental barrier for stochastic localization based approaches, and we explain its physical significance in Remark 7.4.

Remark 1.2. In many models, we have $\beta_{\mathrm{uniq}} < \beta_{\mathrm{SL}} < \beta_{\mathrm{shatter}}$, and β_{SL} is close to β_{shatter}. For example, for the pure p-spin models (where $\gamma_p = 1$ and all the other γ_i are equal to 0), $\beta_{\mathrm{uniq}} \asymp (\log p)^{-1/2}$, while $\beta_{\mathrm{SL}}, \beta_{\mathrm{shatter}} \asymp 1$ and $\beta_{\mathrm{SL}}/\beta_{\mathrm{shatter}}$ is bounded from below by the universal constant $\sqrt{e}/2$. See [HMP24, Remark 1.1, Eq. 1.8].

1.1 Weak Poincaré inequalities and localization schemes

The starting point of our work is a relaxation of Poincaré inequalities, known as *weak Poincaré inequalities*, which can be leveraged to prove mixing from warm starts. To simplify the discussion, we restrict here to the setting of discrete Markov chains. Our main application is to a continuous Markov chain, namely the Langevin diffusion for a spherical spin glass, and we outline the differences in Remark 1.7 below.

Let P_H be a time-reversible Markov chain with stationary distribution μ_H. For any functions $f, g : \Omega \to \mathbb{R}$, define the *Dirichlet form* as $\mathcal{E}(f, g) := \mathbf{E}_{x \sim \mu_H} \mathbf{E}_{y \sim P_H x} (f(x) - f(y))(g(x) - g(y))$. We say P_H satisfies a C-*Poincaré inequality* if for any function $f : \Omega \to \mathbb{R}$:

$$\mathcal{E}(f, f) \geqslant C \cdot \mathbf{Var}[f],$$

for $C \geqslant 1/\mathrm{poly}(N)$. A Poincaré inequality has a classic implication for rapid mixing. In particular, for ν_t as the distribution obtained by running P_H for continuous time t starting at a distribution ν_0, we have:

$$\chi^2(\nu_t \| \mu_H) \leqslant \exp(-Ct) \cdot \chi^2(\nu_0 \| \mu_H).$$

We say P_H satisfies a (C, ε)-weak Poincaré inequality if for any function $f : \Omega \to \mathbb{R}$:

$$\mathcal{E}(f, f) \geqslant C \cdot \mathbf{Var}[f] - \varepsilon \cdot \|f - \mathbf{E}f\|_\infty^2$$

One can derive the following mixing guarantee from a weak Poincaré inequality; see, e.g., [RW01, Theorem 2.1].

$$\chi^2(\nu_t \| \mu_H) \leqslant \exp(-Ct) \cdot \chi^2(\nu_0 \| \mu_H) + \varepsilon \cdot \left\| \frac{\mathrm{d}\nu_0}{\mathrm{d}\mu_H} - 1 \right\|_\infty^2. \tag{2}$$

In particular, if ν_0 is a warm start for μ_H in the sense that $\left\| \frac{\mathrm{d}\nu_0}{\mathrm{d}\mu_H} - 1 \right\|_\infty$ is suitably small, this implies that the Markov chain's output distribution ν_t approximates μ_H.

Since the target measure in one stage of simulated annealing is a warm start for that of the next stage, such a guarantee allows one to inductively argue that simulated annealing succeeds at sampling. We summarize this implication below.

3

Theorem 1.3 (Informal, see Theorem 4.13). *If $P_{\beta H}$ satisfies a weak Poincaré inequality with suitable parameters for every $\beta \in [0, 1]$, then simulated annealing succeeds at sampling from μ_H.*

Localization schemes for weak Poincaré inequalities.   We restrict to the following simple setting: μ_H is a distribution on $\{\pm 1\}^N$. Let P_H be the *Glauber dynamics* Markov chain where in a single step from x, we pick a uniformly random coordinate $i \sim [N]$, and toggle x_i with probability:

$$\frac{\mu_H(x^{\oplus i})}{\mu_H(x) + \mu_H(x^{\oplus i})} \,.$$

A special case of the localization schemes framework is the *spectral independence* framework of Anari, Liu, and Oveis Gharan [ALOG21].

Theorem 1.4 ([AL20, ALOG21]). *The following local-to-global principle reduces proving a Poincaré inequality to establishing bounds on the spectrum of influence matrices. Suppose for **every** $S \subseteq [N]$, and **every** pinning x_S of coordinates in S, the spectral norm of its influence matrix Ψ_{S,x_S} is at most α, then the Glauber dynamics chain satisfies a $n^{-O(\alpha)}$-Poincaré inequality. Here, the influence matrix Ψ_{S,x_S} is an $(n - |S|) \times (n - |S|)$ matrix indexed by vertices $v \notin S$, where*

$$\Psi_{S,x_S}[a, b] := \mathbf{Pr}[x_a = +1 | x_b = +1] - \mathbf{Pr}[x_a = -1 | x_b = +1] \,.$$

While the above theorem has been influential and useful in proving mixing time bounds for a variety of Markov chains relevant to sampling combinatorial structures, the "for every" requirement in the above theorem is quite punishing in average-case settings. For example, in the presence of metastable states, P_H does not satisfy a Poincaré inequality, but may nevertheless satisfy a weak Poincaré inequality. In such cases, there are choices of S and x_S for which Ψ_{S,x_S} has large spectral norm, and the above statement has no implications for the mixing time of P_H.

We give a general local-to-global principle to prove weak Poincaré inequalities. A one-line summary of this local-to-global principle is:

> *Bounded influence over **all** pinnings implies a Poincaré inequality.*

An analogous summary of the local-to-global principle in the present paper is:

> *Bounded influencce over **typical** pinnings implies a **weak** Poincaré inequality.*

To give a more concrete instantiation of our message, our result implies a "softer" version of Theorem 1.4, tolerant to some "bad" pinnings, which we state below.

Theorem 1.5 (Special case of Lemma 6.8 and Remark 6.9). *Let $i_1, \ldots, i_N$ be a random permutation of $[N]$, let $S_t := \{i_1, \ldots, i_t\}$, and let $x \sim \mu_H$. Suppose with probability $1 - \varepsilon$ over the randomness of x and the permutation $i_1, \ldots, i_N$, we have that for every $t \in [N]$, the influence matrix $\Psi_{S_t, x_{S_t}}$ has spectral norm bounded by α. Then, P_H satisfies a $(n^{-O(\alpha)}, O(\varepsilon))$-weak Poincaré inequality.*

Remark 1.6. The reader should think of the spectral norm of $\Psi_{S_t, x_{S_t}}$ as quantifying how much variance of the distribution $\mu_H | x_{S_t}$ is "explained" by revealing $x_{i_{t+1}}$.

Remark 1.7. Theorem 1.5 holds at a more general level, for a large family of *localization schemes*; see [CE22] for examples of localization schemes and further discussion. The localization scheme at play in the above local-to-global principle is process of revealing coordinates of a Gibbs sample x in random order.

In our main application of sampling from a spherical spin glass using simulated annealing of Langevin diffusion, we consider a different localization scheme, *stochastic localization*, where the revealed information at time t is $y_t = tx + B_t$ where $(B_t)_{t \geqslant 0}$ is a standard Brownian motion. Analyzing this localization scheme requires studying *exponential tilts* rather than pinnings of μ_H. The analogous local-to-global principle in this setting is:

> *Bounded covariance over **typical** exponential tilts implies a **weak** Poincaré inequality.*

We defer a technical discussion to Section 2, and refer to Lemma 6.8 and Remark 6.9 for a formal statement.

1.2 Sampling from spherical spin glasses

We now state our main results for sampling from spherical spin glasses. We will encode the coefficients $\gamma_2, \ldots, \gamma_{p_*}$ in (1) into the *mixture function* $\xi(q) = \sum_{p=2}^{p_*} \gamma_p^2 q^p$. Note that the parameter β in (1) can of course be absorbed into the γ_p, so we can state thresholds directly in terms of the function ξ. Physics heuristics [CHS93] suggest that Glauber dynamics and Langevin diffusion, with random initialization, sample from μ_H with vanishing total variation error under the following condition. Note that this and the below conditions take the form of an upper bound on ξ or its derivatives, and therefore demarcate a region of sufficiently high temperature.

$$\xi'(q) < \frac{q}{1-q} \text{ for all } q \in (0,1). \tag{Non-shattering}$$

Recent work by one of the authors, Montanari, and Pham [HMP24] gives an algorithm based on simulating Eldan's stochastic localization process [Eld13, Eld20] (see below), which samples from μ_H with vanishing total variation error under the following condition.

$$\xi''(q) < \frac{1}{(1-q)^2} \text{ for all } q \in [0,1). \tag{SL}$$

Note that this condition implies (Non-shattering), which can be seen by integrating the inequality. [HMP24] also shows a matching hardness result, that for any strictly replica symmetric model (see (Strict RS) below) *not* satisfying (SL), a generalized family of stochastic localization algorithms fails to sample from μ_H.

Our main result is that simulated annealing samples from μ_H in the same regime.

Theorem 1.8 (See Theorem 7.2). *Under* (SL), *with probability at least* $1 - e^{-\Omega(N^{1/5})}$ *over the randomness of H, annealed Langevin dynamics produces a sample whose total variation distance to* μ_H *is at most* $e^{-\Omega(N^{1/5})}$.

As alluded to in the above discussion, the main input to our framework is a high-probability covariance bound on the *random* exponential tilts of the Gibbs measure encountered along the stochastic localization process. Combined with our weak Poincaré inequality framework, this

5

implies that simulated annealing samples from the Gibbs measure. On the way to proving these covariance bounds, we establish a high-probability covariance bound on all spherical spin glasses in the *(strictly) replica symmetric* phase, a high-temperature phase where the model enjoys a certain notion of correlation decay.

$$\xi''(0) < 1 \text{ and } \xi(q) + q + \log(1 - q) < 0 \text{ for all } q \in (0, 1). \tag{Strict RS}$$

Theorem 1.9 (Informal, see Theorem 7.29)**.** *Under* (Strict RS), *with probability* $1 - e^{-\Omega(N^{1/5})}$ *over the randomness of H, $\|\mathbf{Cov}(\mu_H)\|_{\mathsf{op}} = O(1)$.*

This is the first covariance bound to cover the entire replica symmetric phase with higher order interactions, and we believe it is interesting in its own right. This result is sharp: in the complement of the replica symmetric regime, arguing as in [AG24, Proposition 4.2] shows that $\mathbf{E}\|\mathbf{Cov}(\mu_H)\|_{\mathsf{op}}$ is diverging, of order $\Omega(\sqrt{N})$.

The relation between (SL) and (Strict RS) is as follows. First, (Strict RS) follows from (SL) by integrating twice. Second, (SL) is equivalent to the condition that random exponential tilts of μ_H of any magnitude are typically replica symmetric. This is needed for the algorithmic stochastic localization approach of [HMP24], and arises in the current work (where stochastic localization appears as an analysis tool, rather than as an algorithm) for a similar reason, see Remark 7.4.

The connection from Theorem 1.9 to high-probability covariance bounds on the tilted measures encountered along the localization process relies on a reduction developed in [HMP24]. This reduction implies that typically, the vast majority of the mass of these tilted measures live near a certain codimension-2 band passing through a *TAP fixed point*, which behaves like a spin glass in two fewer dimensions. The proof of Theorem 1.9 also builds on tools developed in [HMP24], and by one of the authors and Sellke in [HS23], which together provide high-precision control of partition functions in the replica symmetric regime.

On the other hand, our approach also leads to several improvements over earlier results. First, we obtain a sampler with total variation error $e^{-\Omega(N^{1/5})}$ with probability $1 - e^{-\Omega(N^{1/5})}$, whereas [HMP24] obtains total variation error $N^{-\varepsilon}$ with probability $1 - N^{-\varepsilon}$, for small constant ε. Our total variation error is close to the best possible, as beyond the uniqueness threshold, at least a $e^{-O(N)}$ fraction of μ_H is typically trapped in metastable states [AJ24], which are hard to reach. Moreover, there is no longer a need to encode a mean estimator for the stochastic localization process (see below) directly in the algorithm; running a natural Markov chain is sufficient.

More conceptually, our work gives the first analysis of a Markov chain for this problem that "sees" the benignness of a random initialization and overcomes the uniqueness threshold.

1.3 Weak Poincaré inequalities beyond annealing

The discussion thus far has been focused on proving mixing time bounds for Markov chains initialized at warm starts. In fact, our framework extends beyond this and can be used to prove rapid mixing of a Markov chain initialized at a distribution that "sees" the different components of the target distribution. For instance, consider the simple scenario where the target distribution π is a mixture of two disconnected component distributions, each of which satisfies a (true) Poincaré inequality. The disconnectedness means that the full distribution π does not satisfy a true Poincaré inequality. However, if we initialize at a distribution that splits its mass equally between the two components, we would expect a Markov chain to rapidly mix to the target distribution.

How does one convert this belief to a (generalizable) proof? The key is that while the distribution may not satisfy a Poincaré inequality for *all* functions, a variant of such an inequality does hold for functions encountered along the trajectory of the Markov chain. More concretely, we may prove the following theorem.

Theorem 1.10 (Informal, see Theorem 4.6). *Consider the trajectory $(\nu_t)_{t \geqslant 0}$ of a Markov chain with stationary distribution π, initialized at a distribution ν_0. Suppose that for all $s \leqslant T$,*

$$\mathcal{E}\left(\frac{\mathrm{d}\nu_s}{\mathrm{d}\pi}, \frac{\mathrm{d}\nu_s}{\mathrm{d}\pi}\right) \geqslant \rho_{\mathrm{PI}}\left(\mathbf{Var}_\pi\left[\frac{\mathrm{d}\nu_s}{\mathrm{d}\pi}\right] - \delta\right).$$

Then,

$$\chi^2(\nu_T \| \pi) \leqslant e^{-2\rho_{\mathrm{PI}}T} \chi^2(\nu_0 \| \pi) + \delta.$$

We remark that our earlier equation (2) is a near-immediate consequence of the above. Returning to the above example with two disconnected components, if ν_s placed exactly half its mass on each of the two components, the error δ can be taken to be 0.

For our first application in Section 5, we use this picture of how the initialization can capture "symmetries" in the distribution.

Sampling from mixtures of log-concave distributions with advice. An example of a distribution where we can take advantage of "symmetries" is the following. Suppose we have a distribution π which is a mixture of K distributions

$$\pi = \sum_{i=1}^{K} p_i \pi_i,$$

each of which is well-connected (e.g., satisfies a Poincaré inequality). We do not expect a Markov chain to rapidly mix to π from a worst-case initialization. Does the scenario change if we initialize more cleverly? To be concrete, suppose we are given m samples $x_1, \ldots, x_m$ from π, and initialize our Markov chain at the empirical distribution $\sum_{i=1}^{m} \delta_{x_i}$. If the component measures (π_i) are "far apart" and do not interact with each other, we would expect the Markov chain to rapidly mix from this initialization if the fraction of points in each cluster is (approximately) equal to the correct fraction p_i. On the other hand, if the component measures were very close together, we would expect their mixture to also satisfy a Poincaré inequality.

However, it is unclear how to translate this intuition to a proof. In previous work [KV23], sampling guarantees are provided for this algorithm, but the running time has a doubly exponential dependence on the number of components K. Our second illustration of weak Poincaré inequalities provides high-probability sampling guarantees for this problem, by running Langevin diffusion for time that is polynomial in all parameters involved. We refer the reader to Section 5 for the details of the theorem statement and its (self-contained) proof.

This problem is studied extensively in an independent work of Koehler, Lee, and Vuong [KLV24]. Motivated by the success of *score matching* methods in modern machine learning, they prove that Langevin dynamics and Glauber dynamics converge to the stationary distribution when initialized from the above empirical distribution under similar conditions to our setting, even if the Markov chain updates come from a slightly perturbed distribution (i.e. if they were learned by a score matching algorithm). They also use their techniques to give an efficient algorithm for learning approximately low-rank Ising models.

7

1.4 Related work

Markov chain mixing and localization schemes. The first use of the local-to-global phenomenon in mixing was in the work of [ALGV24] on establishing rapid mixing of the "basis exchange" walk on bases of a matroid, which used the local-to-global theorem for simplicial complexes from [KO20]. Their approach was later formalized into the framework of *spectral independence* [ALOG21], which was widely successful in resolving numerous problems in algorithmic sampling and counting; see [Liu23] for a comprehensive literature survey.

In the world of sampling from continuous distributions, most recent progress on the KLS conjecture on the Poincaré constant of isotropic log-concave distributions (see [LV24] and the recent survey [KL24]) has employed Eldan's stochastic localization [Eld13]. Later, stochastic localization was used in the work of Eldan, Koehler, and Zeitouni [EKZ22] to analyze the Poincaré constant for Glauber dynamics on Ising models. The seemingly unrelated techniques of spectral independence and stochastic localization approaches to analyzing mixing times were unified under the framework of localization schemes [CE22], which, as an application, also simplified the proof of [EKZ22].

Weak Poincaré inequalities. The study of weak Poincaré inequalities was initiated in the work of Aida [Aid98] and Mathieu [Mat06] in the context of proving other functional inequalities. The work of Röckner and Wang [RW01] observed the connection between a Markov chain satisfying a weak Poincaré inequality, and rapid mixing from "sufficiently warm starts". We refer the reader to the monograph of Wang [Wan06, Chapter 4] for a comprehensive treatment of weak Poincaré inequalities and their implications to mixing and concentration.

Weak Poincaré inequalities are also related to the notion of s-conductance, a weakened version of conductance introduced in [LS93] which has been used frequently in the literature on sampling from convex bodies (see [Che23, Section 7.4.2] for a textbook treatment). This connection is explained in [GMT06]. We also refer the reader to [CGG07], which defines a notion of weak log-Sobolev inequality and uses it to derive a rapid mixing result.

The work [AEGP23] gives a sampling algorithm for the ferromagnetic random-field Ising model on a finite domain $D \subseteq \mathbb{Z}^d$, which follows an approach of chaining warm starts similar to the present work, inspired by convex body sampling literature [LS93]. [AEGP23] shows that in a certain parameter regime, the Glauber dynamics for this model satisfy a weak Poincaré inequality. They then construct an increasing sequence of sub-domains $D_0 \subset D_1 \subset \cdots \subset D_T = D$ and show that a sample from the model on D_i can be converted to a warm start for the model on D_{i+1}. Since the weak Poincaré inequality implies mixing from a warm start, this yields a sampling algorithm based on running the Glauber dynamics on this increasing sequence of models.

The work [AJK+21] introduces a related notion of restricted modified log-Sobolev inequality, which implies entropy contraction (without an additive error, in contrast to a weak Poincaré inequality) for all warm starts. This is used to derive optimal mixing times for several Markov chains.

Sampling from random initializations. The separation between worst-case mixing times and mixing from a random initialization has been studied in a variety of other settings. [CDL+12, BGZ24] characterize which product measure initializations enjoy rapid mixing in a temperature range where worst-case mixing is exponential for the Curie-Weiss Potts model. Notably, as discussed in [BGZ24, Section 1.3], their analysis also characterizes mixing from initializations constructed by

simulated annealing. [LS16, LS17] show that a uniform initialization halves the mixing time for Glauber dynamics for the ferromagnetic Ising model on bounded degree graphs, such as the 1D lattice. [GS22] introduces the notion of weak spatial mixing in a phase, and proves that Glauber dynamics for the ferromagnetic Ising model on the 2D lattice has rapid mixing when initialized uniformly at $\pm\vec{1}$. [GS24] uses the same notion to study mixing from a similar random initialization for a certain natural Markov chain for the random cluster model. [BNN24] show rapid mixing for Glauber dynamics for the exponential random graph model when initialized from a carefully chosen Erdős–Rényi random graph.

Sampling from spherical spin glasses and algorithmic stochastic localization. There is a long history of work studying Markov chain dynamics on spin glasses. An important line of work [CHS93, CK93, BCKM98, BADG06, BAGJ20, CCM21, Sel24b] studies the Langevin dynamics for spherical spin glasses on an N-independent time scale. While the Langevin dynamics do not mix on this time scale, these works capture important statistics of the trajectory such as the energy attained by the Langevin dynamics after a given time, and uncover deep phenomena such as *aging*.

Rapid mixing guarantees at sufficiently high temperature were obtained in [GJ19] for the Langevin dynamics for spherical spin glasses, and in [BB19, EKZ22, AJK$^+$22, ABXY24, AJK$^+$24, AKV24] for the Glauber dynamics for the Sherrington–Kirkpatrick model [SK75] and Ising spin glasses. These approaches show mixing from a worst-case initialization via a functional inequality.

Recently, [AMS22, AMS23a] introduced a new sampling algorithm based on simulating Eldan's stochastic localization scheme [Eld13, Eld20]. This approach has since been used in applications such as Bayesian posterior sampling [MW23, MW24], and is closely related to the denoising diffusions method in machine learning [SDWMG15, HJA20, SSDK$^+$21] (see [Mon23] for details). The resulting algorithm samples in a wider range of temperatures, though with the weaker guarantee of vanishing *Wasserstein* rather than total variation error. The recent work [HMP24] improved this guarantee to total variation, and the resulting algorithm succeeds to the same threshold (SL) as in the present work.

Within the algorithmic stochastic localization approach, the main task is to estimate the means of a sequence of exponential tilts of the Gibbs measure, which appear as the drift process of a stochastic differential equation parametrizing the localization process. In [AMS22], this is achieved with an estimator based on approximate message passing (AMP), which is accurate to leading order. [HMP24] develops an improved estimator with a suitable correction term, which improves the algorithm's guarantee from Wasserstein to total variation error.

Covariance bounds for spin glasses. There has been a great deal of recent work on covariance bounds for spin glasses [BXY23, AG24, BSXY24], in part due to the connection between covariance bounds and functional inequalities developed in the localization schemes literature. In particular, [AG24, BSXY24] address the case of the Sherrington–Kirkpatrick (SK) model, and [BXY23] addresses the SK model with external field.

1.5 Open problems

We conclude with several open problems.

Non-sampling guarantees for simulated annealing. While we initiate a study of simulated annealing to attain sampling guarantees, one could ask how to analyze simulated annealing beyond

sampling. In recent work [LMR+24], three of the authors, Liu, and Raghavendra introduce the framework of *locally stationary distributions* to analyze slow-mixing Markov chains, and leverage it to obtain recovery guarantees for the spiked Wigner and stochastic block model inference problems. We start by reiterating [LMR+24, Problems 1.20 and 1.21] — is simulated annealing computationally optimal for random CSPs with planted solutions?

Further, consider the problem of optimizing the Hamiltonian (1) of the mixed p-spin model. Historically, simulated annealing was one of the earliest algorithms developed for this problem [CHKW23]. The works [Mon21, Sub21, AMS21, Sel24a] develop algorithms that are optimal among suitably Lipschitz algorithms [HS22] and conjecturally among all efficient algorithms. The limiting energy obtained by natural Markov chain dynamics is an long-standing question in its own right [CK93], which was solved for pure models in [Sel24b] but is otherwise open. We ask:

Problem 1.11. What energy does simulated annealing obtain when run on the Hamiltonian (1)?

We refer the reader to [MRT04, FFRT21] and references therein for relevant discussion. We also ask the following question, which seems instrumental to making progress towards the above.

Problem 1.12. How does a non-worst-case initialization (such as one constructed by simulated annealing) affect the locally stationary distribution that is reached by a Markov chain?

Along similar lines, we have the following concrete question about understanding Markov chains from non-worst-case initializations.

Worst-case combinatorial optimization via simulated annealing. The paradigm of solving a semidefinite program and rounding its solution has been extremely successful at achieving optimal approximation guarantees for a wide variety of combinatorial optimization problems, especially constraint satisfaction problems [KKMO07, Rag08].

However, on large families of instances (sparse ones for instance), the solutions produced by these SDPs can be refined locally to improve the approximation ratio, but these improvements do not match the corresponding hardness thresholds. For example, for the problem of Max Cut, the classical SDP algorithm [GW95] gives an α_{GW}-approximation for $\alpha_{GW} \approx 0.878$, and a local refinement [HK22] produces an $\alpha_{GW} + \Omega\left(\frac{1}{d^2}\right)$-approximation. On the other hand, it is (UG-)hard [Tre01] to approximate the max-cut better than $\alpha_{GW} + O\left(\frac{1}{\sqrt{d}}\right)$.

Problem 1.13. Does a Markov chain initialized at the SDP solution attain a $\alpha_{GW} + \Omega\left(\frac{1}{\sqrt{d}}\right)$-approximation to the max-cut in a bounded degree graph?

Sampling from spin glasses up to the shattering threshold. It is conjectured that the Langevin diffusion with uniform random initialization samples from spherical p-spin models for inverse temperatures up to the shattering threshold (Non-shattering) [CHS93, CK93]. Similarly, this is conjectured for the Glauber dynamics Markov chain for models over the hypercube $\{\pm 1\}^N$ instead of the sphere S_N, for an analogous shattering threshold. As a start, can we show such guarantees for simulated annealing (as opposed to a fixed-temperature Markov chain from uniform initialization)?

Problem 1.14. Does simulated annealing sample from p-spin models up to the shattering threshold?

The failure of algorithmic stochastic localization beyond the (SL) condition [HMP24, Section 10] suggests that ideas beyond our proof strategy are required to prove the above.

Simulated annealing in more general models. For sampling from the spherical p-spin model, our results show that simulated annealing succeeds in the regime (SL) where algorithmic stochastic localization succeeds. At the level of proofs, these methods are also closely related, as both revolve around suitable control of the localization process: in the algorithmic stochastic localization approach, this is used to construct a mean estimator for the localized measures, and in our approach it is used to bound the localized measures' covariances. These tasks are closely linked; see Remark 7.4.

One question is whether simulated annealing succeeds in more general models. In particular, samplers based on algorithmic stochastic localization have been developed for the Sherrington-Kirkpatrick model in the replica symmetric regime [AMS22, Cel24], p-spin models over the hypercube [AMS23a], and posteriors of spiked matrix models [MW23]. These samplers are proven to have vanishing Wasserstein error, and sampling with vanishing total variation error remains an open problem in these models. It would be interesting to show that simulated annealing achieves this. More speculatively, we may ask if there is a general reduction from a sampling guarantee for algorithmic stochastic localization to one for simulated annealing.

#BIS. A major open problem in the field of approximate counting is settling the complexity of #BIS: where the algorithmic task is to approximate the number of independent sets in a bipartite graph. So far, algorithmic progress for this problem has been limited to restricted classes of graphs, such as lattices & tori [HPR19], and expander graphs [JKP20]. Numerous interesting approximate counting problems have been shown to be #BIS-hard [CGM12, GJ12, CGG$^+$16, GSVY16]. While vanilla Glauber dynamics fails at the corresponding sampling task, it is plausible that a variant of simulated annealing succeeds.

Problem 1.15. Does (a simple variant of) simulated annealing succeed at sampling a uniformly random independent set in a bipartite graph?

Structural guarantees from weak Poincaré inequalities. According to physics heuristics, the Gibbs measure of a spherical mixed p-spin glass between β_{uniq} and β_{shatter} consists of one main cluster accounting for nearly all the mass, and metastable states with exponentially small mass that are poorly connected to the main cluster and each other. We do not prove this picture, but the weak Poincaré inequality we obtain (up to β_{SL}) is sufficient to imply a sampling guarantee for simulated annealing. One open direction is to show that the above picture holds, and that the main cluster satisfies a genuine Poincaré inequality. More generally, one may ask:

Problem 1.16. If a distribution satisfies a weak Poincaré inequality, is it TV-close to a distribution satisfying a true Poincaré inequality?

We note that Lemmas 4.10 and A.4 show a converse of this statement, that if we perturb a distribution satisfying a true Poincaré inequality (for the Langevin diffusion or Glauber dynamics Markov chains), the resulting distribution satisfies a weak Poincaré inequality.

1.6 Organization

In Section 2, we give a technical overview of how we use weak Poincaré inequalities to analyze simulated annealing for our main application of sampling from spherical p-spin distributions.

In Section 3, we cover some basic preliminaries that will be useful. Then, in Section 4, we formally define weak functional inequalities and establish some of their basic properties.

In Section 5, we demonstrate the effectiveness of this framework by showing how to sample from a mixture of distributions satisfying Poincaré inequalities from data-based initializations.

Our main application to spherical p-spin models spans Sections 6 to 8, and requires more background in stochastic localization and spin glass theory. In Section 6, we review some basic properties of stochastic localization and show how to adapt the framework of localization schemes from [CE22] to prove weak functional inequalities. Then, in Section 7, we initiate the discussion of weak Poincaré inequalities for spherical p-spin models. To assist the reader in understanding the proof of a weak Poincaré inequality, we provide a separate technical overview in Subsection 7.1. The rest of Section 7 reduces the proof to proving high-probability covariance bounds for strictly replica-symmetric models with small external field, which is then established in Section 8.

2 Technical overview

Let H be a Hamiltonian on state space Ω, and let μ_H be its Gibbs distribution. Our goal in this section is to describe our strategy to prove that simulated annealing succeeds at sampling. In our application, Ω is the scaled sphere $S_N := \sqrt{N} \cdot \mathbb{S}^{N-1}$, and μ_H comes with an associated Markov chain known as *Langevin diffusion*, which we denote with P_H. For ease of exposition, we restrict the discussion to this setting, though much of it holds in a more general setting.

Definition 2.1 (Simulated annealing, informal). Initialize at the uniform distribution on S_N (which is equal to μ_0), and for each $i \in [m]$, run $P_{\frac{i}{m}H}$ for time T.

For the sequel, we abbreviate $P_{\frac{i}{m}H}$ and $\mu_{\frac{i}{m}H}$ as P_i and μ_i, and we use $P_{i,t}$ to denote running P_i for time t. The strategy to prove that simulated annealing succeeds at sampling is to establish a weak Poincaré inequality for P_i for all i.

Let L be the infinitesimal generator of P_i. For functions $f, g : \Omega \to \mathbb{R}$, we define the *Dirichlet form* $\mathcal{E}(f, g)$ as $\mathbf{E}_{\mu_i}[f \, L \, g]$.

Remark 2.2. In the case of Langevin diffusion for a distribution π, the Dirichlet form can be evaluated as

$$\mathcal{E}(f, g) = \mathbf{E}_{\mu_i} \langle \nabla f, \nabla g \rangle,$$

where ∇ denotes the Euclidean gradient if π is supported on $\mathbb{R}^N$, and the Riemannian gradient on S_N if π is supported on S_N.

As discussed in Section 1, we say a Markov chain satisfies a *weak Poincaré inequality* with parameters (C, ε) if

$$\mathcal{E}(f, f) \geqslant C \cdot \mathbf{Var}[f] - \varepsilon \cdot \|f - \mathbf{E}f\|_\infty^2 - \varepsilon \cdot \sup_{x \in \Omega} \|\nabla f(x)\|^2,$$

12

which implies the following mixing result Theorem 4.6 for the chi-squared divergence; see also [RW01, Theorem 2.1]. Defining ν_t as the distribution after running the Markov chain for time t from initial distribution ν_0, we have

$$\chi^2(\nu_t \| \mu_i) \leqslant \exp(-Ct) \cdot \chi^2(\nu_0 \| \mu_i) + \varepsilon \cdot \left\| \frac{\mathrm{d}\nu_0}{\mathrm{d}\mu_i} - 1 \right\|_\infty^2 + \varepsilon \cdot \sup_{x \in \Omega} \left\| \boldsymbol{\nabla} \frac{\mathrm{d}\nu_0}{\mathrm{d}\mu_i}(x) \right\|^2.$$

Analyzing simulated annealing with weak Poincaré inequalities. To see why the above statement plays well with simulated annealing, imagine plugging in initialization $\nu_0 = \mu_{i-1}$. By selecting the number of annealing steps $m = \mathrm{poly}(N)$, we can ensure $\left\| \frac{\mathrm{d}\nu_0}{\mathrm{d}\mu_i} - 1 \right\|_\infty$ and $\sup_{x \in \Omega} \left\| \boldsymbol{\nabla} \frac{\mathrm{d}\nu_0}{\mathrm{d}\mu_i}(x) \right\|$ are $O(1)$. The guarantee after running the Markov chain for some sufficiently large polynomial time T is then

$$\chi^2(P_{i,T}\mu_{i-1} \| \mu_i) \leqslant O(\varepsilon),$$

which in particular implies

$$d_{\mathrm{TV}}(P_{i,T}\mu_{i-1}, \mu_i) \leqslant O(\sqrt{\varepsilon}).$$

When we combine the above with the data processing inequality, we then get the following guarantee for $\nu_{m,T} := P_{m,T} \cdots P_{2,T} P_{1,T} \nu_0$, the distribution that simulated annealing samples from.

$$d_{\mathrm{TV}}(P_{m,T} \cdots P_{1,T}\mu_0, \mu_m) \leqslant d_{\mathrm{TV}}(P_{m,T} \cdots P_{1,T}\mu_0, P_{m,T}\mu_{m-1}) + d_{\mathrm{TV}}(P_{m,T}\mu_{m-1}, \mu_m)$$
$$\leqslant d_{\mathrm{TV}}(P_{m-1,T} \cdots \mu_0, \mu_{m-1}) + O(\sqrt{\varepsilon}).$$

Applying the above inequality m times tells us that $d_{\mathrm{TV}}(\nu_{m,T}, \mu_m) \leqslant O(\sqrt{\varepsilon} \cdot m)$.

We now turn our attention to the proof technique for showing a weak Poincaré inequality.

How to prove weak Poincaré inequalities. Suppose our goal is to prove a weak Poincaré inequality for a measure π. The high-level strategy in the localization schemes approach for proving a weak Poincaré inequality is to design a *measure decomposition* of π: for some mixture distribution ρ, express π as $\mathbf{E}_{z \sim \rho} \pi_z$. Refer to Subsection 3.1 for a brief review of measure decompositions. Once we have a measure decomposition in hand, establishing the following simple set of inequalities forms the crux of the argument. Let f be a function such that $\mathbf{E}_\pi f = 1$.

1. *Conservation of Dirichlet form.*

$$\mathcal{E}_\pi(f, f) \geqslant \mathbf{E}_{z \sim \rho} \mathcal{E}_{\pi_z}(f, f).$$

   In the case of Langevin diffusion, this is an equality, and in the case of Glauber dynamics, the inequality is true by a generic concavity argument; see, e.g. [AJK+22, Page 19] or [LMRW24, Page 5].

2. *Weak Poincaré inequality for good component measures.*

$$\mathbf{E}_{z \sim \rho} \mathcal{E}_{\pi_z}(f, f) \geqslant C \cdot \mathbf{E}_{z \sim \rho} \mathbf{Var}_{\pi_z}[f] \cdot \mathbf{1}[z \text{ "good"}] - \varepsilon \|f - 1\|_\infty^2 - \varepsilon \|\boldsymbol{\nabla} f\|_\infty^2,$$

   where the "good" π_z are those which satisfy a (c, ε)-weak Poincaré inequality. This inequality follows from the nonnegativity of norms and Dirichlet forms.

13

3. *Approximate conservation of variance.*

$$\mathbf{E}_{z \sim \rho} \mathbf{Var}_{\pi_z}[f] \geqslant \alpha \cdot \mathbf{Var}_{\pi}[f].$$

This is one of the parts that depends nontrivially on π and the decomposition ρ, and we discuss the general proof strategy for this portion based on localization schemes.

4. *High-probability goodness of component measures.*

$$\mathbf{Pr}_{z \sim \rho}[z \text{ "good"}] \geqslant 1 - \varepsilon.$$

This part also requires analyzing the measure decomposition we design. Ideally, the measure decomposition presents us with "simpler" measures than π itself.

Once we have the above inequalities at hand, we get a $(c\alpha, 2\varepsilon)$-weak Poincaré inequality; see Lemma 4.11 for details.

How to construct a good measure decomposition. Henceforth, we restrict our attention to the case where $\pi = \mu_H$, the Gibbs distribution for a spherical mixed p-spin glass model. In the discussion below, we fix H as a typical Hamiltonian, and drop the phrase "with high probability" for events that occur with high probability over the randomness of H.

To construct our measure decomposition, we rely on Eldan's stochastic localization [Eld13]. Our inspiration is the use of stochastic localization as a tool for measure decomposition for proving Poincaré inequalities in the work of Chen and Eldan [CE22]. Stochastic localization is a measure-valued random process $(\mu_t)_{t \geqslant 0}$ described by:

$$\mathrm{d}\mu_t(x) \propto \exp\left(\langle \boldsymbol{y}_t, x \rangle - \frac{t}{2} \|x\|^2 \right) \mathrm{d}\mu_H(x),$$

where $\boldsymbol{y}_t = \sigma + B_t$ where $\sigma \sim \mu_H$ and $(B_t)_{t \geqslant 0}$ is a standard Brownian motion; see [AM22a, Theorem 2] for a proof of why the above description of stochastic localization is equivalent to the more traditional definition via a stochastic differential equation that μ_t obeys.

We run stochastic localization up to a stopping time τ, defined as

$$\tau := \min\{t : 0 \leqslant t \leqslant T, \|\mathbf{Cov}(\mu_t)\| \geqslant K \text{ or } t = T\},$$

where T is chosen as a sufficiently large constant, independent of N. We impose the constraint on the covariance matrix as it is relevant to satisfying approximate conservation of variance: [CE22, Eq. (20)] proves that a measure decomposition based on stochastic localization run for time at most T satisfies approximate conservation of variance with parameter $\alpha = \exp(-KT)$ if $\|\mathbf{Cov}(\mu_t)\|_{\mathrm{op}}$ is bounded by K almost surely. Hence, by construction, we automatically ensure that our measure decomposition satisfies approximate conservation of variance.

For the measure decomposition to ultimately be useful, we also need to argue that the component measures satisfy a weak Poincaré inequality with high probability. Building on technical results in Huang, Montanari, and Pham [HMP24, Section 9.2], we show that the stochastic localization process run up to time T starting at μ_H gives a distribution satisfying an $(\Omega(1), \exp(-\Omega(n)))$-weak Poincaré inequality with probability $1 - \exp(-\Omega(n))$ over the randomness of the stochastic

localization path; see Lemma 7.7 for details. Unfortunately, in the situation where the stochastic localization process stops before T, we do not have a simple way to show a weak Poincaré inequality, and for our analysis, treat z arising from early stopping as "bad".

Thus, we have: $\mathbf{Pr}_{z\sim\rho}[z \text{ "good"}] \geqslant 1 - \exp(-\Omega(n)) - \mathbf{Pr}[\tau < T]$. To bound $\mathbf{Pr}[\tau < T]$, it is sufficient to prove a high-probability covariance norm bound on the entire stochastic localization path for $0 \leqslant t \leqslant T$. Most of the technical work in this paper is devoted to proving this covariance norm bound.

Theorem 2.3 (Informal version of Lemma 7.6). *For a typical H, with probability $1 - e^{-\Omega(n^{1/5})}$ over the randomness of the stochastic localization path, we have $\|\mathbf{Cov}(\mu_t)\|_{\mathsf{op}} < K$.*

The proof of the covariance bound spans Sections 7 and 8; we give a detailed technical overview of how it is proved in Subsection 7.1.

3 Preliminaries

Notation

- We use S_N to denote the scaled $(N-1)$-sphere, $\sqrt{N} \cdot \mathsf{S}^{N-1}$.

- We use ρ to denote the uniform measure over S_N.

- Given $\sigma_1, \sigma_2 \in S_N$, we use $R(\cdot, \cdot)$ to denote the normalized inner product (i.e. the *overlap*)

$$R(\sigma_1, \sigma_2) := \frac{\langle \sigma_1, \sigma_2 \rangle}{N}.$$

- For an interval $I \subseteq [-1, 1]$ and $x \in S_N$, define $\mathsf{Band}(x, I) := \{\sigma \in S_N : R(\sigma, x) \in I\}$.

- We use c to denote small constants whose values may change from line to line, and C to denote similarly fickle large constants.

- Let $f : \Omega \to \mathbb{R}$ be any function. We define $\mathrm{osc}(f) := \sup f - \inf f$.

- Let $f : \Omega \to \mathbb{R}$ be a smooth function. If $\Omega \subseteq \mathbb{R}^N$, then ∇f denotes its Euclidean gradient. If $\Omega \subseteq S_N$, then $\nabla_{\mathsf{sp}} f$ denotes the Riemannian gradient on S_N. When the correct notion of gradient is clear from context, by an abuse of notation we will suppress this distinction and simply write ∇f.

3.1 Measure decompositions

Our framework for proving weak functional inequalities relies on the notion of a measure decomposition.

Definition 3.1 (Measure decomposition). Let π be a distribution on $\mathbb{R}^N$. Let ρ be a mixture distribution, also on $\mathbb{R}^N$, which indexes into a family of mixture components $\{\pi_z\}_{z\in\mathbb{R}^N}$. We say that (ρ, π_z) is a measure decomposition for π if

$$\pi = \mathbf{E}_{z\sim\rho} \pi_z.$$

One reason measure decompositions are useful is that they compose nicely with worst-case functional inequalities, as shown in the following lemma.

Lemma 3.2 ([BB19, AJK$^+$22, CE22]). *Let π be a distribution over $\Omega \subseteq \mathbb{R}^N$, and $\pi = \mathbf{E}_{z \sim \rho} \pi_z$ a measure decomposition of π such that*

- *for all functions f, $\mathbf{E}_{z \sim \rho} \mathbf{Var}_{\pi_z}[f] \geqslant C_{\mathrm{Var}} \mathbf{Var}_\pi[f]$, and*

- *Every π_z satisfies a ρ_{PI}-Poincaré inequality with respect to Langevin diffusion.*

Then, π satisfies a $\rho_{\mathrm{PI}} C_{\mathrm{Var}}$-Poincaré inequality.

In Lemma 4.11, we will show an average-case relaxation of the above result, that π satisfies a weak Poincaré inequality if *most* measures in the decomposition satisfy weak Poincaré inequalities. Then, in Section 6, we construct explicit measure decompositions using the localization schemes framework introduced in [CE22]. This will show weak Poincaré inequalities for our measures of interest.

Besides proving functional inequalities, measure decompositions have also been directly used for sampling and inference (see, e.g., [MW24, LMR$^+$24]).

3.2   Langevin diffusion

In this paper, we study Langevin diffusion on $\mathbb{R}^N$ and the scaled sphere S_N. These definitions can be directly generalized to the setting of Riemannian manifolds, but we do not comment further on this.

Definition 3.3 (Langevin diffusion on $\mathbb{R}^N$). Let π be a distribution on $\mathbb{R}^N$ with density at x proportional to $e^{-V(x)}$ for some function V. The *Langevin diffusion* process with stationary distribution π is the solution to the stochastic differential equation

$$\mathrm{d}Z_t = -\boldsymbol{\nabla} V(Z_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t,$$

where $(B_t)_{t \geqslant 0}$ is a standard Brownian motion.

Definition 3.4 (Langevin diffusion on S_N). Let π be a distribution on S_N with $\mathrm{d}\pi(x) \propto e^{-V(x)}\,\mathrm{d}\rho(x)$, where $V : S_N \to \mathbb{R}$. The Langevin diffusion process with stationary distribution π is the solution to the stochastic differential equation

$$\mathrm{d}Z_t = -\boldsymbol{\nabla}_{\mathsf{sp}} V(Z_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t,$$

where $(B_t)_{t \geqslant 0}$ is a standard spherical Brownian motion. (For a textbook introduction to spherical Brownian motion, see [Hsu02].)

Fact 3.5 ([Che23, Example 1.2.17]). *The Langevin diffusion SDE with stationary distribution π is reversible with respect to π. In particular, the ergodicity of the process implies that $\mathrm{KL}(\mathrm{Law}(Z_t)\|\pi) \xrightarrow{t \to \infty} 0$.*

Furthermore, it is well-known that Langevin diffusion on $\mathbb{R}^N$ with respect to a strongly log-concave stationary distribution converges rapidly.

Definition 3.6. Let π be a distribution over $\mathbb{R}^N$ with density proportional to e^{-V}. π is said to be *α-strongly log-concave* if V is α-strongly convex, that is, $\boldsymbol{\nabla}^2 V \succeq \alpha I$.

Fact 3.7 ([Che23, Theorem 1.2.24]). *Let π be a distribution satisfying a log-Sobolev inequality with constant ρ_{LS}, in that for any differentiable function $f : \mathbb{R}^N \to \mathbb{R}_{>0}$,*

$$\mathbf{E}_\pi \|\boldsymbol{\nabla}\sqrt{f}\|^2 \geqslant \rho_{\mathrm{LS}}\mathrm{Ent}_\pi[f].$$

Then, if π_t is the distribution at time t of Langevin diffusion,

$$\mathrm{KL}(\pi_t\|\pi) \leqslant \mathrm{KL}(\pi_0\|\pi)e^{-\rho_{\mathrm{LS}}\cdot t}.$$

Furthermore, α-strongly log-concave distributions π satisfy a log-Sobolev inequality with constant α.

4 Weak functional inequalities

In this paper, we study continuous-time Markov chains.

