Markov Chains Approximate Message Passing

Amit Rajaraman[*] David X. Wu[†]

December 3, 2025

Abstract

Markov chain Monte Carlo algorithms have long been observed to obtain near-optimal performance in various Bayesian inference settings. However, developing a supporting theory that make these studies rigorous has proved challenging.

In this paper, we study the classical spiked Wigner inference problem, where one aims to recover a planted Boolean spike from a noisy matrix measurement. We relate the recovery performance of Glauber dynamics on the annealed posterior to the performance of Approximate Message Passing (AMP), which is known to achieve Bayes-optimal performance. Our main results rely on the analysis of an auxiliary Markov chain called *restricted Gaussian dynamics* (RGD). Concretely, we establish the following results:

1. RGD can be reduced to an effective one-dimensional recursion which mirrors the evolution of the AMP iterates.
2. From a warm start, RGD rapidly converges to a fixed point in correlation space, which recovers Bayes-optimal performance when run on the posterior.
3. Conditioned on widely believed mixing results for the SK model, we recover the phase transition for non-trivial inference.

*MIT. `amit_r@mit.edu`. Supported by a MathWorks Fellowship.
†UC Berkeley. `david_wu@berkeley.edu`. Supported by NSF Graduate Research Fellowship DGE-2146752.

Contents

1 Introduction

Markov chains are fundamental objects of study which model the dynamics of complex systems arising in statistical physics, computational biology, and social networks. Besides modeling physical systems, they are a highly successful algorithmic tool to sample from high-dimensional distributions in Bayesian inference, statistics, and theoretical computer science.

It has long been empirically observed that canonical Markov chains such as *Glauber dynamics* which attempt to sample from the posterior distribution are quite successful for Bayesian inference. For example, [DKMZ11] observed that for community detection in the stochastic block model, the output of the Glauber dynamics is Bayes optimal. However, there has been a dearth of rigorous results confirming or explaining the algorithmic success of Markov chains in statistical inference. One fundamental reason for the lack of rigorous progress is that, for many posterior distributions of interest, Glauber dynamics provably suffers from slow worst-case mixing. This leads to the following natural question.

Question 1.1. *If we run Glauber dynamics for polynomial time on the posterior distribution of a natural Bayesian inference problem, does it achieve Bayes-optimal performance?*

To make progress on this, we revisit early empirical studies which compared the performance of Glauber dynamics to that of local *message passing* algorithms. Our starting point is [DKMZ11], which found that the performance of Glauber dynamics matches that of the Belief Propagation (BP) algorithm which directly estimates the marginals of the posterior in the stochastic block model.

Since then, message passing algorithms such as BP and its dense variant Approximate Message Passing (AMP) [DMM09, BM11] have been widely applied in a variety of high-dimensional statistics problems, to great success (for a few examples, see [DAM16, MV21, EAMS22, HMP24] and references therein). A particularly appealing property of AMP is that its asymptotic performance can be rigorously characterized by a finite-dimensional recursion known as *state evolution*.

For certain inference problems, one can use state evolution to design AMP algorithms to construct Bayes-optimal estimators. Similarly, we *a priori* know Glauber achieves the same guarantees in exponential time, since by then the dynamics would be able to produce Gibbs samples. Motivated by this, we refine our initial question to ask about connecting the performance of these two classes of algorithms if Glauber is only run for polynomial time.

Question 1.2. *If we run Glauber dynamics for polynomial time, can we prove it performs as well as AMP?*

Our main results make progress on Question 1.2 by proving a formal connection between Markov chains and AMP for the spiked Wigner inference problem, which we define below. We show that the dynamics of the correlation with the spike—the observable of primary interest—are governed by fixed point relations which also characterize the asymptotic performance of AMP for the same problem. See Figure 1 for a schematic diagram of our main results.

The precision of our results require model-specific calculations that we did not attempt to generalize beyond the spiked Wigner setting. However, our high-level proof strategy is fairly generic and we believe it should be applicable to general spiked matrix inference problems.

In the spiked Wigner model, one aims to recover an unknown *signal* $\boldsymbol{x}$ drawn from the prior $\mathrm{Unif}(\{\pm 1\})^{\otimes N}$ given a noisy matrix measurement

$$\boldsymbol{M} = \boldsymbol{W} + \frac{\lambda}{N}\boldsymbol{x}\boldsymbol{x}^\top , \tag{1}$$

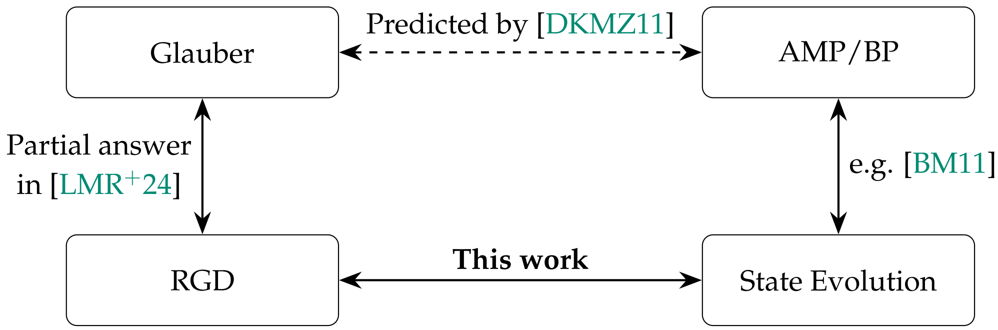


Figure 1: An illustration of our results.

for some signal-to-noise ratio (or SNR) $\lambda \geqslant 0$, and a *noise* matrix $\boldsymbol{W} \sim \mathsf{GOE}(N)$ with $\boldsymbol{W}_{ij} = \boldsymbol{W}_{ji} \sim \mathcal{N}\left(0, \frac{1}{N}\right)$ for $i \neq j$ and $\boldsymbol{W}_{ii} \sim \mathcal{N}(0, \frac{2}{N})$.

The spiked Wigner model has been extensively studied over the last two decades, providing a rich test bed for average-case complexity. Perhaps the first interesting phenomenon uncovered by this line of work is the famous BBP phase transition [HR04, BBP05], which marks the information-theoretic and computational phase transition at $\lambda = 1$. Indeed, it is known that for $\lambda < 1$, $\boldsymbol{M}$ is mutually contiguous with $\boldsymbol{W}$, whereas for $\lambda > 1$ the top eigenvector of $\boldsymbol{M}$ is nontrivially correlated with $\boldsymbol{x}$.

A straightforward calculation shows that the posterior distribution in the spiked Wigner model is of the form

$$\mu_{\beta \boldsymbol{M}}(\sigma) \propto \exp\left(\frac{\beta}{2}\sigma^\top \boldsymbol{M}\sigma\right) = \exp\left(\frac{\beta}{2}\sigma^\top \boldsymbol{W}\sigma + \frac{\beta\lambda}{2N}\langle \sigma, \boldsymbol{x}\rangle^2\right) \tag{2}$$

for $\beta = \lambda$.

Beyond worst-case mixing. [LMR+24] introduced the framework of locally stationary distributions to analyze the performance of Markov chains before they have mixed. Roughly speaking, a locally stationary distribution is any distribution which is approximately invariant under another step of the Markov chain. They are also very easy to sample from—any reversible Markov chain such as Glauber dynamics produces a locally stationary distribution in $\mathrm{poly}(N)$ steps.

For the spiked Wigner problem, they show that for sufficiently large (constant) λ and $\beta < \frac{1}{4}$, the Glauber dynamics run on $\mu_{\beta \boldsymbol{M}}$ for a large polynomial time from an arbitrary initialization reaches a vector that is non-trivially correlated with the spike. To prove this result, they combined the properties of locally stationary distributions with an analysis of a different Markov chain known as *Restricted Gaussian Dynamics* (RGD), which we define below.[1]

Definition 1.3 (Restricted Gaussian dynamics [STL20, LST21, CE22])**.** For $\boldsymbol{M} = \boldsymbol{W} + \frac{\lambda}{N}\boldsymbol{x}\boldsymbol{x}^\top$, *restricted Gaussian dynamics* (RGD) at inverse temperature β is a Markov chain on $\{\pm 1\}^N$ where a transition from σ to $\boldsymbol{\sigma}'$ is given by the following:

- Sample $\boldsymbol{g} \sim \mathcal{N}(0, 1)$ and define $\boldsymbol{z} := \beta\lambda\frac{1}{N}\langle \sigma, \boldsymbol{x}\rangle + \sqrt{\frac{\beta\lambda}{N}}\boldsymbol{g}$.
- Sample $\boldsymbol{\sigma}'$ from the distribution $\mu_{\beta \boldsymbol{W}, \boldsymbol{z}\boldsymbol{x}}$.

[1] The precise definition is not terribly important, so readers seeking intuition can skip the definition without much concern.

It is straightforward to show that RGD has $\mu_{\beta \boldsymbol{M}}$ as its unique stationary distribution. However, the same bottleneck that precludes fast mixing for Glauber also prevents RGD from mixing quickly. In spite of this, we achieve a precise and nearly distributional understanding of the dynamics of RGD, which allows us to deduce several interesting properties of its dynamics. Although RGD is not an implementable algorithm, the tools developed by [LMR$^+$24] allow us to transfer this understanding over to Glauber dynamics in certain temperature regimes.

Before we state our main results, we take a brief excursion to give some standard spin glass definitions.

Definition 1.4. Let $\boldsymbol{W} \sim \mathsf{GOE}(N)$. The Sherrington–Kirkpatrick model (SK model) at inverse temperature $\beta > 0$ and external field $v \in \mathbb{R}^N$ is the distribution over $\{\pm 1\}^N$ with density

$$\mu_{\beta \boldsymbol{W}, v}(\sigma) \propto \exp\left(\frac{\beta}{2} \cdot \sigma^\top \boldsymbol{W} \sigma + \langle v, \sigma \rangle\right) .$$

.

Our results for the performance of Glauber on the spiked Wigner posterior rely on the rapid mixing of Glauber dynamics for the SK model. To simplify the discussion, we will state our results whenever the SK model satisfies the following mixing condition based on modified log-Sobolev inequalities.

Condition 1 (MLSI)**.** *We say that $\beta < 1$ satisfies (MLSI) if with probability $1 - o(1)$ over the random matrix $\boldsymbol{W}$, for all $h \in \mathbb{R}$, $\mu_{\beta \boldsymbol{W}, h\mathbf{1}}$ satisfies a modified log-Sobolev inequality with constant $\Omega(1/N)$.*

A recent breakthrough line of work [EKZ22, CE22, AKV24] has made significant progress towards establishing rapid mixing of Glauber dynamics in the entire high-temperature regime for the SK model. In particular, they establish that (MLSI) holds for $\beta < 0.295$; it is conjectured that (MLSI) holds for the entire high-temperature regime $\beta < 1$ [MPV87]. We are now poised to state our main results.

1.1 Main results

The following theorem establishes that the short-term behavior of the correlation of Glauber iterates are approximate fixed points of an AMP-like fixed point equation in the entire high SNR regime $\lambda > 1$. Furthermore, it establishes a sharp phase transition at $\beta = \frac{1}{\lambda}$ for the success of Glauber dynamics at weak recovery of the spike.

Theorem 1.5 (Informal, see Theorem 5.2)**.** *Let $\boldsymbol{\sigma}_0 \in \{\pm 1\}^N$ be arbitrary, and let $(\boldsymbol{\sigma}_t)_{t \geqslant 0}$ be the trajectory of the Glauber dynamics run with stationary distribution $\mu_{\beta \boldsymbol{M}}$ initialized at $\boldsymbol{\sigma}_0$. Let $T \geqslant \widetilde{\Omega}(N^4)$, and $\widehat{\sigma} = \boldsymbol{\sigma}_{\boldsymbol{t}}$ for a uniformly random $\boldsymbol{t} \sim [0, T]$.*

Then, for any $\varepsilon > 0$ and any $\beta \in (0, 1)$ such that $\beta \neq \frac{1}{\lambda}$ and (MLSI) holds, with high probability over the noise $\boldsymbol{W}$, we have

$$\mathbf{Pr}\left[\left|\frac{1}{N}\left|\langle \widehat{\sigma}, \boldsymbol{x} \rangle\right| - \mathsf{OPT}_{\beta,\lambda}\right| > \frac{1}{N^{1/2-\varepsilon}}\right] = o_N(1) ,$$

for some constant $\mathsf{OPT}_{\beta,\lambda}$ that is the largest fixed point of an explicit AMP-like recursion. Here, $\mathsf{OPT}_{\beta,\lambda} > 0$ if $\beta > \frac{1}{\lambda}$ and $\mathsf{OPT}_{\beta,\lambda} = 0$ if $\beta < \frac{1}{\lambda}$.

Furthermore, $\mathsf{OPT}_{\beta,\lambda}$ *characterizes the correlation of a Gibbs sample:*

$$\mathbf{Pr}_{\sigma\sim\mu_{\beta M}}\left[\left|\frac{1}{N}|\langle\boldsymbol{\sigma},\boldsymbol{x}\rangle| - \mathsf{OPT}_{\beta,\lambda}\right| > \frac{1}{N^{1/2-\varepsilon}}\right] \leqslant e^{-cN^{\varepsilon}}\,.$$

Remark 1.6. If one only cared about obtaining a correlation that is within $o_N(1)$ of $\mathsf{OPT}_{\beta,\lambda}$, the formal version of the above result will show that it suffices to set $T = \widetilde{\Omega}(N^3)$.

The key technical innovation in our work is a much more fine-grained understanding of the RGD Markov chain, which is essential for proving the above theorem. As mentioned earlier, the connection between RGD and the Glauber dynamics is present in prior work [LMR+24]. Although the form of their connection breaks when $\beta > 1$, our understanding of RGD extends to this regime. We believe that a connection between RGD and the Glauber dynamics should also exist in this low-temperature regime, but do not attempt to answer this question in the present work.

Towards illustrating our fine-grained understanding of RGD, we present the following theorem, which provides a direct comparison between RGD and AMP. Indeed, this connection is precisely what yields the fixed point equations alluded to in Theorem 1.5.

Theorem 1.7 (Informal, see Section 2.2)**.** *Let* $\lambda \geqslant \beta > 1$ *and* $\varepsilon > 0$ *independent of* N*. There exists some threshold* $\tau \in (0,1)$ *(depending on* β,λ*) such that the following is true.*

Let $x \in \{\pm 1\}^N$ *be an arbitrary spike, and fix* $\sigma \in \{\pm 1\}^N$ *such that* $\frac{1}{N}|\langle \sigma, x\rangle| > \tau$*. Let* $\boldsymbol{\sigma}^{\mathrm{RGD}}$ *be obtained by taking one step of RGD from* σ*, and* $\boldsymbol{\sigma}^{\mathrm{AMP}}$ *be obtained by taking one step of an AMP-like update from* σ*. Then, with high probability over the noise* $\boldsymbol{W}$*, the following holds:*

$$\left|\frac{1}{N}\langle\boldsymbol{\sigma}^{\mathrm{RGD}}, x\rangle - \frac{1}{N}\langle\boldsymbol{\sigma}^{\mathrm{AMP}}, x\rangle\right| = o_N(1)\,.$$

Remark 1.8. The formal version of this theorem depends on verifying an explicit numerical condition which can easily be simulated on a computer; see Section 2.2 for more details.

In other words, if both algorithms are provided a sufficiently correlated warm start, one step of RGD is close to one step of AMP in terms of correlation. In the formal version, the instantiation σ is also allowed to slightly depend on $\boldsymbol{W}$.

One is naturally led to wonder whether a version of Theorem 1.5 holds true for RGD for $\beta > 1$, so that we can relate the correlation of a locally stationary point of RGD to that of a sample from the posterior? Unfortunately, we expect such results to be false from worst-case initializations. However, we *can* still hope for such results from non-worst-case initializations, specifically from warm starts that already have some non-trivial correlation.

