

Sampling from the Sherrington–Kirkpatrick Model up to Criticality

Hang Du^{*} Brice Huang[†] Sidhanth Mohanty[‡] Amit Rajaraman[§]

October 13, 2026

Abstract

We prove that the simulated annealing algorithm efficiently samples from the Sherrington–Kirkpatrick model at inverse temperature $\beta < 1$ to an arbitrary inverse polynomial accuracy in total variation.

^{*}MIT. hangdu@mit.edu.

[†]Stanford University. bmhuang@stanford.edu.

[‡]Northwestern University. sidhanth.mohanty@northwestern.edu.

[§]MIT. amit.r@mit.edu.

Contents

1	Introduction	1
1.1	Technical overview	2
1.2	Related work	4
1.3	Future directions	8
1.4	AI methodology	9
2	Preliminaries	10
2.1	Notation	10
2.2	Simulated annealing	10
2.3	Stochastic localization	11
3	Warmup: High-probability covariance bounds for the zero-field SK model	13
4	High-probability covariance bounds at nonzero localization time	15
4.1	Proof of Lemma 4.4	17
4.2	Overlap concentration and crude bounds on the cumulants	19
5	Lower bounds for the prediction of the inverse covariance	21
5.1	The case when t is small	21
5.2	The case when t is large	22
5.3	The proof of Lemma 5.5	23
	References	26
A	Tying up loose ends	33
A.1	Controlling error terms	34
A.2	Proof of Lemma 4.6	35
A.3	Proof of Lemma A.1	40
A.4	Proof of Lemma A.2	43

1 Introduction

We study the problem of sampling from the Sherrington–Kirkpatrick (SK) model. This is an Ising model $\mu_{\beta W}$ with the random interaction matrix βW , for $W \sim \text{GOE}(N)$, that is, $W_{ij} = W_{ji} \sim \mathcal{N}(0, \frac{1}{N})$. More explicitly, the density of this distribution is given by

$$\mu_{\beta W}(\sigma) \propto \exp\left(\frac{\beta}{2} \langle \sigma, W \sigma \rangle\right). \quad (1)$$

There is a vast literature dedicated to the study of Ising models, due to their connections to physics, optimization, statistical inference, among others; see [WJ08, MM09, FV17] and references therein. The SK model in particular has provided a rich testbed for techniques, and has been the subject of much work for several decades. Some landmarks in the study of the SK model are the Parisi formula predicting its free energy [Par79, Par80b, Par80a], foundational works towards rigorous study [GT02, Gue03, ASS03], Talagrand’s proof of the Parisi formula [Tal06], and Panchenko’s resolution of the ultrametricity conjecture [Pan13]; we refer the reader to [Tal10, Tal11, Pan12] for more details.

Of special interest has been the task of *sampling* from the SK model in the high-temperature regime $0 \leq \beta < 1$. It is believed that the canonical Glauber dynamics Markov chain mixes rapidly at this temperature [SZ81, SZ82, MPV87b].

Proving this has been a major open problem in the past decade, as outlined in [Sim20] and [BDL⁺26, Problem 15]. We defer a detailed discussion of prior results to Section 1.2.

While rapid mixing of the fixed-temperature Glauber dynamics remains open, in this paper, we study *simulated annealing*, a simple modification of the standard Markov chain Monte Carlo paradigm that is frequently used in practice across a variety of areas, including VLSI design, combinatorial optimization, and scheduling among others; see [KGV83, SS85, LA87, WLL88, JAMS89, YRN94, KAJ94, HJJ03].

The simulated annealing algorithm is a time-dependent Markov chain where each step is from the Glauber dynamics Markov chain, but the inverse temperature starts at 0 and is slowly incremented to β over the course of running the Markov chain. The intuitive picture a reader should keep in mind is that while a Markov chain to sample from a target distribution $\mu_{\beta W}$ may not mix from a generic initialization, it would nevertheless sample from $\mu_{\beta W}$ when initialized at $\mu_{(\beta-\delta)W}$.

Our main result is the following.

Theorem 1.1. *Let $0 \leq \beta < 1$ and let $a > 0$ be a (large) constant. With probability at least $1 - N^{-a}$ over $W \sim \text{GOE}(N)$, the output distribution $\hat{\mu}$ of the simulated annealing algorithm run for $O_{a,\beta}(N \log N)$ steps of the Glauber dynamics per iteration, with inverse temperature gap c/N , run on the SK model $\mu_{\beta W}$, is such that $d_{\text{TV}}(\hat{\mu}, \mu_{\beta W}) \leq N^{-a}$.*

This is the first algorithm sampling from the SK model for all $\beta < 1$ up to an arbitrary inverse polynomial accuracy—the previous best results [Cel24, LSS26] only attain a total variation distance that is slightly $o(1)$. Another contribution is that our algorithm of choice is based on a Markov chain, which has consequences for the spiked Wigner model. In this statistical inference task, for some $x \sim \{\pm 1\}^N$, we are tasked with inferring x from $W + \frac{\lambda}{N} x x^\top$ for $W \sim \text{GOE}(N)$. In particular, using the techniques of [LMR⁺24, RW26], our result makes progress towards establishing that

Markov chains succeed at non-trivial recovery for all $\lambda > 1$, matching the so-called BBP phase transition [BBP05] for when non-trivial recovery is possible. Completing this proof would require a version of the above statement not just for the zero-field model, but even for non-zero external fields aligned with the all-ones direction.

We prove the above by establishing a *weak Poincaré inequality* [Aid98, RW01, Mat06, HMRW24] for the SK model. A distribution π on $\{\pm 1\}^N$ is said to satisfy a (ρ, ε) -weak Poincaré inequality if for all functions f ,

$$\mathcal{E}(f, f) \geq \frac{1}{\rho} (\text{Var}_\pi[f] - \varepsilon \|f\|_\infty^2), \quad (2)$$

where $\mathcal{E}(f, f)$ is the standard Dirichlet form. When $\varepsilon = 0$, this devolves into the usual Poincaré inequality, equivalent to a spectral gap for the Glauber dynamics. For nonzero ε , this yields rapid mixing (down to total variation error $\approx \sqrt{\varepsilon}$) from a warm start. [HMRW24] establishes that if all the $\{\mu_{\alpha W}\}_{0 \leq \alpha \leq \beta}$ satisfy (ρ, ε) weak Poincaré inequalities with ε a sufficiently small subconstant quantity, then this yields guarantees for simulated annealing. We prove Theorem 1.1 by establishing a high-probability weak Poincaré inequality for each $\mu_{\beta W}$, and taking a union bound over the inverse temperatures encountered along the trajectory of simulated annealing.

Theorem 1.2. *Let $0 < \beta < 1$ and $a > 0$ be constants. With probability at least $1 - N^{-a}$ over $W \sim \text{GOE}(N)$, the SK model $\mu_{\beta W}$ satisfies an $(O(N), N^{-a})$ -weak Poincaré inequality.*

To establish this weak Poincaré inequality, we use the *localization schemes* framework of Chen–Eldan [CE22]. We elaborate more on the details of this framework and how we use it in Section 1.1; using techniques from prior works, this reduces the proof of Theorem 1.1 to establishing high-probability covariance bounds for a suitably modified SK model with a random external field—this is the primary technical contribution of our work. A special case of our analysis is a high-probability covariance bound for the zero-field SK model.

Theorem 1.3. *Let $0 < \beta < 1$ and $a > 0$. Then, for some bounded constant $L = L_{\beta, a}$, with probability at least $1 - N^{-a}$ over $W \sim \text{GOE}(N)$, the SK model (1) is such that $\text{Cov}(\mu_{\beta W}) \preceq L \cdot \text{Id}$.*

1.1 Technical overview

In this section, we provide an overview of how we reduce Theorem 1.2 to establishing appropriate covariance bounds using localization schemes. Localization schemes are a principled approach to designing measure decompositions, providing a measure-valued martingale $(\mu_t)_{0 \leq t \leq T}$ with μ_0 being the Gibbs measure. For this decomposition to be helpful in proving a weak Poincaré inequality for the Gibbs measure, we will desire the following two properties:

1. First, the (random) endpoint measure μ_T is “nice” in some sense. In our case, this will mean proving that μ_T satisfies a Poincaré inequality with high probability, say at least $1 - \varepsilon$.
2. Second, we may establish some pathwise control that will allow us to lift this niceness back up, to proving a weak Poincaré inequality for the base measure μ_0 . In our case, this will mean proving that with high probability, say at least $1 - \varepsilon$, over the localization path, the covariances of all the μ_t are spectrally bounded by some constant (independent of N).

Given this, we can execute the following string of inequalities, for $\mathcal{E}$ being the event that the endpoint measure satisfies a Poincaré inequality, so $\Pr[\mathcal{E}] \geq 1 - \varepsilon$:

$$\begin{aligned} \mathcal{E}_\mu(f, f) &\geq \mathbb{E}\mathcal{E}_{\mu_T}(f, f) \\ &\gtrsim \mathbb{E}\text{Var}_{\mu_T}[f] \mathbf{1}_{\mathcal{E}} \\ &\geq \mathbb{E}\text{Var}_{\mu_T}[f] - 4\varepsilon\|f\|_\infty^2 \\ &\gtrsim \text{Var}_\mu[f] - C\varepsilon\|f\|_\infty^2. \end{aligned}$$

In the above $\gtrsim$ hides $\text{poly}(N)$ factors. The non-trivial inequalities here are the second and fourth: the second uses the high-probability Poincaré inequality for the endpoint measure, and the fourth uses the pathwise control. The latter uses Itô's formula to compute the time-derivative of the expected variance, and controls this with the aforementioned "pathwise control" of the (μ_t) measures; we refer the interested reader to [CE22, Section 4] and [HMRW24, Section 6] for more details.

Our localization scheme of choice will be *stochastic localization* [Eld13]. For a measure $\mu_0 = \mu$ on $\{\pm 1\}^N$, stochastic localization is defined by the following measure-valued martingale:

$$d\mu_t(x) = \langle x - \text{mean}(\mu_t), dB_t \rangle \cdot \mu_t(x). \quad (3)$$

[EM22] provides an alternate perspective on this, which will be far more helpful for us. Draw $\sigma \sim \mu$ and a Brownian motion $(B_t)_{t \geq 0}$. Then, the process $(\mu_t)_{t \geq 0}$ is equivalently given by the measure, which at time t , is the posterior on σ conditioned on $z_t = t\sigma + B_t$. At time 0, z_t is uninformative so this is simply the Gibbs measure. On the other hand, at infinite time, $\frac{1}{t}z_t$ is simply equal to σ , and the posterior is then just the Dirac delta on this (random) point x which is marginally distributed according to μ . This posterior measure has an easy explicit formula:

$$\mu_t(y) \propto \exp\left(\frac{\beta}{2}y^\top W y + \langle t\sigma + B_t, y \rangle\right). \quad (4)$$

We now wish to show that $\text{Cov}(\mu_t)$ is bounded with high probability, and that μ_T satisfies a Poincaré inequality with high probability. However, as stated, this is rather intractable: the random variable σ depends on the disorder W in some complicated way, which makes the analysis of μ_t very difficult.

To get around this issue, we shall shift to the so-called *planted* model, first used by Achlioptas-Coja-Oghlan [AC08] in random CSPs. This technique is now standard in the study of stochastic localization on the SK model [EMS22, EMS23, HMP24, HMRW24]. The planted model π is defined by the following random Hamiltonian: draw $x \sim \text{Unif}\{\pm 1\}^N$, and $W \sim \text{GOE}(N)$, and set

$$\pi(y) \propto \exp\left(\frac{\beta}{2}\left(y^\top W y + \frac{\beta}{N}\langle x, y \rangle^2\right)\right).$$

The planted model is helpful for two reasons. The first, perhaps unsurprisingly, is that we may argue about it instead of the original SK model; they are mutually contiguous, in that high-probability events for either Hamiltonian transfer to high-probability events for the other, in a manner made precise in Lemma 2.5. For example, we may use this statement with the event being "the model satisfies a weak Poincaré inequality."

The second, which is more interesting, is that the trajectory of stochastic localization is significantly more tractable in this model. In fact, the posterior on x conditioned on $M := W + \frac{\beta}{N}xx^\top$ is

precisely the distribution π . This allows us to swap out the sample σ used in (4) for the hidden “plant” $\mathbf{x}$! That is, we are interested in models of the form

$$\pi_t(y) \propto \exp \left(\frac{\beta}{2} y^\top \left(\mathbf{W} + \beta \frac{\mathbf{x}\mathbf{x}^\top}{N} \right) y + \langle t\mathbf{x} + \mathbf{g}\sqrt{t}, y \rangle \right),$$

for $\mathbf{g} \sim \mathcal{N}(0, \text{Id})$. We shall prove a high-probability covariance bound for each fixed t (not depending on the disorder), then use a union bound and a simple perturbation argument to translate this to a high-probability covariance bound for all t . The problem then reduces to establishing high-probability covariance bounds for these models.

Theorem 1.4. *Let $0 < \beta < 1$ and let $a > 0$ be a (large) constant. Fix $t > 0$ bounded, possibly decaying with N . For $\mathbf{g} \sim \mathcal{N}(0, \text{Id})$ and $\mathbf{x} \sim \text{Unif}\{\pm 1\}^N$, consider the probability measure μ_t on $\{\pm 1\}^N$ with*

$$\mu_t(y) \propto \exp \left(\frac{\beta}{2} y^\top \left(\mathbf{W} + \beta \frac{\mathbf{x}\mathbf{x}^\top}{N} \right) y + \langle t\mathbf{x} + \mathbf{g}\sqrt{t}, y \rangle \right).$$

Then, there is some constant $L = L_{\beta,a}$ such that with probability $1 - O(N^{-a})$, $\|\mathbf{Cov}(\mu_t)\| \leq L$.

The technical bulk of this paper will go towards proving the above, and we provide an overview of some ideas involved in [Section 3](#).

At a high level, we prove the above as follows. We consider a prediction $\mathbf{K}^{-1}$ for the covariance matrix based on the Thouless–Anderson–Palmer (TAP) approximation to the free energy, the “inverse TAP Hessian”. This matrix has a term linear in the disorder, and a handful of terms involving the mean of the measure. In an entry of the $\mathbf{K} \cdot \mathbf{Cov}$ that involves an element $\mathbf{W}_{ij}$ of the disorder and some cumulant of the Gibbs measure, we may then decouple the cumulant from the disorder using the cavity method, at the cost of introducing a small (random) external field. We can now Taylor expand any given cumulant in terms of this small random external field; this introduces terms involving zero-field SK models (on $N - 1$ spins) and Gaussians independent of the disorder defining these models. Gaussian concentration allows us to control the error in these approximations, using only the crude bounds that are yielded by overlap concentration, which are later refined using other means. Finally, showing that this matrix $\mathbf{K}$ is bounded away from 0 requires other machinery, which we may establish using techniques from previous work [\[Cel24\]](#).

We further believe that many of our techniques to prove covariance bounds should extend to the entire replica-symmetric regime [\[AT78\]](#) given sufficiently good overlap concentration, which might follow from recent results [\[Lop26, KN26\]](#). We have made no attempt to do this extension.

We must also show that at sufficiently large T , π_T satisfies a Poincaré inequality with high probability. This near-immediately follows from the techniques of the previous work of Bandeira–El Alaoui–Rödder [\[BER26\]](#).

Lemma 1.5 ([\[BER26\]](#)). *Let μ_T be the measure obtained by running stochastic localization on $\mu_{\beta\mathbf{W} + \frac{\beta^2}{N}\mathbf{x}\mathbf{x}^\top}$ for bounded time $T > 0$. Then, for a sufficiently large (constant) T , with probability $1 - e^{-cN}$, μ_T has Poincaré constant $O(N)$.*

1.2 Related work

Sampling from spin glasses. It was shown in a series of works [\[BB19, EKZ22, CE22\]](#) that the Glauber dynamics satisfies a Poincaré inequality for $\beta < 1/4$, which was improved on by Anari–Koehler–Vuong [\[AKV24\]](#) to show that it succeeds up to $\beta \approx 0.295$; both of these proofs use

stochastic localization. More recently, the “Bochner method” was used by Wang [Wan26] and in a follow-up work by Boban–Li–Oveis Gharan [BLO26] to establish fast mixing up to $\beta = \frac{1}{2} + \varepsilon$ (for some small constant ε).

An independent line of work pioneered by El Alaoui, Montanari, and Sellke [EMS22] studies efficient sampling via *algorithmic stochastic localization*, based on simulating the stochastic differential equation obtained from the time-reversal of (3). It was first shown in [EMS22, Cel24] that this algorithm achieves (normalized) Wasserstein distance $o(1)$ to the Gibbs measure. The work of Huang, Montanari, and Pham [HMP24] illustrates how these ideas can be used to design algorithms achieving a total variation distance guarantee of $o(1)$ using this technique in the related spherical p -spin model. Recently, works of Davies, Lee, Sandhu, and Shi [DLSS26a, LSS26] extended these ideas to design an algorithm that achieves total variation distance $o(1)$ to the Gibbs measure for the SK model.

The same set of authors has an accompanying line of work [DLSS26b], which builds on these ideas to establish a $(O(N), o(1))$ -weak Poincaré inequality for the Gibbs measure when $\beta < 1/2$. However, their quantitative weakness parameter is not small enough to establish success of simulated annealing, and they must use the Langevin diffusion on another distribution to generate a sufficiently warm start.

Beyond the SK model, the “2-spin” model, the problem of sampling from the more general class of p -spin models has been studied. These are defined by coefficients $(\beta_p)_{0 \leq p \leq p_*}$ for some constant p_* , and have density

$$\mu(y) \propto \exp \left(\sum_{p \geq 0} \frac{\beta_p}{N^{(p-1)/2}} \langle G^{(p)}, y^{\otimes p} \rangle \right),$$

for $G_{i_1, \dots, i_p}^{(p)} \sim \mathcal{N}(0, 1)$ iid. It is predicted that these models exhibit a richer phase diagram, with a true Poincaré inequality for β up to a uniqueness threshold β_{uniq} [BBM96, AJ24, AJK⁺24], that Glauber dynamics initialized uniformly at random succeed up to a shattering threshold β_{sh} [CHS93], and that Markov chain-based methods fail at sampling past β_{sh} [CHS93, AJ24, EMS25, EI 25].

The spherical setting, where the state space is $\sqrt{N} \cdot \mathbb{S}^{N-1}$ instead of $\{\pm 1\}^N$, is generally more tractable. The corresponding 2-spin model does not suffer any phase transition in sampling, with efficient sampling possible at any β [GLLR21], but higher spin models are not fully understood. Of note are [HMP24], which uses algorithmic stochastic localization, and [HMRW24], which analyzes simulated annealing using weak Poincaré inequalities, to sample up to a threshold β_{SL} that, for pure p -spin models, is within a universal constant of the predicted sharp “shattering” threshold β_{sh} . While algorithmic stochastic localization provably cannot do better, it seems plausible that simulated annealing reaches β_{sh} .

Covariance bounds. Proving covariance bounds for the SK Gibbs measure, as in Theorem 1.4, has also been studied, mostly towards the goal of sampling. Talagrand [Tal11] showed that when $\beta < 1$, $\mathbf{E} \|\text{Cov}(\mu_{\beta W})\|_{\text{op}} = O(\log^8 N)$, and explicitly posed as an open question whether $\mathbf{E} \|\text{Cov}(\mu_{\beta W})\|_{\text{op}} = O(1)$ when $\beta < 1$. A more precise question one may ask is whether the “inverse TAP Hessian” $K^{-1} = ((1 + \beta^2)\text{Id} - \beta W)^{-1}$ is an accurate predictor of the covariance of the zero-field SK model. It is believed that with high probability, $\|K \text{Cov} - \text{Id}\|_{\text{op}} = O\left(\frac{1}{\sqrt{N}}\right)$. In fact, a suitable modification of this belief is expected to hold not just for the high-temperature zero-field model, but for replica-symmetric models with a sufficiently large random external field, below the

so-called de Almeida–Thouless line [AT78].

All the results mentioned in the previous paragraph on sampling from the SK model with Markov chains, which establish Poincaré inequalities, also immediately yield (high-probability) covariance bounds, but the best such result only holds for $\beta < \frac{1}{2} + \varepsilon$. El Alaoui–Gaitonde [EG24], building on estimates in [ABSY21], successfully resolved Talagrand’s conjecture and showed that $\mathbb{E} \|\mathbf{K}\mathbf{Cov} - \text{Id}\|_F^2 = O(1)$ for all $\beta < 1$. [BSXY24, BXY23] were able to establish high-probability covariance bounds in a setting where there is an external field h —their regime of (β, h) is a large subset of the predicted replica-symmetric regime [Tal10, Tal11].

Our techniques. We now discuss the works that inspired our techniques. The local-to-global principle reducing a weak Poincaré inequality to high-probability covariance bounds was inspired by the localization schemes based approach to proving Poincaré inequalities for Markov chain [ALOV24, EKZ22, CE22, LMRW24, AKV24]. The use of these methods to establish weak Poincaré inequalities appears in the works of [EEGP23, HMRW24, DLSS26b].

The specific localization scheme of choice—stochastic localization—was introduced by Eldan [Eld13], and subsequently developed extensively in work on the KLS conjecture [LV24, Che21, KL22, SZ26, BKL26].

Our proof of the covariance bounds relies on proving that the covariance is close to a prediction for $\mathbf{K}^{-1}$ that arises from the TAP equations for the free energy of the SK model, developed by Thouless–Anderson–Palmer [TAP77]. To prove that $\mathbf{K}^{-1}$ is bounded, we need to prove that $\mathbf{K}$ is bounded away from 0, for which we use an iterative procedure based on approximate message passing (AMP), a technique now used widely in statistical inference [DMM09, BM11, DAM16, MV21], and in studying spin glasses [Bol14, DS19, Mon19, CT21, EMS21, ES22, HS24].

The proof that the covariance is close to $\mathbf{K}^{-1}$ relies on the classical *cavity method*—comparing a Gibbs distribution of a Hamiltonian to a version where one site is removed—introduced in statistical physics literature by Mézard–Parisi–Virasoro [MPV87a], developed by Aizenman–Sims–Starr [ASS03], and extensively used thereafter by Talagrand [Tal10, Tal11] among others.

Concurrent work by OpenAI. The problem of sampling from the SK model was one among 722 claimed results released very recently by OpenAI. In addition to establishing a spectral gap for $\beta < 1$, they prove cutoff from a worst-case initialization, and establish a mixing time of $n^{5/3+o(1)}$ at $\beta = 1$. We provide a short summary of our interpretation of their proof of a spectral gap [Ope26] below.

They begin essentially by proving a weak Poincaré inequality, albeit with a much better weakness parameter than ours. Using the standard stochastic localization that we do, they show that the covariance is bounded on the localization path with probability $1 - e^{-cN}$, and that the endpoint measure satisfies a Poincaré inequality with probability $1 - e^{-cN}$. We remark that their proof bounding the covariance significantly differs from ours. They show that under a certain “stability” criterion, the distribution has bounded covariance.

The aforementioned stability criterion requires local convexity at *every* approximate fixed point of the TAP free energy—this differs from previous results [Cel24], where it is shown that a late AMP iterate approximates the mean well, and that one has local convexity at this AMP iterate. In particular, their Lemma 5.7 uses this to conclude that there is a unique TAP fixed point.

We remark that such conclusions are expected to be *false* in the p -spin model (in the regime with metastable states); it seems likely that our proof is more flexible in this regard, and indeed a

similar strategy succeeds at establishing a weak Poincaré inequality for p -spin models up to the β_{SL} threshold from [HMP24, Eq. (1.5)].

Given this weak Poincaré inequality, the key technical innovation is that they are then able to boost this to a true Poincaré inequality. To do this, they express the Poincaré inequality in a more fine-grained fashion. For the stopping time τ as in [HMRW24, Lemma 6.5] (which is equal to T with probability $1 - e^{-cN}$), denote by $\mathcal{E}$ the event that $\tau \neq T$ or that μ_T does not satisfy a Poincaré inequality—this is an event on the space of trajectories of stochastic localization. One then has

$$\mathcal{E}(f, f) \geq \rho \cdot (\text{Var}[f] - \mathbf{E}[\mathbf{1}_{\mathcal{E}} \cdot \text{Var}_{\mu_\tau}[f]]) .$$

Their goal now is to show that $\mathbf{E}[\mathbf{1}_{\mathcal{E}} \cdot \text{Var}_{\mu_\tau}[f]]$ is small relative to $\mathcal{E}(f, f)$. While their proof as stated does not express it in this way, it may be interpreted as showing that for any function f that has mean 0 and variance 1 (under μ_0),

$$\mathbf{E}[\mathbf{1}_{\mathcal{E}} \cdot \langle f^2 \rangle_{\mu_\tau}] \leq o(1) + O(1) \cdot \mathcal{E}(f, f) . \quad (5)$$

Given this, one can substitute this bound and rearrange in order to prove a true Poincaré inequality. In contrast, in our work and in [HMRW24], we use the throwaway bound $\mathbf{E}[\mathbf{1}_{\mathcal{E}} \cdot \text{Var}_{\mu_\tau}[f]] \leq \text{osc}(f)^2 \cdot \Pr[\mathcal{E}]$.