Definition 4.1 (Markov semigroup). Let $(X_t)_{t\geqslant 0}$ denote a continuous-time Markov process on *state space* Ω. Let $(P_t)_{t\geqslant 0}$ be the associated *Markov semigroup* operator; P_t acts on functions $f : \Omega \to \mathbb{R}$ via $P_t f(x) = \mathbf{E}[f(X_t)|X_0 = x]$. Throughout, we assume that the semigroup is *reversible* with respect to stationary distribution π. Furthermore, let L denote the *infinitesimal generator* of P_t, i.e., $P_t = e^{-t\mathrm{L}}$. For functions $f, g : \Omega \to \mathbb{R}$, we define the *Dirichlet form* as $\mathcal{E}(f,g) = \mathbf{E}_\pi[f\mathrm{L}g]$.

See e.g. [Che23, Section 1.2] for a textbook treatment. Of particular interest to us are the two settings where the semigroup corresponds to a discrete-time Markov chain or the Langevin diffusion defined in Subsection 3.2. In these cases, the Dirichlet form satisfies the following explicit identities.

Fact 4.2 (Dirichlet form from discrete-time Markov chain). *Let P be the transition matrix of a reversible discrete-time Markov chain with stationary distribution π. We can define an associated continuous-time semigroup operator $(P_t)_{t\geqslant 0}$ by setting $\mathrm{L} = I - P$. The Dirichlet form for the continuous-time dynamics satisfies*

$$\mathcal{E}(f,g) := \mathbf{E}_{x\sim\pi}\mathbf{E}_{y\sim_P x}(f(x) - f(y))(g(x) - g(y)).$$

Here, for a probability distribution μ, we say $x \sim \mu$ to denote a sample x from μ, and we use $y \sim_P x$ for a single transition from x according to P.

Fact 4.3 (Dirichlet form for Langevin diffusion). *We will need the following explicit identities for the Dirichlet form for Langevin diffusion.*

(1) *When $(P_t)_{t\geqslant 0}$ corresponds to Langevin diffusion on $\mathbb{R}^N$ with stationary distribution π, the Dirichlet form is $\mathcal{E}(f,g) = \mathbf{E}_\pi[\langle\boldsymbol{\nabla}f, \boldsymbol{\nabla}g\rangle]$.*

(2) *When $(P_t)_{t\geqslant 0}$ corresponds to Langevin diffusion on S_N with stationary distribution π, the Dirichlet form is $\mathcal{E}(f,g) = \mathbf{E}_\pi[\langle\boldsymbol{\nabla}_{\mathrm{sp}}f, \boldsymbol{\nabla}_{\mathrm{sp}}g\rangle]$.*

Definition 4.4. We say π satisfies a *weak Poincaré inequality* if for some error functional Error : $\mathbb{R}_{>0}^{\Omega} \to \mathbb{R}_{\geqslant 0}$ and $\rho_{\mathrm{PI}} > 0$,

$$\mathbf{Var}_{\pi}[f] \leqslant \frac{1}{\rho_{\mathrm{PI}}} \cdot \mathcal{E}(f, f) + \mathrm{Error}(f) \,.$$

Similarly, we say π satisfies a *weak modified log-Sobolev inequality* if for some error functional Error : $\mathbb{R}_{>0}^{\Omega} \to \mathbb{R}_{\geqslant 0}$ and $\rho_{\mathrm{LS}} \geqslant 0$,

$$\mathrm{Ent}_{\pi}[f] \leqslant \frac{1}{\rho_{\mathrm{LS}}} \cdot \mathcal{E}(f, \log f) + \mathrm{Error}(f).$$

Theorem 4.5. *Consider the trajectory $(\nu_t)_{t \geqslant 0}$ of a reversible continuous-time Markov chain with stationary distribution π, initialized at the distribution ν_0, and suppose that π satisfies a weak MLSI with parameters* Error *and ρ_{LS}. Fix $T > 0$, and set Λ_T to be the distribution on $[0, T]$ with density $\Lambda_T(s) = \frac{e^{\rho_{\mathrm{LS}}}}{e^{\rho_{\mathrm{LS}}T} - 1} \cdot e^{\rho_{\mathrm{LS}}s}$. Then,*

$$\mathrm{KL}(\nu_T \| \pi) \leqslant e^{-\rho_{\mathrm{LS}}T} \mathrm{KL}(\nu_0 \| \pi) + \mathbf{E}_{s \sim \Lambda_T}[\mathrm{Error}(\tfrac{\mathrm{d}\nu_s}{\mathrm{d}\pi})].$$

Proof. Let $f_0 = \frac{\mathrm{d}\nu_0}{\mathrm{d}pi}$, and let $f_t = P_t f_0 = \frac{\mathrm{d}\nu_t}{\mathrm{d}\pi}$ (this last equality holds due to reversibility). For ease of notation, set $\mathrm{Error}_t = \mathrm{Error}(f_t)$ for $t \geqslant 0$. Recalling that $\mathcal{E}(f_t, \log f_t) = -\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{KL}(\nu_t \| \pi)$, the weak MLSI says that

$$-\mathcal{E}(f_t, \log f_t) + \rho_{\mathrm{LS}} \cdot \mathrm{KL}(\nu_t \| \pi) - \rho_{\mathrm{LS}} \cdot \mathrm{Error}_t \leqslant 0,$$

so

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(e^{\rho_{\mathrm{LS}}t} \cdot \mathrm{KL}(\nu_t \| \pi) - \rho_{\mathrm{LS}} \int_0^t e^{\rho_{\mathrm{LS}}s} \mathrm{Error}_s \, \mathrm{d}s \right) \leqslant 0.$$

Therefore,

$$e^{\rho_{\mathrm{LS}}T} \cdot \mathrm{KL}(\nu_T \| \pi) - \rho_{\mathrm{LS}} \int_0^T e^{\rho_{\mathrm{LS}}s} \mathrm{Error}_s \, \mathrm{d}s \leqslant \mathrm{KL}(\nu_0 \| \pi),$$

and

$$\mathrm{KL}(\nu_T \| \pi) \leqslant e^{-\rho_{\mathrm{LS}}T} \mathrm{KL}(\nu_0 \| \pi) + \rho_{\mathrm{LS}} \int_0^T e^{\rho_{\mathrm{LS}}(s-T)} \mathrm{Error}_s.$$

Noting that $\Lambda_T(s) = \frac{\rho_{\mathrm{LS}}}{e^{\rho_{\mathrm{LS}}T} - 1} \cdot e^{\rho_{\mathrm{LS}}s} \geqslant \rho_{\mathrm{LS}} e^{\rho_{\mathrm{LS}}(s-T)}$, the above implies that

$$\mathrm{KL}(\nu_T \| \pi) \leqslant e^{-\rho_{\mathrm{LS}}T} \mathrm{KL}(\nu_0 \| \pi) + \mathbf{E}_{s \sim \Lambda_T}[\mathrm{Error}_s],$$

as desired. $\qquad\square$

By essentially the same proof, we obtain the analogous result for weak Poincaré inequalities.

Theorem 4.6. *Consider the trajectory $(\nu_t)_{t \geqslant 0}$ of a (continuous-time) Markov chain with stationary distribution π, initialized at the distribution ν_0, and suppose that π satisfies a weak Poincaré inequality with parameters* Error *and ρ_{PI}. Fix $T > 0$, and set Λ_T to be the distribution on $[0, T]$ with density $\Lambda_T(s) = \frac{e^{2\rho_{\mathrm{PI}}}}{e^{2\rho_{\mathrm{PI}}T} - 1} \cdot e^{2\rho_{\mathrm{PI}}s}$. Then,*

$$\chi^2(\nu_T \| \pi) \leqslant e^{-2\rho_{\mathrm{PI}}T} \chi^2(\nu_0 \| \pi) + \mathbf{E}_{s \sim \Lambda_T}[\mathrm{Error}(\tfrac{\mathrm{d}\nu_s}{\mathrm{d}\pi})].$$

For the analysis of the annealed Langevin dynamics, we will also require the following definition. For $f : \Omega \to \mathbb{R}$, let $\mathrm{osc}(f) := \sup(f) - \inf(f)$, and let ∇f denote the Riemannian gradient.

18

Definition 4.7 (Weak functional inequalities for Langevin). We say a distribution π on $\Omega \subseteq \mathbb{R}^N$ or $\Omega \subseteq S_N$ satisfies a $(\rho_{\text{PI}}, \varepsilon)$-weak Poincaré inequality if for all differentiable functions f,

$$\mathbf{Var}_\pi[f] \leqslant \frac{1}{\rho_{\text{PI}}} \cdot \mathcal{E}(f, f) + \varepsilon \cdot \left(\text{osc}(f)^2 + \sup_{x \in \Omega} \| \boldsymbol{\nabla} f \|^2 \right).$$

Similarly, we say π satisfies a $(\rho_{\text{LS}}, \varepsilon)$-weak modified log-Sobolev inequality if for all differentiable functions f,

$$\text{Ent}_\pi[f] \leqslant \frac{1}{\rho_{\text{LS}}} \cdot \mathcal{E}(f, \log f) + \varepsilon \cdot \left(\text{osc}(\sqrt{f})^2 + \sup_{x \in \Omega} \| \boldsymbol{\nabla} f \|^2 \right).$$

Remark 4.8. As mentioned in the beginning of this section, by replacing the Riemannian gradient with the discrete gradient, an analogous theory can be developed for annealed Glauber dynamics; see Definition A.1.

Remark 4.9. A weak Poincaré inequality with sufficiently good parameters implies a true Poincaré inequality. Indeed, any low conductance cut limits on the region of valid $(\rho_{\text{PI}}, \varepsilon)$. Hence, by Cheeger, one can conclude that Langevin satisfies a true Poincaré inequality, with some loss in parameters.

We shall typically use weak Poincaré inequalities with functions f that have expectation 1, where we bound $\text{osc}(f) \leqslant 2\|f - 1\|_\infty$.

4.1 Properties of weak functional inequalities

In this section, we state some crucial properties of weak functional inequalities for Langevin diffusion on $\mathbb{R}^N$ or S_N. With minor modifications, the same results hold for Glauber dynamics on finite state spaces; see Appendix A for formal details.

Lemma 4.10. *Let π be a distribution on $\mathbb{R}^N$ or S_N satisfying a ρ_{PI}-Poincaré inequality for Langevin diffusion, and π' a distribution such that $d_{\text{TV}}(\pi, \pi') \leqslant \delta$. Then, π' satisfies a $\left(\rho_{\text{PI}}, \delta \max(\rho_{\text{PI}}^{-1}, 1) \right)$-weak Poincaré inequality for Langevin diffusion.*

Proof. There exists a coupling $\mathcal{C}$ of (π, π') such that for $(x, x') \sim \mathcal{C}$, $\mathbf{Pr}[x \neq x'] \leqslant \delta$. Thus,

$$\begin{aligned} \mathcal{E}_{\pi'}(f, f) = \mathbf{E}_{\pi'} \| \boldsymbol{\nabla} f \|^2 \\ \geqslant \mathbf{E}_\pi \| \boldsymbol{\nabla} f \|^2 - \delta \sup \| \boldsymbol{\nabla} f \|^2 \\ \geqslant \rho_{\text{PI}} \mathbf{Var}_\pi[f] - \delta \sup \| \boldsymbol{\nabla} f \|^2. \end{aligned}$$

Let $I = [\inf f, \sup f]$. Note that $\mathbf{Var}_\pi[f] = \inf_{a \in I} \mathbf{E}_\pi[(f - a)^2]$. For each $a \in I$,

$$\mathbf{E}_\pi[(f - a)^2] \geqslant \mathbf{E}_{\pi'}[(f - a)^2] - \delta \cdot \text{osc}(f)^2,$$

and therefore

$$\mathbf{Var}_\pi[f] \geqslant \mathbf{Var}_{\pi'}[f] - \delta \cdot \text{osc}(f)^2. \tag{3}$$

Combining with the above shows

$$\mathcal{E}_{\pi'}(f, f) \geqslant \rho_{\text{PI}} \mathbf{Var}_{\pi'}[f] - \delta \left(\rho_{\text{PI}} \cdot \text{osc}(f)^2 + \sup \| \boldsymbol{\nabla} f \|^2 \right). \qquad \square$$

As foreshadowed previously, measure decompositions compose well with weak functional inequalities. Indeed, the following lemma can be viewed as a relaxation of the setup to prove genuine functional inequalities (cf. Lemma 3.2).

Lemma 4.11. *Let π be a distribution over $\mathbb{R}^N$ or S_N, and $\pi = \mathbf{E}_{z \sim \rho} \pi_z$ a measure decomposition of π such that*

- *for all functions f, $\mathbf{E}_{z \sim \rho} \mathbf{Var}_{\pi_z}[f] \geqslant C_{\mathrm{Var}} \mathbf{Var}_\pi[f]$, and*

- *with probability $1 - \eta$ over $z \sim \rho$, π_z satisfies a $(\rho_{\mathrm{PI}}, \delta)$-weak Poincaré inequality with respect to Langevin diffusion.*

Then, π satisfies a $\left(\rho_{\mathrm{PI}} C_{\mathrm{Var}}, \frac{\delta + \eta}{C_{\mathrm{Var}}} \right)$-weak Poincaré inequality.

Proof. Let us say that z is *good* if π_z satisfies a weak Poincaré inequality, and f be a function. Then,

$$
\begin{aligned}
\mathcal{E}_\pi(f, f) &= \mathbf{E}_{z \sim \rho} \mathcal{E}_{\pi_z}(f, f) \\
&\geqslant \mathbf{E}_{z \sim \rho} \mathcal{E}_{\pi_z}(f, f) \mathbf{1}_{z \text{ is good}} \\
&\geqslant \mathbf{E}_{z \sim \rho} \rho_{\mathrm{PI}} \mathbf{Var}_{\pi_z}[f] \mathbf{1}_{z \text{ is good}} - \delta \rho_{\mathrm{PI}} \cdot (\mathrm{osc}(f)^2 + \sup \|\nabla f\|^2) \\
&= \mathbf{E}_{z \sim \rho} \rho_{\mathrm{PI}} \mathbf{Var}_{\pi_z}[f] - \delta \rho_{\mathrm{PI}} \cdot (\mathrm{osc}(f)^2 + \sup \|\nabla f\|^2) - \mathbf{E}_{z \sim \rho} \rho_{\mathrm{PI}} \mathbf{Var}_{\pi_z}[f] \mathbf{1}_{z \text{ is not good}} \\
&\geqslant \rho_{\mathrm{PI}} \mathbf{E}_{z \sim \rho} \mathbf{Var}_{\pi_z}[f] - (\delta \rho_{\mathrm{PI}} + \eta \rho_{\mathrm{PI}}) \cdot (\mathrm{osc}(f)^2 + \sup \|\nabla f\|^2) \\
&\geqslant C_{\mathrm{Var}} \rho_{\mathrm{PI}} \mathbf{Var}_\pi[f] - (\delta \rho_{\mathrm{PI}} + \eta \rho_{\mathrm{PI}}) \cdot (\mathrm{osc}(f)^2 + \sup \|\nabla f\|^2) .
\end{aligned}
$$

The desired follows. $\qquad \square$

4.2 Weak Poincaré inequalities and annealed Markov chains

The notion of weak functional inequalities defined in Definition 4.7 can be naturally applied in the context of simulated annealing, which we now define.

Definition 4.12 (Annealing scheme). Let H be a Hamiltonian over Ω, and $(\mu_\beta)_{\beta \geqslant 0}$ the class of distributions over Ω with $\mu_\beta(\sigma) \propto e^{\beta H(\sigma)}$. For each $\beta \leqslant \beta_0$, let $P = P_\beta$ be a (reversible and ergodic) Markov chain with stationary distribution μ_β.

An *(inverse) temperature schedule* is any function $\beta : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$. An *annealing scheme* $\mathcal{A}$ is the time-inhomogeneous Markov chain such that at time t, one applies the Markov chain $P_{\beta(t)}$.

Of interest is the temperature schedule of the form $t \mapsto \delta \cdot \lfloor \frac{t}{T} \rfloor$, with the chain being run for time $T \cdot \left(\frac{\beta_0}{\delta} + 1 \right)$.

Theorem 4.13. *Let $T, \delta > 0$ such that $k_0 := \frac{\beta_0}{\delta}$ is an integer. Suppose that for each $\beta = k\delta$ for $0 \leqslant k \leqslant k_0$, μ_β satisfies a $(\rho_{\mathrm{PI}}, \varepsilon)$-weak Poincaré inequality for P_β. Consider the annealing scheme given by schedule $t \mapsto \delta \cdot \lfloor \frac{t}{T} \rfloor$, run for total time $T \cdot \left(\frac{\beta_0}{\delta} + 1 \right)$. Let ν be the output distribution of this annealing scheme. Then,*

$$
d_{\mathrm{TV}}(\nu, \mu_{\beta_0}) \leqslant \frac{\beta_0}{\delta} \cdot \left[(1 + \delta \sup \|\nabla H\|) e^{2\delta \|H\|_\infty} - 1 \right] \cdot O \left(e^{-2\rho_{\mathrm{PI}} T} + \varepsilon \right)^{1/2} .
$$

Remark 4.14. Setting $\varepsilon = 0$ and $\delta = \beta_0$ matches the guarantees of [CE22] (after applying Pinsker's inequality).

Proof. We shall prove the above using a simple inductive argument – our goal will be to show that initialized at μ_β, the $P_{\beta+\delta}$ Markov chain run for time T yields a distribution sufficiently close (in total variation distance) to $\mu_{\beta+\delta}$. The total variation distance between the distribution that the annealed Markov chain outputs and the true distribution μ_{β_0} is then upper bounded by the sum of these total variation errors.

Let $\nu^{(r,k)}$ be the distribution obtained by running the annealed Markov chain initialized with $\mu_{r\delta}$ until inverse temperature $k\delta$. In particular, $\nu^{(r,k)}$ corresponds to the result of running our annealed Markov chain for $T(k-r)$ time, and $\nu^{(k,k)} = \mu_{k\delta}$. We are interested in bounding $d_{\text{TV}}\left(\nu^{(0,k_0)}, \mu_{\beta_0}\right)$. We have

$$d_{\text{TV}}\left(\nu^{(0,k_0)}, \mu_{\beta_0}\right) = d_{\text{TV}}\left(\nu^{(k_0-1,k_0)}, \mu_{\beta_0}\right) + \sum_{1 \leqslant r \leqslant k_0-1}\left(d_{\text{TV}}\left(\nu^{(r-1,k_0)}, \mu_{\beta_0}\right) - d_{\text{TV}}\left(\nu^{(r,k_0)}, \mu_{\beta_0}\right)\right)$$

$$\leqslant d_{\text{TV}}\left(\nu^{(k_0-1,k_0)}, \mu_{\beta_0}\right) + \sum_{1 \leqslant r \leqslant k_0-1} d_{\text{TV}}\left(\nu^{(r-1,k_0)}, \nu^{(r,k_0)}\right) \qquad \text{(Triangle inequality)}$$

$$\leqslant d_{\text{TV}}\left(\nu^{(k_0-1,k_0)}, \mu_{\beta_0}\right) + \sum_{1 \leqslant r \leqslant k_0-1} d_{\text{TV}}\left(\nu^{(r-1,r)}, \nu^{(r,r)}\right) \qquad \text{(Data processing)}$$

$$= \sum_{1 \leqslant r \leqslant k_0} d_{\text{TV}}\left(\nu^{(r-1,r)}, \mu_{r\delta}\right).$$

We now turn to controlling the error functional $\text{osc}(f)^2 + \sup \|\nabla f\|^2$. Fix an arbitrary β, and set f to be the likelihood ratio $\frac{d\mu_\beta}{d\mu_{\beta+\delta}}$. Then,

$$\|f - 1\|_\infty \leqslant \left\|\frac{e^{-\delta H}}{\mathbf{E}_{\mu_{\beta+\delta}} e^{-\delta H}} - 1\right\|_\infty$$

$$\leqslant \left\|\frac{e^{-\delta H} - 1}{\mathbf{E}_{\mu_{\beta+\delta}} e^{-\delta H}}\right\|_\infty + \left|\frac{1}{\mathbf{E}_{\mu_{\beta+\delta}} e^{-\delta H}} - 1\right|$$

$$\leqslant \frac{e^{\delta\|H\|_\infty} - 1}{e^{-\delta\|H\|_\infty}} + \frac{e^{\delta\|H\|_\infty} - 1}{e^{-\delta\|H\|_\infty}} \leqslant 2 \cdot (e^{2\delta\|H\|_\infty} - 1).$$

Hence, $\text{osc}(f) \leqslant 4 \cdot (e^{2\delta\|H\|_\infty} - 1)$. Next, a simple computation yields

$$\|\nabla f\| = \frac{\delta e^{-\delta H}}{\mathbf{E}_{\mu_{\beta+\delta}} e^{-\delta H}} \|\nabla H\|$$

$$\leqslant 2\delta \cdot e^{2\delta\|H\|_\infty} \|\nabla H\|,$$

so we have $\sup \|\nabla f\| \leqslant 2\delta \cdot e^{2\delta\|H\|_\infty} \sup \|\nabla H\|$.

Since each $\mu_{r\delta}$ satisfies a $(\rho_{\text{PI}}, \varepsilon)$-weak Poincaré inequality, Theorem 4.6 with the above calculation implies that

$$d_{\text{TV}}\left(\nu^{(r-1,r)}, \mu_{r\delta}\right)^2 \leqslant \chi^2\left(\nu^{(r-1,r)} \| \mu_{r\delta}\right)$$

$$\leqslant e^{-2\rho_{\text{PI}}T} \cdot \chi^2\left(\mu_{(r-1)\delta} \| \mu_{r\delta}\right) + \varepsilon \cdot \left(16(e^{2\delta\|H\|_\infty} - 1)^2 + 4(\delta e^{2\delta\|H\|_\infty} \sup \|\nabla H\|)^2\right)$$

$$\leqslant \left(16(e^{2\delta\|H\|_\infty} - 1)^2 + 4(\delta e^{2\delta\|H\|_\infty} \sup \|\nabla H\|)^2\right)\left(e^{-2\rho_{\text{PI}}T} + \varepsilon\right).$$

Plugging this back into the earlier sequence of equations completes the proof. $\qquad\qquad\square$

Remark 4.15. While the proof above has been stated for the annealing scheme where at time t the Hamiltonian is of the form $\sigma \mapsto \beta(t) \cdot H(\sigma)$, the proof immediately extends to essentially any annealing scheme that changes the Hamiltonian "slowly", in that if H_t is the Hamiltonian at time t, $\|H_{t+T} - H_t\|_\infty \leqslant \delta$ for all t. A concrete example of such a scheme that might work better than the vanilla annealing is that which at time t has as Hamiltonian $\sigma \mapsto H(\beta(t) \cdot \sigma)$.

5   Vignette: Sampling from mixture models with advice

We are interested in the following question.

> Let π be a distribution over $\mathbb{R}^N$ with density proportional to e^{-V}. Given oracle access to the gradient ∇V, when is it possible to efficiently produce samples that are close (in total variation distance) to π?

We begin with an overview of existing results towards the above question. Recall from Fact 3.7 that for distributions satisfying a Poincaré inequality, such as strongly log-concave distributions, Langevin diffusion enjoys rapid mixing. Beyond this setting, however, very little is known. [BCE+22, CWZZ24] prove certain "local mixing" guarantees for Langevin diffusion on non-log-concave distributions, but these do immediately not translate to sampling guarantees. The works [GLR18, LRG18, GTC24] use Langevin diffusion-based algorithms to sample from mixtures of log-concave distributions. Furthermore, the first of these papers proves that it is hard to sample from a mixture of two Gaussian distributions with distinct covariance matrices given access to just the gradient ∇V.

In [KV23], the first theoretical guarantees are provided for a new model designed to circumvent this issue, where in addition to being given access to the gradient ∇V, we are also given "advice" in the form of m samples from the distribution (also see [NHH+20] and [Hin10, GLZ+18, XLZW16] for related discussion). In particular, they show that when the stationary distribution is a mixture of constantly many strongly log-concave distributions, Langevin diffusion initialized at the empirical measure on the advice gets close to the stationary distribution. However, their dependence on the number of components K is doubly exponential. The main result in this section improves the doubly exponential dependence to a polynomial one for any mixture of distributions satisfying Poincaré inequalities. Similar results are obtained by Koehler, Lee, & Vuong [KLV24].

Theorem 5.1. *Let $\varepsilon, \delta \in (0, 1)$, and let π a mixture*

$$\pi = \sum_{i=1}^{K} p_i \pi_i$$

of distributions $(\pi_i)_{i=1}^K$, where each π_i satisfies a Poincaré inequality with constant (at least) ρ_{PI}. Further assume that $p_i \geqslant p_*$ for all i. Let ν_0 be a random distribution over $\mathbb{R}^N$ such that $\mathbf{E}\nu_0 = \pi$, in that for any measurable subset A of $\mathbb{R}^N$, $\mathbf{E}\nu_0(A) = \pi(A)$. Set

$$m = \Omega\left(\frac{\log(1/\delta)}{p_* \varepsilon^2}\right).$$

Let $v_1, \ldots, v_m$ be iid draws from v_0, and v the uniform mixture over the $(v_i)_{i=1}^m$. Further suppose that with probability at least $1 - \delta$, $\chi^2(v_i \| \pi) \leqslant M$. Denoting by μ_T the distribution attained by running Langevin diffusion for time T initialized v, it holds that

$$\mathbf{Pr}\left[\chi^2(\mu_T \| \pi) \leqslant \varepsilon\right] \geqslant 1 - O(\delta),$$

for $T = \Omega\left(\frac{1}{\rho_{\mathrm{PI}}} \log\left(\frac{M}{\varepsilon}\right)\right)$, where the probability is over the draws of v_i.

Remark 5.2. One should think of v_0 as being the point mass distribution supported on a (random) sample drawn from π. Alternatively, one can think of v_0 as being the distribution obtained by drawing a sample x_0 according to π, then running Langevin diffusion for a short amount of time — doing this would make the χ^2-divergence $\chi^2(v_0 \| \pi)$ finite. We also remark that a version of this proof goes through if we have that each π_i satisfies a log-Sobolev inequality instead of a Poincaré inequality, working with KL divergences instead.

Proof of Theorem 5.1. The idea of the proof will be to show that up to some additive error depending on the samples, π does satisfy a Poincaré inequality with respect to the distributions along the path of Langevin diffusion initialized at the empirical distribution. This error corresponds to how imbalanced the samples are in terms of the mixture weights — a straightforward concentration argument using Bernstein's inequality then shows that this error is small, so the χ^2 divergence essentially decays exponentially fast, as if π satisfied a true Poincaré inequality.

Let f_t be the Radon-Nikodym derivative of μ_t (obtained by running Langevin diffusion initialized at v) with respect to π. By definition, we have

$$\begin{aligned}
\chi^2(\mu_t \| \pi) &= \mathbf{E}_\pi[f_t^2] - 1 \\
&= \sum_{i=1}^K p_i\left(\mathbf{E}_{\pi_i}[f_t^2] - 1\right) \\
&= \sum_{i=1}^K p_i \mathbf{Var}_{\pi_i}[f_t] + \sum_{i=1}^K p_i\left(\mathbf{E}_{\pi_i}[f_t]^2 - 1\right).
\end{aligned}$$

Because each π_i satisfies a Poincaré inequality, the first term is bounded as

$$\sum_{i=1}^K p_i \mathbf{Var}_{\pi_i}[f_t] \leqslant \frac{1}{\rho_{\mathrm{PI}}} \sum_{i=1}^K p_i \mathbf{E}_{\pi_i} \|\nabla f_t\|^2 = \frac{1}{\rho_{\mathrm{PI}}} \mathbf{E}_\pi \|\nabla f_t\|^2.$$

Consequently,

$$\chi^2(\mu_t \| \pi) \leqslant \frac{1}{\rho_{\mathrm{PI}}} \cdot \mathbf{E}_\pi \|\nabla f_t\|^2 + \sum_{i=1}^K p_i\left(\mathbf{E}_{\pi_i}[f_t]^2 - 1\right). \tag{4}$$

Theorem 4.6 then yields that

$$\begin{aligned}
\chi^2(\mu_T \| \pi) &\leqslant \chi^2(\mu_0 \| \pi) \cdot e^{-\rho_{\mathrm{PI}} T} + \mathbf{E}_{s \sim \Lambda_T}\left[\sum_{i=1}^K p_i\left(\mathbf{E}_{\pi_i}[f_t]^2 - 1\right)\right] \\
&\leqslant M e^{-\rho_{\mathrm{LS}} T} + \mathbf{E}_{s \sim \Lambda_T}\left[\sum_{i=1}^K p_i\left(\mathbf{E}_{\pi_i}[f_t]^2 - 1\right)\right].
\end{aligned}$$

Above, we use that because the KL divergence to π of each of the ν_i is at most M, so is that of the mixture $\mu_0 = \nu$.

To conclude, we shall establish tail bounds on

$$\mathbf{E}_{s \sim \Lambda_T} \left[\sum_{i=1}^{K} p_i \left(\mathbf{E}_{\pi_i}[f_s]^2 - 1 \right) \right].$$

For $1 \leqslant j \leqslant m$, let $f_s^{(j)}$ be the Radon-Nikodym derivative of $\mu_s^{(j)}$ with respect to π, where $\mu_s^{(j)}$ is the distribution obtained by running Langevin diffusion for time s initialized at ν_j. It is not difficult to see that $f_s = \frac{1}{m} \sum_{j=1}^{m} f_s^{(j)}$.

First, for fixed s and j, we use the fact that the $(\mathbf{E}_{\pi_i}[f_s^{(j)}])_j$ are independent mean 1 random variables, with Hoeffding's inequality, to get tail bounds for $\mathbf{E}_{\pi_i}[f_s]^2 - 1$. We may use this to bound a certain Orlicz norm of this random variable — this bound on the norm also transfers to $\mathbf{E}_{s \sim \Lambda_T} \left[\sum_{i=1}^{K} p_i \left(\mathbf{E}_{\pi_i}[f_s]^2 - 1 \right) \right]$ as it is a convex combination of random variables with bounded Orlicz norm. This immediately yields the desired tail bound.

Fix s and i. To start, we have the almost sure bounds

$$\frac{1}{p_*} = \frac{1}{p_*} \mathbf{E}_{\pi}[f_s^{(j)}] = \frac{1}{p_*} \sum_{r=1}^{K} p_r \mathbf{E}_{\pi_r}[f_s^{(j)}] \geqslant \mathbf{E}_{\pi_i}[f_s^{(j)}] \geqslant 0.$$

Note that because the expected ν_j is equal to π, $\mathbf{E}_{\nu_j} \mathbf{E}_{\pi_i}[f_s^{(j)}] = 1$ for any j. Furthermore, because $\mathbf{E}_{\pi_i}\left[f_s^{(j)}\right]$ is a mean 1 random variable which is bounded in $\left[0, \frac{1}{p_*}\right]$, its variance is at most $\frac{1}{p_*}$ (see e.g. [BD00]). Bernstein's inequality implies that

$$\mathbf{Pr}\left[|\mathbf{E}_{\pi_i} f_s - 1| > t\right] = \mathbf{Pr}\left[\left|\frac{1}{m} \sum_{i=1}^{m} \mathbf{E}_{\pi_i}\left[f_s^{(j)}\right] - 1\right| > t\right] \leqslant 2 \exp\left(-\frac{m p_*}{2} \cdot \frac{t^2}{1+t}\right).$$

Thus, for any $t > 0$,

$$\mathbf{Pr}\left[\left|\mathbf{E}_{\pi_i}[f_s]^2 - 1\right| > t\right] \leqslant \mathbf{Pr}\left[|\mathbf{E}_{\pi_i} f_s - 1| > \frac{t}{2(1+\sqrt{t})}\right]$$

$$\leqslant 2 \exp\left(-\frac{m p_*}{8} \cdot \frac{\left(\frac{t}{1+\sqrt{t}}\right)^2}{1 + \frac{t}{1+\sqrt{t}}}\right).$$

Now, consider the Orlicz norm $\|\cdot\|_\psi$ associated to the above family of tail bounds. As mentioned earlier, standard machinery may be used to go from the above tail bounds to a bound on the norm $\left\|\mathbf{E}_{\pi_i}[f_s]^2 - 1\right\|_\psi$. Convexity of the norm yields the same bound on $\left\|\mathbf{E}_{s \sim \Lambda_T} \sum_{i=1}^{K} p_i \left(\mathbf{E}_{\pi_i}[f_s]^2 - 1\right)\right\|_\psi$. Translating this back to a tail bound, we get that

$$\mathbf{Pr}\left[\left|\mathbf{E}_{s \sim \Lambda_T} \sum_{i=1}^{K} p_i \left(\mathbf{E}_{\pi_i}[f_s]^2 - 1\right)\right| > \frac{\varepsilon}{2}\right] \leqslant 2 \exp\left(-\frac{m p_* \varepsilon^2}{10}\right) \leqslant \delta.$$

Conditioning on the above event not happening, we get that

$$\chi^2(\mu_T \| \pi) \leqslant \chi^2(\mu_0 \| \pi) \cdot e^{-\rho_{\mathrm{PI}} \cdot T} + \frac{\varepsilon}{2} \leqslant \varepsilon,$$

as desired. $\qquad \square$

6 Stopped localization schemes

6.1 Localization schemes

We review some basic notions for the localization schemes framework introduced in [CE22].

Definition 6.1 (Linear-tilt localization scheme). Let $\mu = \mu_0$ be a probability measure, $(\mu_t)_{t \in \mathbb{Z}_{\geqslant 0}}$ be a localization process. A linear-tilt localization scheme is one where μ_t is defined by

$$\mu_{t+1}(x) = \mu_t(x) \left(1 + \langle x - \mathbf{m}(\mu_t), Z_t \rangle\right)$$

where Z_t is a random variable with $\mathbf{E}[Z_t|\mu_t] = 0$ and $\mathbf{m}(\mu_t)$ denotes the mean of μ_t.

For our main application to p-spin models, we will focus on a continuous-time version of linear-tilt localization known as stochastic localization [Eld13].

Definition 6.2 (Stochastic localization). Let μ be a probability measure on $\Omega \subseteq \mathbb{R}^N$, $(B_t)_{t \geqslant 0}$ be a standard Brownian motion on $\mathbb{R}^N$. The stochastic localization process with driving matrix $(C_t)_{t \geqslant 0}$ is a localization process $(\mu_t)_{t \geqslant 0}$ with $\mu_0 = \mu$ and

$$\mu_t(x) \propto \mu_0(x) \exp\left(-\tfrac{1}{2}\langle x, \Sigma_t x \rangle + \langle y_t, x \rangle\right),$$

where $\Sigma_t = \int_0^t C_s^2 \, ds$ and $y_t = \int_0^t C_s^2 \mathbf{m}(\mu_s) \, ds + C_s \, dB_s$.

A crucial property of these localization schemes is that establishing (approximate) conservation of variance reduces to bounding the covariance matrices of the intermediate distributions μ_t.

Lemma 6.3 (Conservation of variance for linear-tilt [CE22, Claim 22]). *Let $(\mu_t)_{t \in \mathbb{Z}_{\geqslant 0}}$ be a linear-tilt localization process. Suppose that for all $t \leqslant T$ we have*

$$\left\| \mathbf{Cov}(Z_t|\mu_t)^{1/2} \cdot \mathbf{Cov}(\mu_t) \cdot \mathbf{Cov}(Z_t|\mu_t)^{1/2} \right\|_{\mathrm{op}} \leqslant K_t,$$

where $K_t \in [0,1]$. Then for any function φ,

$$\frac{\mathbf{EVar}_{\mu_T}[\varphi]}{\mathbf{Var}_\mu[\varphi]} \geqslant \prod_{t=0}^{T-1}(1 - K_t).$$

Lemma 6.4 (Conservation of variance for stochastic localization). *Let $(\mu_t)_{t \geqslant 0}$ be a stochastic localization process with driving matrix $(C_t)_{t \geqslant 0}$. Suppose that for all $t \leqslant T$ we have*

$$\left\| C_t^{1/2} \cdot \mathbf{Cov}(\mu_t) \cdot C_t^{1/2} \right\|_{\mathrm{op}} \leqslant K_t$$

where $K_t \in [0,1]$. Then for any function φ,

$$\frac{\mathbf{EVar}_{\mu_T}[\varphi]}{\mathbf{Var}_\mu[\varphi]} \geqslant e^{-\int_0^T K_t \, dt}.$$

6.2 Proving weak Poincaré inequalities using stopped localization schemes

To apply Theorem 4.13, we required weak Poincaré inequalities for the measures of interest. To show these, we next introduce a generic tool to prove these using Lemma 4.11, building on the localization schemes framework introduced in Subsection 6.1. Let μ be a distribution. Using a localization scheme, we would like to design a measure decomposition $\mu = \mathbf{E}_{z \sim \rho} \mu_z$ such that

- for all functions f, $\mathbf{Var}_\pi[f] \leqslant C_{\mathrm{Var}} \mathbf{E}_{z \sim \rho} \mathbf{Var}_{\pi_z}[f]$, and

- with probability $1 - \eta$ over $z \sim \rho$, π_z satisfies a $(\rho_{\mathrm{PI}}, \delta)$-weak Poincaré inequality.

One way to ensure the first condition — approximate conservation of variance — is to simply stop the localization scheme whenever it fails to hold. Indeed, the following lemma immediately follows from Lemma 6.3.

Lemma 6.5. *Let $\mu = \mu_0$ be a measure, and let $(\mu_t)_{t \in \mathbb{Z}_{\geqslant 0}}$ be a linear-tilt localization process defined by*

$$\mu_{t+1}(x) = \mu_t(x) \left(1 + \langle x - \mathbf{m}(\mu_t), Z_t \rangle\right)$$

for some random variable Z_t with $\mathbf{E}[Z_t | \mu_t] = 0$. Let $T > 0$ be an arbitrary stopping time and $0 \leqslant K_t < 1$ for each $t \geqslant 0$, and consider the stopping time

$$\tau = T \wedge \inf_{t \geqslant 0} \left\{ \left\| \mathbf{Cov}(Z_t | \mu_t)^{1/2} \cdot \mathbf{Cov}(\mu_t) \cdot \mathbf{Cov}(Z_t | \mu_t)^{1/2} \right\|_{\mathrm{op}} \geqslant K_t \right\}.$$

Then, for any function φ,

$$\frac{\mathbf{EVar}_{\mu_\tau}[\varphi]}{\mathbf{Var}_\mu[\varphi]} \geqslant \prod_{t \geqslant 0} (1 - K_t).$$

Similarly, we have the following lemma for stochastic localization, which follows from Lemma 6.4.

Lemma 6.6. *Let $\mu = \mu_0$ be a measure, and $(\mu_t)_{t \geqslant 0}$ be a stochastic localization process with driving matrix $(C_t)_{t \geqslant 0}$. Let $T, K > 0$ be constant parameters, and consider the stopping time*

$$\tau = T \wedge \inf_{t \geqslant 0} \left\{ \left\| C_t^{1/2} \cdot \mathbf{Cov}(\mu_t) \cdot C_t^{1/2} \right\|_{\mathrm{op}} \geqslant K \right\}.$$

Then,

$$\frac{\mathbf{EVar}_{\mu_\tau}[\varphi]}{\mathbf{Var}_\mu[\varphi]} \geqslant e^{-TK}.$$

Remark 6.7. The localization process in the above lemmas can depend on φ, and need not be a linear-tilt localization. The more general requirement is that

$$\frac{\mathbf{E}\left[\mathbf{Var}_{\mu_{t+1}}[\varphi] | \mu_t\right]}{\mathbf{Var}_{\mu_t}[\varphi]} \geqslant K_t \quad \text{or} \quad \frac{1}{\mathbf{Var}_{\mu_t}[\varphi]} \cdot \frac{\mathrm{d}}{\mathrm{d}s} \mathbf{E}\left[\mathbf{Var}_{\mu_s}[\varphi] \mid \mu_t\right]\bigg|_{s=t} \geqslant K.$$

This can always be achieved by stopping the localization process whenever these conditions fail to hold.

With these elements in hand, we now show how to prove a weak Poincaré inequality using stopped localization schemes.

Lemma 6.8. *Let $\mu = \mu_0$ be a measure, and $(\mu_t)_{t \geqslant 0}$ be a stochastic localization process with driving matrix $(C_t)_{t \geqslant 0}$. Let $T, K > 0$ be constant parameters. Suppose that with probability $1 - \eta_1$, it holds that $\left\| C_t^{1/2} \cdot \mathbf{Cov}(\mu_t) \cdot C_t^{1/2} \right\|_{\mathrm{op}} < K$ for all $t \in [0, T]$. Further suppose that with probability $1 - \eta_2$, μ_T satisfies a $(\rho_{\mathrm{PI}}, \delta)$-weak Poincaré inequality. Then, μ satisfies a $\left(\rho_{\mathrm{PI}} e^{-TK}, e^{TK} (\delta + \eta_1 + \eta_2) \right)$-weak Poincaré inequality.*

Proof. As in Lemma 6.6, define the stopping time

$$\tau = T \wedge \inf_{t \geqslant 0} \left\{ \left\| C_t^{1/2} \cdot \mathbf{Cov}(\mu_t) \cdot C_t^{1/2} \right\|_{\mathrm{op}} \geqslant K \right\}.$$

Consider the measure decomposition $\mu = \mathbf{E}\mu_\tau$. By Lemma 6.6, this decomposition is variance-conserving with parameter e^{-K}. By the hypothesis of the lemma, $\tau = T$ with probability $1 - \eta_1$, and μ_1 satisfies a weak Poincaré inequality with probability $1 - \eta_2$. Consequently, μ_τ satisfies a weak Poincaré inequality with probability at least $1 - \eta_1 - \eta_2$. Lemma 4.11 completes the proof. $\square$

Remark 6.9. An analogous lemma to the above holds if $(\mu_t)_{t \in \mathbb{Z}_{\geqslant 0}}$ is any linear-tilt localization process.

While it will not be used in this paper, we note that a similar method proves a weak Poincaré inequality for a natural Markov chain associated to a localization scheme. This includes for example the restricted Gaussian dynamics; see [CE22] for several other examples.

Lemma 6.10. *Let $\mu = \mu_0$ be a measure, and $(\mu_t)_{t \geqslant 0}$ be a stochastic localization process with driving matrix $(C_t)_{t \geqslant 0}$. Let $T, K > 0$ be constant parameters. Consider the Markov chain P given by $P_{x \to y} = \mathbf{E}\left[\frac{\mu_T(x)\mu_T(y)}{\mu_0(x)} \right]$. Define the stopping time*

$$\tau = T \wedge \inf_{t \geqslant 0} \left\{ \left\| C_t^{1/2} \cdot \mathbf{Cov}(\mu_t) \cdot C_t^{1/2} \right\|_{\mathrm{op}} \geqslant K \right\}.$$

If $\tau = T$ with probability at least $1 - \delta$, then P satisfies a $(e^{-TK}, \delta e^{TK})$-weak Poincaré inequality.

Proof. For the Markov chain P, the Dirichlet form is given by $\mathcal{E}_P(f, f) = \mathbf{EVar}_{\mu_T}[f]$ (see, e.g., [CE22, Proposition 19]). We then have the chain of inequalities

$$\begin{aligned} \mathbf{EVar}_{\mu_T}[f] &\geqslant \mathbf{EVar}_{\mu_\tau}[f] - \delta \mathrm{osc}(f)^2 \\ &\geqslant e^{-TK} \mathbf{Var}_{\mu_0}[f] - \delta \mathrm{osc}(f)^2. \end{aligned}$$

The first inequality here is immediate since

$$\mathbf{EVar}_{\mu_\tau}[f] - \mathbf{EVar}_{\mu_T}[f] = \mathbf{EVar}_{\mu_\tau}[f]\mathbf{1}_{\tau \neq T} \leqslant \mathrm{osc}(f)^2 \mathbf{Pr}[\tau \neq T].$$

The second inequality follows from Lemma 6.4. $\square$

Remark 6.11. As in Remark 6.7, the above lemma can be generalized to localization schemes other than stochastic localization.

7 Sampling from spherical p-spin models

In this section, we prove that simulated annealing samples from spherical spin glass models for models satisfying (SL). Recall that $S_N = \sqrt{N} \cdot \mathbb{S}^{N-1}$. For $\gamma_2, \gamma_3, \ldots, \gamma_{p_*} \geqslant 0$, the mixed p-spin Hamiltonian $H_N : S_N \to \mathbb{R}$ is defined by

$$H_N(\sigma) := \sum_{p \geqslant 2} \frac{\gamma_p}{N^{(p-1)/2}} \sum_{i_1,\ldots,i_p=1}^{N} g_{i_1,\ldots,i_p} \sigma_{i_1} \cdots \sigma_{i_p}, \tag{5}$$

for i.i.d. samples $g_{i_1,\ldots,i_p}$ from $\mathcal{N}(0,1)$. This is the gaussian process on $\mathbb{R}^N$ with covariance

$$\mathbf{E} H_N(\sigma^1) \cdot H_N(\sigma^2) = N \cdot \xi\Big(R(\sigma^1, \sigma^2)\Big),$$

where we recall the mixture function ξ is defined by $\xi(s) = \sum_{p=2}^{p_*} \gamma_p^2 s^p$. The algorithm we will study is the following simple annealing scheme for Langevin diffusion.