Theorem 1.9 (Informal, see Theorem 5.1)**.** *Fix* $\lambda > 1$*,* $K > 0$ *large,* $\varepsilon > 0$ *small, and* $\beta > \frac{1}{\lambda}$*. There exists a threshold* $\tau \in (0,1)$ *such that the following holds. With high probability over the noise* $\boldsymbol{W}$*, the following holds. Let* $\sigma_0 \in \{\pm 1\}^n$*, and assume that* $|R(\sigma_0, \mathbf{1})| > \tau$*. Suppose that we run the RGD Markov chain at inverse temperature* β *from* σ_0 *for* $T \geqslant \omega(\log N)$ *steps to arrive at distribution* ν_T*. Suppose* $T \leqslant N^K$*. Then,*

$$\mathbf{E}_{\sigma_T\sim\nu_T}\left[\left||R\left(\boldsymbol{\sigma}_T, \boldsymbol{x}\right)| - \mathsf{OPT}_{\beta,\lambda}\right|\right] \leqslant O\left(\frac{1}{N^{1/2-\varepsilon}}\right).$$

While the requirement that the initialization has sufficiently large correlation may seem like too strong a requirement, we emphasize that this is not the case. If one ran an annealed version of the RGD Markov chain where β is slowly incremented, the resulting trajectory of correlations is large enough that this lower bound condition is satisfied. Accordingly, we believe this might provide a path to showing that annealed *Glauber* dynamics attains a correlation of $\mathsf{OPT}_{\beta,\lambda}$.

Remark 1.10. Again, the formal version of the above theorem requires establishing two explicit numerical conditions, which can easily be simulated on a computer.

Remark 1.11. We draw attention here to the (perhaps) surprising fact that simulations suggest the correlation $\mathsf{OPT}_{\beta,\lambda}$ with the spike increases *past* $\beta = \lambda$. This is a rather subtle point: while the posterior mean, corresponding to $\mu_{\lambda \boldsymbol{M}}$ (and AMP for that matter) attains the MMSE, this does not imply that samples from the posterior are Bayes-optimal with respect to correlation. What *is* true is that the posterior mean is Bayes optimal for the *normalized squared correlation*, i.e. it is the estimator $\widehat{x}$ which maximizes $\mathbf{E}\left[\frac{\langle \widehat{x}, x\rangle^2}{\|\widehat{x}\|^2}\right]$.

En route to proving the above result, we prove a new high-precision estimate for the mean magnetization for the SK model under weak external field, which may be of independent interest. These estimates enable us to overcome technical difficulties with worst-case analysis of the RGD dynamics for $\beta < 1$.

Theorem 1.12 (Informal, see Lemma 6.1)**.** *Let $\beta < 1$ and $\varepsilon > 0$ sufficiently small. With probability $1 - o(1)$ over $W \sim \mathsf{GOE}(N)$, for all $h = O(N^{-(1/4+\varepsilon)})$,*

$$\frac{1}{N}\mathbf{E}_{\boldsymbol{x} \sim \mu_{\beta W, h\mathbf{1}}} \langle \boldsymbol{x}, \mathbf{1} \rangle = \mathbf{E}_{\boldsymbol{g} \sim \mathcal{N}(0,1)} \left[\tanh\left(\beta \boldsymbol{g}\sqrt{q} + h\right)\right] \left(1 + O\left(\frac{1}{N^{\varepsilon}}\right)\right) + O\left(\frac{1}{N^{1/2+\varepsilon}}\right),$$

where q is the unique solution to $q = \mathbf{E}[\tanh^2(\beta \boldsymbol{g}\sqrt{q} + h)]$.

In particular, this result establishes the correct asymptotics for the mean magnetization up to the phase transition in weak external field established by [DW23].

1.2 Related work

Spiked Wigner model. The spiked Wigner model has provided a rich test bed for average-case algorithms and complexity. Indeed, the basic question of efficient detection and recovery of the spike has been studied through the lens of spectral algorithms [BBP05], low-degree polynomials [PWBM18], and message passing [DAM16]. Interestingly, these efficient algorithms all succeed in polynomial time once $\lambda > 1$. Furthermore, when $\lambda < 1$, the mutual information between the observation $\boldsymbol{M}$ and the signal $\boldsymbol{x}$ turns out to be $o(1)$ [LM17], indicating that one cannot hope to design an estimator $\widehat{x} \in \{\pm 1\}^N$ that is non-trivially correlated with the spike $\boldsymbol{x}$.

For more general priors beyond $\mathrm{Unif}(\{\pm 1\}^N)$, the story becomes significantly richer. For example, there exist sparse priors where there exists a statistical-computational gap: there is an intermediate regime of SNRs (λ_c, λ) where recovery is information theoretically possible, but no known efficient algorithm succeeds at weak recovery. [EAKJ20] pin down the information-theoretic threshold at $\lambda = \lambda_c$ for general bounded i.i.d. priors. The algorithmic side of special cases such as sparse PCA [DM14] and non-negative PCA [MR15] have also been extensively studied. The problem has also been generalized to the tensor setting [MR14, WEM19, BGJ20, KX25], where there is an interesting gap in performance between naive local algorithms and methods based on the sum-of-squares hierarchy or the Kikuchi method.

In terms of sampling from the posterior, due to the sign symmetry of the posterior creating a bottleneck in distributions, Glauber suffers an exponential worst-case mixing time. On the flip side, it has been observed that random initialization or simulated annealing appears to empirically succeed at the goal of sampling from the posterior [DKMZ11]. Nevertheless, it has remained challenging to give provable guarantees about the performance of sampling algorithms, even under algorithmic warm starts. A notable positive sampling result in this direction is [MW23], which uses a different diffusion-based algorithm to sample from the posterior in Wasserstein distance when the SNR λ is a sufficiently large constant.

Bayes optimal estimation. There is by now a large body of work characterizing Bayes optimal estimation for mean-squared error. Indeed, past work has designed algorithms that attain mean-squared error within $\frac{\mathrm{polylog}(N)}{\sqrt{N}}$ of the MMSE [RV18, LW22], as well as generalizations to constant-rank spikes, and other priors [Mio17, MV21]. These aforementioned algorithms are based on AMP—one first computes the top eigenvector of $\boldsymbol{M}$ to obtain a vector that has non-trivial correlation with the spike $\boldsymbol{x}$, then refines this estimate by iteratively applying the AMP update. To our knowledge, these algorithms based on AMP are the only algorithms that are known to obtain a mean-squared error within $o(1)$ of the MMSE in the entire high SNR regime $\lambda > 1$.

Typically, one considers the inference setting where knowledge of the full generative model is assumed. Perhaps confusingly, this is also referred to as the "Bayes-optimal setting". In [AKUZ19], they study a mismatched setting for spiked Wigner where the knowledge of λ is not assumed. In fact, in Section 2.2 we will show that the performance of their AMP algorithm corresponds with the fixed point equations for RGD.

In certain scenarios, Markov chain-based algorithms (or a sample from the posterior) might be preferred to AMP, since they have the ability to capture second-order information, like the variance of the correlation with the spike. Furthermore, samples can be used more flexibly to compute general posterior expectations.

Dynamics of Markov chains before mixing. Many physically and algorithmically interesting phenomena manifest only when Markov chain dynamics do *not* mix quickly, such as *aging* [CK93] and *metastability* [BDH16, BJ24]. In the context of *spin glasses*, physicists have made many predictions about the nonequilibrium behavior of natural Markov chains such as Langevin dynamics and Glauber dynamics. For the pure spherical p-spin glass, the seminal work of Cugliandolo and Kurchan [CK93] derived a heuristic set of integro-differential equations which govern the behavior of Langevin dynamics when run for $O(1)$ time. This was later rigorously proven for $T = O(1)$ by [BDG06] and the dynamics of the energy observable was eventually established for $T = \exp(O(N))$ by [Sel24]. In recent work, [DGPZ25] rigorously prove that the energy and overlap of the sequential scan block dynamics for the SK model are dictated by an analogous set of integro-difference equations for $O(N)$ timescales.

[LSS22] characterized the overlap of Langevin dynamics for spiked matrix posteriors in the large system limit $N \to \infty$ for *fixed* times t. The analogous result in discrete time would correspond to asymptotically controlling the overlap for $O(N)$ steps of Glauber dynamics. In contrast, we give control on any sufficiently large polynomial time scale for Glauber. [BGP24] give an extremely precise characterization of trajectory of Langevin diffusion for the posterior of the multi-spike tensor PCA problem given polynomially many samples of the spiked tensor (the spike being fixed, and the noise tensor i.i.d. across all samples). Concretely, in the rank-one matrix setting, they establish strong recovery of the spike $\boldsymbol{x}$ when the SNR $\lambda = N^{\alpha}$ for any $\alpha > 0$. Their proofs are based on careful analysis of the SDE governing the correlations, which are inaccessible in the discrete setting.

Locally stationary distributions. Before a recent work of Liu, Mohanty, Raghavendra, Rajaraman, and Wu [LMR+24], it was not rigorously known whether Glauber dynamics on the spiked Wigner posterior achieves nontrivial recovery of the spike. As mentioned earlier in the introduction, they introduce the framework of locally stationary distributions to analyze the performance of slow-mixing Markov chains on inference problems such as spiked Wigner.

Our results directly improve on theirs in the following concrete ways:

(i) We achieve tight temperature and SNR thresholds for our analysis, conditional on widely believed mixing conditions for the Sherrington–Kirkpatrick (SK) model.

(ii) Wherever SK mixing holds, we nail down the exact correlation that Glauber dynamics achieves for any SNR λ, as the solution to a simple fixed point equation. In particular, widely believed mixing conditions for the SK model recover the BBP transition $\lambda > 1$.

(iii) In particular, we unconditionally recover weak recovery in the regime $\lambda > \frac{1}{0.295}$ for the success of Glauber, based on the rigorous SK mixing results of [AKV24].

(iv) As in Remark 1.6, we improve the runtime guarantees by an N^2 factor.

Locally stationary distributions have since been used in other settings to understand the behavior of Markov chains before mixing. Notably, [BCV25] introduces a definition of local stationarity for quantum Hamiltonians, and uses it to show that all metastable states satisfy certain structural guarantees. [RSS+25] also find an interesting application of locally stationary distributions in the context of stochastic backtracking algorithms for LLM reasoning.

1.3 Organization

In Section 2, we provide a technical overview of our techniques to understand the behavior of RGD, and its connection to AMP. In Section 3, we list a few open questions that seem important to fleshing out our understanding of RGD and Glauber. In Section 4, we cover some basic preliminaries that will be useful.

In Section 5, we prove our main theorems about restricted Gaussian dynamics. Next, in Section 6, we establish novel high-precision estimates for the mean magnetization of a high-temperature SK model under weak external field, which are crucial for analyzing RGD. Finally, in Section 7, we describe some results analyzing the one-dimensional recursion governing the behavior of RGD.

2 Technical Overview

Our main results crucially depend on a very precise understanding of the restricted Gaussian dynamics. As it turns out, its study transparently clarifies the fixed point structure and its relationship to AMP. Using the machinery of locally stationary distributions [LMR+24, Lemma 3.7], we can then transfer our understanding of RGD to Glauber dynamics; see Theorem 5.2 for more details.

The technical overview is organized as follows. We will first define RGD and describe the high-level reduction of why one should expect RGD to boil down to a one-dimensional recursion on correlation. Then, in Section 2.1, we show how spin glass machinery can be leveraged to explicitly understand what this one-dimensional recursion is. Once we have written down the explicit recursion, we can readily connect it to AMP; this is carried out in Section 2.2. Finally, we describe some qualitative properties of this recursion, such as the nature of its fixed points, in Section 2.3.

First off, observe that due to the rotational symmetry of the noise matrix and prior, we may assume without loss of generality that the spike is the all-ones vectors $\mathbf{1}$. In this case, the scaled posterior (2) of the spiked Wigner model becomes

$$\mu_{\beta \boldsymbol{M}}(\sigma) \propto \exp\left(\frac{\beta}{2}\sigma^\top \boldsymbol{M}\sigma\right) = \exp\left(\frac{\beta}{2}\sigma^\top \boldsymbol{W}\sigma + \frac{\beta\lambda}{2N}\langle \sigma, \mathbf{1}\rangle^2\right).$$

We start by explaining where RGD arises from, and the connection between the posterior and the SK model. It arises from the Hubbard–Stratonovich transform [Str57, Hub59] (also see [LMRW24, Theorem 3.12]). Let $\boldsymbol{\sigma} \sim \mu_{\beta W + \frac{\beta\lambda}{N}\mathbf{1}\mathbf{1}^\top}$, and $\boldsymbol{g} \sim \mathcal{N}(0,1)$. Set $\boldsymbol{z} = \frac{\beta\lambda}{N}\langle \boldsymbol{\sigma}, \mathbf{1}\rangle + \sqrt{\frac{\beta\lambda}{N}}\boldsymbol{g}$. Then, the posterior distribution on $\boldsymbol{\sigma}$ conditioned on $\boldsymbol{z}$ is simply

$$\begin{aligned}\mathbf{Pr}[\sigma \mid z] &\propto \mu_{\beta \boldsymbol{M}}(\sigma)\cdot \exp\left(-\frac{N}{2\beta\lambda}\left(z - \frac{\beta\lambda}{N}\langle \sigma, \mathbf{1}\rangle\right)\right)\\ &\propto \exp\left(\frac{\beta}{2}\sigma^\top \boldsymbol{W}\sigma + z\langle \sigma, \mathbf{1}\rangle\right) = \mu_{\beta \boldsymbol{W}, z\mathbf{1}},\end{aligned}$$

and the quadratic term from the spike in the Hamiltonian is now an external field. In other words,

$$\mu_{\beta W + \frac{\beta\lambda}{N}\mathbf{1}\mathbf{1}^\top} = \mathbf{E}_{\boldsymbol{z}}\mu_{\beta W, \boldsymbol{z}\mathbf{1}}.$$

The RGD Markov chain (Definition 1.3) is then simply the noising-denoising process associated to the above measure decomposition: if at $\sigma \in \{\pm 1\}^N$, draw $\boldsymbol{g} \sim \mathcal{N}(0,1)$, then move to $\boldsymbol{\sigma}' \sim \mu_{\beta \boldsymbol{W}, \left(\frac{\beta\lambda}{N}\langle \sigma, \mathbf{1}\rangle + \sqrt{\frac{\beta\lambda}{N}}\boldsymbol{g}\right)\mathbf{1}}$. It is not difficult to show, given this perspective of a noising-denoising process, that RGD has stationary distribution equal to $\mu_{\beta \boldsymbol{M}}$. The usefulness of RGD is that it brings forth the importance of the correlation with the spike, which is the observable we ultimately care about.

Reduction to one dimension. So far, we have written down the N-dimensional RGD chain, whose evolution explicitly depends on the correlation with the spike. To make RGD more tractable and eventually recover the AMP update, we would like to reduce the dynamics to a finite-dimensional one.

This motivates us to consider the related P^{RGD} Markov chain on $[-1,1]$, which tracks the (normalized) correlation of the N-dimensional RGD iterate with the spike. In this Markov chain, if starting at $y \in \mathbb{R}$, we draw $\boldsymbol{g} \sim \mathcal{N}(0,1)$, set $\boldsymbol{z} = \beta\lambda y + \sqrt{\frac{\beta\lambda}{N}}\boldsymbol{g}$, draw $\boldsymbol{\sigma} \sim \mu_{\beta \boldsymbol{W}, \boldsymbol{h}\mathbf{1}}$, and finally move to $\boldsymbol{z}' = \frac{1}{N}\langle \boldsymbol{\sigma}, \mathbf{1}\rangle$. The stationary distribution of this Markov chain is just the (normalized) marginal of $\mu_{\beta \boldsymbol{M}}$ along the all-ones direction.

So far, this 1-dimensional chain still goes through an N-dimensional SK Gibbs sample $\boldsymbol{\sigma}$; in particular, it depends on the $N \times N$ matrix $\boldsymbol{W}$.Therefore, to reduce P^{RGD} to an effective 1-dimensional recursion, we must establish two things:

- Concentration of $\boldsymbol{z}'$ around the mean correlation. More explicitly, we would like to show that with high probability, the disorder $\boldsymbol{W}$ is such that for all the encountered external fields $\boldsymbol{z}$, the correlation of the random draw $\boldsymbol{\sigma}$ is concentrated around the mean correlation $\langle \boldsymbol{\sigma} \rangle$.
- A scalar formula for the mean correlation of a Gibbs sample from an SK model with external field.

Both of these points are tractable using spin glass machinery in the high temperature regime for SK, which we describe shortly. As we shall soon see, the scalar formula that governs the mean correlation is what crucially leads to state evolution equations.

2.1 The mean recursion and fixed points structure

Let us now see more concretely how the fixed point relation arises from the analysis of P^{RGD}.