To prove this, they consider the auxiliary stochastic localization process defined as follows. Denote by ν_0 the distribution with density $\mu_0 \cdot f^2$. For some draw X and $Y_t = tX + B_t$, let $\mathbb{P}$ be the law of the posterior on X under the prior $X \sim \mu_0$, defining a measure-valued process $(\mu_t)_{t \geq 0}$. Analogously, let $\mathbb{Q}$ be the law of the posteriors under the prior $X \sim \nu_0$. On a common observation path $\mathcal{F}_t = \sigma((Y_s)_{s \leq t})$, Bayes' rule yields that the law of ν_t is simply that of $\frac{\mu_t \cdot f^2}{\langle f^2 \rangle_{\mu_t}}$, and so for any time t ,

$$\frac{d\mathbb{Q}_t}{d\mathbb{P}_t} = \langle f^2 \rangle_{\mu_t} . \quad (6)$$

Consequently,

$$\mathbf{E}[\mathbf{1}_{\mathcal{E}} \langle f^2 \rangle_{\mu_\tau}] = \mathbb{Q}_\tau(\mathcal{E}) .$$

The goal now is to port the established bound $\mathbb{P}_\tau(\mathcal{E}) \leq e^{-cN}$ over to a useful bound on $\mathbb{Q}_\tau(\mathcal{E})$. To do this, they show that the KL divergence

$$\text{KL}(\mathbb{Q}_\tau \| \mathbb{P}_\tau) = \text{Ent}[\langle f^2 \rangle_{\mu_\tau}] = \text{Ent}_{\mu_0}[f^2] - \mathbf{E}[\text{Ent}_{\mu_\tau}[f^2]]$$

is $o(N) + O(N) \cdot \mathcal{E}(f, f)$. This divergence can be expressed using Itô's formula in terms of the trajectory of stochastic localization, which yields

$$\text{KL}(\mathbb{Q}_\tau \| \mathbb{P}_\tau) = \frac{1}{2} \mathbf{E}_{X \sim \nu_0, B} \int_0^\tau \|\text{mean}(\nu_t) - \text{mean}(\mu_t)\|^2 ,$$

where the measures μ_t and ν_t are coupled by using the same observation path, as in (6). The remainder of the proof is in two steps. First, they replace $\text{mean}(\mu_t)$ with the TAP approximation, which is more tractable. The second step, which is more involved, requires going through the proof that the TAP fixed point approximates the mean and tracking how reweighing by f^2 influences the proof. In particular, they “Taylor expand” f out in terms of its discrete derivatives to order 1. These discrete derivative terms may then be bounded in terms of the Dirichlet form (and the variance),

yielding the second term in (5); the first term arises from errors in the TAP approximation. We refer the reader to [Ope26, Proposition 7.3] for more details.

Organization. We begin in Section 2 with an overview of the techniques we will require. Section 3 provides a self-contained proof of the high-probability zero-field covariance bound Theorem 1.3, whose approach we build on in the remainder of the paper. Section 4 establishes the high-probability covariance bound Theorem 1.4 at nonzero localization time, conditioned on the inverse prediction K being positive and bounded away from 0. Section 5 shows that the inverse prediction K is indeed bounded away from 0 with high probability. In Appendix A, we prove a handful of lemmas omitted from the main body.

1.3 Future directions

We propose a few directions for further study.

Mixing in SK model with a field. An interesting extension of the results in this paper would be to establish mixing in the SK model under a field up to the critical threshold.

Problem 1.6 (Mixing under the de Almeida–Thouless line). In the presence of a field $h \cdot \mathbf{1}$ for $h \in \mathbb{R}$, for $W \sim \text{GOE}(N)$, prove that simulated annealing samples from the Ising model $\mu_{\beta W, h, \mathbf{1}}$ in polynomial time as long as $0 \leq \beta < \beta_*(h)$, where $\beta_*(h)$ is described by the de Almeida–Thouless line from [AT78].

A qualitative version of the above is due to [BER26]. Beyond settling a conjectured spin glass threshold, the above question is of interest in the connection between sampling from the SK model, and the efficacy of Markov chains at optimal inference in the spiked Wigner model via a connection made by [LMR⁺24, RW26].

Counting vs. sampling in average-case models. A conventional wisdom in the classical sampling literature is that the algorithmic task of computing a partition function has roughly the same complexity as sampling from a Gibbs distribution, and there is a formal equivalence [JVV86, SJ89] in that an algorithm for computing the partition function for a family of Hamiltonians gives a sampling algorithm for the corresponding Gibbs distributions.

In the average-case setting, the situation is murkier. As a running example, consider the spherical p -spin glass model, where the Hamiltonian is a degree- p random polynomial whose coefficients are drawn independently from a Gaussian distribution. It is believed that Markov chain-based methods should fail to sample from the Gibbs distribution when the inverse temperature exceeds the *shattering threshold* β_{sh} , and the best sampling algorithms work up to a constant of the threshold [HMP24, HMRW24]. On the other hand, an algorithm for computing the partition function was given well beyond β_{sh} up to the “replica symmetric threshold” in a work of [BHL⁺26], which suggests that spin glass models may exhibit phases where sampling is hard, but computing the partition function is tractable. However, this story was complicated by the recent work [ZSSJ26], which uses a quantum-inspired algorithm to sample in the shattered phase.

We consider it an interesting direction to understand what the true algorithmic threshold is for sampling, as well as what the threshold is for Markov chains.

Certifiable sampling. We say that a sampling algorithm for an Ising model is *certifiable* if there is a $\text{poly}(N)$ -time verifier that takes as input the interaction matrix of an Ising model, and outputs

either yes or abstain, and further outputs yes with high probability over the disorder. When the verifier outputs yes, the sampling algorithm must additionally (approximately) sample from the target distribution.

Many of the previous results for mixing in Ising models [EKZ22, CE22, AKV24, Wan26] that yielded samplers for the SK model when $\beta < 1/2$ in fact give certifiable samplers since the proofs of mixing rely on spectral conditions on the interaction matrix, or other closely related matrices. The proof of mixing in this paper relies significantly more on the randomness of the disorder compared to previous works, and hence does not yield an efficiently checkable condition.

Problem 1.7. What is the threshold β_* such that there is a certifiable sampler from the SK model at inverse temperature $\beta < \beta_*$?

We anticipate that certifiable samplers are a useful primitive for certain downstream applications of sampling to statistical inference and privacy, where it may be essential to be certain that one is indeed sampling from the desired target distribution. Beyond applications, if β_* is indeed equal to 1, we anticipate that it would yield significantly simpler proofs of the results in this paper. In a world where β_* is *not* equal to 1, an interesting takeaway is that the proof of success of any sampler for the SK model necessarily strongly exploits the distributional assumption of the input.

1.4 AI methodology

This paper is the result of a long-running project on sampling from the high-temperature SK model up to $\beta < 1$ using simulated annealing, starting early 2025. Our original approach was centered on the TAP Hessian $\mathbf{K}$, and tried to show that for constant p , one has the bound $\mathbf{E} \text{Tr} \left((Z \cdot (\mathbf{K} \text{Cov} - \text{Id}))^{2p} \right) \leq \frac{1}{N^{p-1}} \cdot (\mathbf{E} Z)^{2p}$, by manually expanding out this expectation and witnessing various cancellations. At nonzero localization time, this would further involve “TAP planting” in order to have a tractable approximation to the mean.

We established this in the zero-field case for $p = 1$, and for $p = 2$, used a computer program to bound various lower-order terms in this expansion (for context, we would require $p \geq 12$ in order to get meaningful guarantees for simulated annealing). We shortly discovered that this expectation is not bounded all the way up to $\beta = 1$, due to low-probability tail events. Nevertheless, we had a good sense of the kind of high-probability event that one might be able to condition on in order to get a useful high-probability covariance bound using this proof.

In mid-2026, we tried using a higher-order prediction for the covariance matrix using the Plefka expansion [Ple82, GY91]: this is a more precise prediction $\mathbf{K}^{(\ell)}$ for the inverse covariance featuring higher-order corrections, and we tried showing that $\mathbf{E} \left\| Z \cdot (\mathbf{K}^{(\ell)} \text{Cov} - \text{Id}) \right\|_{\text{F}}^2 \leq \frac{1}{N^{\ell-1}}$. This plan was successfully executed for some small ℓ , but it was not sufficient for the simulated annealing application.

At the launch of GPT-5.5, we attempted using it to establish the bounds described above. These attempts were not fruitful, and produced nothing beyond the earlier results that had been proved by hand. This process was repeated with the launch of GPT-5.6 Sol, which produced a 150-page document: this document featured a high-probability zero-field covariance bound using the Plefka expansion down to an arbitrarily small inverse polynomial error probability using the previously sketched approaches, but failed to prove similar results in the non-zero field case.

When GPT-6 Astra Pro launched, all the previous documents were fed in with a proof strategy, and it produced an 85-page document with a complete proof of a weak Poincaré inequality strong enough to yield guarantees for simulated annealing. However, this proof deviated from the prompt, and regressed from using the Plefka expansion to the earlier idea of $\text{Tr} \left((Z \cdot (\mathbf{K}\text{Cov} - \text{Id}))^{2p} \right)$ after conditioning on a high-probability event, in addition to using the so-called “ $1/N$ -expansion,” e.g. [CGTvH26]. Furthermore, at nonzero localization time, it continued to use the recommended approach of TAP planting to have a tractable approximation of the mean. Various parts of this proof were digested over the following week. We note that this proof did successfully establish that the covariance prediction is $O\left(\frac{1}{\sqrt{N}}\right)$ -close to the covariance, a bound that the present manuscript does *not* establish. In a nutshell, this proof considered two expansions of the inverse covariance, one in β and the other in $1/N$; showing the well-definedness of this requires establishing analyticity of certain functions. It then analyzed various terms in the β -expansion, then transferred these bounds over to useful statements for the coefficients of the $1/N$ -expansion.

After digestion, the authors provided suggestions to simplify the proof based on their understanding of the zero-field proof, attempting to take more advantage of the aforementioned analyticity. This produced a 33-page document containing a proof similar to that in the present manuscript (which, notably, does not use this analyticity). Over the subsequent weeks, the authors digested this proof and simplified numerous parts of it, resulting in the present document: these efforts utilized many incarnations of GPT-6 Astra, ranging in power from Medium to Ultra.

GPT-6 Astra was additionally used in typesetting, the detection of bugs, and finding references. Tautologically, the authors take all responsibility for this manuscript.

2 Preliminaries

2.1 Notation

- We use C to denote appropriately large universal constants whose value may change from line to line, and c to denote similarly fickle small constants.
- Given a symmetric matrix M , we denote by $\|M\|_{\text{op}}$ or $\|M\|$ its largest singular value, and by $\|M\|_{\text{F}}$ its Frobenius norm.
- Given a Gibbs measure μ over $\{\pm 1\}^N$, we denote by $\langle f(\sigma) \rangle_{\mu}$ the Gibbs average of f with respect to the measure μ . When the measure is clear from context, we denote it simply as $\langle f(\sigma) \rangle$ or $\langle f \rangle$.
- Given tensors $M \in (\mathbb{R}^N)^{\otimes s}$ and $K \in (\mathbb{R}^N)^{\otimes t}$ for some integer $t \leq s$, we denote by $M[K] \in (\mathbb{R}^N)^{\otimes (s-t)}$ the tensor

$$(M[K])_{i_1, \dots, i_{s-t}} = \sum_{i_{s-t+1}, \dots, i_s} M_{i_1, \dots, i_s} \cdot K_{i_{s-t+1}, \dots, i_s}.$$

2.2 Simulated annealing

In this section, let $H : \{\pm 1\}^N \rightarrow \mathbb{R}$ be a Hamiltonian, and let $\mu_{\beta H}(x) \propto \exp(\beta H(x))$ be the corresponding Gibbs distribution.

We begin by defining Glauber dynamics.

Definition 2.1 (Glauber dynamics). We describe the *Glauber dynamics* Markov chain for sampling from $\mu_{\beta H}$ via a single step of the Markov chain.

Let x be the current state. Pick $i \sim [N]$ uniformly at random, and transition $x' \sim \mu_{\beta H}|_{x_{-i}}$ where x_{-i} refers to x with the i -th coordinate erased.

In the simulated annealing algorithm, one runs Glauber dynamics but changes β according to a schedule.

Definition 2.2 (Simulated annealing). Let $(\beta_i)_{i \in [T]} > 0$ be a sequence of inverse temperatures and $x \in \{\pm 1\}^N$ be an initialization. The simulated annealing algorithm with schedule $(\beta_i)_{i \in [T]}$ runs T steps of Glauber dynamics, where the t -th step is taken according to Glauber dynamics for $\mu_{\beta_t H}$.

We rely on the fact that a weak Poincaré inequality for $\mu_{\beta H}$ implies that simulated annealing samples; see, e.g., [HMRW24, Theorem 4.13].

Lemma 2.3 (Simulated annealing from weak Poincaré inequalities). Let $0 = \beta_0 < \beta_1 < \dots < \beta_L = \beta$. Suppose that each $\mu_{\beta_j H}$ satisfies a (ρ, ε) -weak Poincaré inequality. Suppose for each $j \in [L]$ and some sufficiently small constant $c > 0$:

$$\beta_j - \beta_{j-1} \leq \frac{c}{H_{\max} - H_{\min}},$$

Starting from a uniformly random initialization, run m Glauber updates at each inverse temperature β_j . Then the output distribution ν satisfies

$$d_{\text{TV}}(\nu, \mu_{\beta H}) \leq O\left(L\left(e^{-\frac{m}{2\rho}} + \sqrt{\varepsilon}\right)\right).$$

2.3 Stochastic localization

To establish a weak Poincaré inequality, we use stochastic localization (see, e.g., (3)) to (approximately) decompose the Gibbs distribution into measures that each satisfy a Poincaré inequality. In particular, we use the following consequence of [HMRW24, Lemmas 6.8 and A.6].

Lemma 2.4 (Localization criterion). Let $(\mu_t)_{0 \leq t \leq T}$ be stochastic localization initialized at a fixed distribution μ on $\{\pm 1\}^N$. Suppose that, with probability at least $1 - \delta$, the following two events hold:

- $\sup_{0 \leq t \leq T} \|\text{Cov}(\mu_t)\|_{\text{op}} \leq K$.
- μ_T satisfies a Poincaré inequality with constant ρ_T .

Then μ satisfies a $(\rho_T e^{KT}, C\delta e^{KT})$ -weak Poincaré inequality.

Thus, it remains to control the covariance along the localization path. To obtain a handle on the distributions encountered along a localization path, we will shift to the so-called “planted” model in order to make the localization process more tractable.

Lemma 2.5. Let $0 \leq \beta < 1$. Let μ_{null} be the law over measures defined by a Hamiltonian $\mathbf{H}_{\text{null}}$ obtained by drawing $\mathbf{W} \sim \text{GOE}(N)$ and setting $H_{\text{null}}(\sigma) = \frac{\beta}{2} \cdot \sigma^\top \mathbf{W} \sigma$. Let μ_{pl} be the law over measures defined by a Hamiltonian $\mathbf{H}_{\text{pl}}$ obtained by drawing $\mathbf{W} \sim \text{GOE}(N)$ and $\mathbf{x} \sim \{\pm 1\}^N$, then setting $\mathbf{H}_{\text{pl}}(\sigma) = \frac{\beta}{2} \cdot \sigma^\top \mathbf{W} \sigma + \frac{\beta^2}{2N} \cdot \langle \sigma, \mathbf{x} \rangle^2$. Then, μ_{null} and μ_{pl} are mutually contiguous. More quantitatively, for any event $\mathcal{E}$ over the space of measures on $\{\pm 1\}^N$,

$$c\mu_{\text{pl}}(\mathcal{E})^2 \leq \mu_{\text{null}}(\mathcal{E}) \leq C\sqrt{\mu_{\text{pl}}(\mathcal{E})}.$$

Proof. Let us focus on the second inequality. We have

$$\mu_{\text{null}}(\mathcal{E}) = \mathbf{E}_{\text{pl}} \left[\mathbf{1}_{\mathcal{E}} \cdot \frac{d\mu_{\text{null}}}{d\mu_{\text{pl}}} \right] \leq \sqrt{\mu_{\text{pl}}(\mathcal{E})} \cdot \sqrt{\mathbf{E}_{\text{pl}} \left[\left(\frac{d\mu_{\text{null}}}{d\mu_{\text{pl}}} \right)^2 \right]} = \sqrt{\mu_{\text{pl}}(\mathcal{E})} \cdot \sqrt{\mathbf{E}_{\text{null}} \left[\frac{d\mu_{\text{null}}}{d\mu_{\text{pl}}} \right]}.$$

It remains to show that second quantity is $O(1)$. Note now that for any prescribed interaction matrix W (without the scaling by β),

$$\begin{aligned} \frac{d\mu_{\text{pl}}}{d\mu_{\text{null}}}(W \mid \mathbf{x} = x) &= \frac{\exp \left(-\frac{N}{4} \left\| W - \frac{\beta}{N} x x^\top \right\|_F^2 \right)}{\exp \left(-\frac{N}{4} \|W\|_F^2 \right)} \propto \exp \left(\frac{\beta}{2} x^\top W x \right), \text{ so} \\ \frac{d\mu_{\text{pl}}}{d\mu_{\text{null}}}(W) &\propto \sum_{x \in \{\pm 1\}^N} \exp \left(\frac{\beta}{2} x^\top W x \right) = Z_{\beta W}. \end{aligned}$$

Our goal is then simply to show that

$$\mathbf{E} \left[\frac{1}{Z_{\beta W}} \right] \lesssim \frac{1}{\mathbf{E}[Z_{\beta W}]}. \quad \square$$

This estimate is easily obtained by integrating the lower tail described in [Tal11, Theorem 11.2.1]. The first inequality is similar and requires us to show that $\mathbf{E}[Z_{\beta W}^2] \lesssim \mathbf{E}[Z_{\beta W}]^2$, which is far easier to establish. $\square$

By the characterization of stochastic localization given by El Alaoui and Montanari [EM22], along with contiguity Lemma 2.5, and the fact that $(\mathbf{x}, \mathbf{H}_{\text{pl}})$ has the same law as sampling $\mathbf{H}_{\text{pl}}$ and then sampling $\mathbf{x}$ from $\mu_{\beta M}$ for $M := \mathbf{W} + \beta \frac{\mathbf{x}\mathbf{x}^\top}{N}$, it suffices to study the covariance matrix of the Ising model $\mu_t := \mu_{\beta M, t\mathbf{x} + \mathbf{B}_t}$.

Theorem 1.4. Let $0 < \beta < 1$ and let $a > 0$ be a (large) constant. Fix $t > 0$ bounded, possibly decaying with N . For $\mathbf{g} \sim \mathcal{N}(0, \text{Id})$ and $\mathbf{x} \sim \text{Unif}\{\pm 1\}^N$, consider the probability measure μ_t on $\{\pm 1\}^N$ with

$$\mu_t(y) \propto \exp \left(\frac{\beta}{2} y^\top \left(\mathbf{W} + \beta \frac{\mathbf{x}\mathbf{x}^\top}{N} \right) y + \langle t\mathbf{x} + \mathbf{g}\sqrt{t}, y \rangle \right).$$

Then, there is some constant $L = L_{\beta, a}$ such that with probability $1 - O(N^{-a})$, $\|\mathbf{Cov}(\mu_t)\| \leq L$.

We now describe how to complete the proof of the main theorem.

Proof of Theorem 1.1. We need the following stronger version of Theorem 1.4 and prove that for any constant $a > 0$:

$$\Pr_{W,x} \left[\Pr_{(\mu_t)_{t \in [0,T]} \sim \text{SL}} \left[\sup_{0 \leq t \leq T} \|\mathbf{Cov}(\mu_t)\|_{\text{op}} \leq K \right] \geq 1 - O(N^{-a}) \right] \geq 1 - O(N^{-a}). \quad (7)$$

The proof of this closely follows that of [HMRW24, Lemma 6.7], which is based on using Theorem 1.4 to obtain a high probability bound for a given time, and then taking a union bound over a net of times.

With (7) in hand, we can use Lemma 2.4 to obtain a weak Poincaré inequality for each temperature, take a union bound over temperatures, and apply Lemma 2.3 to complete the proof. $\square$

3 Warmup: High-probability covariance bounds for the zero-field SK model

In this section, we sketch some ideas in the proof of Theorem 1.4 in the setting where $h = 0$.

Theorem 1.3. *Let $0 < \beta < 1$ and $a > 0$. Then, for some bounded constant $L = L_{\beta,a}$, with probability at least $1 - N^{-a}$ over $\mathbf{W} \sim \text{GOE}(N)$, the SK model (1) is such that $\mathbf{Cov}(\mu_{\beta\mathbf{W}}) \preceq L \cdot \text{Id}$.*

Since the diagonal entries of $\mathbf{W}$ do not affect the covariance, we may set these entries to 0 throughout this section. Define

$$K = (1 + \beta^2)\text{Id} - \beta\mathbf{W}.$$

Note that K is PSD and bounded away from 0 with high probability, as $\lambda_{\max}(\mathbf{W})$ converges in probability to 2.

We start by using the “cavity formula” to express the covariance. Given an external field $h \in \mathbb{R}^n$, denote by $m(h)$ the mean of the SK model with interaction matrix $\mathbf{W}$ and external field h . Similarly, given $i \in [N]$, denote by $m^{(i)}(h)$ the mean of the SK model (on $N - 1$ spins) with the i th spin removed, with interaction matrix $\mathbf{W}_{-i,-i}$ and external field h . Similarly denote by $\mathbf{Cov}^{(i)}(h) \in \mathbb{R}^{n \times n}$ the covariance matrix of the cavity model, with the i th row and column equal to 0. In particular, we denote $\mathbf{Cov}^{(i)} = \mathbf{Cov}^{(i)}(\mathbf{0})$. For ease of notation, let us also define

$$d^i = m^{(i)}(\beta\mathbf{W}_i) \text{ and } D^i = \mathbf{Cov}^{(i)}(\beta\mathbf{W}_i).$$

These are respectively the mean and covariance in the cavity model where the i th spin is pinned to +1. The law of total covariance then yields

$$\mathbf{Cov} = \begin{pmatrix} 1 & (d^i)^\top \\ d^i & D^i + d^i(d^i)^\top \end{pmatrix}.$$

We may expand out the matrix $K\mathbf{Cov} - \text{Id}$ using the above cavity expansion—see Section 4 for more details, where the analogue of this decomposition is described for nonzero localization time.

Lemma 3.1. *For each row i , we may write*

$$(K\mathbf{Cov} - \text{Id})_i = \alpha^i \mathbf{Cov}_i + Y_i - Z_i,$$

where $Y_{i,i} = Z_{i,i} = 0$, $\mathbf{Cov}^{(i)} = \mathbf{Cov}^{(i)}(0)$, and

$$\begin{aligned}\alpha^i &= \beta^2 - \beta \langle \mathbf{W}_i, d^i \rangle \\ Y_{i,-i} &= (d^i - \beta \mathbf{Cov}^{(i)} \mathbf{W}_i)^\top \\ Z_{i,-i} &= \beta \mathbf{W}_i^\top (D^i - \mathbf{Cov}^{(i)}) .\end{aligned}$$

Our next goal will be to show that $\alpha^i = o(1)$, $\|Y\|_F^2 = O(1)$, and $\|Z\|_F^2 = O(1)$ with high probability. This would complete the proof, since we then have

$$\cdot \mathbf{Cov} = (K - \text{diag}(\alpha))^{-1} (\text{Id} + Y - Z) \preceq O(1) .$$

We first control $\|Y\|_F^2$. The idea is to Taylor expand the function $m^{(i)}$ defining d^i , which yields an accurate approximation because the external field $\beta \mathbf{W}_i$ is small (with entries of order $N^{-1/2}$). In particular, since $\mathbf{Cov}^{(i)}(h) = \nabla_h m^{(i)}(h)$, the first order Taylor approximation around $h = 0$ is

$$d^i \approx \beta \mathbf{Cov}^{(i)} \mathbf{W}_i ,$$

which is exactly the other term in $Y_{i,-i}$. We can control the error in this approximation by Taylor expanding further; the error depends on the next nonzero term of the expansion. Before carrying out the expansion, we note that the terms in the Taylor expansion take the form of cumulants of the partition function. Let $Z^{(i)}(h)$ be the SK partition function at external field h , where spin i and its interactions are removed. Let $\kappa_r^{(i)}(h) = \nabla_h^r \log Z^{(i)}(h)$. It is standard that $\kappa_1^{(i)}(h) = m^{(i)}(h)$. If we write $\kappa_r^{(i)} = \kappa_r^{(i)}(0)$, then (because $\kappa_1^{(i)} = \kappa_3^{(i)} = 0$)

$$d^i \approx \sum_{r=1}^3 \frac{\beta^{r-1}}{(r-1)!} \kappa_r^{(i)} [\mathbf{W}_i^{\otimes(r-1)}] = \beta \mathbf{Cov}^{(i)} \mathbf{W}_i ,$$

with squared ℓ^2 error of order $\|\kappa_4^{(i)}(h) [\mathbf{W}_i^{\otimes 3}]\|_2^2$. We discuss the bounding of this error shortly—this will be the complicated part of the proof.

We can bound $\|Z\|_F^2$ near-identically, comparing $D^i = \mathbf{Cov}^{(i)}(\beta \mathbf{W}_i)$ to $\mathbf{Cov}^{(i)} = \mathbf{Cov}^{(i)}(0)$ by Taylor expansion of $\mathbf{Cov}^{(i)} = \kappa_2^{(i)}$. We similarly find that each row has norm $\|Z_i\|^2 = O(1/N)$ with very high probability, and so $\|Z\|_F^2 = O(1)$.