Definition 7.1 (Annealed Langevin diffusion). Let $\delta_N, T_N > 0$ be parameters possibly depending on N. For any $\beta \geqslant 0$, let $\mu_\beta := \mu_{\beta H_N}$, where H_N is the p-spin Hamiltonian. Annealed Langevin diffusion is the annealing scheme $\mathcal{A}$ where $\beta(t) = \delta_N \lfloor t/T_N \rfloor$ and P_β is the Langevin semigroup operator for Gibbs distribution μ_β. In words, $\mathcal{A}$ keeps β constant for time T_N and then increments β by δ_N.

Theorem 7.2. *Let H_N be a mixed p-spin Hamiltonian whose mixture function ξ satisfies (SL), which we recall below:*

$$\xi''(q) < \frac{1}{(1-q)^2} \text{ for all } q \in [0,1).$$

Let μ be the associated Gibbs measure over the scaled sphere S_N, with

$$\mathrm{d}\mu(\sigma) \propto \exp(H_N(\sigma)) \, \mathrm{d}\rho(\sigma).$$

With probability $1 - e^{-cN^{1/5}}$ over the randomness of H_N, the following holds. For some parameters $\delta_N = O(N^{-4/5})$, $T_N = \Omega(N^{1/5})$, the output measure ν of the the annealed Langevin diffusion scheme with these parameters satisfies

$$d_{\mathrm{TV}}(\nu, \mu) \leqslant e^{-cN^{1/5}}.$$

Remark 7.3. We expect that the error $e^{-cN^{1/5}}$ can be improved to e^{-cN}, matching the fact that a $e^{-O(N)}$ fraction of the Gibbs measure is typically trapped in metastable states between the uniqueness and shattering thresholds [AJ24]. However, we will not pursue this improvement in this paper.

Remark 7.4. The condition (SL) is a fundamental barrier for stochastic localization, both as an algorithm and a proof technique. As was essentially shown in [HMP24, Section 10], for models satisfying (Strict RS) but not (SL), the means $\mathbf{m}(\mu_t)$ along the localization process do not move stably, in the sense that there exist time intervals of width $o(1)$ in which $\mathbf{m}(\mu_t)$ moves by $\Omega(N^{1/2})$. (The condition (Strict RS) is an artifact of the proof, and it is expected that the mean continues to move non-stably beyond the regime (Strict RS)). In the setting of [HMP24], this implies that their algorithmic simulation of the localization process fails, because approximate message passing will not estimate the mean at some times. In our setting, this implies that the covariance $\mathbf{Cov}(\mu_t)$, which arises as the derivative of $\mathbf{m}(\mu_t)$, is genuinely not bounded in operator norm at some times, and thus the main input to our framework does not hold.

Remark 7.5. While the result above is stated for the continuous time Langevin *diffusion*, the results therein can be adapted to the discretized Langevin Monte Carlo algorithm using standard tools, à la [Che23, Part II], to obtain a polynomial time sampling algorithm.

To prove the above, we shall use Theorem 4.13 in conjunction with Lemma 6.8. For the remainder of this section, let $(\gamma_p)_{p \geqslant 2}$ be a sequence of weights such that the associated mixture function ξ satisfies the condition (SL).

Notation (Measure decomposition for p-spin models). Let $\mu_{H_N} = \mu_0$. For a large constant time T, let $(\mu_t)_{0 \leqslant t \leqslant T}$ be the stochastic localization process with driving matrix Id (see Definition 6.2).

Lemma 7.6 (Covariance bound on stochastic localization path). *There exist constants c, K, depending only on ξ, such that for any constant $T > 0$ the following holds with probability at least $1 - e^{-cN^{1/5}}$ over the randomness of H_N. If $(\mu_t)_{0 \leqslant t \leqslant T}$ is the (random) trajectory of stochastic localization initialized at $\mu_0 = \mu_{H_N}$, with probability $1 - e^{-cN^{1/5}}$, $\|\mathbf{Cov}(\mu_t)\|_{\mathsf{op}} < K$ for all $0 \leqslant t \leqslant T$. In other words,*

$$\mathbf{Pr}_{H_N} \left[\mathbf{Pr}_{(\mu_t)|H_N} \left[\|\mathbf{Cov}(\mu_t)\|_{\mathsf{op}} < K \text{ for all } 0 \leqslant t \leqslant T \right] \geqslant 1 - e^{-cN^{1/5}} \right] \geqslant 1 - e^{-cN^{1/5}}.$$

Lemma 7.7 (Weak Poincaré inequality for endpoint distributions). *There exists a constant T depending only on ξ such that the following holds with probability at least $1 - e^{-cN}$ over the randomness of H_N. With probability at least $1 - e^{-cN}$, the (random) measure μ_T satisfies a (c, e^{-cN})-weak Poincaré inequality. In other words,*

$$\mathbf{Pr}_{H_N} \left[\mathbf{Pr}_{\mu_T | H_N} \left[\mu_T \text{ satisfies a } (c, e^{-cN})\text{-weak Poincaré inequality} \right] \geqslant 1 - e^{-cN} \right] \geqslant 1 - e^{-cN}.$$

Let us first see how these two lemmas imply the main theorem.

Proof of Theorem 7.2. Fix some $0 \leqslant \beta \leqslant 1$. Note that the Hamiltonian βH_N has mixture function $\xi_\beta(s) = \xi(\beta^2 s)$, and if ξ satisfies (SL) then ξ_β does as well. Plugging in Lemmas 7.6 and 7.7 into Lemma 6.8 implies that with probability at least $1 - e^{-cN^{1/5}}$, $\mu_{\beta H_N}$ satisfies a $(c, e^{-cN^{1/5}})$-weak Poincaré inequality. A union bound implies that with probability $1 - e^{-cN^{1/5}}$, for all β encountered along the annealing schedule, $\mu_{\beta H_N}$ satisfies a $(c, e^{-cN^{1/5}})$-weak Poincaré inequality.

By [HS22, Proposition 2.3], with probability $1 - e^{-cN}$, $\|\nabla H_N\|_\infty = O(\sqrt{N})$. The same argument implies that with probability $1 - e^{-cN}$, $\|H_N\|_\infty = O(N)$. With probability $1 - e^{-cN^{1/5}}$, all three of these events occur, and Theorem 4.13 completes the proof. $\square$

We conclude this subsection by proving Lemma 7.7.

Proof of Lemma 7.7. Let $\mathrm{d}\mu_t(\sigma) \propto e^{H_{N,T}(\sigma)} \, \mathrm{d}\sigma$ for $H_{N,T}(\sigma) = H_N(\sigma) + \langle y, \sigma \rangle$. Let

$$S_N(y) = \{\sigma \in S_N : R(y, \sigma) > 0\}.$$

Let $U \in \mathbb{R}^{N \times (N-1)}$ be a matrix whose columns are an orthonormal basis of the orthogonal complement of y. Let $\widehat{y} = \sqrt{N} y / \|y\|$ be y (which is a.s. nonzero) scaled to length $\sqrt{N}$, and define the map $\sigma_y(\rho) : \mathbb{R}^{N-1} \to S_N(y)$ by

$$\sigma_y(\rho) = \frac{\widehat{y} + U\rho}{\sqrt{1 + R(\rho, \rho)}}.$$

This is the inverse of the map that first stereographically projects $S_N(\boldsymbol{y})$ from the origin to $\widehat{\boldsymbol{y}} + \boldsymbol{U}\mathbb{R}^{N-1}$, the plane tangent to S_N at $\widehat{\boldsymbol{y}}$, and then maps the resulting point to coordinates given by $\boldsymbol{U}$. Let $\varepsilon_0 = 0.1$, and $A = \{\rho \in \mathbb{R}^{N-1} : \|\rho\|^2 \leqslant \varepsilon_0 N\}$, and note that

$$\sigma_{\boldsymbol{y}}(A) = \{\sigma \in S_N : R(\sigma, \widehat{\boldsymbol{y}}) \geqslant (1 + \varepsilon_0)^{-1/2}\}$$

is a spherical cap around $\widehat{\boldsymbol{y}}$. Let $A' := \sigma_{\boldsymbol{y}}(A)$. By arguments in [HMP24, Subsection 9.2], there exists a measure ν (denoted $\tilde{\nu}^{\mathrm{proj}}_{H_N, \boldsymbol{y}}$, see Eq. 2.10 therein) such that the following holds with probability $1 - e^{-cN}$.

- The push-forward of $\nu_{|A}$ through $\sigma_{\boldsymbol{y}}$ coincides with $(\mu_T)_{|A'}$. (Lemma 9.5 therein.)

- $\mu_T(A') = 1 - e^{-cN}$. (Lemma 9.6 therein states this with $1 - o_N(1)$ in place of $1 - e^{-cN}$, but the proof implies bound $1 - e^{-cN}$, as this is the bound given by Proposition 5.12 used therein.)

- $\nu(A) = 1 - e^{-cN}$. (Corollary 9.7 therein, modulo the same issue of $1 - o_N(1)$ versus $1 - e^{-cN}$, which is addressed the same way.)

- ν is $\Omega(1)$-strongly log-concave. (Proposition 9.8 therein.)

By the well-known Bakry-Émery condition (see, e.g., [Che23, Section 1.2.3]), on this event ν satisfies a ρ_{PI}-Poincaré inequality for some $\rho_{\mathrm{PI}} = \Omega(1)$. We will transfer this inequality to a $(\rho_{\mathrm{PI}}, e^{-cN})$-weak Poincaré inequality for μ_T. Consider a smooth test function $f : S_N \to \mathbb{R}$ and let $\widetilde{f} : \mathbb{R}^{N-1} \to \mathbb{R}$ be defined by $\widetilde{f} = f \circ \sigma_{\boldsymbol{y}}$. Since $d_{\mathrm{TV}}(\mu_T, (\mu_T)_{|A'}) = e^{-cN}$ and $d_{\mathrm{TV}}(\nu, \nu_{|A}) = e^{-cN}$, and $\mathrm{osc}(f') \leqslant \mathrm{osc}(f)$, arguing as in (3) shows

$$\mathbf{Var}_{\mu_T}(f) \leqslant \mathbf{Var}_{(\mu_T)_{|A'}}(f) + e^{-cN}\mathrm{osc}(f)$$
$$= \mathbf{Var}_{\nu_{|A}}(\widetilde{f}) + e^{-cN}\mathrm{osc}(f)$$
$$\leqslant \mathbf{Var}_{\nu}(\widetilde{f}) + 2e^{-cN}\mathrm{osc}(f).$$

By the Poincaré inequality for ν and the definition of the Dirichlet form for Langevin diffusion,

$$\mathbf{Var}_{\nu}(\widetilde{f}) \leqslant \frac{1}{\rho_{\mathrm{PI}}} \cdot \mathcal{E}_{\nu}(\widetilde{f}, \widetilde{f}) = \frac{1}{\rho_{\mathrm{PI}}} \cdot \mathbf{E}_{\nu}[\|\nabla \widetilde{f}\|^2]$$

By [HMP24, Proof of Lemma 9.5], the map $\sigma_{\boldsymbol{y}}$ has Jacobian $J_{\sigma_{\boldsymbol{y}}}$ satisfying $\|J_{\sigma_{\boldsymbol{y}}}\|_{\mathrm{op}} \leqslant 1$, and thus for all $\rho \in \mathbb{R}^{N-1}$,

$$\|\nabla \widetilde{f}(\rho)\| = \|\nabla (f \circ \sigma_{\boldsymbol{y}})(\rho)\| \leqslant \|\nabla f(\sigma_{\boldsymbol{y}}(\rho))\|.$$

It follows that

$$\mathbf{E}_{\nu}[\|\nabla \widetilde{f}\|^2] \leqslant \mathbf{E}_{\nu_{|A}}[\|\nabla \widetilde{f}\|^2] + e^{-cN} \sup \|\nabla \widetilde{f}\|^2$$
$$\leqslant \mathbf{E}_{(\mu_T)_{|A'}}[\|\nabla f\|^2] + e^{-cN} \sup \|\nabla f\|^2$$
$$\leqslant \mathbf{E}_{\mu_T}[\|\nabla f\|^2] + 2e^{-cN} \sup \|\nabla f\|^2$$

Combining the above shows

$$\mathbf{Var}_{\mu_T}(f) \leqslant \frac{1}{\rho_{\mathrm{PI}}} \mathcal{E}_{\mu_T}(f, f) + 2e^{-cN}\left(\mathrm{osc}(f) + \frac{1}{\rho_{\mathrm{PI}}} \sup \|\nabla f\|^2\right).$$

The result follows by adjusting c. $\qquad\square$

7.1 Technical overview for covariance bounds

The proof of the main theorem has boiled down to Lemma 7.6 — we now give a high-level overview of our proof strategy for this. We wish to show that with very high probability $(1 - e^{-\Omega(N^{1/5})})$, the covariance is bounded along the entire path $(\mu_t)_{0 \leqslant t \leqslant T}$ of stochastic localization. By performing a union bound over time and a standard perturbation argument, it suffices to show that for a fixed time $t \in [0, T]$, μ_t has bounded covariance with very high probability.

To do this, we recall an alternate view of stochastic localization [AM22b]. The measure at time t of stochastic localization (with the identity driving matrix) is given as follows. First, draw $\sigma \sim \mu_{H_N}$, and independently $g \sim \mathcal{N}(0, I_N)$. Then, μ_t has the same law as $\mu_{H_N, t\sigma + \sqrt{t}g}$, in that

$$\mu_{H_N, t\sigma + \sqrt{t}g}(\widetilde{\sigma}) \propto \exp\left(H_N(\widetilde{\sigma}) + \langle t\sigma + \sqrt{t}g, \widetilde{\sigma}\rangle\right).$$

As written, the covariance of this distribution is difficult to analyze — the sample σ has very complicated correlations with the disorder of the Hamiltonian H_N, making it intractable.

The planting trick. To deal with this, we will use the *planting trick* introduced by Achlioptas and Coja-Oghlan [ACO08]. The application of this method in the context of stochastic localization is by now standard [AMS22, AMS23b, HMP24], and we review the main ideas for the reader's convenience.

Definition 7.8 (Planted p-spin model). The planted measure μ_{pl} is a joint law over a Hamiltonian H_N and a *spike* $x \in S_N$ given by

$$d\mu_{\text{pl}}(H_N, x) \propto \exp\left(H_N(x)\right) \cdot d\rho(x) \cdot d\mu_{\text{null}}(H_N),$$

where ρ is the uniform measure over S_N and μ_{null} is the law over p-spin Hamiltonians with mixture function ξ. We frequently abuse notation to let $\mu_{\text{pl}}(H_N)$ denote the marginal of μ_{pl} on H_N.

To provide further intuition for the above definition, consider the following alternate sampling interpretation of the planted model, which describes the distribution of x conditioned on H_N.

Fact 7.9. *Consider the following inference problem. We start by sampling the spike $x \sim S_N$, sample $G^{(p)}$ as a rank-p tensor with iid $\mathcal{N}(0, 1)$ entries for $p \geqslant 2$, and for each p let $M^{(p)} = -G^{(p)} + \frac{\gamma_p}{N^{(p-1)/2}} x^{\otimes p}$.*

Then, the posterior on x after observing the tensors $(M^{(p)})_{p \geqslant 2}$ is of the form $\mu(x = \sigma \mid (M^{(p)})) \propto \exp(H_N(\sigma))$, where

$$H_N(\sigma) = \sum_{p \geqslant 2} \frac{\gamma_p}{N^{(p-1)/2}} \langle M^{(p)}, \sigma^{\otimes p}\rangle.$$

Then, the joint law of (H_N, x) is μ_{pl}.

The above says that conditioned on H_N, the distribution of x (according to μ_{pl}) is simply distributed as a sample according to μ_{H_N}. That is, the spike x resulting in a Hamiltonian $H_N \sim \mu_{\text{pl}}$ is exchangeable with a sample from μ_{H_N}.

The latter of these interpretations will be very useful for us. When dealing with the measure at time t of stochastic localization applied to the p-spin model, the primary issue was that it was unclear how to deal with the sample σ drawn from the Gibbs distribution. However, if we could

work with the planted p-spin model, this issue would be absent. Indeed, the exchangability of the spike and a sample implies that the law of μ_t applied to the planted model is given by

$$\mu_{H_N, tx + \sqrt{t}g}(\widetilde{\sigma}) \propto \exp\left(H_N(\widetilde{\sigma}) + \langle tx + \sqrt{t}g, \widetilde{\sigma} \rangle \right),$$

where x is the spike hidden in H_N. This decouples the randomness of the external field $tx + \sqrt{t}g$ and the disorder of the Hamiltonian H_N that arises from the Gaussians $(G^{(p)})_{p \geqslant 2}$.

As was shown in [HMP24, Corollary 3.5] and recalled just below, the planted and null models are mutually contiguous. Thus high-probability statements from one model transfer to the other, and it suffices to study the planted model.

For all models satisfying (Strict RS), the measures $\mu_{\mathsf{null}}(H_N)$ and $\mu_{\mathsf{pl}}(H_N)$ from Definition 7.8 are mutually contiguous, i.e., for any sequence of events $\mathcal{E}_N$, $\mu_{\mathsf{null}}(\mathcal{E}_N) \to 0$ whenever $\mu_{\mathsf{pl}}(\mathcal{E}_N) \to 0$.

The transfer from the p-spin model to the planted model may then be carried out by setting

$$\mathcal{E}_N = \left\{ H_N : \mathbf{Pr}_{\substack{\sigma \sim \mu_{H_N} \\ g \sim \mathcal{N}(0, I_N)}} \left[\left\| \mathbf{Cov}\left(\mu_{H_N, t\sigma + \sqrt{t}g} \right) \right\| > K \right] < e^{-cN^{1/5}} \right\}.$$

This event is very complicated in the null model, but exchangeability makes it tractable in the planted model. In the actual proof, we will require a stronger (quantitative) version of mutual contiguity; see Proposition 7.16 for details.

Now, we must understand what the Hamiltonian in the planted model looks like conditioned on the spike.

Fact 7.10. *Consider the following process: sample $x \sim S_N$, $\widetilde{H}_N \sim \mu_{\mathsf{null}}$, and define H_N by $H_N(\sigma) = \widetilde{H}_N(\sigma) + N \cdot \xi(R(x, \sigma))$. Then, the joint law of (H_N, x) is μ_{pl}.*

Consequently, our goal is to bound the covariance of the distribution

$$\mu_t(\sigma) \propto \exp\left(\widetilde{H}_N(\sigma) + N \cdot \xi\left(R\left(x, \sigma\right)\right) + \langle tx + \sqrt{t}g, \sigma \rangle \right).$$

for $\widetilde{H}_N \sim \mu_{\mathsf{null}}$ with mixture function ξ. Now, define ξ_t by $\xi_t(s) = \xi(s) + ts$, and extend the definition of the p-spin model (5) to allow a random linear term. Then,

$$\mu_t(\sigma) \propto \exp\left(\underbrace{\widetilde{H}_{N,t}(\sigma) + N \cdot \xi_t(R(x, \sigma))}_{H_{N,t}(\sigma)} \right),$$

where $\widetilde{H}_{N,t} \sim \mu_{\mathsf{null}}$ with mixture function ξ_t.

The TAP planted model. We now turn to controlling the covariance matrix of these models. As we will see below, it is relatively easier to bound the covariance matrix (in fact, the second moment matrix) of a model with zero or small external field. However, for any time $t > 0$, $H_{N,t}$ has an external field. We will use a method developed in [HMP24] to reduce to the case of a model with zero or small external field.

Let $m^{\text{true}} = \mathbf{m}(\mu_t)$. The main intuition of this reduction is that the Gibbs measure concentrates near a codimension-2 band passing through m^{true} and orthogonal to m^{true} and x, and furthermore *the model on this band is essentially a replica symmetric model with no external field.* Moreover, one expects that both $R(m^{\text{true}}, m^{\text{true}})$ and $R(m^{\text{true}}, x)$ concentrate near a value $q_* = q_*(t)$ defined by $\xi'_t(q_*) = \frac{q_*}{1-q_*}$.

However, m^{true} is a complicated function of $H_{N,t}$, so it is a priori difficult to reason about the joint distribution of $(m^{\text{true}}, H_{N,t})$. Thus, this reduction is formally carried out by conditioning on a *TAP fixed point* m^{TAP}, which will serve as a proxy for m^{true}. Define the TAP free energy

$$\mathcal{F}_{\text{TAP}}(m) = H_{N,t}(m) + \frac{N}{2} \cdot \theta(R(m,m)) + \frac{N}{2} \log(1 - R(m,m)),$$

where

$$\theta(s) = \xi(1) - \xi(s) - (1-s)\xi'(s).$$

As shown in [HMP24], for sufficiently small constant $\iota > 0$, with probability $1 - e^{-cN}$ $\mathcal{F}_{\text{TAP}}$ has a unique critical point m^{TAP} in the region $\mathcal{S}_\iota$ defined by $R(m,m), R(m,x) \in [q_* - \iota, q_* + \iota]$. Due to the existence and uniqueness of m^{TAP}, it becomes possible to relate $H_{N,t}$ to a "TAP-planted model" where one samples m^{TAP} first, and then samples $H_{N,t}$ conditional on $\nabla \mathcal{F}_{\text{TAP}}(m) = 0$:

Lemma 7.11 (See Lemma 7.21; essentially due to [HMP24]). *For any small constant $\iota > 0$, the following holds. For any $H_{N,t}$-measurable event $\mathcal{E}$, if*

$$\sup_{m^{\text{TAP}} \in \mathcal{S}_\iota} \mathbf{Pr}(\mathcal{E} | \nabla \mathcal{F}_{\text{TAP}}(m^{\text{TAP}}) = 0) \to 0,$$

then $\mathbf{Pr}(\mathcal{E}) \to 0$.

Crucially, the conditional law of $H_{N,t}$ in the TAP-planted model is very tractable, as (for a fixed m^{TAP}) $\nabla \mathcal{F}_{\text{TAP}}(m^{\text{TAP}}) = 0$ amounts to a linear constraint on the Gaussian process $H_{N,t}$. The resulting explicit conditional law of $H_{N,t}$ is described in Lemma 7.23.

Remark 7.12. While it will not be relevant to our purposes, [AMS22, AMS23b, HMP24] have shown that m^{TAP} typically approximates m^{true} well, in the sense that $\|m^{\text{true}} - m^{\text{TAP}}\|^2 = O(1)$, thereby justifying the heuristic that m^{TAP} is a proxy for m^{true}.

Remark 7.13. The idea of reducing to a TAP-planted model has also been used beyond the setting of sampling from spherical spin glasses. In the recent work [Hua24], an analogous reduction is used to obtain the capacity of the Ising perceptron. In this application, passage to the TAP-planted model is used to tightly control a partition function rather than to bound a covariance matrix.

Consequently, we can now work within the TAP-planted model. Let H_{TAP} denote the Hamiltonian $H_{N,t}$ after conditioning on x and $\nabla \mathcal{F}_{\text{TAP}}(m^{\text{TAP}}) = 0$. As

$$\mathbf{Cov}(\mu_{H_{\text{TAP}}}) \preceq \mathbf{E}_{\sigma \sim \mu_{H_{\text{TAP}}}}(\sigma - \mathbf{v})(\sigma - \mathbf{v})^\top$$

for any $\mathbf{v} \in \mathbb{R}^N$ (with equality at $\mathbf{v} = m^{\text{true}}$), it suffices to control the operator norm of

$$\mathbf{E}_{\sigma \sim \mu_{H_{\text{TAP}}}}(\sigma - m^{\text{TAP}})(\sigma - m^{\text{TAP}})^\top.$$

Reduction to slices of the sphere. Next, to control the covariance, let us decompose the sphere into codimension-2 slices

$$S(a,b) := \left\{ \sigma \in S_N : R(\sigma, \boldsymbol{m}) = \left(1 + \frac{a}{\sqrt{N}} \right) R(\boldsymbol{m}, \boldsymbol{m}), \ R(\sigma, \boldsymbol{x}) = \left(1 + \frac{b}{\sqrt{N}} \right) R(\boldsymbol{x}, \boldsymbol{m}) \right\},$$

with the central slice centered at $\boldsymbol{m} = \boldsymbol{m}^{\mathsf{TAP}}$. Let $\mu_{a,b}$ be the measure $\mu_{H_{\mathsf{TAP}}}$ conditioned to lie in the codimension-2 slice $S(a,b)$.

The concentration of the Gibbs measure described in the previous section implies that, viewed as random variables of a sample $\sigma \sim \mu_{H_{\mathsf{TAP}}}$, a and b are well-concentrated around 0. Let $v_{a,b}$ be the center of $S(a,b)$. The covariance of the distribution may be bounded as

$$\begin{aligned}
\mathbf{Cov}(\mu) &\preceq \mathbf{E}(\sigma - \boldsymbol{m})(\sigma - \boldsymbol{m})^\top \\
&= \mathbf{E}_{(a,b)} \mathbf{E}_{\sigma \sim \mu_{a,b}} (\sigma - v_{a,b} + v_{a,b} - \boldsymbol{m})(\sigma - v_{a,b} + v_{a,b} - \boldsymbol{m})^\top \\
&\preceq 2 \mathbf{E}_{(a,b)} \mathbf{E}_{\sigma \sim \mu_{a,b}} (\sigma - v_{a,b})(\sigma - v_{a,b})^\top + 2 \mathbf{E}_{(a,b)}(v_{a,b} - \boldsymbol{m})(v_{a,b} - \boldsymbol{m})^\top \\
&\preceq 2 \mathbf{E}_{(a,b)} \mathbf{E}_{\sigma \sim \mu_{a,b}} (\sigma - v_{a,b})(\sigma - v_{a,b})^\top + 2 \mathbf{E}_{(a,b)} O\left(a^2 + b^2\right).
\end{aligned} \tag{6}$$

One can interpret $v_{a,b}$ as explaining the variation within the slice originating from the $\boldsymbol{m}$ and $\boldsymbol{x}$ directions. Hence, as alluded to in the previous discussion about the TAP planted model, the key fact is that under $\mu_{a,b}$, the recentered sample $\sigma - v_{a,b}$ is a sample from a spherical spin glass in two lower dimensions, as can be shown by calculating the covariance of the (conditioned) Gaussian process H_{TAP} restricted to this slice. This verification is carried out in Corollary 7.26.

These codimension-2 models have the crucial property that the spherical spin glass on the slice $a = b = 0$ is a model satisfying (Strict RS) with no external field (i.e. degree-1 term), while nearby slices have a small (random) external field of magnitude $\sqrt{a^2 + b^2}$. In particular, the first term of (6) requires bounding the second moment of a Gibbs sample from a strictly RS model with small (random) external field. As a result, (6) would be bounded if we proved the following.

1. Let H_N be the Hamiltonian of a slightly generalized mixed p-spin model, where we allow the mixture function ξ to have a small linear term $\gamma_1 q$ (in our proofs we allow $\gamma_1^2 \leqslant N^{-4/5}$), such that the non-degree-1 part $\xi_{\sim 1}$ of ξ satisfies (Strict RS). Then, with high probability,

$$\left\| \mathbf{E}_{\mu_{H_N}} \sigma \sigma^\top \right\|_{\mathrm{op}} = O\left(1 + \gamma_1^2 N\right).$$

   Much of Section 8 is dedicated to showing this.

2. The second moments of a and b are $O(1)$. In fact, we will show in Lemma 7.37 that they are essentially $O(1)$-subgaussian.

Let us start by explaining how to show subgaussianity.

Subgaussianity of a, b. The distribution ν of (a,b) is given by

$$\nu(a,b) \propto \exp\left(\log \widehat{Z}_{a,b} + \frac{N-4}{2} \log\left(1 - \frac{\|v_{a,b}\|^2}{N} \right) + H_{\mathsf{TAP}}(v_{a,b}) \right).$$

Here, the first term $\log \widehat{Z}_{a,b}$ is the free energy of the $(N-2)$-dimensional p-spin model $\mu_{a,b}$, obtained by restricting $\mu_{H_{\mathsf{TAP}}}$ to the slice $S(a,b)$ and rescaling the distribution to lie on S_{N-2}. The second term

34

is an effective decrement in the free energy caused by the radius of the sphere $S(a, b)$ shrinking for larger values of a and b. The third term is an effective increment in the free energy coming from the energy of H_{TAP} at the center of the slice $S(a, b)$.

For a fixed (a, b), the only random quantities in the definition of ν are the first and third terms. In Theorem 7.29, proved in Section 8, we show that the first term may essentially be approximated by a deterministic function of the mixture function of $\mu_{a,b}$, at the cost of incurring a small $O(1)$ error. We do not elaborate on the details of this proof in the technical overview; it is similar to that used to bound the covariance (which we explain shortly). The third term is similar, and is a deterministic function plus a small Gaussian, whose variance is $O(a^2 + b^2)$.

Given these bounds, we may show that the distribution ν is strongly log-concave at 0 with high probability over the randomness of H_{TAP}. A simple perturbation argument then implies that ν is strongly log-concave in a macroscopic neighborhood of 0, implying subgaussianity.

Covariance bound for strictly RS models with small external fields. The covariance bound has now boiled down to bounding $\|M\|_{\text{op}}$, for $M = \mathbf{E}_{\mu_{H_N}}[\sigma\sigma^\top]$ the second moment matrix of model satisfying Eq. (Strict RS) with small external field. Note that M is a H_N-measurable random variable.

The proof proceeds in two high level steps, which we carry out in Section 8.

1. We show using the second moment method that with positive probability over H_N, $\|M\|_{\text{op}}$ is bounded.

2. Using a much simpler argument, we can show that $\|M\|_{\text{op}}$ is essentially $O(N^{-1/10})$-Lipschitz in the disorder. Hence, by gaussian concentration, it concentrates very well around its expectation (which is $O(1)$ by the positive probability bound).

Let us elaborate a bit more on the proof of the first point above. It turns out that, under the condition (Strict RS) with small external field, the leading order contribution to M comes from the degree-2 part of the Hamiltonian $H_{N,2}(\sigma) = \frac{\gamma_2}{N^{1/2}} \sum_{i,j} g_{i,j}\sigma_i\sigma_j$. We will ultimately reduce the study of the covariance matrix of μ_{H_N} to that of $\mu_{H_{N,2}}$, and then show boundedness of $\mathbf{Cov}(\mu_{H_{N,2}})$ using random matrix theory. A similar strategy of isolating the degree-2 component of H_N was used to study the partition function and magnetization of strictly RS models in [HMP24].

Degree-2 behavior. Let us discuss the typical behavior of the covariance of $\mu_{H_{N,2}}$. Define the degree-2 Gibbs measure

$$\mathrm{d}\mu_{H_{N,2}}(\sigma) \propto \exp(H_{N,2}(\sigma))\,\mathrm{d}\rho(\sigma),$$

with corresponding partition function $Z_{N,2} = \int \exp(H_{N,2}(\sigma))\,\mathrm{d}\rho(\sigma)$. This is the spherical Sherrington-Kirkpatrick model with interaction matrix $A = \nabla^2 H_N(0)$; note that A is a scaled GOE matrix. Observe that if we shift A by a constant multiple of the identity γId_N, the measure does not change, as it is supported on S_N. The crucial observation is the following:

> For a careful choice of γ, the measure $\mathrm{d}\mu_{H_{N,2}}(\sigma) \propto \exp\left(-\frac{1}{2}\langle\sigma, (\gamma\text{Id}_N - A)\sigma\rangle\right)\mathrm{d}\rho(\sigma)$ looks like a Gaussian with covariance $(\gamma\text{Id}_N - A)^{-1}$.

In fact, we will see that it suffices to pick $\gamma = 1 + \xi''(0)$. The typical value of $\|x\|_2^2$, where $x \sim \mathcal{N}(0, \gamma\text{Id}_N - A)^{-1}$, is equal to $\text{Tr}(\gamma\text{Id}_N - A)^{-1}$. By approximating this trace using the semicircle

law for the eigenvalues of A and the explicit choice of γ, we see that $\|x\|_2^2 \approx N$, which justifies the heuristic that this Gaussian approximates the spherical distribution $\mu_{H_{N,2}}$.

For the above discussion to be well-defined, we require that $\gamma \mathrm{Id}_N - A$ is positive definite, which can only occur if the maximum eigenvalue of A is bounded above by $\gamma = 1 + \xi''(0)$. By standard concentration inequalities about the maximum eigenvalue of a GOE matrix, this holds with a constant margin with exponentially good probability. Thus, at least for typical realizations of $H_{N,2}$, the covariance will have bounded operator norm. To make this rigorous, we will use the Laplace transform to precisely control the moments of the overlaps, as was previously done in [BL16, HMP24].

Reduction to degree-2. Below, we give some justification for why one should expect to be able to reduce to the degree-2 behavior. We will heuristically argue this by showing that the partition function Z_N is essentially controlled by the degree-2 portion.

To simplify the discussion, let us assume that we are in a $2 + p$ spin model, so that $H_N(\sigma) = H_{N,2}(\sigma) + H_{N,p}(\sigma)$, where $H_{N,p}(\sigma) = \frac{\gamma_p}{N^{(p-1)/2}} \sum_{i_1,\dots,i_p=1}^{N} g_{i_1,\dots,i_p} \sigma_{i_1} \cdots \sigma_{i_p}$. The corresponding mixture function decomposes as $\xi(q) = \gamma_2^2 q^2 + \xi_{\sim 2}(q)$, so that $\xi_{\sim 2}(q) = \gamma_p^2 q^p$ corresponds to the non degree-2 part of the mixture function. It turns out that, once we condition on $H_{N,2}$ (and hence the value of $Z_{N,2}$), the full partition function Z_N is essentially deterministic. Indeed, we will show in Proposition 8.2 that with very high probability,

$$Z_N \approx Z_{N,2} e^{N\xi_{\sim 2}(1)/2}$$

To see why this is reasonable, let us consider the first two moments of Z_N conditioned on the degree-2 Hamiltonian $H_{N,2}$. Indeed, let $\mathbf{E}_{\sim 2}$ denote expectation with respect to $H_{N,p}$ conditioned on $H_{N,2}$. Standard gaussian MGF calculations yield $\mathbf{E}_{\sim 2} Z_N = e^{N\xi_{\sim 2}(1)/2} Z_{N,2}$ and

$$\mathbf{E}_{\sim 2}[Z_N^2] = Z_{N,2}^2 e^{N\xi_{\sim 2}(1)} \int \exp\left(N\xi_{\sim 2}(R(\sigma^1, \sigma^2))\right) \mathrm{d}\rho^{\otimes 2}(\sigma^1, \sigma^2)$$

$$= (\mathbf{E}_{\sim 2} Z_N)^2 \int \exp\left(N\gamma_p^2 R(\sigma^1, \sigma^2)^p\right) \mathrm{d}\rho^{\otimes 2}(\sigma^1, \sigma^2).$$

At sufficiently high temperatures, the typical overlap behavior $R(\sigma^1, \sigma^2) \asymp N^{-1/2}$, where σ^1, σ^2 are iid draws from μ_{H_N}. This matches the overlap behavior at infinite temperature, where the Gibbs distribution is uniform. Then, pretending that $R(\sigma^1, \sigma^2) = cN^{-1/2}$ for all σ^1, σ^2, we obtain that

$$\mathbf{E}_{\sim 2}[Z_N^2] \approx (\mathbf{E}_{\sim 2} Z_N)^2 \exp\left(c^p \gamma_p^2 N^{1-p/2}\right).$$

Since $p \geqslant 3$, it follows that, conditional on $H_{N,2}$, the conditional variance of Z_N is tiny compared to its conditional expectation. In summary, we see that the higher degree portions of the partition function have negligible contributions to the fluctuations of Z_N, so that the typical behavior of Z_N is controlled by $Z_{N,2}$.

Turning now to the covariance bound, we will control the (i,j)th covariance entry $M_{i,j} := \int \sigma_i \sigma_j e^{H_N(\sigma)} \mathrm{d}\rho(\sigma)$. A crucial fact is that, by rotational invariance of the sphere and gaussians, we can rotate to the eigenbasis of $A = \nabla^2 H_N(0)$ so that A becomes diagonal. When A is diagonal, one can in fact show that

$$\mathbf{E}_{\sim 2}[M_{i,j}^2] \lesssim \frac{1}{N^2} (\mathbf{E}_{\sim 2} M_{i,j})^2,$$

where $\mathbf{E}_{\sim 2} M_{i,j}$ can be interpreted as (up to normalization) the predicted (i,j)th covariance entry by just looking at the degree-2 randomness; see Propositions 8.17 and 8.18 for details. It follows that the Frobenius norm error of the true covariance compared to the degree-2 covariance is $O(1)$. Combined with the typical behavior of the degree-2 covariance being essentially the diagonal matrix $((1 + \xi''(0))\mathrm{Id}_N - A)^{-1}$, we conclude an $O(1)$ covariance bound for μ_{H_N}.

Although this direct moment approach can be made rigorous at sufficiently high temperature, it will not cover the entire regime (Strict RS) of our main theorem. To deal with this, we will use the *free energy typical truncation* recently introduced by [HS23]. The main idea is that, while pairs σ^1, σ^2 with overlap $R(\sigma^1, \sigma^2) \asymp N^{-1/2}$ do not necessarily dominate the second moment $\mathbf{E}[Z_N^2]$ throughout the regime (Strict RS), there is a truncation $\widetilde{Z}_N$ accounting for nearly all of Z_N, whose second moment is dominated by such pairs. We defer the details to the following sections.

7.2 Null models, planted models, and contiguity

As described in the technical overview, we will need a quantitative strengthening of contiguity between the null and planted models. For convenience, let us restate the definition of the planted model.

Definition 7.8 (Planted p-spin model)**.** The planted measure μ_{pl} is a joint law over a Hamiltonian H_N and a *spike* $x \in S_N$ given by

$$\mathrm{d}\mu_{\mathsf{pl}}(H_N, x) \propto \exp\left(H_N(x)\right) \cdot \mathrm{d}\rho(x) \cdot \mathrm{d}\mu_{\mathsf{null}}(H_N),$$

where ρ is the uniform measure over S_N and μ_{null} is the law over p-spin Hamiltonians with mixture function ξ. We frequently abuse notation to let $\mu_{\mathsf{pl}}(H_N)$ denote the marginal of μ_{pl} on H_N.

Remark 7.14 (Interpretation of planted model)**.** Equivalently, the planted measure μ_{pl} can be described as follows.

- Sample $x \sim S_N$.

- Sample $\widetilde{H}_N \sim \mu_{\mathsf{null}}$.

- Define H_N by $H_N(\sigma) = \widetilde{H}_N(\sigma) + N \cdot \xi(R(x, \sigma))$.

The following Bayesian interpretation of μ_{pl} will make the planted model amenable to explicit calculation. For (x, H_N) sampled from μ_{pl}, the posterior distribution $x | H_N$ is described by the density:

$$\mathrm{d}\mu_{x|H_N}(\sigma) \propto \exp(H_N(\sigma)) \, \mathrm{d}\rho(\sigma) .$$

Therefore, the distribution of (H_N, σ) for $\sigma \sim \mu_{H_N}$ is identical to that of (H_N, x).

In order to show the probability bound of $1 - e^{-cN^{1/5}}$ in Lemma 7.6, we will prove the following quantitative strengthening of mutual contiguity, under the following quantitative strict RS condition. Note that, since the proof of Theorem 7.2 union bounds over $\mathrm{poly}(N)$ many values of β, quantitative control of the error in Lemma 7.6 is needed to carry out the proof.

Condition 7.15 (ε-strict replica symmetry)**.** We say ξ is ε-strictly replica symmetric if for all $q \in (0,1)$,

$$\frac{1}{q^2} \cdot (\xi(q) + q + \log(1-q)) \leqslant -\varepsilon/2. \tag{7}$$

Under this assumption, we prove the following quantitative contiguity result in Section 8.

Proposition 7.16 (Quantitative contiguity). *Under Condition 7.15, there exists $c = c(\varepsilon) > 0$ such that for any event $\mathcal{E}$, if $\mu_{\mathsf{pl}}(\mathcal{E}) = p$, then $\mu_{\mathsf{null}}(\mathcal{E}) \leqslant e^{-cN^{1/5}} + e^{\frac{1}{c}\sqrt{\log \frac{2}{p}}} p$.*

Thus, from now on, we work under the planted model. One reason the planted model is easier to work with is because of the following lemma, which provides a simple description of the distribution of μ_t (by describing the distribution of the external field at time t) in the planted model.

Lemma 7.17. *Let μ_t be the distribution after running stochastic localization with the Id driving matrix for time t initialized at μ_{H_N}. Then μ_t arises as the Gibbs distribution of the Hamiltonian $H_{N,t}(\sigma)$:*

$$H_{N,t}(\sigma) = H_N(\sigma) + \langle \boldsymbol{y}_t, \sigma \rangle,$$

where

$$(H_N, \boldsymbol{y}_t) \overset{\mathrm{law}}{=} (H_N, t\boldsymbol{x} + \sqrt{t}\boldsymbol{g}),$$

where $\boldsymbol{x} \sim S_N$, $H_N \sim \mu_{\mathsf{pl}}(\cdot|\boldsymbol{x})$, and $\boldsymbol{g} \sim \mathcal{N}(0, \mathrm{Id}_N)$.

Notation ($\mu_{\mathsf{pl},t}, \xi_t(q), \gamma(q)$). We will use $\mu_{\mathsf{pl},t}$ to denote the distribution of the pair $(H_{N,t}, \boldsymbol{x})$, $\xi_t(q) = \xi(q) + tq$ to refer to the mixture function of $H_{N,t}$, and $\gamma(q)$ to refer to the function $q\xi'_t(q)$.

In the subsequent sections, we will prove a high probability covariance bound for μ_t at a *fixed* time t under the planted model.

Lemma 7.18. *There exist universal constants c, T, K, such that for any $t \in [0, T]$, with probability at least $1 - e^{-cN^{1/5}}$ over the randomness of H_N drawn from $\mu_{\mathsf{pl},t}$, we have $\|\mathbf{Cov}(\mu_t)\| \leqslant K$.*

We now have all the necessary ingredients to prove the covariance bound along the entire localization path for the null model.

Proof of Lemma 7.6. Define $\mathcal{T}$ as the discrete set $\{iT/\delta : 1 \leqslant i \leqslant 1/\delta, i \in \mathbb{Z}\}$ for $\delta = N^{-100}$. We will prove:

$$\mathbf{Pr}_{H_N}\left[\mathbf{Pr}_{(\mu_t)|H_N}\left[\|\mathbf{Cov}(\mu_t)\|_{\mathsf{op}} < K \text{ for all } t \in \mathcal{T}\right] \geqslant 1 - e^{-cN^{1/5}}\right] \geqslant 1 - e^{-cN^{1/5}}.$$

A simple continuity argument can be used to derive the desired statement from the above. By taking a union bound over all elements of $\mathcal{T}$, along with Proposition 7.16 and Lemma 7.18, we can conclude:

$$\mathbf{E}_{H_N}\mathbf{Pr}_{(\mu_t)|H_N}\left[\|\mathbf{Cov}(\mu_t)\|_{\mathsf{op}} > K \text{ for some } t \in \mathcal{T}\right] \leqslant e^{-2cN^{-1/5}}.$$

The resulting statement then follows from Markov's inequality on the random variable

$$\mathbf{Pr}_{(\mu_t)|H_N}\left[\|\mathbf{Cov}(\mu_t)\|_{\mathsf{op}} > K \text{ for some } t \in \mathcal{T}\right]. \qquad \square$$

7.3 TAP planted models

In this section, we formally introduce the TAP planted model and relate it to the planted model from the previous section.

Definition 7.19. Let H_N be a planted Hamiltonian with mixture function ξ, and define

$$\theta(s) = \xi(1) - \xi(s) - (1-s)\xi'(s).$$

The associated *TAP free energy* is defined by

$$\mathcal{F}_{\mathsf{TAP}}(m) = H_N(m) + \frac{N}{2} \cdot \theta\left(R(m, m)\right) + \frac{N}{2} \cdot \log\left(1 - R(m, m)\right).$$

While the TAP free energy is interesting for a multitude of reasons, we will be interested in it because its fixed points provide a good proxy for the mean. Furthermore, the linearity of the TAP free energy in the Gaussian coefficients of the Hamiltonian provides certain desirable properties (that using the true mean would not allow).