Define the mean magnetization $\boldsymbol{m}_h = \frac{1}{N}\mathbf{E}_{\sigma \sim \mu_{\beta W,h}}\langle \boldsymbol{\sigma}, \mathbf{1}\rangle$, and for $\boldsymbol{\sigma} \sim \mu_{\beta W,h}$, let $\widetilde{\boldsymbol{g}}_h$ be the random variable $\sqrt{N} \cdot \left(\frac{1}{N}\langle \boldsymbol{\sigma}, \mathbf{1}\rangle - \boldsymbol{m}_h\right)$. Under this notation, we can write down the following description of P^{RGD}: from $z \in \mathbb{R}$,

1. draw $\boldsymbol{g} \sim \mathcal{N}(0,1)$, set $\boldsymbol{h} = \beta\lambda z + \sqrt{\frac{\beta\lambda}{N}}\boldsymbol{g}$,
2. move to $\boldsymbol{z}' = \boldsymbol{m}_{\boldsymbol{h}} + \frac{1}{\sqrt{N}}\widetilde{\boldsymbol{g}}_{\boldsymbol{h}}$.

For simplicity of exposition, let us first assume $z = \Omega(1)$; we will comment on the case of small z later. First, observe by gaussian concentration, with very high probability we have $\boldsymbol{h} \approx \beta\lambda z$.

To simplify the recursion further, we need concentration of $\boldsymbol{z}'$ around $\boldsymbol{m}_h$, as well as an explicit formula for $\boldsymbol{m}_h$. The key observation is that such properties are tractable to prove if the SK model $\mu_{\beta W,h}$ is in the high temperature regime. We first need to introduce some standard spin glass quantities.

Definition 4.11 (Overlap constant)**.** For $\beta \geqslant 0$ and $h > 0$, we define $q = q_{\beta,h}$ as the unique solution to the recursion

$$q = \mathbf{E}_{\boldsymbol{g} \sim \mathcal{N}(0,1)}\left[\tanh^2\left(\beta \boldsymbol{g}\sqrt{q} + h\right)\right].$$

We also define $q_{\beta,h} = 0$ for $h = 0$.

We can now define the high-temperature region for the SK model, first predicted by de Almeida and Thouless [dAT78].[2]

[2] It is not rigorously known whether the AT line exactly characterizes the high temperature regime. However, this has been numerically verified, so in the interest of exposition we brush this issue aside for now.

Definition 4.12 (AT line)**.** A non-negative pair (β, h) is said to satisfy the Almeida–Thouless (AT) condition, or to be above the AT line, if

$$\beta^2 \mathbf{E}\,\mathrm{sech}^4\Big(\beta z\sqrt{q_{\beta,h}} + h\Big) < 1\,. \tag{AT}$$

We show that if (AT) is satisfied, the correlation $\langle \boldsymbol{\sigma}, \mathbf{1}\rangle$ of a Gibbs sample $\boldsymbol{\sigma} \sim \mu_{\beta W,h}$ concentrates around the mean correlation. In other words, we have $\boldsymbol{z}' \approx \boldsymbol{m_h}$ with high probability. Hence, we conclude that a single step of P^{RGD} can be approximated by the map $z \mapsto \boldsymbol{m}_{\beta\lambda z}$.

Turning now to the mean formula, in Lemma 6.1, we establish that with high probability over $\boldsymbol{W}$, for all $h = O(1)$ where (AT) is satisfied,

$$\boldsymbol{m}_h \approx \mathbf{E}_{g\sim\mathcal{N}(0,1)}\big[\tanh\Big(\beta \boldsymbol{g}\sqrt{q_{\beta,h}} + h\Big)\big], \tag{3}$$

where we recall q from Definition 4.11. In particular, we establish tighter control on $\boldsymbol{m}_h$ when $h = o(1)$ than is implied by existing literature, which is crucial for analyzing the dynamics near $h = 0$.

Thus, we have reduced a single step of P^{RGD} to the deterministic iteration

$$z_{t+1} \leftarrow \mathbf{E}_{g\sim\mathcal{N}(0,1)}\left[\tanh\left(\beta \boldsymbol{g}\sqrt{q} + \beta\lambda z_t\right)\right]\,,$$

where $q = q_{\beta,\beta\lambda z_t}$.

This suggests that the limiting behavior of P^{RGD} is given by fixed point solutions to the following set of equations:

$$\begin{aligned} z &= \mathbf{E}_{g\sim\mathcal{N}(0,1)}\left[\tanh\left(\beta \boldsymbol{g}\sqrt{q} + \beta\lambda z\right)\right]\,,\\ q &= \mathbf{E}_{g\sim\mathcal{N}(0,1)}\left[\tanh^2\left(\beta \boldsymbol{g}\sqrt{q} + \beta\lambda z\right)\right]\,. \end{aligned} \tag{4}$$

Whether P^{RGD} provably converges to these fixed points is not completely clear, and we defer further discussion on this matter to Section 2.3.

Let us now address the case of small z. When $z = o(1)$, and especially if z is $\Theta(\frac{1}{\sqrt{N}})$, the picture is more delicate, as then the random quantities fluctuate at the same scale as z itself. In Theorem 5.1 we use the anticoncentration of $\boldsymbol{g}$ to show that in $\Omega(\log N)$ steps, we reach the region $z = \Omega(1)$ with constant probability.

2.2 RGD simulates AMP

Let us now draw the explicit connection to AMP fixed points. The state evolution fixed points (see Lemma 4.10 for a precise statement) that govern the late-stage performance of Bayes AMP for spiked Wigner [DAM16, MV21] read

$$\begin{aligned} \mu &= \lambda \mathbf{E}_{g\sim\mathcal{N}(0,1)}\big[\tanh\big(\lambda \boldsymbol{g}\sigma + \lambda^2\sigma^2\big)\big]\\ \sigma^2 &= \mathbf{E}_{g\sim\mathcal{N}(0,1)}\Big[\tanh^2\big(\lambda \boldsymbol{g}\sigma + \lambda^2\sigma^2\big)\Big]\,. \end{aligned}$$

Let (μ, σ^2) be the unique fixed point to the SE equations. We claim that $(z, q) = (\frac{\mu}{\lambda}, \sigma^2)$ is also a solution to (4) for $\beta = \lambda$. Indeed, using that $\mathbf{E}\tanh\big(\sqrt{\gamma}\boldsymbol{g} + \gamma\big) = \mathbf{E}\tanh^2\big(\sqrt{\gamma}\boldsymbol{g} + \gamma\big)$ for any $\gamma \geqslant 0$, we see that $\mu = \lambda\sigma^2$. Therefore our putative assignment satisfies $z = q$, and some straightforward algebra yields that (z, q) is a valid solution to the fixed point equations.

As it turns out, the update equations for RGD given by (4) are nearly identical to the state evolution equations for mismatched AMP for the spiked Wigner model [AKUZ19]. Specifically, we have $\Delta_0 = \frac{1}{\lambda^2}$, $\Delta = \frac{1}{\beta\lambda}$, $\eta_t(\cdot) = \tanh(\cdot)$. Then some straightforward algebra, combined with [AKUZ19, Theorem 1], yields the SE recursion

$$\begin{aligned} M_{t+1} &= \mathbf{E}\left[\tanh\left(\beta \boldsymbol{g}\sqrt{Q_t} + \beta\lambda M_t\right)\right] \\ Q_{t+1} &= \mathbf{E}\left[\tanh^2(\beta \boldsymbol{g}\sqrt{Q_t} + \beta\lambda M_t)\right]. \end{aligned}$$

The direct comparison is just identifying $M_t \leftarrow z_t$ and $Q_t \leftarrow q_t$. The only difference is that (M_t, Q_t) update synchronously, whereas our recursion sets $q_t = q_t(z_t)$ and therefore collapses into an effective one-dimensional recursion on just z_t.

Remark 2.1. A subtle point here is that even for $\beta = \lambda$, the mismatched AMP is *not* syntactically equivalent to Bayes AMP; in order to derive Bayes AMP, one must artificially enforce that $z = q$.

2.3 Qualitative behavior of the fixed point equations

The qualitative behavior of the correlation recursion is determined by the fixed points of the recursion, as well as the convergence to said fixed points.

Convergence is nontrivial to guarantee, and indeed the proof of convergence for Bayes AMP does not immediately generalize to the mismatched setting where $\beta \neq \lambda$. To this end, in Proposition 7.1 we establish that so long as $(\beta, \beta\lambda z)$ satisfy (AT), the map $z \mapsto \boldsymbol{m}_{\beta\lambda z}$ is strictly increasing in z. For any $\beta < 1$, (AT) is automatically satisfied, and in this setting we are able to fully characterize convergence for the recursion starting from any z.

Turning now to the fixed points, one can immediately identify $z = 0$ as one fixed point of this recursion. As it turns out, this map $z \mapsto \boldsymbol{m}_{\beta\lambda z}$ behaves differently depending on β, as illustrated in Figure 2:

(i) If β is too small (smaller than $\frac{1}{\lambda}$), the signal is washed out and 0 is the only fixed point of this recursion. Furthermore, it is a stable fixed point.

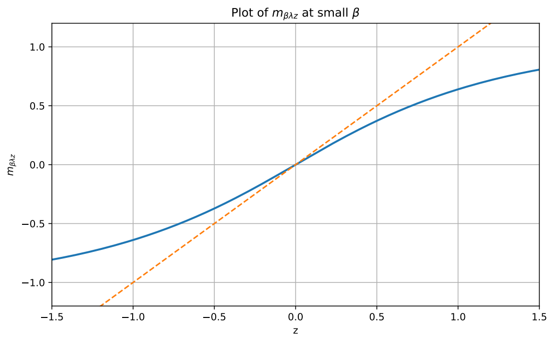
(ii) If β is of intermediate magnitude (between $\frac{1}{\lambda}$ and 1), in addition to an unstable fixed point at 0, this function has stable fixed points at $\pm\mathsf{OPT}_{\beta,\lambda}$.

(iii) If $\beta > 1$, the formula for the mean correlation $\boldsymbol{m}_{\beta\lambda z}$ is only valid if z is sufficiently large. Indeed, if z is close to 0, (AT) is violated, and the corresponding SK model is known to be in the low-temperature replica-symmetry breaking (RSB) regime. Hence, the recursion is only valid for large z, and in this region there are at most two stable fixed points $\pm\mathsf{OPT}_\beta$.

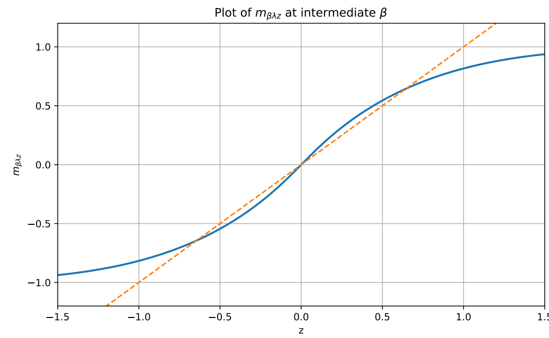
(iv) Finally, if β is very large, the RSB region for z is so large that it swallows the fixed points that are present at smaller β.

We make the above predictions for the convergence and fixed points rigorous in Lemma 7.5. Figure 2 suggests two kinds of results one could hope to show.

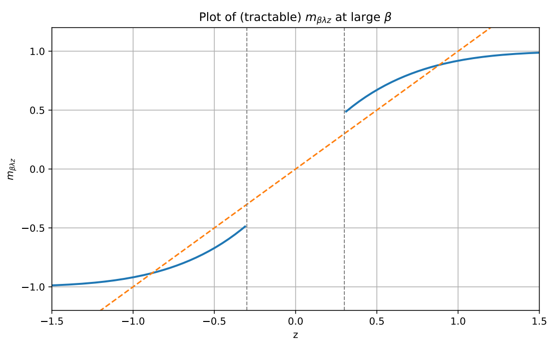
(a) First, that for any β, the RGD Markov chain reaches the stable OPT_β fixed point from a sufficiently warm start which puts z_0 in the high-temperature, replica-symmetric (RS) regime.



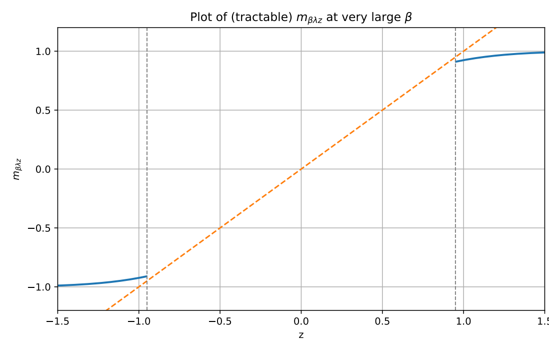
(a) $\boldsymbol{m}_{\beta\lambda z}$ at $\beta < \frac{1}{\lambda}$: 0 is the only fixed point.



(b) $\boldsymbol{m}_{\beta\lambda z}$ at intermediate $\frac{1}{\lambda} < \beta < 1$: non-trivial fixed points appear.



(c) $\boldsymbol{m}_{\beta\lambda z}$ at $1 < \beta < \lambda$: the 0 fixed point is swallowed by intractability.



(d) $\boldsymbol{m}_{\beta\lambda z}$ at $\beta \gg \lambda$: all fixed points are swallowed by intractability.

Figure 2: Behavior of the RGD recursion for different values of β.

That is, if the initialization z_0 of P^{RGD} is such that $(\beta, \beta\lambda z_0)$ satisfies (AT), then the correlation shoots to the stable fixed point OPT_β in a small constant number of steps of RGD. This is essentially immediate from classical analyses of fixed point recursions, see Lemma 5.4 for details.

(b) Second, that for $\beta < 1$, the RGD Markov chain reaches the stable OPT_β fixed point from an *arbitrary* initialization. In the case that $\beta < \frac{1}{\lambda}$, 0 is the only stable fixed point, so for now we assume that $\beta \in (\frac{1}{\lambda}, 1)$. Since the previous point implies that the correlation goes to OPT_β once it becomes sufficiently large, the interesting part here is showing that the Markov chain successfully escapes the unstable fixed point at 0. We establish this in Lemma 5.6, where we show that the noise $\boldsymbol{g}$ and $\widetilde{\boldsymbol{g}}$, which are typically $O\left(\frac{1}{\sqrt{N}}\right)$, are anticoncentrated enough to cause the correlation to escape 0.

Remark 2.2. As a simple illustrative example, we can carry out the same analysis on the RGD dynamics for the Curie-Weiss model at inverse temperature β and external field h. In this case $\boldsymbol{m_h}$ can be computed exactly as $\tanh(\boldsymbol{h})$, and so we see that RGD converges to the stable fixed point of the mean-field recursion $m = \tanh(\beta m + h)$.

3 Conjectures and open questions

One natural question is whether we can extend our analysis to other inference problems. At least for the case of spiked matrix problems, the general recipe seems fairly adaptable. For example, consider a more general spiked matrix inference problem with a different prior and noise model (i.e. not GOE). The main obstacles here are (1) establishing replica symmetry (specifically overlap concentration) for this more general spin glass, and using it to show sample magnetization concentration à la Lemma 6.2 and (2) proving a mean correlation formula, analogous to Lemma 6.1.

Question 3.1. *Can one establish a kind of universality result for our fixed point characterization that does not rely on model-specific calculations?*

It would also be interesting to see if one could push the limits of our analysis to the multi-spike setting and achieve results analogous to [BGP24] for Glauber dynamics.

Another interesting (but open-ended) question is: what other classes of inference problems can be analyzed using locally stationary distributions? The success of RGD in the present analysis is rather contingent on the spiked matrix structure; in other inference problems, it is unclear what Markov chain should be used as a proxy to understand Glauber dynamics.

Question 3.2. *Identify broader inference problems that can be analyzed through the lens of locally stationary distributions.*

We next turn to open questions regarding the spiked Wigner problem in particular. One fundamental barrier to applying our techniques to understand Glauber for $\beta > 1$ is that the transfer of local stationarity (Lemma 4.8) does not work in this setting, even if we assume mixing guarantees in the entire conjecture high-temperature region (AT). In particular, it is unclear how to prove a version of Lemma 4.8 where we only assume that the component measure satisfies an MLSI with high probability, under the assumption that the initialization from which the locally stationary distribution is obtained lives in this high-temperature region.

Question 3.3. *Assuming mixing for SK everywhere where (β, h) satisfy (AT), can we extend our analysis to characterize the recovery performance for annealed Glauber dynamics up to $\beta \in (1, \beta_\lambda)$ for some $\beta_\lambda > \lambda$?*

It seems likely that if the above question were answered affirmatively, this would provide an alternate algorithm that attains Bayes-optimal performance in the entire high-SNR regime $\lambda > 1$.