The fact that $\alpha^i = o(1)$ also follows by Taylor expanding d^i as above. We have

$$\langle \mathbf{W}_i, d^i \rangle \approx \beta \mathbf{Cov}^{(i)} [\mathbf{W}_i^{\otimes 2}] ,$$

with error of order $|\kappa_4^{(i)} [\mathbf{W}_i^{\otimes 4}]|$. We finally have that with high probability,

$$\mathbf{Cov}^{(i)} [\mathbf{W}_i^{\otimes 2}] = \frac{\text{Tr } \mathbf{Cov}^{(i)}}{N} + o(1) = 1 + o(1) ,$$

because, by overlap concentration, the conditional variance of this random variable is bounded by

$$\text{Var} \left[\mathbf{Cov}^{(i)} [\mathbf{W}_i^{\otimes 2}] \middle| \mathbf{W}_{-i,-i} \right] \leq \frac{2 \|\mathbf{Cov}^{(i)}\|_F^2}{N^2} \leq O\left(\frac{1}{N}\right)$$

with high probability, and its conditional expectation is exactly $\frac{1}{N} \text{Tr } \mathbf{Cov}^{(i)}$. Gaussian concentration can then be used to obtain an error probability bound of $e^{-c\sqrt{N}}$: we have for a sufficiently small constant ε that

$$\Pr \left[\left| \mathbf{Cov}^{(i)}[W_i^{\otimes 2}] - \frac{\text{Tr } \mathbf{Cov}^{(i)}}{N} \right| > \varepsilon \right] \leq 2 \exp \left(-c \min \left\{ \frac{\varepsilon^2 N^2}{\|M\|_F^2}, \frac{\varepsilon N}{\|M\|_{\text{op}}} \right\} \right) \leq e^{-c\sqrt{N}},$$

Now, how do we show that the error $\|\kappa_4[W_i^{\otimes 3}]\|_2^2$ to be small. We may establish this by controlling its high moments using Wick's formula, which various produces terms involving trace bounds and other cumulants. For example, one has

$$\mathbb{E}_{W_i} \|\kappa_4[W_i^{\otimes 3}]\|^2 = \frac{1}{N^3} \cdot \left(6\|\kappa_4\|_F^2 + 9\| \underbrace{\kappa_4[\text{Id}]}_{=-2\text{Cov}^2} \|_F^2 \right),$$

and overlap concentration allows us to bound both of these terms by $O(N^2)$ (in general $\|\kappa_r\|_F^2 \leq N^{r/2}$), and thus bound the expectation by $O(1/N)$. However, bounding constant moments in this way and using overlap concentration will only allow us to establish that for constant p , $\|\mathbf{Cov}\|_{\text{op}} = O_{L^p}(1)$: this only yields a high-probability bound on the norm of the covariance by $O(N^\delta)$.

Nevertheless, this already suffices to get better bounds on the second term in the expectation. Indeed, earlier we bounded $\|\mathbf{Cov}^2\|_F \leq \|\mathbf{Cov}\|_F^2 = O_{L^p}(N)$, but we may now bound $\|\mathbf{Cov}^2\|_F \leq \|\mathbf{Cov}\|_F \cdot \|\mathbf{Cov}\|_{\text{op}} = O_{L^p}(\sqrt{N})$, which yields a high-probability bound for the second term. Controlling the other term requires us to bound $\|\kappa_4\|_F^2$ by something better than $O(N^2)$; an improved bound of this form will also suffice to obtain the necessary better bounds on higher moments. This will require some more ideas, and we defer a detailed discussion to the actual proof, which is presented in [Section A.4](#).

4 High-probability covariance bounds at nonzero localization time

In this section, we establish [Theorem 1.4](#), restated for convenience.

Theorem 1.4. *Let $0 < \beta < 1$ and let $a > 0$ be a (large) constant. Fix $t > 0$ bounded, possibly decaying with N . For $g \sim \mathcal{N}(0, \text{Id})$ and $x \sim \text{Unif}\{\pm 1\}^N$, consider the probability measure μ_t on $\{\pm 1\}^N$ with*

$$\mu_t(y) \propto \exp \left(\frac{\beta}{2} y^\top \left(\mathbf{W} + \beta \frac{\mathbf{x}\mathbf{x}^\top}{N} \right) y + \langle t\mathbf{x} + g\sqrt{t}, y \rangle \right).$$

Then, there is some constant $L = L_{\beta,a}$ such that with probability $1 - O(N^{-a})$, $\|\mathbf{Cov}(\mu_t)\| \leq L$.

We will use the convention that the diagonal of $\mathbf{M}$ is zeroed out.

We now establish some notation. Set:

$$m = \langle \sigma \rangle, \quad v_i = 1 - m_i^2, \quad q_N = \frac{\|m\|^2}{N},$$

and define the TAP Hessian

$$K := \text{diag}(v_i^{-1}) - \beta \mathbf{M} + \beta^2(1 - q_N)\text{Id} - \frac{2\beta^2}{N} m m^\top. \quad (8)$$

We will bound $K\mathbf{Cov} - \text{Id}$ with high probability and combine this with a lower bound on the smallest eigenvalue of K .

Lemma 4.1 (Operator norm of the residual). *For any constant $a > 0$, with probability $1 - O(N^{-a})$, we have*

$$\|K\mathbf{Cov} - \text{Id}\|_{\text{op}} \leq O(1).$$

Lemma 4.2 (Smallest eigenvalue of the TAP Hessian). *There is a constant $\gamma > 0$, depending only on β and T , such that for any constant $a > 0$,*

$$\Pr[K \not\geq \gamma] \leq N^{-a}.$$

Corollary 4.3 (Covariance bound). *For some constant L , for any constant $a > 0$, we have*

$$\Pr[\|\mathbf{Cov}\|_{\text{op}} \geq L] \leq N^{-a}.$$

The proof of [Lemma 4.1](#) depends on the following estimates. Let $f_i := (K\mathbf{Cov} - \text{Id})_{i,-i}$, and let $\lambda_i := \beta^2(1 - q_N) - \beta d_i^\top \mathbf{M}_{i,-i}$ where $d_i := \frac{\mathbf{Cov}_{i,-i}}{v_i}$.

Lemma 4.4. *For every constant $a > 0$, there is a constant C_a such that*

$$\Pr\left[\sum_{i \in [N]} \|f_i - \lambda_i \mathbf{Cov}_{-i,i}\|_2^2 > C_a\right] \leq N^{-a}$$

We prove the following estimate for the diagonal part.

Lemma 4.5 (Diagonal terms). *We have $(K\mathbf{Cov} - \text{Id})_{ii} = \lambda_i v_i - \frac{2\beta^2 m_i}{N} (\mathbf{Cov} m)_i$. Further, for every $a > 0$ we have:*

$$\Pr\left[\max_{i \in [N]} |\lambda_i| > N^{-1/4}\right] \leq O(N^{-a}).$$

Hence,

$$\max_{i \in [N]} |(K\mathbf{Cov} - \text{Id})_{ii} - \lambda_i v_i| \leq 2\beta^2.$$

We defer the proofs of [Lemmas 4.4](#) and [4.5](#) to [Section 4.1](#), and prove [Lemma 4.1](#) below.

Proof of Lemma 4.1. We begin by conditioning on the high probability events from [Lemmas 4.2, 4.4](#) and [4.5](#), which simultaneously hold with probability $1 - O(N^{-a})$. Now define

$$D = \text{diag}(\lambda_i), \quad R = (K - D)\mathbf{Cov} - \text{Id}.$$

Then we have

$$\|R - \text{diag}(R)\|_F^2 = \sum_{i=1}^N \|f_i - \lambda_i \mathbf{Cov}_{-i,i}\|_2^2 \leq C.$$

By the identity in [Lemma 4.5](#) and the fact that $\mathbf{Cov}_{ii} = v_i$, we have for all i that

$$|R_{ii}| = |(K\mathbf{Cov} - \text{Id})_{ii} - (D\mathbf{Cov})_{ii}| \leq |(\lambda_i v_i + 2\beta^2) - (\lambda_i(1 - m_i^2))| = 2\beta^2,$$

and consequently,

$$\|R\|_{\text{op}} \leq \|R - \text{diag}(R)\|_F + \max_i |R_{ii}| \leq O(1).$$

Noting now that $K \geq \gamma$ for some constant γ , and $D \leq o(1)$, this implies that $\|\mathbf{Cov}\|_{\text{op}} = O(1)$, and thus [Lemma 4.1](#) and [Corollary 4.3](#). $\square$

4.1 Proof of Lemma 4.4

We begin by establishing some notational conventions.

- The gradient ∇ is always taken exclusively with respect to the external field. In particular, for a function F , we will use $\nabla F(g)$ to denote $\nabla_h F(h)|_{h=g}$.
- We will use the symbol Δ exclusively to denote remainder terms arising from Taylor expansions of various functions, with the subscript denoting the function that the term comes from.

Fix a site i , and consider the distribution $\mu^{(i)}$ obtained by removing its field and all its interactions with the other sites. Let $u_i \in \mathbb{R}^{N-1}$ be the field on the remaining sites, and let $F^{(i)}(u_i)$ be its free energy. Write

$$m^{(i)} = \nabla F^{(i)}(u_i), \quad \mathbf{Cov}^{(i)} = \nabla^2 F^{(i)}(u_i), \quad \kappa_r^{(i)} = \nabla^r F^{(i)}(u_i),$$

and

$$m^{(i,\pm)} = \nabla F^{(i)}(u_i \pm \beta \mathbf{M}_{i,-i}), \quad \mathbf{Cov}^{(i,\pm)} = \nabla^2 F^{(i)}(u_i \pm \beta \mathbf{M}_{i,-i}).$$

Denote

$$\overline{\mathbf{Cov}}_i := \frac{1+m_i}{2} \mathbf{Cov}^{(i,+)} + \frac{1-m_i}{2} \mathbf{Cov}^{(i,-)},$$

and observe that

$$d_i = \frac{\mathbf{Cov}_{i,-i}}{v_i} = \frac{m^{(i,+)} - m^{(i,-)}}{2}.$$

By the law of total covariance we obtain the block decomposition

$$\mathbf{Cov} = \begin{pmatrix} v_i & v_i d_i^\top \\ v_i d_i & \overline{\mathbf{Cov}}_i + v_i d_i d_i^\top \end{pmatrix}, \quad (9)$$

where the i th coordinate is written first. Denoting $\alpha = \beta^2(1 - q_N)$, observe that

$$\begin{aligned} f_i &= \underbrace{d_i}_{(\text{diag}(v^{-1})\mathbf{Cov})_{i,-i}} - \underbrace{\beta(\overline{\mathbf{Cov}}_i + v_i d_i d_i^\top) \mathbf{M}_{i,-i}}_{(\beta \mathbf{M} \mathbf{Cov})_{i,-i}} \\ &+ \underbrace{\alpha v_i d_i}_{(\alpha \mathbf{Cov})_{i,-i}} - \underbrace{\frac{2\beta^2 m_i}{N} (\mathbf{Cov} \cdot m)_{-i}}_{((2\beta^2/N) m m^\top \mathbf{Cov})_{i,-i}}. \end{aligned} \quad (10)$$

Rearranging the above using (9) and plugging in the definition of λ_i ,

$$f_i - \lambda_i \mathbf{Cov}_{i,-i} = d_i - \beta \overline{\mathbf{Cov}}_i \mathbf{M}_{i,-i} - \frac{2\beta^2 m_i}{N} (\mathbf{Cov} \cdot m)_{-i}$$

By the Taylor series expansion for $m^{(i,\pm)}$ and $\mathbf{Cov}^{(i,\pm)}$, we have:

$$m^{(i,\pm)} = \nabla F^{(i)}(u_i \pm \beta \mathbf{M}_{i,-i}) = m^{(i)} \pm \beta \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i} + \frac{\beta^2}{2} \kappa_3^{(i)} [M_{i,-i}^{\otimes 2}] \pm \frac{\beta^3}{6} \kappa_4^{(i)} [M_{i,-i}^{\otimes 3}] + \Delta_{m,i,\pm},$$

$$\mathbf{Cov}^{(i,\pm)} = \nabla^2 F^{(i)}(u_i \pm \beta \mathbf{M}_{i,-i}) = \mathbf{Cov}^{(i)} \pm \beta \kappa_3^{(i)} [\mathbf{M}_{i,-i}] + \frac{\beta^2}{2} \kappa_4^{(i)} [\mathbf{M}_{i,-i}^{\otimes 2}] + \Delta_{\mathbf{Cov},i,\pm},$$

we obtain the corresponding Taylor expansion of $\overline{\mathbf{Cov}}_i$ and d_i as

$$d_i = \frac{m^{(i,+)} - m^{(i,-)}}{2} = \beta \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i} + \frac{\beta^3}{6} \kappa_4^{(i)} [\mathbf{M}_{i,-i}^{\otimes 3}] + \Delta_{d,i},$$

$$\overline{\mathbf{Cov}}_i = \frac{1+m_i}{2} \mathbf{Cov}^{(i,+)} + \frac{1-m_i}{2} \mathbf{Cov}^{(i,-)} = \mathbf{Cov}^{(i)} + \beta m_i \kappa_3^{(i)} [\mathbf{M}_{i,-i}] + \frac{\beta^2}{2} \kappa_4^{(i)} [\mathbf{M}_{i,-i}^{\otimes 2}] + \Delta_{\mathbf{Cov},i}.$$

By the above,

$$\begin{aligned} d_i - \beta \overline{\mathbf{Cov}}_i \mathbf{M}_{i,-i} &= \left(\beta \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i} + \frac{\beta^3}{6} \kappa_4^{(i)} [\mathbf{M}_{i,-i}^{\otimes 3}] \right) \\ &\quad - \left(\beta \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i} + \beta^2 m_i \kappa_3^{(i)} [\mathbf{M}_{i,-i}^{\otimes 2}] + \frac{\beta^3}{2} \kappa_4^{(i)} [\mathbf{M}_{i,-i}^{\otimes 3}] \right) + \Delta_{f,i} \\ &= -\beta^2 m_i \kappa_3^{(i)} [\mathbf{M}_{i,-i}^{\otimes 2}] - \frac{\beta^3}{3} \kappa_4^{(i)} [\mathbf{M}_{i,-i}^{\otimes 3}] + \Delta_{f,i}, \end{aligned}$$

where $\Delta_{f,i} = \Delta_{d,i} - \beta \Delta_{\mathbf{Cov},i} \mathbf{M}_{i,-i}$. Therefore

$$f_i = \lambda_i \mathbf{Cov}_{-i,i} - \frac{\beta^3}{3} \kappa_4^{(i)} [\mathbf{M}_{i,-i}^{\otimes 3}] - \beta^2 m_i \kappa_3^{(i)} [\mathbf{M}_{i,-i}^{\otimes 2}] - \frac{2\beta^2 m_i}{N} (\mathbf{Cov} \cdot m)_{-i} + \Delta_{f,i}. \quad (11)$$

We may write:

$$\kappa_3^{(i)} [\mathbf{M}_{i,-i}^{\otimes 2}] = \frac{1}{N} \kappa_3^{(i)} [\text{Id}] + \kappa_3^{(i)} \left[\mathbf{M}_{i,-i}^{\otimes 2} - \frac{1}{N} \text{Id} \right].$$

We will show that the first term of the above nearly cancels the fourth term of (11), and that the second term is negligible.

Observe that

$$\frac{1}{N} \kappa_3^{(i)} [\text{Id}] = \frac{1}{N} \nabla \text{Tr} \mathbf{Cov}^{(i)} = \frac{1}{N} \nabla \left(N - 1 - \|m^{(i)}\|_2^2 \right) = -\frac{2}{N} \mathbf{Cov}^{(i)} \cdot m^{(i)}.$$

Plugging the above into (11) gives

$$\begin{aligned} f_i - \lambda_i \mathbf{Cov}_{-i,i} &= \frac{2\beta^2 m_i}{N} \left(\mathbf{Cov}^{(i)} \cdot m^{(i)} - (\mathbf{Cov} \cdot m)_{-i} \right) \\ &\quad - \beta^2 m_i \kappa_3^{(i)} \left[\mathbf{M}_{i,-i}^{\otimes 2} - \frac{1}{N} \text{Id} \right] - \frac{\beta^3}{3} \kappa_4^{(i)} [\mathbf{M}_{i,-i}^{\otimes 3}] + \Delta_{f,i}. \end{aligned}$$

The below estimates [Lemmas 4.6](#) and [4.7](#), whose proofs we defer to [Appendix A](#), complete the proof.

Lemma 4.6. *For every constant $a > 0$, the following bounds hold for all $i \in [N]$ with probability $1 - O_\varepsilon(N^{-a})$:*

$$\begin{aligned} \left\| \frac{2\beta^2 m_i}{N} \left(\mathbf{Cov}^{(i)} \cdot m^{(i)} - (\mathbf{Cov} \cdot m)_{-i} \right) \right\|_2 &\leq \frac{C}{\sqrt{N}} \\ \left\| \frac{\beta^3}{3} \kappa_4^{(i)} [\mathbf{M}_{i,-i}^{\otimes 3}] \right\|_2 &\leq \frac{C}{\sqrt{N}} \end{aligned}$$

$$\left\| \beta^2 m_i \kappa_3^{(i)} \left[\mathbf{M}_{i,-i}^{\otimes 2} - \frac{1}{N} \text{Id} \right] \right\|_2 \leq \frac{C}{\sqrt{N}}$$

$$\max_{i \in [N]} |\lambda_i| \leq N^{-1/4}.$$

Lemma 4.7. *For any $a, \varepsilon > 0$, with probability $1 - O(N^{-a})$, we have*

$$\|\Delta_{d,i}\|_2 \leq O(N^{-1+\varepsilon}), \quad \|\Delta_{\text{Cov},i} \mathbf{M}_{i,-i}\|_2 \leq O(N^{-3/4+\varepsilon}).$$

and consequently, $\|\Delta_{f,i}\|_2 \leq \|\Delta_{d,i}\|_2 + \beta \|\Delta_{\text{Cov},i} \mathbf{M}_{i,-i}\|_2 \leq O(N^{-3/4+\varepsilon})$.

4.2 Overlap concentration and crude bounds on the cumulants

In this section, we establish overlap concentration for the tilted model, and use this to establish some crude bounds on the Frobenius norms of the cumulants.

Lemma 4.8. *Let $-1 \leq s \leq 1$ fixed, and $t > 0$ bounded. Consider the Ising model μ with interaction matrix $\beta \mathbf{M}$, for $\mathbf{M} = \mathbf{W} + \frac{\beta}{N} \mathbf{x} \mathbf{x}^\top$ (with $\mathbf{x} \sim \{\pm 1\}^N$), and external field*

$$\mathbf{h} = \left(t + \frac{s\beta^2 \mathbf{s}}{N} \right) \cdot \mathbf{x} + \mathbf{g} \sqrt{t + \frac{s^2 \beta^2}{N}}, \quad (12)$$

with $\mathbf{s} \sim \{\pm 1\}$ and $\mathbf{g} \sim \mathcal{N}(0, \text{Id})$. Then, for some explicit deterministic constant q , we have

$$\mathbf{E} \langle |R_{12} - q|^{2k} \rangle \lesssim N^{-k}.$$

The same concentration holds for overlaps of a sample with the spike $\mathbf{x}$.

We bound the desired cumulants in the following corollary.

Lemma 4.9. *Let $-1 \leq s \leq 1$ fixed, and $t > 0$ bounded. Let a be a (large) constant. Consider the Ising model μ with interaction matrix $\beta \mathbf{M}$, for $\mathbf{M} = \mathbf{W} + \frac{\beta}{N} \mathbf{x} \mathbf{x}^\top$ (with $\mathbf{x} \sim \{\pm 1\}^N$), and external field*

$$\mathbf{h} = \left(t + \frac{s\beta^2 \mathbf{s}}{N} \right) \cdot \mathbf{x} + \mathbf{g} \sqrt{t + \frac{s^2 \beta^2}{N}},$$

with $\mathbf{s} \sim \{\pm 1\}$ and $\mathbf{g} \sim \mathcal{N}(0, \text{Id})$. Then, for any constants $r \geq 2, a > 0$,

$$\mathbf{E} \|\kappa_r(\mu)\|_F^{2a} \lesssim N^{ra/2}.$$

Proof. For $2r$ replicas $\sigma_1, \dots, \sigma_{2r}$ from the Gibbs measure, and denoting $R_{ab} = \frac{1}{N} \langle \sigma_a, \sigma_b \rangle$ for shorthand, because κ_r is a cumulant, we may express $\|\kappa_r(\mu)\|_F^2 = N^r \cdot \langle P(R_{ab} - q) \rangle$ for some fixed homogeneous degree- r polynomial P depending on r but not N . [Lemma 4.8](#) then completes the proof, since we may bound

$$\mathbf{E} \|\kappa_r(\mu)\|_F^{2a} = N^{ra} \cdot \mathbf{E} \langle P(R_{ab} - q) \rangle^a \lesssim N^{ra} \cdot N^{-ra/2},$$

where in the final step we use Hölder's inequality on P to reduce it to bounding $\mathbf{E} \langle |R_{12} - q|^{ra} \rangle \leq N^{-ra/2}$. $\square$

The $q = q\left(t + \frac{s^2\beta^2}{N}\right)$ above is defined as the unique solution to the recursion

$$q(r) = \mathbf{E}_{\mathbf{Z} \sim \mathcal{N}(0,1)} \tanh^2\left(r + \beta^2 q(r) + \mathbf{Z} \sqrt{r + \beta^2 q(r)}\right). \quad (13)$$

In the next section, we will have to bound the above for small r .

The hardest parts of the proof can be black-boxed away to the following lemma due to El Alaoui–Krzakala–Jordan.

Lemma 4.10 ([AKJ17, Remark after Theorem 8]). *Let $\tau > 0$ bounded. Consider the Ising model μ with interaction matrix $\beta \mathbf{M}$ for $\mathbf{M} = \mathbf{W} + \frac{\beta}{N} \mathbf{x} \mathbf{x}^\top$ and external field*

$$\mathbf{h} = \tau \mathbf{x} + \sqrt{\tau} \mathbf{g}. \quad (14)$$

Then, for some explicit deterministic constant $q = q(\tau)$, we have

$$\mathbf{E} \left\langle |R_{12} - q|^{2k} \right\rangle \lesssim N^{-k}.$$

Remark 4.11. Although [AKJ17] states their estimate for τ bounded away from zero, its proof is uniform as $\tau \downarrow 0$. The primary ingredients that require verification are (i) $q(\tau) = \Theta(\tau)$ as $\tau \downarrow 0$, (ii) that the preliminary concentration estimate of [AKJ17, Proposition 13] holds with uniform constants as $\tau \downarrow 0$, and (iii) the matrix $(\text{Id} - t\mathbf{A})$ in Eq. (47) is uniformly invertible. We omit the details thereof.

We shall transfer this statement to the model of interest using a contiguity statement, much like how we transitioned to the planted model earlier in Lemma 2.5.

Lemma 4.12. *Define by ν_1 the law over external fields $\mathbf{h}_1$ defined by (12) and by ν_2 the law over $\mathbf{h}_2$ defined by (14), for $\tau = t + \frac{s^2\beta^2}{N}$. Then, for any constant p and non-negative measurable function F ,*

$$\mathbf{E}_{\nu_1}[F] \lesssim \sqrt{\mathbf{E}_{\nu_2}[F^2]}.$$

Proof. It suffices to prove that $\mathbf{E}_{\mathbf{h} \sim \nu_2} \left[\frac{d\nu_1}{d\nu_2}(\mathbf{h})^2 \right] \lesssim 1$. Indeed, we then have

$$\mathbf{E}_{\nu_1}[F] = \mathbf{E}_{\mathbf{h} \sim \nu_2} \left[F(\mathbf{h}) \cdot \frac{d\nu_1}{d\nu_2}(\mathbf{h}) \right] \leq \sqrt{\mathbf{E}_{\nu_2}[F^2]} \cdot \sqrt{\mathbf{E}_{\mathbf{h} \sim \nu_2} \left[\left(\frac{d\nu_1}{d\nu_2} \right)^2 \right]}$$

by the Cauchy–Schwarz inequality. To prove this, we can establish this bound after conditioning on $\mathbf{x}$ and s . Assume we fix $s = +1$ (the proof when $s = -1$ is identical), so we have

$$\begin{aligned} \mathbf{h}_1 &\sim \mathcal{N}\left(\mathbf{x} \cdot \left(t + \frac{s\beta^2}{N}\right), t + \frac{s^2\beta^2}{N}\right) \text{ and} \\ \mathbf{h}_2 &\sim \mathcal{N}\left(\mathbf{x} \cdot \left(t + \frac{s^2\beta^2}{N}\right), t + \frac{s^2\beta^2}{N}\right). \end{aligned}$$

A straightforward calculation on these Gaussians, then shows that

$$\mathbf{E}_{\mathbf{h} \sim \nu_2} \left[\frac{d\nu_1}{d\nu_2}(\mathbf{h})^2 \right] \lesssim \exp\left(\frac{1}{t + \frac{s^2\beta^2}{N}} \cdot \left\| \mathbf{x} \cdot \left(\left(t + \frac{s\beta^2}{N}\right) - \left(t + \frac{s^2\beta^2}{N}\right) \right) \right\|^2\right)$$

$$\lesssim \exp\left(\frac{4N}{s^2\beta^2} \cdot \frac{\|\mathbf{x}\|^2 \cdot s^2\beta^4}{N^2}\right) = e^{4\beta^2}.$$

The lemma follows. $\square$

We now complete the proof as follows.

Proof of Lemma 4.8. Let μ_1 be the (random) model defined in Lemma 4.8, and μ_2 the (random) model defined in Lemma 4.10. We then have by Lemma 4.12 that

$$\mathbf{E}_{\nu_1} \langle |R_{12} - q|^{2k} \rangle \lesssim \sqrt{\left(\mathbf{E}_{\nu_2} \langle |R_{12} - q|^{2k} \rangle^2\right)} \leq \sqrt{\left(\mathbf{E}_{\nu_2} \langle |R_{12} - q|^{4k} \rangle\right)} \lesssim \frac{1}{N^k},$$

as desired, where the middle inequality is Jensen's. For the overlap concentration with a spike, we note that by exchangeability and the posterior structure of the distribution, Lemma 4.10 holds if one of the samples is replaced with the spike. $\square$

5 Lower bounds for the prediction of the inverse covariance

Recall the definition of the matrix K from (8),

$$K = \text{diag}\left(\frac{1}{1 - m_i^2}\right) - \beta \mathbf{M} + \beta^2(1 - q_N)\text{Id} - \frac{2\beta^2}{N}mm^\top$$

The goal of this subsection will be to prove the following more precise version of Lemma 4.2.

Lemma 5.1. *Let $T > 0$ large. There exists a constant $\gamma = \gamma_{\beta,T}$ such that for any constant a and $t \in [0, T]$,*

$$\Pr[K_t \not\geq \gamma] \leq N^{-a}. \quad (15)$$

5.1 The case when t is small

We start with the regime where t is small.