Fact 7.20 ([HMP24, Fact 4.2]). *Let ξ be a mixture function satisfying the condition* (SL). *For any $t \in [0, \infty)$, let $\xi_t(q) = \xi(q) + tq$. Then there is a unique solution in $[0, 1)$, which we denote $q_* = q_*(t)$, to*

$$\xi'_t(q_*) = \frac{q_*}{1 - q_*}.$$

Lemma 7.21. *For any $K > 0$, sufficiently small (constant) $\iota > 0$ and $x \in S_N$:*

$$\mathbf{Pr}_{\mu_{\mathsf{pl},t}}\left[\left\|\mathbf{Cov}(\mu_{H_{N,t}})\right\|_{\mathsf{op}} \geqslant K\right]$$

$$\leqslant C \cdot \sup_{m \in \mathcal{S}_\iota} \mathbf{Pr}_{\mu_{\mathsf{pl},t}|x}\left[\left\|\mathbf{Cov}(\mu_{H_{N,t}})\right\|_{\mathsf{op}} \geqslant K \wedge \mathcal{E}_\iota \mid \nabla\mathcal{F}_{\mathsf{TAP}}(m) = 0\right]^{1/2} + 2e^{-cN},$$

where

$$\mathcal{S}_\iota = \mathcal{S}_\iota(x) := \left\{m \in \mathbb{R}^N : |R(m, m) - q_*|, |R(m, x) - q_*| < \iota\right\},$$

and $\mathcal{E}_\iota$ is the event that $\mathcal{F}_{\mathsf{TAP}}$ has a unique critical point m^{TAP} in $\mathcal{S}_\iota$, and that

$$\mathbf{Pr}_{\sigma \sim \mu_{H_{N,t}}}\left[R(\sigma, m^{\mathsf{TAP}}), R(\sigma, x) \in [q_* - \iota, q_* + \iota]\right] \geqslant 1 - e^{-cN}.$$

Proof. The above statement is effectively due to [HMP24, Propositions 4.4(d) and 4.5(a)]. For the reader's convenience, we include the steps to arriving at the above statement. For any event $\mathcal{E}$ (and in particular, for the event $\mathcal{E}$ defined in [HMP24, Proposition 4.4]), we have:

$$\mathbf{Pr}\left[\left\|\mathbf{Cov}(\mu_{H_{N,t}})\right\|_{\mathsf{op}} \geqslant K\right] \leqslant \mathbf{Pr}\left[\left\|\mathbf{Cov}(\mu_{H_{N,t}})\right\|_{\mathsf{op}} \geqslant K \wedge \mathcal{E}_\iota \wedge \mathcal{E}\right] + \mathbf{Pr}\left[\overline{\mathcal{E}}\right] + \mathbf{Pr}\left[\overline{\mathcal{E}_\iota}\right].$$

The desired statement follows by observing that $\mathbf{Pr}\left[\overline{\mathcal{E}_\iota}\right] \leqslant e^{-cN}$ by [HMP24, Proposition 4.5(a)], $\mathbf{Pr}\left[\overline{\mathcal{E}}\right] \leqslant e^{-cN}$ by [HMP24, Proposition 4.4], and applying [HMP24, Proposition 4.4(d)] with $X = \mathbf{1}\left[\left\|\mathbf{Cov}(\mu_{H_{N,t}})\right\|_{\mathsf{op}} \geqslant K \wedge \mathcal{E}_\iota\right]$. $\qquad\square$

Lemma 7.21 reduces our task to studying the covariance matrix in a *conditional* planted model.

Notation (μ_{TAP}, H_{TAP}, q_m, q_x). For $x \sim S_N$ and $m \in \mathbb{R}^N$, we consider the distribution $\mu_{\mathsf{TAP},x,m}$ of H_{TAP} for $H_{\mathsf{TAP}} \sim (\mu_{\mathsf{pl},t}|x, \nabla\mathcal{F}_{\mathsf{TAP}}(m) = 0)$. We use q_m and q_x to refer to $R(m,m)$ and $R(m,x)$ respectively.

Lemma 7.22. *Let $x \in S_N$, let S_ι be as in Lemma 7.21, and let $m \in S_\iota$. Then for an absolute constant $K > 0$,*

$$\mathbf{Pr}_{H_{\mathsf{TAP}} \sim \mu_{\mathsf{TAP},x,m}}[\|\mathbf{Cov}(\mu_{H_{\mathsf{TAP}}})\| \geqslant K \wedge \mathcal{E}_\iota] \leqslant e^{-cN^{1/5}}.$$

Proof of Lemma 7.18. The statement is immediate from Lemmas 7.21 and 7.22. □

The rest of this section is dedicated to proving Lemma 7.22. As a first step, we determine the law of the typical Hamiltonian sampled from μ_{TAP}. We prove the following lemma in Appendix B — it follows by routine calculations, using the form of the law of a Gaussian process conditioned on the value of a linear function of it. Recall that $\tilde{\xi}_t(q) = \xi(q) + tq$.

Lemma 7.23. *The law of Hamiltonian $H_{\mathsf{TAP}} \sim \mu_{\mathsf{TAP},x,m}$ is described by a Gaussian process $(H_{\mathsf{TAP}}(\sigma))_{\sigma \in S_N}$ defined by*

$$\mathbf{E}\, H_{\mathsf{TAP}}(\sigma) = N\tilde{\xi}_t(R(x,\sigma)) - \langle x, v(\sigma) \rangle \cdot \tilde{\xi}_t'(q_x) - \frac{\tilde{\xi}_t'(R(m,\sigma))}{\gamma'(q_m)} \cdot \langle m, \sigma \rangle \cdot \left(\theta'(q_m) - \frac{1}{1-q_m} \right)$$

$$\frac{1}{N}\mathbf{Cov}(H_{\mathsf{TAP}}(\sigma), H_{\mathsf{TAP}}(\sigma')) = \tilde{\xi}_t(R(\sigma,\sigma')) - R(\sigma,\sigma')\frac{\tilde{\xi}_t'(R(m,\sigma))\tilde{\xi}_t'(R(m,\sigma'))}{\tilde{\xi}_t'(q_m)}$$

$$+ \frac{\tilde{\xi}_t''(q_m)}{\gamma'(q_m)\tilde{\xi}_t'(q_m)}\gamma(R(m,\sigma))\gamma(R(m,\sigma')),$$

where

$$v(\sigma) := \frac{\tilde{\xi}_t'(R(m,\sigma))}{\tilde{\xi}_t'(q_m)}\left[I - \frac{\tilde{\xi}_t''(q_m)}{\gamma'(q_m)} \cdot \frac{mm^\top}{N} \right]\sigma$$

$$\gamma(q) := q \cdot \tilde{\xi}_t'(q).$$

For the proofs below, it will also be helpful to consider Hamiltonians with a linear term representing an external field. For a sequence $\gamma_1, \gamma_2, \ldots, \gamma_{p_*}$, consider the following generalization of H_N from (5):

$$H_N(\sigma) := \sum_{p \geqslant 1} \frac{\gamma_p}{N^{(p-1)/2}} \sum_{i_1,\ldots,i_p=1}^{N} g_{i_1,\ldots,i_p}\sigma_{i_1}\cdots\sigma_{i_p}. \tag{8}$$

This has mixture function

$$\xi(s) = \sum_{p \geqslant 1} \gamma_p^2 q^p.$$

We will write $\xi_{\sim 1}(s) = \sum_{p \geqslant 2} \gamma_p^2 q^p$ for the part of ξ with degree at least 2, and extend Condition 7.15 to such ξ as follows.

Condition 7.24 (ε-strict replica symmetry). We say ξ is ε-strictly replica symmetric if $\gamma_1^2 \leqslant N^{-4/5}$ and $\xi_{\sim 1}$ satisfies Condition 7.15.

7.4 Slices in TAP planted models

For succinctness, we shall fix $x \in S_N$ and $m \in S_\iota(x)$, and use μ_{TAP} to refer to the distribution $\mu_{\text{TAP},x,m}$. For $H_{\text{TAP}} \sim \mu_{\text{TAP}}$, we are interested in bounding the covariance of $\mu_{H_{\text{TAP}}}$. To reason about $\mu_{H_{\text{TAP}}}$, we write it as a mixture of distributions over $(N-2)$-dimensional slices of S_N. For $a, b \in \mathbb{R}$, we define

$$S(a,b) := \left\{ \sigma \in S_N : R(\sigma, m) = \left(1 + \frac{a}{\sqrt{N}} \right) q_m, \, R(\sigma, x) = \left(1 + \frac{b}{\sqrt{N}} \right) q_x \right\}.$$

Let $r_{a,b}$ refer to the radius of this slice, which is equal to

$$r_{a,b} = \sqrt{1 - q_m \left(1 + \frac{a}{\sqrt{N}} \right)^2 - \frac{q_m q_x^2}{q_m - q_x^2} \left(\frac{a - b}{\sqrt{N}} \right)^2}.$$

Note in particular that

$$q_m \left(1 + \frac{a}{\sqrt{N}} \right)^2 + \frac{q_m q_x^2}{q_m - q_x^2} \left(\frac{a - b}{\sqrt{N}} \right)^2 \geq q_m \left(1 + \frac{a}{\sqrt{N}} \right)^2. \tag{9}$$

We refer to the uniform distribution on this slice as $\rho_{a,b}$, and the partition function on the slice as

$$Z_{a,b} := \mathbf{E}_{\sigma \sim \rho_{a,b}} \exp(H_{\text{TAP}}(\sigma)).$$

With this definition, the partition function of the original Hamiltonian is given by

$$Z = \Lambda_N \int Z_{a,b} r_{a,b}^{N-4} \, \mathrm{d}(a,b)$$

for some fixed number Λ_N depending only on N.

Remark 7.25. To see why we scale by $r_{a,b}^{N-4}$, observe that when H_{TAP} is the constant-0 Hamiltonian, the resulting distribution on the sphere should be uniform. The distribution restricted to each slice must also be uniform. However, not all slices are weighted equally — slice that have smaller radii must be downweighted accordingly, with this weighting proportional to $r_{a,b}^{N-4}$ for $S(a,b)$.[1]

Use ν to refer to the distribution over (a,b) where $\mathrm{d}\nu(a,b) \propto Z_{a,b} r_{a,b}^{N-4} \, \mathrm{d}(a,b)$, and $\mu_{a,b}$ to refer to the distribution $\mu_{H_{\text{TAP}}}$ restricted to $S(a,b)$. Now, we can write $\mu_{H_{\text{TAP}}}$ as the following mixture:

$$\mu_{H_{\text{TAP}}} = \mathbf{E}_{(a,b) \sim \nu} \mu_{a,b}.$$

We will need coarse understanding of the tails of ν, and fine understanding of the distribution $\mu_{a,b}$ for small a and b.

Now, let us probe the distribution $\mu_{a,b}$.

$$\frac{\mathrm{d}\mu_{a,b}}{\mathrm{d}\rho_{a,b}}(\sigma) = \frac{\exp(H_{\text{TAP}}(\sigma))}{\mathbf{E}_{\sigma \sim \rho_{a,b}} \exp(H_{\text{TAP}}(\sigma))}$$

Since $S(a,b)$ can be naturally identified with S_{N-2}, the first step to understanding $\mu_{a,b}$ is to express it as a p-spin model on S_{N-2}. To do so, we will verify that *some* Hamiltonian that gives rise to $\mu_{a,b}$

---

[1] The constant of proportionality here is something depending only on N.

has a mixture function that is given by a polynomial in the overlap. We can write any $\sigma \in S(a,b)$ as:

$$\sigma = \sqrt{N} \cdot v(a,b) + \underbrace{\sqrt{1 - \|v(a,b)\|^2}}_{r_{a,b}} \sigma_\perp$$

for $\sigma_\perp \in S_N$ orthogonal to m and x, and for $v(a,b)$ in the span of m and x. Let Q be an isometric linear transformation that maps S_{N-2} to $S_N \perp \{m, x\}$. We can write $H_{\mathsf{TAP}}(\sigma) = H_{\mathsf{TAP}}(v(a,b) + r_{a,b}Q\tau)$ (where $\tau \in S_{N-2}$). The following is a consequence of Lemma 7.23, and is proved in Appendix B.

Corollary 7.26. *For a fixed choice of a and b, the Gaussian process* $(H_{\mathsf{TAP}}(v(a,b) + r_{a,b}Q\tau))_{\tau \in S_{N-2}}$ *is described by the following law.*

- *Let $H_{a,b}$ be a spherical p-spin Hamiltonian with mixture function $\xi_{a,b}$ given by:*

$$\xi_{a,b}(s) := \xi_t\left(\|v(a,b)\|^2 + r_{a,b}^2 s\right) - \xi_t\left(\|v(a,b)\|^2\right) - s \cdot \frac{r_{a,b}^2 \xi_t'\left(q_m \cdot \left(1 + \frac{a}{\sqrt{N}}\right)\right)^2}{\xi_t'(q_m)}.$$

- *Let $V(a,b) := \xi_t\left(\|v(a,b)\|^2\right) - \|v(a,b)\|^2 \cdot \frac{\xi_t'\left(\left(1 + \frac{a}{\sqrt{N}}\right)q_m\right)^2}{\xi_t'(q_m)} + \frac{\xi_t''(q_m)}{\gamma'(q_m)\xi_t'(q_m)} \cdot \gamma\left(\left(1 + \frac{a}{\sqrt{N}}\right)q_m\right)^2$.*

The law of $H_{\mathsf{TAP}}(v(a,b) + r_{a,b}Q\tau)$ is the same as that of $H_{a,b}(\tau) + \sqrt{N} \cdot g_{a,b} + \mathbf{E}_{\mu_{\mathsf{TAP}}} H_{\mathsf{TAP}}(v(a,b) + r_{a,b}Q\tau)$ where $g_{a,b}$ is a centered Gaussian of variance $V(a,b)$ independent of $H_{a,b}$.

Now, H_{TAP} is described by the collection $(H_{a,b}, g_{a,b})_{a,b}$. This is *not* an independent collection of random variables. The only structural properties of this collection we will use are:

- For each $a, b \in \mathbb{R}$, we have $H_{a,b}$ and $g_{a,b}$ are independent.

- For any a, b, we have $g_{a,b} = g_{0,0} + \hat{g}_{a,b}$, where $\hat{g}_{a,b}$ is a centered Gaussian of variance $O\left(\frac{a^4 + b^4}{N^2}\right)$.

We will first give an explicit form for $v(a,b)$.

Fact 7.27. *We have*

$$\sqrt{N} \cdot v(a,b) = m \cdot \left(1 + \frac{aq_m - bq_x^2}{\sqrt{N}(q_m - q_x^2)}\right) + \frac{q_m q_x}{q_m - q_x^2} x \cdot \left(\frac{b - a}{\sqrt{N}}\right).$$

Lemma 7.28. *There exists $\varepsilon = \varepsilon(\xi) > 0$ such that for all $t \geq 0$ and all $|a|, |b| \leq \varepsilon N^{1/10}$, the mixture function $\xi_{a,b}$ defined in Corollary 7.26 (recall this implicitly depends on t) is ε-strictly replica symmetric (Condition 7.24).*

Proof. We will first show $\xi_{a,b}'(0) \leq N^{-4/5}$. We calculate:

$$\xi_{a,b}'(0) = r_{a,b}^2 \xi_t'\left(\|v(a,b)\|^2\right) - \frac{r_{a,b}^2 \xi_t'\left(q_m\left(1 + \frac{a}{\sqrt{N}}\right)\right)^2}{\xi_t'(q_m)}$$

$$\leq \xi_t'\left(q_m + \frac{2aq_m}{\sqrt{N}} + O\left(\frac{a^2 + b^2}{N}\right)\right) - \frac{\xi_t'\left(q_m\left(1 + \frac{a}{\sqrt{N}}\right)\right)^2}{\xi_t'(q_m)}.$$

42

Here, the inequality follows from the fact that $\|v(a,b)\|^2 = q_m \left(1 + \frac{a}{\sqrt{N}}\right)^2 + O\left(\frac{a^2+b^2}{N}\right)$, ξ'_t is non-decreasing, and $r^2_{a,b} \leqslant 1$. Now, we have, using the $O(1)$-Lipschitzness of ξ'_t,

$$\xi'_t \left(\|v(a,b)\|^2\right) = \xi'_t(q_m) + \xi''_t(q_m) \cdot \frac{2aq_m}{\sqrt{N}} + O\left(\frac{a^2+b^2}{N}\right)$$

and

$$\xi'_t \left(q_m \left(1 + \frac{a}{\sqrt{N}}\right)\right)^2 = \xi'_t(q_m)^2 + \frac{2aq_m}{\sqrt{N}} \cdot \xi''_t(q_m) \cdot \xi'_t(q_m) + O\left(\xi'_t(q_m) \cdot \frac{a^2+b^2}{N}\right).$$

Thus $\check{\xi}'_{a,b}(0) = O(\frac{a^2+b^2}{N})$. Setting ε sufficiently small ensures $\check{\xi}'_{a,b}(0) \leqslant N^{-4/5}$. Next, we show $(\check{\xi}_{a,b})_{\sim 1}$ satisfies Condition 7.15. Since ξ satisfies (SL), there exists sufficiently small $\varepsilon = \varepsilon(\xi)$ such that $\xi''(q) \leqslant \frac{1-\varepsilon}{(1-q)^2}$ for all $q \in [0,1)$. Then,

$$\begin{aligned}
\check{\xi}''_{a,b}(q) &= \left(1 - \|v(a,b)\|^2\right)^2 \xi''_t\left(\|v(a,b)\|^2 + \left(1 - \|v(a,b)\|^2\right)q\right) \\
&\leqslant \left(1 - \|v(a,b)\|^2\right)^2 \cdot \frac{1-\varepsilon}{\left(1 - \|v(a,b)\|^2 - \left(1 - \|v(a,b)\|^2\right)q\right)^2} \\
&= \frac{1-\varepsilon}{(1-q)^2} \leqslant \frac{1}{(1-q)^2} - \varepsilon.
\end{aligned}$$

Integrating twice shows

$$(\check{\xi}_{a,b})_{\sim 1}(q) + q + \log(1-q) \leqslant \frac{1}{2}\varepsilon q^2. \qquad \square$$

We are now ready to bound $\mathbf{Cov}(\mu_{H_{\mathsf{TAP}}})$. First, recall that for any distribution μ over $\mathbb{R}^N$ and any vector $v \in \mathbb{R}^N$, we have $\mathbf{Cov}(\mu) \preceq \mathbf{E}_{\sigma \sim \mu}(\sigma - v)(\sigma - v)^\top$. Thus it suffices to bound

$$\begin{aligned}
&\mathbf{E}_{\sigma \sim \mu_{H_{\mathsf{TAP}}}}(\sigma - m)(\sigma - m)^\top \\
&= \mathbf{E}_{(a,b) \sim \nu} \mathbf{E}_{\sigma \sim \mu_{a,b}}(\sigma - v(a,b) + v(a,b) - m)(\sigma - v(a,b) + v(a,b) - m)^\top \\
&\preceq 2\mathbf{E}_{(a,b) \sim \nu} \mathbf{E}_{\sigma \sim \mu_{a,b}}(\sigma - v(a,b))(\sigma - v(a,b))^\top + 2\mathbf{E}_{(a,b) \sim \nu}(v(a,b) - m)(v(a,b) - m)^\top.
\end{aligned}$$
(10)

We will bound the spectral norm of each of the above terms below. We employ the following statement for proving the desired bounds, proved in Section 8.

Theorem 7.29. *Suppose $\varepsilon > 0$, and ξ is ε-strictly replica symmetric (Condition 7.24). Then, there exist $c = c(\varepsilon)$ and $C = C(\varepsilon)$ such that the following hold with probability $1 - e^{-cN^{1/5}}$.*

(a) We have $\nabla^2 H_N(0) \preceq (1 + \xi''(0) - \varepsilon^2/8) I_N$ and

$$\left| \log Z_N - \frac{N\xi(1)}{2} - \frac{N\xi''(0)}{4} - \frac{\log(1-\xi''(0))}{2} + \frac{1}{2}\log\det\left((1+\xi''(0))I_N - \nabla^2 H_N(0)\right) \right| \leqslant 1.$$

(b) The Gibbs measure satisfies $\left\|\mathbf{E}_{\mu_{H_N}}\sigma\sigma^\top\right\|_{\mathsf{op}} \leqslant C(1 + \gamma_1^2 N).$

Bounding the first term. Observe that:

$$\mathbf{E}_{\sigma \sim \mu_{a,b}}(\sigma - v(a,b))(\sigma - v(a,b))^\top = r_{a,b}^2 Q_{a,b} \mathbf{E}_{\tau \sim \mu_{H_{a,b}}} \tau \tau^\top Q_{a,b}^\top .$$

Thus,

$$\left\| \mathbf{E}_{\sigma \sim \mu_{a,b}}(\sigma - v(a,b))(\sigma - v(a,b))^\top \right\| = r_{a,b}^2 \left\| \mathbf{E}_{\tau \sim \mu_{H_{a,b}}} \tau \tau^\top \right\| \leqslant C r_{a,b}^2 (1 + a^2 + b^2). \tag{11}$$

where the inequality follows from Lemma 7.28 and Theorem 7.29.

Our next goal is to control the fluctuations of (a, b). By definition, if $\widehat{Z}_{a,b}$ is the partition function of the distribution with Hamiltonian $H_{a,b}$ defined by $H_{a,b}(\tau) = H_{\mathsf{TAP}}(v(a,b) + r_{a,b}Q\tau)$, we have

$$v(a,b) \propto \exp\left(\log \widehat{Z}_{a,b} + (N - 4) \log r_{a,b} + \mathbf{E}_{\mu_{\mathsf{TAP}}} H_{\mathsf{TAP}}(v(a,b)) + \sqrt{N} g_{a,b} \right). \tag{12}$$

We will show that $(a, b) \sim v$, conditioned on $|a|, |b| \leqslant \varepsilon N^{1/10}$ for ε as in Lemma 7.28, is subgaussian with variance $O(1)$. We will also show in Lemma 7.36 that v places very little mass outside the set $|a|, |b| \leqslant \varepsilon N^{1/10}$.

Lemma 7.30. *Let ε be as in Lemma 7.28. On an event with probability $1 - e^{-cN^{1/5}}$, the following holds. The density of (a, b) under v, conditioned on $|a|, |b| \leqslant \varepsilon N^{1/10}$, is given by*

$$v(a,b) \propto \exp\left(N\widehat{E}_{a,b} + \sqrt{N} g_{a,b} + \mathrm{Error}_{a,b}^{(1)} + \mathrm{Error}_{a,b}^{(2)} + \Delta_{a,b} \right),$$

where $|\Delta_{a,b}| \leqslant 1$ and

$$\widehat{E}_{a,b} = \frac{\xi_{a,b}(1)}{2} + \log r_{a,b} + \frac{1}{N} \mathbf{E}_{\mu_{\mathsf{TAP}}} H_{\mathsf{TAP}}(v(a,b)) - \frac{\xi_t(1)}{2}$$

$$= \frac{1}{2} \left(\log r_{a,b}^2 - \xi_t(\|v(a,b)\|^2) - r_{a,b}^2 \cdot \frac{\xi_t'\left(q_m \left(1 + \frac{a}{\sqrt{N}} \right) \right)^2}{\xi_t'(q_m)} \right)$$

$$- \gamma(q_x) \cdot \frac{\xi_t'\left(q_m \left(1 + \frac{a}{\sqrt{N}} \right) \right)}{\xi_t'(q_m)} \cdot \left(\left(1 + \frac{b}{\sqrt{N}} \right) - q_m \cdot \frac{\xi_t''(q_m)}{\gamma'(q_m)} \cdot \left(1 + \frac{a}{\sqrt{N}} \right) \right)$$

$$+ \xi_t \left(q_x \left(1 + \frac{b}{\sqrt{N}} \right) \right) + \frac{\gamma\left(q_m \left(1 + \frac{a}{\sqrt{N}} \right) \right)}{\gamma'(q_m)} \cdot \left((1 - q_m) \xi_t''(q_m) + \frac{1}{1 - q_m} \right).$$

for the error terms

$$\mathrm{Error}_{a,b}^{(1)} = \left(\log \widehat{Z}_{a,b} - \frac{N\xi_{a,b}(1)}{2} - \frac{N\xi_{a,b}''(0)}{4} + \frac{1}{2} \log \det \left((1 + \xi_{a,b}''(0))\mathrm{Id} - \nabla^2 H_N(v(a,b)) \right) \right) - 4 \log r_{a,b}$$

and

$$\mathrm{Error}_{a,b}^{(2)} = \frac{N\xi_{a,b}''(0)}{4} - \frac{1}{2} \log \det \left((1 + \xi_{a,b}''(0))\mathrm{Id} - \nabla^2 H_N(v(a,b)) \right)$$

The above follows by expanding out all the terms in the expression (12) for the density of v and evaluating the term $\log \widehat{Z}_{a,b}$ using Theorem 7.29, which applies because the models $\xi_{a,b}$ for $|a|, |b| \leqslant \varepsilon N^{1/10}$ are ε-strictly replica symmetric by Lemma 7.28. Because the conclusion of Theorem 7.29 holds with probability $1 - e^{-cN^{1/5}}$, we may evaluate $\log \widehat{Z}_{a,b}$ over a $1/\mathrm{poly}(N)$-net of such (a, b) via a union bound, and then infer the estimate for all such a, b by a standard continuity argument.

Lemma 7.31. $\nabla \widehat{E}_{a,b}\big|_{(a,b)=(0,0)} = 0$.

Lemma 7.32. *There exist constants $\eta, \varepsilon > 0$ such that for all $|a|, |b| \leqslant \varepsilon\sqrt{N}$, $N\nabla^2 \widehat{E}_{a,b} \preceq -\eta\mathrm{Id}$.*

The above follow from routine calculations, which we defer to Appendix B.

Lemma 7.33. *For every constant $\iota > 0$, there is a constant c such that with probability $1 - e^{-cN}$, for all a, b, we have $|g_{a,b} - g_{0,0}| \leqslant \iota\frac{a^2+b^2}{\sqrt{N}}$.*

Lemma 7.34. *With probability $1 - e^{-cN^{1/5}}$, $|\mathrm{Error}_{a,b}^{(1)}| = O(1)$ uniformly for all $|a|, |b| < \varepsilon N^{1/10}$, for ε as in Lemma 7.28.*

Proof. We shall show this very high probability bound for a fixed a, b. Constructing a net over the relevant a, b and performing a union bound over this net allows us to extend this to a uniform bound for all a, b; we omit the details. We may write the error term as

$$\mathrm{Error}_{a,b}^{(1)} = \log \widehat{Z}_{a,b} - \frac{N\xi_{a,b}(1)}{2} - \frac{\log\left(1 - \xi_{a,b}''(0)\right)}{2}$$
$$- \frac{N\xi_{a,b}''(0)}{4} + \frac{1}{2}\log\det\left((1 + \xi_{a,b}''(0))\mathrm{Id} - \nabla^2 H_{a,b}(0)\right) + O(1).$$

Due to the bound on a and b, the above is $O(1)$ with very high probability by Theorem 7.29(a). The desideratum follows. □

Lemma 7.35. *For any sufficiently small $\iota > 0$, with probability at least $1 - e^{-cN}$, $\left|\mathrm{Error}_{a,b}^{(2)} - \mathrm{Error}_{0,0}^{(2)}\right| = O(1)$ for all $a, b < \iota N^{1/4}$.*

We relegate the proof of the above to the appendix Appendix B. The idea of the proof is that the Hessian $\nabla^2 H_{a,b}(0)$ does not deviate too much for small variations in a, b – the first order terms in the deviation end up being cancelled by the $\xi_{a,b}''(0)/2$ term, while the second order terms are $O(1)$.

Lemma 7.36. *Let $\mathcal{E}_\iota$ be as in Lemma 7.21. With probability $1 - e^{-cN^{1/5}}$, either $\mathcal{E}_\iota$ does not hold, or the following holds. For ε as in Lemma 7.28,*

$$\mathbf{Pr}_{(a,b)\sim\nu}\left[|a| \leqslant \varepsilon N^{1/10} \text{ and } |b| \leqslant \varepsilon N^{1/10}\right] \geqslant 1 - e^{-cN^{1/5}}.$$

Proof. On the event $\mathcal{E}_\iota$, we have

$$\mathbf{Pr}_{(a,b)\sim\nu}\left[|a| > \iota N^{1/2} \text{ or } |b| > \iota N^{1/2}\right] \leqslant e^{-cN}.$$

Thus, let

$$T = \left\{(a, b) \in \mathbb{R}^2 : |a| \in [\varepsilon N^{1/10}, \iota N^{1/2}] \text{ or } |b| \in [\varepsilon N^{1/10}, \iota N^{1/2}]\right\}.$$

It suffices to show that $\mathbf{Pr}_{(a,b)\sim\nu}[(a, b) \in T] \leqslant e^{-cN^{1/5}}$ with probability $1 - e^{-cN^{1/5}}$. Recall the density of $(a, b) \sim \nu$ is given by (12), and that $\mathbf{E}\widehat{Z}_{a,b} = e^{N\xi_{a,b}(1)/2}$. Thus, for $\widehat{\mathbf{E}}$ denoting expectation with respect to the $\widehat{Z}_{a,b}$ alone,

$$\widehat{\mathbf{E}}\int_T \exp\left(\log\widehat{Z}_{a,b} + (N-4)\log r_{a,b} + \mathbf{E}_{\mu_{\mathrm{TAP}}}H_{\mathrm{TAP}}(v(a,b)) + \sqrt{N}g_{a,b}\right)\mathrm{d}(a,b)$$

$$= \int_T \exp\left(N\widehat{E}_{a,b} + \frac{N\check{\xi}_t(1)}{2} + \sqrt{N}g_{a,b} - 4\log r_{a,b}\right) \mathrm{d}(a,b).$$

On the event in Lemma 7.33, we have, for any constant $\iota > 0$,

$$\sqrt{N}g_{a,b} \leqslant \sqrt{N}g_{0,0} + \iota(a^2 + b^2).$$

By Lemma 7.32,

$$N\widehat{E}_{a,b} \leqslant N\widehat{E}_{0,0} - \eta(a^2 + b^2).$$

Combining shows that

$$N\widehat{E}_{a,b} + \sqrt{N}g_{a,b} - 4\log r_{a,b} \leqslant N\widehat{E}_{0,0} + \sqrt{N}g_{0,0} - \frac{\eta}{2}(a^2 + b^2) + O(1).$$

Combining shows

$$\widehat{\mathbf{E}} \int_T \exp\left(\log\widehat{Z}_{a,b} + (N-4)\log r_{a,b} + \mathbf{E}_{\mu_{\mathsf{TAP}}} H_{\mathsf{TAP}}(v(a,b)) + \sqrt{N}g_{a,b}\right) \mathrm{d}(a,b)$$
$$\leqslant e^{-cN^{1/5}} \exp\left(N\widehat{E}_{0,0} + \frac{N\check{\xi}_t(1)}{2} + \sqrt{N}g_{0,0}\right)$$

and therefore with probability $1 - e^{-cN^{1/5}/2}$ over the $\widehat{Z}_{a,b}$,

$$\int_T \exp\left(\log\widehat{Z}_{a,b} + (N-4)\log r_{a,b} + \mathbf{E}_{\mu_{\mathsf{TAP}}} H_{\mathsf{TAP}}(v(a,b)) + \sqrt{N}g_{a,b}\right) \mathrm{d}(a,b)$$
$$\leqslant e^{-cN^{1/5}/2} \exp\left(N\widehat{E}_{0,0} + \frac{N\check{\xi}_t(1)}{2} + \sqrt{N}g_{0,0}\right) \tag{13}$$

On the other hand, Lemma 7.30 implies that with probability $1 - e^{-cN^{1/5}}$,

$$\log\widehat{Z}_{0,0} + (N-4)\log r_{0,0} + \mathbf{E}_{\mu_{\mathsf{TAP}}} H_{\mathsf{TAP}}(v(0,0)) + \sqrt{N}g_{0,0}$$
$$= N\widehat{E}_{0,0} + \frac{N\check{\xi}_t(1)}{2} + \sqrt{N}g_{0,0} + \mathrm{Error}_{0,0}^{(1)} + \mathrm{Error}_{0,0}^{(2)} + O(1),$$

and Lemma 7.34 implies $|\mathrm{Error}_{0,0}^{(1)}| = O(1)$ with probability $1 - e^{-cN^{1/5}}$. Furthermore, Lemma 8.3 below implies that $|\mathrm{Error}_{0,0}^{(2)}| \leqslant N^{1/10}$ with probability $1 - e^{-cN^{1/5}}$. Thus

$$\log\widehat{Z}_{0,0} + (N-4)\log r_{0,0} + \mathbf{E}_{\mu_{\mathsf{TAP}}} H_{\mathsf{TAP}}(v(0,0)) + \sqrt{N}g_{0,0}$$
$$\geqslant N\widehat{E}_{0,0} + \frac{N\check{\xi}_t(1)}{2} + \sqrt{N}g_{0,0} - 2N^{1/10},$$

and standard continuity arguments imply that for $T' = \{(a,b) : |a|, |b| \leqslant N^{-10}\}$,

$$\int_{T'} \exp\left(\log\widehat{Z}_{a,b} + (N-4)\log r_{a,b} + \mathbf{E}_{\mu_{\mathsf{TAP}}} H_{\mathsf{TAP}}(v(a,b)) + \sqrt{N}g_{a,b}\right) \mathrm{d}(a,b)$$
$$\geqslant e^{-3N^{1/10}} \exp\left(N\widehat{E}_{0,0} + \frac{N\check{\xi}_t(1)}{2} + \sqrt{N}g_{0,0}\right).$$

Comparing with (13) implies the conclusion, after adjusting c. $\qquad\square$

Lemma 7.37. *With probability $1 - e^{-cN^{1/5}}$, either $\mathcal{E}_\iota$ does not hold or the following holds. There exists a random variable X over $\mathbb{R}^2$ (coupled with v) such that the following holds for $(a, b) \sim v$.*

(a) With probability at least $1 - e^{-cN^{1/5}}$, $X = (a, b)$.

(b) X has mean $O(1)$ and is $O(1)$-subgaussian.

Proof. This is an immediate corollary of Lemmas 7.31 to 7.36, setting X to be the random variable that is equal to (a, b) if $|a|, |b| < \varepsilon N^{1/10}$, and 0 otherwise. $\qquad\square$

We are now finally prepared to bound $\mathbf{Cov}(\mu)$.

Lemma 7.22. *Let $x \in S_N$, let $\mathcal{S}_\iota$ be as in Lemma 7.21, and let $m \in \mathcal{S}_\iota$. Then for an absolute constant $K > 0$,*

$$\mathbf{Pr}_{H_{\mathsf{TAP}} \sim \mu_{\mathsf{TAP},x,m}} [\|\mathbf{Cov}(\mu_{H_{\mathsf{TAP}}})\| \geqslant K \wedge \mathcal{E}_\iota] \leqslant e^{-cN^{1/5}}.$$

Proof. Note that $|a|, |b| \leqslant 2\sqrt{N}$ almost surely. Let X be as in Lemma 7.37. This lemma implies that with probability at least $1 - e^{-cN^{1/5}}$ over the randomness of the Hamiltonian,

$$\mathbf{E}_{(a,b)\sim v}\left[a^2 + b^2\right] = \mathbf{E}[\|X\|^2] + \mathbf{E}_{(a,b)\sim v}\left[\mathbf{1}[X \neq (a,b)](a^2 + b^2)\right]$$
$$\leqslant O(1) + \mathbf{Pr}(X \neq (a,b)) \cdot 8N = O(1).$$

Thus, by plugging in (11) along with this observation into (10), we get that the following holds with probability at least $1 - e^{-cN^{1/5}}$

$$\|\mathbf{Cov}(\mu)\| \leqslant 2C\mathbf{E}_{(a,b)\sim v}(1 + a^2 + b^2) + \|2\mathbf{E}_{(a,b)\sim v}(v(a,b) - m)(v(a,b) - m)^\top\|$$
$$\leqslant O(1) + 2\mathbf{E}_{(a,b)\sim v}\|v(a,b) - m\|^2$$
$$\leqslant O(1) + 2\mathbf{E}_{(a,b)\sim v}\left[O(a^2 + b^2)\right]$$
$$\leqslant O(1). \qquad\qquad\square$$

8 High-probability covariance bound of replica symmetric spherical spin glass

In this section we prove the main technical input to the proofs in Subsection 7.4. This takes the form of a high-probability bound on the partition function and covariance matrix (in fact, second moment matrix) of a spherical spin glass in the replica symmetric phase.

In this section, we let H_N be defined as in (8), with a linear term corresponding to an external field:

$$H_N(\sigma) := \sum_{p \geqslant 1} \frac{\gamma_p}{N^{(p-1)/2}} \sum_{i_1,\dots,i_p=1}^N g_{i_1,\dots,i_p} \sigma_{i_1} \cdots \sigma_{i_p}.$$

We recall $\xi_{\sim 1}(q) = \sum_{p \geqslant 2} \gamma_p^2 q^p$ denotes the part of ξ without the linear term, and let

$$\xi_{\sim 2}(q) = \gamma_1^2 q + \sum_{p \geqslant 3} \gamma_p^2 q^p$$

denote the part of ξ excluding the degree 2 term.

The results in this section hold under the following condition, which we restate for reference.

Condition 7.24 (ε-strict replica symmetry). We say ξ is ε-strictly replica symmetric if $\gamma_1^2 \leqslant N^{-4/5}$ and $\xi_{\sim 1}$ satisfies Condition 7.15.

Throughout this section, we treat $\varepsilon > 0$ as a constant and let $O_\varepsilon(1)$ denote a quantity bounded depending on ε.

Theorem 7.29. *Suppose $\varepsilon > 0$, and ξ is ε-strictly replica symmetric (Condition 7.24). Then, there exist $c = c(\varepsilon)$ and $C = C(\varepsilon)$ such that the following hold with probability $1 - e^{-cN^{1/5}}$.*

(a) We have $\nabla^2 H_N(0) \preceq (1 + \xi''(0) - \varepsilon^2/8)I_N$ and

$$\left| \log Z_N - \frac{N\xi(1)}{2} - \frac{N\xi''(0)}{4} - \frac{\log(1 - \xi''(0))}{2} + \frac{1}{2}\log\det\left((1 + \xi''(0))I_N - \nabla^2 H_N(0)\right) \right| \leqslant 1.$$

(b) The Gibbs measure satisfies $\|\mathbf{E}_{\mu_{H_N}} \sigma\sigma^\top\|_{\mathsf{op}} \leqslant C(1 + \gamma_1^2 N)$.

In the below proofs, we allow the constants c and C to change from line to line, but they will always be uniform in ε. We always set C sufficiently large depending on ε, and then c sufficiently small depending on ε, C.

Theorem 7.29 will be proved through the following pair of propositions. We introduce the degree-2 Hamiltonian

$$H_{N,2}(\sigma) := \frac{\gamma_2}{N^{1/2}} \sum_{i_1,i_2=1}^{N} g_{i_1,i_2}\sigma_{i_1}\sigma_{i_2} = \frac{1}{2}\langle\nabla^2 H_N(0)\sigma, \sigma\rangle. \tag{14}$$

Similarly let $H_{N,\sim 2}(\sigma) = H_N(\sigma) - H_{N,2}(\sigma)$ be the non degree-2 part of $H_N(\sigma)$. Define the degree-2 Gibbs measure and partition function by

$$d\mu_{H_{N,2}}(\sigma) = \frac{\exp(H_{N,2}(\sigma))}{Z_{N,2}}\,d\rho(\sigma), \qquad Z_{N,2} = \int_{S_N} \exp(H_{N,2}(\sigma))\,d\rho(\sigma).$$

Throughout this section, we will let $\mathbf{E}$ denote expectation with respect to the disorder coefficients $g_{i_1,\ldots,i_p}$, while $\langle\cdot\rangle$ denotes averaging with respect to $\sigma \sim \mu_{H_N}$ (or several i.i.d. samples $\sigma^1, \sigma^2, \ldots$ from this measure). Similarly, let $\langle\cdot\rangle_2$ denote Gibbs average with respect to $\mu_{H_{N,2}}$.

Note that $\nabla^2 H_N(0)$ depends on H_N only through $H_{N,2}$.

Proposition 8.1 (Concentration of degree-2 partition function; proved in Subsection 8.1). *With probability $1 - e^{-cN}$ over $H_{N,2}$, we have $\nabla^2 H_N(0) \preceq (1 + \xi''(0) - \varepsilon^2/8)I_N$ and*

$$\left| \log Z_{N,2} - \frac{N\xi''(0)}{2} - \frac{\log(1 - \xi''(0))}{2} + \frac{1}{2}\log\det\left((1 + \xi''(0))I - \nabla^2 H_N(0)\right) \right| \leqslant 1/2.$$

The following is proved in Subsections 8.2 and 8.3.

Proposition 8.2. *There is a $H_{N,2}$-measurable event with probability $1 - e^{-cN^{1/5}}$ on which the following holds with probability $1 - e^{-cN^{1/5}}$ over $H_{N,\sim 2}$.*

1. The partition functions $Z_N, Z_{N,2}$ satisfy

$$\left| \log \frac{Z_N}{Z_{N,2}} - \frac{N\xi_{\sim 2}(1)}{2} \right| \leqslant 1/2.$$

48

2. *The Gibbs measure satisfies $\|\langle \sigma\sigma^\top \rangle\|_{\mathsf{op}} \leqslant C(1 + \gamma_1^2 N)$.*

Proof of Theorem 7.29. Immediate from Propositions 8.1 and 8.2, since $\xi(1) = \xi_{\sim 2}(1) + \frac{1}{2}\xi''(0)$. □

We also show the following concentration of the log determinant in Theorem 7.29.

Lemma 8.3 (Proved in Subsection 8.1). *There exists a $H_{N,2}$-measurable random variable X that the following holds.*

1. *With probability $1 - e^{-cN}$, $X = \log\det\big((1 + \xi''(0))I - \nabla^2 H_N(0)\big) - N\xi''(0)/2$.*

2. *X has mean $O_\varepsilon(1)$ and is $O_\varepsilon(1)$-subgaussian.*

This implies the quantitative contiguity between the planted and null models, which we restate below for convenience.

Proposition 7.16 (Quantitative contiguity). *Under Condition 7.15, there exists $c = c(\varepsilon) > 0$ such that for any event $\mathcal{E}$, if $\mu_{\mathsf{pl}}(\mathcal{E}) = p$, then $\mu_{\mathsf{null}}(\mathcal{E}) \leqslant e^{-cN^{1/5}} + e^{\frac{1}{c}\sqrt{\log\frac{2}{p}}}p$.*

Proof. Let $\mathcal{E}_{\mathsf{good}}$ be intersection of the event in Theorem 7.29, the event

$$X = \log\det\big((1 + \xi''(0))I - \nabla^2 H_N(0)\big) - N\xi''(0)/2$$

from Lemma 8.3, and the event $X \leqslant t$, for some $t > 0$ to be determined. Then, after adjusting $c = c(\varepsilon)$ as necessary,

$$\mu_{\mathsf{null}}(\overline{\mathcal{E}_{\mathsf{good}}}) \leqslant e^{-cN^{1/5}} + \mathbb{P}(X > t) \leqslant e^{-cN^{1/5}} + e^{-c(t - \frac{1}{c})_+^2}.$$

Note that $\log \mathbf{E} Z_N = N\xi(1)/2$, while on the event $\mathcal{E}_{\mathsf{good}}$,

$$\log Z_N = \frac{N\xi(1) - X + O_\varepsilon(1)}{2} \geqslant \frac{N\xi(1) - t - \frac{1}{c}}{2}.$$

Thus $\frac{\mathbf{E} Z_N}{Z_N} \leqslant e^{\frac{1}{2}(t + \frac{1}{c})}$. So,

$$\mu_{\mathsf{null}}(\mathcal{E}) \leqslant \mu_{\mathsf{null}}(\overline{\mathcal{E}_{\mathsf{good}}}) + \int \frac{\mathbf{E} Z_N}{Z_N}\mathbf{1}[H_N \in \mathcal{E} \cap \mathcal{E}_{\mathsf{good}}]\,\mathrm{d}\mu_{\mathsf{pl}}(H_N)$$
$$\leqslant e^{-cN^{1/5}} + e^{-c(t - \frac{1}{c})_+^2} + e^{\frac{1}{2}(t + \frac{1}{c})}p.$$

We then take $t = \frac{1}{c} + \sqrt{\frac{1}{c}\log\frac{1}{p}}$, so that this is bounded by

$$e^{-cN^{1/5}} + \left(1 + e^{\frac{1}{c} + \sqrt{\frac{1}{c}\log\frac{1}{p}}}\right)p.$$

Further adjusting c proves the desired bound. □

8.1 Concentration of degree-2 partition function

We write $A = \frac{1}{2}\nabla^2 H_N(0) = \frac{\sqrt{\xi''(0)}}{2} M$. It is straightforward to check that M is distributed as a sample from $\mathrm{GOE}(N)$.

Fact 8.4. *We have $\xi''(0) \leqslant 1 - \varepsilon$.*

Proof. Writing (7) as

$$\xi_{\sim 1}(q) + q + \log(1 - q) \leqslant -\varepsilon q^2/2$$

and Taylor expanding around $q = 0$ implies the result. $\qquad\square$

Fact 8.5. *With probability $1 - e^{-cN}$, $\nabla^2 H_N(0) \preceq (1 + \xi''(0) - \varepsilon^2/8)I$.*

Proof. With probability $1 - e^{-cN}$, we have $\lambda_{\max}(M) \leqslant 2 + \varepsilon^2/8$. Then,

$$
\begin{aligned}
\lambda_{\max}\left((1 + \xi''(0) - \varepsilon^2/8)I - \nabla^2 H_N(0)\right) &\geqslant 1 + \xi''(0) - \varepsilon^2/8 - \sqrt{\xi''(0)}(2 + \varepsilon^2/8) \\
&= (1 - \sqrt{\xi''(0)})^2 - \varepsilon^2(1 + \sqrt{\xi''(0)})/8 \\
&> \varepsilon^2/4 - \varepsilon^2/4 = 0
\end{aligned}
$$

by Fact 8.4. $\qquad\square$

For $\gamma \in (\lambda_{\max}(A), +\infty)$, define

$$G(\gamma) = \gamma - \frac{1}{2N} \log \det(\gamma I - A). \tag{15}$$

Note that

$$G'(\gamma) = 1 - \frac{1}{2N} \mathrm{Tr}(\gamma I - A)^{-1}$$

is continuous and increasing, with $\lim_{\gamma \downarrow \lambda_{\max}(A)} G'(\gamma) = -\infty$ and $\lim_{\gamma \uparrow +\infty} G'(\gamma) = 1$. Thus G' has a unique root γ_* in $(\lambda_{\max}(A), +\infty)$. The following lemma is a consequence of [HMP24, Lemma 7.3], which is proved by an analysis of a Laplace transform of the free energy also used in [BL16].