We note that [AKUZ19] makes some predictions for the behavior of these models in the RSB regime, i.e. beyond β_λ.

For technical reasons, the current techniques which involve a transfer of local stationarity between Glauber and RGD necessarily incur some $\mathrm{poly}(N)$ overhead in the runtime. On the other hand, based on physics heuristics and the rigorous results of [Sel24, DGPZ25], one would expect the dynamics to achieve nonzero correlation in even $O(N)$ steps.

Conjecture 3.4. *Glauber dynamics succeeds at weak recovery in (near-)linear runtime.*

One path towards achieving such a guarantee is to establish exponential contraction of the local stationarity parameter, perhaps by establishing a weak Poincaré inequality [HMRW24] for symmetric test functions. Indeed, we believe that our techniques to analyze RGD might help in showing that annealed Glauber dynamics succeeds at sampling from the posterior (and potentially even from lower-temperature models, up to the aforementioned threshold $\beta_\lambda > \lambda$). Indeed, while the "conservation of variance" property used to show sampling guarantees for annealed Glauber dynamics is typically reliant on proving certain covariance bounds (e.g. [HMRW24, Lemma 6.8]) which is often technically challenging, perhaps this can be sidestepped if one instead directly tried to establish a weak Poincaré inequality for the (one-dimensional!) RGD Markov chain.

Conjecture 3.5. *For $\beta \leqslant \beta_\lambda$, annealed Glauber dynamics samples from the scaled posterior $\mu_{\beta \boldsymbol{M}}$ in* $\mathrm{poly}(N)$*, or even $O(N \log N)$ time.*

This threshold β_λ is defined as the largest temperature β at which there is a nonzero fixed point z to the recursion (4) such that $(\beta, \beta\lambda z)$ satisfies (AT). As shown in Proposition 7.1, this threshold is in fact greater than λ.

4 Preliminaries

4.1 Notation

- Given $\sigma_1, \sigma_2 \in \{\pm 1\}^N$, we denote $R(\sigma_1, \sigma_2) = \frac{1}{N}\langle \sigma_1, \sigma_2 \rangle$ to be the normalized correlation between the two.
- We use c to denote small positive constants whose value may change from line to line, and C to denote similarly fickle large constants.

4.2 Markov chain basics

Definition 4.1 (Ising model)**.** Let $J \in \mathbb{R}^{n \times n}$ be a symmetric *interaction matrix* and $h \in \mathbb{R}^n$ an *external field*. The *Ising model* corresponding to J and h is the probability distribution $\mu_{J,h}$ on $\{\pm 1\}^n$, where

$$\mu_{J,h}(x) \propto \exp\left(\frac{1}{2} x^\top J x + \langle h, x \rangle\right).$$

We drop the h from the subscript when it is equal to 0.

We also recall the standard definitions of functional inequalities which imply rapid mixing. For any functions $f, g : \Omega \to \mathbb{R}$, define the *Dirichlet form* $\mathcal{E}(f, g) := \mathbf{E}_{\boldsymbol{x} \sim \pi} \mathbf{E}_{\boldsymbol{y} \sim P\boldsymbol{x}} (f(\boldsymbol{x}) - f(\boldsymbol{y}))(g(\boldsymbol{x}) - g(\boldsymbol{y}))$.

Definition 4.2 (Poincaré inequality)**.** We say P satisfies a *Poincaré inequality* with constant C if for any function $f : \Omega \to \mathbb{R}$:

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

Definition 4.3 (Modified log-Sobolev inequality)**.** We say P satisfies a *modified log-Sobolev inequality* (MLSI) with constant C if for any function $f : \Omega \to \mathbb{R}_{>0}$,

$$\mathcal{E}_P(f, \log f) \geqslant C \cdot \mathrm{Ent}[f]\,,$$

where $\mathrm{Ent}[f] := \mathbf{E}_\pi[f \log f] - \mathbf{E}_\pi f \log \mathbf{E}_\pi f$.

4.3 Locally stationary distributions

We now give the relevant definitions for locally stationary distributions.

Definition 4.4 ([LMR+24, Definition 1.2])**.** Let P be a time-reversible ergodic Markov chain on a finite state space Ω, with associated stationary distribution π. A probability measure ν on Ω is said to be ε-*locally stationary* with respect to P if

$$\mathcal{E}\left(f, \log f\right) := \sum_{x,y\in\Omega} P_{x\to y}\left(f(x) - f(y)\right) \cdot \log \frac{f(x)}{f(y)} \leqslant \varepsilon\,.$$

One can also show that at a typical time t, *any* reversible Markov chain yields a locally stationary distribution.

Lemma 4.5. *Let ν_0 be an arbitrary distribution, $\nu_t = P_t\nu_0$, and set $f_t = \frac{\mathrm{d}\nu_t}{\pi}$. Also set $f = \mathbf{E}_{t\sim[0,T]} f_t$. Then,*

$$\mathcal{E}\left(f, \log f\right) \leqslant \frac{\log(1/\mathrm{KL}(\nu\|\pi))}{T} \leqslant \frac{\log(1/\pi_{\min})}{T}\,.$$

In particular, $\nu := \mathbf{E}_{t\sim[0,T]}\nu_t$ is $\frac{\log(1/\pi_{\min})}{T}$-locally stationary with respect to π.

Proof. This immediately follows from [LMR+24, Lemma 3.1] (also see [BCV25, Lemma II.1]), coupled with the observation that the Dirichlet form $f \mapsto \mathcal{E}(f, \log f)$ is convex. □

We will also need to use the fact that a locally stationary distribution does not change much in TV distance if we apply P for a few steps.

Lemma 4.6 (TV stability, essentially [LMR+24, Lemma 3.2])**.** *Let P be a reversible Markov chain with stationary distribution π, ν an arbitrary distribution which is ε-locally stationary. Then,*

$$d_{\mathrm{TV}}\left(\nu, P^T\nu\right) \leqslant \frac{T\sqrt{\varepsilon}}{2}.$$

Proof. When $T = 1$, this is [LMR+24, Lemma 3.2]. The data processing inequality implies that for any $k \geqslant 1$, $d_{\mathrm{TV}}\left(P^k\nu, P^{k+1}\nu\right) \leqslant \frac{\sqrt{\varepsilon}}{2}$. The claim follows on applying the triangle inequality. □

Corollary 4.7. *Let $\phi : \Omega \to \mathbb{R}$ be a bounded function on the state space of a Markov chain P, and ν an ε-locally stationary measure. Then,*

$$\left|\mathbf{E}_{\boldsymbol{x}\sim\nu}[\phi(\boldsymbol{x})] - \mathbf{E}_{\boldsymbol{x}\sim P^T\nu}[\phi(\boldsymbol{x})]\right| \leqslant \|\phi\|_\infty \cdot T\sqrt{\varepsilon}.$$

We will also need to transfer local stationarity between Glauber and RGD.

Lemma 4.8 ([LMR+24, Lemma 3.7])**.** *Let P be the Markov chain associated to a measure decomposition $\pi = \mathbf{E}_{z\sim\rho}\pi_z$. Let $f : \{\pm1\}^n \to \mathbb{R}_{>0}$ be any function and set τ such that $\min_{x\in\{\pm1\}^n} f(x) > \exp(-\tau)$ or $\max_{x\in\{\pm1\}^n} f(x) < \exp(\tau)$. For $\delta := \inf_z \rho_{\mathrm{LS}}(\pi_z)$, we have*

$$\mathcal{E}_P(f, \log f) \leqslant O\left(\frac{\tau}{\delta}\right) \cdot \mathcal{E}_\pi(f, \log f)\,.$$

4.4 Approximate Message Passing and State Evolution

For simplicity, we only recall the specific version of AMP used for spiked Wigner [DAM16].

Definition 4.9. Set $m^{-1} = x^0 = 0$, and for $t \geqslant 0$ we update

$$\begin{aligned} m^t &= \tanh(x^t)\,, \\ \mathsf{b}_t &= \frac{\lambda^2}{N}\sum_{i=1}^{N}(1-\tanh^2(x_i^t)) \\ x^{t+1} &= \lambda \boldsymbol{M} m^t - \mathsf{b}_t m^{t-1} \end{aligned}$$

where $\tanh(\cdot)$ is applied entrywise.

State evolution [BM11] predicts the empirical distribution of the coordinates of x_t. We record the special case relevant to our setting: the correlation of the AMP iterate with the planted signal $\boldsymbol{x}$ can be predicted by a scalar recursion.

Lemma 4.10 (State evolution for correlation)**.** *Fix a time $t \geqslant 0$. Let (μ_t, σ_t) be defined by the scalar recursion initialized with $(\mu_0, \sigma_0) = (0, 0)$:*

$$\begin{aligned} \mu_{t+1} &= \lambda \mathbf{E}_{g\sim\mathcal{N}(0,1)}[\tanh(\lambda \boldsymbol{g}\sigma_t + \lambda^2\sigma_t^2)] \\ \sigma_{t+1}^2 &= \mathbf{E}_{g\sim\mathcal{N}(0,1)}[\tanh^2(\lambda \boldsymbol{g}\sigma_t + \lambda^2\sigma_t^2)]\,. \end{aligned}$$

Let $X_0 \sim \mathrm{Unif}\{\pm 1\}, Z_0 \sim \mathcal{N}(0,1)$. As $N \to \infty$, we have the following convergence in W_2:

$$\frac{1}{N}\langle m^t, \boldsymbol{x}\rangle \to \mathbf{E}[X_0(\mu_t X_0 + \sigma_t Z_0)]\,.$$

4.5 The Sherrington–Kirkpatrick model

In this subsection, we introduce several important parameters in the SK model.

Definition 4.11 (Overlap constant)**.** For $\beta \geqslant 0$ and $h > 0$, we define $q = q_{\beta,h}$ as the unique solution to the recursion

$$q = \mathbf{E}_{g\sim\mathcal{N}(0,1)}\left[\tanh^2\left(\beta \boldsymbol{g}\sqrt{q} + h\right)\right]\,.$$

We also define $q_{\beta,h} = 0$ for $h = 0$.

It is known that in the high-temperature replica-symmetric phase $\beta < 1$, the overlap between two samples from the corresponding SK model concentrates around the above number.

Definition 4.12 (AT line)**.** A non-negative pair (β, h) is said to satisfy the Almeida–Thouless (AT) condition, or to be above the AT line, if

$$\beta^2 \mathbf{E}\,\mathrm{sech}^4\left(\beta z\sqrt{q_{\beta,h}} + h\right) < 1\,. \tag{AT}$$

Observe that if $\beta < 1$, (β, h) satisfies (AT) for any $h \in \mathbb{R}_{\geqslant 0}$. A very important question in the study of the SK model is when the model is replica-symmetric: when does the overlap between two independent samples from the Gibbs measure concentrate?

Definition 4.13. We say that (β, h) satisfies *exponential overlap concentration* if for some constant C independent of N, with high probability over $\boldsymbol{W} \sim \mathsf{GOE}(N)$,

$$\mathbf{E}_{\sigma^1,\sigma^2 \sim \mu_{\beta W, h1}} \left[\exp \left(\frac{N \left(\frac{1}{N} \langle \sigma^1, \sigma^2 \rangle - q_{\beta,h} \right)^2}{C} \right) \right] \leqslant 2 \,. \tag{overlap-conc}$$

It is believed that the replica-symmetric region coincides with the set of (β, h) satisfying (AT).

Condition 2 (RS-AT)**.** *β is said to satisfy* (RS-AT) *if the following holds with high probability over $\boldsymbol{W}$. Let $h_* = \inf\{h \geqslant 0 : (\beta, h)$ satisfies* (AT)*$\}$. Then for all $h > h_*$, $\mu_{\beta \mathbf{W},h}$ satisfies* (overlap-conc)*.*

Remark 4.14. At a high level, this condition is saying that being above the AT line implies a strong form of replica symmetry, and that the AT region is monotonic in h. The latter condition has been numerically verified, but the authors are unaware of a rigorous statement of this form.

Remark 4.15. Significant progress has been made towards establishing that for the SK model, (β, h) satisfies (AT) if and only if the corresponding model is replica-symmetric. On the negative side, it was shown by Toninelli [Ton02] that if (β, h) does *not* satisfy (AT), the corresponding model is replica-symmetric breaking. On the positive side, it was shown by Talagrand [Tal11, Theorem 13.7.1] that exponential overlap concentration holds for a very large subregion of the area above the AT line, which is conjectured to coincide with the entirety of the region above the AT line. Making this conjecture rigorous requires proving an explicit numerical condition. We refrain from elaborating as the details are not directly relevant to our work, and instead refer the interested reader to Talagrand's text, as well as the discussion in [JT17, p.8]. We remark that it is known that Talagrand's subregion contains the entire high-temperature regime $\beta < 1$, see e.g. [BXY23, p.3] and [Che21].

5 The dynamics of RGD

The goal of this section will be to understand the trajectory of RGD—we will establish that RGD rapidly converges to fixed points of the one-dimensional recursion of interest. Recall from the technical overview that for $\lambda > 1$ and $\beta > 0$, this recursion is given by the update function f_β defined by

$$f_{\beta,\lambda}(z) = \mathbf{E}\left[\tanh\left(\beta g \sqrt{q_{\beta,\beta\lambda z}} + \beta\lambda z\right)\right] \,. \tag{5}$$

When λ is clear from context, we refer to $f_{\beta,\lambda}$ as simply f_β.

Often in this section, we will make the following assumption about the structure of f_β.

Condition 3 (AT-fixed-pt)**.** *For $\lambda > 1$ and $\beta > 0$, we say that (β, λ) satisfies* (AT-fixed-pt) *if there exists a fixed point $\mathsf{OPT}_{\beta,\lambda}$ of $f_{\beta,\lambda}$ such that $(\beta, \beta\lambda\mathsf{OPT}_{\beta,\lambda})$ satisfies* (AT)*.*

This condition has some direct consequences about the recursion, which we state below in Lemma 5.3. One might wonder when (AT-fixed-pt) holds; in fact from the definition of the AT line one immediately concludes that it holds for $\beta < 1$. In Lemma 7.6, we are also able to show that it holds at the posterior temperature where $\beta = \lambda$. It is an interesting open question whether one can rigorously prove that (AT-fixed-pt) holds for $\beta \in (1, \lambda)$.

Let us now state the main result for this section.

Theorem 5.1 (Fixed-temperature RGD)**.** *Fix $\lambda > 1$, $K > 0$ large, $\varepsilon > 0$ small, and $\beta > \frac{1}{\lambda}$. With high probability over the noise $\boldsymbol{W}$, the following holds. Let $x_0 \in \{\pm 1\}^n$, and assume that $(\beta, \beta\lambda R(x_0, \mathbf{1}))$ satisfies (AT), β satisfies (RS-AT), and (β, λ) satisfies (AT-fixed-pt). Suppose that we run the RGD Markov chain at inverse temperature β from x_0 for $T \geqslant \omega(\log N)$ steps to arrive at distribution ν_T. Suppose $T \leqslant N^K$. Then,*

$$\mathbf{E}_{\boldsymbol{x}_T\sim\nu_T}\left[\left||R\left(\boldsymbol{x}_T,\mathbf{1}\right)| - \mathsf{OPT}_{\beta,\lambda}\right|\right] \leqslant O\left(\frac{1}{N^{1/2-\varepsilon}}\right).$$

Combining Theorem 5.1 with the properties of locally stationary distributions, we obtain the following control over the recovery performance of Glauber when $\beta < 1$.