Lemma 5.2 ([Mai07, Corollary 1.2]). (15) holds for $t = 0$.

Lemma 5.3. *There is some small constant $t_0 = t_0(\beta)$ such that (15) holds for all $0 \leq t < t_0$.*

Proof. We have

$$\begin{aligned} K_t - K_0 &= \left(\text{diag}\left(\frac{1}{1 - m_i^2}\right) - \text{Id}\right) - \beta^2 q_N \text{Id} - \frac{2\beta^2}{N}mm^\top \\ &\succeq -3\beta^2 q_N. \end{aligned}$$

It suffices to show that for any $\varepsilon > 0$, for sufficiently small t , $q_N = \frac{\|m\|^2}{N} \leq \varepsilon$ with high probability. This would allow us to choose $\varepsilon = \frac{\gamma_0}{12\beta^2}$, for γ_0 the lower bound yielded by Lemma 5.2. We have

$$\frac{\|m\|^2}{N} = \langle R_{12} \rangle = q + \langle R_{12} - q \rangle.$$

By the overlap concentration of Lemma 4.8, we have that with probability $1 - N^{-a}$, $\langle R_{12} - q \rangle \leq \varepsilon$. It remains to establish that $q = q(t)$ is bounded, for $q(t)$ defined as in (13). Consider f defined by $f(r) = \mathbf{E} \tanh^2(r + \sqrt{r}\mathbf{Z})$, so $q(t) = f(t + \beta^2 q(t))$. We may show that $f'(r) = \mathbf{E} \text{sech}^4(r + \sqrt{r}\mathbf{Z}) \leq 1$, so $q(t) \leq t + \beta^2 q(t)$. Rearranging proves the desideratum for t sufficiently small. $\square$

5.2 The case when t is large

The bulk of the proof requires understanding the mean when t is bounded away from 0.

Lemma 5.4. *For t bounded away from 0, (15) holds.*

Fortunately, much of the hard work may be black-boxed away to pre-existing papers by Celentano [Cel24] and El Alaoui–Montanari–Sellke [EMS22]. Let us do some setup before describing exactly what these results are.

Recall

$$K(m) = \text{diag} \left(\frac{1}{1 - m_i^2} \right) - \beta \mathbf{M} + \beta^2 (1 - q_N) \text{Id} - \frac{2\beta^2}{N} m m^\top.$$

Throughout this section, let $\varepsilon > 0$ be a small constant and t bounded away from 0. The goal of this section will be to prove the following.

Lemma 5.5. *Let $\varepsilon > 0$ a small constant, $a > 0$ a large constant, and t be bounded away from 0. There exists an approximation $\hat{m} = \hat{m}(W, g)$ of the mean of μ_t such that*

- (a) $\Pr \left[\|\hat{m} - m\|_2 \geq \varepsilon \sqrt{N} \right] \leq N^{-a},$
- (b) *for some constant γ , with probability $1 - o(1)$, for all $u \in (-1, 1)^N$ with $\|\hat{m} - u\| \leq \varepsilon \sqrt{N}$, $K(u) \succeq \gamma$, and*
- (c) *$\hat{m}$ is a $O_\varepsilon(\sqrt{N})$ -Lipschitz function of W and an $O(1)$ -Lipschitz function of g with high probability. More quantitatively, if $\|W\|, \|W'\|$ are bounded by some constant, we have $\|\hat{m}(W, g) - \hat{m}(W', g')\| \leq L\sqrt{N}\|W - W'\| + L\|g - g'\|$ for some constant L depending on ε but not N .*

Before proving the above, let us describe how to complete the proof—this requires us to boost the probability in (b) from $1 - o(1)$ to $1 - N^{-a}$. We will do this using Gaussian concentration.

Lemma 5.4. *For t bounded away from 0, (15) holds.*

Proof. Define the set of “good” disorder $\mathcal{E}_{\text{good}}$ by

$$\mathcal{E}_{\text{good}} = \left\{ (W, g) : \text{for all } u \text{ with } \|\hat{m}(W, g) - u\| \leq 2\varepsilon\sqrt{N}, K_{W,g}(u) \succeq 2\gamma \right\},$$

so $\Pr [\mathcal{E}_{\text{good}}] \geq 1/2$ by Lemma 5.5(b). First, observe that by definition,

$$\|K_{W,g}(u) - K_{W',g'}(u)\|_2 \leq \|W - W'\|_2. \quad (16)$$

Suppose that we want to establish that $K_{W,g}(m)$ is bounded away from 0. Further suppose that we had, for some $(W', g') \in \mathcal{E}_{\text{good}}$, that

- (i) $\|W - W'\|_2 + \frac{1}{\sqrt{N}}\|g - g'\|_2 \leq \gamma,$
- (ii) $\|m(W, g) - \hat{m}(W, g)\| \leq \varepsilon\sqrt{N},$ and
- (iii) $\|\hat{m}(W, g) - \hat{m}(W', g')\| \leq \varepsilon\sqrt{N}.$

This would complete the proof. Indeed, (ii) and (iii) imply that $\|m(W, g) - \hat{m}(W', g')\| \leq 2\varepsilon\sqrt{N}$, so the definition of $\mathcal{E}_{\text{good}}$ gives that $K_{W', g'}(m(W, g)) \geq 2\gamma$. (16) then implies that $K_{W, g}(m(W, g)) \geq K_{W', g'}(m(W, g)) - \|W - W'\|_2 - \frac{1}{\sqrt{N}}\|g - g'\|_2 \geq \gamma$, as desired.

It remains to show that the three conditions hold with high probability. (ii) holds with probability $1 - N^{-a}$ by Lemma 5.5(a), and the Lipschitzness of Lemma 5.5(c) reduces (iii) to a statement like (i). To conclude, it suffices to establish that for any small constant η , with high probability over W , there exists $(W', g') \in \mathcal{E}_{\text{good}}$ such that $\|W - W'\| + \frac{1}{\sqrt{N}}\|g - g'\| \leq \eta$. This in fact occurs with probability $1 - e^{-cN}$ using Gaussian isoperimetry, e.g. [BLM13, Theorem 10.15]—this implies that for any set S of standard Gaussian measure at least $1/2$, $\Pr_{g \sim \mathcal{N}(0,1)} [\inf_{x \in S} \|g - x\| \geq r] \leq e^{-r^2/2}$, completing the proof. $\square$

5.3 The proof of Lemma 5.5

In this section, we prove the primary technical input to the covariance bound when t is bounded away from 0.

Lemma 5.5. *Let $\varepsilon > 0$ a small constant, $a > 0$ a large constant, and t be bounded away from 0. There exists an approximation $\hat{m} = \hat{m}(W, g)$ of the mean of μ_t such that*

- (a) $\Pr \left[\|\hat{m} - m\|_2 \geq \varepsilon\sqrt{N} \right] \leq N^{-a}$,
- (b) for some constant γ , with probability $1 - o(1)$, for all $u \in (-1, 1)^N$ with $\|\hat{m} - u\| \leq \varepsilon\sqrt{N}$, $K(u) \geq \gamma$, and
- (c) $\hat{m}$ is a $O_\varepsilon(\sqrt{N})$ -Lipschitz function of W and an $O(1)$ -Lipschitz function of g with high probability. More quantitatively, if $\|W\|, \|W'\|$ are bounded by some constant, we have $\|\hat{m}(W, g) - \hat{m}(W', g')\| \leq L\sqrt{N}\|W - W'\| + L\|g - g'\|$ for some constant L depending on ε but not N .

Towards transparency, let us explain how the approximation $\hat{m}$ is defined—it will not be too important to understand the details of all the terms here. Set $\mathbf{h} = t\mathbf{x} + \mathbf{B}_t$ for the external field. Initialize

$$\hat{m}^{(-1)} = 0 \text{ and } \hat{m}^{(0)} = \tanh(\mathbf{h}).$$

For $k \geq 0$, define

$$\hat{m}^{(k+1)} = \tanh \left(\beta \mathbf{M} \hat{m}^{(k)} + \mathbf{h} - \beta^2 \left(1 - \frac{\|\hat{m}^{(k)}\|^2}{N} \right) \hat{m}^{(k-1)} \right),$$

where $\tanh$ is applied entrywise. We will choose $\hat{m} = \hat{m}^{(k_*)}$ for some large constant k_* depending on ε but not N .

Proof of Lemma 5.5(a,c). (c) is immediate from the definition of $\hat{m}$, since we are choosing a large constant iterate of this process. For (a), we wish to show that $m^\top \hat{m}$ concentrates. Let $\mathcal{D} = (\mathbf{M}, \mathbf{h})$, so $m = \mathbf{E}[\mathbf{x} \mid \mathcal{D}]$. That is,

$$\frac{m^\top \hat{m}}{N} = \mathbf{E} \left[\frac{\mathbf{x}^\top \hat{m}}{N} \mid \mathcal{D} \right].$$

For any fixed iterate k , [Cel24, Proposition 3] establishes that $\frac{x^\top \hat{m}}{N} \rightarrow q_k$ in probability, for some explicit constant q_k . Consider the random variable $\rho = \frac{x^\top \hat{m}}{N}$, conditioned on x , defined in terms of the disorder W and the noise in the external field g . Using the Kirschbraun extension theorem and the Lipschitzness of (c), we may define a function $\tilde{m}$ that is $O(\sqrt{N})$ -Lipschitz in W everywhere, and $\tilde{m} = \hat{m}$ with probability $1 - e^{-cN}$. Define $\tilde{\rho} = \frac{x^\top \tilde{m}}{N}$. Gaussian concentration yields that

$$\Pr[|\tilde{\rho} - \mathbf{E}\tilde{\rho}| > s] \leq e^{-cNs^2},$$

and $\rho \rightarrow q_k$ in probability also implies that for sufficiently large k , for constant δ ,

$$\Pr[|\rho - q_k| > \delta] \leq e^{-cN}.$$

Therefore,

$$\Pr\left[\left|\frac{m^\top \hat{m}}{N} - q_k\right| \geq \delta\right] \leq e^{-cN}$$

To conclude, we write

$$\frac{\|m - \hat{m}\|^2}{N} = \frac{\|m\|^2}{N} + \frac{\|\hat{m}\|^2}{N} - 2 \cdot \frac{\hat{m}^\top m}{N}.$$

Overlap concentration (Lemma 4.8) implies that $\left|\frac{\|m\|^2}{N} - q\right| \leq \delta$ with probability $1 - N^{-a}$, and we have just shown that $\left|\frac{\hat{m}^\top m}{N} - q_k\right| \leq \delta$ with probability $1 - N^{-a}$. An argument very similar to the Lipschitzness one above also yields that $\left|\frac{\|\hat{m}\|^2}{N} - q_k\right| \leq \delta$, using [Cel24, Proposition 3], with probability $1 - N^{-a}$. Taking k large enough that $|q_k - q| \leq \delta$, and adjusting the constants establishes (a). $\square$

The remainder of this section will be spent on (b). In fact, [Cel24, Theorem 2(i)] establishes that $K(\hat{m}) + \frac{2\beta^2}{N} \hat{m} \hat{m}^\top \succeq \gamma$ with high probability, and our proof will be spent showing that the extra rank-one term is innocuous.

The proof will follow the same blueprint as [Cel24], but there will be an extra term (corresponding to the rank-one term) that will slightly influence the final step of the proof.

Proof of Lemma 5.5(b). Throughout this proof, we denote by δ_k a small constant which goes to 0 as the iterate k_* defining $\hat{m}$ goes to ∞ . First off, the fact that $\|u - \hat{m}\|$ is small means that we only need to argue about the “mixed” matrix,

$$\tilde{K}(u) := \text{diag}\left(\frac{1}{1 - u_i^2}\right) - \beta M + \beta^2 \left(1 - \frac{\|u\|^2}{N}\right) \text{Id} - \frac{2\beta^2}{N} \hat{m} \hat{m}^\top.$$

Let $v \in \mathbb{S}^{N-1}$ be an arbitrary vector (possibly depending on the disorder in some uncontrolled way), so we are interested in showing the (strict) negativity of

$$-v^\top \tilde{K}(u) v = \beta v^\top W v + R(u, v),$$

for some function $R = R_W$. [Cel24, Corollary 1], the “post-AMP Sudakov–Fernique inequality,” establishes that the first term may be replaced as

$$\text{plimsup}_{N \rightarrow \infty} \sup_{u, v} \left\{ \beta v^\top W v + R(u, v) \right\} \leq \text{plimsup}_{N \rightarrow \infty} \sup_{u, v} \left\{ \frac{2\beta}{\sqrt{N}} \langle g_{\text{AMP}}(v), v \rangle + R(u, v) \right\}$$

for some explicit function g_{AMP} defined by the disorder, whose definition we will refrain from introducing. Let $I \sim [N]$ be a random index, and define $M = \hat{m}_I$, $U = u_I$, and $V = \sqrt{N}v_I$. The extra rank-one term $\frac{2\beta^2}{N} \langle v, \hat{m} \rangle^2$ is then precisely $2\beta^2 \mathbf{E}[MV]^2$. Assume wlog that $x = 1$.

The next step is to move to appropriate *limiting* distributions for all these random variables. At the cost of some small error δ_k that goes to 0 as the iterate k (defining $\hat{m}$), this allows us to work with idealized distributions that are explicitly defined in terms of Gaussians. For example, we can pretend that M is distributed as $\tanh(t + \beta^2 q + z\sqrt{t + \beta^2 q})$ for $z \sim \mathcal{N}(0, 1)$, up to some small error.

The upshot, which is the content of [Cel24, Sections 4.2–4.4], is essentially [Cel24, Eq. (26)] with an extra term corresponding to the rank-one part. This is as follows.

$$\mathcal{J}(U, V) = 2\beta \mathbf{E}[B_\rho V] - \mathbf{E} \left[\frac{V^2}{1 - U^2} \right] + \beta^2 (\mathbf{E}[V])^2 - \beta^2 (1 - q) + \underbrace{2\beta^2 q \rho^2}_{\text{new term}},$$

where we have the constraints

$$\mathbf{E}[V^2] = 1, \quad |U| \leq 1, \quad \mathbf{E}[SV] = \rho\sqrt{q}, \quad \|U - S\|_2 \leq \varepsilon + \delta_k$$

and for some choice of $z_1, z_2, z_3 \sim \mathcal{N}(0, 1)$ that are independent from each other,

$$\begin{aligned} S &= \tanh(t + \beta^2 q + z_1 \beta \sqrt{q} + z_2 \sqrt{t}) , \\ B_\rho &= \rho z_1 + z_3 \sqrt{1 - \rho^2}. \end{aligned} \tag{17}$$

In particular, $\mathbf{E}[S^2] = q$, and $\rho = \frac{\mathbf{E}[SV]}{\sqrt{q}} \in [-1, 1]$. Here, S plays the role of the typical coordinate of $\hat{m}$. The next step is to also replace the U with S , suffering a small constant loss.

Towards this, we have for any choice of a, b, α that

$$\begin{aligned} \mathcal{J} &= 2\beta \mathbf{E}[B_\rho V] - \mathbf{E} \left[\frac{V^2}{1 - U^2} \right] + \beta^2 (\mathbf{E}[V])^2 - \beta^2 (1 - q) + 2\beta^2 q \rho^2 \\ &\quad + a \left(\frac{\mathbf{E}[SV]}{\sqrt{q}} - \rho \right) + b (\mathbf{E}[V] - \mathbf{E}[V]) + \alpha (\mathbf{E}[V^2] - 1) \\ &= \mathbf{E} \left[V \left(2\beta B_\rho + \frac{aS}{\sqrt{q}} + b \right) - V^2 \left(\frac{1}{1 - U^2} - \alpha \right) \right] \\ &\quad + \beta^2 (\mathbf{E}[V])^2 - \beta^2 (1 - q) + 2\beta^2 q \rho^2 - a\rho - b\mathbf{E}[V] - \alpha \\ &\leq \frac{1}{4} \mathbf{E} \left[\frac{\left(2\beta B_\rho + \frac{aS}{\sqrt{q}} + b \right)^2}{\frac{1}{1 - U^2} - \alpha} \right] + \beta^2 (\mathbf{E}[V])^2 - \beta^2 (1 - q) + 2\beta^2 q \rho^2 - a\rho - b\mathbf{E}[V] - \alpha. \end{aligned}$$

The final inequality here is by the AM-GM inequality: we have $cv - dv^2 \leq \frac{c^2}{4d}$. Note now that if $0 \leq \alpha \leq 1/2$, we may replace U with S in the first expression, at the cost of a small loss. After doing this, we are left with an extremely explicit calculation in terms of Gaussians, since essentially all dependence on V has been removed. Now denote $\eta = \mathbf{E}[V]$.

The result is a modification [Cel24, Eq. (27)], which asks to show that

$$\inf_{0 \leq \alpha \leq 1/2} \mathcal{L}_\alpha := \inf_{\substack{a, b \in \mathbb{R} \\ 0 \leq \alpha \leq 1/2}} \frac{1}{4} \mathbf{E} \left[\frac{\left(2\beta B_\rho + \frac{aS}{\sqrt{q}} + b \right)^2}{\frac{1}{1 - S^2} - \alpha} \right] + \beta^2 \eta^2 - \beta^2 (1 - q) + 2\beta^2 q \rho^2 - a\rho - b\eta - \alpha$$

is negatively bounded away from 0. In fact, it will suffice to show that $\mathcal{L}_0 \leq 0$ —the remainder of the proof follows [Cel24, Section 4.4], which uses a simple continuity argument to transform this to a strict negative margin by setting α to be a small positive constant. To do this, we set $a = 2\beta^2(2q\rho - \eta\sqrt{q})$ and $b = 2\beta^2\eta$. This then simplifies to

$$\mathcal{L}_0 \leq -\beta^2(1 - \beta^2\nu) (q\rho^2 + (\eta - \rho\sqrt{q})^2) ,$$

for $\nu = \mathbb{E}[1 - S^2]$. We omit the details of this calculation—it amounts to integrating out z_3 , using Gaussian integration by parts to explicitly compute terms like $\mathbb{E}[(1 - S^2)z_1^2]$ (which show up when expanding out the parenthesised terms in the expectation), and completing appropriate squares—such calculations are very similar to those done in [Cel24, pages 16 and 17]. $\square$

Acknowledgments

We would like to thank David X. Wu and Mark Sellke for many technical discussions during earlier stages of this project. S.M. and A.R. would like to thank the Simons Institute for the Theory of Computing at Berkeley for their hospitality and hosting them at the “Spectral Theory Beyond Graphs” program, where part of the work was conducted. H.D. was partially supported by an NSF-Simons research collaboration grant (award number 2031883). B.H. was supported by a Stanford Science Fellowship and NSF Mathematical Sciences Postdoctoral Fellowship.

References

- [ABSY21] Arka Adhikari, Christian Brennecke, Per von Soosten, and Horng-Tzer Yau. Dynamical approach to the TAP equations for the Sherrington–Kirkpatrick model. *Journal of Statistical Physics*, 183(3):35, 2021. [6](#)
- [AC08] 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. [3](#)
- [Aid98] Shigeki Aida. Uniform positivity improving property, Sobolev inequalities, and spectral gaps. *Journal of functional analysis*, 158(1):152–185, 1998. [2](#)
- [AJ24] Gérard Ben Arous and Aukosh Jagannath. Shattering versus metastability in spin glasses. *Communications on Pure and Applied Mathematics*, 77(1):139–176, 2024. [5](#)
- [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. [5](#)
- [AKJ17] Ahmed El Alaoui, Florent Krzakala, and Michael I Jordan. Finite size corrections and likelihood ratio fluctuations in the spiked Wigner model. *arXiv preprint arXiv:1710.02903*, 2017. [20](#)

- [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. [4](#), [6](#), [9](#)
- [ALOV24] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-Concave Polynomials II: High-Dimensional Walks and an FPRAS for Counting Bases of a Matroid. *Annals of Mathematics*, 199(1):259–299, 2024. [6](#)
- [ASS03] Michael Aizenman, Robert Sims, and Shannon L. Starr. An Extended Variational Principle for the Sherrington–Kirkpatrick Spin-Glass Model. *Physical Review B*, 68(21):214403, 2003. [1](#), [6](#)
- [AT78] Jairo RL de Almeida and David J Thouless. Stability of the Sherrington-Kirkpatrick solution of a spin glass model. *Journal of Physics A: Mathematical and General*, 11(5):983, 1978. [4](#), [6](#), [8](#)
- [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. [4](#)
- [BBM96] Alain Barrat, Raffaella Burioni, and Marc Mézard. Dynamics within metastable states in a mean-field spin glass. *Journal of Physics A: Mathematical and General*, 29(5):L81–L87, 1996. [5](#)
- [BBP05] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. 2005. [2](#)
- [BDL⁺26] Afonso S Bandeira, Daniil Dmitriev, Kevin Lucca, Petar Nizić-Nikolac, and Almut Rödder. Randomstrasse101: Open Problems of 2025. *arXiv preprint arXiv:2603.29571*, 2026. [1](#)
- [BER26] Afonso S Bandeira, Ahmed El Alaoui, and Almut Rödder. Mixing of Glauber Dynamics on High Overlap Gibbs Measures. *arXiv preprint arXiv:2607.06813*, 2026. [4](#), [8](#), [33](#)
- [BHL⁺26] Ferenc Bencs, Brice Huang, Daniel Z. Lee, Kuikui Liu, and Guus Regts. On Zeros and Algorithms for Disordered Systems: Mean-Field Spin Glasses. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing (STOC)*, pages 508–517, 2026. [8](#)
- [BKL26] Pierre Bizeul, Boaz Klartag, and Joseph Lehec. Presenting a proof of the Kannan–Lovász–Simonovits conjecture. *arXiv preprint arXiv:2610.05474*, 2026. [6](#)
- [BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013. [23](#)
- [BLO26] Mathews Boban, Anqi Li, and Shayan Oveis Gharan. Rank-1-perturbed trickledown theorems: Mixing time of Glauber dynamics for the Sherrington-Kirkpatrick model up to $\beta \leq \frac{1}{2} + \varepsilon$. *arXiv preprint arXiv:2609.13138*, 2026. [5](#)
- [BM11] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011. [6](#)

- [Bol14] Erwin Bolthausen. An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Communications in Mathematical Physics*, 325(1):333–366, 2014. [6](#)
- [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):9, 2024. [6](#)
- [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. [6](#)
- [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. [2](#), [3](#), [4](#), [6](#), [9](#), [33](#)
- [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. [1](#), [4](#), [5](#), [6](#), [22](#), [24](#), [25](#), [26](#)
- [CGTvH26] Chi-Fang Chen, Jorge Garza-Vargas, Joel A Tropp, and Ramon van Handel. A New Approach to Strong Convergence. *Annals of Mathematics*, 203(2):555–602, 2026. [10](#)
- [Che21] Yuansi Chen. An Almost Constant Lower Bound of the Isoperimetric Coefficient in the KLS Conjecture. *Geometric and Functional Analysis*, 31(1):34–61, 2021. [6](#)
- [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(2):257–271, 1993. [5](#)
- [CT21] Wei-Kuo Chen and Si Tang. On convergence of the cavity and Bolthausen’s TAP iterations to the local magnetization. *Communications in Mathematical Physics*, 386(2):1209–1242, 2021. [6](#)
- [DAM16] Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the binary stochastic block model. pages 185–189, 07 2016. [6](#)
- [DLSS26a] Ewan Davies, Holden Lee, Juspreet Singh Sandhu, and Jonathan Shi. Potential Hessian Ascent III: Sampling the Sherrington–Kirkpatrick Model at $\beta < 1/2$. *arXiv preprint arXiv:2605.03718*, 2026. [5](#)
- [DLSS26b] Ewan Davies, Holden Lee, Juspreet Singh Sandhu, and Jonathan Shi. Weak Poincaré Inequalities via Approximate Stochastic Localization: Application to Sampling the Sherrington–Kirkpatrick Model. *arXiv preprint arXiv:2607.08160*, 2026. [5](#), [6](#)
- [DMM09] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009. [6](#)

- [DS19] Jian Ding and Nike Sun. Capacity lower bound for the Ising perceptron. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 816–827, 2019. [6](#)
- [EEGP23] Ahmed El Alaoui, Ronen Eldan, Reza Gheissari, and Arianna Piana. Fast relaxation of the random field Ising dynamics. *arXiv preprint arXiv:2311.06171*, 2023. [6](#)
- [EG24] 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](#)
- [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. [4](#), [6](#), [9](#)
- [El 25] Ahmed El Alaoui. Near-optimal shattering in the Ising pure p-spin and rarity of solutions returned by stable algorithms. *Stochastic Processes and their Applications*, page 104792, 2025. [5](#)
- [Eld13] Ronen Eldan. Thin Shell Implies Spectral Gap Up to Polylog via a Stochastic Localization Scheme. *Geometric and Functional Analysis*, 23(2):532–569, March 2013. [3](#), [6](#)
- [EM22] Ahmed El Alaoui and Andrea Montanari. An information-theoretic view of stochastic localization. *IEEE Transactions on Information Theory*, 68(11):7423–7426, 2022. [3](#), [12](#)
- [EMS21] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of Mean-Field Spin Glasses. *The Annals of Probability*, 49(6):2922–2960, 2021. [6](#)
- [EMS22] 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. [3](#), [5](#), [22](#)
- [EMS23] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from Mean-Field Gibbs Measures via Diffusion Processes. *arXiv preprint arXiv:2310.08912*, 2023. [3](#)
- [EMS25] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Shattering in pure spherical spin glasses. *Communications in Mathematical Physics*, 406(5):111, 2025. [5](#)
- [ES22] Ahmed El Alaoui and Mark Sellke. Algorithmic Pure States for the Negative Spherical Perceptron. *Journal of Statistical Physics*, 189(2):27, 2022. [6](#)
- [FV17] Sacha Friedli and Yvan Velenik. Statistical Mechanics of Lattice Systems. A Concrete Mathematical Introduction. 2017. [1](#)
- [GLLR21] Rong Ge, Holden Lee, Jianfeng Lu, and Andrej Risteski. Efficient sampling from the Bingham distribution. In *Algorithmic Learning Theory*, pages 673–685. PMLR, 2021. [5](#)