Lemma 8.6. *With probability $1 - e^{-cN}$ over $H_{N,2}$,*

$$Z_{N,2} = (1 + O(N^{-c}))\sqrt{\frac{2}{G''(\gamma_*)}}(2e)^{-N/2}\exp(NG(\gamma_*)). \tag{16}$$

Proof. Recalling (14), we have

$$H_{N,2}(\sigma) = \frac{\sqrt{\xi''(0)}}{2}\langle M\sigma, \sigma \rangle,$$

and Fact 8.4 implies the factor $\sqrt{\xi''(0)}/2$ is bounded away from $1/2$. Then [HMP24, Lemma 7.3] (with $u = 0$) implies the result. $\qquad\square$

Define $\gamma_0 = (1 + \xi''(0))/2$. The next lemma shows that, although the variable γ_* in (16) is random, we may approximate it deterministically by γ_0.

Lemma 8.7. *For sufficiently large C depending on ε, and sufficiently small c depending on ε, C, with probability $1 - e^{-cN}$ the following holds for all $\gamma \in [\gamma_0 - N^{-1/2}, \gamma_0 + N^{-1/2}]$.*

1. $|G'(\gamma_0)| \leqslant 1/(C\sqrt{N})$.

2. $\left|\frac{G''(\gamma)}{2/(1-\zeta''(0))} - 1\right| \leqslant N^{-1/3}$.

Proof. Let $\mathrm{d}\rho_{\mathrm{sc}}(x) = \frac{1}{2\pi}\mathbf{1}[|x| \leqslant 2]\sqrt{4-x^2}\,\mathrm{d}x$ denote Wigner's semicircle law, and

$$f_1(x) = 1 - \frac{1}{1+\zeta''(0) - 2\sqrt{\zeta''(0)}x}, \qquad f_2(x) = \frac{2}{(1+\zeta''(0)-2\sqrt{\zeta''(0)}x)^2}.$$

For $k \in [2]$, let

$$L_k = \int f_k(x)\,\mathrm{d}\rho_{\mathrm{sc}}(x).$$

We will show that with probability $1 - e^{-cN}$, for each $k \in [2]$,

$$\left|G^{(k)}(\gamma_0) - L_k\right| \leqslant \frac{1}{C\sqrt{N}}. \tag{17}$$

Recall that $M \sim \mathrm{GOE}(N)$. For $f : \mathbb{R} \to \mathbb{R}$, define the spectral trace

$$\mathrm{Tr}\, f(M) = \sum_{i=1}^{N} f(\lambda_i(M)).$$

Note that $G^{(k)}(\gamma_0) = N^{-1} \cdot \mathrm{Tr}\, f_k(M)$. Define

$$\widetilde{f}_k(x) = f_k(\min(x, 2 + \varepsilon^2/8)).$$

By the proof of Fact 8.5, $1 + \zeta''(0) - 2\sqrt{\zeta''(0)}x \geqslant \varepsilon^2/8$ for $x \leqslant 2 + \varepsilon^2/8$, so $\widetilde{f}_k$ is $O_\varepsilon(1)$-Lipschitz. Moreover, $\lambda_{\max}(M) \leqslant 2 + \varepsilon^2/8$ with probability $1 - e^{-cN}$, and on this event $\mathrm{Tr}\, f_k(M) = \mathrm{Tr}\, \widetilde{f}_k(M)$.

By [GZ00, Lemma 1.2(b)], if we write $M_{i,i} = \sqrt{2/N}Z_{i,i}$, $M_{i,j} = \sqrt{1/N}Z_{i,j}$, then $\mathrm{Tr}\, \widetilde{f}_k(M)$ is a $O_\varepsilon(1)$-Lipschitz function of the standard gaussians $(Z_{i,j})_{1\leqslant i\leqslant j\leqslant N}$. Thus $\mathrm{Tr}\, \widetilde{f}_k(M)$ is $O_\varepsilon(1)$-subgaussian, i.e.

$$\mathbb{P}(|\mathrm{Tr}\, \widetilde{f}_k(M) - \mathbf{E}\,\mathrm{Tr}\, \widetilde{f}_k(M)| \geqslant t) \leqslant 2e^{-t^2/C}$$

for some $C = O_\varepsilon(1)$. By [BY05, Theorem 1.1],

$$\mathrm{Tr}\, \widetilde{f}_k(M) - NL_k$$

converges in distribution to a gaussian with mean and variance $O_\varepsilon(1)$. Combined with subgaussianity of $\mathrm{Tr}\, \widetilde{f}_k(M)$, this implies

$$|\mathbf{E}\,\mathrm{Tr}\, \widetilde{f}_k(M) - NL_k| = O_\varepsilon(1).$$

It follows that (after possibly increasing $C = O_\varepsilon(1)$),

$$\mathbb{P}(|\mathrm{Tr}\, \widetilde{f}_k(M) - NL_k| \geqslant t) \leqslant 2e^{-(t-C)_+^2/C}.$$

Thus

$$\mathbb{P}(|G^{(k)} - L_k| \geqslant t) \leqslant \mathbb{P}(\mathrm{Tr}\, f_k(M) \neq \mathrm{Tr}\, \widetilde{f}_k(M)) + \mathbb{P}(|\mathrm{Tr}\, \widetilde{f}_k(M) - NL_k| \geqslant Nt)$$
$$\leqslant e^{-cN} + 2e^{-(Nt-C)_+^2/C}.$$

Plugging in $t = 1/(C\sqrt{N})$ proves (17). Next, direct calculations show $L_1 = 0$, $L_2 = \frac{2}{1-\xi''(0)}$. The former directly implies conclusion (1), and the latter implies

$$\left| G''(\gamma_0) - \frac{2}{1 - \xi''(0)} \right| \leqslant \frac{1}{C\sqrt{N}}.$$

Moreover, on the probability $1 - e^{-cN}$ event that $\lambda_{\max}(M) \leqslant 2 + \varepsilon^2/8$, $G^{(3)}(\gamma) = O_\varepsilon(1)$ for all $\gamma \in [\gamma_0 - N^{-1/2}, \gamma_0 + N^{-1/2}]$. This implies the conclusion (2). $\qquad\square$

Lemma 8.3 is proved by the same method, and we present the proof here.

Proof of Lemma 8.3. Let

$$f_0(x) = \log\left(1 + \xi''(0) - \sqrt{\xi''(0)}x\right).$$

An elementary calculation shows that

$$L_0 := \int f_0(x) \, d\rho_{sc}(X) = \xi''(0)/2.$$

Proceeding as in the above proof, we have

$$\log \det\left((1 + \xi''(0))I - \sqrt{\xi''(0)}\nabla^2 H_N(0) \right) = \operatorname{Tr} f_0(M).$$

If we take $\widetilde{f}_0(x) = f_0(\min(x, 2 + \varepsilon^2/8))$, then $\operatorname{Tr} f_0(M) = \operatorname{Tr} \widetilde{f}_0(M)$ with probability $1 - e^{-cN}$. The same proof shows $\operatorname{Tr} \widetilde{f}_0(M)$ is $O_\varepsilon(1)$-subgaussian, and

$$\left|\mathbf{E} \operatorname{Tr} \widetilde{f}_0(M) - NL_0\right| = O_\varepsilon(1).$$

Thus we may take $X = \operatorname{Tr} \widetilde{f}_0(M) - NL_0 = \operatorname{Tr} \widetilde{f}_0(M) - N\xi''(0)/2$. $\qquad\square$

Proof of Proposition 8.1. The assertion $\nabla^2 H_N(0) \preceq (1 + \xi''(0) - \varepsilon^2/8)I_N$ is proved in Fact 8.5. Suppose the events in Lemmas 8.6 and 8.7 occur. Since γ_* is the solution to $G'(\gamma_*) = 0$, we have

$$|\gamma_0 - \gamma_*| \leqslant |G'(\gamma_0)| \cdot \frac{1 - \xi''(0)}{2(1 - N^{-1/3})} \leqslant \frac{1}{C\sqrt{N}}.$$

So,

$$N|G(\gamma_0) - G(\gamma_*)| \leqslant \frac{N}{2}|\gamma_0 - \gamma_*|^2 \sup_{\gamma \in [\gamma_0 - N^{-1/2}, \gamma_0 + N^{-1/2}]} G''(\gamma) \leqslant \frac{1}{C^2} \cdot \frac{2(1 + N^{-1/3})}{1 - \xi''(0)} \leqslant \frac{3}{C^2\varepsilon}.$$

Moreover, $\xi''(\gamma_0)/\xi''(\gamma_*) = 1 + O(N^{-1/3})$. Combining with Lemma 8.6 shows that, for some Δ satisfying $|\Delta| \leqslant \frac{3}{C^2\varepsilon}$,

$$Z_{N,2} = (1 + O(N^{-c}))e^\Delta \sqrt{\frac{2}{G''(\gamma_0)}} (2e)^{-N/2} \exp(NG(\gamma_0))$$

$$= (1 + O(N^{-c}))e^\Delta \sqrt{1 - \xi''(0)} (2e)^{-N/2} \exp(N\gamma_0) \det\left(\gamma_0 I - \frac{1}{2}\nabla^2 H_N(0)\right)^{-1/2}$$

$$= (1 + O(N^{-c}))e^\Delta \sqrt{1 - \xi''(0)} \exp\left(N\xi''(0)/2\right) \det\left((1 + \xi''(0))I - \nabla^2 H_N(0)\right)^{-1/2}.$$

Taking a logarithm and setting C sufficiently large concludes the proof. $\qquad\square$

Finally, the following concentration estimates for samples from $\mu_{H_{N,2}}$ will be useful in the sequel. This is proved similarly to [HMP24, Lemma 7.5], and we defer the proof to Appendix B.3.

Let $v_1, \ldots, v_N$ denote the (unit) eigenvectors of $\nabla^2 H_N(0)$. These are well defined on the almost sure event that all eigenvalues of $\nabla^2 H_N(0)$ have multiplicity 1.

Proposition 8.8. *With probability $1 - e^{-cN}$ over $H_{N,2}$, the following holds. Let $\sigma, \sigma^1, \sigma^2 \sim \mu_{H_{N,2}}$, and let $W = \langle \sigma, v_i \rangle$ for any $i \in [N]$, or $W = \langle \sigma^1, \sigma^2 \rangle / \sqrt{N}$. Then:*

1. *For any $0 \leqslant t \leqslant N^{1/5}$, $\mathbb{P}(|W| \geqslant t) \leqslant 3e^{-ct^2}$.*

2. *For any $k \in 2\mathbb{N}$, there exists $C_k > 0$ independent of N such that $\langle W^k \rangle_2 \leqslant C_k$.*

In particular, part (2) implies $\|\langle \sigma\sigma^\top \rangle_2\|_{\mathrm{op}} \leqslant C$.

8.2 Conditional positive probability bounds for non degree-2 part

In this subsection, we prove the following propositions, which establish a weaker version of Proposition 8.2 with positive instead of high probability.

Proposition 8.9. *There is a $H_{N,2}$-measurable event with probability $1 - e^{-cN^{1/5}}$ on which, with probability $1 - N^{-1/15}$ over $H_{N,\sim 2}$,*

$$\left| \log \frac{Z_N}{Z_{N,2}} - \frac{N\xi_{\sim 2}(1)}{2} \right| = O(N^{-1/15}).$$

Proposition 8.10. *There is a $H_{N,2}$-measurable event with probability $1 - e^{-cN^{1/5}}$ on which, with probability $1/2$ over $H_{N,\sim 2}$,*

$$\left\| \frac{Z_N}{Z_{N,2}e^{N\xi_{\sim 2}(1)/2}} \langle \sigma\sigma^\top \rangle - \langle \sigma\sigma^\top \rangle_2 \right\|_F^2 \leqslant C(1 + \gamma_1^4 N^2).$$

In conjunction with Propositions 8.8 and 8.9, the above immediately implies a positive probability bound on the second moment matrix $\langle \sigma\sigma^\top \rangle$.

Both propositions rely on the following truncation to Z_N developed in [HS23], which allows one to estimate Z_N via the second moment method throughout the strictly RS regime. As shown in the following lemma, this truncation does not significantly affect the first moment; at the same time, it will force the second moment to be dominated by pairs of nearly-orthogonal points.

Lemma 8.11. *The following holds for sufficiently small $c > 0$ depending on ε. Let*

$$T = T(H_N) := \left\{ \sigma \in S_N : \int_{S_N} \mathbf{1}[|R(\sigma,\tau)| \geqslant N^{-2/5}]e^{H_N(\tau)}\,d\rho(\tau) \leqslant e^{N\xi(1)/2 - cN^{1/5}} \right\}.$$

Then, we have:

$$\mathbf{E}\int_{S_N} \mathbf{1}[\sigma \notin T]e^{H_N(\sigma)}\,d\rho(\sigma) \leqslant e^{N\xi(1)/2 - cN^{1/5}}, \tag{18}$$

$$\mathbf{E}\int_{S_N} \mathbf{1}[\sigma \notin T]e^{H_{N,2}(\sigma)}\,d\rho(\sigma) \leqslant e^{N\xi''(0)/4 - cN^{1/5}}, \tag{19}$$

$$\mathbf{E}\int_{S_N} \mathbf{1}[\sigma^1 \notin T, |R(\sigma^1,\sigma^2)| \leqslant 3N^{-2/5}]e^{H_N(\sigma^1)+H_N(\sigma^2)}\,d\rho^{\otimes 2}(\sigma^1,\sigma^2) \leqslant e^{N\xi(1) - cN^{1/5}}, \tag{20}$$

$$\mathbf{E}\int_{S_N} \mathbf{1}[\sigma^1 \notin T, |R(\sigma^1,\sigma^2)| \leqslant 3N^{-2/5}]e^{H_{N,2}(\sigma^1)+H_N(\sigma^2)}\,d\rho^{\otimes 2}(\sigma^1,\sigma^2) \leqslant e^{N\xi(1)/2 + N\xi''(0)/4 - cN^{1/5}}. \tag{21}$$

The proof of the above lemma is very similar to [HS23, Proposition 3.1] and [HMP24, Lemma 7.9], and we defer it to Appendix B.3. As a corollary, we can get control on the first two moments of Z_N with respect to the randomness in $H_{N,\sim 2}$. To this end, let $\mathbf{E}_{\sim 2}$ denote expectation with respect to $H_{N,\sim 2}$.

Corollary 8.12. *There is a $H_{N,2}$-measurable event with probability $1 - e^{-cN^{1/5}}$ on which the following holds. For*

$$\widetilde{T} = \widetilde{T}(H_{N,\sim 2}) := \left\{ \sigma \in S_N : \langle \mathbf{1}[|R(\sigma,\tau)| \geqslant N^{-2/5}] e^{H_{N,\sim 2}(\tau)} \rangle_2 \leqslant e^{N\xi_{\sim 2}(1)/2 - cN^{1/5}} \right\}, \tag{22}$$

where the Gibbs average is with respect to $\tau \sim \langle \cdot \rangle_2$, we have

$$\mathbf{E}_{\sim 2} \left\langle \mathbf{1}[\sigma \notin \widetilde{T}] e^{H_{N,\sim 2}(\sigma)} \right\rangle_2 \leqslant e^{N\xi_{\sim 2}(1)/2 - cN^{1/5}},$$

$$\mathbf{E}_{\sim 2} \left\langle \mathbf{1}[\sigma \notin \widetilde{T}] \right\rangle_2 \leqslant e^{-cN^{1/5}},$$

$$\mathbf{E}_{\sim 2} \left\langle \mathbf{1}[\sigma^1 \notin \widetilde{T}, |R(\sigma^1,\sigma^2)| \leqslant 3N^{-2/5}] e^{H_{N,\sim 2}(\sigma^1) + H_{N,\sim 2}(\sigma^2)} \right\rangle_2 \leqslant e^{N\xi_{\sim 2}(1) - cN^{1/5}},$$

$$\mathbf{E}_{\sim 2} \left\langle \mathbf{1}[\sigma^1 \notin \widetilde{T}, |R(\sigma^1,\sigma^2)| \leqslant 3N^{-2/5}] e^{H_{N,\sim 2}(\sigma^2)} \right\rangle_2 \leqslant e^{N\xi_{\sim 2}(1)/2 - cN^{1/5}}.$$

Proof. By Proposition 8.1 and Lemma 8.3, with probability $1 - e^{-cN^{2/5}}$ over $H_{N,2}$,

$$Z_{N,2} \geqslant e^{N\xi''(0)/4 - cN^{1/5}/2}.$$

On this event, for $\sigma \in T$ where T is as in Lemma 8.11,

$$\left\langle \mathbf{1}[|R(\sigma,\tau)| \geqslant N^{-2/5}] e^{H_{N,\sim 2}(\tau)} \right\rangle_2 = \frac{1}{Z_{N,2}} \int_{S_N} \mathbf{1}[|R(\sigma,\tau)| \geqslant N^{-2/5}] e^{H_N(\tau)} \, d\rho(\tau)$$

$$\leqslant e^{N\xi_{\sim 2}(1)/2 - cN^{1/5}/2}.$$

Here we recall $\xi(1)/2 - \xi''(0)/4 = \xi_{\sim 2}(1)/2$. So, $\sigma \in \widetilde{T}(H_{N,\sim 2}, c/2)$, where this denotes $\widetilde{T}$ defined with $c/2$ in place of c. Therefore $T \subseteq \widetilde{T}(H_{N,\sim 2}, c/2)$.

By Markov's inequality and Lemma 8.11, with probability $1 - e^{-cN^{1/5}/4}$ over $H_{N,2}$,

$$\mathbf{E}_{\sim 2} \int_{S_N} \mathbf{1}[\sigma \notin T] e^{H_N(\sigma)} \, d\rho(\sigma) \leqslant e^{N\xi(1)/2 - 3cN^{1/5}/4}$$

On the intersection of these events,

$$\mathbf{E}_{\sim 2} \left\langle \mathbf{1}[\sigma \notin \widetilde{T}(H_{N,\sim 2}, c/2)] e^{H_{N,\sim 2}(\sigma)} \right\rangle_2 \leqslant \mathbf{E}_{\sim 2} \left\langle \mathbf{1}[\sigma \notin T] e^{H_{N,\sim 2}(\sigma)} \right\rangle_2$$

$$= \frac{1}{Z_{N,2}} \mathbf{E}_{\sim 2} \int_{S_N} \mathbf{1}[\sigma \notin T] e^{H_N(\sigma)} \, d\rho(\sigma)$$

$$\leqslant e^{N\xi_{\sim 2}(1)/2 - cN^{1/5}/4}.$$

The first conclusion follows by adjusting c, and the other two conclusions follow similarly. $\square$

For the rest of this subsection, we condition on a realization of $H_{N,2}$ satisfying the following good event.

Definition 8.13. Let E_2 denote the $H_{N,2}$-measurable event that the events in Proposition 8.8 and Corollary 8.12 hold. This occurs with probability $1 - e^{-cN^{1/5}}$.

We can now prove Proposition 8.9.

Proof of Proposition 8.9. Let $\widetilde{T}$ be as in Corollary 8.12. We can write

$$\frac{Z_N}{Z_{N,2}} = \langle e^{H_{N,\sim 2}(\sigma)} \rangle_2 = X_1 + X_2, \tag{23}$$

where

$$X_1 = \left\langle \mathbf{1}[\sigma \in \widetilde{T}] e^{H_{N,\sim 2}(\sigma)} \right\rangle_2, \qquad\qquad X_2 = \left\langle \mathbf{1}[\sigma \notin \widetilde{T}] e^{H_{N,\sim 2}(\sigma)} \right\rangle_2.$$

We will show that that X_2 is much smaller than $\mathbf{E}_{\sim 2} X_1$ with high probability, and then control the fluctuations of X_1. For all $\sigma \in S_N$, $\mathbf{E}_{\sim 2}[e^{H_{N,\sim 2}(\sigma)}] = e^{N\xi_{\sim 2}(1)/2}$, so Corollary 8.12 implies

$$(1 - e^{-cN^{1/5}}) e^{N\xi_{\sim 2}(1)/2} \leqslant \mathbf{E}_{\sim 2}[X_1] \leqslant e^{N\xi_{\sim 2}(1)/2}. \tag{24}$$

On the other hand, by Corollary 8.12 and Markov's inequality, with probability $1 - e^{-cN^{1/5}/2}$ over $H_{N,\sim 2}$,

$$X_2 \leqslant e^{N\xi_{\sim 2}(1)/2 - cN^{1/5}/2}, \tag{25}$$

so $X_2 \leqslant e^{-cN^{1/5}/2} \mathbf{E}_{\sim 2} X_1$, as desired.

We now control the fluctuations of X_1 by estimating

$$\mathbf{Var}_{\sim 2}[X_1] := \mathbf{E}_{\sim 2}[X_1^2] - \mathbf{E}_{\sim 2}[X_1]^2.$$

Then, for $\sigma^1, \sigma^2 \sim \mu_{H_{N,2}}$,

$$\mathbf{E}_{\sim 2}[X_1^2] = \mathbf{E}_{\sim 2} \left\langle \mathbf{1}[\sigma^1, \sigma^2 \in \widetilde{T}] e^{H_{N,\sim 2}(\sigma^1) + H_{N,\sim 2}(\sigma^2)} \right\rangle_2 \leqslant \mathbf{E}_{\sim 2}[Y_1] + \mathbf{E}_{\sim 2}[Y_2],$$

where

$$Y_1 = \left\langle \mathbf{1}[|R(\sigma^1, \sigma^2)| \leqslant N^{-2/5}] e^{H_{N,\sim 2}(\sigma^1) + H_{N,\sim 2}(\sigma^2)} \right\rangle_2,$$
$$Y_2 = \left\langle \mathbf{1}[\sigma^1 \in \widetilde{T}, |R(\sigma^1, \sigma^2)| \geqslant N^{-2/5}] e^{H_{N,\sim 2}(\sigma^1) + H_{N,\sim 2}(\sigma^2)} \right\rangle_2. \tag{26}$$

By the definition of $\widetilde{T}$ and (24),

$$\mathbf{E}_{\sim 2}[Y_2] \leqslant \mathbf{E}_{\sim 2} \left\langle \mathbf{1}[\sigma^1 \in \widetilde{T}] e^{H_{N,\sim 2}(\sigma^1)} \right\rangle_2 e^{N\xi_{\sim 2}(1)/2 - cN^{1/5}} \leqslant e^{N\xi_{\sim 2}(1) - cN^{1/5}}. \tag{27}$$

We further calculate

$$\mathbf{E}_{\sim 2}[Y_1] = e^{N\xi_{\sim 2}(1)} \left\langle \mathbf{1}[|R(\sigma^1, \sigma^2)| \leqslant N^{-2/5}] e^{N\xi_{\sim 2}(R(\sigma^1, \sigma^2))} \right\rangle_2.$$

Recall that in Theorem 7.29, we assumed $\gamma_1^2 \leqslant N^{-4/5}$. Thus, for $|R| \leqslant N^{-2/5}$,

$$\xi_{\sim 2}(R) = \gamma_1^2 R + O(R^3) = O(N^{-6/5}).$$

It follows that $\mathbf{E}_{\sim 2}[Y_1] \leqslant (1 + O(N^{-1/5})) e^{N\xi_{\sim 2}(1)}$.

Combining the above estimates shows $\mathbf{E}_{\sim 2}[X_1^2] \leqslant (1 + O(N^{-1/5}))e^{N\xi_{\sim 2}(1)}$. Further combining with the lower bound in (24) shows

$$\mathbf{Var}_{\sim 2}[X_1] = O(N^{-1/5})e^{N\xi_{\sim 2}(1)}.$$

By Chebyshev's inequality, with probability $1 - N^{-1/15}/2$,

$$|X_1 - \mathbf{E}_{\sim 2}[X_1]| = O(N^{-1/15})e^{N\xi_{\sim 2}(1)/2}$$

Union bounding with the event in (25), and recalling (24), we conclude that with probability $1 - N^{-1/15}$,

$$\frac{Z_N}{Z_{N,2}} = (1 + O(N^{-1/15}))e^{N\xi_{\sim 2}(1)/2}. \qquad \square$$

The above proof also implies the following estimate, which will be useful in the sequel.

Corollary 8.14. *On the event E_2 (Definition 8.13), with probability $1 - e^{-cN^{1/5}}$ over $H_{N,\sim 2}$,*

$$\int \mathbf{1}[|R(\sigma^1, \sigma^2)| \geqslant N^{-2/5}]e^{H_N(\sigma^1) + H_N(\sigma^2)} \, d\rho^{\otimes 2}(\sigma^1, \sigma^2) \leqslant Z_{N,2}^2 e^{N\xi_{\sim 2}(1) - cN^{1/5}}.$$

Proof. Dividing through by $Z_{N,2}^2$, it suffices to show, for $\sigma^1, \sigma^2 \sim \mu_{H_{N,2}}$,

$$\left\langle \mathbf{1}[|R(\sigma^1, \sigma^2)| \geqslant N^{-2/5}]e^{H_{N,\sim 2}(\sigma^1) + H_{N,\sim 2}(\sigma^2)} \right\rangle_2 \leqslant e^{N\xi_{\sim 2}(1) - cN^{1/5}}.$$

The left-hand side is bounded by $Y_2 + Y_3$, where Y_2 is as in (26) and

$$Y_3 := \left\langle \mathbf{1}[\sigma^1 \notin \widetilde{T}]e^{H_{N,\sim 2}(\sigma^1) + H_{N,\sim 2}(\sigma^2)} \right\rangle_2 = \left\langle \mathbf{1}[\sigma \notin \widetilde{T}]e^{H_{N,\sim 2}(\sigma)} \right\rangle_2 \left\langle e^{H_{N,\sim 2}(\sigma)} \right\rangle_2.$$

By (27), $\mathbf{E}_{\sim 2}[Y_2] \leqslant e^{N\xi_{\sim 2}(1) - cN^{1/5}}$. By Corollary 8.12,

$$\mathbf{E}_{\sim 2} \left\langle \mathbf{1}[\sigma \notin \widetilde{T}]e^{H_{N,\sim 2}(\sigma)} \right\rangle_2 \leqslant e^{N\xi_{\sim 2}(1)/2 - cN^{1/5}}, \qquad \mathbf{E}_{\sim 2}\langle e^{H_{N,\sim 2}(\sigma)} \rangle_2 \leqslant e^{N\xi_{\sim 2}(1)/2}.$$

So, the following estimates each hold with probability $1 - e^{-cN^{1/5}/4}$ over $H_{N,\sim 2}$:

$$Y_2 \leqslant e^{N\xi_{\sim 2}(1) - cN^{1/5}/2}, \qquad \left\langle \mathbf{1}[\sigma \notin \widetilde{T}]e^{H_{N,\sim 2}(\sigma)} \right\rangle_2 \leqslant e^{N\xi_{\sim 2}(1)/2 - 3cN^{1/5}/4},$$

$$\langle e^{H_{N,\sim 2}(\sigma)} \rangle_2 \leqslant e^{N\xi_{\sim 2}(1)/2 + cN^{1/5}/4}.$$

The conclusion follows on the intersection of these events, after adjusting c. $\qquad \square$

We now turn to the proof of Proposition 8.10. By rotational invariance of gaussians, we may assume $\nabla^2 H_N(0)$ is diagonal while keeping the law of $H_{N,\sim 2}$ unchanged. For $i, j \in [N]$, define

$$X_{i,j} = \left\langle \sigma_i \sigma_j \left(e^{H_{N,\sim 2}(\sigma) - N\xi_{\sim 2}(1)/2} - \mathbf{1}[i = j] \right) \right\rangle_2,$$

and note that this equals the (i, j) entry of the matrix appearing in Proposition 8.10. For $\sigma \in \mathbb{R}^N$ and $i \in [N]$, let $\sigma_{\sim i} \in \mathbb{R}^{N-1}$ denote σ with coordinate i omitted. Similarly, for $i \neq j$, let $\sigma_{\sim i,j} \in \mathbb{R}^{N-2}$

denote σ with coordinates i and j omitted, and by slight abuse of notation let $\sigma_{\sim i,i} = \sigma_{\sim i}$. For $i, j \in [N]$ (possibly with $i = j$) define analogously to $\widetilde{T}$

$$\widetilde{T}_{i,j} := \left\{ \sigma \in S_N : \left\langle \mathbf{1}[|R(\sigma_{\sim i,j}, \tau_{\sim i,j})| \geqslant 2N^{-2/5}] e^{H_{N,\sim 2}(\tau)} \right\rangle_2 \leqslant e^{N\xi_{\sim 2}(1)/2 - cN^{1/5}} \right\}, \qquad (28)$$

where we recall the Gibbs average is with respect to $\tau \sim \langle \cdot \rangle_2$. Then define

$$\widetilde{X}_{i,j} = \left\langle \mathbf{1}[|\sigma_i|, |\sigma_j| \leqslant \log N, \sigma \in \widetilde{T}_{i,j}] \sigma_i \sigma_j \left(e^{H_{N,\sim 2}(\sigma) - N\xi_{\sim 2}(1)/2} - \mathbf{1}[i = j] \right) \right\rangle_2$$

$$\widehat{X}_{i,j} = \left\langle \mathbf{1}[|\sigma_i^1|, |\sigma_j^1|, |\sigma_i^2|, |\sigma_j^2| \leqslant \log N, |R(\sigma_{\sim i,j}^1, \sigma_{\sim i,j}^2)| \leqslant 2N^{-2/5}] \right.$$

$$\left. \sigma_i^1 \sigma_j^1 \sigma_i^2 \sigma_j^2 \left(e^{H_{N,\sim 2}(\sigma^1) - N\xi_{\sim 2}(1)/2} - \mathbf{1}[i = j] \right) \left(e^{H_{N,\sim 2}(\sigma^2) - N\xi_{\sim 2}(1)/2} - \mathbf{1}[i = j] \right) \right\rangle_2.$$

Note that $\widehat{X}_{i,j}$ is the contribution to $X_{i,j}^2$ coming from σ^1, σ^2 that are both not localized to coordinate i or j and have small overlap. The following two lemmas reduce the task of controlling $X_{i,j}^2$ to bounding $\mathbf{E}_{\sim 2}\widehat{X}_{i,j}$. They are proved by manipulating the typicality truncations $\widetilde{T}$ and $\widetilde{T}_{i,j}$ similarly to the proofs above; we defer these proofs to Appendix B.3.

Lemma 8.15. *For each $i, j \in [N]$, with probability $1 - e^{-c\log^2 N}$ over $H_{N,\sim 2}$,*

$$X_{i,j}^2 \leqslant 2\widetilde{X}_{i,j}^2 + e^{-c\log^2 N}.$$

Lemma 8.16. *For each $i, j \in [N]$,*

$$\mathbf{E}_{\sim 2}\widetilde{X}_{i,j}^2 \leqslant \mathbf{E}_{\sim 2}\widehat{X}_{i,j} + e^{-cN^{1/5}}.$$

We now turn to bounding the $\mathbf{E}_{\sim 2}\widehat{X}_{i,j}$. This is achieved by the following pair of propositions.

Proposition 8.17. *For any $i \in [N]$, we have $\mathbf{E}_{\sim 2}\widehat{X}_{i,i} \leqslant C(N\gamma_1^4 + N^{-1})$.*

Proposition 8.18. *For any distinct $i, j \in [N]$, we have $\mathbf{E}_{\sim 2}\widehat{X}_{i,j} \leqslant C(\gamma_1^4 + N^{-2})$.*

Throughout the next two proofs, $\langle \cdot \rangle_2$ denotes expectation w.r.t. $\sigma^1, \sigma^2 \sim \langle \cdot \rangle_2$, and we write $R = R(\sigma^1, \sigma^2)$, $R_{\sim i} = R(\sigma_{\sim i}^1, \sigma_{\sim i}^2)$, and $R_{\sim i,j} = R(\sigma_{\sim i,j}^1, \sigma_{\sim i,j}^2)$.

Proof of Proposition 8.17. By direct calculation,

$$\mathbf{E}_{\sim 2}\widehat{X}_{i,i} = \mathbf{E}_{\sim 2}\left\langle \mathbf{1}[|\sigma_i^1|, |\sigma_i^2| \leqslant \log N, |R_{\sim i}| \leqslant 2N^{-2/5}] \right.$$

$$\left. (\sigma_i^1)^2 (\sigma_i^2)^2 \left(e^{H_{N,\sim 2}(\sigma^1) - N\xi_{\sim 2}(1)/2} - 1 \right) \left(e^{H_{N,\sim 2}(\sigma^2) - N\xi_{\sim 2}(1)/2} - 1 \right) \right\rangle_2$$

$$= \left\langle \mathbf{1}[|\sigma_i^1|, |\sigma_i^2| \leqslant \log N, |R_{\sim i}| \leqslant 2N^{-2/5}] (\sigma_i^1)^2 (\sigma_i^2)^2 \left(e^{N\xi_{\sim 2}(R)} - 1 \right) \right\rangle_2.$$

In view of Proposition 8.8, σ_i^1 and σ_i^2 are subgaussian of scale $O(1)$ and R is subgaussian of scale $O(N^{-1/2})$. We will see that the above integral is dominated by $|\sigma_i^1| \asymp |\sigma_i^2| \asymp 1$ and $|R| \asymp N^{-1/2}$, in

57

which case Taylor expanding $e^{N\xi_{\sim 2}(R)}$ shows this integral has the desired scale. Formally, we can write the above integral as $Y_{i,i}^{(1)} + Y_{i,i}^{(2)}$, where

$$Y_{i,i}^{(1)} = \left\langle \mathbf{1}[|\sigma_i^1|, |\sigma_i^2| \leqslant \log N, |R_{\sim i}| \leqslant 2N^{-1/2}\log N](\sigma_i^1)^2(\sigma_i^2)^2 \left(e^{N\xi_{\sim 2}(R)} - 1\right)\right\rangle_2$$

$$Y_{i,i}^{(2)} = \left\langle \mathbf{1}[|\sigma_i^1|, |\sigma_i^2| \leqslant \log N, 2N^{-1/2}\log N \leqslant |R_{\sim i}| \leqslant 2N^{-2/5}](\sigma_i^1)^2(\sigma_i^2)^2 \left(e^{N\xi_{\sim 2}(R)} - 1\right)\right\rangle_2$$

are the contributions from $|R_{\sim i}|$ smaller and larger than $2N^{-1/2}\log N$.

We first address $Y_{i,i}^{(2)}$. Note that on the event in the indicator in $Y_{i,i}^{(2)}$, $|R| \leqslant 3N^{-2/5}$. Thus, as $\gamma_1^2 \leqslant N^{-4/5}$,

$$N|\xi_{\sim 2}(R)| \leqslant N\gamma_1^2 R + O(NR^3) = O(N^{-1/5}).$$

It follows that $|e^{N\xi_{\sim 2}(R)} - 1| \leqslant 1$. Then, by Cauchy–Schwarz,

$$\begin{aligned}
Y_{i,i}^{(2)} &\leqslant \left\langle \mathbf{1}[|R| \geqslant N^{-1/2}\log N](\sigma_i^1)^2(\sigma_i^2)^2\right\rangle_2 \\
&\leqslant \left\langle \mathbf{1}[|R| \geqslant N^{-1/2}\log N]\right\rangle_2^{1/2} \left\langle (\sigma_i^1)^2(\sigma_i^2)^2\right\rangle_2^{1/2} \\
&\leqslant e^{-c\log^2 N} \cdot O(1) = e^{-c\log^2 N},
\end{aligned} \tag{29}$$

where we have used Proposition 8.8(1) for the tail probability and Proposition 8.8(2) for the coordinate moments.

Next we turn to $Y_{i,i}^{(1)}$. On the event in the indicator in $Y_{i,i}^{(1)}$, $|R| \leqslant 3N^{-1/2}\log N$, so

$$N|\xi_{\sim 2}(R)| \leqslant N\gamma_1^2 R + O(NR^3) = O(N^{-3/10}\log N),$$

where we recall $\gamma_1^2 \leqslant N^{-4/5}$. Thus, Taylor expanding the exponential and $\xi_{\sim 2}$,

$$\begin{aligned}
Y_{i,i}^{(1)} &= \left\langle \mathbf{1}[|\sigma_i^1|, |\sigma_i^2| \leqslant \log N, |R_{\sim i}| \leqslant 2N^{-1/2}\log N](\sigma_i^1)^2(\sigma_i^2)^2 \right. \\
&\qquad \left. \left(N\xi_{\sim 2}(R) + \tfrac{1}{2}N^2\xi_{\sim 2}(R)^2 + \tfrac{1}{6}N^3\xi_{\sim 2}(R)^3\right)\right\rangle_2 + O(N^{-6/5}\log^8 N) \\
&= \left\langle \mathbf{1}[|\sigma_i^1|, |\sigma_i^2| \leqslant \log N, |R_{\sim i}| \leqslant 2N^{-1/2}\log N](\sigma_i^1)^2(\sigma_i^2)^2 \right. \\
&\qquad \left. \left(N(\gamma_1^2 R + \gamma_3^2 R^3 + \gamma_4^2 R^4) + \tfrac{1}{2}N^2(\gamma_1^2 R + \gamma_3^2 R^3)^2 + \tfrac{1}{6}N^3\gamma_1^6 R^3\right)\right\rangle_2 + O(N^{-11/10}\log^9 N).
\end{aligned}$$

By exchangeability of $(\sigma_i^1, -\sigma_i^1)$, $(\sigma_i^2, -\sigma_i^2)$, and $(R_{\sim i}, -R_{\sim i})$, all the odd degree in R terms vanish, leaving

$$Y_{i,i}^{(1)} = \tfrac{1}{2}N^2\gamma_1^4 Q_2 + (N\gamma_4^2 + N^2\gamma_1^2\gamma_3^2)Q_4 + \tfrac{1}{2}N^2\gamma_3^4 Q_6 + o(N^{-1}),$$

and where we have introduced the notation

$$Q_k := \left\langle \mathbf{1}[|\sigma_i^1|, |\sigma_i^2| \leqslant \log N, |R_{\sim i}| \leqslant 2N^{-1/2}\log N](\sigma_i^1)^2(\sigma_i^2)^2 R^k\right\rangle_2.$$

By Cauchy–Schwarz and Proposition 8.8, for each $k \in \{2, 4, 6\}$,

$$Q_k \leqslant \left\langle (\sigma_i^1)^4(\sigma_i^2)^4\right\rangle_2^{1/2} \langle R^{2k}\rangle_2^{1/2} = O(N^{-k/2}).$$

This implies $Y_{i,i}^{(1)} \leqslant C(N\gamma_1^4 + N^{-1})$. Combining with the bound (29) on $Y_{i,i}^{(2)}$ implies the result. $\qquad \square$

Proof of Proposition 8.18. We calculate as above

$$\mathbf{E}_{\sim 2}\widehat{X}_{i,j} \leqslant \mathbf{E}_{\sim 2}\Big\langle \mathbf{1}[|\sigma_i^1|, |\sigma_j^1|, |\sigma_i^2|, |\sigma_j^2| \leqslant \log N, |R_{\sim i,j}| \leqslant 2N^{-2/5}]$$

$$\sigma_i^1 \sigma_j^1 \sigma_i^2 \sigma_j^2 e^{H_{N,\sim 2}(\sigma^1) + H_{N,\sim 2}(\sigma^2) - N\xi_{\sim 2}(1)}\Big\rangle_2$$

$$= \Big\langle \mathbf{1}[|\sigma_i^1|, |\sigma_j^1|, |\sigma_i^2|, |\sigma_j^2| \leqslant \log N, |R_{\sim i,j}| \leqslant 2N^{-2/5}]\sigma_i^1 \sigma_j^1 \sigma_i^2 \sigma_j^2 e^{N\xi_{\sim 2}(R)}\Big\rangle_2.$$

Our strategy for evaluating this will be similar as above, except that because this integral contains $\sigma_i^1 \sigma_j^1 \sigma_i^2 \sigma_j^2$ instead of $(\sigma_i^1)^2 (\sigma_i^2)^2$, we will need to expand the exponential more carefully to obtain cancellations in these terms. Formally, we write the above integral as $Y_{i,j}^{(1)} + Y_{i,j}^{(2)}$ for

$$Y_{i,j}^{(1)} = \Big\langle \mathbf{1}[|\sigma_i^1|, |\sigma_j^1|, |\sigma_i^2|, |\sigma_j^2| \leqslant \log N, |R_{\sim i,j}| \leqslant 2N^{-1/2}\log N]\sigma_i^1 \sigma_j^1 \sigma_i^2 \sigma_j^2 e^{N\xi_{\sim 2}(R)}\Big\rangle_2,$$

$$Y_{i,j}^{(2)} = \Big\langle \mathbf{1}[|\sigma_i^1|, |\sigma_j^1|, |\sigma_i^2|, |\sigma_j^2| \leqslant \log N, 2N^{-1/2}\log N \leqslant |R_{\sim i,j}| \leqslant 2N^{-2/5}]\sigma_i^1 \sigma_j^1 \sigma_i^2 \sigma_j^2 e^{N\xi_{\sim 2}(R)}\Big\rangle_2.$$

Identically to the previous proof, on the event in the indicator in $Y_{i,j}^{(2)}$ we have $N|\xi_{\sim 2}(R)| = O(N^{-1/5})$, so $e^{N\xi_{\sim 2}(R)} \leqslant 2$. Then, by Cauchy–Schwarz and Proposition 8.8,

$$|Y_{i,j}^{(2)}| \leqslant 2\Big\langle \mathbf{1}[|R| \geqslant N^{-1/2}\log N]|\sigma_i^1 \sigma_j^1 \sigma_i^2 \sigma_j^2|\Big\rangle_2$$

$$\leqslant 2\langle \mathbf{1}[|R| \geqslant N^{-1/2}\log N]\rangle_2^{1/2} \Big\langle (\sigma_i^1)^4 (\sigma_i^2)^4 \Big\rangle_2^{1/4} \Big\langle (\sigma_j^1)^4 (\sigma_j^2)^4 \Big\rangle_2^{1/4}$$

$$\leqslant 2e^{-c\log^2 N} \cdot O(1) \cdot O(1) = e^{-c\log^2 N}.$$

To address $Y_{i,j}^{(1)}$, define $\Delta_i = \sigma_i^1 \sigma_i^2/N$ and $\Delta_j = \sigma_j^1 \sigma_j^2/N$, the contributions to R coming from the ith and jth coordinate, respectively. Then, by exchangeability of $(\sigma_i^1, -\sigma_i^1)$ and $(\sigma_j^1, -\sigma_j^1)$,

$$4Y_{i,j}^{(1)} = \Big\langle \mathbf{1}[|\sigma_i^1|, |\sigma_j^1|, |\sigma_i^2|, |\sigma_j^2| \leqslant \log N, |R_{\sim i,j}| \leqslant 2N^{-1/2}\log N]$$

$$\sigma_i^1 \sigma_j^1 \sigma_i^2 \sigma_j^2 (e^{N\xi_{\sim 2}(R_{\sim i,j}+\Delta_i+\Delta_j)} - e^{N\xi_{\sim 2}(R_{\sim i,j}+\Delta_i-\Delta_j)} - e^{N\xi_{\sim 2}(R_{\sim i,j}-\Delta_i+\Delta_j)} + e^{N\xi_{\sim 2}(R_{\sim i,j}-\Delta_i-\Delta_j)})\Big\rangle_2.$$

Note that on the event in this indicator, $|R_{\sim i,j} \pm \Delta_i \pm \Delta_j| \leqslant 3N^{-1/2}\log N$. Define

$$\kappa(x) = e^{N\xi_{\sim 2}(x)},$$

and note that

$$\sup_{|x|\leqslant 3N^{-1/2}\log N} \kappa^{(4)}(x) = \sup_{|x|\leqslant 3N^{-1/2}\log N} \Big(N\xi_{\sim 2}^{(4)}(x) + 4N^2\xi_{\sim 2}'(x)\xi_{\sim 2}^{(3)}(x) + 3N^2\xi_{\sim 2}''(x)^2 +$$

$$+ 6N^3\xi_{\sim 2}'(x)^2\xi_{\sim 2}''(x) + N^4\xi_{\sim 2}'(x)^4\Big)\kappa(x) = O(N^{6/5}),$$

where we have used that $\sup_{|x|\leqslant 3N^{-1/2}\log N} \kappa(x) \leqslant 2$ and $\gamma_1^2 \leqslant N^{-4/5}$. Since $|\Delta_i|, |\Delta_j| \leqslant N^{-1}\log^2 N$ on the event in the indicator, for $s_i, s_j \in \{\pm 1\}$,

$$e^{N\xi_{\sim 2}(R_{\sim i,j}+s_i\Delta_i+s_j\Delta_j)} = \kappa(R_{\sim i,j}) + \kappa'(R_{\sim i,j})(s_i\Delta_i + s_j\Delta_j) + \tfrac{1}{2}\kappa''(R_{\sim i,j})(s_i\Delta_i + s_j\Delta_j)^2$$