Theorem 5.2 (Fixed-temperature Glauber)**.** *Let $\beta \in (\frac{1}{\lambda}, 1)$ satisfy (MLSI) and $\varepsilon > 0$. With high probability over $\boldsymbol{W}$, the following holds. Let $\boldsymbol{\sigma}_0 \in \{\pm 1\}^N$ be arbitrary, and let $(\boldsymbol{\sigma}_t)_{t\geqslant 0}$ be the trajectory of the Glauber dynamics run with stationary distribution $\mu_{\beta\boldsymbol{M}}$ initialized at $\boldsymbol{\sigma}_0$. Let $T \geqslant \widetilde{\Omega}(N^{4-2\varepsilon})$, and $\widehat{\boldsymbol{\sigma}} = \boldsymbol{\sigma}_{\boldsymbol{t}}$ for a uniformly random $\boldsymbol{t} \sim [0, T]$. Then, there exists an explicit constant $\mathsf{OPT}_{\beta,\lambda} > 0$ such that*

$$\mathbf{Pr}\left[\left|\frac{1}{N}\left|\langle\widehat{\boldsymbol{\sigma}}, \boldsymbol{x}\rangle\right| - \mathsf{OPT}_{\beta,\lambda}\right| > \frac{C}{N^{1/2-\varepsilon}}\right] = o_N(1)\,.$$

Proof. Let us condition on the very high probability event that $\|\boldsymbol{W}\|_{\mathrm{op}} = O(1)$. Let ν denote the distribution $\mathbf{E}_{\boldsymbol{t}\sim[0,T]}\nu_{\boldsymbol{t}}$, and f its relative density with respect to $\mu_{\beta\boldsymbol{M}}$. By Lemma 4.5, we have that $\mathcal{E}(f, \log f) \leqslant O(\frac{N}{T})$. Applying Lemma 4.8 yields that $\mathcal{E}_{\mathrm{RGD}}(f_t, \log f_t) \leqslant O(\frac{\tau}{\delta}\cdot\frac{N}{T})$. We can crudely bound $\tau = O(N)$ because $\|\beta\boldsymbol{M}\|_{\mathrm{op}} = O(1)$, and since we assumed that β satisfies (MLSI), we have $\delta = \Omega(\frac{1}{N})$, yielding that

$$\mathcal{E}_{\mathrm{RGD}}(f_t, \log f_t) \leqslant O\left(\frac{N^3}{T}\right).$$

We can now apply Corollary 4.7 to ν with the test function $\phi(x) = \left||R(x,\mathbf{1})| - \mathsf{OPT}_{\beta,\lambda}\right|$, to conclude that

$$\left|\mathbf{E}_{\nu}\left[\left|R(\boldsymbol{x},\mathbf{1}) - \mathsf{OPT}_{\beta,\lambda}\right|\right] - \mathbf{E}_{(P^{\mathrm{RGD}})^{C\log N}\nu}\left[\left|R(\boldsymbol{x},\mathbf{1}) - \mathsf{OPT}_{\beta,\lambda}\right|\right]\right| \leqslant \widetilde{O}\left(\sqrt{\frac{N^3}{T}}\right).$$

Now since $\beta < 1$, all three of the conditions (AT), (RS-AT), and (AT-fixed-pt) are known to hold. Therefore Theorem 5.1 yields the appropriate control on the latter expectation, and we conclude by applying Markov. $\square$

We shall prove Theorem 5.1 in two steps. First, in Lemma 5.4, we shall prove the claim assuming a warm start initialization where $|R(x_0, \mathbf{1})| = \Omega(1)$; such a warm start is necessary for $\beta > 1$ since the AT condition is nontrivial there. We then show in Lemma 5.6 that when $\beta < 1$, RGD escapes the unstable fixed point where $R(x_0, \mathbf{1}) = 0$.

For use later in this section, we also note the following consequences of (AT-fixed-pt), proved in Section 7.

Lemma 5.3. *Let $\lambda > 1$. If $\beta < 1$ or $\beta = \lambda$, (β, λ) satisfies (AT-fixed-pt). Suppose that (β, λ) satisfies (AT-fixed-pt). Then,*

(i) *if h is such that $(\beta, \beta\lambda h)$ satisfies (AT), $f'_{\beta,\lambda}(h) > 0$,*

(ii) *$\mathsf{OPT}_{\beta,\lambda}$ is the unique fixed point h of $f_{\beta,\lambda}$ such that $(\beta, \beta\lambda h)$ satisfies (AT), and*

(iii) *$\mathsf{OPT}_{\beta,\lambda}$ is a stable fixed point of $f_{\beta,\lambda}$.*

5.1 RGD from a warm start reaches the stable fixed point

In this section, we will show that running RGD on the posterior of spiked Wigner attains the Gibbs correlation, provided that the Markov chain is initialized at a point (or distribution) with typically large correlation. Our analysis will require RGD to be run for at least a sufficiently large constant number of steps, but the analysis extends to any large polynomial number of steps. In contrast, existing nonasymptotic analyses of AMP only work for sublinearly many steps.

Lemma 5.4. *Fix $\lambda > 1$, $K > 0$ large, $\varepsilon > 0$ small, and $\beta > \frac{1}{\lambda}$. With high probability over the noise $\mathbf{W}$, the following holds. Let $x_0 \in \{\pm 1\}^n$, and assume that $(\beta, \beta\lambda|R(x_0, \mathbf{1})|)$ satisfies (AT), and $|R(x_0, \mathbf{1})| = \Omega(1)$. Also assume that β satisfies (RS-AT) and (β, λ) satisfies (AT-fixed-pt).[3] Suppose that we run the RGD Markov chain at inverse temperature β from x_0 for $T \geqslant \omega(\log N)$ steps to arrive at distribution ν_T. Suppose $T \leqslant N^K$. Then,*

$$\mathbf{E}_{\boldsymbol{x}_T \sim \nu_T}\left[\left||R\left(\boldsymbol{x}_T, \mathbf{1}\right)| - \mathsf{OPT}_{\beta,\lambda}\right|\right] \leqslant O\left(\frac{1}{N^{1/2-\varepsilon}}\right).$$

Remark 5.5. If one only desired to have correlation within a(n arbitrarily small) constant of OPT_β, our proof (specifically (7)) will show that it suffices to run RGD for a (large) constant number of steps.

Proof. For ease of notation, let $c > 0$ be a sufficiently small constant such that (β, c) satisfies (AT), and let

$$\mathcal{H} = \{h \in [-2\beta\lambda, 2\beta\lambda] : (\beta, h) \text{ satisfies (AT) and } |h| > c\} .$$

Observe that $\mathcal{H}$ contains $\beta\lambda R(x_0, \mathbf{1})$. Consider the one-dimensional projection of the RGD walk, initialized at $z_0 = R(x_0, \mathbf{1})$. One step of this one-dimensional walk is defined by walking from $\boldsymbol{z}_t$ as

$$\boldsymbol{z}_{t+1} = R(\boldsymbol{x}_{t+1}, \mathbf{1}) \,, \text{ where } \boldsymbol{x}_{t+1} \sim \mu_{\beta W, \beta\lambda \boldsymbol{z}_t + \sqrt{\frac{\beta\lambda}{n}} \boldsymbol{g}_t} ,$$

where each $\boldsymbol{g}_t$ is iid drawn from $\mathcal{N}(0,1)$. Let us condition on the following high-probability events:

(i) For every $t < T$, $|\boldsymbol{g}_t| < N^\varepsilon$. By standard Gaussian tail bounds, this occurs with probability $T \cdot e^{-\Omega(N^\varepsilon)} = e^{-cN^\varepsilon}$.

(ii) For all $h \in \mathcal{H}$ and constant k,

$$\mathbf{E}_{\sigma \sim \mu_{\beta W, h}} \left(R(\boldsymbol{\sigma}, \mathbf{1}) - \mathbf{E}\left[\tanh\left(\beta\sqrt{q_{\beta,h}}\boldsymbol{g} + h\right)\right]\right)^{2k} < \frac{1}{N^k} .$$

This follows from Lemma 6.2, along with the assumption that β satisfies (RS-AT).

We claim that, on the above events, for all $t < T$, $\beta\lambda \boldsymbol{z}_t$ is bounded away from the AT line by a constant. The base case $t = 0$ is true by assumption. Next, since $\beta\lambda \boldsymbol{z}_t$ is bounded away from the AT line by a constant, (i) implies that $\beta\lambda \boldsymbol{z}_t + \sqrt{\frac{\beta\lambda}{n}} \boldsymbol{g}_t$ does so as well. (ii) then implies that with high probability, for all $t < T$,

$$\left| \boldsymbol{z}_{t+1} - f_\beta\left(\boldsymbol{z}_t + \sqrt{\frac{1}{\beta\lambda N}} \boldsymbol{g}_t\right) \right| \leqslant O\left(\frac{1}{N^{1/2-\varepsilon}}\right) .$$

[3] Here and below, satisfying (AT) means we satisfy the AT line inequality with some constant gap

The Lipschitzness of $f_{\beta,\lambda}$ with (i) implies that

$$\left|z_{t+1} - f_\beta(z_t)\right| \leqslant O\left(\frac{1}{N^{1/2-\varepsilon}}\right). \tag{6}$$

Note in particular that by the monotonicity assumption in (RS-AT), if the above is true, then $\beta\lambda z_{t+1}$ also satisfies (AT) with a constant gap. Indeed, by Lemma 5.3(i), if $z_t < \mathsf{OPT}_{\beta,\lambda}$, we have $f_\beta(z_t) > z_t$ so it satisfies (AT). On the other hand, if $z_t > \mathsf{OPT}_{\beta,\lambda}$, then $f_\beta(z_t) > \mathsf{OPT}_{\beta,\lambda}$, so it satisfies (AT) due to (AT-fixed-pt). As a result, we conclude that for all $t < T$, using f_β is valid as an approximation to the RGD update at step t.

Next, we show that z_t converges to a small constant interval around OPT_β in a constant number of steps. For starters, since we have assumed (AT-fixed-pt), Lemma 5.3(iii) implies that $\mathsf{OPT} = \mathsf{OPT}_{\beta,\lambda}$ is the unique nonzero fixed point of f_β in the AT region, and that it is stable. Let $\eta > 0$ such that f'_β is bounded by some $L < 1$ on $[\mathsf{OPT} - \eta, \mathsf{OPT} + \eta]$.

There exists a large constant C and random variables $(\delta_t)_{0\leqslant t\leqslant T}$ such that $\delta_t < \frac{C}{N^{1/2-\varepsilon}}$ almost surely for all $t < T$ and $z_{t+1} = f_\beta(z_t + \delta_t)$. We claim that after some constant number of steps τ, $z_\tau \in [\mathsf{OPT} - \eta/2, \mathsf{OPT} + \eta/2]$. Indeed, set

$$\text{gap} = \inf_{h\in\mathcal{H}\setminus[\mathsf{OPT}-\eta/2,\mathsf{OPT}+\eta/2]} \left|f_\beta(h) - h\right|,$$

a constant bounded away from 0 due to the definition of $\mathcal{H}$ and the boundedness of f'_β near OPT. Then, while $z_t \in \mathcal{H} \setminus [\mathsf{OPT} - \eta/2, \mathsf{OPT} + \eta/2]$, since f_β is increasing (Lemma 5.3(i)), we have

$$|z_{t+1} - \mathsf{OPT}| < |z_t - \mathsf{OPT}| - \text{gap}.$$

Furthermore, once z_t reaches $[\mathsf{OPT} - \eta/2, \mathsf{OPT} + \eta/2]$, because $\delta_t = o(1)$, it is easy to check that $f'_\beta(z_t) < L$ for all future times $t > \tau$. For all such times, we have

$$|z_{t+1} - \mathsf{OPT}| = \left|f_\beta(z_t + \delta_t) - f_\beta(\mathsf{OPT})\right| \leqslant L\,|z_t + \delta_t - \mathsf{OPT}| \leqslant L|z_t - \mathsf{OPT}| + L \cdot \frac{C}{N^{1/2-\varepsilon}}.$$

Iterating this inequality yields that

$$|z_{\tau+t} - \mathsf{OPT}| \leqslant L^t\,|z_\tau - \mathsf{OPT}| + \frac{L}{1-L} \cdot \frac{C}{N^{1/2-\varepsilon}}. \tag{7}$$

It follows that once we set $T = \Omega(\log N)$,

$$|z_T - \mathsf{OPT}| \leqslant O\left(\frac{1}{N^{1/2-\varepsilon}}\right)$$

as desired. □

5.2 High-temperature RGD escapes the unstable fixed point

This section is dedicated to proving the following lemma about RGD escaping from the trivial fixed point at 0 correlation.

Lemma 5.6. *Let $\lambda > 1$, $\beta \in (\frac{1}{\lambda}, 1)$, and $\varepsilon > 0$ be a sufficiently small constant. Let $x_0 \in \{\pm 1\}^n$ arbitrary, and $(x_t)_{t \geqslant 0}$ the (random) trajectory of RGD initialized at x_0. Then,*

$$\mathbf{Pr}\left[\max_{0 \leqslant t \leqslant T} \frac{1}{N} |\langle x_T, \mathbf{1} \rangle| \leqslant \varepsilon\right] \leqslant e^{-\Omega(T/\log N)}.$$

Lemma 5.6 almost immediately follows from the following more general lemma about appropriate stochastic processes escaping unstable fixed points.

Lemma 5.7. *Let $g : \mathbb{R} \to \mathbb{R}$ be a smooth function such that for some constants $L > 1$ and $\varepsilon > 0$, $|g(x)| > L|x|$ for all $|x| < 2\varepsilon$. Let $X_0 \sim \nu_0$ for some arbitrary distribution ν_0 on $\mathbb{R}$. Suppose we have three sequences $(X_t, Y_t, Z_t)_{t \geqslant 0}$ of random variables on a common probability space. Assume that*

(i) *for all $t \geqslant 0$, Z_t has law $\mathcal{N}\left(0, \Theta\left(\frac{1}{N}\right)\right)$, and is independent of $\sigma((X_s)_{0 \leqslant s \leqslant t-1}, (Y_s)_{0 \leqslant s \leqslant t})$,*

(ii) $\mathbf{E}\left[Y_t^2 \mid (X_s, Y_s, Z_s)_{0 \leqslant s \leqslant t-1}\right] \leqslant \frac{C}{N}$.

(iii) *(X_t) is defined by*
$$X_{t+1} = g(X_t) + Y_t + Z_t.$$

Then,
$$\mathbf{Pr}\left[\max_{0 \leqslant t \leqslant T} |X_t| \leqslant \varepsilon\right] \leqslant e^{-\Omega(T/\log N)}.$$

Proof. Our proof proceeds in two stages. We first show that the dynamics escape zero to a "lukewarm" start, and after this exponential growth kicks in to make X_t grow to $\Omega(1)$.

To be more precise, let $\kappa > 0$ be a (large) constant to be fixed later. We shall show the following two statements, where $c > 0$ is a sufficiently small constant:

$$\mathbf{Pr}\left[|X_{t+1}| \geqslant \frac{\kappa}{\sqrt{N}} \mid (X_s, Y_s, Z_s)_{0 \leqslant s \leqslant t}\right] \geqslant c, \tag{8}$$

$$\mathbf{Pr}\left[\max_{t \leqslant s \leqslant t + C\log N} |X_s| \geqslant \varepsilon \mid (X_s, Y_s, Z_s)_{0 \leqslant s \leqslant t-1}, |X_t| \geqslant \frac{\kappa}{\sqrt{N}}\right] \geqslant c. \tag{9}$$

Putting these two together with the independence structure of the (Y_s, Z_s) immediately yields the result.

Let us start by proving (8), which is a consequence of anticoncentration and concentration of Z_t. Indeed, we have

$$\begin{aligned}
&\mathbf{Pr}\left[|X_{t+1}| \geqslant \frac{\kappa}{\sqrt{N}} \mid (X_s, Y_s)_{0 \leqslant s \leqslant t}, (Z_s)_{0 \leqslant s \leqslant t-1}, |g(X_t) + Y_t| \geqslant \frac{2\kappa}{\sqrt{N}}\right] \\
&\qquad \geqslant \mathbf{Pr}\left[|_t| \leqslant \frac{\kappa}{\sqrt{N}} \mid (X_s, Y_s)_{0 \leqslant s \leqslant t}, (Z_s)_{0 \leqslant s \leqslant t-1}, |g(X_t) + Y_t| \geqslant \frac{2\kappa}{\sqrt{N}}\right] \geqslant c, \text{ and} \\
&\mathbf{Pr}\left[|X_{t+1}| \geqslant \frac{\kappa}{\sqrt{N}} \mid (X_s, Y_s)_{0 \leqslant s \leqslant t}, (Z_s)_{0 \leqslant s \leqslant t-1}, |g(X_t) + Y_t| \leqslant \frac{2\kappa}{\sqrt{N}}\right] \\
&\qquad \geqslant \mathbf{Pr}\left[|Z_t| \geqslant \frac{3\kappa}{\sqrt{N}} \mid (X_s, Y_s)_{0 \leqslant s \leqslant t}, (Z_s)_{0 \leqslant s \leqslant t-1}, |g(X_t) + Y_t| \leqslant \frac{2\kappa}{\sqrt{N}}\right] \geqslant c,
\end{aligned}$$

implying (8).