- [GT02] Francesco Guerra and Fabio Lucio Toninelli. The Thermodynamic Limit in Mean Field Spin Glass Models. *Communications in Mathematical Physics*, 230(1):71–79, 2002. [1](#)
- [Gue03] Francesco Guerra. Broken Replica Symmetry Bounds in the Mean Field Spin Glass Model. *Communications in Mathematical Physics*, 233(1):1–12, 2003. [1](#)
- [GY91] Antoine Georges and Jonathan S Yedidia. How to expand around mean-field theory using high-temperature expansions. *Journal of Physics A: Mathematical and General*, 24(9):2173–2192, 1991. [9](#)
- [HJJ03] Darrall Henderson, Sheldon H. Jacobson, and Alan W. Johnson. The Theory and Practice of Simulated Annealing. In Fred Glover and Gary A. Kochenberger, editors, *Handbook of Metaheuristics*, pages 287–319. Kluwer Academic Publishers, Boston, 2003. [1](#)
- [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](#), [7](#), [8](#)
- [HMRW24] Brice Huang, Sidhanth Mohanty, Amit Rajaraman, and David X Wu. Weak Poincaré Inequalities, Simulated Annealing, and Sampling from Spherical Spin Glasses. *arXiv preprint arXiv:2411.09075*, 2024. [2](#), [3](#), [5](#), [6](#), [7](#), [8](#), [11](#), [13](#)
- [HS24] Brice Huang and Mark Sellke. Optimization Algorithms for Multi-Species Spherical Spin Glasses. *Journal of Statistical Physics*, 191:29, 2024. [6](#)
- [JAMS89] David S. Johnson, Cecilia R. Aragon, Lyle A. McGeoch, and Catherine Schevon. Optimization by Simulated Annealing: An Experimental Evaluation; Part I, Graph Partitioning. *Operations Research*, 37(6):865–892, 1989. [1](#)
- [JVV86] Mark R. Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random Generation of Combinatorial Structures from a Uniform Distribution. *Theoretical Computer Science*, 43:169–188, 1986. [8](#)
- [KAJ94] Christos Koulamas, S. R. Antony, and R. Jaen. A Survey of Simulated Annealing Applications to Operations Research Problems. *Omega*, 22(1):41–56, 1994. [1](#)
- [KGV83] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598):671–680, 1983. [1](#)
- [KL22] Bo’az Klartag and Joseph Lehec. Bourgain’s Slicing Problem and KLS Isoperimetry up to Polylog. *Geometric and Functional Analysis*, 32(5):1134–1159, 2022. [6](#)
- [KN26] Seiichiro Kusuoka and Shuta Nakajima. A quantitative replica-symmetric bound of Sherrington–Kirkpatrick model in the entire de Almeida–Thouless region. *arXiv preprint arXiv:2608.23413*, 2026. [4](#)
- [LA87] Peter J. M. van Laarhoven and Emile H. L. Aarts. *Simulated Annealing: Theory and Applications*, volume 37 of *Mathematics and Its Applications*. D. Reidel Publishing Company, Dordrecht, 1987. [1](#)

- [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. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 203–215. IEEE, 2024. [1](#), [8](#)
- [LMRW24] Kuikui Liu, Sidhanth Mohanty, Amit Rajaraman, and David X Wu. Fast Mixing in Sparse Random Ising Models. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 120–128. IEEE, 2024. [6](#)
- [Lop26] Patrick Lopatto. Replica symmetry up to the de Almeida-Thouless line in the Sherrington-Kirkpatrick model. *arXiv preprint arXiv:2604.11921*, 2026. [4](#)
- [LSS26] Holden Lee, Juspreet Singh Sandhu, and Jonathan Shi. Potential Hessian Ascent IV: Sampling the Sherrington-Kirkpatrick model at $\beta < 1$, 2026. [1](#), [5](#)
- [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. [6](#)
- [Mai07] Mylène Maïda. Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. *Electronic Journal of Probability*, 12(41):1131–1150, 2007. [21](#)
- [Mat06] P Mathieu. Quand l’inegalite log-Sobolev implique l’inegalite de trou spectral. In *Séminaire de Probabilités XXXII*, pages 30–35. Springer, 2006. [2](#)
- [MM09] Marc Mézard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, 01 2009. [1](#)
- [Mon19] Andrea Montanari. Optimization of the Sherrington–Kirkpatrick Hamiltonian. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1417–1433. IEEE, 2019. [6](#)
- [MPV87a] Marc Mézard, Giorgio Parisi, and Miguel A Virasoro. SK model: The replica solution without replicas. *Spin Glass Theory and Beyond: An Introduction to the Replica Method and its Applications*. Edited by Mezard M et al. Published by World Scientific Press, pages 232–237, 1987. [6](#)
- [MPV87b] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, volume 9 of *World Scientific Lecture Notes in Physics*. World Scientific, 1987. [1](#)
- [MV21] Andrea Montanari and Ramji Venkataramanan. Estimation of low-rank matrices via approximate message passing. 2021. [6](#)
- [Ope26] OpenAI. A Spectral Gap Throughout the High-Temperature Sherrington–Kirkpatrick Phase, September 2026. Preprint, September 24, 2026. [6](#), [8](#)
- [Pan12] Dmitry Panchenko. The Sherrington–Kirkpatrick Model: An Overview. *Journal of Statistical Physics*, 149(2):362–383, 2012. [1](#)

- [Pan13] Dmitry Panchenko. The Parisi Ultrametricity Conjecture. *Annals of Mathematics*, 177(1):383–393, 2013. [1](#)
- [Par79] Giorgio Parisi. Infinite Number of Order Parameters for Spin-Glasses. *Physical Review Letters*, 43(23):1754–1756, 1979. [1](#)
- [Par80a] Giorgio Parisi. A Sequence of Approximated Solutions to the S-K Model for Spin Glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115–L121, 1980. [1](#)
- [Par80b] Giorgio Parisi. The Order Parameter for Spin Glasses: A Function on the Interval 0–1. *Journal of Physics A: Mathematical and General*, 13(3):1101–1112, 1980. [1](#)
- [Ple82] Timm Plefka. Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model. *Journal of Physics A: Mathematical and general*, 15(6):1971–1978, 1982. [9](#)
- [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. [2](#)
- [RW26] Amit Rajaraman and David X Wu. Markov Chains Approximate Message Passing. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing*, pages 1192–1199, 2026. [1](#), [8](#)
- [Sim20] Simons Institute for the Theory of Computing. Rapid Mixing of Glauber Dynamics for the Sherrington–Kirkpatrick Model. <https://wiki.simons.berkeley.edu/doku.php?id=hd20:glauber-dynamics-for-sk>, 2020. Polymath project, program on Probability, Geometry, and Computation in High Dimensions. Contact: Andrea Montanari. [1](#)
- [SJ89] Alistair Sinclair and Mark Jerrum. Approximate Counting, Uniform Generation and Rapidly Mixing Markov Chains. *Information and Computation*, 82(1):93–133, 1989. [8](#)
- [SS85] Carl Sechen and Alberto Sangiovanni-Vincentelli. The TimberWolf Placement and Routing Package. *IEEE Journal of Solid-State Circuits*, 20(2):510–522, 1985. [1](#)
- [SZ81] H. Sompolinsky and Annette Zippelius. Dynamic Theory of the Spin-Glass Phase. *Physical Review Letters*, 47(5):359–362, 1981. [1](#)
- [SZ82] H. Sompolinsky and Annette Zippelius. Relaxational Dynamics of the Edwards–Anderson Model and the Mean-Field Theory of Spin-Glasses. *Physical Review B*, 25(11):6860–6875, 1982. [1](#)
- [SZ26] Zhao Song and Xinzhi Zhang. An $O(1)$ Bound for the KLS Constant. *arXiv preprint arXiv:2610.01447*, 2026. [6](#)
- [Tal06] Michel Talagrand. The Parisi Formula. *Annals of Mathematics*, 163(1):221–263, 2006. [1](#)
- [Tal10] Michel Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*, volume 54. Springer Science & Business Media, 2010. [1](#), [6](#)

- [Tal11] Michel Talagrand. *Mean Field Models for Spin Glasses: Volume II: Advanced replica-symmetry and low temperature*, volume 55. Springer Science & Business Media, 2011. [1](#), [5](#), [6](#), [12](#)
- [TAP77] David J Thouless, Philip W Anderson, and Robert G Palmer. Solution of ‘Solvable Model of a Spin Glass’. *Philosophical Magazine*, 35(3):593–601, 1977. [6](#)
- [Wan26] Sihan Wang. Optimal Mixing of Glauber Dynamics for the Sherrington-Kirkpatrick Model at $\beta < 1/2$. *arXiv preprint arXiv:2608.22159*, 2026. [5](#), [9](#)
- [WJ08] Martin J. Wainwright and Michael I. Jordan. *Graphical Models, Exponential Families, and Variational Inference*, volume 1 of *Foundations and Trends in Machine Learning*. Now Publishers Inc, 2008. [1](#)
- [WLL88] D. F. Wong, H. W. Leong, and C. L. Liu. *Simulated Annealing for VLSI Design*. Kluwer Academic Publishers, 1988. [1](#)
- [YRN94] Takeshi Yamada, Bruce E. Rosen, and Ryohei Nakano. Critical Block Simulated Annealing for Job Shop Scheduling. *Transactions of the Institute of Electrical Engineers of Japan, Part C*, 114(4):476–482, 1994. [1](#)
- [ZSSJ26] Leo Zhou, Noah Shetty, Mark Sellke, and Stephen P. Jordan. Gibbs Sampling in the Shattered Phase by Decoded Quantum Interferometry, 2026. [8](#)

A Tying up loose ends

We start by proving [Lemma 1.5](#), which is essentially an immediate consequence of the techniques of [\[BER26\]](#).

Lemma 1.5 ([\[BER26\]](#)). *Let μ_T be the measure obtained by running stochastic localization on $\mu_{\beta W + \frac{\beta^2}{N} \mathbf{x} \mathbf{x}^\top}$ for bounded time $T > 0$. Then, for a sufficiently large (constant) T , with probability $1 - e^{-cN}$, μ_T has Poincaré constant $O(N)$.*

Proof. Assume without loss of generality that $\mathbf{x} = \mathbf{1}$. We start by noting that [\[BER26, Lemma 2.3\]](#) holds for sufficiently large localization time. Indeed, choose δ such that the first condition of [\[BER26, Theorem 1.1\]](#) is satisfied, and every small submatrix has sufficiently small operator norm, which occurs with probability $1 - e^{-cN}$. The external field is of the form $T\mathbf{x} + g\sqrt{T}$. For sufficiently large T , with probability $1 - e^{-cN}$, we have $\|g\sqrt{T}\| \ll T\sqrt{N}$. Consequently, a sample from this measure has at least $(1 - \frac{\delta}{2})N$ positive spins, which by [\[BER26, Lemma 2.2\]](#) yields a correlation bound (for the base model as well as the pinned models).

To conclude, we use this correlation bound until we have pinned $(1 - \delta)N$ of the coordinates, and we are left with an Ising model (with some external field) whose interaction matrix has bounded spectral radius; [\[CE22, Corollary 51\]](#) yields a Poincaré inequality for these measures, and the spectral independence framework allows us to lift this back to a Poincaré inequality for μ_T . $\square$

A.1 Controlling error terms

Over the remainder of this section, we prove [Lemmas 4.6](#) and [4.7](#). For $s \in [-1, 1]$, let $\kappa_{r,s}^{(i)}$ denote $\nabla^r F^{(i)}(u_i + s\beta \mathbf{M}_{i,-i})$ —in words, the r th cumulant of the model where the i th site is removed, and the field is shifted by $s\beta \mathbf{M}_{i,-i}$. More generally for $B \subseteq [N]$, we will use $\kappa_{r,s}^{(B)}$ to denote $\nabla^r F^{(i)}(u_B + s\beta \sum_{i \in B} \mathbf{M}_{i,-i})$ where u_B refers to the external field on the remaining sites.

We will rely on the following estimates for the cumulant, whose proof we defer to [Section A.3](#).

Lemma A.1. *For any fixed integers $r \geq 3$ and $k, \ell, p \geq 0$ satisfying $2k + \ell \leq r$, and fixed $s \in [-1, 1]$,*

$$\mathbf{E} \left\| \kappa_{r,s}^{(i)} \left[\text{Id}^{\otimes k} \otimes \mathbf{M}_{i,-i}^{\otimes \ell} \right] \right\|_F^p \leq c_p N^{p(r/4 - \ell/2)}.$$

Further, the same bound continues to hold if $\kappa_{r,s}^{(i)}$ is replaced with $\kappa_{r,s}^{(B)}$ where $|B|$ is a constant independent of N .

We are now ready to prove [Lemma 4.7](#), restated for convenience.

Lemma 4.7. *For any $a, \varepsilon > 0$, with probability $1 - O(N^{-a})$, we have*

$$\|\Delta_{d,i}\|_2 \leq O(N^{-1+\varepsilon}), \quad \|\Delta_{\text{Cov},i} \mathbf{M}_{i,-i}\|_2 \leq O(N^{-3/4+\varepsilon}).$$

and consequently, $\|\Delta_{f,i}\|_2 \leq \|\Delta_{d,i}\|_2 + \beta \|\Delta_{\text{Cov},i} \mathbf{M}_{i,-i}\|_2 \leq O(N^{-3/4+\varepsilon})$.

Proof. By Taylor's theorem, the remainder term admits the following description:

$$\begin{aligned} \Delta_{d,i} &= \frac{\beta^5}{48} \int_0^1 (1-s)^4 (\kappa_{6,s}^{(i)} + \kappa_{6,-s}^{(i)}) [\mathbf{M}_{i,-i}^{\otimes 5}] ds, \\ \Delta_{\text{Cov},i} \mathbf{M}_{i,-i} &= \frac{\beta^3 m_i}{4} \int_0^1 (1-s)^2 (\kappa_{5,s}^{(i)} + \kappa_{5,-s}^{(i)}) [\mathbf{M}_{i,-i}^{\otimes 4}] ds \\ &\quad + \frac{\beta^4}{12} \int_0^1 (1-s)^3 (\kappa_{6,s}^{(i)} + \kappa_{6,-s}^{(i)}) [\mathbf{M}_{i,-i}^{\otimes 5}] ds. \end{aligned}$$

By [Lemma A.1](#) and triangle inequality, we have:

$$\mathbf{E} \|\Delta_{d,i}\|_F^p \leq CN^{-p}, \quad \mathbf{E} \|\Delta_{\text{Cov},i} \mathbf{M}_{i,-i}\|_F^p \leq CN^{-3p/4}.$$

The conclusion follows by applying Markov's inequality to the above for an appropriately large choice of p depending on ε . $\square$

We now turn our attention to proving [Lemma 4.6](#). A key ingredient is the following set of improved cumulant bounds for particular special cases. We defer the proof to [Section A.4](#).

Lemma A.2 (Improved cumulant bounds). *There exists an absolute constant $C > 0$ such that the following holds.*

(a) *For all $p \geq 2$, there exists $C_p > 0$ such that*

$$\max \left\{ \mathbf{E} \|\kappa_3\|_F^p, \max_{i \in [N]} \mathbf{E} \|\kappa_3^{(i)}\|_F^p \right\} \leq C_p N^{p/2}, \quad \max \left\{ \mathbf{E} \|\kappa_4\|_F^p, \max_{i \in [N]} \mathbf{E} \|\kappa_4^{(i)}\|_F^p \right\} \leq C_p N^{3p/4}.$$

(b) For all $p \geq 2$, there exists $C_p > 0$ such that

$$\max \left\{ \mathbf{E} \|\mathbf{Cov}\|_{\text{op}}^p, \max_{i \in [N]} \mathbf{E} \|\mathbf{Cov}^{(i)}\|_{\text{op}}^p \right\} \leq C_p.$$

(c) For all $a > 0$, there exists $C_a > 0$ such that

$$\Pr \left[\max \left\{ \|\mathbf{Cov}\|_F, \max_{i \in [N]} \|\mathbf{Cov}^{(i)}\|_F \right\} > C\sqrt{N} \right] \leq C_a N^{-a}.$$

(d) For all $a > 0$, there exists $C_a > 0$ such that

$$\Pr \left[\max \left\{ \|\kappa_3\|_F, \max_{i \in [N]} \|\kappa_3^{(i)}\|_F \right\} > C\sqrt{N} \right] \leq C_a N^{-a}.$$

We now prove [Lemma 4.6](#).

A.2 Proof of [Lemma 4.6](#)

Lemma 4.6. For every constant $a > 0$, the following bounds hold for all $i \in [N]$ with probability $1 - O_\varepsilon(N^{-a})$:

$$\begin{aligned} \left\| \frac{2\beta^2 m_i}{N} \left(\mathbf{Cov}^{(i)} \cdot m^{(i)} - (\mathbf{Cov} \cdot m)_{-i} \right) \right\|_2 &\leq \frac{C}{\sqrt{N}} \\ \left\| \frac{\beta^3}{3} \kappa_4^{(i)} \left[\mathbf{M}_{i,-i}^{\otimes 3} \right] \right\|_2 &\leq \frac{C}{\sqrt{N}} \\ \left\| \beta^2 m_i \kappa_3^{(i)} \left[\mathbf{M}_{i,-i}^{\otimes 2} - \frac{1}{N} \text{Id} \right] \right\|_2 &\leq \frac{C}{\sqrt{N}} \\ \max_{i \in [N]} |\lambda_i| &\leq N^{-1/4}. \end{aligned}$$

Bound on $\left\| \frac{2\beta^2 m_i}{N} \left(\mathbf{Cov}^{(i)} \cdot m^{(i)} - (\mathbf{Cov} \cdot m)_{-i} \right) \right\|$. First, observe that:

$$\nabla_{u_i} \left[\text{Tr } \mathbf{Cov} - \text{Tr } \mathbf{Cov}^{(i)} \right] = 2 \left(\mathbf{Cov}^{(i)} m^{(i)} - (\mathbf{Cov} \cdot m)_{-i} \right).$$

By (9), we have:

$$\begin{aligned} \text{Tr } \mathbf{Cov} - \text{Tr } \mathbf{Cov}^{(i)} &= \frac{1+m_i}{2} \text{Tr } \kappa_{2,+}^{(i)} + \frac{1-m_i}{2} \text{Tr } \kappa_{2,-}^{(i)} - \text{Tr } \kappa_2^{(i)} + v_i(1 + \|d_i\|_2^2) \\ &= \frac{\text{Tr}(\kappa_{2,+}^{(i)} + \kappa_{2,-}^{(i)})}{2} + m_i \frac{\text{Tr}(\kappa_{2,+}^{(i)} - \kappa_{2,-}^{(i)})}{2} - \text{Tr } \kappa_2^{(i)} + v_i(1 + \|d_i\|_2^2). \end{aligned} \quad (18)$$

In preparation to compute the gradient of the above, observe that:

$$\nabla_{u_i} m_i = v_i d_i, \quad \nabla_{u_i} v_i = -2m_i v_i d_i, \quad \nabla_{u_i} d_i = \frac{\kappa_{2,+}^{(i)} - \kappa_{2,-}^{(i)}}{2}, \quad \nabla_{u_i} \text{Tr } \kappa_{2,\pm}^{(i)} = \kappa_{3,\pm}^{(i)} [\text{Id}]$$

We differentiate the three terms in

$$\text{Tr } \mathbf{Cov} - \text{Tr } \mathbf{Cov}^{(i)} = \left(\frac{\text{Tr}(\kappa_{2,+}^{(i)} + \kappa_{2,-}^{(i)})}{2} - \text{Tr } \kappa_2^{(i)} \right) + \frac{m_i}{2} \text{Tr}(\kappa_{2,+}^{(i)} - \kappa_{2,-}^{(i)}) + v_i(1 + \|d_i\|_2^2)$$

separately. First,

$$\nabla_{u_i} \left(\frac{\text{Tr}(\kappa_{2,+}^{(i)} + \kappa_{2,-}^{(i)})}{2} - \text{Tr} \kappa_2^{(i)} \right) = \frac{1}{2} (\kappa_{3,+}^{(i)} + \kappa_{3,-}^{(i)} - 2\kappa_3^{(i)}) [\text{Id}].$$

Next, by the product rule,

$$\nabla_{u_i} \left(\frac{m_i}{2} \text{Tr}(\kappa_{2,+}^{(i)} - \kappa_{2,-}^{(i)}) \right) = \frac{v_i}{2} \text{Tr}(\kappa_{2,+}^{(i)} - \kappa_{2,-}^{(i)}) d_i + \frac{m_i}{2} (\kappa_{3,+}^{(i)} - \kappa_{3,-}^{(i)}) [\text{Id}].$$

Finally,

$$\begin{aligned} \nabla_{u_i} (v_i (1 + \|d_i\|_2^2)) &= (1 + \|d_i\|_2^2) \nabla_{u_i} v_i + 2v_i (\nabla_{u_i} d_i)^\top d_i \\ &= -2m_i v_i (1 + \|d_i\|_2^2) d_i + v_i (\kappa_{2,+}^{(i)} - \kappa_{2,-}^{(i)}) d_i. \end{aligned}$$

From the above, we obtain:

$$\begin{aligned} (\mathbf{Cov}^{(i)} m^{(i)} - (\mathbf{Cov} m)_{-i}) &= \frac{1}{4} (\kappa_{3,+}^{(i)} + \kappa_{3,-}^{(i)} - 2\kappa_3^{(i)}) [\text{Id}] + \frac{m_i}{4} (\kappa_{3,+}^{(i)} - \kappa_{3,-}^{(i)}) [\text{Id}] \\ &\quad + \frac{v_i}{4} \text{Tr}(\kappa_{2,+}^{(i)} - \kappa_{2,-}^{(i)}) d_i - m_i v_i (1 + \|d_i\|_2^2) d_i + \frac{v_i}{2} (\kappa_{2,+}^{(i)} - \kappa_{2,-}^{(i)}) d_i. \end{aligned}$$

We bound the five terms on the right-hand side in order.

The first term. By Taylor's formula, we have:

$$\frac{1}{4} (\kappa_{3,+}^{(i)} + \kappa_{3,-}^{(i)} - 2\kappa_3^{(i)}) [\text{Id}] = \frac{\beta^2}{4} \int_0^1 (1-s) (\kappa_{5,s}^{(i)} + \kappa_{5,-s}^{(i)}) [\text{Id} \otimes \mathbf{M}_{i,-i}^{\otimes 2}] ds.$$

By [Lemma A.1](#), the above is at most $O(N^{1/4+\varepsilon})$ with probability $1 - O(N^{-a})$.

The second term. By Taylor's theorem:

$$\frac{m_i}{4} (\kappa_{3,+}^{(i)} - \kappa_{3,-}^{(i)}) [\text{Id}] = \frac{\beta m_i}{2} \kappa_4^{(i)} [\text{Id} \otimes \mathbf{M}_{i,-i}] + \frac{\beta^3 m_i}{8} \int_0^1 (1-s)^2 (\kappa_{6,s}^{(i)} + \kappa_{6,-s}^{(i)}) [\text{Id} \otimes \mathbf{M}_{i,-i}^{\otimes 3}] ds.$$

By [Lemma A.1](#), the second term is $O(N^\varepsilon)$ with probability $1 - O(N^{-a})$.

To control the first term, observe that we may write:

$$\kappa_4^{(i)} [\text{Id} \otimes \mathbf{M}_{i,-i}] = \kappa_4^{(i)} [\text{Id}] \mathbf{M}_{i,-i}.$$

Since $\kappa_4^{(i)} [\text{Id}]$ is also the second derivative of $\text{Tr} \mathbf{Cov}^{(i)}$, we see that it is equal to $-2(\mathbf{Cov}^{(i)})^2 - 2\kappa_3^{(i)} [m^{(i)}]$. First, notice that $\|(\mathbf{Cov}^{(i)})^2\|_F \leq \sqrt{N} \|\mathbf{Cov}^{(i)}\|_{\text{op}}^2$. By [Lemma A.2](#), we have $\mathbb{E} \left\| (\mathbf{Cov}^{(i)})^2 \right\|_F^p \leq c_p N^{p/2}$.