59

$$+ \tfrac{1}{6}\kappa^{(3)}(R_{\sim i,j})(s_i\Delta_i + s_j\Delta_j)^3 + O(N^{6/5}) \cdot (2N^{-1}\log^2 N)^4.$$

It follows that

$$e^{N\xi_{\sim 2}(R_{\sim i,j}+\Delta_i+\Delta_j)} - e^{N\xi_{\sim 2}(R_{\sim i,j}+\Delta_i-\Delta_j)} - e^{N\xi_{\sim 2}(R_{\sim i,j}-\Delta_i+\Delta_j)} + e^{N\xi_{\sim 2}(R_{\sim i,j}-\Delta_i-\Delta_j)}$$
$$= 4\kappa''(R_{\sim i,j})\Delta_i\Delta_j + O(N^{-14/5}\log^8 N),$$

and thus

$$Y_{i,j}^{(1)} = \left\langle \mathbf{1}[|\sigma_i^1|,|\sigma_j^1|,|\sigma_i^2|,|\sigma_j^2| \leqslant \log N, |R_{\sim i,j}| \leqslant 2N^{-1/2}\log N]\sigma_i^1\sigma_j^1\sigma_i^2\sigma_j^2\Delta_i\Delta_j\kappa''(R_{\sim i,j})\right\rangle_2$$
$$+ O(N^{-14/5}\log^{12} N)$$
$$= \left\langle \mathbf{1}[|\sigma_i^1|,|\sigma_j^1|,|\sigma_i^2|,|\sigma_j^2| \leqslant \log N, |R_{\sim i,j}| \leqslant 2N^{-1/2}\log N]\right.$$
$$\left.(\sigma_i^1\sigma_j^1\sigma_i^2\sigma_j^2)^2 \left(N^{-1}\xi_{\sim 2}''(R_{\sim i,j}) + \xi_{\sim 2}'(R_{\sim i,j})^2\right) e^{N\xi_{\sim 2}(R_{\sim i,j})}\right\rangle_2 + o(N^{-2}).$$

On the event in this indicator, $e^{N\xi_{\sim 2}(R_{\sim i,j})} \leqslant 2$, and therefore $\xi_{\sim 2}'$ and $\xi_{\sim 2}''$ can be Taylor expanded to obtain

$$Y_{i,j}^{(1)} \leqslant Y_{i,j}^{(3)} + Y_{i,j}^{(4)} + o(N^{-2}),$$

where

$$Y_{i,j}^{(3)} = 6\gamma_3^2 N^{-1}\left\langle \mathbf{1}[|\sigma_i^1|,|\sigma_j^1|,|\sigma_i^2|,|\sigma_j^2| \leqslant \log N, |R_{\sim i,j}| \leqslant 2N^{-1/2}\log N]\right.$$
$$\left.(\sigma_i^1\sigma_j^1\sigma_i^2\sigma_j^2)^2 R_{\sim i,j}e^{N\xi_{\sim 2}(R_{\sim i,j})}\right\rangle_2,$$

$$Y_{i,j}^{(4)} = \mathbf{E}_{\mu_{H_{N,2}}}\left\langle \mathbf{1}[|\sigma_i^1|,|\sigma_j^1|,|\sigma_i^2|,|\sigma_j^2| \leqslant \log N, |R_{\sim i,j}| \leqslant 2N^{-1/2}\log N]\right.$$
$$\left.(\sigma_i^1\sigma_j^1\sigma_i^2\sigma_j^2)^2 \left(12\gamma_4^2 N^{-1}R_{\sim i,j}^2 + (\gamma_1^2 + 3\gamma_3^2 R_{\sim i,j}^2)^2\right) e^{N\xi_{\sim 2}(R_{\sim i,j})}\right\rangle_2$$

On the event in these indicators, we further have

$$|R^2 - R_{\sim i,j}^2| = |R - R_{\sim i,j}||R + R_{\sim i,j}| \leqslant (|\Delta_i| + |\Delta_j|) \cdot 5N^{-1/2}\log N = O(N^{-3/2}\log^3 N). \qquad (30)$$

From this it readily follows that

$$Y_{i,j}^{(4)} \leqslant 2\left\langle \mathbf{1}[|\sigma_i^1|,|\sigma_j^1|,|\sigma_i^2|,|\sigma_j^2| \leqslant \log N, |R_{\sim i,j}| \leqslant 2N^{-1/2}\log N]\right.$$
$$\left.(\sigma_i^1\sigma_j^1\sigma_i^2\sigma_j^2)^2 \left(12\gamma_4^2 N^{-1}R^2 + (\gamma_1^2 + 3\gamma_3^2 R^2)^2\right)\right\rangle_2 + o(N^{-2})$$
$$= 2\gamma_1^4\widetilde{Q}_0 + (12\gamma_1^2\gamma_3^2 + 24\gamma_4^2 N^{-1})\widetilde{Q}_2 + 18\gamma_3^2\widetilde{Q}_4 + o(N^{-2}),$$

where

$$\widetilde{Q}_k = \left\langle \mathbf{1}[|\sigma_i^1|,|\sigma_j^1|,|\sigma_i^2|,|\sigma_j^2| \leqslant \log N, |R_{\sim i,j}| \leqslant 2N^{-1/2}\log N](\sigma_i^1\sigma_j^1\sigma_i^2\sigma_j^2)^2 R^k\right\rangle_2.$$

By Cauchy–Schwarz and Proposition 8.8, for each $k \in \{0, 2, 4\}$,

$$\widetilde{Q}_k \leqslant \left\langle (\sigma_i^1)^8 (\sigma_i^2)^8 \right\rangle_2^{1/4} \left\langle (\sigma_j^1)^8 (\sigma_j^2)^8 \right\rangle_2^{1/4} \langle R^{2k} \rangle_2^{1/2} = O(N^{-k/2}).$$

This implies $Y_{i,j}^{(4)} \leqslant C(\gamma_1^4 + N^{-2})$. To control $Y_{i,j}^{(3)}$, we recall that $|N\xi_{\sim 2}(R_{\sim i,j})| = O(N^{-3/10} \log N)$ and Taylor expand the exponential:

$$Y_{i,j}^{(3)} = 6\gamma_3^2 N^{-1} \Big\langle \mathbf{1}[|\sigma_i^1|, |\sigma_j^1|, |\sigma_i^2|, |\sigma_j^2| \leqslant \log N, |R_{\sim i,j}| \leqslant 2N^{-1/2} \log N]$$

$$(\sigma_i^1 \sigma_j^1 \sigma_i^2 \sigma_j^2)^2 R_{\sim i,j} (1 + N\xi_{\sim 2}(R_{\sim i,j})) \Big\rangle_2 + o(N^{-2})$$

$$= 6\gamma_3^2 N^{-1} \Big\langle \mathbf{1}[|\sigma_i^1|, |\sigma_j^1|, |\sigma_i^2|, |\sigma_j^2| \leqslant \log N, |R_{\sim i,j}| \leqslant 2N^{-1/2} \log N]$$

$$(\sigma_i^1 \sigma_j^1 \sigma_i^2 \sigma_j^2)^2 (R_{\sim i,j} + N\gamma_1^2 R_{\sim i,j}^2 + N\gamma_3^2 R_{\sim i,j}^4) \Big\rangle_2 + o(N^{-2}).$$

By exchangeability of $(R_{\sim i,j}, -R_{\sim i,j})$, the contribution of the term $R_{\sim i,j}$ vanishes. By (30), we can further estimate $R_{\sim i,j}^2$ with R^2, obtaining

$$Y_{i,j}^{(3)} = 6\gamma_3^2 \Big\langle \mathbf{1}[|\sigma_i^1|, |\sigma_j^1|, |\sigma_i^2|, |\sigma_j^2| \leqslant \log N, |R_{\sim i,j}| \leqslant 2N^{-1/2} \log N]$$

$$(\sigma_i^1 \sigma_j^1 \sigma_i^2 \sigma_j^2)^2 (\gamma_1^2 R^2 + \gamma_3^2 R^4) \Big\rangle_2 + o(N^{-2})$$

$$= 6\gamma_1^2 \gamma_3^2 \widetilde{Q}_2 + 6\gamma_3^4 \widetilde{Q}_4 + o(N^{-2}) \leqslant C(\gamma_1^2 N^{-1} + N^{-2}).$$

Combining all of the above estimates concludes the proof. $\square$

Proof of Proposition 8.10. By a union bound, the event in Lemma 8.15 holds for all $i, j \in [N]$ with probability $1 - e^{-c \log^2 N}$ (over $H_{N, \sim 2}$). On this event,

$$\left\| \frac{Z_N}{Z_{N,2} e^{N\xi_{\sim 2}(1)/2}} \langle \sigma\sigma^\top \rangle - \langle \sigma\sigma^\top \rangle_2 \right\|_F^2 = \sum_{i,j=1}^N X_{i,j}^2 \leqslant 2 \sum_{i,j=1}^N \widetilde{X}_{i,j}^2 + e^{-c \log^2 N}. \tag{31}$$

Combining Lemma 8.16 and Propositions 8.17 and 8.18 shows that

$$\mathbf{E}_{\sim 2} \sum_{i,j=1}^N \widetilde{X}_{i,j}^2 \leqslant 2C(N^2 \gamma_1^4 + 1) + e^{-cN^{1/5}} \leqslant 3C(N^2 \gamma_1^4 + 1).$$

Thus, with probability $2/3$ over $H_{N, \sim 2}$, $\sum_{i,j=1}^N \widetilde{X}_{i,j}^2 \leqslant 9C(N^2 \gamma_1^4 + 1)$. Combining with (31) and taking a final union bound shows that with probability $1/2$ over $H_{N, \sim 2}$,

$$\left\| \frac{Z_N}{Z_{N,2} e^{N\xi_{\sim 2}(1)/2}} \langle \sigma\sigma^\top \rangle - \langle \sigma\sigma^\top \rangle_2 \right\|_F^2 \leqslant 18C(N^2 \gamma_1^4 + 1) + e^{-c \log^2 N} \leqslant 20C(N^2 \gamma_1^4 + 1).$$

The result follows after adjusting C. $\square$

8.3 From positive to very high probability

In this section, we boost the positive probability bound on the second moment matrix to a very high probability bound. To this end, we will show that an appropriate proxy function for the second moment matrix is very Lipschitz. This will imply the desired concentration by standard gaussian concentration.

Let $g \in \mathbb{R}^{N+N^3+N^4+\cdots+N^{p_*}}$ be the vectorized collection of all gaussian interactions corresponding to $H_{N,\sim 2}$. Throughout, as in the previous subsection, we will condition on the event E_2 from Definition 8.13 over $H_{N,2}$, which holds with probability $1 - e^{-cN^{1/5}}$. We define the following functions of g:

$$F_1(g) = B(1 + \gamma_1^2 N) \cdot \left(1 - \left| \log \frac{Z_N}{Z_{N,2}} - \frac{N\zeta_{\sim 2}(1)}{2} \right| \right)$$

$$F_2(g) = B(1 + \gamma_1^2 N) - \left\| \langle \sigma\sigma^\top \rangle \right\|_{\mathrm{op}}$$

$$F(g) = \max(\min(F_1(g), F_2(g)), 0) \,,$$

where B is a sufficiently large constant (specified in the proof of Lemma 8.19). If we can show that with high probability over g, $\min(F_1(g), F_2(g)) \geq 0$, then the conclusion follows. Indeed, from $F_2(g) \geq 0$, we immediately obtain $\left\| \langle \sigma\sigma^\top \rangle \right\|_{\mathrm{op}} \leq B(1 + \gamma_1^2 N)$. F_1 allows control over the free energy of the p-spin model in terms of that of the corresponding 2-spin model. This gives good control over the overlaps (in a manner to be made precise shortly), which is crucial for establishing the high probability statement. It is also important earlier in this section, in showing that the free energy of the p-spin model concentrates well.

Towards this, we first start with the positive probability statement, which was essentially established in the previous subsections.

Lemma 8.19. *There exists a constant $B > 0$ such that with probability at least $\frac{1}{3}$, we have $F(g) \geq \frac{B}{2}(1 + \gamma_1^2 N)$.*

Proof. By Proposition 8.9, with probability $1 - O(N^{-1/15})$ over g, we have $F_1(g) \geq \frac{B}{2}(1 + \gamma_1^2 N)$, so $\frac{Z_N}{Z_{N,2}e^{N\zeta_{\sim 2}(1)/2}} \geq e^{-1/2}$. Intersecting this with the event from Proposition 8.10 implies that with probability at least $\frac{1}{3}$,

$$\left\| \langle \sigma\sigma^\top \rangle \right\|_{\mathrm{op}} \leq e^{1/2} \left\| \langle \sigma\sigma^\top \rangle_2 \right\|_{\mathrm{op}} + e^{1/2}\sqrt{C(1 + \gamma_1^4 N^2)}$$

$$\leq \frac{B}{2}(1 + \gamma_1^2 N) \,,$$

where we have used E_2 to apply Proposition 8.8 and after appropriately picking B. $\square$

Let $\mathcal{E}$ denote the $H_{N,2}$-measurable event from Corollary 8.14:

$$\left\{ g : \int \mathbf{1}[|R(\sigma^1, \sigma^2)| \geq N^{-2/5}] e^{H_N(\sigma^1) + H_N(\sigma^2)} \, d\rho^{\otimes 2}(\sigma^1, \sigma^2) \leq Z_{N,2}^2 e^{N\zeta_{\sim 2}(1) - cN^{1/5}} \right\},$$

which holds with probability $1 - e^{-cN^{1/5}}$ over g. The key observation is that this gives us good control on the overlaps.

Lemma 8.20. *On $\mathcal{E}$, if $F_1(g) \geqslant 0$, then for any $0 \leqslant p \leqslant \log^2 N$, we have*

$$\left\| \langle \sigma^{\otimes p} \rangle \right\|_F^2 \leqslant O(N^{3p/5}).$$

Proof. By splitting up the expectation based on whether $|R(\sigma^1, \sigma^2)| \geqslant N^{-2/5}$, on $\mathcal{E}$ we have

$$
\begin{aligned}
\left\| \langle \sigma^{\otimes p} \rangle \right\|_F^2 &= \langle N R(\sigma^1, \sigma^2) \rangle^p \\
&\leqslant N^{3p/5} + N^{2p} \frac{1}{Z_N^2} \int \mathbf{1}[|R(\sigma^1, \sigma^2)| \geqslant N^{-2/5}] e^{H_N(\sigma^1) + H_N(\sigma^2)} \, d\rho^{\otimes 2}(\sigma^1, \sigma^2) \\
&\leqslant N^{3p/5} + N^{2p} \frac{Z_{N,2}^2}{Z_N^2} e^{N \xi_{\sim 2}(1) - c N^{1/5}} \qquad\qquad\qquad\qquad\text{(Definition of } \mathcal{E}) \\
&\leqslant N^{3p/5} + N^{2p} e^{1 - c N^{1/5}},
\end{aligned}
$$

where the last line used $F_1(g) \geqslant 0$. Since $p \leqslant \log^2 N$, the above quantity is $O(N^{3p/5})$, as desired. $\qquad\square$

The above is a crucial input to prove Lipschitzness of F on $\mathcal{E}$.

Lemma 8.21. *The function F is $O((1 + \gamma_1^2 N) N^{-1/10})$-Lipschitz restricted to $\mathcal{E}$.*

Before we prove this, let us see how it implies Proposition 8.2, restated for convenience.

Proposition 8.2. *There is a $H_{N,2}$-measurable event with probability $1 - e^{-cN^{1/5}}$ on which the following holds with probability $1 - e^{-cN^{1/5}}$ over $H_{N,\sim 2}$.*

1. *The partition functions $Z_N, Z_{N,2}$ satisfy*

$$\left| \log \frac{Z_N}{Z_{N,2}} - \frac{N \xi_{\sim 2}(1)}{2} \right| \leqslant 1/2.$$

2. *The Gibbs measure satisfies $\| \langle \sigma \sigma^\top \rangle \|_{\mathsf{op}} \leqslant C(1 + \gamma_1^2 N)$.*

Proof of Proposition 8.2. By Kirszbraun's extension theorem, we can extend F to $\widetilde{F}$ such that each $\widetilde{F}$ has the same Lipschitz constant as F and agrees with F on $\mathcal{E}$. We can now apply gaussian concentration to $\widetilde{F}$ to conclude that

$$\Pr\left[|\widetilde{F}(g) - \mathbf{E}\widetilde{F}(g)| \geqslant \tfrac{B}{4}(1 + \gamma_1^2 N) \right] \geqslant 1 - e^{-cN^{1/5}}. \tag{32}$$

By Lemma 8.19, with probability at least $\frac{1}{3}$, we have $F(g) \geqslant \frac{B}{2}(1 + \gamma_1^2 N)$. Upon further intersection with $\mathcal{E}$ (where $\widetilde{F}(g) = F(g)$) and the event from (32), we conclude $\mathbf{E}\widetilde{F}(g) \geqslant \frac{B}{4}(1 + \gamma_1^2 N)$. Thus,

$$
\begin{aligned}
\Pr[F(g) = 0] &\leqslant \Pr[\mathcal{E}^c] + \Pr[\widetilde{F}(g) = 0] \\
&\leqslant e^{-cN^{1/5}} + \Pr\left[|\widetilde{F}(g) - \mathbf{E}\widetilde{F}(g)| \geqslant \frac{B(1 + \gamma_1^2 N)}{4} \right] \\
&\leqslant e^{-cN^{1/5}},
\end{aligned}
$$

after adjusting c.

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Finally, let us prove the Lipschitz bound.

Proof of Lemma 8.21. The set $\mathcal{E}$ is a convex set in $\boldsymbol{g}$. Indeed, $e^{H_N(\sigma)}$ is a convex function of $\boldsymbol{g}$, so the LHS of the inequality $\mathcal{E}$ is convex in $\boldsymbol{g}$, whereas the RHS does not depend on $\boldsymbol{g}$, and sublevel sets of convex functions are convex. Furthermore, F is absolutely continuous (hence differentiable almost everywhere), so to prove F is Lipschitz on $\mathcal{E}$ it suffices to bound $\|\boldsymbol{\nabla} F\|$ on $\mathcal{E}$, wherever it is defined.

The easier case is if $\min(F_1(\boldsymbol{g}), F_2(\boldsymbol{g})) < 0$. In this case, $F(\boldsymbol{g}) = 0$ in an open neighborhood of $\boldsymbol{g}$, so $\boldsymbol{\nabla} F(\boldsymbol{g}) = 0$ identically. Therefore, for the rest of the proof, assume $\min(F_1(\boldsymbol{g}), F_2(\boldsymbol{g})) \geqslant 0$. We will compute the gradient of the F_i's, and to simplify the calculation, we will take the gradient with respect to $\boldsymbol{g}_p \in \mathbb{R}^{N^p}$ corresponding to the degree-p disorder in $H_{N,\sim 2}$.

For $F_1(\boldsymbol{g})$, note that its only dependence on $\boldsymbol{g}$ is via $\log Z_N$, we have

$$\left\|\boldsymbol{\nabla}_{\boldsymbol{g}_p} F_1(\boldsymbol{g})\right\| = B(1 + \gamma_1^2 N)\left\|\boldsymbol{\nabla}_{\boldsymbol{g}_p} \log Z_N\right\| = B(1 + \gamma_1^2 N) \cdot \frac{\gamma_p}{N^{(p-1)/2}}\left\|\langle \sigma^{\otimes p}\rangle\right\|_F,$$

and since $F_1(\boldsymbol{g}) \geqslant 0$, we can apply Lemma 8.20 to conclude that

$$
\begin{aligned}
\left\|\boldsymbol{\nabla}_{\boldsymbol{g}} \log Z_N\right\|^2 &\lesssim \sum_{p \in [p_*]\setminus\{2\}} \gamma_p^2 N^{-(p-1)} \cdot N^{3p/5} \\
&\leqslant \gamma_1^2 \cdot N^{3/5} + \sum_{p \geqslant 3} \gamma_p^2 N^{1-2p/5} \\
&\lesssim \gamma_1^2 \cdot N^{3/5} + N^{-1/5}
\end{aligned}
\tag{33}
$$

Since $\gamma_1^2 \leqslant N^{-4/5}$, we conclude that $\left\|\boldsymbol{\nabla}_{\boldsymbol{g}} F_1(\boldsymbol{g})\right\| \lesssim (1 + \gamma_1^2 N)N^{-1/10}$, as desired.

Turning now to $F_2(\boldsymbol{g})$, we observe that $\left\|\langle \sigma\sigma^\top\rangle\right\|_{\mathrm{op}} = \langle\langle u, \sigma\rangle^2\rangle$, where u is the top eigenvector of $\langle \sigma\sigma^\top\rangle$ with $\|u\|_2 = 1$. By the envelope theorem, we can evaluate the gradient with u fixed. For any $v \in \mathbb{R}^{N^p}$ with $\|v\|_2 = 1$, we will upper bound $\langle v, \boldsymbol{\nabla}_{\boldsymbol{g}_p}\langle\langle u, \sigma\rangle^2\rangle\rangle$. Applying the quotient rule yields

$$
\begin{aligned}
\langle v, \boldsymbol{\nabla}_{\boldsymbol{g}_p}\langle\langle u, \sigma\rangle^2\rangle\rangle &= \langle\langle u, \sigma\rangle^2\langle v, \boldsymbol{\nabla}_{\boldsymbol{g}_p} H(\sigma)\rangle\rangle - \langle\langle u, \sigma\rangle^2\rangle\langle\langle v, \boldsymbol{\nabla}_{\boldsymbol{g}_p} H(\sigma)\rangle\rangle \\
&= \frac{\gamma_p}{N^{(p-1)/2}}\left(\langle\langle u, \sigma\rangle^2\langle v, \sigma^{\otimes p}\rangle\rangle - \langle\langle u, \sigma\rangle^2\rangle\langle\langle v, \sigma^{\otimes p}\rangle\rangle\right).
\end{aligned}
$$

Consider the first term $\langle\langle u, \sigma\rangle^2\langle v, \sigma^{\otimes p}\rangle\rangle$. Using Hölder's inequality with $q = 1 + \log N$ and $q' = 1 + \frac{1}{\log N}$, we see

$$
\begin{aligned}
\langle\langle u, \sigma\rangle^2\langle v, \sigma^{\otimes p}\rangle\rangle &\leqslant \left\langle\langle u, \sigma\rangle^{2q'}\right\rangle^{1/q'}\langle\langle v, \sigma^{\otimes p}\rangle^q\rangle^{1/q} \\
&\leqslant N^{1/\log N}\langle\langle u, \sigma\rangle^2\rangle\langle v^{\otimes q}, \langle\sigma^{\otimes pq}\rangle\rangle^{1/q} \\
&\lesssim (1 + \gamma_1^2 N)\langle v^{\otimes q}, \langle\sigma^{\otimes pq}\rangle\rangle^{1/q} \\
&\leqslant (1 + \gamma_1^2 N) \cdot N^{3p/10},
\end{aligned}
$$

where in the second to last line we have used $F_2(\boldsymbol{g}) \geqslant 0$ to apply the bound $\langle\langle u, \sigma\rangle^2\rangle \leqslant O(1 + \gamma_1^2 N)$, and in the last line we have used $F_1(\boldsymbol{g}) \geqslant 0$, along with $pq \leqslant O(\log N)$, to apply Lemma 8.20. The same argument upper bounds the contribution of the second term as $O\left((1 + \gamma_1^2 N)N^{3p/10}\right)$. These bounds (aside from the common factor of $O(1 + \gamma_1^2 N)$, which we can pull out), exactly match the ones used in the calculation as carried out for $F_1(\boldsymbol{g})$ in (33). Hence, the same argument ultimately yields $\left\|\boldsymbol{\nabla}_{\boldsymbol{g}} F_2(\boldsymbol{g})\right\| \lesssim (1 + \gamma_1^2 N)N^{-1/10}$, completing the proof. $\qquad\square$

References

[ABXY24]   Arka Adhikari, Christian Brennecke, Changji Xu, and Horng-Tzer Yau. Spectral gap estimates for mixed p-spin models at high temperature. *Probability Theory and Related Fields*, pages 1–29, 2024. 9

[ACO08]   Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 793–802. IEEE, 2008. 31

[AEGP23]   Ahmed El Alaoui, Ronen Eldan, Reza Gheissari, and Arianna Piana. Fast relaxation of the random field Ising dynamics. *arXiv preprint arXiv:2311.06171*, 2023. 2, 8

[AG24]   Ahmed El Alaoui and Jason Gaitonde. Bounds on the covariance matrix of the Sherrington–Kirkpatrick model. *Electronic Communications in Probability*, 29:1–13, 2024. 6, 9

[Aid98]   Shigeki Aida. Uniform positivity improving property, Sobolev inequalities, and spectral gaps. *Journal of functional analysis*, 158(1):152–185, 1998. 8

[AJ24]   Gérard Ben Arous and Aukosh Jagannath. Shattering versus metastability in spin glasses. *Communications on Pure and Applied Mathematics*, 77(1):139–176, 2024. 2, 6, 28

[AJK+21]   Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence II: optimal sampling and concentration via restricted modified log-Sobolev inequalities. *arXiv preprint arXiv:2111.03247*, 2021. 8

[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 *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1418–1430, 2022. 1, 9, 13, 16

[AJK+24]   Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Universality of spectral independence with applications to fast mixing in spin glasses. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 5029–5056. SIAM, 2024. 9

[AKV24]   Nima Anari, Frederic Koehler, and Thuy-Duong Vuong. Trickle-Down in Localization Schemes and Applications. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1094–1105, 2024. 9

[AL20]   Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1198–1211, 2020. 4

[ALGV24]   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. *Annals of Mathematics*, 199(1):259–299, 2024. 1, 8

[ALOG21]   Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. *SIAM Journal on Computing*, (0):FOCS20–1, 2021. 1, 4, 8

[AM22a]   Ahmed El Alaoui and Andrea Montanari. An information-theoretic view of stochastic localization. *IEEE Transactions on Information Theory*, 68(11):7423–7426, 2022. 14

[AM22b]   Ahmed El Alaoui and Andrea Montanari. An information-theoretic view of stochastic localization. *IEEE Transactions on Information Theory*, 68(11):7423–7426, 2022. 31

[AMS21]   Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *The Annals of Probability*, 49(6):2922–2960, 2021. 10

[AMS22]   Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from the Sherrington-Kirkpatrick Gibbs measure via algorithmic stochastic localization. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 323–334. IEEE, 2022. 9, 11, 31, 33

[AMS23a]   Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from Mean-Field Gibbs Measures via Diffusion Processes. *arXiv preprint arXiv:2310.08912*, 2023. 9, 11

[AMS23b]   Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Shattering in pure spherical spin glasses. *arXiv preprint arXiv:2307.04659*, 2023. 2, 31, 33

[BADG06]   Gérard Ben Arous, Amir Dembo, and Alice Guionnet. Cugliandolo-Kurchan equations for dynamics of spin-glasses. *Probability theory and related fields*, 136(4):619–660, 2006. 9

[BAGJ20]   Gérard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Bounding flows for spherical spin glass dynamics. *Communications in Mathematical Physics*, 373:1011–1048, 2020. 9

[BB19]   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. 9, 16

[BCE+22]   Krishna Balasubramanian, Sinho Chewi, Murat A Erdogdu, Adil Salim, and Shunshi Zhang. Towards a theory of non-log-concave sampling: first-order stationarity guarantees for Langevin Monte Carlo. In *Conference on Learning Theory*, pages 2896–2923. PMLR, 2022. 22

[BCKM98]   Jean-Philippe Bouchaud, Leticia F Cugliandolo, Jorge Kurchan, and Marc Mézard. Out of equilibrium dynamics in spin-glasses and other glassy systems. *Spin glasses and random fields*, 12:161, 1998. 9

[BD00]   Rajendra Bhatia and Chandler Davis. A Better Bound on the Variance. *The American Mathematical Monthly*, 107(4):353–357, 2000. 24

[BGZ24]   Antonio Blanca, Reza Gheissari, and Xusheng Zhang. Mean-field Potts and random-cluster dynamics from high-entropy initializations. *arXiv preprint arXiv:2404.13014*, 2024. 8

[BL16]   Jinho Baik and Ji Oon Lee. Fluctuations of the free energy of the spherical Sherrington–Kirkpatrick model. *Journal of Statistical Physics*, 165:185–224, 2016. 36, 50

[BNN24]   Guy Bresler, Dheeraj Nagaraj, and Eshaan Nichani. Metastable mixing of Markov chains: Efficiently sampling low temperature exponential random graphs. *The Annals of Applied Probability*, 34(1A):517–554, 2024. 9

[BSXY24]   Christian Brennecke, Adrien Schertzer, Changji Xu, and Horng-Tzer Yau. The two point function of the SK model without external field at high temperature. *Probab. Math. Phys*, 5:131–175, 2024. 9

[BXY23]   Christian Brennecke, Changji Xu, and Horng-Tzer Yau. Operator norm bounds on the correlation matrix of the SK model at high temperature. *arXiv preprint arXiv:2307.12535*, 2023. 9

[BY05]   ZD Bai and Jianfeng Yao. On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli*, 11(6):1059–1092, 2005. 51

[CCM21]   Michael Celentano, Chen Cheng, and Andrea Montanari. The high-dimensional asymptotics of first order methods with random data. *arXiv preprint arXiv:2112.07572*, 2021. 9

[CDL+12]   Paul Cuff, Jian Ding, Oren Louidor, Eyal Lubetzky, Yuval Peres, and Allan Sly. Glauber dynamics for the mean-field Potts model. *Journal of Statistical Physics*, 149:432–477, 2012. 8

[CE22]   Yuansi Chen and Ronen Eldan. Localization Schemes: A Framework for Proving Mixing Bounds for Markov Chains. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 110–122, Los Alamitos, CA, USA, nov 2022. IEEE Computer Society. 1, 5, 8, 12, 14, 16, 20, 25, 27, 73

[Cel24]   Michael Celentano. Sudakov–Fernique post-AMP, and a new proof of the local convexity of the TAP free energy. *The Annals of Probability*, 52(3):923–954, 2024. 11

[CGG07]   Patrick Cattiaux, Ivan Gentil, and Arnaud Guillin. Weak logarithmic Sobolev inequalities and entropic convergence. *Probability theory and related fields*, 139:563–603, 2007. 8

[CGG+16]   Jin-Yi Cai, Andreas Galanis, Leslie Ann Goldberg, Heng Guo, Mark Jerrum, Daniel Štefankovič, and Eric Vigoda. #BIS-hardness for 2-spin systems on bipartite bounded degree graphs in the tree non-uniqueness region. *Journal of Computer and System Sciences*, 82(5):690–711, 2016. 11

[CGM12]   Prasad Chebolu, Leslie Ann Goldberg, and Russell Martin. The complexity of approximately counting stable matchings. *Theoretical Computer Science*, 437:35–68, 2012. 11

[Che23]   Sinho Chewi. Log-concave sampling. *Book draft available at https://chewisinho. github. io*, 2023. 8, 16, 17, 29, 30

[CHKW23] Sergio Caracciolo, Alexander Hartmann, Scott Kirkpatrick, and Martin Weigel. Simu-
          lated annealing, optimization, searching for ground states. In *Spin Glass Theory and
          Far Beyond: Replica Symmetry Breaking After 40 Years*, pages 1–20. World Scientific,
          2023. 10

[CHS93]   Andrea Crisanti, Heinz Horner, and H J Sommers. The spherical p-spin interaction
          spin-glass model: the dynamics. *Zeitschrift für Physik B Condensed Matter*, 92:257–271,
          1993. 2, 5, 9, 10

[CK93]   Leticia F Cugliandolo and Jorge Kurchan. Analytical solution of the off-equilibrium
          dynamics of a long-range spin-glass model. *Physical Review Letters*, 71(1):173, 1993. 9,
          10

[CWZZ24] Xiang Cheng, Bohan Wang, Jingzhao Zhang, and Yusong Zhu. Fast Conditional
          Mixing of MCMC Algorithms for Non-log-concave Distributions. *Advances in Neural
          Information Processing Systems*, 36, 2024. 22

[DF91]   Martin Dyer and Alan Frieze. Computing the volume of convex bodies: a case
          where randomness provably helps. *Probabilistic combinatorics and its applications*,
          44(123-170):0754–68052, 1991. 2

[DFK91]   Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm
          for approximating the volume of convex bodies. *Journal of the ACM (JACM)*, 38(1):1–
          17, 1991. 2

[DSC96]   Persi Diaconis and Laurent Saloff-Coste. Logarithmic Sobolev inequalities for finite
          Markov chains. *The Annals of Applied Probability*, 6(3), 8 1996. 74

[EKZ22]   Ronen Eldan, Frederic Koehler, and Ofer Zeitouni. A spectral condition for spectral
          gap: fast mixing in high-temperature Ising models. *Probability theory and related fields*,
          182(3):1035–1051, 2022. 1, 8, 9

[Eld13]   Ronen Eldan. Thin shell implies spectral gap up to polylog via a stochastic local-
          ization scheme. *Geometric and Functional Analysis*, 23(2):532–569, 2013. 5, 8, 9, 14,
          25

[Eld20]   Ronen Eldan. Taming correlations through entropy-efficient measure decompositions
          with applications to mean-field approximation. *Probability Theory and Related Fields*,
          176(3):737–755, 2020. 5, 9

[FFRT21] Giampaolo Folena, Silvio Franz, and Federico Ricci-Tersenghi. Gradient descent dy-
          namics in the mixed p-spin spherical model: finite-size simulations and comparison
          with mean-field integration. *Journal of Statistical Mechanics: Theory and Experiment*,
          2021(3):033302, 2021. 10

[GJ12]   Leslie Ann Goldberg and Mark Jerrum. Approximating the partition function of the
          ferromagnetic Potts model. *Journal of the ACM (JACM)*, 59(5):1–31, 2012. 11

[GJ19]   Reza Gheissari and Aukosh Jagannath. On the spectral gap of spherical spin glass dynamics. In *Annales de l'institut Henri Poincare (B) Probability and Statistics*, volume 55, pages 756–776. Institute of Mathematical Statistics, 2019. 2, 9

[GJK23]   David Gamarnik, Aukosh Jagannath, and Eren C Kızıldağ. Shattering in the Ising Pure *p*-Spin Model. *arXiv preprint arXiv:2307.07461*, 2023. 2

[GLR18]   Rong Ge, Holden Lee, and Andrej Risteski. Simulated tempering Langevin Monte Carlo II: An improved proof using soft Markov chain decomposition. *arXiv preprint arXiv:1812.00793*, 2018. 22

[GLZ+18]   Ruiqi Gao, Yang Lu, Junpei Zhou, Song-Chun Zhu, and Ying Nian Wu. Learning Generative ConvNets via Multi-grid Modeling and Sampling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9155–9164, 2018. 22

[GMT06]   Sharad Goel, Ravi Montenegro, and Prasad Tetali. Mixing Time Bounds via the Spectral Profile. *Electronic Journal of Probability*, 11, 2006. 8

[GS22]   Reza Gheissari and Alistair Sinclair. Low-temperature ising dynamics with random initializations. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1445–1458, 2022. 9

[GS24]   Reza Gheissari and Alistair Sinclair. Spatial mixing and the random-cluster dynamics on lattices. *Random Structures & Algorithms*, 64(2):490–534, 2024. 9

[GSVY16]   Andreas Galanis, Daniel Stefankovic, Eric Vigoda, and Linji Yang. Ferromagnetic Potts model: Refined #BIS-hardness and related results. *SIAM Journal on Computing*, 45(6):2004–2065, 2016. 11

[GTC24]   Wei Guo, Molei Tao, and Yongxin Chen. Provable benefit of annealed langevin monte carlo for non-log-concave sampling. *arXiv preprint arXiv:2407.16936*, 2024. 22

[GW95]   Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995. 10

[GZ00]   A Guionnet and O Zeitouni. Concentration of the spectral measure for large matrices. *Electronic Communications in Probability*, 5:119–136, 2000. 51, 84

[Hin10]   Geoffrey Hinton. A Practical Guide to Training Restricted Boltzmann Machines. *Momentum*, 9(1):926, 2010. 22

[HJA20]   Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 9

[HK22]   Jun-Ting Hsieh and Pravesh K Kothari. Approximating Max-Cut on Bounded Degree Graphs: Tighter Analysis of the FKL Algorithm. *arXiv preprint arXiv:2206.09204*, 2022. 10

[HMP24]   Brice Huang, Andrea Montanari, and Huy Tuan Pham. Sampling from Spherical Spin Glasses in Total Variation via Algorithmic Stochastic Localization. *arXiv preprint arXiv:2404.15651*, 2024. 3, 5, 6, 9, 11, 14, 28, 30, 31, 32, 33, 35, 36, 39, 50, 53, 54, 86

[HPR19]   Tyler Helmuth, Will Perkins, and Guus Regts. Algorithmic Pirogov-Sinai theory. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 1009–1020, 2019. 11

[HS22]   Brice Huang and Mark Sellke. Tight Lipschitz Hardness for optimizing Mean Field Spin Glasses. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 312–322. IEEE, 2022. 10, 29

[HS23]   Brice Huang and Mark Sellke. A Constructive Proof of the Spherical Parisi Formula. *arXiv preprint arXiv:2311.15495*, 2023. 6, 37, 53, 54, 87, 88, 90

[Hsu02]   Elton P Hsu. *Stochastic analysis on manifolds*. Number 38. American Mathematical Soc., 2002. 16

[Hua24]   Brice Huang. Capacity threshold for the Ising perceptron. *arXiv preprint arXiv:2404.18902*, 2024. 33

[JKP20]   Matthew Jenssen, Peter Keevash, and Will Perkins. Algorithms for #BIS-hard problems on expander graphs. *SIAM Journal on Computing*, 49(4):681–710, 2020. 11

[JSV04]   Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM (JACM)*, 51(4):671–697, 2004. 2

[KKMO07]   Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007. 10

[KL24]   Bo'az Klartag and Joseph Lehec. Isoperimetric inequalities in high-dimensional convex sets. *arXiv preprint arXiv:2406.01324*, 2024. 8

[KLS97]   Ravi Kannan, László Lovász, and Miklós Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures & Algorithms*, 11(1):1–50, 1997. 2

[KLV24]   Frederic Koehler, Holden Lee, and June Vuong. Efficiently learning and sampling multimodal distributions with data-based initialization. *Personal communication*, 2024. 7, 22

[KO20]   Tali Kaufman and Izhar Oppenheim. High order random walks: Beyond spectral gap. *Combinatorica*, 40:245–281, 2020. 8

[KV23]   Frederic Koehler and Thuy-Duong Vuong. Sampling Multimodal Distributions with the Vanilla Score: Benefits of Data-Based Initialization. *arXiv preprint arXiv:2310.01762*, 2023. 7, 22

[Liu23]   Kuikui Liu. *Spectral Independence a New Tool to Analyze Markov Chains*. PhD thesis, University of Washington, 2023. 8

[LMR⁺24]   Kuikui Liu, Sidhanth Mohanty, Prasad Raghavendra, Amit Rajaraman, and David X Wu. Locally Stationary Distributions: A Framework for Analyzing Slow-Mixing Markov Chains. *arXiv preprint arXiv:2405.20849*, 2024. 10, 16

[LMRW24] Kuikui Liu, Sidhanth Mohanty, Amit Rajaraman, and David X Wu. Fast Mixing in Sparse Random Ising Models. *arXiv preprint arXiv:2405.06616*, 2024. 13

[LRG18]   Holden Lee, Andrej Risteski, and Rong Ge. Beyond log-concavity: Provable guarantees for sampling multi-modal distributions using simulated tempering langevin monte carlo. *Advances in neural information processing systems*, 31, 2018. 22

[LS90]   László Lovász and Miklós Simonovits. The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. In *Proceedings [1990] 31st annual symposium on foundations of computer science*, pages 346–354. IEEE, 1990. 2

[LS93]   László Lovász and Miklós Simonovits. Random walks in a convex body and an improved volume algorithm. *Random structures & algorithms*, 4(4):359–412, 1993. 8

[LS16]   Eyal Lubetzky and Allan Sly. Information percolation and cutoff for the stochastic Ising model. *Journal of the American Mathematical Society*, 29(3):729–774, 2016. 9

[LS17]   Eyal Lubetzky and Allan Sly. Universality of cutoff for the Ising model. 2017. 9

[LV24]   Yin Tat Lee and Santosh S Vempala. Eldan's stochastic localization and the KLS conjecture: Isoperimetry, concentration and mixing. *Annals of Mathematics*, 199(3):1043–1092, 2024. 8

[Mat06]   P Mathieu. Quand l'inegalite log-Sobolev implique l'inegalite de trou spectral. In *Séminaire de Probabilités XXXII*, pages 30–35. Springer, 2006. 8

[Mon21]   Andrea Montanari. Optimization of the Sherrington–Kirkpatrick Hamiltonian. *SIAM Journal on Computing*, (0):FOCS19–1, 2021. 10

[Mon23]   Andrea Montanari. Sampling, diffusions, and stochastic localization. *arXiv preprint arXiv:2305.10690*, 2023. 9

[MRT04]   Andrea Montanari and Federico Ricci-Tersenghi. Cooling-schedule dependence of the dynamics of mean-field glasses. *Physical Review B—Condensed Matter and Materials Physics*, 70(13):134406, 2004. 10

[MW23]   Andrea Montanari and Yuchen Wu. Posterior sampling from the spiked models via diffusion processes. *arXiv preprint arXiv:2304.11449*, 2023. 9, 11

[MW24]   Andrea Montanari and Yuchen Wu. Provably efficient posterior sampling for sparse linear regression via measure decomposition. *arXiv preprint arXiv:2406.19550*, 2024. 9, 16

[NHH+20]   Erik Nijkamp, Mitch Hill, Tian Han, Song-Chun Zhu, and Ying Nian Wu. On the Anatomy of MCMC-Based Maximum Likelihood Learning of Energy-Based Models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5272–5280, 2020. 22

[Rag08]   Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 245–254, 2008. 10

[RW01]   Michael Röckner and Feng-Yu Wang. Weak Poincaré inequalities and L2-convergence rates of Markov semigroups. *Journal of Functional Analysis*, 185(2):564–603, 2001. 3, 8, 13

[SDWMG15]   Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. PMLR, 2015. 9

[Sel24a]   Mark Sellke. Optimizing mean field spin glasses with external field. *Electronic Journal of Probability*, 29:1–47, 2024. 10

[Sel24b]   Mark Sellke. The threshold energy of low temperature Langevin dynamics for pure spherical spin glasses. *Communications on Pure and Applied Mathematics*, 77(11):4065–4099, 2024. 9, 10

[SK75]   David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical review letters*, 35(26):1792, 1975. 9

[SSDK+21]   Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*, 2021. 9

[Sub21]   Eliran Subag. Following the ground states of full-RSB spherical spin glasses. *Communications on Pure and Applied Mathematics*, 74(5):1021–1044, 2021. 10

[Tre01]   Luca Trevisan. Non-approximability results for optimization problems on bounded degree instances. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 453–461, 2001. 10

[Wan06]   Feng-Yu Wang. *Functional inequalities Markov semigroups and spectral theory*. Elsevier, 2006. 8

[XLZW16]   Jianwen Xie, Yang Lu, Song-Chun Zhu, and Yingnian Wu. A Theory of Generative ConvNet. In *International conference on machine learning*, pages 2635–2644. PMLR, 2016. 22

A Annealed Glauber dynamics on discrete domains

In this section, we collect the analogous results for weak functional inequalities for Glauber dynamics.

Definition A.1 (Weak Poincaré for the hypercube). We say π on $\{\pm 1\}^n$ satisfies a $(\rho_{\mathrm{PI}}, \varepsilon)$-weak Poincaré inequality for Glauber dynamics if for all functions f,

$$\mathbf{Var}_\pi[f] \leqslant \frac{1}{\rho_{\mathrm{PI}}} \cdot \mathcal{E}(f, f) + \varepsilon \cdot \mathrm{osc}(f)^2.$$

Similarly, we say π satisfies a $(\rho_{\mathrm{LS}}, \varepsilon)$-weak modified log-Sobolev inequality if for all functions f,

$$\mathrm{Ent}_\pi[f] \leqslant \frac{1}{\rho_{\mathrm{LS}}} \cdot \mathcal{E}(f, \log f) + \varepsilon \cdot \mathrm{osc}(\sqrt{f})^2.$$

Remark A.2. The above definition is related to the continuous setting by using the discrete gradient, which can be bounded by $\mathrm{osc}(f)^2$.

First, we will need a concavity property for the Dirichlet form for Glauber dynamics, which is well-known. We provide a proof for the sake of self-containedness.