Let us next prove (9). Consider the following event, parametrized by some constant $\iota > 0$ that we shall fix later:

$$\mathcal{E} = \left\{ (Y_s, Z_s)_{s \geqslant t} : |Y_s + Z_s| \leqslant \frac{\iota\kappa}{\sqrt{N}} \cdot (s - t + 1) \right\} .$$

Observe that if $\iota\kappa$ is sufficiently large, the above holds with nonzero probability. Indeed, we have $\mathbf{E}\left[(Y_t + Z_t)^2 \mid (Y_s, Z_s)_{0 \leqslant s \leqslant t-1} \right] \leqslant \frac{C}{N}$, so Markov's inequality followed by a union bound gives that

$$\mathbf{Pr}\left[\mathcal{E} \mid (X_s, Y_s, Z_s)_{0 \leqslant s \leqslant t-1} \right] \geqslant 1 - \sum_{j \geqslant 0} \frac{C}{\iota^2 \kappa^2 (j+1)^2} , \tag{10}$$

which is bounded away from 0 for $\iota\kappa$ sufficiently large.

Now, towards (9), we claim the following for sufficiently small ι and some constant $\widetilde{L} \in (1, L)$ bounded away from 1 (to be set later): if $\mathcal{E}$ holds, $|X_{t+j}| \leqslant \varepsilon$, and $|X_{t+j}| \geqslant \widetilde{L}^j \cdot \frac{\kappa}{\sqrt{N}}$, then $|X_{t+j+1}| \geqslant \widetilde{L}^{j+1} \cdot \frac{\kappa}{\sqrt{N}}$. It is not difficult to see that this implies (9): all the X_s being smaller than ε would contradict the exponential growth prescribed by the previous sentence if the constant C is taken large enough that $\widetilde{L}^{C \log N} \cdot \frac{\kappa}{\sqrt{N}} \geqslant \varepsilon$. To complete the proof, we have that if the events described earlier hold,

$$\begin{aligned} |X_{t+j+1}| &= |g(X_{t+j}) + Y_{t+j} + Z_{t+j}| \\ &\geqslant |g(X_{t+j})| - |Y_{t+j} + Z_{t+j}| \\ &\geqslant L \cdot \widetilde{L}^j \cdot \frac{\kappa}{\sqrt{N}} - \frac{\iota\kappa}{\sqrt{N}} \cdot (j+1) \\ &\geqslant \widetilde{L}^j \cdot \frac{\kappa}{\sqrt{N}} \cdot \left(L - \iota \cdot \sup_{j \geqslant 0} \frac{j+1}{\widetilde{L}^j} \right) . \end{aligned}$$

To conclude, choose the constants such that

- $\widetilde{L} = \frac{L+1}{2}$ is bounded away from 1,
- ι is small enough that $L - \iota \cdot \inf_{j \geqslant 0} \frac{j+1}{\widetilde{L}^j} \geqslant \widetilde{L}$, and
- κ is large enough that the probability in (10) is bounded away from 0. □

Let us go back and prove Lemma 5.6.

Proof of Lemma 5.6. We shall prove a similar statement not for the correlation, but for the external fields $(\boldsymbol{h}_t)_{t \geqslant 0}$ encountered along the trajectory of RGD (see Item 2). Concretely, this process is defined as follows, starting from arbitrary $h = \boldsymbol{h}_0 \in \mathbb{R}$.

1. sample $\boldsymbol{z} = f_{\beta,\lambda}(h) + \frac{1}{\sqrt{N}} \widetilde{\boldsymbol{g}}_h$, where we recall $\widetilde{\boldsymbol{g}}_h$ is the deviation from the expected magnetization of a Gibbs sample from $\mu_{\beta \boldsymbol{W}, h}$.
2. draw $\boldsymbol{g} \sim \mathcal{N}\left(0, \frac{1}{\beta\lambda N}\right)$, and move to $\boldsymbol{h}' = \boldsymbol{z} + \boldsymbol{g}$.

Lemma 6.1 yields that with probability $1 - o(1)$, we may write

$$\boldsymbol{h}_{t+1} = f_{\beta,\lambda}(\boldsymbol{h}_t) \left(1 + O\left(\frac{1}{N^{\varepsilon}} \right) \right) + O\left(\frac{1}{N^{1/2+\varepsilon}} \right) + \widetilde{\boldsymbol{g}}_{\boldsymbol{h}_t} + \boldsymbol{g}_t .$$

Let us verify that we can apply Lemma 5.7 with $X_t = \boldsymbol{h}_t$, $Y_t = O\left(\frac{1}{N^{1/2+\varepsilon}}\right) + \widetilde{g}_h$, and $Z_t = g_t$.

By definition, g_t is independent of the collection $\{\widetilde{g}_{h_t}, (\widetilde{g}_{h_s}, g_s)_{0 \leqslant s \leqslant t-1}\}$. The bound on the conditional second moment of $\widetilde{g}_h$ follows from Lemma 6.2—recall that for $\beta < 1$, as mentioned in Remark 4.15, (β, h) satisfies (overlap-conc) for all h. The lower bound of $|f_{\beta,\lambda}(x)| \left(1 + O\left(\frac{1}{N^\varepsilon}\right)\right) > L|x|$ for small x follows from the instability of $f_{\beta,\lambda}$ at 0 for $\beta \in \left(\frac{1}{\lambda}, 1\right)$: this is proved in Proposition 7.1. The desideratum follows on using Lemma 5.7. $\square$

We may now prove Theorem 5.1, restated for convenience.

Theorem 5.1 (Fixed-temperature RGD)**.** *Fix $\lambda > 1$, $K > 0$ large, $\varepsilon > 0$ small, and $\beta > \frac{1}{\lambda}$. With high probability over the noise $\mathbf{W}$, the following holds. Let $x_0 \in \{\pm 1\}^n$, and assume that $(\beta, \beta\lambda R(x_0, \mathbf{1}))$ satisfies (AT), β satisfies (RS-AT), and (β, λ) satisfies (AT-fixed-pt). Suppose that we run the RGD Markov chain at inverse temperature β from x_0 for $T \geqslant \omega(\log N)$ steps to arrive at distribution ν_T. Suppose $T \leqslant N^K$. Then,*

$$\mathbf{E}_{\boldsymbol{x}_T \sim \nu_T}\left[\left||R(\boldsymbol{x}_T, \mathbf{1})| - \mathsf{OPT}_{\beta,\lambda}\right|\right] \leqslant O\left(\frac{1}{N^{1/2-\varepsilon}}\right).$$

Proof. If $\beta > 1$, this immediately follows from Lemma 5.4, since the AT line is non-trivial there. For $\beta < 1$, this follows from Lemmas 5.4 and 5.6. $\square$

6 High-temperature mean magnetization estimates in the SK model

In the RGD recursion for spiked Wigner, the measure decomposition is

$$\mu_{\beta M,0} = \mathbf{E}_{\substack{x \sim \mu_{\beta M,0} \\ g \sim N(0,1)}}\left[\mu_{\beta W, (\frac{\beta\lambda}{n}\langle \mathbf{1}, x\rangle + \sqrt{\beta\lambda/n}\, g)\mathbf{1}}\right].$$

The following explicit formula allows us to determine the deterministic scalar recursion which determines the behavior of locally stationary RGD, in the high-temperature regime $\beta < 1$.

Lemma 6.1 (High-temperature magnetization concentration for all fields)**.** *Let $\beta < 1$, $\varepsilon > 0$ sufficiently small, and $K > 0$ large independent of N. With probability $1 - o(1)$ over $W \sim \mathsf{GOE}(N)$, for all $0 \leqslant h \leqslant K$,*

(i) if $h < N^{-(1/4+\varepsilon)}$,

$$\frac{1}{N}\mathbf{E}_{\boldsymbol{x} \sim \mu_{\beta W, h\mathbf{1}}}\langle \boldsymbol{x}, \mathbf{1}\rangle = \mathbf{E}_{g \sim \mathcal{N}(0,1)}\left[\tanh\left(\beta g\sqrt{q} + h\right)\right]\left(1 + O\left(\frac{1}{N^\varepsilon}\right)\right) + O\left(\frac{1}{N^{1/2+\varepsilon}}\right).$$

(ii) if $h > N^{-(1/4+\varepsilon)}$,

$$\frac{1}{N}\mathbf{E}_{\boldsymbol{x} \sim \mu_{\beta W, h\mathbf{1}}}\langle \boldsymbol{x}, \mathbf{1}\rangle = \mathbf{E}_{g \sim \mathcal{N}(0,1)}\left[\tanh\left(\beta g\sqrt{q} + h\right)\right] + O\left(\frac{1}{N^{1/2-\varepsilon}}\right).$$

where $q = q_h$ is the functional order parameter defined by

$$q = \mathbf{E}_{g \sim \mathcal{N}(0,1)}\left[\tanh^2\left(\beta g\sqrt{q} + h\right)\right].$$

Note that for $h > N^{-(1/4+\varepsilon)}$, a bound of the form of (i) is strictly stronger than that in (ii).

The above theorem will be a corollary of two estimates for the mean magnetization. The first is essentially already present in [Han07] (also see [CT21]).

Lemma 6.2 (essentially [Han07])**.** *Let (β, h) satisfy (overlap-conc). Then, for any constant k,*

$$\mathbf{E}\left[(R(\sigma,\mathbf{1}) - q_1)^{2k}\right] \leqslant O\left(\frac{1}{N^k}\right),$$

where we denote $q_1 = \mathbf{E}_{g\sim\mathcal{N}(0,1)}\left[\tanh\left(\beta g\sqrt{q} + h\right)\right]$.

The proof is identical to that of Corollary 4.1, Theorem 5.1, and Theorem 6.1 in [Han07]—we observe that all these proofs of upper bounds on the moments go through assuming overlap concentration, even if we do not have a central limit theorem for the overlap (such a CLT *is* required for the stronger Theorem 1.2 in Hanen's paper, which proves a CLT for the magnetization).

However, this lemma is not useful in the regime where h is tiny, say $\Theta\left(N^{-1/2}\right)$, since q_1 itself is $O(h)$, and the error is at the same scale as the estimate. To get around this, we prove a mean magnetization formula that is more precise in the small h regime, albeit only with a lower moment that will suffice for our purposes.

Lemma 6.3 (Magnetization estimates under weak external field)**.** *Let $\delta > 0$ and $h < N^{-\alpha}$ for some $\alpha > 1/4$. Then,*

$$\mathbf{E}\left[\left|\frac{1}{N}\sum_{i=1}^{N}\langle\sigma_i\rangle - q_1\right|^{2-\delta}\right] \leqslant O\left(\max\left\{\frac{1}{N}, \frac{h}{\sqrt{N}}\right\}\right)^{2-\delta}.$$

We will defer the proof of Lemma 6.3 to Section 6.1. We may now prove Lemma 6.1.

Proof. For ease of notation, let $\boldsymbol{m}(h)$ be the random variable $\frac{1}{N}\sum_{i=1}^{N}\langle\sigma_i\rangle$ under external field h. It is not difficult to show that $\boldsymbol{m}(h)$ is non-decreasing in h: indeed, its derivative with respect to h is the variance of the magnetization in the corresponding model. For $h > N^{-\left(\frac{1}{4}+\varepsilon\right)}$, Markov's inequality applied to Lemma 6.2 yields that

$$\mathbf{Pr}\left[|\boldsymbol{m}(h) - q_1(h)| > \frac{1}{N^{\frac{1}{2}-\varepsilon}}\right] < \frac{1}{N^2}. \tag{11}$$

On the other hand, for $h < N^{-\left(\frac{1}{4}+\varepsilon\right)}$, Markov's inequality applied to Lemma 6.3 yields that for some c_1 we will fix shortly,

$$\mathbf{Pr}\left[|\boldsymbol{m}(h) - q_1(h)| > \max\left\{\frac{h}{N^{\frac{1}{2}-\frac{c_1}{2-\delta}}}, \frac{1}{N^{1-\frac{c_1}{2-\delta}}}\right\}\right] < O\left(N^{-c_1}\right).$$

Now, choose $\delta = \varepsilon$ and c_1 such that $\frac{c_1}{2-\varepsilon} = \frac{1}{2} - \varepsilon$, so appealing to Lemma 7.4, the above reduces to

$$\mathbf{Pr}\left[|\boldsymbol{m}(h) - q_1(h)| > \max\left\{\frac{q_1(h)}{N^{\varepsilon}}, \frac{1}{N^{1/2+\varepsilon}}\right\}\right] < O\left(N^{-c_1}\right). \tag{12}$$

For $c_2 = \frac{1}{2} + \varepsilon$, let $S = \left\{q_1^{-1}\left(iN^{-c_2}\right) : 0 \leqslant i \leqslant KN^{c_2}\right\}$; note that S is well defined because q_1 is strictly increasing in h. We shall perform a union bound over S. For sufficiently small ε, it is plainly true that $N^{-c_1} \cdot N^{c_2} = o(1)$, since $c_1 \approx 1$ and $c_2 \approx \frac{1}{2}$. Consequently, (12) yields that with probability $1 - o(1)$, for all $h \in S$ with $h < N^{-\left(\frac{1}{4}+\varepsilon\right)}$,

$$|\boldsymbol{m}(h) - q_1(h)| < O\left(\max\left\{\frac{q_1(h)}{N^\varepsilon}, \frac{1}{N^{1/2+\varepsilon}}\right\}\right).$$

On the other hand, (11) yields that with probability $1 - o(1)$, for all $h \in S$ with $h > N^{-\left(\frac{1}{4}+\varepsilon\right)}$,

$$|\boldsymbol{m}(h) - q_1(h)| < \frac{1}{N^{1/2-\varepsilon}}.$$

This bound further extends from S to all h bounded by K: indeed, suppose $h \leqslant K$ and $h_1 < h < h_2$ such that $h_1, h_2 \in S$, $|q_1(h_1) - q_1(h)| < N^{-c_2}$ and $|q_1(h_2) - q_1(h)| < N^{-c_2}$. Then, if $h < N^{-\left(\frac{1}{4}+\varepsilon\right)}$,

$$\boldsymbol{m}(h) \geqslant \boldsymbol{m}(h_1) \geqslant q_1(h_1) - O\left(\frac{q_1(h_1)}{N^\varepsilon}\right) - O\left(\frac{1}{N^{1/2+\varepsilon}}\right) \geqslant q_1(h) - O\left(\frac{q_1(h)}{N^\varepsilon}\right) - O\left(\frac{1}{N^{1/2+\varepsilon}}\right),$$

where we have used the fact that q_1 is strictly increasing in h. Similarly, we have

$$\boldsymbol{m}(h) \leqslant q_1(h) + O\left(\frac{q_1(h)}{N^\varepsilon}\right) + O\left(\frac{1}{N^{1/2+\varepsilon}}\right).$$

An identical argument works for the alternate error bound when $h > N^{-\left(\frac{1}{4}+\varepsilon\right)}$. □

6.1 High-precision mean magnetization estimates under weak external field

For this section, let $h < N^{-\alpha}$ for some $\alpha > 1/4$, and $\beta < 1$. We also define q, q_1 by the recursions

$$q = \mathbf{E}_{z\sim\mathcal{N}(0,1)}\left[\tanh^2\left(\beta\sqrt{q}z + h\right)\right],$$
$$q_1 = \mathbf{E}_{z\sim\mathcal{N}(0,1)}\left[\tanh\left(\beta\sqrt{q}z + h\right)\right].$$

We also denote Z to be the partition function of the SK model with external field h.

In this section, we will prove the following lemma.