For every symmetric matrix B with $\|B\|_F = 1$, we have:

$$\begin{aligned} |\langle \kappa_3^{(i)} [m^{(i)}], B \rangle|^2 &\leq \left((m^{(i)})^\top \mathbf{Cov}^{(i)} m^{(i)} \right) \cdot \text{Var}_{\sigma \sim \mu_{-i}} \left[(\sigma - m^{(i)})^\top B (\sigma - m^{(i)}) \right] \\ &= \left((m^{(i)})^\top \mathbf{Cov}^{(i)} m^{(i)} \right) \left(\kappa_4^{(i)} [B \otimes B] + 2 \text{Tr}(\mathbf{Cov}^{(i)} B \mathbf{Cov}^{(i)} B) \right) \\ &\leq N \|\mathbf{Cov}^{(i)}\|_{\text{op}} \left(\|\kappa_4^{(i)}\|_F + 2 \|\mathbf{Cov}^{(i)}\|_{\text{op}}^2 \right), \end{aligned} \tag{19}$$

where both inequalities are by Cauchy-Schwarz, and the fact that $\|m_i\|^2 \leq N$. By choosing the extremal B and applying [Lemma A.2](#), we obtain

$$\mathbf{E} \left\| \kappa_3^{(i)} [m^{(i)}] \right\|_F^p = O(N^{7p/8}).$$

Also, $\|(\mathbf{Cov}^{(i)})^2\|_F \leq \sqrt{N} \|\mathbf{Cov}^{(i)}\|_{\text{op}}^2$. Consequently,

$$\mathbf{E} \left\| \kappa_4^{(i)} [\text{Id}] \right\|_F^p \leq O(N^{7p/8}).$$

Returning to the goal of obtaining a bound on $\left\| \kappa_4^{(i)} [\text{Id}] \mathbf{M}_{i,-i} \right\|$, we get:

$$\left\| \kappa_4^{(i)} [\text{Id}] \mathbf{M}_{i,-i} \right\|_2 \leq \left\| \kappa_4^{(i)} [\text{Id}] \mathbf{W}_{i,-i} \right\|_2 + \frac{1}{N} \left\| \kappa_4^{(i)} [\text{Id}] \mathbf{x}_{i,-i} \right\|_2$$

By Gaussian concentration, we have:

$$\mathbf{E} \left\| \kappa_4^{(i)} [\text{Id}] \mathbf{W}_{i,-i} \right\|_2^p \leq c_p N^{3p/8},$$

and by using $\|M\|_{\text{op}} \leq \|M\|_F$ for any matrix M combined with $\|\mathbf{x}_{i,-i}\| \leq \sqrt{N}$, we have:

$$\frac{1}{N^p} \mathbf{E} \left\| \kappa_4^{(i)} [\text{Id}] \mathbf{x}_{i,-i} \right\|_2^p \leq O(N^{3p/8}).$$

Putting the above together gives

$$\mathbf{E} \left\| \kappa_4^{(i)} [\text{Id}] \mathbf{M}_{i,-i} \right\|_2^p \leq O(N^{3p/8}).$$

The third, fourth, and fifth terms. We have:

$$\begin{aligned} \text{Tr}(\kappa_{2,+}^{(i)} - \kappa_{2,-}^{(i)}) &= \beta \int_{-1}^1 \kappa_{3,s}^{(i)} [\text{Id} \otimes \mathbf{M}_{i,-i}] ds, \\ \kappa_{2,+}^{(i)} - \kappa_{2,-}^{(i)} &= \beta \int_{-1}^1 \kappa_{3,s}^{(i)} [\mathbf{M}_{i,-i}] ds, \end{aligned}$$

where both quantities on the right-hand side have the p -th moment of their Frobenius norm bounded by $O(N^{p/4})$ by [Lemma A.1](#). Using $|m_i|, v_i \leq 1$, and $\mathbf{E} \|d_i\|_2^p \leq c_p$, we obtain:

$$\begin{aligned} \mathbf{E} \left\| v_i \cdot \text{Tr}(\kappa_{2,+}^{(i)} - \kappa_{2,-}^{(i)}) d_i \right\|_2^p &\leq O(N^{p/4}), \\ \mathbf{E} \left\| \frac{2m_i v_i}{N} (1 + \|d_i\|_2^2) d_i \right\|_2^p &\leq O(1), \\ \mathbf{E} \left\| v_i (\kappa_{2,+}^{(i)} - \kappa_{2,-}^{(i)}) d_i \right\|_2^p &= O(N^{p/4}). \end{aligned}$$

Final bound. Combining the five estimates along with Markov's inequality on the p -th moment for a constant p depending on a proves

$$\frac{2}{N} \left\| \mathbf{Cov}^{(i)} m^{(i)} - (\mathbf{Cov} m)_{-i} \right\|_2 = O\left(\frac{1}{\sqrt{N}}\right)$$

with probability $1 - O(N^{-a})$. Since $\beta^2 m_i \leq 1$, the desired bound follows.

Bound on $\left\| \frac{\beta^3}{3} \kappa_4^{(i)} [\mathbf{M}_{i,-i}^{\otimes 3}] \right\|_2$. We can write:

$$\begin{aligned} \kappa_4^{(i)} [\mathbf{M}_{i,-i}^{\otimes 3}] &= \frac{3}{N} \kappa_4^{(i)} [\text{Id}] \mathbf{M}_{i,-i} + \kappa_4^{(i)} [\mathbf{W}_{i,-i}^{\otimes 3}] - \frac{3}{N} \kappa_4^{(i)} [\text{Id}] \mathbf{W}_{i,-i} \\ &\quad + \frac{3\beta x_i}{N} \kappa_4^{(i)} [(\mathbf{W}_{i,-i}^{\otimes 2} - N^{-1} \text{Id}) \otimes \mathbf{x}_{-i}] + \frac{3\beta^2}{N^2} \kappa_4^{(i)} [\mathbf{W}_{i,-i} \otimes \mathbf{x}_{-i}^{\otimes 2}] + \frac{\beta^3}{N^3} \kappa_4^{(i)} [\mathbf{x}_{-i}^{\otimes 3}] \end{aligned}$$

Using [Lemma A.2](#), our bound on moments of $\kappa_4^{(i)} [\text{Id}] \mathbf{M}_{i,-i}$, and moment bounds for Gaussians, one may see that:

$$\mathbf{E} \left\| \kappa_4^{(i)} [\mathbf{M}_{i,-i}^{\otimes 3}] \right\|_2^p \leq O(N^{-5p/8}).$$

An application of Markov's inequality for an appropriate choice of p implies the desired bound.

Bound on $\left\| \beta^2 m_i \kappa_3^{(i)} [\mathbf{M}_{i,-i}^{\otimes 2} - \frac{1}{N} \text{Id}] \right\|_2$. First, observe that we may write:

$$\mathbf{M}_{i,-i}^{\otimes 2} - \frac{1}{N} \text{Id} = \left(\mathbf{W}_{i,-i}^{\otimes 2} - \frac{1}{N} \text{Id} \right) + \frac{\beta x_i}{N} (\mathbf{x}_{-i} \otimes \mathbf{W}_{i,-i} + \mathbf{W}_{i,-i} \otimes \mathbf{x}_{-i}) + \frac{\beta^2}{N^2} \mathbf{x}_{-i}^{\otimes 2}.$$

To analyze contracting $\kappa_3^{(i)}$ via each of the above terms, we find it convenient to define the following matrix V constructed by choosing

$$V_{a,b} := \left\langle \kappa_3^{(i)} [e_a], \kappa_3^{(i)} [e_b] \right\rangle.$$

By symmetry of the cumulant tensor, one can also see that $V = \sum_a \left(\kappa_3^{(i)} [e_a] \right)^2$. A key identity related to V that can be checked by direct computation is that for any vector u , $\left\| \kappa_3^{(i)} [u] \right\|_F = \sqrt{u^\top V u}$.

First term. A standard computation proves that $\left\| \kappa_3^{(i)} [\mathbf{W}_{i,-i}^{\otimes 2} - \frac{1}{N} \text{Id}] \right\|_F^2$ is equal to $\frac{2}{N^2} \text{Tr } V$. To get a handle on this random variable, we calculate moments of its fluctuation, i.e. moments of

$$\left\| \kappa_3^{(i)} [\mathbf{W}_{i,-i}^{\otimes 2} - \frac{1}{N} \text{Id}] \right\|_F^2 - \frac{2}{N^2} \text{Tr } V.$$

Since the above is a degree-4 polynomial of gaussian random variables, we may use hypercontractivity to bound the p -th moment in terms of its variance. We may compute its variance using Wick's formula and obtain:

$$\begin{aligned} &\mathbf{E} \left(\left\| \kappa_3^{(i)} [\mathbf{W}_{i,-i}^{\otimes 2} - \frac{1}{N} \text{Id}] \right\|_F^2 - \frac{2}{N^2} \text{Tr } V \right)^2 \\ &= \frac{1}{N^4} \left(8 \sum_{a,b} \left\langle \kappa_3^{(i)} [e_a], \kappa_3^{(i)} [e_b] \right\rangle^2 + 16 \sum_{a,b} \text{Tr} \left(\kappa_3^{(i)} [e_a] \kappa_3^{(i)} [e_b] \kappa_3^{(i)} [e_a] \kappa_3^{(i)} [e_b] \right) + 32 \text{Tr } V^2 \right) \\ &\leq \frac{56}{N^4} \text{Tr } V^2 \leq O \left(\frac{\text{Tr } V \cdot \|V\|_{\text{op}}}{N^4} \right). \end{aligned}$$

where in the first inequality we used that for any pair of symmetric matrices X, Y we have $\text{Tr } XYXY \leq \text{Tr } X^2 Y^2$. By a similar argument to the proof of (19) and [Lemma A.2](#) respectively, we see that:

$$\mathbf{E} \|V\|_{\text{op}}^p \leq O(N^{3p/4}), \quad \mathbf{E} (\text{Tr } V)^p \leq O(N^p).$$

By the above, we see that the p -th moment of the fluctuation for fixed constant p is $O(N^{-9p/8})$.

Observe that $\text{Tr } V = \left\| \kappa_3^{(i)} \right\|_F^2$, which is $O(N)$ with probability $1 - O(N^{-a})$ for any constant a , and so we have:

$$\left\| \kappa_3^{(i)} \left[\mathbf{W}_{i,-i}^{\otimes 2} - \frac{1}{N} \text{Id} \right] \right\|_F \leq O\left(\frac{1}{\sqrt{N}}\right).$$

Second and third terms. We now treat the second and third terms.

$$\begin{aligned} \mathbf{E}_{\mathbf{W}} \left\| \frac{\beta \mathbf{x}_i}{N} \kappa_3^{(i)} [\mathbf{x}_{-i} \otimes \mathbf{W}_{i,-i}] \right\|_F^p &\leq \frac{c_p}{N^{3p/2}} \left\| \kappa_3^{(i)} [\mathbf{x}_{-i}] \right\|_F^p = \frac{c_p}{N^{3p/2}} \left(\mathbf{x}_{-i}^\top V \mathbf{x}_{-i} \right)^{p/2} \leq \frac{c_p}{N^p} \|V\|_{\text{op}}^{p/2}. \\ \left\| \frac{\beta^2}{N^2} \kappa_3^{(i)} [\mathbf{x}_{-i}^{\otimes 2}] \right\|_F &= \frac{\beta^2}{N^2} \left\| \kappa_3^{(i)} [\mathbf{x}_{-i}] \mathbf{x}_{-i} \right\|_F \leq \frac{\beta^2}{N^{3/2}} \left\| \kappa_3^{(i)} [\mathbf{x}_{-i}] \right\|_F = \frac{\beta^2}{N^{3/2}} \sqrt{\mathbf{x}_{-i}^\top V \mathbf{x}_{-i}} \leq \frac{\beta^2}{N} \|V\|_{\text{op}}^{1/2}. \end{aligned}$$

The desired bound follows from aforementioned bounds on moments of $\|V\|_{\text{op}}$.

Bound on λ_i . We may express λ_i exactly as:

$$\lambda_i = \frac{\beta^2}{N} \left(\text{Tr } \mathbf{Cov} - \text{Tr } \mathbf{Cov}^{(i)} \right) - \beta^2 \left(\mathbf{M}_{i,-i}^\top \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i} - \frac{1}{N} \text{Tr } \mathbf{Cov}^{(i)} \right) - \beta \mathbf{M}_{i,-i}^\top (d_i - \beta \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i}).$$

We now bound each of the above three terms.

Observe that the first term is:

$$\begin{aligned} |\text{Tr } \mathbf{Cov} - \text{Tr } \mathbf{Cov}^{(i)}| &= \left| 1 - m_i^2 + \left\| m^{(i)} \right\|^2 - \left\| m_{-i} \right\|^2 \right| \\ &\leq 1 + \left| \left\langle m^{(i)} - m_{-i}, m^{(i)} + m_{-i} \right\rangle \right| \\ &\leq 1 + 2\sqrt{N} \left\| m^{(i)} - m_{-i} \right\|. \end{aligned}$$

We have that:

$$m_{-i} - m^{(i)} = \frac{1+m_i}{2} (m^{(i,+)} - m^{(i)}) + \frac{1-m_i}{2} (m^{(i,-)} - m^{(i)}),$$

By a Taylor expansion, we see:

$$m^{(i,\pm)} - m^{(i)} = \pm \beta \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i} + \beta^2 \int_0^1 (1-s) \kappa_{3,\pm}^{(i)} [\mathbf{M}_{i,-i}^{\otimes 2}] ds.$$

By [Lemma A.1](#), the p -th moment of the above for fixed p is $O(1)$. Thus,

$$\mathbf{E} \left(\frac{\beta^2}{N} \left(\text{Tr } \mathbf{Cov} - \text{Tr } \mathbf{Cov}^{(i)} \right) \right)^p \leq O(N^{-p/2}).$$

For the second term, observe that:

$$\begin{aligned} \mathbf{M}_{i,-i}^\top \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i} - \frac{1}{N} \text{Tr } \mathbf{Cov}^{(i)} &= \left(\mathbf{W}_{i,-i}^\top \mathbf{Cov}^{(i)} \mathbf{W}_{i,-i} - \frac{1}{N} \text{Tr } \mathbf{Cov}^{(i)} \right) + \frac{2\beta \mathbf{x}_i}{N} \mathbf{W}_{i,-i}^\top \mathbf{Cov}^{(i)} \mathbf{x}_{-i} + \frac{\beta^2}{N^2} \mathbf{x}_{-i}^\top \mathbf{Cov}^{(i)} \mathbf{x}_{-i}. \end{aligned}$$

By bounds on gaussian moments, we have:

$$\begin{aligned} \left(\mathbf{E}_W \left| \mathbf{W}_{i,-i}^\top \mathbf{Cov}^{(i)} \mathbf{W}_{i,-i} - \frac{1}{N} \text{Tr} \mathbf{Cov}^{(i)} \right|^p \right)^{1/p} &\leq \frac{C_p}{N} \|\mathbf{Cov}^{(i)}\|_F, \\ \left(\mathbf{E}_W \left| \frac{2\beta \mathbf{x}_i}{N} \mathbf{W}_{i,-i}^\top \mathbf{Cov}^{(i)} \mathbf{x}_{-i} \right|^p \right)^{1/p} &\leq \frac{C_p}{N^{3/2}} \|\mathbf{Cov}^{(i)} \mathbf{x}_{-i}\|_2 \leq \frac{C_p}{N} \|\mathbf{Cov}^{(i)}\|_F. \end{aligned}$$

By Cauchy–Schwarz, we have:

$$\frac{\beta^2}{N^2} |\mathbf{x}_{-i}^\top \mathbf{Cov}^{(i)} \mathbf{x}_{-i}| \leq \frac{\beta^2}{N} \|\mathbf{Cov}^{(i)}\|_F.$$

Thus, by [Lemma 4.9](#) we may conclude that:

$$\mathbf{E} \left| \beta^2 \left(\mathbf{M}_{i,-i}^\top \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i} - \frac{1}{N} \text{Tr} \mathbf{Cov}^{(i)} \right) \right|^p \leq C_p N^{-p/2}.$$

For the third term, first observe that by performing a Taylor expansion we get:

$$\beta \mathbf{M}_{i,-i}^\top (d_i - \beta \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i}) = \frac{\beta^4}{4} \int_0^1 (1-s)^2 (\kappa_{4,s}^{(i)} + \kappa_{4,-s}^{(i)}) [\mathbf{M}_{i,-i}^{\otimes 4}] ds.$$

By [Lemma A.1](#), we get:

$$\mathbf{E} \left| \beta \mathbf{M}_{i,-i}^\top (d_i - \beta \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i}) \right|^p \leq O(N^{-p}). \quad (20)$$

Combining the estimates for the three terms gives:

$$\mathbf{E} |\lambda_i|^p \leq O(N^{-p/2}),$$

and Markov's inequality along with a union bound for appropriate choice of p proves the desired statement.

A.3 Proof of [Lemma A.1](#)

We will prove the statement for $\kappa_{r,s}^{(i)}$. The analogous statement for $\kappa_{r,s}^{(B)}$ follows an identical proof.

Fix $s \in [-1, 1]$. Recall that the goal is to prove:

$$\mathbf{E} \left\| \kappa_{r,s}^{(i)} [\text{Id}^{\otimes k} \otimes \mathbf{M}_{i,-i}^{\otimes \ell}] \right\|_F^p \leq C_p N^{p(r/4 - \ell/2)}.$$

First, observe that we may write:

$$\left\| \kappa_{r,s}^{(i)} [\text{Id}^{\otimes k} \otimes \mathbf{M}_{i,-i}^{\otimes \ell}] \right\|_F \leq C_\ell \sum_{\ell_1 + \ell_2 = \ell} \left\| \kappa_{r,s}^{(i)} \left[\text{Id}^{\otimes k} \otimes \mathbf{W}_{i,-i}^{\otimes \ell_1} \otimes \frac{\mathbf{x}_{-i}^{\otimes \ell_2}}{N^{\ell_2}} \right] \right\|_F$$

for some C_ℓ that only depends on ℓ . To obtain a handle on the above (up to a loss of a constant depending on p and ℓ), it suffices to study a single summand.

An ingredient in obtaining a bound on a single summand is the following bound in the special case when $\ell_1 = 0$, whose proof we defer until the end of this proof.

Lemma A.3. For every fixed $r \geq 3$, $p \geq 2$, and nonnegative integers k and ℓ such that $2k + \ell \leq r$, we have:

$$\left(\mathbf{E} \left\| \kappa_{r,s}^{(i)} \left[\text{Id}^{\otimes k} \otimes \mathbf{x}_{-i}^{\otimes \ell} \right] \right\|_F^p \right)^{1/p} \leq C N^{(r+\ell)/4}.$$

In particular, we perform a proof by induction, where the above is the base case. Henceforth, for the induction hypothesis, assume that there is some choice of $\ell_1 \geq 0$ for which the desired statement holds.

In the rest of this argument ℓ_2 is fixed, and we denote the cumulant tensor of interest by

$$\kappa_r(k, \ell_1) := \kappa_{r,s}^{(i)} \left[\text{Id}^{\otimes k} \otimes \mathbf{W}_{i,-i}^{\otimes \ell_1} \otimes \frac{\mathbf{x}_{-i}^{\otimes \ell_2}}{N^{\ell_2}} \right].$$

We split the term of interest into a sum of its mean and a fluctuation term, and control each term separately. In particular, we have:

$$\kappa_r(k, \ell_1) = \mathbf{E}_{\mathbf{W}_{i,-i}} \kappa_r(k, \ell_1) + (\kappa_r(k, \ell_1) - \mathbf{E}_{\mathbf{W}_{i,-i}} \kappa_r(k, \ell_1)).$$

Controlling $\mathbf{E} \kappa_r(k, \ell_1)$. Observe that by gaussian integration by parts, we have:

$$\mathbf{E} \kappa_r(k, \ell_1) = \frac{\ell_1 - 1}{N} \mathbf{E} \kappa_r(k+1, \ell_1 - 2) + \frac{s\beta}{N} \mathbf{E} \kappa_{r+1}(k+1, \ell_1 - 1),$$

following the convention that $\kappa_r(k, -1) = 0$ for any r, k . By treating [Lemma A.3](#) as a base case for an induction argument, we obtain:

$$\mathbf{E}_{\mathbf{W}_{i,-i}, \mathbf{x}} \left\| \mathbf{E}_{\mathbf{W}_{i,-i}} \kappa_{r,s}^{(i)} \left[\text{Id}^{\otimes k} \otimes \mathbf{W}_{i,-i}^{\otimes \ell_1} \otimes \frac{\mathbf{x}_{-i}^{\otimes \ell_2}}{N^{\ell_2}} \right] \right\|_F^p \leq C_p N^{p(r/4 - 3\ell_2/4 - \ell_1/2)}. \quad (21)$$

It remains to control the fluctuation term.

Controlling $\kappa_r(k, \ell_1) - \mathbf{E} \kappa_r(k, \ell_1)$. To control the p -th moment of the fluctuation, we treat it as a function of $\mathbf{W}_{i,-i}$, and apply the L^p Gaussian Poincaré inequality, which says that for any smooth function $F : \mathbb{R}^m \rightarrow \mathbb{R}^n$, we have:

$$\mathbf{E}_g \|F(g) - \mathbf{E} F(g)\|_F^p \leq C_p \mathbf{E}_g \|\nabla F(g)\|_F^p.$$

As a result, we have:

$$\mathbf{E}_{\mathbf{W}_{i,-i}} \|\kappa_r(k, \ell_1) - \mathbf{E} \kappa_r(k, \ell_1)\|_F^p \leq \frac{C_p}{\sqrt{N}^p} \mathbf{E}_{\mathbf{W}_{i,-i}} \|\nabla \kappa_r(k, \ell_1)\|_F^p. \quad (22)$$

Now, we may compute the gradient as:

$$\nabla \kappa_r(k, \ell_1) = \ell_1 \kappa_r(k, \ell_1 - 1) + s\beta \kappa_{r+1}(k, \ell_1) \quad (23)$$

By [Eq. \(21\)](#) to [\(23\)](#), along with the induction hypothesis, we have:

$$\mathbf{E} \|\kappa_r(k, \ell_1) - \mathbf{E} \kappa_r(k, \ell_1)\|_F^p \leq C_p N^{p(r/4 - 3\ell_2/4 - \ell_1/2)} + \frac{C_p}{\sqrt{N}^p} \mathbf{E} \|\kappa_{r+1}(k, \ell_1)\|_F^p$$

Iterating the above inequality $2\ell_1$ times, we obtain:

$$\mathbf{E} \|\kappa_r(k, \ell_1) - \mathbf{E}\kappa_r(k, \ell_1)\|_F^p \leq C_p N^{p(r/4 - 3\ell_2/4 - \ell_1/2)} + \frac{C_p}{N^{p\ell_1}} \mathbf{E} \|\kappa_{r+2\ell_1}(k, \ell_1)\|_F^p$$

By Cauchy-Schwarz, we may bound the second term as:

$$\mathbf{E} \|\kappa_{r+2\ell_1}(k, \ell_1)\|_F^p \leq \mathbf{E} \left\| \kappa_{r+2\ell_1} \left[\text{Id}^{\otimes k} \otimes \frac{\mathbf{x}_{-i}^{\otimes \ell_2}}{N^{\ell_2}} \right] \right\|_F^p \cdot \|\mathbf{W}_{i,-i}\|_F^{\ell_1 p}.$$

By another application of Cauchy-Schwarz, [Lemma A.3](#), and Gaussian concentration, we may bound the above by $C_p N^{p(r/4 - \ell_1/2 - 3\ell_2/4)}$, which completes the proof.

We are now ready to prove [Lemma A.3](#).

Lemma A.3. *For every fixed $r \geq 3$, $p \geq 2$, and nonnegative integers k and ℓ such that $2k + \ell \leq r$, we have:*

$$\left(\mathbf{E} \left\| \kappa_{r,s}^{(i)} \left[\text{Id}^{\otimes k} \otimes \mathbf{x}_{-i}^{\otimes \ell} \right] \right\|_F^p \right)^{1/p} \leq C N^{(r+\ell)/4}.$$

Proof. For simplicity, assume that $k + \ell > 0$; the same ideas extend to the setting where $k = \ell = 0$. We write $Y := \kappa_{r,s}^{(i)}[\text{Id}^{\otimes k} \otimes \mathbf{x}_{-i}^{\otimes \ell}]$. For ease of notation, set $d = r - 2k - \ell$. We drop the cavity and shift notation, denoting by F the corresponding free energy and by $\mathbf{x}$ the cavity vector $\mathbf{x}_{-i}$. Denote $n = N - 1$. We now use the definition of κ_r as an order- r gradient of F in order to simplify the expression for Y . We have, for example, that $(\kappa_3[\text{Id}])_a = \sum_b \partial_{h_a} \partial_{h_b}^2 F = \partial_{h_a} \Delta F$, so $\kappa_3[\text{Id}] = \nabla \Delta F$. Similarly, contracting by $\mathbf{x}$ can be expressed in terms of the gradient—for example, $(\kappa_3[\mathbf{x}])_{ab} = \sum_c \partial_{h_a} \partial_{h_b} \cdot \mathbf{x}_c \partial_{h_c} F$, so $\kappa_3[\mathbf{x}] = \nabla^2(\mathbf{x} \cdot \nabla)F$. More generally, we have

$$Y = \nabla^{r-2k-\ell} \Delta^k (\mathbf{x} \cdot \nabla)^\ell F.$$

We will start by establishing that the k Laplacians and ℓ inner gradients will eventually result in a homogenous degree- $k + \ell$ polynomial P in centered overlaps, with $Y = N^{k+\ell} \nabla^d \langle P \rangle$.