Fact A.3. *Let π be a distribution on $\{\pm 1\}^n$, and $\pi = \mathbf{E}_{z \sim \rho} \pi_z$ a measure decomposition of π. Then $\mathcal{E}_\pi(f, f) \geqslant \mathbf{E}_{z \sim \rho} \mathbf{E}_{\pi_z}(f, f)$.*

Proof. For Glauber dynamics on the hypercube, we have

$$\mathcal{E}_\pi(f, f) = \frac{1}{n} \sum_{\|x-y\|_1 = 2} \frac{\pi(x)\pi(y)}{\pi(x) + \pi(y)} (f(x) - f(y))^2$$

$$= \frac{1}{n} \sum_{\|x-y\|_1 = 2} \frac{\mathbf{E}_{z \sim \rho} \pi_z(x) \mathbf{E}_{z \sim \rho} \pi_z(y)}{\mathbf{E}_{z \sim \rho} \pi_z(x) + \mathbf{E}_{z \sim \rho} \pi_z(y)} (f(x) - f(y))^2$$

$$\geqslant \mathbf{E}_{z \sim \rho} \mathcal{E}_{\pi_z}(f, f),$$

where the last line follows from concavity of the map $(a, b) \mapsto \frac{ab}{a+b}$ for $a, b > 0$. $\qquad \square$

The following lemma transfers a true Poincaré inequality on π to a weak Poincaré inequality on π' for Glauber dynamics on the hypercube.

Lemma A.4. *Let π, π' be distributions on $\{\pm 1\}^n$ such that π satisfies a ρ_{PI}-Poincaré inequality for Glauber dynamics and $d_{\mathrm{TV}}(\pi, \pi') \leqslant \delta$. Then, π' satisfies a weak $(\rho_{\mathrm{PI}}, 2\delta)$-Poincaré inequality for Glauber dynamics.*

Proof. Again, there exists a coupling $\mathcal{C}$ of (π, π') such that for $(x, x') \sim \mathcal{C}$, $\mathbf{Pr}[x \neq x'] \leqslant \delta$. The main difference is that the Dirichlet form comparison can be bounded in terms of $\mathrm{osc}(f)^2$. Arguing as before yields

$$\mathcal{E}_{\pi'}(f, f) \geqslant \mathcal{E}_\pi(f, f) - \delta \cdot \mathrm{osc}(f)^2$$

$$\geqslant \rho_{\mathrm{PI}} \cdot \mathbf{Var}_\pi[f] - \delta \cdot \mathrm{osc}(f)^2$$

$$\geqslant \rho_{\mathrm{PI}} \cdot \mathbf{Var}_{\pi'}[f] - \delta \cdot \mathrm{osc}(f)^2 (1 + \rho_{\mathrm{PI}}).$$

$\qquad \square$

Remark A.5. The above two results also hold more generally if P is the Markov chain associated to a Doob localization scheme (cf. [CE22, Section 2.3]), such as when P is Glauber dynamics for a general product domain.

Lemma A.6. *Let π be a distribution over $\{\pm 1\}^n$, and $\pi = \mathbf{E}_{z\sim\rho}\pi_z$ a measure decomposition of π such that*

- *for all functions f, $\mathbf{E}_{z\sim\rho}\mathbf{Var}_{\pi_z}[f] \geqslant C_{\mathrm{Var}}\mathbf{Var}_\pi[f]$, and*

- *with probability $1 - \eta$ over $z \sim \rho$, π_z satisfies a $(\rho_{\mathrm{PI}}, \delta)$-weak Poincaré inequality with respect to Glauber.*

Then, π satisfies a $\left(\rho_{\mathrm{PI}}C_{\mathrm{Var}}, \frac{\delta+\eta}{C_{\mathrm{Var}}}\right)$-weak Poincaré inequality.

Proof. The proof is the same as that of Lemma 4.11, except in the Langevin case we have $\mathcal{E}_\pi(f,f) = \mathbf{E}_{z\sim\rho}\mathcal{E}_{\pi_z}(f,f)$, whereas here we apply Fact A.3 to get the desired inequality. $\qquad\square$

Finally, we record the following simple observation connecting weak functional inequalities in discrete domains.

Fact A.7. *Let π be a distribution on a finite state space Ω, and set $C_\pi = \frac{1-2\pi_{\min}}{\log(1/\pi_{\min}-1)}$. If π satisfies a $(\rho_{\mathrm{PI}}, \varepsilon)$-weak Poincaré inequality, then π also satisfies a $\left(4\rho_{\mathrm{PI}}C_\pi, \frac{\varepsilon}{C_\pi}\right)$-weak MLSI and a $(\rho_{\mathrm{PI}}C_\pi, \frac{\varepsilon}{C_\pi})$-weak LSI.*

Proof. For finite state spaces, it is well-known that the LSI of the complete graph Markov chain $P_{K(\pi)}$ has $\rho_{\mathrm{LS}} = \frac{1-2\pi_{\min}}{\log(1/\pi_{\min}-1)}$ (see e.g., [DSC96]). Furthermore, observe that $\mathcal{E}_{P_{K(\pi)}}(f,f) = \mathbf{Var}_\pi[f]$. Hence,

$$\frac{1}{\rho_{\mathrm{PI}}}\mathcal{E}(f,f) \geqslant \mathbf{Var}_\pi[f] - \varepsilon \cdot \mathrm{osc}(f)^2 \qquad\qquad \text{(Weak PI)}$$

$$\geqslant C_\pi\mathrm{Ent}_\pi[f^2] - \varepsilon \cdot \mathrm{osc}(f)^2, \qquad\qquad \text{(LSI of } P_{K(\pi)}\text{)}$$

which establishes the weak LSI. For the weak MLSI, one applies the inequality $4\mathcal{E}(f,f) \leqslant \mathcal{E}(f^2, \log f^2)$, whose proof reduces to checking the two-variable inequality $4(\sqrt{u} - \sqrt{v})^2 \leqslant (u-v)\log\frac{u}{v}$ for positive u, v. $\qquad\square$

B Deferred calculations for spherical spin glasses

B.1 The TAP Hamiltonian

In this subsection, we will prove Lemma 7.23, which we restate for convenience.

Lemma 7.23. *The law of Hamiltonian $H_{\mathrm{TAP}} \sim \mu_{\mathrm{TAP},x,m}$ is described by a Gaussian process $(H_{\mathrm{TAP}}(\sigma))_{\sigma\in S_N}$ defined by*

$$\mathbf{E}\,H_{\mathrm{TAP}}(\sigma) = N\xi_t(R(x,\sigma)) - \langle x, v(\sigma)\rangle \cdot \xi_t'(q_x) - \frac{\xi_t'(R(m,\sigma))}{\gamma'(q_m)} \cdot \langle m, \sigma\rangle \cdot \left(\theta'(q_m) - \frac{1}{1-q_m}\right)$$

$$\frac{1}{N}\mathbf{Cov}\big(H_{\mathrm{TAP}}(\sigma), H_{\mathrm{TAP}}(\sigma')\big) = \xi_t\big(R(\sigma,\sigma')\big) - R(\sigma,\sigma')\frac{\xi_t'(R(m,\sigma))\xi_t'(R(m,\sigma'))}{\xi_t'(q_m)}$$

$$+ \frac{\xi_t''(q_m)}{\gamma'(q_m)\xi_t'(q_m)}\gamma(R(m,\sigma))\gamma(R(m,\sigma')),$$

where

$$v(\sigma) := \frac{\xi_t'(R(\boldsymbol{m},\sigma))}{\xi_t'(q_{\boldsymbol{m}})} \left[I - \frac{\xi_t''(q_{\boldsymbol{m}})}{\gamma'(q_{\boldsymbol{m}})} \cdot \frac{\boldsymbol{m}\boldsymbol{m}^\top}{N} \right] \sigma$$

$$\gamma(q) := q \cdot \xi_t'(q) \,.$$

To prove the above, we will need the following formulas for any p-spin Hamiltonian H_N with mixture function ξ.

Fact B.1. *For any $u, v, m \in \mathbb{R}^N$, we have:*

$$\frac{1}{N}\mathbf{E}\langle u, \boldsymbol{\nabla} H_N(m)\rangle\langle v, \boldsymbol{\nabla} H_N(m)\rangle = R(u,v)\xi'(R(m,m)) + R(m,u)R(m,v)\xi''(R(m,m)) \,.$$

Proof. Once we write the derivative as its definition as a limit, the order of the limit and the expectation operator can be swapped by the dominated convergence theorem.

$$\frac{1}{N}\mathbf{E}\langle u, \boldsymbol{\nabla} H_N(m)\rangle\langle v, \boldsymbol{\nabla} H_N(m)\rangle = \frac{1}{N}\mathbf{E}\lim_{\delta,\varepsilon\to 0} \frac{H_N(m+\delta u) - H_N(m)}{\delta} \cdot \frac{H_N(m+\varepsilon v) - H_N(m)}{\varepsilon}$$

$$= \frac{1}{N}\lim_{\delta,\varepsilon\to 0}\frac{1}{\delta\varepsilon}\mathbf{E}(H_N(m+\delta u) - H_N(m)) \cdot (H_N(m+\varepsilon v) - H_N(m))$$

$$= \lim_{\delta,\varepsilon\to 0}\frac{1}{\delta\varepsilon}[\xi(R(m+\delta u, m+\varepsilon v)) - \xi(R(m+\delta u, m)) - \xi(R(m, m+\varepsilon v)) + \xi(R(m,m))]$$

$$= R(u,v)\xi'(R(m,m)) + R(m,u)R(m,v)\xi''(R(m,m)) \,. \qquad \square$$

Fact B.2. *For any $u, v, m \in \mathbb{R}^N$, we have:*

$$\frac{1}{N}\mathbf{E}\langle u, \boldsymbol{\nabla} H_N(m)\rangle H_N(v) = R(u,v)\xi'(R(m,v)) \,.$$

The proof of the above is analogous to the proof of [Fact B.1](), and hence omitted. We now prove [Lemma 7.23]().

Proof of [Lemma 7.23](). The distribution of $H_{\mathsf{TAP}}(\sigma)$ is the same as that of $H_{N,t}(\sigma)|x, \boldsymbol{\nabla}\mathcal{F}_{\mathsf{TAP}}(\boldsymbol{m}) = 0$. Recall that $H_{N,t}(\sigma) = N\xi_t(R(x,\sigma)) + \widetilde{H}(\sigma)$ where $\widetilde{H}(\sigma)$ is a centered Gaussian process. Next, observe that conditioning on $\boldsymbol{\nabla}\mathcal{F}_{\mathsf{TAP}}(\boldsymbol{m}) = 0$ is the same as conditioning on

$$\boldsymbol{\nabla}\widetilde{H}(\boldsymbol{m}) = -x \cdot \xi_t'(q_x) - \boldsymbol{m} \cdot \left(\theta'(q_{\boldsymbol{m}}) - \frac{1}{1 - q_{\boldsymbol{m}}} \right) \,. \tag{34}$$

Observe that $(H_{\mathsf{TAP}}(\sigma))_{\sigma \in S_N}$ is a Gaussian process, since it is obtained by conditioning on another Gaussian process satisfying affine constraints. First, observe that we can write

$$H_{\mathsf{TAP}}(\sigma) = N\xi_t(R(x,\sigma)) + \widetilde{H}_{\mathsf{TAP}}(\sigma) \tag{35}$$

where $\widetilde{H}_{\mathsf{TAP}}(\sigma) = \widetilde{H}(\sigma)|x, \boldsymbol{\nabla}\mathcal{F}_{\mathsf{TAP}}(\boldsymbol{m}) = 0$. To understand the behavior of $\widetilde{H}_{\mathsf{TAP}}(\sigma)$, we break $\widetilde{H}(\sigma)$ into a sum of two terms: one term for its projection onto the space $U := \left\{ \left\langle \boldsymbol{\nabla}\widetilde{H}(\boldsymbol{m}), u \right\rangle : u \in \mathbb{R}^N \right\}$, and the part that is orthogonal to U, and thus independent of $\boldsymbol{\nabla}\widetilde{H}(\boldsymbol{m})$. Concretely, let us write

$$\widetilde{H}(\sigma) = \left\langle \boldsymbol{\nabla}\widetilde{H}(\boldsymbol{m}), v(\sigma) \right\rangle + \left(\widetilde{H}(\sigma) - \left\langle \boldsymbol{\nabla}\widetilde{H}(\boldsymbol{m}), v(\sigma) \right\rangle \right) \,. \tag{36}$$

This is true for any $v(\sigma)$, but we have set up the definition such that the second summand is independent of $\nabla \widetilde{H}(m)$. To verify this, since these two random variables are each mean 0, it suffices to check that for any $u \in S_N$,

$$\mathbf{E}\left\langle \nabla \widetilde{H}(m), u \right\rangle \left(\widetilde{H}(\sigma) - \left\langle \nabla \widetilde{H}(m), v(\sigma) \right\rangle \right) = 0.$$

By Facts B.1 and B.2, the left-hand-side of the above is:

$$R(u,\sigma)\xi_t'(R(m,\sigma)) - R(u,v(\sigma))\xi_t'(q_m) - R(m,u)R(m,v(\sigma))\xi_t''(q_m).$$

We would like $v(\sigma)$ to be such that this is 0 for *all* u. Setting u orthogonal to m and σ shows that we must have $v(\sigma)$ in the subspace spanned by m and σ.

Suppose that $v(\sigma) = \alpha\sigma + \beta m$. Then, plugging this into the above requires that

$$\begin{aligned}
0 &= R(\sigma,u)\xi_t'(R(m,\sigma)) - (\alpha R(\sigma,u) + \beta R(m,u))\, \xi_t'(q_m) - R(m,u)\,(\alpha R(m,\sigma) + \beta q_m)\, \xi_t''(q_m) \\
&= R(\sigma,u)\,(\xi_t'(R(m,\sigma)) - \alpha\xi_t'(q_m)) \\
&\quad - R(m,u)\,(\beta\xi_t'(q_m) - \alpha R(m,\sigma)\xi_t''(q_m) - \beta q_m \xi_t''(q_m))\,.
\end{aligned}$$

Since this is true for all u, each of these two terms must be 0. That is,

$$\alpha = \frac{\xi_t'(R(m,\sigma))}{\xi_t'(q_m)}$$

and

$$\beta = -\alpha \cdot \frac{R(m,\sigma)\xi_t''(q_m)}{\xi_t'(q_m) + q_m \xi_t''(q_m)},$$

so

$$\begin{aligned}
v(\sigma) &= \frac{\xi_t'(R(m,\sigma))}{\xi_t'(q_m)} \left(\sigma - m \cdot \frac{R(m,\sigma)\xi_t''(q_m)}{\xi_t'(q_m) + q_m \xi_t''(q_m)} \right) \\
&= \frac{\xi_t'(R(m,\sigma))}{\xi_t'(q_m)} \left(\mathrm{Id} - \frac{R(m,\sigma)\xi_t''(q_m)}{\xi_t'(q_m) + q_m \xi_t''(q_m)} \cdot \frac{mm^\top}{N} \right) \sigma
\end{aligned}$$

as defined.

Now returning to (36), when we condition on x and $\nabla \mathcal{F}_{\mathsf{TAP}}(m) = 0$, by plugging in (34), we get

$$\begin{aligned}
\widetilde{H}_{\mathsf{TAP}}(\sigma) = {}& -\langle x, v(\sigma) \rangle \cdot \xi_t'(q_x) - \langle m, v(\sigma) \rangle \cdot \left(\theta'(q_m) - \frac{1}{1 - q_m} \right) \\
& + \left(\widetilde{H}(\sigma) - \left\langle \nabla \widetilde{H}(m), v(\sigma) \right\rangle \right) \tag{37}
\end{aligned}$$

We use $\widehat{H}(\sigma)$ to denote the random variable $\widetilde{H}(\sigma) - \left\langle \nabla \widetilde{H}(m), v(\sigma) \right\rangle$, whose distribution remains unaffected by the conditioning, as this random variable is independent of x and the event $\nabla \mathcal{F}_{\mathsf{TAP}}(m) = 0$. Since $\widehat{H}(\sigma)$ is centered, our expression for $\mathbf{E}\, H_{\mathsf{TAP}}(\sigma)$ follows from (35) and (37), and the observation that

$$R(m,v(\sigma)) = \frac{\xi_t'(R(m,\sigma))}{\gamma'(q_m)} \cdot R(m,\sigma).$$

It remains to compute $N^{-1}\mathbf{Cov}(H_{\mathsf{TAP}}(\sigma), H_{\mathsf{TAP}}(\sigma'))$ for any $\sigma, \sigma' \in S_N$. Observe that this is equal to $N^{-1}\mathbf{E}\widehat{H}(\sigma)\widehat{H}(\sigma')$. By Facts B.1 and B.2, we have that this is equal to:

$$\xi_t\big(R(\sigma, \sigma')\big) - R(v(\sigma), \sigma')\xi_t'(R(\boldsymbol{m}, \sigma')) - R(v(\sigma'), \sigma)\xi_t'(R(\boldsymbol{m}, \sigma))$$
$$+ R(v(\sigma), v(\sigma'))\xi_t'(q_m) + R(\boldsymbol{m}, v(\sigma))R(\boldsymbol{m}, v(\sigma'))\xi''(q_m).$$

The formula for the covariance can be obtained from the above by expanding $v(\sigma)$. □

Next, we look at the mixture function of these "TAP planted distributions on slices".

Corollary 7.26. *For a fixed choice of a and b, the Gaussian process $(H_{\mathsf{TAP}}(v(a,b) + r_{a,b}Q\tau))_{\tau \in S_{N-2}}$ is described by the following law.*

- *Let $H_{a,b}$ be a spherical p-spin Hamiltonian with mixture function $\xi_{a,b}$ given by:*

$$\xi_{a,b}(s) := \xi_t\Big(\|v(a,b)\|^2 + r_{a,b}^2 s\Big) - \xi_t\Big(\|v(a,b)\|^2\Big) - s \cdot \frac{r_{a,b}^2 \xi_t'\Big(q_m \cdot \big(1 + \frac{a}{\sqrt{N}}\big)\Big)^2}{\xi_t'(q_m)}.$$

- *Let $V(a,b) := \xi_t\Big(\|v(a,b)\|^2\Big) - \|v(a,b)\|^2 \cdot \frac{\xi_t'\big((1 + \frac{a}{\sqrt{N}})q_m\big)^2}{\xi_t'(q_m)} + \frac{\xi_t''(q_m)}{\gamma'(q_m)\xi_t'(q_m)} \cdot \gamma\Big(\big(1 + \frac{a}{\sqrt{N}}\big)q_m\Big)^2.$*

The law of $H_{\mathsf{TAP}}(v(a,b) + r_{a,b}Q\tau)$ is the same as that of $H_{a,b}(\tau) + \sqrt{N} \cdot g_{a,b} + \mathbf{E}_{\mu_{\mathsf{TAP}}}H_{\mathsf{TAP}}(v(a,b) + r_{a,b}Q\tau)$ where $g_{a,b}$ is a centered Gaussian of variance $V(a,b)$ independent of $H_{a,b}$.

Proof. Let $\tau, \tau' \in S_{N-2}$, and

$$\sigma = v(a,b) + r_{a,b}Q\tau \text{ and } \sigma' = v(a,b) + r_{a,b}Q\tau'.$$

Recall from Lemma 7.23 that

$$N^{-1}\mathbf{Cov}\left(H_{\mathsf{TAP}}(\sigma), H_{\mathsf{TAP}}(\sigma')\right)$$
$$= \xi_t'(R(\sigma,\sigma')) - R(\sigma,\sigma')\frac{\xi_t'(R(\boldsymbol{m},\sigma))\xi_t'(R(\boldsymbol{m},\sigma'))}{\xi_t'(q_m)} + \frac{\xi_t''(q_m)}{\gamma'(q_m)\xi_t'(q_m)}\gamma(R(\boldsymbol{m},\sigma))\gamma(R(\boldsymbol{m},\sigma')).$$

By the definition of σ and σ', we have $R(\boldsymbol{m}, \sigma) = R(\boldsymbol{m}, \sigma') = R(\boldsymbol{m}, v(a,b)) = \big(1 + \frac{a}{\sqrt{N}}\big)q_m$, and $R(\sigma,\sigma') = R\Big(\|v(a,b)\|^2 + r_{a,b}^2 R(\tau,\tau')\Big)$, since Q is an isometry. As a result,

$$N^{-1}\mathbf{Cov}\left(H_{\mathsf{TAP}}(\sigma), H_{\mathsf{TAP}}(\sigma')\right)$$

$$= \xi_t'(\|v(a,b)\|^2 + r_{a,b}^2 R(\tau,\tau')) - (\|v(a,b)\|^2 + r_{a,b}^2 R(\tau,\tau'))\frac{\xi_t'\big((1 + \frac{a}{\sqrt{N}})q_m\big)^2}{\xi_t'(q_m)}$$

$$+ \frac{\xi_t''(q_m)}{\gamma'(q_m)\xi_t'(q_m)}\gamma\Big(\big(1 + \frac{a}{\sqrt{N}}\big)q_m\Big)^2.$$

This may be written as

$$N^{-1}\mathbf{Cov}\left(H_{\mathsf{TAP}}(\sigma), H_{\mathsf{TAP}}(\sigma')\right) = \xi_{a,b}(R(\tau,\tau')) + V(a,b).$$

This implies that $H_{\text{TAP}}(\sigma)$ is equal to $H_{a,b}(\tau) + g_{a,b}$ for some Gaussian process $(H_{a,b}(\tau))_{\tau \in S_{N-2}}$, where $g_{a,b}$ is a centered Gaussian of variance $V(a,b)$. To complete the proof, we must show that the correlation structure of $H_{a,b}$ can be achieved by a p-spin model with mixture function $\xi_{a,b}$. To do this, it suffices to show that $\xi_{a,b}$ is indeed a valid mixture function, in that $\xi_{a,b}^{(p)}(0) \geqslant 0$ for all $p \geqslant 1$, and $\xi_{a,b}(0) = 0$. The latter of these is clearly true by construction. The former is easily seen to be true for $p \geqslant 2$, since for such p,

$$\xi_{a,b}^{(p)}(0) = r_{a,b}^{2p} \xi_t^{(p)}(\|v(a,b)\|^2) \geqslant 0$$

since ξ_t is a valid mixture function. For $p = 1$,

$$\xi_{a,b}'(0) = r_{a,b}^2 \cdot \left(\xi_t'\left(\|v(a,b)\|^2\right) - \frac{\xi_t'\left(q_m\left(1 + \frac{a}{\sqrt{N}}\right)\right)^2}{\xi_t'(q_m)} \right)$$

$$\overset{(9)}{\geqslant} r_{a,b}^2 \cdot \left(\xi_t'\left(q_m\left(1 + \frac{a}{\sqrt{N}}\right)^2\right) - \frac{\xi_t'\left(q_m\left(1 + \frac{a}{\sqrt{N}}\right)\right)^2}{\xi_t'(q_m)} \right)$$

$$= \frac{r_{a,b}^2}{\xi_t'(q_m)} \cdot \left(\xi_t'\left(q_m\left(1 + \frac{a}{\sqrt{N}}\right)^2\right) \xi_t'(q_m) - \xi_t'\left(q_m\left(1 + \frac{a}{\sqrt{N}}\right)\right)^2 \right) \geqslant 0.$$

In the first inequality above, we use the fact that ξ_t' is non-decreasing. The final inequality is an application of Cauchy-Schwarz. $\qquad \square$

B.2 Understanding concentration around the codimension-2 slice

Next, we bound the variance of $g_{a,b} - g_{0,0}$.

Lemma 7.33. *For every constant $\iota > 0$, there is a constant c such that with probability $1 - e^{-cN}$, for all a, b, we have $|g_{a,b} - g_{0,0}| \leqslant \iota \frac{a^2 + b^2}{\sqrt{N}}$.*

Proof of Lemma 7.33. The strategy is to prove that for any $a, b, a', b' \in \mathbb{R}$ of magnitude $\ll N^{1/4}$, $g_{a,b} - g_{a',b'}$ is a Gaussian of variance $O\left(\cdot \|v(a,b) - v(a',b')\|^4\right) = O\left(\frac{(a-a')^4 + (b-b')^4}{N^2}\right)$. The desideratum then immediately follows by applying Slepian's lemma on $(|g_{a,b} - g_{0,0}|)_{a,b}$ comparing it to the Gaussian process $\langle G, (v(a,b) - v(0,0))(v(a,b) - v(0,0))^\top \rangle$ for a standard Gaussian matrix G.

We carry out the calculation for $a', b' = 0$; the general case follows similarly. We have $g_{a,b} = N^{-1/2}\left(H_{\text{TAP}}(\sqrt{N} \cdot v(a,b)) - \mathbf{E}H_{\text{TAP}}(\sqrt{N} \cdot v(a,b))\right)$, and $g_{0,0} = N^{-1/2}(H_{\text{TAP}}(m) - \mathbf{E}H_{\text{TAP}}(m))$. Clearly, $g_{a,b} - g_{0,0}$ is a centered Gaussian process. As in the proof of Lemma 7.23, we have that the distribution of $H_{\text{TAP}}(\sigma) - \mathbf{E}H_{\text{TAP}}(\sigma)$ is the same as that of

$$\widetilde{H}(\sigma) - \left\langle \boldsymbol{\nabla}\widetilde{H}(m), v(\sigma) \right\rangle,$$

where $\widetilde{H}$ is a Hamiltonian distributed according to the mixture function ξ_t. Thus, the distribution of $g_{a,b} - g_{0,0}$ is:

$$\widetilde{H}(u(a,b)) - \left\langle \boldsymbol{\nabla}\widetilde{H}(m), v(u(a,b)) \right\rangle - \widetilde{H}(m) + \left\langle \boldsymbol{\nabla}\widetilde{H}(m), v(m) \right\rangle.$$

For brevity, we denote $v(a, b)$ as $m + \varepsilon$. We express $\tilde{H}(m + \varepsilon)$ in its Taylor expansion, and we get:

$$\sqrt{N}(g_{a,b} - g_{0,0}) = \sum_{i \geqslant 1} \frac{1}{i!} \left\langle D_i \tilde{H}(m), \varepsilon^{\otimes i} \right\rangle - \left\langle \nabla \tilde{H}(m), v(m + \varepsilon) - v(m) \right\rangle. \tag{38}$$

Expanding out $v(m + \varepsilon) - v(m)$ ultimately yields:

$$v(m + \varepsilon) - v(m) = \varepsilon + R(m, \varepsilon)\varepsilon \frac{\xi_t''(q_m)}{\gamma'(q_m)} - \frac{R(m, \varepsilon)^2 \xi_t''(q_m)^2}{\xi_t'(q_m)\gamma'(q_m)} m.$$

Plugging in the above into Eq. (38) gives:

$$\sqrt{N}(g_{a,b} - g_{0,0}) = \sum_{i \geqslant 2} \frac{1}{i!} \left\langle D_i \tilde{H}(m), \varepsilon^{\otimes i} \right\rangle$$

$$- \left\langle \nabla \tilde{H}(m), \varepsilon \right\rangle R(m, \varepsilon) \frac{\xi_t''(q_m)}{\gamma'(q_m)} - \left\langle \nabla \tilde{H}(m), m \right\rangle \frac{R(m, \varepsilon)^2 \xi_t''(q_m)^2}{\xi_t'(q_m)\gamma'(q_m)}$$

We have an explicit expression for ε:

$$\varepsilon = \frac{aq_m - bq_x^2}{\sqrt{N}(q_m - q_x^2)} m + \frac{q_m q_x}{q_m - q_x^2} \left(\frac{b - a}{\sqrt{N}} \right) x.$$

This explicit expression can be used to obtain the following bounds on the variances of the above terms:

$$\mathbf{Var} \left[\frac{1}{i!} \left\langle D_i \tilde{H}(m), \varepsilon^{\otimes i} \right\rangle \right] \leqslant \frac{O(a^{2i} + b^{2i})}{N^{i-1}}$$

$$\mathbf{Var} \left[\left\langle \nabla \tilde{H}(m), \varepsilon \right\rangle R(m, \varepsilon) \frac{\xi_t''(q_m)}{\gamma'(q_m)} \right] = \frac{O(a^4 + b^4)}{N}$$

$$\mathbf{Var} \left[\left\langle \nabla \tilde{H}(m), m \right\rangle \frac{R(m, \varepsilon)^2 \xi_t''(q_m)}{\xi_t'(q_m)\gamma'(q_m)} \right] \leqslant \frac{O(a^4 + b^4)}{N}$$

The expression for $\sqrt{N}(g_{a,b} - g_{0,0})$ only involves a constant number of terms, and since the first term enumerates over $i \geqslant 2$, and since $|a|, |b| \leqslant \sqrt{N}$, we have an overall bound of $\frac{O(a^4 + b^4)}{N}$. Dividing by $\sqrt{N}$ gives the desired variance bound. $\qquad \square$

Lemma 7.31. $\nabla \hat{E}_{a,b} \big|_{(a,b)=(0,0)} = 0.$

Proof. Recall

$$\hat{E}_{a,b} = \frac{1}{2} \left(\underbrace{\log r_{a,b}^2 - \xi_t(\|v(a,b)\|^2) - r_{a,b}^2 \cdot \frac{\xi_t' \left(q_m \left(1 + \frac{a}{\sqrt{N}} \right) \right)^2}{\xi_t'(q_m)}}_{(I)} \right.$$

$$\left. + \underbrace{\xi_t \left(q_x \left(1 + \frac{b}{\sqrt{N}} \right) \right) + \frac{\gamma \left(q_m \left(1 + \frac{a}{\sqrt{N}} \right) \right)}{\gamma'(q_m)} \cdot \left((1 - q_m)\xi_t''(q_m) + \frac{1}{1 - q_m} \right)}_{(II)} \right)$$

79

$$
-\frac{\gamma(q_x)}{\xi_t'(q_m)} \cdot \underbrace{\xi_t'\left(q_m\left(1+\frac{a}{\sqrt{N}}\right)\right) \cdot \left(\left(1+\frac{b}{\sqrt{N}}\right) - q_m \cdot \frac{\xi_t''(q_m)}{\gamma'(q_m)} \cdot \left(1+\frac{a}{\sqrt{N}}\right)\right)}_{\text{(III)}}.
$$

Because

$$
\|v(a,b)\|^2 = q_m\left(1+\frac{a}{\sqrt{N}}\right)^2 + \frac{q_m q_x}{q_m - q_x^2}\cdot\left(\frac{a-b}{\sqrt{N}}\right)^2,
$$

we have

$$
\boldsymbol{\nabla}\|v(a,b)\|^2\big|_{(a,b)=(0,0)} = -\boldsymbol{\nabla} r_{a,b}^2\big|_{(a,b)=(0,0)} = \left(\frac{2q_m}{\sqrt{N}},0\right).
$$

We also have $r_{0,0}^2 = 1 - q_m$ and $\|v(0,0)\|^2 = q_m$. Let us start by computing the derivative with respect to a. We have

$$
\sqrt{N}\cdot\partial_a(\text{III})\big|_{(a,b)=(0,0)} = \xi_t''(q_m)\cdot q_m\cdot\left(1-\frac{q_m\xi_t''(q_m)}{\gamma'(q_m)}\right) + \xi_t'(q_m)\cdot\left(-q_m\cdot\frac{\xi_t''(q_m)}{\gamma'(q_m)}\right)
$$
$$
= \frac{q_m\xi_t''(q_m)}{\gamma'(q_m)}\left(\gamma'(q_m) - q_m\xi_t''(q_m) - \xi_t'(q_m)\right) = 0.
$$

Next,

$$
\sqrt{N}\cdot\partial_a(\text{I})\big|_{(a,b)=(0,0)}
$$
$$
= \frac{1}{r_{0,0}^2}\cdot(-2q_m) - \xi_t'(\|v(0,0)\|^2)\cdot(2q_m) - (-2q_m)\cdot\left(\frac{\xi_t'(q_m)^2}{\xi_t'(q_m)}\right) - r_{0,0}^2\cdot\frac{2\xi_t'(q_m)\cdot\xi_t''(q_m)\cdot q_m}{\xi_t'(q_m)}
$$
$$
= \frac{-2q_m}{1-q_m} - 2q_m\xi_t'(q_m) + 2q_m\xi_t'(q_m) - 2q_m(1-q_m)\xi_t''(q_m)
$$
$$
= \frac{-2q_m}{1-q_m} - 2q_m(1-q_m)\xi_t''(q_m).
$$

Finally,

$$
\sqrt{N}\cdot\partial_a(\text{II})\big|_{(a,b)=(0,0)}
$$
$$
= \frac{\gamma'(q_m)\cdot q_m}{\gamma'(q_m)}\cdot\left((1-q_m)\xi_t''(q_m) + \frac{1}{1-q_m}\right)
$$
$$
= -\frac{1}{2}\cdot\sqrt{N}\cdot\partial_a(\text{I})\big|_{(a,b)=(0,0)}
$$

as desired. The derivative with respect to b is much simpler, since the derivative of $r_{a,b}^2$ with respect to b is 0 at $(0,0)$. Consequently, $\partial_b(\text{I})\big|_{(a,b)=(0,0)} = 0$, $\partial_b(\text{II})\big|_{(a,b)=(0,0)} = q_x\xi_t'(q_x) = \gamma(q_x)$, and $\partial_b(\text{III})\big|_{(a,b)=(0,0)} = \xi_t'(q_m)$, completing the proof. $\square$

Lemma 7.32. *There exist constants $\eta, \varepsilon > 0$ such that for all $|a|, |b| \leqslant \varepsilon\sqrt{N}$, $N\boldsymbol{\nabla}^2\widehat{E}_{a,b} \preceq -\eta\text{Id}$.*

Proof. For ease of notation, define $\widetilde{E}_{a,b} = \widehat{E}_{\sqrt{N}a,\sqrt{N}b}$, $\widetilde{r}_{a,b} = r_{\sqrt{N}a,\sqrt{N}b}$, and $\widetilde{v}(a,b) = v(\sqrt{N}a,\sqrt{N}b)$. As in the previous lemma, recall

$$
\widetilde{E}_{a,b} = \frac{1}{2}\left(\log\widetilde{r}_{a,b}^2 - \xi_t(\|\widetilde{v}(a,b)\|^2) - \widetilde{r}_{a,b}^2\cdot\frac{\xi_t'(q_m(1+a))^2}{\xi_t'(q_m)}\right)
$$

$$+ \xi_t \left(q_x \left(1 + b \right) \right) + \frac{\gamma \left(q_m \left(1 + a \right) \right)}{\gamma'(q_m)} \cdot \left(\left(1 - q_m \right) \xi_t''(q_m) + \frac{1}{1 - q_m} \right)$$

$$- \frac{\gamma(q_x)}{\xi_t'(q_m)} \cdot \xi_t' \left(q_m \left(1 + a \right) \right) \cdot \left(\left(1 + b \right) - q_m \cdot \frac{\xi_t''(q_m)}{\gamma'(q_m)} \cdot \left(1 + a \right) \right).$$

Because the Hessian is Lipschitz in all the parameters involved, it suffices to prove the negative definiteness of the Hessian at $(0, 0)$, under the assumption that $q_m = q_x = q$, where q (formerly denoted $q_*(t)$) satisfies $\xi_t'(q) = \frac{q}{1-q}$. Under these constraints, we have $\gamma(q) = q\xi_t'(q) = \frac{q^2}{1-q}$ and $\gamma'(q) = \xi_t'(q) + q\xi_t''(q) = \frac{q}{1-q} \left(1 + (1 - q)\xi_t''(q) \right)$. $\widetilde{E}_{a,b}$ simplifies as

$$\widetilde{E}_{a,b} = \frac{1}{2} \underbrace{\left(\log \widetilde{r}_{a,b}^2 - \xi_t(\|\widetilde{v}(a,b)\|^2) - \widetilde{r}_{a,b}^2 \cdot \frac{\xi_t' \left(q \left(1 + a \right) \right)^2}{\xi_t'(q)} \right)}_{(I)}$$

$$+ \underbrace{\xi_t \left(q \left(1 + b \right) \right) + \frac{\gamma \left(q \left(1 + a \right) \right)}{q} \cdot \frac{1 + (1 - q)^2 \xi_t''(q)}{1 + (1 - q)\xi_t''(q)}}_{(II)}$$

$$- \underbrace{q \left(1 + b \right) \cdot \xi_t' \left(q \left(1 + a \right) \right) + q^2 \left(1 + a \right) \cdot \frac{\xi_t''(q)}{\gamma'(q)}}_{(III)}.$$

We have $\partial_b^2(III)\big|_{(0,0)} = 0$, and

$$\partial_b^2(II)\big|_{(0,0)} = \xi_t''(q) \cdot q^2.$$

We have that $\partial_b \widetilde{r}_{a,b}^2\big|_{(0,0)} = 0$, and $\partial_b^2 \widetilde{r}_{a,b}^2\big|_{(0,0)} = \frac{-2q^2}{1-q}$. Consequently,

$$\partial_b^2(I)\big|_{(0,0)} = \frac{1}{2} \left(\frac{1}{r_{0,0}^2} \cdot \partial_b^2 \widetilde{r}_{a,b}^2\big|_{(0,0)} - \xi_t'(\|v(0,0)\|^2) \cdot \partial_b^2 \|\widetilde{v}(a,b)\|^2\big|_{(0,0)} - \partial_b^2 \widetilde{r}_{a,b}^2\big|_{(0,0)} \cdot \xi_t'(q) \right)$$

$$= \frac{-q^2}{(1-q)^2}.$$

It follows that

$$\partial_b^2 \widehat{E}_{a,b}\big|_{(0,0)} = \xi_t''(q) \cdot q^2 - \frac{q^2}{(1-q)^2}.$$

Similarly, we have $\partial_a \partial_b(II)\big|_{(0,0)} = 0$, and

$$\partial_a \partial_b(III)\big|_{(0,0)} = \xi_t''(q) \cdot q^2.$$

We have that $\partial_a \partial_b \widetilde{r}_{a,b}^2\big|_{(0,0)} = \frac{-2q^2}{1-q}$. Consequently,

$$\partial_a \partial_b(I)\big|_{(0,0)} = \frac{1}{2} \left(\frac{1}{r_{0,0}^2} \cdot \partial_a \partial_b \widetilde{r}_{a,b}^2\big|_{(0,0)} - \xi_t'(\|v(0,0)\|^2) \cdot \partial_a \partial_b \|\widetilde{v}(a,b)\|^2\big|_{(0,0)} - \partial_a \partial_b \widetilde{r}_{a,b}^2\big|_{(0,0)} \cdot \xi_t'(q) \right)$$

$$= \frac{-q^2}{(1-q)^2}.$$

It follows that

$$\partial_a \partial_b \widehat{E}_{a,b}\Big|_{(0,0)} = \xi_t''(q) \cdot q^2 - \frac{q^2}{(1-q)^2}.$$

It remains to compute the second derivative with respect to a. We have

$$\partial_a^2 (\text{III})\Big|_{(0,0)} = q^3 \xi_t'''(q).$$

We also have

$$\partial_a^2 (\text{II})\Big|_{(0,0)} = q\gamma''(q) \cdot \frac{1 + (1-q)^2 \xi_t''(q)}{1 + (1-q)\xi_t''(q)}$$

$$= q \cdot \frac{1 + (1-q)^2 \xi_t''(q)}{1 + (1-q)\xi_t''(q)} \cdot \left(2\xi_t''(q) + q\xi_t'''(q)\right)$$

$$\leqslant q^2 \xi_t'''(q) + 2q\xi_t''(q).$$

We have $\partial_a^2 \widetilde{r}_{a,b}^2\Big|_{(0,0)} = -\left(2q + \frac{2q^2}{1-q}\right) = -\frac{2q}{1-q}$ and $\partial_a \widetilde{r}_{a,b}^2\Big|_{(0,0)} = -2q$. Finally,

$$\partial_a^2 (\text{I})\Big|_{(0,0)} = \frac{1}{r_{0,0}^2} \cdot \frac{-2q}{1-q} - \frac{1}{r_{0,0}^4} \cdot (-2q)^2 - \xi_t'(q) \cdot \frac{-2q}{1-q} - \xi_t''(q) \cdot (-2q)^2 - \frac{-2q}{1-q} \cdot \xi_t'(q)$$

$$- 2 \cdot (-2q) \cdot \frac{2\xi_t'(q)\xi_t''(q)q}{\xi_t'(q)} - (1-q) \cdot \frac{2q^2(\xi_t'(q)\xi_t'''(q) + \xi_t''(q)^2)}{\xi_t'(q)}$$

$$= \frac{-2q}{(1-q)^2} - \frac{4q^2}{(1-q)^2} - 4q^2\xi_t''(q) + 8q^2\xi_t''(q) - 2q^2(1-q)\xi_t'''(q) - 2q^2(1-q) \cdot \frac{\xi_t''(q)^2}{\xi_t'(q)}$$

$$= \frac{-2q(2q+1)}{(1-q)^2} + 4q^2\xi_t''(q) - 2q^2(1-q)\xi_t'''(q) - 2q(1-q)^2\xi_t''(q)^2.$$

Therefore,

$$\partial_a^2 \widehat{E}_{a,b}\Big|_{(0,0)} \leqslant -q^3\xi_t'''(q) + q^2\xi_t'''(q) + 2q\xi_t''(q) - \frac{q(2q+1)}{(1-q)^2}$$

$$+ 2q^2\xi_t''(q) - q^2(1-q)\xi_t'''(q) - q(1-q)^2\xi_t''(q)^2$$

$$= \frac{-q(2q+1)}{(1-q)^2} + 2q\xi_t''(q) + 2q^2\xi_t''(q) - q(1-q)^2\xi_t''(q)^2.$$

To conclude, let us check that the Hessian is negative definite. Because $\xi_t''(q) \cdot q^2 - \frac{q^2}{(1-q)^2} < 0$ by the SL condition (SL), it suffices to check that

$$\partial_a^2 \widehat{E}_{a,b}\Big|_{(0,0)} < q^2\xi_t''(q) - \frac{q^2}{(1-q)^2}.$$

This is true if and only if

$$q(1-q)^2\xi_t''(q)^2 - q(q+2)\xi_t''(q) + \frac{q(q+1)}{(1-q)^2} > 0.$$

It is not difficult to see that this is true if $\xi_t''(q)$ is less than the smaller root of the above quadratic, which is equal to

$$\frac{(q+2) - \sqrt{(q+2)^2 - 4(q+1)}}{2(1-q)^2} = \frac{1}{(1-q)^2}.$$

This is true by the SL condition (SL), concluding the proof. $\qquad \square$

Next, we shall prove Lemma 7.35. Recall the definition

$$\text{Error}_{a,b}^{(2)} = \frac{N\xi_{a,b}''(0)}{4} - \frac{1}{2}\log\det\left((1+\xi_{a,b}''(0))\text{Id} - \nabla^2 H_{a,b}(0)\right),$$

where $\nabla^2 H_{a,b}(0)$ is equal to the restriction of $r_{a,b}^2 \cdot \nabla^2 H_{\text{TAP}}(v(a,b))$ restricted to the codimension-2 subspace orthogonal to m and x.

Lemma 7.35. *For any sufficiently small $\iota > 0$, with probability at least $1 - e^{-cN}$, $\left|\text{Error}_{a,b}^{(2)} - \text{Error}_{0,0}^{(2)}\right| = O(1)$ for all $a, b < \iota N^{1/4}$.*

Let us start by computing the correlation structure of the random matrices $\nabla^2 H_{a,b}(0)$. Note that $\nabla^2 H_{a,b}(0)$ is an $(N-2)$-dimensional GOE matrix scaled by $\sqrt{\xi_{a,b}''(0)}$.

Fact B.3. *For σ^1, σ^2,*

$$\frac{1}{N}\mathbb{E}\langle\nabla^2 H_{\text{TAP}}(\sigma^1), u^1 \otimes u^2\rangle\langle\nabla^2 H_{\text{TAP}}(\sigma^2), v^1 \otimes v^2\rangle$$

$$= \xi_t''\left(R(\sigma^1, \sigma^2)\right) \cdot \left(R(u^1, v^2) \cdot R(u^2, v^1) + R(u^1, v^1) \cdot R(u^2, v^2)\right).$$

In particular,

$$\frac{1}{N}\mathbb{E}\langle\nabla^2 H_{\text{TAP}}(\sigma^1), \nabla^2 H_{\text{TAP}}(\sigma^2)\rangle = \xi_t''\left(R(\sigma^1, \sigma^2)\right).$$

The above follows from calculations similar to those involved in the proofs of Facts B.1 and B.2; we omit the details.

Proof of Lemma 7.35. We shall prove the statement for a fixed a, b; a union bound over a, b implies the boundedness for all $a, b \leqslant \iota N^{1/4}$.