Lemma 6.3 (Magnetization estimates under weak external field)**.** *Let $\delta > 0$ and $h < N^{-\alpha}$ for some $\alpha > 1/4$. Then,*

$$\mathbf{E}\left[\left|\frac{1}{N}\sum_{i=1}^N \langle\sigma_i\rangle - q_1\right|^{2-\delta}\right] \leqslant O\left(\max\left\{\frac{1}{N}, \frac{h}{\sqrt{N}}\right\}\right)^{2-\delta}.$$

Lemma 6.4. $\mathbf{E}\left[Z^2\left(\frac{1}{N}\sum_{i=1}^N \langle\sigma_i\rangle - q_1\right)^2\right] \leqslant \mathbf{E}[Z]^2 \cdot O\left(\max\left\{\frac{1}{N}, \frac{h}{\sqrt{N}}\right\}\right)^2.$

Proof of Lemma 6.3. We have by Hölder's inequality and [DW23, Corollary 3.5] (using the fact that $\alpha > 1/4$) that

$$\mathbf{E}\left[\left|\frac{1}{N}\sum_{i=1}^N \langle\sigma_i\rangle - q_1\right|^{2-\delta}\right] \leqslant \mathbf{E}\left[Z^2 \cdot \left(\frac{1}{N}\sum_{i=1}^N \langle\sigma_i\rangle - q_1\right)^2\right]^{\frac{2-\delta}{2}} \cdot \mathbf{E}\left[\frac{1}{Z^{4/\delta-2}}\right]^{\delta/2}$$

$$\leqslant \mathbf{E}[Z]^{2-\delta} \cdot O\left(\max\left\{\frac{1}{N}, \frac{h}{\sqrt{N}}\right\}\right)^{2-\delta} \cdot \frac{1}{\mathbf{E}[Z]^{2-\delta}}$$
$$= O\left(\max\left\{\frac{1}{N}, \frac{h}{\sqrt{N}}\right\}\right)^{2-\delta},$$

as desired. □

We dedicate the remainder of this section to proving Lemma 6.4. Using replicas, we can write

$$\mathbf{E}\left[Z^2\left(\frac{1}{N}\sum_{i=1}^{N}\langle\sigma_i\rangle - q_1\right)^2\right] = \sum_{\sigma,\rho}\mathbf{E}[e^{H_N(\sigma)+H_N(\rho)}(m(\sigma)-q_1)(m(\rho)-q_1)],$$

where $H_N(\sigma) = \frac{\beta}{2}\langle\sigma, \mathbf{W}\sigma\rangle + h\langle\sigma, 1\rangle$, so that $\mathbf{Cov}(H_N(\sigma), H_N(\rho)) = \frac{\beta^2}{2}NR(\sigma,\rho)^2$. It follows that $H_N(\sigma)+H_N(\rho)$ is distributed as $\mathcal{N}\left(h\langle\sigma+\rho,\mathbf{1}\rangle, N\beta^2\left(1+R(\sigma,\rho)^2\right)\right)$, and so,

$$\mathbf{E}\left[e^{H_N(\sigma)+H_N(\rho)}\right] = e^{\frac{N\beta^2}{2}\left(1+R(\sigma,\rho)^2\right)+h\langle\sigma+\rho,\mathbf{1}\rangle}.$$

Since the summand in fact has no dependence on the disorder, we have reduced the moment calculation to an evaluation of the double sum

$$\sum_{\sigma,\rho\in\{\pm1\}^N} e^{\frac{N\beta^2}{2}(1+R(\sigma,\rho)^2)+h\langle\sigma+\rho,\mathbf{1}\rangle}(m(\sigma)-q_1)(m(\rho)-q_1)$$
$$= (\mathbf{E}Z)^2(2\cosh(h))^{-2N}\sum_{\sigma,\rho} e^{\frac{N\beta^2}{2}R(\sigma,\rho)^2+h\langle\sigma+\rho,\mathbf{1}\rangle}(m(\sigma)-q_1)(m(\rho)-q_1),$$

where we have used that $\mathbf{E}[Z] = (2\cosh(h))^N e^{\frac{\beta^2 N}{4}}$. This explains the dependence on $(\mathbf{E}Z)^2$.

It thus remains to compute the double sum. To this end, we will use the HS transform, followed by the Laplace method.

Fact 6.5 (Hubbard–Stratonovich transform)**.** *For any $c > 0$, we have*

$$e^{cx^2} = \frac{1}{\sqrt{c\pi}}\int_{\mathbb{R}} e^{-\frac{z^2}{c}+2zx}\,\mathrm{d}z$$

Applying the HS transform with $c = \frac{2\beta^2}{N}$ and $x = \frac{\langle\sigma,\rho\rangle}{2}$, we obtain that

$$e^{\frac{N\beta^2}{2}R(\sigma,\rho)^2} = e^{\frac{\beta^2}{2N}\langle\sigma,\rho\rangle^2} = \sqrt{\frac{N}{2\pi\beta^2}}\int_{\mathbb{R}} e^{-\frac{N}{2\beta^2}z^2+z\langle\sigma,\rho\rangle}\,\mathrm{d}z,$$

so that the double sum becomes

$$\sqrt{\frac{N}{2\pi\beta^2}}\cdot\int_{\mathbb{R}}\exp\left(-\frac{N}{2\beta^2}z^2\right)\sum_{\sigma,\rho}\exp(z\langle\sigma,\rho\rangle + h\langle\sigma+\rho,\mathbf{1}\rangle)(m(\sigma)-q_1)(m(\rho)-q_1)\,\mathrm{d}z.$$

Observe that given z, the pairs (σ_i,ρ_i) are independent of each other, and hence this sum over σ,ρ is tractable to explicitly compute. In particular, consider the distribution p on $\{\pm1\}^2$, with

$p(a,b) = \frac{1}{C(z,h)} \exp(zab + ha + hb)$. Then one can check the normalization constant is $C(z,h) = 2(e^z \cosh(2h) + e^{-z})$, and we have

$$\sum_{\sigma,\rho} \exp(z \langle \sigma, \rho \rangle + h \langle \sigma + \rho, \mathbf{1} \rangle)(m(\sigma) - q_1)(m(\rho) - q_1) = C(z,h)^N \cdot \mathbf{E}_{p^{\otimes N}} \left[(m(\sigma) - q_1)\,(m(\rho) - q_1) \right] .$$

Now, expanding it out, we have

$$(m(\sigma) - q_1)(m(\rho) - q_1) = \frac{1}{N^2} \sum_{i \neq j} (\sigma_i - q_1)(\rho_j - q_1) + \frac{1}{N^2} \sum_i (\sigma_i - q_1)(\rho_i - q_1) ,$$

and therefore,

$$\begin{aligned} f_N(z) &:= \mathbf{E}_{p^{\otimes N}} [(m(\sigma) - q_1)(m(\rho) - q_1)] \\ &= \frac{N(N-1)}{N^2} (m(z) - q_1)^2 + \frac{1}{N} (k(z) - 2q_1 m(z) + q_1^2) \\ &= (m(z) - q_1)^2 + \frac{1}{N} \left(k(z) - m(z)^2 \right) \\ &= (m(z) - q_1)^2 + \frac{1}{N} \left(k(z) - q \right) + \frac{1}{N} \left(q - q_1^2 \right) + \frac{1}{N} \left(q_1^2 - m(z)^2 \right) \end{aligned}$$

where $m(z) := \mathbf{E}_p[a] = \mathbf{E}_p[b]$ and $k(z) := \mathbf{E}_p[ab]$. Thus, our goal is to show that

$$(2\cosh(h))^{-2N} \cdot \sqrt{\frac{N}{2\pi\beta^2}} \cdot \int_{\mathbb{R}} \exp\left(N \left[-\frac{z^2}{2\beta^2} + \log\left(2(e^z \cosh(2h) + e^{-z})\right) \right] \right) f_N(z) \, \mathrm{d}z \leqslant O\left(\frac{h^2}{N} \right) .$$

Rearranging, simplifying, and summarizing, we would like to show that

$$\mathbf{E}_{z \sim \mathcal{N}\left(0, \frac{\beta^2}{N}\right)} \left[\left(\frac{e^z \cosh(2h) + e^{-z}}{\cosh(2h) + 1} \right)^N f_N(z) \right] \leqslant O\left(\frac{h^2}{N} \right) ,$$

where an explicit calculation yields that

$$\begin{aligned} m(z) &= \frac{e^z \sinh(2h)}{e^z \cosh(2h) + e^{-z}} \text{ and} \\ k(z) &= \frac{e^z \cosh(2h) - e^{-z}}{e^z \cosh(2h) + e^{-z}} . \end{aligned}$$

Let us massage this expression to deal with the first part of the integrand. Set

$$g_N(z) = \frac{\cosh(z) + \sinh(h)^2 e^z}{\left(1 + \sinh(h)^2\right) e^{\tanh(h)^2 \cdot z} \left(e^{\frac{z^2}{2}\left(1 - \tanh(h)^4\right)} \right)} .$$

Then,

$$(*) = \mathbf{E}_{z \sim \mathcal{N}\left(0, \frac{\beta^2}{N}\right)} \left[g_N(\boldsymbol{z})^N \cdot e^{N \tanh(h)^2 \boldsymbol{z}} \cdot e^{\frac{N\boldsymbol{z}^2}{2}\left(1 - \tanh(h)^4\right)} \cdot f_N(\boldsymbol{z}) \right]$$

Define $\gamma > 0$ by $\frac{1}{\gamma^2} = \frac{1}{\beta^2} - \left(1 - \tanh(h)^4\right)$, and $\mu = \gamma^2 \tanh(h)^2$, so

$$(*) = \frac{\gamma}{\beta} \cdot \mathbf{E}_{z \sim \mathcal{N}\left(0, \frac{\gamma^2}{N}\right)} \left[g_N(\boldsymbol{z})^N \cdot e^{N \tanh(h)^2 \boldsymbol{z}} \cdot f_N(\boldsymbol{z}) \right]$$

$$
\begin{aligned}
&= \frac{\gamma}{\beta} \cdot e^{\frac{1}{2}\cdot N\gamma^2 \tanh(h)^4} \cdot \mathbf{E}_{z\sim\mathcal{N}\left(\mu,\frac{\gamma^2}{N}\right)} \left[g_N(z)^N \cdot f_N(z)\right] \\
&\asymp \mathbf{E}_{z\sim\mathcal{N}\left(\mu,\frac{\gamma^2}{N}\right)} \left[g_N(z)^N \cdot f_N(z)\right] ,
\end{aligned}
$$

where the final line uses the fact that $h \ll N^{-1/4}$. Why is this useful? By the choice of g, we in fact have $g(0) = 1$, $g^{(1)}(0) = g^{(2)}(0) = 0$, and $g^{(3)}(0) = -2\tanh(h)^2\left(1 - \tanh(h)^4\right) = O(h^2)$. Thus, since z is typically $\lesssim \frac{1}{\sqrt{N}}$, $g_N(z)$ is typically $1 + O\left(\frac{h^2}{N^{3/2}}\right)$, and so $g_N(z)^N$ is typically $1 + O\left(\frac{h^2}{\sqrt{N}}\right)$, essentially allowing us to ignore it in the integral.

More concretely, for our purposes (where we do not care about precise constants), it suffices to use the looser Cauchy–Schwarz bound, yielding

$$
(*) \lesssim \mathbf{E}\left[g_N(z)^{2N}\right]^{1/2} \cdot \mathbf{E}\left[f_N(z)^2\right]^{1/2}
$$

To control this, we will show

(i) $\mathbf{E}\left[(m(z) - q_1)^{2k}\right] \leqslant O\left(\frac{h^{2k}}{N^k}\right)$ for all constant k, proved in Proposition 6.6,

(ii) $\mathbf{E}\left[(k(z) - q)^{2k}\right] \leqslant O\left(\frac{1}{N^k}\right)$ for all constant k, proved in Proposition 6.7, and

(iii) $\mathbf{E}\left[g_N(z)^{2N}\right] = O(1)$, proved in Proposition 6.8.

These essentially immediately yield Lemma 6.4.

6.2 A series of Gaussian expectations

Proposition 6.6 (Controlling the magnetization)**.** *For any constant k,*

$$
\mathbf{E}_{z\sim\mathcal{N}\left(\mu,\frac{\gamma^2}{N}\right)} \left[\left(\frac{e^z\sinh(2h)}{e^z\cosh(2h) + e^{-z}} - q_1\right)^{2k}\right] \leqslant O\left(\frac{h^2}{N}\right)^k .
$$

Proof. Let us write

$$
\begin{aligned}
&\mathbf{E}\left[\left(\frac{e^z\sinh(2h)}{e^z\cosh(2h) + e^{-z}} - q_1\right)^{2k}\right] \\
&\quad\lesssim \mathbf{E}\left[\left(\frac{e^z\sinh(2h)}{e^z\cosh(2h) + e^{-z}} - \mathbf{E}\left[\frac{e^z\sinh(2h)}{e^z\cosh(2h) + e^{-z}}\right]\right)^{2k}\right] + \left(\mathbf{E}\left[\frac{e^z\sinh(2h)}{e^z\cosh(2h) + e^{-z}}\right] - q_1\right)^{2k} \\
&\quad= O\left(h^{2k}\right) \mathbf{E}\left[\left(\frac{e^z}{e^z\cosh(2h) + e^{-z}} - \mathbf{E}\left[\frac{e^z}{e^z\cosh(2h) + e^{-z}}\right]\right)^{2k}\right] + \left(\mathbf{E}\left[\frac{e^z\sinh(2h)}{e^z\cosh(2h) + e^{-z}}\right] - q_1\right)^{2k} .
\end{aligned}
$$

The first term is easy to deal with: because $z \mapsto \frac{e^z}{e^z\cosh(2h)+e^{-z}}$ is an $O(1)$-Lipschitz function, Gaussian concentration of Lipschitz functions yields that the associated $2k$th moment is bounded by $O\left(\frac{\gamma^2}{N}\right)^k$.

For the second term, because $q = O(h^2)$,

$$
q_1 = \mathbf{E}_{g\sim\mathcal{N}(0,1)}\left[\tanh\left(\beta g\sqrt{q} + h\right)\right] = h + O(h^3).
$$

On the other hand, setting $f(z) = \frac{e^z \sinh(2h)}{e^z \cosh(2h)+e^{-z}}$, so

$$\begin{aligned} f(0) &= \tanh(h) = h + O(h^3)\,, \\ f'(0) &= O(h)\,, \text{ and} \\ f''(0) &= O\left(h^3\right)\,, \end{aligned}$$

we have

$$\begin{aligned} \mathbf{E}\left[f(\boldsymbol{z})\right] &= f(0) + f'(0)\mu + \frac{1}{2}f''(0)\left(\mu^2 + \frac{\gamma^2}{N}\right) + O\left(\frac{h}{\sqrt{N}}\right) \\ &= h + O(h)\cdot\gamma^2\tanh(h)^2 - O(h^3)\cdot\left(\mu^2 + \frac{\gamma^2}{N}\right) + O\left(\frac{h}{\sqrt{N}}\right) \\ &= h + O\left(\frac{h}{\sqrt{N}}\right). \end{aligned}$$

Consequently,

$$\left|\mathbf{E}\left[\frac{e^{\boldsymbol{z}}\sinh(2h)}{e^{\boldsymbol{z}}\cosh(2h)+e^{-\boldsymbol{z}}}\right] - q_1\right| = O\left(\frac{h}{\sqrt{N}}\right),$$

completing the proof. □

Proposition 6.7 (Controlling the overlap)**.** *For any constant k,*

$$\mathbf{E}_{z\sim\mathcal{N}\left(\mu,\frac{\gamma^2}{N}\right)}\left[\left(\frac{e^z\cosh(2h)-e^{-z}}{e^z\cosh(2h)+e^{-z}} - q\right)^{2k}\right] \leqslant O\left(\frac{1}{N^k}\right).$$

Proof. Our strategy will essentially be identical to that of the previous proof. Again, because $z \mapsto k(z)$ is an $O(1)$-Lipschitz function, we have

$$\mathbf{E}\left[\left(\frac{e^z\cosh(2h)-e^{-z}}{e^z\cosh(2h)+e^{-z}} - q\right)^{2k}\right] \lesssim O\left(\frac{1}{N^k}\right) + \left(\mathbf{E}\left[\frac{e^z\cosh(2h)-e^{-z}}{e^z\cosh(2h)+e^{-z}}\right] - q\right)^{2k}$$

We have $k(z) = \tanh\left(z + \frac{1}{2}\ln\cosh(2h)\right) = \tanh\left(z + h^2 + O(h^4)\right)$. As a result,

$$\begin{aligned} \mathbf{E}\left[k(\boldsymbol{z})\right] &= \mathbf{E}\left[\boldsymbol{z} + h^2 - \frac{1}{3}\boldsymbol{z}^3\right] + O\left(h^4\right) \\ &= \mu + h^2 - \frac{1}{3}\cdot\mu^3 - \mu\cdot\frac{\gamma^2}{N} + O\left(h^4\right) \\ &= \mu + h^2 + O\left(\frac{1}{N}\right). \end{aligned}$$

Let us also figure out the leading terms in the Taylor series of q. We have $q = O(h^2)$, so suppose that $q = ah^2 + O(h^4)$. Expanding out the recursion defining q, using the Taylor series expansion $\tanh^2(x) = x^2 + O(x^4)$, we have

$$q = \mathbf{E}_{g\sim\mathcal{N}(0,1)}\left[\beta\sqrt{q}\boldsymbol{g} + h\right] = \beta^2 q + h^2 + O\left(h^4\right).$$

It follows that $q = \frac{1}{1-\beta^2} \cdot h^2 + O(h^4)$. We also have

$$\begin{aligned}
\mu + h^2 &= \tanh(h)^2 \cdot \frac{1}{\frac{1}{\beta^2} - \left(1 - \tanh(h)^4\right)} + h^2 \\
&= h^2 \left(\frac{1}{\frac{1}{\beta^2} - 1} + 1 \right) + O\left(\frac{1}{N}\right) \\
&= \frac{1}{1-\beta^2} \cdot h^2 + O\left(\frac{1}{N}\right).
\end{aligned}$$

Therefore,

$$|\mathbf{E}\left[k(\boldsymbol{z})\right] - q| = O\left(\frac{1}{N}\right),$$

completing the proof. □

Proposition 6.8. *It holds that* $\mathbf{E}_{z \sim \mathcal{N}\left(0, \frac{\beta^2}{N}\right)}\left[g_N(\boldsymbol{z})^{2N}\right] = 1 + o_N(1)$.