Suppose that $k + \ell > 0$. Let us start by understanding how a Laplacian/gradient affects the free energy. For starters, we have

$$\begin{aligned} \Delta F &= \text{Tr } \mathbf{Cov} = n(1 - \langle R_{12} \rangle) = \text{constant} - n \langle \bar{R}_{12} \rangle, \\ (\mathbf{x} \cdot \nabla) F &= \langle \mathbf{x}, m \rangle = n \langle R_{1\star} \rangle = \text{constant} + n \langle R_{1\star} \rangle. \end{aligned}$$

Because $r \geq 3$, the constant terms will ultimately vanish. Suppose that after m contraction operations, we are at some polynomial $n^m \langle P_m \rangle$. A routine calculation shows that taking the Laplacian/gradient results in the Gibbs average of a polynomial of degree larger by 1. We execute this calculation for the gradient; the calculation for the Laplacian is similar. Suppose that P_m is some degree- m polynomial in centered overlaps among v replicas (and the plant).

$$\begin{aligned} (\mathbf{x} \cdot \nabla) \langle P_m(\sigma^1, \dots, \sigma^v) \rangle &= \left\langle \mathbf{x}, \left\langle P_m \cdot \sum_{a=1}^v (\sigma^j - \sigma^{v+1}) \right\rangle \right\rangle \\ &= n \cdot \left\langle P_m \cdot \sum_{a=1}^v (R_{j\star} - R_{(v+1)\star}) \right\rangle \end{aligned}$$

$$= n \cdot \left\langle P_m \cdot \sum_{a=1}^v \left(\bar{R}_{j^*} - \bar{R}_{(v+1)^*} \right) \right\rangle.$$

This is the Gibbs average of a homogenous degree- $(m+1)$ polynomial in centered overlaps! It follows that when $k + \ell > 0$,

$$Y = n^{k+\ell} \nabla^{r-2k-\ell} \langle P \rangle$$

for some polynomial P in centered overlaps. We may now use the same rules to expand out the gradient:

$$\left(\nabla^d \langle P(\sigma^1, \dots, \sigma^v) \rangle \right) = \left\langle P(\sigma^1, \dots, \sigma^v) \cdot \bigotimes_{a=v}^{v+d-1} \left(\sum_{b=1}^a (\sigma_b - \sigma_{a+1}) \right) \right\rangle,$$

and thus, using the replica trick to write $\langle G(\sigma) \rangle^2 = \langle G(\sigma) G(\tilde{\sigma}) \rangle$ for a fresh replica $\tilde{\sigma}$, we have

$$\begin{aligned} \|Y\|_F^2 &= n^{2(k+\ell)} \sum_{i_1, \dots, i_d} \left\langle P(\sigma^1, \dots, \sigma^v) \cdot \prod_{a=v}^{v+d-1} \left(\sum_{b=1}^a (\sigma_b - \sigma_{a+1}) \right) \right\rangle_{i_{a-v+1}}^2 \\ &= n^{2(k+\ell)} \sum_{i_1, \dots, i_d} \left\langle P(\sigma^1, \dots, \sigma^v) \cdot P(\tilde{\sigma}^1, \dots, \tilde{\sigma}^v) \cdot \prod_{a=v}^{v+d-1} \left(\sum_{b=1}^a (\sigma_b - \sigma_{a+1}) \right) \left(\sum_{b=1}^a (\tilde{\sigma}_b - \tilde{\sigma}_{a+1}) \right) \right\rangle_{i_{a-v+1}} \\ &= n^{2(k+\ell)} \left\langle P(\sigma^1, \dots, \sigma^v) \cdot P(\tilde{\sigma}^1, \dots, \tilde{\sigma}^v) \cdot \prod_{a=v}^{v+d-1} \left\langle \sum_{b=1}^a (\sigma_b - \sigma_{a+1}), \sum_{b=1}^a (\tilde{\sigma}_b - \tilde{\sigma}_{a+1}) \right\rangle \right\rangle. \end{aligned}$$

Note now that this final product term may be expressed as a degree- d product of *centered* replicas—this again uses the fact that the sum of the coefficients is 0. Summarizing, we have

$$\|Y\|_F^2 = n^{2(k+\ell)+d} \cdot \langle P \rangle = n^{r+\ell} \langle P \rangle.$$

for some homogenous degree- $2(k+\ell) + d$ polynomial P in centered overlaps. Noting $2k + \ell + d = r$ and using replicas again, we have

$$\|Y\|_F^{2p} = n^{p(r+\ell)} \cdot \langle Q \rangle$$

for some homogenous degree- $p(r+\ell)$ polynomial Q . To complete the proof, we simply use Hölder's inequality on each term of the polynomial to reduce bounding it to bounding large moments of centered replicas—this is the content of [Lemma 4.8](#), which yields that $\mathbf{E} \langle Q \rangle \leq C n^{-p(r+\ell)/2}$, and thus that $\mathbf{E} \|Y\|_F^{2p} \leq C N^{p(r+\ell)/2}$. $\square$

A.4 Proof of [Lemma A.2](#)

Throughout this subsection, $B \subseteq [N]$ is a set with $|B| \leq 4$; no intuition is lost by considering $B = \emptyset$. Define the TAP map $T^{(B)} : (-1, 1)^{[N] \setminus B} \rightarrow \mathbb{R}^{[N] \setminus B}$ by

$$T_i^{(B)}(m) = \operatorname{arctanh}(m_i) - \beta \mathbf{M}_{i, -(B \cup \{i\})}^\top m_{-(B \cup \{i\})} + \frac{\beta^2}{N} (N - |B| - \|m\|_2^2) m_i.$$

Let $F^{(B)}(u)$ be the free energy on spins $[N] \setminus B$ with interaction $\beta \mathbf{M}_{-B, -B}$ and external field $u \in \mathbb{R}^{[N] \setminus B}$. For $q \geq 1$, let $\kappa_q^{(B)}(u) = \nabla^q F^{(B)}(u)$; in particular $\kappa_1^{(B)}(u) = m^{(B)}(u)$ and $\kappa_2^{(B)}(u) = \mathbf{Cov}^{(B)}(u)$ are the mean and covariance of this Gibbs measure with external field u .

Recall that the planted model in [Theorem 1.4](#) has external field $h = tx + \sqrt{t}g$. Whenever we omit the argument u by writing e.g. $\mathbf{Cov}^{(B)}$, we evaluate at $u = h_{-B}$.

Throughout we use ∇ to denote derivatives with respect to u , while D denotes derivatives with respect to the input m of $T^{(B)}$, evaluated at $m = m^{(B)}(h_{-B})$. In particular, for $K^{(B)}$ defined in (8), we have $DT^{(B)}(m) = K^{(B)}$.

The TAP heuristic predicts that $T^{(B)}(m^{(B)}(u)) \approx u$, which motivates the definition of the residual

$$r^{(B)}(u) = T^{(B)}(m^{(B)}(u)) - u. \quad (24)$$

The starting point of the proof of [Lemma A.2](#) is the following set of identities, which are proved directly by differentiating [Eq. \(24\)](#) with respect to u .

Fact A.4. *We have*

$$\begin{aligned} K^{(B)}\mathbf{Cov}^{(B)} - \text{Id} &= \nabla r^{(B)}, \\ K^{(B)}\kappa_3^{(B)} &= \nabla^2 r^{(B)} - D^2 T^{(B)}(m)\{\mathbf{Cov}^{(B)}, \mathbf{Cov}^{(B)}\}, \\ K^{(B)}\kappa_4^{(B)} &= \nabla^3 r^{(B)} - \sum_{\text{Sym}} D^2 T^{(B)}(m)\{\kappa_3^{(B)}, \mathbf{Cov}^{(B)}\} - D^3 T^{(B)}(m)\{\mathbf{Cov}^{(B)}, \mathbf{Cov}^{(B)}, \mathbf{Cov}^{(B)}\}. \end{aligned}$$

Remark A.5. To clarify the notation, for $q = 2, 3, 4$, $\kappa_q^{(B)}$ is a tensor of order q , and $K^{(B)}$ acts by matrix multiplication on the first coordinate of $\kappa_q^{(B)}$:

$$(K^{(B)}\kappa_q^{(B)})_{ij_1 \dots j_{q-1}} = \sum_{a \notin B} K_{ia}^{(B)} (\kappa_q^{(B)})_{aj_1 \dots j_{q-1}}.$$

For $q = 1, 2, 3$, $\nabla^q r^{(B)}$ is a tensor of order $q + 1$, where the first coordinate indexes the entries of $r^{(B)}$ while the remaining q coordinates index the q derivatives. Similarly, $D^q T^{(B)}(m)$ is a tensor of order $q + 1$ where the first coordinate indexes the entries of $T^{(B)}$. In the notation $D^2 T^{(B)}(m)\{\mathbf{Cov}^{(B)}, \mathbf{Cov}^{(B)}\}$, the matrices $\mathbf{Cov}^{(B)}$ act by matrix multiplication on all but the first coordinate of $D^2 T^{(B)}(m)$:

$$\left(D^2 T^{(B)}(m)\{\mathbf{Cov}^{(B)}, \mathbf{Cov}^{(B)}\} \right)_{ijk} = \sum_{a, b \notin B} \frac{\partial^2 T_i^{(B)}}{\partial z_a \partial z_b}(m) \mathbf{Cov}_{aj}^{(B)} \mathbf{Cov}_{bk}^{(B)},$$

We define $D^3 T^{(B)}(m)\{\mathbf{Cov}^{(B)}, \mathbf{Cov}^{(B)}, \mathbf{Cov}^{(B)}\}$ similarly. In $D^2 T^{(B)}(m)\{\kappa_3^{(B)}, \mathbf{Cov}^{(B)}\}$, $\kappa_3^{(B)}$ and $\mathbf{Cov}^{(B)}$ contract with all but the first coordinates of $D^2 T^{(B)}(m)$:

$$\left(D^2 T^{(B)}(m)\{\kappa_3^{(B)}, \mathbf{Cov}^{(B)}\} \right)_{ijkl} = \sum_{a, b \notin B} \frac{\partial^2 T_i^{(B)}}{\partial z_a \partial z_b}(m) (\kappa_3^{(B)})_{ajk} \mathbf{Cov}_{bl}^{(B)}.$$

Finally, $\sum_{\text{Sym}}$ sums cyclically over the last three coordinates of the resulting 4-tensor:

$$\left(\sum_{\text{Sym}} D^2 T^{(B)}(m)\{\kappa_3^{(B)}, \mathbf{Cov}^{(B)}\} \right)_{ijkl} = \sum_{abc \in \{jkl, k\ell j, \ell jk\}} \left(D^2 T^{(B)}(m)\{\kappa_3^{(B)}, \mathbf{Cov}^{(B)}\} \right)_{iabc}$$

The main step in the proof of [Lemma A.2](#) is to obtain suitably sharp L^p bounds on the Frobenius norms of all the right-hand sides in [Fact A.4](#). In particular, for $|B| \leq 3$, this can be translated into L^p bounds on $\|\kappa_3^{(B)}\|_F$, $\|\kappa_4^{(B)}\|_F$, and $\|\mathbf{Cov}^{(B)}\|_{\text{op}}$: on the event $K^{(B)} \succeq \gamma$, which holds with overwhelming probability by [Lemma 4.2](#), these are bounded by $\gamma^{-1}\|K^{(B)}\kappa_3^{(B)}\|_F$, $\gamma^{-1}\|K^{(B)}\kappa_4^{(B)}\|_F$, and $\gamma^{-1}\|K^{(B)}\mathbf{Cov}^{(B)}\|_{\text{op}}$, while the L^p contribution from the event $K^{(B)} \not\succeq \gamma$ can be controlled by crude polynomial bounds on $\|\kappa_3^{(B)}\|_F$, $\|\kappa_4^{(B)}\|_F$, and $\|\mathbf{Cov}^{(B)}\|_{\text{op}}$.

For $|B| \leq 3$ and $i \in [N] \setminus B$, recall from [Section 4](#) that $v_i^{(B)} = 1 - (m_i^{(B)})^2$, $q_N^{(B)} = \|m^{(B)}\|^2/N$, and

$$d_i^{(B)} = \frac{\mathbf{Cov}_{i,-(B \cup \{i\})}^{(B)}}{v_i^{(B)}} = \frac{m^{(B,i,+)} - m^{(B,i,-)}}{2}, \quad \lambda_i^{(B)} = \beta^2 \left(1 - \frac{|B|}{N} - q_N^{(B)} \right) - \beta \mathbf{M}_{i,-(B \cup \{i\})}^\top d_i^{(B)}.$$

Introduce the notation

$$\phi_i^{(B)}(s) = F^{(B \cup \{i\})}(u^{(B \cup \{i\})} + s\beta \mathbf{M}_{i,-(B \cup \{i\})}). \quad (25)$$

Also for $s \in \{\pm 1\}$, define

$$m^{(B,i,s)} = \nabla F^{(B \cup \{i\})}(u^{(B \cup \{i\})} + s\beta \mathbf{M}_{i,-(B \cup \{i\})}),$$

which is the mean of the Gibbs measure on spins $[N] \setminus (B \cup \{i\})$, where spins $\{\sigma_j : j \in B\}$ are removed and σ_i is set to s . The following lemma provides a description of $r^{(B)}$, which will be useful for controlling its derivatives appearing in [Fact A.4](#).

Lemma A.6. *For $|B| \leq 3$ and $i \in [N] \setminus B$, we have $r_i^{(B)} = \zeta_i^{(B)} + m_i^{(B)}\lambda_i^{(B)}$, where*

$$\zeta_i^{(B)} = \frac{\phi_i^{(B)}(1) - \phi_i^{(B)}(-1) - (\phi_i^{(B)})'(1) - (\phi_i^{(B)})'(-1)}{2} = -\frac{1}{4} \int_{-1}^1 (1-s^2)(\phi_i^{(B)})^{(3)}(s) ds \quad (26)$$

Proof. We provide the proof for $B = \emptyset$ and omit the superscript $(\emptyset)$; the case of general B is analogous. For $s \in \{\pm 1\}$, the contribution $Z_i(s)$ to the partition function from $\sigma_i = s$ is $\exp(h_i s + \phi_i(s))$. Thus,

$$\text{arctanh}(m_i) - h_i = \frac{1}{2} \log \frac{1+m_i}{1-m_i} - h_i = \frac{1}{2} \log \frac{Z_i(+1)}{Z_i(-1)} - h_i = \frac{\phi_i(+1) - \phi_i(-1)}{2}. \quad (27)$$

Differentiating [Eq. \(25\)](#) yields, for $s \in \{\pm 1\}$,

$$\phi_i'(s) = \beta \mathbf{M}_{i,-i}^\top m^{(i,s)}.$$

Combining with $m_{-i} = \frac{1+m_i}{2}m^{(i,+)} + \frac{1-m_i}{2}m^{(i,-)}$ yields

$$\beta \mathbf{M}_{i,-i}^\top m_{-i} = \frac{1+m_i}{2}\phi_i'(1) + \frac{1-m_i}{2}\phi_i'(-1) = \frac{\phi_i'(1) + \phi_i'(-1)}{2} + m_i \cdot \beta \mathbf{M}_{i,-i}^\top d_i.$$

Substituting the last display and [Eq. \(27\)](#) into $r_i = T_i(m) - h_i$ yields $r_i = \zeta_i + m_i\lambda_i$. The second equality in [Eq. \(26\)](#) follows from integration by parts. $\square$

Using [Lemma A.6](#), the right-hand sides in [Fact A.4](#) can be expanded as a combination of $d_i^{(B)}$ and derivatives in $m_i^{(B)}, \zeta_i^{(B)}, \lambda_i^{(B)}$. In the following lemma, we collect the L^p estimates we need for these quantities. The proof is largely by Taylor expanding these quantities in terms of contractions of the $\kappa_{r,s}^{(i)}$ and applying [Lemma A.1](#); we defer the proof to the end of this subsection.

Lemma A.7. *For every fixed $p \geq 2$, there exists C_p such that for all $|B| \leq 3$ and $i \in [N] \setminus B$,*

$$\mathbf{E} \|d_i^{(B)} - \beta \mathbf{Cov}^{(B \cup \{i\})} M_{i, -(B \cup \{i\})}\|_2^p \leq C_p N^{-p/2},$$

and for $j = 1, 2, 3$,

$$\begin{aligned} \mathbf{E} \|(v_i^{(B)})^{-1} \nabla^j m_i^{(B)}\|_F^p &\leq C_p N^{p(j-1)/4}, \\ \mathbf{E} \|\nabla^j \zeta_i^{(B)}\|_F^p &\leq C_p N^{p(j-3)/4}. \end{aligned}$$

Finally, for $j = 0, 1, 2, 3$,

$$\mathbf{E} \|\nabla^j \lambda_i^{(B)}\|_F^p \leq C_p \left(N^{p(j-3)/4} + N^{-p} \mathbf{E} \|\kappa_{j+2}^{(B \cup \{i\})}\|_F^p \right).$$

and in particular $\mathbf{E} |\lambda_i^{(B)}|^p \leq C_p N^{-p/2}$.

Corollary A.8. *For all $p \geq 2$, there exists $C_p > 0$ such that for all $|B| \leq 3$,*

$$\mathbf{E} \|\nabla^2 r^{(B)}\|_F^p \leq C_p N^{p/2}, \quad \mathbf{E} \|\nabla^3 r^{(B)}\|_F^p \leq C_p N^{3p/4}.$$

Proof. By [Lemma 4.9](#), for each $j = 1, 2, 3$,

$$\mathbf{E} \|\kappa_{j+2}^{(B \cup \{i\})}\|_F^p \leq C_p N^{p(j+2)/4},$$

and so the estimate on $\nabla^j \lambda_i^{(B)}$ in [Lemma A.7](#) gives

$$\mathbf{E} \|\nabla^j \lambda_i^{(B)}\|_F^p \leq C_p N^{p(j-2)/4}$$

after adjusting C_p . Recall from [Lemma A.6](#) that $r_i^{(B)} = \zeta_i^{(B)} + m_i^{(B)} \lambda_i^{(B)}$. The result follows by differentiating this identity and applying the last display and the remaining estimates from [Lemma A.7](#). $\square$

The following lemma directly implies [Lemma A.2\(a\)](#), which corresponds to the case $|B| \leq 1$; we allow more general B for use in the rest of the proof of [Lemma A.2](#). It also provides stronger bounds on $\|\kappa_3^{(B)}\|_F$ and $\|\kappa_4^{(B)}\|_F$ than obtained in [Lemma 4.9](#), which strengthens the estimate on $\nabla^j \lambda_i^{(B)}$ in [Lemma A.7](#); this will be useful in [Corollary A.10](#) below.

Lemma A.9. *For all $p \geq 2$, there exists $C_p > 0$ such that for all $|B| \leq 3$,*

$$\mathbf{E} \|\kappa_3^{(B)}\|_F^p \leq C_p N^{p/2}, \quad \mathbf{E} \|\kappa_4^{(B)}\|_F^p \leq C_p N^{3p/4}.$$

In particular, [Lemma A.2\(a\)](#) follows as the case $|B| \leq 1$.

Proof. We provide the proof for $B = \emptyset$ and omit the superscript $(\emptyset)$; the case of general B is analogous. Most of the proof will be devoted to the estimates

$$\mathbf{E}\|D^2T(m)\{\mathbf{Cov}, \mathbf{Cov}\}\|_F^p \leq C_p N^{p/2}, \quad (28)$$

$$\mathbf{E}\|D^2T(m)\{\kappa_3, \mathbf{Cov}\}\|_F^p \leq C_p N^{3p/4}, \quad (29)$$

$$\mathbf{E}\|D^3T(m)\{\mathbf{Cov}, \mathbf{Cov}, \mathbf{Cov}\}\|_F^p \leq C_p N^{p/2}. \quad (30)$$

We first explain why they imply the conclusion. Together with [Corollary A.8](#), these imply

$$\mathbf{E}\|K\kappa_3\|_F^p \leq C_p N^{p/2}, \quad \mathbf{E}\|K\kappa_4\|_F^p \leq C_p N^{3p/4}.$$

Let $\mathcal{E} = \{K \succeq \gamma \text{Id}\}$, where $\gamma > 0$ is given by [Lemma 4.2](#). On $\mathcal{E}$, for each $q \in \{3, 4\}$,

$$\|\kappa_q\|_F \leq \|K^{-1}\|_{\text{op}} \|K\kappa_q\|_F \leq \gamma^{-1} \|K\kappa_q\|_F.$$

while on the complement of $\mathcal{E}$, we can use the crude bound $\|\kappa_q\|_F \leq CN^{q/2}$. By the above and [Lemma 4.2](#),

$$\mathbf{E}\|\kappa_3\|_F^p \leq \gamma^{-p} \mathbf{E}\|K\kappa_3\|_F^p + \mathbf{Pr}[\mathcal{E}^c] C_p N^{3p/2} \leq C_p N^{p/2},$$

and similarly $\mathbf{E}\|\kappa_4\|_F^p \leq C_p N^{3p/4}$. We now turn to the proofs of [Eq. \(28\)](#) to [\(30\)](#). Write

$$A(z) = \text{arctanh}(z), \quad U(z) = -\frac{\beta^2}{N} \|z\|^2 z, \quad (31)$$

where arctanh is applied entry-wise. Note that $T - A - U$ is linear in m . It therefore suffices to establish [Eq. \(28\)](#) to [\(30\)](#) with A and U in place of T .

We first treat A . Set

$$y_i = v_i^{-1} \nabla m_i, \quad z_i = v_i^{-1} \nabla^2 m_i.$$

Since $\nabla m = \mathbf{Cov}$ and $\nabla^2 m = \kappa_3$, direct differentiation gives

$$\begin{aligned} (D^2 A(m)\{\mathbf{Cov}, \mathbf{Cov}\})_i &= 2m_i y_i^{\otimes 2}, \\ (D^2 A(m)\{\kappa_3, \mathbf{Cov}\})_i &= 2m_i z_i \otimes y_i, \\ (D^3 A(m)\{\mathbf{Cov}, \mathbf{Cov}, \mathbf{Cov}\})_i &= 2(1 + 3m_i^2) y_i^{\otimes 3}. \end{aligned}$$

By [Lemma A.7](#) with $j = 1, 2$, for every fixed $p \geq 2$,

$$\mathbf{E}\|y_i\|_2^p \leq C_p, \quad \mathbf{E}\|z_i\|_F^p \leq C_p N^{p/4}.$$

Together with the trivial estimate $|m_i| \leq 1$ and Hölder's inequality, we obtain for all i

$$\begin{aligned} \mathbf{E}\|(D^2 A(m)\{\mathbf{Cov}, \mathbf{Cov}\})_i\|_F^p &\leq C_p, \\ \mathbf{E}\|(D^2 A(m)\{\kappa_3, \mathbf{Cov}\})_i\|_F^p &\leq C_p N^{p/4}, \\ \mathbf{E}\|(D^3 A(m)\{\mathbf{Cov}, \mathbf{Cov}, \mathbf{Cov}\})_i\|_F^p &\leq C_p. \end{aligned}$$

For any tensor $H \in (\mathbb{R}^N)^{\otimes q}$ with slices $H_i \in (\mathbb{R}^N)^{\otimes(q-1)}$ along the first index, the inequality

$$\mathbf{E}\|H\|_F^p = \mathbf{E}\left(\sum_i \|H_i\|_F^2\right)^{p/2} \leq N^{p/2-1} \sum_i \mathbf{E}\|H_i\|_F^p$$

therefore yields [Eq. \(28\)](#) to [\(30\)](#) with A in place of T . We next treat U . Direct differentiation gives

$$\begin{aligned} (D^2 U(m)\{\mathbf{Cov}, \mathbf{Cov}\})_{ijk} &= -\frac{2\beta^2}{N} \left(m_i (\mathbf{Cov}^2)_{jk} + \mathbf{Cov}_{ij} (\mathbf{Cov} m)_k + \mathbf{Cov}_{ik} (\mathbf{Cov} m)_j \right), \\ (D^2 U(m)\{\kappa_3, \mathbf{Cov}\})_{ijk\ell} &= -\frac{2\beta^2}{N} \left(m_i \sum_a (\kappa_3)_{ajk} \mathbf{Cov}_{a\ell} + (\kappa_3)_{ijk} (\mathbf{Cov} m)_\ell + \mathbf{Cov}_{i\ell} \sum_a m_a (\kappa_3)_{ajk} \right), \\ (D^3 U(m)\{\mathbf{Cov}, \mathbf{Cov}, \mathbf{Cov}\})_{ijk\ell} &= -\frac{2\beta^2}{N} \left(\mathbf{Cov}_{ij} (\mathbf{Cov}^2)_{k\ell} + \mathbf{Cov}_{ik} (\mathbf{Cov}^2)_{j\ell} + \mathbf{Cov}_{i\ell} (\mathbf{Cov}^2)_{jk} \right). \end{aligned} \quad (32)$$

The triangle inequality and Cauchy–Schwarz imply

$$\begin{aligned} \|D^2 U(m)\{\mathbf{Cov}, \mathbf{Cov}\}\|_F &\leq \frac{6\beta^2}{N} \|m\|_2 \|\mathbf{Cov}\|_F^2, \\ \|D^2 U(m)\{\kappa_3, \mathbf{Cov}\}\|_F &\leq \frac{6\beta^2}{N} \|m\|_2 \|\kappa_3\|_F \|\mathbf{Cov}\|_F, \\ \|D^3 U(m)\{\mathbf{Cov}, \mathbf{Cov}, \mathbf{Cov}\}\|_F &\leq \frac{6\beta^2}{N} \|\mathbf{Cov}\|_F^3. \end{aligned}$$

By [Lemma 4.9](#),

$$\mathbf{E} \|\mathbf{Cov}\|_F^p \leq C_p N^{p/2}, \quad \mathbf{E} \|\kappa_3\|_F^p \leq C_p N^{3p/4}.$$

Combining these with $\|m\|_2 \leq \sqrt{N}$ and Hölder’s inequality gives [Eq. \(28\)](#) to [\(30\)](#) with U in place of T . $\square$

Corollary A.10. *For all $p \geq 2$, there exists $C_p > 0$ such that for all $|B| \leq 2$,*

$$\mathbf{E} \|\nabla r^{(B)}\|_F^p \leq C_p, \quad \mathbf{E} \|\nabla^2 r^{(B)}\|_F^p \leq C_p N^{p/4}.$$

Proof. Combining the estimate on $\nabla^j \lambda_i^{(B)}$ in [Lemma A.7](#) (with $j = 1, 2$) with [Lemma A.9](#) yields

$$\mathbf{E} \|\nabla \lambda_i^{(B)}\|_F^p \leq C_p N^{-p/2}, \quad \mathbf{E} \|\nabla^2 \lambda_i^{(B)}\|_F^p \leq C_p N^{-p/4}.$$

The result follows by differentiating $r_i^{(B)} = \zeta_i^{(B)} + m_i^{(B)} \lambda_i^{(B)}$ and applying the last display and the remaining estimates from [Lemma A.7](#). $\square$

Lemma A.11. *For all $p \geq 2$, there exists $C_p > 0$ such that for all $|B| \leq 2$,*

$$\mathbf{E} \|\mathbf{Cov}^{(B)}\|_{\text{op}}^p \leq C_p.$$

In particular, [Lemma A.2\(b\)](#) follows as the case $|B| \leq 1$.