Recalling that $\nabla^2 H_{a,b}(0) = r_{a,b}^2 \nabla^2 H_{\text{TAP}}(\sqrt{N} \cdot v(a,b))$ is a GOE matrix scaled by $\sqrt{\xi_{a,b}''(0)}$. By Fact B.3,

$$\frac{1}{N}\mathbb{E}\langle\nabla^2 H_{a,b}(0), \nabla^2 H_{0,0}(0)\rangle = r_{a,b}^2 \cdot r_{0,0}^2 \cdot \frac{1}{N}\mathbb{E}\langle\nabla^2 H_{\text{TAP}}(\sqrt{N} \cdot v(a,b)), \nabla^2 H_{\text{TAP}}(\sqrt{N} \cdot v(0,0))\rangle$$

$$= r_{a,b}^2 \cdot r_{0,0}^2 \cdot \xi_t''\left(\langle v(a,b), v(0,0)\rangle\right).$$

For comparison, we have

$$\frac{1}{N}\mathbb{E}\|\nabla^2 H_{a,b}(0)\|_F^2 = \xi_{a,b}''(0) = r_{a,b}^4 \cdot \xi_t''\left(\|v(a,b)\|^2\right).$$

For succinctness of notation, let

$$\begin{pmatrix} \alpha_1 & \rho \\ \rho & \alpha_2 \end{pmatrix} = \begin{pmatrix} r_{0,0}^4\xi_t''(\|v(0,0)\|^2) & r_{a,b}^2 r_{0,0}^2\xi_t''\left(\langle v(a,b), v(0,0)\rangle\right) \\ r_{a,b}^2 r_{0,0}^2\xi_t''\left(\langle v(a,b), v(0,0)\rangle\right) & r_{a,b}^4\xi_t''(\|v(a,b)\|^2) \end{pmatrix}$$

be the covariance structure of the scaled GOE matrices $\nabla^2 H_{0,0}(0)$ and $\nabla^2 H_{a,b}(0)$. It is not difficult to see that $\alpha_2 = \alpha_1 + O\left(\frac{a^2+b^2}{N}\right)$, and ρ is between α_1 and α_2. Also note that $\alpha_1 = \xi_{0,0}''(0)$ and $\alpha_2 = \xi_{a,b}''(0)$. Then, for some choice of GOE matrices G and $\widetilde{G}$, we may write

$$\nabla^2 H_{0,0}(0) = \sqrt{\alpha_1}G$$

$$\nabla^2 H_{a,b}(0) = \frac{\rho}{\sqrt{\alpha_1}} G + \sqrt{\alpha_2 - \frac{\rho^2}{\alpha_1}} \widetilde{G}.$$

We thus have

$$2\left(\text{Error}_{0,0}^{(2)} - \text{Error}_{a,b}^{(2)}\right) = \frac{N\alpha_1}{2} - \frac{N\alpha_2}{2} - \log \det \left(\underbrace{(1+\alpha_1)\,\text{Id} - \sqrt{\alpha_1}G}_{M_1} \right)$$

$$+ \log \det \left((1+\alpha_2)\text{Id} - \frac{\rho}{\sqrt{\alpha_1}}G - \sqrt{\alpha_2 - \frac{\rho^2}{\alpha_1}} \widetilde{G} \right).$$

We may write the matrix inside the final log det as

$$(1+\alpha_2)\text{Id} - \frac{\rho}{\sqrt{\alpha_1}}G - \sqrt{\alpha_2 - \frac{\rho^2}{\alpha_1}} \widetilde{G}$$

$$= ((1+\alpha_1)\text{Id} - \sqrt{\alpha_1}G) + \underbrace{(\alpha_2 - \alpha_1)\text{Id} - \left(\frac{\rho}{\sqrt{\alpha_1}} - \sqrt{\alpha_1}\right)G}_{M_2} - \underbrace{\sqrt{\alpha_2 - \frac{\rho^2}{\alpha_1}} \widetilde{G}}_{M_3}.$$

The difference of the two log det terms is thus equal to

$$\log \det \left(\text{Id} + \underbrace{M_1^{-1/2}M_2M_1^{-1/2} + M_1^{-1/2}M_3M_1^{-1/2}}_{M} \right).$$

Observe that M_3 is a scaled GOE matrix independent of M_1 (and M_2). Taylor expanding the above, we shall control the trace and Frobenius norm of M. It may be verified that the higher-order terms, corresponding to higher Schatten norms, are $O(1)$. To control the trace and Frobenius norm, we shall essentially control their values in expectation. Standard concentration arguments for GOE matrices, along the lines of Lemma 8.6 using [GZ00, Lemma 1.2(b) and Corollary 1.6(b)], allow us to assume (with probability $1 - e^{-cN}$) that the eigenvalues of G are distributed according to the semicircular distribution up to some small Wasserstein perturbation. That is, with probability $1 - e^{-cN}$, denoting by $\lambda_i(G)$ the eigenvalues of G,

$$\text{Tr}\, M_1^{-1/2}M_2M_1^{-1/2} = \sum_{1 \leqslant i \leqslant N} \frac{(\alpha_2 - \alpha_1) - \left(\frac{\rho}{\sqrt{\alpha_1}} - \sqrt{\alpha_1}\right)\lambda_i(G)}{(1+\alpha_1) - \sqrt{\alpha_1}\lambda_i(G)}$$

$$= N \cdot \int \frac{(\alpha_2 - \alpha_1) - \left(\frac{\rho}{\sqrt{\alpha_1}} - \sqrt{\alpha_1}\right)u}{(1+\alpha_1) - \sqrt{\alpha_1}u} \, d\mu_{\text{sc}}(u) + O(1)$$

$$= N(\alpha_2 - \rho) + O(1),$$

where the final equality follows from the standard semicircle integral $\int \frac{1}{x-u}\, d\mu_{\text{sc}}(u) = \frac{1}{2}\left(x - \sqrt{x^2 - 4}\right)$. On the other hand, because $\widetilde{G}$ is independent of G,

$$\text{Tr}\, M_1^{-1/2}M_3M_1^{-1/2} = O(1)$$

with probability $1 - e^{-cN}$. Let us next control the Frobenius norms of these matrices. Again, because $\widetilde{G}$ is independent of G, with very high probability,

$$\|M\|_F^2 = O(1) + \|M_1^{-1/2} M_2 M_1^{-1/2}\|_F^2 + \|M_1^{-1/2} M_3 M_1^{-1/2}\|_F^2.$$

The first squared Frobenius norm is equal to

$$\sum \left(\frac{(\alpha_2 - \alpha_1) - \left(\frac{\rho}{\sqrt{\alpha_1}} - \sqrt{\alpha_1} \right) \lambda_i(G)}{(1 + \alpha_1) - \sqrt{\alpha_1} \lambda_i(G)} \right)^2.$$

Let ι such that $(1 + \alpha_1) - (2 + \iota)\sqrt{\alpha_1} > \iota$ (this uses strict replica symmetry). Then, with probability $1 - e^{-cN}$, $|\lambda_i(G)| \leqslant 2 + \iota$ for all i. Conditioned on this event happening, and recalling that $\alpha_2 - \alpha_1 = O\left(\frac{a^2 + b^2}{N} \right)$, the above is $\frac{O(a^2 + b^2)^2}{N}$. This is $O(1)$ for all choices of $a, b \leqslant \iota N^{1/4}$.

We must next control the squared Frobenius norm of $M_1^{-1/2} M_3 M_1^{-1/2}$. Let us condition on a typical realization of M_1: all its eigenvalues are smaller than $2 + \iota$ in magnitude, and the empirical spectral distribution is Wasserstein-close to the semicircle law in the same sense as the previous section (where we controlled the trace), in that

$$\sum \frac{1}{(1 + \alpha_1) - \sqrt{\alpha_1} \lambda_i(G)} = N \cdot \int \frac{1}{(1 + \alpha_1) - \sqrt{\alpha_1} u} \, d\mu_{sc}(u) + O(1) = N + O(1)$$

Because M_3 is independent of M_1, it suffices to control the expected Frobenius norm of the matrix – the true realization concentrates around its expectation to additive $O(1)$ factors. It is not difficult to see that this expectation is equal to

$$\frac{1}{N} \cdot \left(\alpha_2 - \frac{\rho^2}{\alpha_1} \right) \cdot \left(\sum \frac{1}{(1 + \alpha_1) - \sqrt{\alpha_1} \lambda_i(G)} \right)^2$$

$$= N \cdot \left(\alpha_2 - \frac{\rho^2}{\alpha_1} \right) \cdot \left(\int \frac{1}{(1 + \alpha_1) - \sqrt{\alpha_1} u} \, d\mu_{sc}(u) \right)^2 + O(1)$$

$$= N \cdot \left(\alpha_2 - \frac{\rho^2}{\alpha_1} \right) + O(1).$$

Putting the pieces together and returning to the Taylor expansion, we get that with very high probability,

$$2 \left(\text{Error}_{0,0}^{(2)} - \text{Error}_{a,b}^{(2)} \right)$$

$$= \frac{N\alpha_1}{2} - \frac{N\alpha_2}{2} + \log \det (\text{Id} + M)$$

$$= O(1) + \frac{N\alpha_1}{2} - \frac{N\alpha_2}{2} + \text{Tr}(M) - \frac{1}{2}\|M\|_F^2$$

$$= O(1) + N \cdot \left(\frac{\alpha_1}{2} - \frac{\alpha_2}{2} + (\alpha_2 - \rho) - \frac{1}{2}\left(\alpha_2 - \frac{\rho^2}{\alpha_1} \right) \right)$$

$$= O(1) + N \cdot \left(\frac{\alpha_1}{2} - \rho + \frac{\rho^2}{2\alpha_1} \right)$$

$$= O(1) + \frac{N}{2} \cdot \frac{(\alpha_1 - \rho)^2}{\alpha_1}.$$

Because $\alpha_1 - \rho = O\left(\frac{a^2 + b^2}{N} \right)$, this is $O(1)$ for $a, b \leqslant \iota N^{1/4}$, completing the proof. $\qquad\square$

B.3 Moment calculations for covariance bounds

We first prove subgaussian concentration for the covariance of the degree-2 part.

Proof of Proposition 8.8. Part (2) follows from part (1) by a standard tail integration argument. Indeed, the random variable W is bounded by $N^{1/2}$, so the contribution to $\mathbf{E}[W^k]$ from the event $|W| \geqslant N^{1/5}$ is bounded by

$$N^{k/2} \mathbb{P}(|W| \geqslant N^{1/5}) \leqslant N^{k/2} e^{-cN^{2/5}},$$

which is vanishing for any constant k. So, we focus on proving part (1). In the case $W = \langle \sigma^1, \sigma^2 \rangle / \sqrt{N}$, this is a special case of [HMP24, Lemma 7.5] (where we take $u = 0$). We consider the case $W = \langle \sigma, v_i \rangle$. Recall that

$$H_{N,2}(\sigma) = \langle A\sigma, \sigma \rangle = \frac{\sqrt{\xi''(0)}}{2} \langle M\sigma, \sigma \rangle$$

where $M \sim \mathsf{GOE}(N)$. For $0 \leqslant s \leqslant N^{1/5} \log N$, we will evaluate

$$\int e^{H_{N,2}(\sigma)} \, d\rho(\sigma) \qquad \text{and} \qquad \int e^{H_{N,2}(\sigma) + s\langle v_i, \sigma \rangle} \, d\rho(\sigma)$$

using [HMP24, Lemma 7.3]. We recall the function $G : (\lambda_{\max}(A), +\infty)$ defined in (15), which we copy below for convenience, and define $\widetilde{G}$ by

$$G(\gamma) = \gamma - \frac{1}{2N} \log \det(\gamma I - A), \qquad \widetilde{G}(\gamma) = G(\gamma) + \frac{s^2}{4N(\gamma - \lambda_i(A))}.$$

Recall from below (15) that G' has a unique root γ_* on $(\lambda_{\max}(A), +\infty)$. By the same argument, $\widetilde{G}'$ has a unique root $\widetilde{\gamma}_*$ on the same interval. As argued in the proof of Lemma 8.6, the conditions of [HMP24, Lemma 7.3] apply. Applying this lemma with $u = 0$ and $u = sv_i$, respectively, shows that with probability $1 - e^{-cN}$,

$$\int e^{H_{N,2}(\sigma)} \, d\rho(\sigma) = (1 + O(N^{-c})) \sqrt{\frac{2}{G''(\gamma_*)}} (2e)^{-N/2} \exp(NG(\gamma_*)),$$

$$\int e^{H_{N,2}(\sigma) + s\langle v_i, \sigma \rangle} \, d\rho(\sigma) = (1 + O(N^{-c})) \sqrt{\frac{2}{\widetilde{G}''(\widetilde{\gamma}_*)}} (2e)^{-N/2} \exp\left(N\widetilde{G}(\widetilde{\gamma}_*) \right). \tag{39}$$

Suppose further the probability $1 - e^{-cN}$ event in Lemma 8.7 holds. As argued in the proof of Proposition 8.1, $|\gamma_* - \gamma_0| \leqslant \frac{1}{C\sqrt{N}}$. Also, with probability $1 - e^{-cN}$, $\lambda_{\max}(A) \leqslant \frac{\sqrt{\xi''(0)}}{2}(2 + \varepsilon^2/8)$.

We will show that on the intersection of these events $|\gamma_* - \widetilde{\gamma}_*| = O(N^{-3/5} \log^2 N)$. First note that

$$\gamma_* - \lambda_{\max}(A) \geqslant \gamma_0 - \lambda_{\max}(A) - \frac{1}{C\sqrt{N}} \geqslant \frac{1 + \xi''(0) - (2 + \varepsilon^2/8)\sqrt{\xi''(0)}}{2} - \frac{1}{C\sqrt{N}} \geqslant \varepsilon^2/32$$

is bounded below by a constant, as in the proof of Fact 8.5. Since $G'(\gamma_*) = 0$ and $\widetilde{G}'(\gamma) = G'(\gamma) - \frac{s^2}{4N(\gamma - \lambda_i(A))^2}$, we have $\widetilde{\gamma}_* \geqslant \gamma_*$ and so $\widetilde{\gamma}_* - \lambda_i(A)$ is also bounded below by a constant. Thus, as $s \leqslant N^{1/5} \log N$, we have

$$0 \geqslant \widetilde{G}'(\gamma_*) = -\frac{s^2}{4N(\gamma - \lambda_i(A))^2} = O(N^{-3/5} \log^2 N).$$

By direct computation,

$$\widetilde{G}''(\gamma) = G''(\gamma) + \frac{s^2}{8N(\gamma - \lambda_i(A))^2}.$$

Since $\widetilde{\gamma}_* \geqslant \gamma_*$, we have $\gamma - \lambda_i(A)$ is bounded below by a constant for all $\gamma \in [\gamma_*, \widetilde{\gamma}_*]$. We claim that $\widetilde{\gamma}_* \in [\gamma_0 - N^{-1/2}, \gamma_0 + N^{-1/2}]$. Combining both conclusions of Lemma 8.7, we obtain that $G''(\gamma) \geqslant \Omega_\varepsilon(1)$ for $\gamma \in [\gamma_0 - N^{-1/2}, \gamma_0 + N^{-1/2}]$, and consequently the desired claim on $\widetilde{\gamma}_*$ holds.

We thus have $\widetilde{G}''(\gamma) = O_\varepsilon(1)$ in the same interval, and consequently we can conclude the stronger statement $|\gamma_* - \widetilde{\gamma}_*| = O(N^{-3/5} \log^2 N)$. Furthermore, since $\widetilde{G}^{(3)}(\gamma) = O_\varepsilon(1)$, the same logic allows us to also conclude

$$\widetilde{G}''(\widetilde{\gamma}_*)/\widetilde{G}''(\gamma_*) = 1 + O(N^{-3/5} \log^2 N).$$

By Taylor expanding $\widetilde{G}$ around $\widetilde{\gamma}_*$, we see

$$N|\widetilde{G}(\gamma_*) - \widetilde{G}(\widetilde{\gamma}_*)| \leqslant \frac{N}{2}|\gamma_* - \widetilde{\gamma}_*|^2 \sup_{\gamma \in [\gamma_0 - N^{-1/2}, \gamma_0 + N^{-1/2}]} G''(\gamma) = O(N^{-1/5} \log^4 N).$$

The above two displays allow us to replace instances of $\widetilde{\gamma}_*$ with γ_* in (39), yielding

$$\int e^{H_{N,2}(\sigma) + s\langle v_i, \sigma \rangle} \, d\rho(\sigma) = (1 + O(N^{-c})) \sqrt{\frac{2}{\widetilde{G}''(\gamma_*)}} (2e)^{-N/2} \exp\left(N\widetilde{G}(\gamma_*)\right),$$

and thus

$$\langle e^{s\langle v_i, \sigma \rangle} \rangle_2 = \frac{\int e^{H_{N,2}(\sigma) + s\langle v_i, \sigma \rangle} \, d\rho(\sigma)}{\int e^{H_{N,2}(\sigma)} \, d\rho(\sigma)}$$

$$= (1 + O(N^{-c})) \sqrt{\frac{G''(\gamma_*)}{\widetilde{G}''(\gamma_*)}} \exp\left(N(\widetilde{G}(\gamma_*) - G(\gamma_*))\right)$$

$$= (1 + O(N^{-c})) \exp\left(s^2/(4(\gamma_* - \lambda_i(A)))\right) = (1 + O(N^{-c})) \exp\left(cs^2\right),$$

where the last two steps again use that $\gamma_* - \lambda_i(A)$ is bounded away from 0. The tail estimate on $W = \langle v_i, \sigma \rangle$ now follows from a standard Chernoff bound. $\qquad\square$

We next turn to Lemma 8.11, proving each part in turn.

Proof of Lemma 8.11, Eq. (18). We reproduce Eq. (18) below for convenience:

$$\mathbf{E} \int_{S_N} \mathbf{1}[\sigma \notin T(H_N)] e^{H_N(\sigma)} \, d\rho(\sigma) \leqslant e^{N\xi(1)/2 - cN^{1/5}}.$$

The proof follows [HS23, Proposition 3.1], except with more precise control of overlaps between $N^{-2/5}$ and a small constant. By symmetry of the sphere, for any deterministic $x \in S_N$,

$$\mathbf{E} \int_{S_N} \mathbf{1}[\sigma \notin T(H_N)] e^{H_N(\sigma)} \, d\rho(\sigma) = \mathbf{E}\left[\mathbf{1}[x \notin T(H_N)] e^{H_N(x)}\right]. \tag{40}$$

Let $\mu_{\mathsf{pl}}(\cdot|x)$ denote the planted model (Definition 7.8) conditional on spike x. A Gaussian change of measure calculation implies that the right-hand side of (40) equals

$$e^{N\xi(1)/2} \mathbb{P}_{H_N^{x,\mathrm{I}} \sim \mu_{\mathsf{pl}}(\cdot|x)}[x \notin T(H_N^{x,\mathrm{I}})].$$

Thus it suffices to show

$$\mathbb{P}_{H_N^{x,\mathrm{I}} \sim \mu_{\mathrm{pl}}(\cdot|x)}[x \notin T(H_N^{x,\mathrm{I}})] \leqslant e^{-cN^{1/5}}.$$

Recall (Remark 7.14) that a sample $H_N^{x,\mathrm{I}} \sim \mu_{\mathrm{pl}}(\cdot|x)$ can be generated by

$$H_N^{x,\mathrm{I}}(\sigma) = N\xi(R(x,\sigma)) + \widetilde{H}_N(\sigma), \tag{41}$$

where $\widetilde{H}_N \sim \mu_{\mathrm{null}}$. Furthermore, from the definition, $x \in T(H_N^{x,\mathrm{I}})$ is equivalent to

$$\int_{S_N} \mathbf{1}[|R(x,\tau)| \geqslant N^{-2/5}] e^{H_N^{x,\mathrm{I}}(\tau)} \, \mathrm{d}\rho(\tau) \leqslant e^{N\xi(1)/2 - cN^{1/5}} \tag{42}$$

We will show this occurs with probability at least $1 - e^{-cN^{1/5}}$. Let ψ denote the probability density of $R(x,\tau) \in [-1,1]$, where τ is sampled from the Haar measure on S_N. Then it is known that

$$\psi(q) = \frac{1}{Z_\psi}(1-q^2)^{(N-3)/2}$$

where $Z_\psi = \Theta(N^{-1/2})$. Define the codimension-1 band

$$\mathrm{Band}(q) = \mathrm{Band}(q; x) := \{\tau \in S_N : R(x,\tau) = q\}$$

and let

$$Z^{x,\mathrm{I}}(q) = \int_{\mathrm{Band}(q)} e^{H_N^{x,\mathrm{I}}(\tau)} \, \mathrm{d}\rho_q(\tau),$$

where ρ_q is the Haar measure on $\mathrm{Band}(q)$, normalized so that $\rho_q(\mathrm{Band}(q)) = \psi(q)$. Then the left-hand side of (42) is equal to

$$\int_{N^{-2/5} \leqslant |q| \leqslant 1} Z^{x,\mathrm{I}}(q) \, \mathrm{d}q.$$

An application of Guerra's interpolation as in [HS23, Lemma 3.3] shows that for any $q \in [-1,1]$ and constant $\eta > 0$, with probability $1 - e^{-cN}$,

$$\frac{1}{N} \log Z^{x,\mathrm{I}}(q) \leqslant \frac{1}{2}\left(\xi(1) + \xi(|q|) + |q| + \log(1-|q|)\right) + \eta.$$

Since $\xi_{\sim 1}$ is ε-strictly replica symmetric and $\gamma_1^2 \leqslant N^{-4/5}$, this implies

$$\frac{1}{N} \log Z^{x,\mathrm{I}}(q) \leqslant \frac{\xi(1)}{2} - \frac{\varepsilon q^2}{4} + 2\eta.$$

Let $\delta > 0$ be small depending on ε, and η small depending on δ. This implies that for any $|q| \geqslant \delta$, with probability $1 - e^{-cN}$,

$$\frac{1}{N} \log Z^{x,\mathrm{I}}(q) \leqslant \frac{\xi(1)}{2} - \frac{\varepsilon\delta^2}{8}.$$

Taking a union bound over a N^{-1}-net of $|q| \geqslant \delta$ as in [HS23, Lemma 3.4] implies that with probability $1 - e^{-cN}$,

$$\int_{\delta \leqslant |q| \leqslant 1} Z^{x,\mathrm{I}}(q) \, \mathrm{d}q \leqslant e^{N\xi(1)/2 - cN}. \tag{43}$$

We address the remaining range of q by a first moment bound. Note that

$$\mathbf{E} \int_{N^{-2/5} \leqslant |q| \leqslant \delta} Z^{x,\mathrm{I}}(q) \, \mathrm{d}q = e^{N\xi(1)/2} \int_{N^{-2/5} \leqslant |q| \leqslant \delta} e^{N\xi(q)} \psi(q) \, \mathrm{d}q. \tag{44}$$

Recall from Fact 8.4 that $\xi''(0) \leqslant 1 - \varepsilon$. Thus, for sufficiently small δ, for all $|q| \leqslant \delta$,

$$\xi_{\sim 1}(q) + \frac{1}{2} \log(1 - q^2) \leqslant -\varepsilon q^2/4.$$

Thus, for all $N^{-2/5} \leqslant |q| \leqslant \delta$,

$$\frac{1}{N} \log \left(e^{N\xi(q)} \psi(q) \right) = \xi(q) + \frac{1}{N} \log \psi(q) = \gamma_1^2 q + \xi_{\sim 1}(q) + \frac{1}{2} \log(1 - q^2) + O(N^{-1} \log N)$$
$$\leqslant \gamma_1^2 q - \varepsilon q^2/4 + O(N^{-1} \log N) \leqslant -\varepsilon N^{-4/5}/8.$$

Combining with (44) shows

$$\mathbf{E} \int_{N^{-2/5} \leqslant |q| \leqslant \delta} Z^{x,\mathrm{I}}(q) \, \mathrm{d}q \leqslant e^{N\xi(1)/2 - cN^{1/5}},$$

so by Markov's inequality, with probability $1 - e^{-cN^{1/5}/2}$,

$$\int_{N^{-2/5} \leqslant |q| \leqslant \delta} Z^{x,\mathrm{I}}(q) \, \mathrm{d}q \leqslant e^{N\xi(1)/2 - cN^{1/5}/2}.$$

Combining with (43) proves (42) after adjusting c. □

Proof of Lemma 8.11, Eq. (19). By the same argument leading to (42), it suffices to prove

$$\int_{S_N} \mathbf{1}[|R(x, \tau)| \geqslant N^{-2/5}] e^{H_N^{x,\mathrm{II}}(\tau)} \, \mathrm{d}\rho(\tau) \leqslant e^{N\xi(1)/2 - cN^{1/5}} \tag{45}$$

holds with probability at least $1 - e^{-cN^{1/5}}$, where now

$$H_N^{x,\mathrm{II}}(\sigma) = N\gamma_2^2 R(x, \sigma)^2 + \widetilde{H}_N(\sigma),$$

i.e. we have replaced the spike in $H_N^{x,\mathrm{I}}$ with only its degree-2 part. Then $H_N^{x,\mathrm{II}}(\sigma) \leqslant H_N^{x,\mathrm{I}}(\sigma)$ almost surely for all σ such that $R(x, \sigma) \geqslant 0$, so (42) implies

$$\int_{S_N} \mathbf{1}[R(x, \tau) \geqslant N^{-2/5}] e^{H_N^{x,\mathrm{II}}(\tau)} \, \mathrm{d}\rho(\tau) \leqslant e^{N\xi(1)/2 - cN^{1/5}}$$

with probability $1 - e^{-cN}$. Moreover, by symmetry of the degree-2 spike,

$$\int_{S_N} \mathbf{1}[R(x, \tau) \geqslant N^{-2/5}] e^{H_N^{x,\mathrm{II}}(\tau)} \, \mathrm{d}\rho(\tau) \overset{d}{=} \int_{S_N} \mathbf{1}[R(x, \tau) \leqslant -N^{-2/5}] e^{H_N^{x,\mathrm{II}}(\tau)} \, \mathrm{d}\rho(\tau).$$

This implies (45) and thus (19). □

Proof of Lemma 8.11, Eq. (20). This follows from a slightly more complex form of the same strategy, where the planted Hamiltonian now has two spikes. For $x^1, x^2 \in S_N$, let

$$H_N^{x^1,x^2,\mathrm{III}}(\sigma) = N\xi(R(x^1,\sigma)) + N\xi(R(x^2,\sigma)) + \tilde{H}_N(\sigma).$$

By the same gaussian change of measure argument as above, the expectation in the left-hand side of (20) equals

$$e^{N\xi(1)} \int \mathbf{1}[|R(\sigma^1,\sigma^2)| \leqslant 3N^{-2/5}] \mathbb{P}(\sigma^1 \notin T(H_N^{\sigma^1,\sigma^2,\mathrm{III}})) \, d\rho^{\otimes 2}(\sigma^1,\sigma^2).$$

Thus it suffices to show that for all $x^1, x^2 \in S_N$ with $|R(x^1,x^2)| \leqslant 3N^{-2/5}$,

$$\int_{S_N} \mathbf{1}[|R(x^1,\tau)| \geqslant N^{-2/5}] e^{H_N^{x^1,x^2,\mathrm{III}}(\tau)} \, d\rho(\tau) \leqslant e^{N\xi(1)/2 - cN^{1/5}} \tag{46}$$

with probability at least $1 - e^{-cN^{1/5}}$. Let $\lambda = R(x^1,x^2) \in [-3N^{-2/5}, 3N^{2/5}]$ and

$$x^2 = \lambda x^1 + \sqrt{1-\lambda^2} x_\perp^2,$$

where $x_\perp^2 \in S_N$ and $R(x^1, x_\perp^2) = 0$. Let ψ_2 denote the probability density of $(R(x^1,\tau), R(x_\perp^2,\tau)) \in [1,1]^2$, where τ is sampled from the Haar measure on S_N. It is known that

$$\psi_2(q) = \frac{\mathbf{1}[q_1^2 + q_2^2 \leqslant 1]}{Z_{\psi_2}} (1 - q_1^2 - q_2^2)^{(N-4)/2}$$

where $Z_\psi = \Theta(N^{-1})$. Define the codimension-2 band

$$\mathsf{Band}(q_1,q_2) = \mathsf{Band}(q_1,q_2; x^1, x^2) := \{\tau \in S_N : R(x^1,\tau) = q_1, R(x_\perp^2,\tau) = q_2\}.$$

and let

$$Z^{x^1,x^2,\mathrm{III}}(q_1,q_2) = \int_{\mathsf{Band}(q_1,q_2)} e^{H_N^{x^1,x^2,\mathrm{III}}(\tau)} \, d\rho_{q_1,q_2}(\tau),$$

where ρ_{q_1,q_2} is the Haar measure on $\mathsf{Band}(q_1,q_2)$, normalized so that $\rho_{q_1,q_2}(\mathsf{Band}(q_1,q_2)) = \psi_2(q_1,q_2)$. Then the left-hand side of (46) is equal to

$$\int \mathbf{1}[|q_1| \geqslant N^{-2/5}] Z^{x^1,x^2,\mathrm{III}}(q_1,q_2) \, d(q_1,q_2). \tag{47}$$

Note that

$$\frac{1}{N} \log Z^{x^1,x^2,\mathrm{III}}(q_1,q_2) = \xi(q_1) + \xi\left(\lambda q_1 + \sqrt{1-\lambda^2} q_2\right) + \frac{1}{N} \log \int_{\mathsf{Band}(q_1,q_2)} e^{\tilde{H}_N(\tau)} \, d\rho_{q_1,q_2}(\tau).$$

Let $\tilde{q} = \tilde{q}(q_1,q_2) := \sqrt{q_1^2 + q_2^2}$. Applying Guerra's interpolation as in [HS23, Lemma 3.3] shows that for any $\eta > 0$, with probability $1 - e^{-cN}$

$$\frac{1}{N} \log \int_{\mathsf{Band}(q_1,q_2)} e^{\tilde{H}_N(\tau)} \, d\rho_{q_1,q_2}(\tau) \leqslant \frac{1}{2} \left(\xi(1) - \xi(\tilde{q}) + \tilde{q} + \log(1 - \tilde{q})\right) + \eta.$$

and thus

$$\frac{1}{N} \log Z^{x^1,x^2,\mathrm{III}}(q_1,q_2) \leqslant \xi(q_1) + \xi\left(\lambda q_1 + \sqrt{1-\lambda^2} q_2\right) + \frac{1}{2} \left(\xi(1) - \xi(\tilde{q}) + \tilde{q} + \log(1 - \tilde{q})\right) + \eta$$

$$\leqslant \xi_{\sim 1}(|q_1|) + \xi_{\sim 1}(|q_2|) + \frac{1}{2}\left(\xi(1) - \xi_{\sim 1}(\tilde{q}) + \tilde{q} + \log(1 - \tilde{q})\right) + 2\eta.$$

Since $\xi_{\sim 1}$ only includes terms that are degree 2 or larger, $\xi_{\sim 1}(|q_1|) + \xi_{\sim 1}(|q_2|) \leqslant \xi_{\sim 1}(\tilde{q})$. Thus the last display is bounded by

$$\frac{1}{2}\left(\xi(1) + \xi_{\sim 1}(\tilde{q}) + \tilde{q} + \log(1 - \tilde{q})\right) + 2\eta \leqslant \frac{\xi(1)}{2} - \frac{\varepsilon \tilde{q}^2}{4} + 2\eta.$$

Arguing as in the proof of equation (18) then shows that for any $\delta > 0$ depending only on ε,

$$\int \mathbf{1}[\tilde{q}(q_1, q_2) > \delta] Z^{x^1, x^2, \text{III}}(q_1, q_2)\, \mathrm{d}(q_1, q_2) \leqslant e^{N\xi(1)/2 - cN} \tag{48}$$

with probability $1 - e^{-cN}$. The remaining part of the integral (47) has expectation

$$\mathbf{E} \int \mathbf{1}[|q_1| \geqslant N^{-2/5}, \tilde{q}(q_1, q_2) \leqslant \delta] Z^{x^1, x^2, \text{III}}(q_1, q_2)\, \mathrm{d}(q_1, q_2)$$
$$= e^{N\xi(1)/2} \int \mathbf{1}[|q_1| \geqslant N^{-2/5}, \tilde{q}(q_1, q_2) \leqslant \delta] e^{N\xi(q_1) + N\xi(\lambda q_1 + \sqrt{1 - \lambda^2} q_2)} \psi_2(q_1, q_2)\, \mathrm{d}(q_1, q_2). \tag{49}$$

Recall $\gamma_1^2 \leqslant N^{-4/5}$. For all (q_1, q_2) in this indicator,

$$\frac{1}{N} \log\left(e^{N\xi(q_1) + N\xi(\lambda q_1 + \sqrt{1 - \lambda^2} q_2)} \psi_2(q_1, q_2)\right)$$
$$= \xi(q_1) + \xi\left(\lambda q_1 + \sqrt{1 - \lambda^2} q_2\right) + \frac{1}{2} \log(1 - \tilde{q}^2) + O(N^{-1} \log N)$$
$$\leqslant 2N^{-4/5} \tilde{q} - \frac{1}{2}(1 - \xi''(0)) \tilde{q}^2 + O(\tilde{q}^3 + N^{-1} \log N)$$
$$\leqslant -\frac{\varepsilon \tilde{q}^2}{2} + 2N^{-4/5} \tilde{q} + O(\tilde{q}^3 + N^{-1} \log N).$$

Since $N^{-2/5} \leqslant \tilde{q} \leqslant \delta$, for δ sufficiently small depending on ε this is bounded by $-cN^{-4/5}$. Combining with (49) shows that with probability $1 - e^{-cN^{1/5}}$,

$$\int \mathbf{1}[|q_1| \geqslant N^{-2/5}, \tilde{q}(q_1, q_2) \leqslant \delta] Z^{x^1, x^2, \text{III}}(q_1, q_2)\, \mathrm{d}(q_1, q_2) \leqslant e^{N\xi(1)/2 - cN^{1/5}}.$$

Further combining with (48) completes the proof. □

Proof of Lemma 8.11, Eq. (21). This is proved identically to equation (20), except with spiked Hamiltonian

$$H_N^{x^1, x^2, \text{IV}}(\sigma) = N\gamma_2^2 R(x^1, \sigma)^2 + N\xi(R(x^2, \sigma)) + \tilde{H}_N(\sigma),$$

i.e. the spike involving x^1 is replaced with just its degree-2 component. The same argument applies and we omit details. □

We turn to the proofs of Lemmas 8.15 and 8.16. The following fact will be useful in the proofs of both lemmas.

Fact B.4. *If $\sigma \in S_N$ satisfies $|\sigma_i|, |\sigma_j| \leqslant \log N$ and $\sigma \notin \tilde{T}_{i,j}$, then $\sigma \notin \tilde{T}$.*

Proof. Consider any $\tau \in S_N$ satisfying $|R(\sigma_{\sim i,j}, \tau_{\sim i,j})| \geqslant 2N^{-2/5}$. Since $|\tau_i|, |\tau_j| \leqslant N^{1/2}$,

$$|R(\sigma, \tau)| \geqslant |R(\sigma_{\sim i,j}, \tau_{\sim i,j})| - \frac{|\sigma_i||\tau_i| + |\sigma_j||\tau_j|}{N} \geqslant N^{-2/5}.$$

Thus the expectation over τ in (22) is larger than the expectation over τ in (28), and so $\sigma \notin \widetilde{T}$. □

Proof of Lemma 8.15. We can write $X_{i,j} = \widetilde{X}_{i,j} + \widetilde{X}_{i,j}^{(1)} + \widetilde{X}_{i,j}^{(2)}$, where

$$\widetilde{X}_{i,j}^{(1)} = \left\langle \mathbf{1}[|\sigma_i|, |\sigma_j| \leqslant \log N, \sigma \notin \widetilde{T}_{i,j}] \sigma_i \sigma_j \left(e^{H_{N,\sim 2}(\sigma) - N\xi_{\sim 2}(1)/2} - \mathbf{1}[i = j] \right) \right\rangle_2,$$
$$\widetilde{X}_{i,j}^{(2)} = \left\langle \mathbf{1}[|\sigma_i| \vee |\sigma_j| > \log N] \sigma_i \sigma_j \left(e^{H_{N,\sim 2}(\sigma) - N\xi_{\sim 2}(1)/2} - \mathbf{1}[i = j] \right) \right\rangle_2.$$

By using $(a + b)^2 \leqslant 2a^2 + 2b^2$, we deduce that

$$X_{i,j}^2 \leqslant 2\widetilde{X}_{i,j}^2 + 2(\widetilde{X}_{i,j}^{(1)} + \widetilde{X}_{i,j}^{(2)})^2,$$

so the rest of the proof is dedicated to showing that $|\widetilde{X}_{i,j}^{(1)} + \widetilde{X}_{i,j}^{(2)}| \leqslant e^{-c \log^2 N}$ with probability $1 - e^{-c \log^2 N}$. To do so, we will simply apply Markov to control $|\widetilde{X}_{i,j}^{(1)}|$ and $|\widetilde{X}_{i,j}^{(2)}|$.

By Fact B.4 and using $|\sigma_i|, |\sigma_j| \leqslant \sqrt{N}$,

$$|\widetilde{X}_{i,j}^{(1)}| \leqslant N \left\langle \mathbf{1}[|\sigma_i|, |\sigma_j| \leqslant \log N, \sigma \notin \widetilde{T}_{i,j}] \left(e^{H_{N,\sim 2}(\sigma) - N\xi_{\sim 2}(1)/2} + \mathbf{1}[i = j] \right) \right\rangle_2$$
$$\leqslant N \left\langle \mathbf{1}[\sigma \notin \widetilde{T}] \left(e^{H_{N,\sim 2}(\sigma) - N\xi_{\sim 2}(1)/2} + \mathbf{1}[i = j] \right) \right\rangle_2.$$

By the first two equations from Corollary 8.12, $\mathbf{E}_{\sim 2}|\widetilde{X}_{i,j}^{(1)}| \leqslant e^{-cN^{1/5}}$. Furthermore,

$$\mathbf{E}_{\sim 2}|\widetilde{X}_{i,j}^{(2)}| \leqslant N\mathbf{E}_{\sim 2} \left\langle (\mathbf{1}[|\sigma_i| > \log N] + \mathbf{1}[|\sigma_j| > \log N]) \left(e^{H_{N,\sim 2}(\sigma) - N\xi_{\sim 2}(1)/2} + \mathbf{1}[i = j] \right) \right\rangle_2$$
$$\leqslant 2N \left\langle \mathbf{1}[|\sigma_i| > \log N] + \mathbf{1}[|\sigma_j| > \log N] \right\rangle_2 \leqslant e^{-c \log^2 N}$$

by Proposition 8.8(1). By Markov's inequality, with probability $1 - e^{-c \log^2 N}$, $|\widetilde{X}_{i,j}^{(1)}| + |\widetilde{X}_{i,j}^{(2)}| \leqslant e^{-c \log^2 N}$, after adjusting c as necessary. This concludes the proof. □

Proof of Lemma 8.16. We can write $\widetilde{X}_{i,j}^2 = \widehat{X}_{i,j} + \widehat{X}_{i,j}^{(1)} - \widehat{X}_{i,j}^{(2)}$, where

$$\widehat{X}_{i,j}^{(1)} = \Big\langle \mathbf{1}[|\sigma_i^1|, |\sigma_j^1|, |\sigma_i^2|, |\sigma_j^2| \leqslant \log N, \sigma^1, \sigma^2 \in \widetilde{T}_{i,j}, |R(\sigma_{\sim i,j}^1, \sigma_{\sim i,j}^2)| > 2N^{-2/5}]$$
$$\sigma_i^1 \sigma_j^1 \sigma_i^2 \sigma_j^2 \left(e^{H_{N,\sim 2}(\sigma^1) - N\xi_{\sim 2}(1)/2} - \mathbf{1}[i = j] \right) \left(e^{H_{N,\sim 2}(\sigma^2) - N\xi_{\sim 2}(1)/2} - \mathbf{1}[i = j] \right) \Big\rangle_2,$$
$$\widehat{X}_{i,j}^{(2)} = \Big\langle \mathbf{1}[|\sigma_i^1|, |\sigma_j^1|, |\sigma_i^2|, |\sigma_j^2| \leqslant \log N, (\sigma^1 \notin \widetilde{T}_{i,j} \vee \sigma^2 \notin \widetilde{T}_{i,j}), |R(\sigma_{\sim i,j}^1, \sigma_{\sim i,j}^2)| \leqslant 2N^{-2/5}]$$
$$\sigma_i^1 \sigma_j^1 \sigma_i^2 \sigma_j^2 \left(e^{H_{N,\sim 2}(\sigma^1) - N\xi_{\sim 2}(1)/2} - \mathbf{1}[i = j] \right) \left(e^{H_{N,\sim 2}(\sigma^2) - N\xi_{\sim 2}(1)/2} - \mathbf{1}[i = j] \right) \Big\rangle_2.$$

By the same argument as before, it suffices to control $\mathbf{E}_{\sim 2}|\widehat{X}_{i,j}^{(1)}|$ and $\mathbf{E}_{\sim 2}|\widehat{X}_{i,j}^{(2)}|$. Note that almost surely,

$$|\widehat{X}_{i,j}^{(1)}| \leqslant N^2 \Big\langle \mathbf{1}[|\sigma_i^1|, |\sigma_j^1|, |\sigma_i^2|, |\sigma_j^2| \leqslant \log N, \sigma^1, \sigma^2 \in \widetilde{T}_{i,j}, |R(\sigma_{\sim i,j}^1, \sigma_{\sim i,j}^2)| > 2N^{-2/5}]$$

$$\left(e^{H_{N,\sim 2}(\sigma^1) - N\xi_{\sim 2}(1)/2} + 1\right) \left(e^{H_{N,\sim 2}(\sigma^2) - N\xi_{\sim 2}(1)/2} + 1\right) \Big\rangle_2 \leqslant N^2(\widehat{X}_{i,j}^{(3)} + \widehat{X}_{i,j}^{(4)})$$

where

$$\widehat{X}_{i,j}^{(3)} = \Big\langle \mathbf{1}[\sigma^1 \in \widetilde{T}_{i,j}, |R(\sigma_{\sim i,j}^1, \sigma_{\sim i,j}^2)| > 2N^{-2/5}]e^{H_{N,\sim 2}(\sigma^1) + H_{N,\sim 2}(\sigma^2) - N\xi_{\sim 2}(1)} \Big\rangle_2,$$

$$\widehat{X}_{i,j}^{(4)} = \Big\langle \mathbf{1}[|\sigma_i^1|, |\sigma_j^1|, |\sigma_i^2|, |\sigma_j^2| \leqslant \log N, |R(\sigma_{\sim i,j}^1, \sigma_{\sim i,j}^2)| > 2N^{-2/5}]$$

$$\left(e^{H_{N,\sim 2}(\sigma^1) - N\xi_{\sim 2}(1)/2} + e^{H_{N,\sim 2}(\sigma^2) - N\xi_{\sim 2}(1)/2} + 1\right) \Big\rangle_2.$$

By definition of $\widetilde{T}_{i,j}$,

$$\widehat{X}_{i,j}^{(3)} \leqslant \Big\langle \mathbf{1}[\sigma^1 \in \widetilde{T}_{i,j}]e^{H_{N,\sim 2}(\sigma^1) - N\xi_{\sim 2}(1)/2 - cN^{1/5}} \Big\rangle_2,$$

and thus $\mathbf{E}_{\sim 2}\widehat{X}_{i,j}^{(3)} \leqslant e^{-cN^{1/5}}$. Furthermore,

$$\widehat{X}_{i,j}^{(4)} \leqslant \Big\langle \mathbf{1}[|R(\sigma^1, \sigma^2)| > N^{-2/5}](e^{H_{N,\sim 2}(\sigma^1) - N\xi_{\sim 2}(1)/2} + e^{H_{N,\sim 2}(\sigma^2) - N\xi_{\sim 2}(1)/2} + 1) \Big\rangle_2,$$

and thus

$$\mathbf{E}_{\sim 2}\widehat{X}_{i,j}^{(4)} \leqslant 3 \Big\langle \mathbf{1}[|R(\sigma^1, \sigma^2)| > N^{-2/5}] \Big\rangle_2 \leqslant e^{-cN^{1/5}}$$

by Proposition 8.8(1). Combining shows $\mathbf{E}_{\sim 2}|\widehat{X}_{i,j}^{(1)}| \leqslant e^{-cN^{1/5}}$. Similarly

$$|\widehat{X}_{i,j}^{(2)}| \leqslant N^2 \Big\langle \mathbf{1}[|\sigma_i^1|, |\sigma_j^1|, |\sigma_i^2|, |\sigma_j^2| \leqslant \log N, (\sigma^1 \notin \widetilde{T}_{i,j} \vee \sigma^2 \notin \widetilde{T}_{i,j}),$$

$$|R(\sigma_{\sim i,j}^1, \sigma_{\sim i,j}^2)| \leqslant 2N^{-2/5}] \left(e^{H_{N,\sim 2}(\sigma^1) - N\xi_{\sim 2}(1)/2} + 1\right) \left(e^{H_{N,\sim 2}(\sigma^2) - N\xi_{\sim 2}(1)/2} + 1\right) \Big\rangle_2$$

By Fact B.4, on the indicator in this expectation, $\sigma^1, \sigma^2 \in \widetilde{T}$. Moreover

$$|R(\sigma^1, \sigma^2)| \leqslant |R(\sigma_{\sim i,j}^1, \sigma_{\sim i,j}^2)| + \frac{|\sigma_i^1||\sigma_i^2| + |\sigma_j^1||\sigma_j^1|}{N} \leqslant 3N^{-2/5}.$$

Thus

$$|\widehat{X}_{i,j}^{(2)}| \leqslant 2N^2 \Big\langle \mathbf{1}[\sigma^1 \notin \widetilde{T}, |R(\sigma^1, \sigma^2)| \leqslant 3N^{-2/5}]$$

$$\left(e^{H_{N,\sim 2}(\sigma^1) - N\xi_{\sim 2}(1)/2} + 1\right) \left(e^{H_{N,\sim 2}(\sigma^2) - N\xi_{\sim 2}(1)/2} + 1\right) \Big\rangle_2$$

and Corollary 8.12 implies $\mathbf{E}_{\sim 2}|\widehat{X}_{i,j}^{(2)}| \leqslant e^{-cN^{1/5}}$. □