Proof. We provide the proof for $B = \emptyset$ and omit the superscript $(\emptyset)$; the case of general B is analogous. Let $\mathcal{E} = \{K \succeq \gamma \text{Id}\}$, where $\gamma > 0$ is given by [Lemma 4.2](#). On $\mathcal{E}$, the identity $K\mathbf{Cov} = \text{Id} + \nabla r$ from [Fact A.4](#) gives

$$\|\mathbf{Cov}\|_{\text{op}} \leq \|K^{-1}\|_{\text{op}} \|\text{Id} + \nabla r\|_{\text{op}} \leq \gamma^{-1} (1 + \|\nabla r\|_F).$$

On the complement of $\mathcal{E}$, we use the crude bound

$$\|\mathbf{Cov}\|_{\text{op}} \leq \text{Tr } \mathbf{Cov} = \sum_i (1 - m_i^2) \leq N.$$

Thus,

$$\begin{aligned}\mathbf{E}\|\mathbf{Cov}\|_{\text{op}}^p &\leq \gamma^{-p}\mathbf{E}\left[(1 + \|\nabla r\|_F)^p\right] + N^p\Pr[\mathcal{E}^c] \\ &\leq 2^{p-1}\gamma^{-p}\left(1 + \mathbf{E}\|\nabla r\|_F^p\right) + N^p\Pr[\mathcal{E}^c].\end{aligned}$$

By [Corollary A.10](#) and [Lemma 4.2](#), this is bounded by C_p . □

Lemma A.12. *There exists a universal $C > 0$ such that for all $a > 0$, there exists $C_a > 0$ such that*

$$\Pr\left[\max_{|B|\leq 2}\|\mathbf{Cov}^{(B)}\|_F > C\sqrt{N}\right] \leq C_a N^{-a}.$$

In particular, [Lemma A.2\(c\)](#) follows by restricting the maximum to $|B| \leq 1$.

Proof. We will show that there exists a universal $C > 0$ such that for all $a > 0$, there exists $C_a > 0$ such that uniformly over B with $|B| \leq 2$,

$$\Pr[\|\mathbf{Cov}^{(B)}\|_F > C\sqrt{N}] \leq C_a N^{-a}. \quad (33)$$

The conclusion follows from [Eq. \(33\)](#) by a union bound (and adjusting a). We provide the proof of [Eq. \(33\)](#) for $B = \emptyset$ and omit the superscript $(\emptyset)$; the case of general B is analogous.

Let $\gamma > 0$ be given by [Lemma 4.2](#) and set $\mathcal{E} = \{K \succeq \gamma \text{Id}\}$. On $\mathcal{E}$, [Fact A.4](#) gives

$$\|\mathbf{Cov}\|_F \leq \|K^{-1}\|_{\text{op}}\|\text{Id} + \nabla r\|_F \leq \gamma^{-1}\left(\sqrt{N} + \|\nabla r\|_F\right).$$

Take $p \geq \max(2, 2a)$. By Markov's inequality, [Corollary A.10](#), and [Lemma 4.2](#),

$$\begin{aligned}\Pr\left[\|\mathbf{Cov}\|_F > 2\gamma^{-1}\sqrt{N}\right] &\leq \Pr[\mathcal{E}^c] + \Pr\left[\|\nabla r\|_F > \sqrt{N}\right] \\ &\leq \Pr[\mathcal{E}^c] + N^{-p/2}\mathbf{E}\|\nabla r\|_F^p \leq C_a N^{-a}.\end{aligned}$$

This proves [Eq. \(33\)](#) with $C = 2\gamma^{-1}$. □

Lemma A.13. *For all $p \geq 2$, there exists $C_p > 0$ such that for all $|B| \leq 1$,*

$$\max_{i \in [N] \setminus B} \mathbf{E} \left| \|d_i^{(B)}\|_2^2 - \frac{\beta^2}{N} \|\mathbf{Cov}^{(B \cup \{i\})}\|_F^2 \right|^p \leq C_p N^{-p/2}.$$

Proof. We provide the proof for $B = \emptyset$ and omit the superscript $(\emptyset)$; the case of general B is analogous. Fix $i \in [N]$ and set

$$\epsilon_i = d_i - \beta \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i}, \quad H_i = (\mathbf{Cov}^{(i)})^2.$$

Note that $v_i^{-1} \nabla m_i = v_i^{-1} \mathbf{Cov}^{(i)}$ has 1 in its i -th entry, and its remaining entries are d_i . Thus [Lemma A.7](#) implies, for every $p \geq 2$,

$$\mathbf{E}\|d_i\|_2^p \leq C_p, \quad \mathbf{E}\|\epsilon_i\|_2^p \leq C_p N^{-p/2}.$$

Note that

$$\|d_i\|_2^2 - \beta^2 \mathbf{M}_{i,-i}^\top H_i \mathbf{M}_{i,-i} = \|d_i\|_2^2 - \|d_i - \epsilon_i\|_2^2 = 2\langle d_i, \epsilon_i \rangle - \|\epsilon_i\|_2^2.$$

By Hölder's inequality,

$$\mathbf{E} \left| \|d_i\|_2^2 - \beta^2 \mathbf{M}_{i,-i}^\top H_i \mathbf{M}_{i,-i} \right|^p \leq C_p N^{-p/2}. \quad (34)$$

Conditional on $(\mathbf{x}, \mathbf{W}_{-i,-i}, \mathbf{g})$, the matrix H_i is fixed and

$$\mathbf{M}_{i,-i} = \mathbf{W}_{i,-i} + \eta_i, \quad \eta_i = \frac{\beta \mathbf{x}_i \mathbf{x}_{-i}}{N},$$

where $\mathbf{W}_{i,-i} \sim \mathcal{N}(0, \text{Id}/N)$ and $\|\eta_i\|_2 \leq \frac{\beta}{\sqrt{N}}$. Writing $\mathbf{E}_i$ for expectation over $\mathbf{W}_{i,-i}$ under this conditioning, standard Gaussian quadratic- and linear-form moment bounds give

$$\mathbf{E}_i \left| \mathbf{M}_{i,-i}^\top H_i \mathbf{M}_{i,-i} - \frac{\text{Tr } H_i}{N} \right|^p \leq \frac{C_p}{N^p} \|H_i\|_F^p + \frac{\|H_i \eta_i\|_2^p}{N^{p/2}} + |\eta_i^\top H_i \eta_i|^p \leq \frac{C_p}{N^p} \|H_i\|_F^p.$$

Since $\|H_i\|_F \leq \sqrt{N} \|\mathbf{Cov}^{(i)}\|_{\text{op}}^2$, [Lemma A.11](#) implies

$$\mathbf{E} \left| \mathbf{M}_{i,-i}^\top H_i \mathbf{M}_{i,-i} - \frac{\text{Tr } H_i}{N} \right|^p \leq C_p N^{-p/2}.$$

Combining with [Eq. \(34\)](#) and $\text{Tr } H_i = \|\mathbf{Cov}^{(i)}\|_F^2$ proves the result. $\square$

Proof of [Lemma A.2\(d\)](#). We will show that there exists a universal $C > 0$ such that for all $a > 0$, there exists $C_a > 0$ such that uniformly over B with $|B| \leq 1$,

$$\Pr[\|\kappa_3^{(B)}\|_F > C\sqrt{N}] \leq C_a N^{-a}. \quad (35)$$

The conclusion follows by a union bound (and adjusting a). We provide the proof of [Eq. \(35\)](#) for $B = \emptyset$ and omit the superscript $(\emptyset)$; the case of general B is analogous.

Let A, U be defined as in [\(31\)](#); as observed in the proof of [Lemma A.9](#), $T - A - U$ is linear. By [Fact A.4](#),

$$K\kappa_3 = \nabla^2 r - D^2 A(m) \{\mathbf{Cov}, \mathbf{Cov}\} - D^2 U(m) \{\mathbf{Cov}, \mathbf{Cov}\}. \quad (36)$$

By [Eq. \(32\)](#) and the estimates $\|m\|_2 \leq \sqrt{N}$, $\|\mathbf{Cov}\|_F \leq \sqrt{N} \|\mathbf{Cov}\|_{\text{op}}$,

$$\|D^2 U(m) \{\mathbf{Cov}, \mathbf{Cov}\}\|_F \leq \frac{6\beta^2}{N} \|m\|_2 \|\mathbf{Cov}\|_{\text{op}} \|\mathbf{Cov}\|_F \leq 6\beta^2 \|\mathbf{Cov}\|_{\text{op}}^2.$$

Together with [Corollary A.10](#) and [Lemma A.11](#), this implies, for every fixed $p \geq 2$,

$$\max \left\{ \mathbf{E} \|\nabla^2 r\|_F^p, \mathbf{E} \|D^2 U(m) \{\mathbf{Cov}, \mathbf{Cov}\}\|_F^p \right\} \leq C_p N^{p/4}.$$

By Markov's inequality with $p \geq \max(2, 4a)$,

$$\Pr \left(\max(\|\nabla^2 r\|_F, \|D^2 U(m) \{\mathbf{Cov}, \mathbf{Cov}\}\|_F) \geq \sqrt{N} \right) \leq C_a N^{-a}. \quad (37)$$

We turn to controlling the term $\|D^2 A(m) \{\mathbf{Cov}, \mathbf{Cov}\}\|_F$. As noted in the proof of [Lemma A.13](#), the i -th coordinate of $v_i^{-1} \nabla m_i$ is 1, and the remaining coordinates are d_i . So,

$$\|D^2 A(m) \{\mathbf{Cov}, \mathbf{Cov}\}\|_F^2 = 4 \sum_i m_i^2 \|v_i^{-1} \nabla m_i\|_2^4 \leq 4 \sum_i (1 + \|d_i\|_2^2)^2.$$

Set $q_i = \beta^2 \|\mathbf{Cov}^{(i)}\|_F^2 / N$ and

$$\Gamma = \sum_i \left| \|d_i\|_2^2 - q_i \right|^2.$$

By Jensen's inequality and [Lemma A.13](#), for every $p \geq 1$,

$$\mathbf{E} \Gamma^p \leq N^{p-1} \sum_i \mathbf{E} \left| \|d_i\|_2^2 - q_i \right|^{2p} \leq C_p.$$

By Markov's inequality with $p \geq \max(1, a)$, $\Pr(\Gamma \leq N) \geq 1 - C_a N^{-a}$. [Lemma A.12](#) provides a universal constant C_0 such that

$$\Pr(\max_i q_i \leq \beta^2 C_0^2) \geq 1 - C_a N^{-a}.$$

On the intersection of these two events,

$$\|D^2 A(m) \{\mathbf{Cov}, \mathbf{Cov}\}\|_F^2 \leq 4 \sum_i (1 + \|d_i\|_2^2)^2 \leq 8 \sum_i (1 + q_i)^2 + 8\Gamma \leq 8N ((1 + \beta^2 C_0^2)^2 + 1).$$

This provides an absolute C such that

$$\Pr(\|D^2 A(m) \{\mathbf{Cov}, \mathbf{Cov}\}\|_F > C\sqrt{N}) \leq C_a N^{-a}.$$

Recalling [\(36\)](#), a union bound with [Eq. \(37\)](#) shows

$$\Pr(\|K\kappa_3\|_F > C\sqrt{N}) \leq C_a N^{-a},$$

after adjusting C and C_a . Combining with [Lemma 4.2](#) proves the result. $\square$

We finally present the deferred proof of [Lemma A.7](#).

Proof of Lemma A.7. We provide the proof for $B = \emptyset$ and omit the superscript $(\emptyset)$; the case of general B is analogous. The proof proceeds in five steps, and uses only [Lemmas A.1](#) and [4.9](#) and Gaussian moment estimates.

Step 1: cavity expansion for d_i . Let $\epsilon_i = d_i - \beta \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i}$. In this step we prove

$$\mathbf{E} \|\epsilon_i\|_2^p \leq C_p N^{-p/2}, \tag{38}$$

which is the first estimate of the lemma. For $s' \in \{\pm 1\}$, Taylor expansion gives

$$m^{(i,s')} = \nabla F^{(i)}(u_i + s' \beta \mathbf{M}_{i,-i}) = \nabla F^{(i)}(u_i) + s' \beta \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i} + \frac{\beta^2}{2} \kappa_3^{(i)} [\mathbf{M}_{i,-i}^{\otimes 2}] + \epsilon_{i,s'},$$

where

$$\epsilon_{i,s'} = \frac{s' \beta^3}{2} \int_0^1 (1-s)^2 \kappa_{4,ss'}^{(i)} [\mathbf{M}_{i,-i}^{\otimes 3}] \, ds.$$

Thus

$$d_i = \frac{m^{(i,+)} - m^{(i,-)}}{2} = \beta \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i} + \frac{\epsilon_{i,+} - \epsilon_{i,-}}{2},$$

and

$$\epsilon_i = \frac{\epsilon_{i,+} - \epsilon_{i,-}}{2} = \frac{\beta^3}{4} \int_0^1 (1-s)^2 \left(\kappa_{4,s}^{(i)} + \kappa_{4,-s}^{(i)} \right) [\mathbf{M}_{i,-i}^{\otimes 3}] \, ds. \tag{39}$$

Applying [Lemma A.1](#) with $r = 4, k = 0$, and $\ell = 3$ yields [Eq. \(38\)](#).

Step 2: moments of d_i and its derivatives. In this step, we prove the auxiliary estimate

$$\mathbf{E} \|\nabla^j d_i\|_F^p \leq C_p N^{pj/4} \quad (40)$$

for $j = 0, 1, 2, 3$. Conditional on $(\mathbf{x}, \mathbf{W}_{-i,-i}, \mathbf{g})$, the matrix $\mathbf{Cov}^{(i)}$ is fixed and

$$\mathbf{M}_{i,-i} = \mathbf{W}_{i,-i} + \eta_i, \quad \eta_i = \frac{\beta \mathbf{x}_i \mathbf{x}_{-i}}{N},$$

where $\mathbf{W}_{i,-i} \sim \mathcal{N}(0, \text{Id}/N)$ and $\|\eta_i\|_2 \leq \beta/\sqrt{N}$. Write $\mathbf{E}_i$ for expectation over $\mathbf{W}_{i,-i}$ under this conditioning. Gaussian moments give

$$\mathbf{E}_i \|\mathbf{Cov}^{(i)} \mathbf{M}_{i,-i}\|_2^p \leq \frac{C_p}{N^{p/2}} \|\mathbf{Cov}^{(i)}\|_F^p + \|\mathbf{Cov}^{(i)} \eta_i\|_2^p \leq \frac{C_p}{N^{p/2}} \|\mathbf{Cov}^{(i)}\|_F^p.$$

Taking expectation and using the covariance bound from [Lemma 4.9](#), we obtain

$$\mathbf{E} \|\mathbf{Cov}^{(i)} \mathbf{M}_{i,-i}\|_2^p \leq C_p$$

Together with [Eq. \(38\)](#) this implies [Eq. \(40\)](#) for $j = 0$. For $j = 1, 2, 3$, differentiating $d_i = (m^{(i,+)} - m^{(i,-)})/2$ and then integrating in s gives

$$\nabla^j d_i = \frac{\kappa_{j+1,1}^{(i)} - \kappa_{j+1,-1}^{(i)}}{2} = \frac{\beta}{2} \int_{-1}^1 \kappa_{j+2,s}^{(i)} [\mathbf{M}_{i,-i}] \, ds.$$

Applying [Lemma A.1](#) with $r = j + 2, k = 0$, and $\ell = 1$ proves [Eq. \(40\)](#).

Step 3: derivatives of m_i and ζ_i . In this step we prove the lemma's estimates on $(v_i)^{-1} \nabla^j m_i$ and $\nabla^j \zeta_i$. Set $a_i = \text{arctanh}(m_i)$. By the definition of d_i ,

$$\nabla a_i = v_i^{-1} \nabla m_i = v_i^{-1} \mathbf{Cov}_i = (1, d_i),$$

where we write the i -th coordinate first. Differentiating this identity and using [Eq. \(40\)](#) gives

$$\mathbf{E} \|\nabla^j a_i\|_F^p \leq C_p N^{p(j-1)/4} \quad (41)$$

for $j = 1, 2, 3$. Since $m_i = \tanh a_i$ and $v_i = 1 - m_i^2$, we calculate

$$\frac{\tanh'(a_i)}{v_i} = 1, \quad \frac{\tanh''(a_i)}{v_i} = -2m_i, \quad \frac{\tanh'''(a_i)}{v_i} = 6m_i^2 - 2.$$

Thus the chain rule gives

$$\begin{aligned} v_i^{-1} \nabla m_i &= \nabla a_i, \\ v_i^{-1} \nabla^2 m_i &= \nabla^2 a_i - 2m_i (\nabla a_i)^{\otimes 2}, \\ v_i^{-1} \nabla^3 m_i &= \nabla^3 a_i - 2m_i \sum_{\text{Sym}} \nabla^2 a_i \otimes \nabla a_i + (6m_i^2 - 2) (\nabla a_i)^{\otimes 3}. \end{aligned}$$

Since $|m_i| \leq 1$, all coefficients are bounded. Applying [Eq. \(41\)](#) proves the desired estimates on $v_i^{-1} \nabla^j m_i$. Finally, differentiating the integral representation in [Eq. \(26\)](#) yields

$$\nabla^j \zeta_i = -\frac{\beta^3}{4} \int_{-1}^1 (1 - s^2) \kappa_{j+3,s}^{(i)} [\mathbf{M}_{i,-i}^{\otimes 3}] \, ds.$$

Applying [Lemma A.1](#) with $r = j + 3$, $k = 0$, and $\ell = 3$ proves the claimed estimates on $\nabla^j \zeta_i$.

Step 4: cavity estimate for covariance trace. Set $\Xi_i = \text{Tr } \mathbf{Cov} - \text{Tr } \mathbf{Cov}^{(i)}$. In this step, we prove the auxiliary estimate

$$\mathbf{E} \|\nabla^j \Xi_i\|_F^p \leq C_p N^{p(j+1)/4} \quad (42)$$

for $j = 0, 1, 2, 3$. Note that here, ∇ includes derivatives in coordinate i of the external field $u = h$, even though $\mathbf{Cov}^{(i)}$ does not depend on this coordinate.

Let σ be a sample from the Gibbs measure. The law of total covariance yields

$$\begin{aligned} \Xi_i &= \text{Tr } \mathbf{E}[\mathbf{Cov}(\sigma|\sigma_i)] - \text{Tr } \mathbf{Cov}^{(i)} + \text{Tr } \mathbf{Cov}(\mathbf{E}[\sigma|\sigma_i]) \\ &= \frac{1+m_i}{2} \kappa_{2,1}^{(i)} [\text{Id}] + \frac{1-m_i}{2} \kappa_{2,-1}^{(i)} [\text{Id}] - \kappa_2^{(i)} [\text{Id}] + v_i(1 + \|d_i\|_2^2) = \frac{1}{2} \Xi_{i,1} + \frac{m_i}{2} \Xi_{i,2} + v_i(1 + \|d_i\|_2^2), \end{aligned}$$

where

$$\Xi_{i,1} = \left(\kappa_{2,1}^{(i)} + \kappa_{2,-1}^{(i)} - 2\kappa_2^{(i)} \right) [\text{Id}], \quad \Xi_{i,2} = \left(\kappa_{2,1}^{(i)} - \kappa_{2,-1}^{(i)} \right) [\text{Id}].$$

For $j = 0, 1, 2, 3$, Taylor expansion in s gives

$$\begin{aligned} \nabla^j \Xi_{i,1} &= \left(\kappa_{j+2,1}^{(i)} + \kappa_{j+2,-1}^{(i)} - 2\kappa_{j+2}^{(i)} \right) [\text{Id}] = \beta^2 \int_0^1 (1-s) \left(\kappa_{j+4,s}^{(i)} + \kappa_{j+4,-s}^{(i)} \right) [\text{Id} \otimes \mathbf{M}_{i,-i}^{\otimes 2}] \, ds, \\ \nabla^j \Xi_{i,2} &= \left(\kappa_{j+2,1}^{(i)} - \kappa_{j+2,-1}^{(i)} \right) [\text{Id}] = \beta \int_{-1}^1 \kappa_{j+3,s}^{(i)} [\text{Id} \otimes \mathbf{M}_{i,-i}] \, ds. \end{aligned}$$

Applying [Lemma A.1](#) with $(r, k, \ell) = (j+4, 1, 2)$ and $(j+3, 1, 1)$ respectively, and integrating in s , bounds

$$\mathbf{E} \|\nabla^j \Xi_{i,1}\|_F^p \leq C_p N^{pj/4}, \quad \mathbf{E} \|\nabla^j \Xi_{i,2}\|_F^p \leq C_p N^{p(j+1)/4}.$$

From the conclusion of step 3 and the trivial bound $v_i \leq 1$, we have $\mathbf{E} \|\nabla^j m_i\|_F^p \leq C_p N^{p(j-1)/4}$ for $j = 1, 2, 3$. Differentiating $v_i = 1 - m_i^2$ gives the same bound for $\nabla^j v_i$; thus

$$\max \left\{ \mathbf{E} \|\nabla^j m_i\|_F^p, \mathbf{E} \|\nabla^j v_i\|_F^p \right\} \leq C_p N^{p(j-1)/4} \quad (43)$$

for $j = 1, 2, 3$. Combining [Eq. \(40\)](#) and [\(43\)](#) with $0 \leq v_i \leq 1$ yields

$$\mathbf{E} \left\| \nabla^j (v_i(1 + \|d_i\|_2^2)) \right\|_F^p \leq C_p N^{pj/4}$$

for $j = 0, 1, 2, 3$. Together this proves [Eq. \(42\)](#).

Step 5: derivatives of λ_i . In this step we prove the lemma's estimates on $\nabla^j \lambda_i$. Using $\text{Tr } \mathbf{Cov} = N - \|m\|_2^2$ and the definitions $\epsilon_i = d_i - \beta \mathbf{Cov}^{(i)} \mathbf{M}_{i,-i}$ from Step 1 and $\Xi_i = \text{Tr } \mathbf{Cov} - \text{Tr } \mathbf{Cov}^{(i)}$ from Step 4, we obtain

$$\lambda_i = \frac{\beta^2}{N} \text{Tr } \mathbf{Cov} - \beta \mathbf{M}_{i,-i}^\top d_i = \frac{\beta^2}{N} \Xi_i - \beta^2 \mathbf{Cov}^{(i)} \left[\mathbf{M}_{i,-i}^{\otimes 2} - \frac{\text{Id}}{N} \right] - \beta \mathbf{M}_{i,-i}^\top \epsilon_i.$$

Differentiating with respect to the external field gives

$$\nabla^j \lambda_i = \frac{\beta^2}{N} \nabla^j \Xi_i - \beta^2 \kappa_{j+2}^{(i)} \left[\mathbf{M}_{i,-i}^{\otimes 2} - \frac{\text{Id}}{N} \right] - \beta \nabla^j (\mathbf{M}_{i,-i}^\top \epsilon_i), \quad (44)$$

for $j = 0, 1, 2, 3$. Since $\mathbf{M}_{i,-i}$ does not depend on the external field, differentiating Eq. (39) gives

$$\nabla^j(\mathbf{M}_{i,-i}^\top \epsilon_i) = \frac{\beta^3}{4} \int_0^1 (1-s)^2 \left(\kappa_{j+4,s}^{(i)} + \kappa_{j+4,-s}^{(i)} \right) [\mathbf{M}_{i,-i}^{\otimes 4}] \, ds.$$

By Lemma A.1 with $r = j + 4, k = 0$, and $\ell = 4$,

$$\mathbf{E} \|\nabla^j(\mathbf{M}_{i,-i}^\top \epsilon_i)\|_F^p \leq C_p N^{p(j-4)/4}. \quad (45)$$

We will bound the $\nabla^j \Xi_i$ term in Eq. (44) using Eq. (42). We now turn to the $\kappa_{j+2}^{(i)}$ term. Recall from Step 2 the notation $\mathbf{M}_{i,-i} = \mathbf{W}_{i,-i} + \eta_i$, and that $\mathbf{E}_i$ denotes expectation over $\mathbf{W}_{i,-i}$ conditional on $(\mathbf{x}, \mathbf{W}_{-i,-i}, \mathbf{g})$. Then

$$\mathbf{M}_{i,-i}^{\otimes 2} - \frac{\text{Id}}{N} = \left(\mathbf{W}_{i,-i}^{\otimes 2} - \frac{\text{Id}}{N} \right) + \mathbf{W}_{i,-i} \otimes \eta_i + \eta_i \otimes \mathbf{W}_{i,-i} + \eta_i^{\otimes 2}.$$

For any tensor H of order at least two measurable with respect to $(\mathbf{x}, \mathbf{W}_{-i,-i}, \mathbf{g})$, Gaussian quadratic- and linear-form moment bounds give

$$\mathbf{E}_i \left\| H \left[\mathbf{M}_{i,-i}^{\otimes 2} - \frac{\text{Id}}{N} \right] \right\|_F^p \leq C_p \left(\frac{1}{N} + \frac{\|\eta_i\|_2}{\sqrt{N}} + \|\eta_i\|_2^2 \right)^p \|H\|_F^p \leq \frac{C_p}{N^p} \|H\|_F^p.$$

The last inequality uses $\|\eta_i\|_2 \leq \beta/\sqrt{N}$ from Step 2. Taking $H = \kappa_{j+2}^{(i)}$, which is fixed under the conditioning, and taking expectation yields

$$\mathbf{E} \left\| \kappa_{j+2}^{(i)} \left[\mathbf{M}_{i,-i}^{\otimes 2} - \frac{\text{Id}}{N} \right] \right\|_F^p \leq \frac{C_p}{N^p} \mathbf{E} \|\kappa_{j+2}^{(i)}\|_F^p.$$

Combining this with Eq. (42), (44) and (45) proves

$$\mathbf{E} \|\nabla^j \lambda_i\|_F^p \leq C_p \left(N^{p(j-3)/4} + N^{-p} \mathbf{E} \|\kappa_{j+2}^{(i)}\|_F^p \right)$$

for $j = 0, 1, 2, 3$, as desired. If $j = 0$, Lemma 4.9 gives $\mathbf{E} \|\kappa_2^{(i)}\|_F^p \leq C_p N^{p/2}$, and we obtain the final assertion $\mathbf{E} |\lambda_i|^p \leq C_p N^{-p/2}$ of the lemma. $\square$