Proof. Recall

$$g_N(z) = \frac{\cosh(z) + \sinh(h)^2 e^z}{\left(1 + \sinh(h)^2\right) e^{\tanh(h)^2 \cdot z} \left(e^{\frac{z^2}{2}\left(1 - \tanh(h)^4\right)} \right)}.$$

Let $\kappa = \tanh(h)^2 = \frac{\sinh(h)^2}{1+\sinh(h)^2}$, so we have

$$g_N(z) = \frac{\cosh(z) + \kappa \sinh(z)}{e^{\kappa z + \frac{z^2}{2} \cdot (1-\kappa^2)}}.$$

First note that the terms in the Taylor expansion up to order 2 cancel out—indeed, for small z, we have $\cosh(z) + \kappa \sinh(z) = 1 + \kappa z + \frac{z^2}{2} + O(z^3)$, and

$$e^{\kappa z + \frac{z^2}{2} \cdot (1-\kappa^2)} = 1 + \kappa z + (1-\kappa^2) \cdot \frac{z^2}{2} + \frac{1}{2}(\kappa z)^2 + O(z^3) = 1 + \kappa z + \frac{z^2}{2} + O(z^3).$$

The claim near-immediately follows. For large z, the $e^{(1-\kappa^2)z^2/2}$ term in the denominator dominates, so we may establish a uniform bound of $g_N(z) \leqslant 1 + O(|z|^3)$. We may thus bound $\mathbf{E}\left[g_N(\boldsymbol{z})^{2N}\right] \leqslant \mathbf{E}\left[\left(1 + O(|\boldsymbol{z}|^3)\right)^{2N}\right] \leqslant 1 + O\left(\frac{1}{\sqrt{N}}\right)$, as desired. □

7 Towards understanding the RGD recursion

7.1 Towards understanding the fixed points in the (AT) region

In this section, we analyze some salient properties of the fixed point equations governing RGD and AMP. While we do not show that there exists a unique fixed point to $f_{\beta,\lambda}$ defined in (5) in the (AT) region, we are able to show that there is *at most* one fixed point in this region, and that this fixed point must be stable.

Proposition 7.1. *Fix $\lambda > 1$ and $\beta \geqslant 0$. Set $h_* = \inf\{h \geqslant 0 : (\beta, h)$ satisfies (AT)$\}$. Suppose that for all $h > h_*$, (β, h) satisfies (AT). Then, for $f_\beta = f_{\beta,\lambda}$ defined as in (5),*

(i) *If $\beta < \frac{1}{\lambda}$, then $h = 0$ is the only fixed point for $f_\beta(h)$. Furthermore, $h = 0$ is a stable fixed point.*

(ii) *If $\frac{1}{\lambda} < \beta < 1$, there are exactly two non-negative fixed points for $f_\beta(h)$ at 0 and $\mathsf{OPT}_{\beta,\lambda}$. Furthermore, $h = 0$ is an unstable fixed point and $h = \mathsf{OPT}_{\beta,\lambda}$ is a stable fixed point.*

(iii) *If $\beta > 1$, there is at most one positive fixed point $h > 0$ such that (β, h) satisfies (AT). Furthermore, if h exists, it is a stable fixed point of f_β. Finally, for $\beta = \lambda$, such a fixed point does indeed exist.*

(iv) *For $h > \frac{h_*}{\beta\lambda}$, $f'_\beta(h) > 0$.*

Given the above, Lemma 5.3 is near-immediate.

Lemma 5.3. *Let $\lambda > 1$. If $\beta < 1$ or $\beta = \lambda$, (β, λ) satisfies (AT-fixed-pt). Suppose that (β, λ) satisfies (AT-fixed-pt). Then,*

(i) *if h is such that $(\beta, \beta\lambda h)$ satisfies (AT), $f'_{\beta,\lambda}(h) > 0$,*

(ii) *$\mathsf{OPT}_{\beta,\lambda}$ is the unique fixed point h of $f_{\beta,\lambda}$ such that $(\beta, \beta\lambda h)$ satisfies (AT), and*

(iii) *$\mathsf{OPT}_{\beta,\lambda}$ is a stable fixed point of $f_{\beta,\lambda}$.*

Proof. (i, ii, iii) near-immediately follow from Proposition 7.1(iii, iv). The fact that (AT-fixed-pt) is satisfied if $\beta < 1$ or $\beta = \lambda$ is a consequence of Proposition 7.1(i, ii, iii). ☐

We shall prove Proposition 7.1 over the course of this section. We start by noting the following simple fact.

Fact 7.2. *For any $a, b > 0$,*

$$2a\mathbf{E}[\operatorname{sech}^2(ag + b)\tanh(ag + b)] = -\mathbf{E}_{g\sim\mathcal{N}(0,1)}[\operatorname{sech}^2(ag + b) \cdot g] > 0.$$

Proof. The first equality is just Stein's lemma. For the second, we use symmetry to express

$$\mathbf{E}_{g\sim\mathcal{N}(0,1)}[\operatorname{sech}^2(ag + b) \cdot g] = \frac{1}{2} \cdot \mathbf{E}\left[g\left(\operatorname{sech}^2(b + ag) - \operatorname{sech}^2(b - ag)\right)\right] < 0\,.$$

If $g > 0$ we have $|b + ag| > |b - ag|$, and if $g < 0$, we have $|b - ag| > |b + ag|$. Since $\operatorname{sech}^2(\cdot)$ is even and decreasing for positive arguments, the claim follows. ☐

We also need the following facts about the recursion defining the overlap constant q.

Lemma 7.3. *Let $\beta > 0$ and $h \geqslant 0$. Consider the solution to the fixed point equation*

$$q = q(h) = \mathbf{E}[\tanh^2(\beta g\sqrt{q} + h)]\,.$$

Then the following hold.

(i) *If $h > 0$, there is a unique solution $q = q(h) > 0$. If $h = 0$ and $\beta < 1$, then $q = 0$ is the unique solution to the above recursion, and $q'(0) = 0$.*

(ii) *If* (β, h) *satisfy* (AT), *then* q *is differentiable. Denoting* $\boldsymbol{X} = \beta \boldsymbol{g}\sqrt{q} + h$, *we have*

$$q'(h) = \frac{2\mathbf{E}[\tanh(\boldsymbol{X})\,\mathrm{sech}^2(\boldsymbol{X})]}{1 - \beta^2 \mathbf{E}[\mathrm{sech}^4(\boldsymbol{X}) - 2\,\mathrm{sech}^2(\boldsymbol{X})\tanh^2(\boldsymbol{X})]}\,.$$

Consequently, $q'(h) > 0$.

Proof. The uniqueness of q in the mentioned regimes is [Gue01] (also see [Tal10, Proposition 1.3.8]). For the second point, define the shorthand $\boldsymbol{T} = \tanh\big(\beta \boldsymbol{g}\sqrt{q} + h\big)$ and $\boldsymbol{S} = \mathrm{sech}\big(\beta \boldsymbol{g}\sqrt{q} + h\big)$. We can implicitly differentiate through h to obtain

$$\begin{aligned} q' &= \mathbf{E}[2\boldsymbol{T}\boldsymbol{S}^2(\beta \boldsymbol{g}(\sqrt{q})' + 1)] \\ &= 2\mathbf{E}[\boldsymbol{T}\boldsymbol{S}^2] + \frac{\beta q'}{\sqrt{q}}\mathbf{E}[\boldsymbol{g}\boldsymbol{T}\boldsymbol{S}^2] \\ &= 2\mathbf{E}[\boldsymbol{T}\boldsymbol{S}^2] + \beta^2 q' \mathbf{E}[\boldsymbol{S}^4 - 2\boldsymbol{S}^2\boldsymbol{T}^2]\,. \qquad \text{(Stein's lemma)} \end{aligned}$$

Solving for q' yields

$$q' = \frac{2\mathbf{E}[\boldsymbol{T}\boldsymbol{S}^2]}{1 - \beta^2\mathbf{E}[\boldsymbol{S}^4 - 2\boldsymbol{S}^2\boldsymbol{T}^2]}\,.$$

For (β, h) satisfying (AT), the denominator is strictly positive (thus justifying the well-definedness of q'). The numerator is strictly positive by Fact 7.2. $\square$

We will also need the following estimates for the behavior of q and q_1 for small h.

Lemma 7.4. *Let* $h \geqslant 0$ *and* $\beta < 1$. *Also let* $q(h)$ *be the unique solution to* $q = \mathbf{E}[\tanh^2(\beta \boldsymbol{g}\sqrt{q} + h)]$ *and define* $q_1(h) = \mathbf{E}[\tanh\Big(\beta \boldsymbol{g}\sqrt{q(h)} + h\Big)]$. *Then*

$$\frac{h^2}{1-\beta^2} - \frac{2h^4}{(1-\beta^2)^3} \leqslant q(h) \leqslant \frac{h^2}{1-\beta^2}$$
$$h - \frac{h^2}{3(1-\beta^2)} \leqslant q_1(h) \leqslant h$$

Proof. The inequality for q is [DW23, Lemma 1.2]. For the inequality on q_1, we first note that $\tanh x \leqslant x$, so with $Y = \beta\sqrt{q}\boldsymbol{z} + h$, we have $q_1 = \mathbf{E}\tanh Y \leqslant \mathbf{E}Y = h$. Furthermore, since $x - \tanh x \leqslant \frac{x^2}{3}$, we obtain $\mathbf{E}[Y - \tanh Y] \leqslant \frac{1}{3}\mathbf{E}Y^2 = \frac{1}{3}(h^2 + \beta^2 q)$. Combining with the upper bound on q, we obtain

$$q_1 \geqslant h - \frac{h^2}{3(1-\beta^2)}\,,$$

as desired. $\square$

We are now ready to prove some nice properties of f_β.

Lemma 7.5. *Fix* $\lambda > 1$ *and* $\beta > 0$. *Consider the function* f_β *defined as in* (5):

$$f_\beta(h) = \mathbf{E}_{g\sim\mathcal{N}(0,1)} \tanh\left(\beta g\sqrt{q_{\beta,\beta\lambda h}} + \beta\lambda h\right)\,.$$

Then,

(i) *If $(\beta, \beta\lambda h)$ satisfies (AT), the function $m \mapsto \frac{f_\beta(m)}{m}$ is strictly decreasing at h.*

(ii) *If $(\beta, \beta\lambda h)$ satisfies (AT), f_β is increasing at h.*

(iii) *0 is a stable fixed point f_β if $\beta < \frac{1}{\lambda}$, and is an unstable fixed point if $\frac{1}{\lambda} < \beta < 1$.*

Proof. Let $\boldsymbol{X} = \boldsymbol{X}(h) = \beta\lambda h + \beta \boldsymbol{g}\sqrt{q_{\beta,\beta\lambda h}}$, and introduce the function $r = r(h) = q_{\beta,\beta\lambda h}$. By Stein's lemma, we have

$$\begin{aligned} f_\beta'(h) &= \mathbf{E}_{\boldsymbol{g}\sim\mathcal{N}(0,1)}\left[\operatorname{sech}^2(\boldsymbol{X}) \cdot \left(\beta\boldsymbol{g} \cdot \frac{\mathrm{d}}{\mathrm{d}h}\sqrt{r} + \beta\lambda\right)\right] \\ &= \mathbf{E}\left[\operatorname{sech}^2(\boldsymbol{X})\left(\beta\boldsymbol{g}\frac{r'}{2\sqrt{r}} + \beta\lambda\right)\right] \\ &= \beta\lambda\mathbf{E}[\operatorname{sech}^2(\boldsymbol{X})] - \beta^2 r'\mathbf{E}\left[\operatorname{sech}^2(\boldsymbol{X})\tanh(\boldsymbol{X})\right]. \end{aligned}$$

Here, we have implicitly used the differentiability of $\sqrt{r(h)}$, which we justify now.

First, by Lemma 7.3(ii), if $\beta < 1$, r is differentiable and strictly positive for all $h > 0$. Let us now compute the limit as $h \to 0^+$. By Lemmas 7.3 and 7.4, we have $r'(h) = \beta^2\lambda^2 \cdot (1 + o_h(1))\frac{2h}{1-\beta^2}$ and $r(h) = \beta^2\lambda^2 \cdot \frac{h^2}{1-\beta^2}(1 + o_h(1))$. Hence,

$$\lim_{h\to 0^+} \frac{r'(h)}{\sqrt{r(h)}} = \sqrt{\frac{2\beta^2\lambda^2}{1-\beta^2}}.$$

It follows by symmetry that $f_\beta'(h)$ is well defined for $\beta < 1$ and all $h \in \mathbb{R}$.

If $\beta < 1$ and $h = 0$, then $r(h) = 0$ by Lemma 7.3(i), so $\boldsymbol{X} = 0$ almost surely and the above simplifies to

$$f_\beta'(0) = \beta\lambda\,,$$

which implies (iii).

Let us next establish (ii). Abbreviate $\boldsymbol{S} = \operatorname{sech}(\boldsymbol{X})$ and $\boldsymbol{T} = \tanh(\boldsymbol{X})$. We have by Lemma 7.3(ii) and (AT) that

$$\begin{aligned} f_\beta'(h) &= \beta\lambda\mathbf{E}\boldsymbol{S}^2 - \beta^2 \cdot \beta\lambda \cdot \frac{2\mathbf{E}[\boldsymbol{T}\boldsymbol{S}^2]}{1 - \beta^2\mathbf{E}\left[\boldsymbol{S}^4 - 2\boldsymbol{S}^2\boldsymbol{T}^2\right]} \cdot \mathbf{E}[\boldsymbol{T}\boldsymbol{S}^2] \\ &\geqslant \beta\lambda\left(\mathbf{E}\boldsymbol{S}^4 - \frac{\mathbf{E}[\boldsymbol{S}^2\boldsymbol{T}]^2}{\mathbf{E}[\boldsymbol{S}^2\boldsymbol{T}^2]}\right) \geqslant 0\,, \end{aligned}$$

where the final inequality is the Cauchy–Schwarz inequality.

Finally, consider (i). Set $g(h) = \frac{f_\beta(h)}{h}$. We have

$$\begin{aligned} h^2 \cdot g'(h) &= f_\beta'(h)h - f_\beta(h) \\ &= \mathbf{E}[\beta\lambda h\boldsymbol{S}^2] - \beta^2 h r'\mathbf{E}\left[\boldsymbol{S}^2\boldsymbol{T}\right] - \mathbf{E}[\boldsymbol{T}]\,. \end{aligned} \tag{13}$$

For fixed h, define the function $F(z) = \mathbf{E}\tanh\left(\beta\lambda z + \beta\boldsymbol{g}\sqrt{q_{\beta,\beta\lambda h}}\right)$, and denote $\boldsymbol{Y} = \beta\lambda z + \beta\boldsymbol{g}\sqrt{q_{\beta,\beta\lambda h}}$. Note that F is *not* the same as f_β, since we have fixed $q = q_{\beta,\beta\lambda h}$ (in particular, $q \neq q_{\beta,\beta\lambda z}$). We may then compute for $z > 0$ that

$$F'(z) = \beta\lambda\mathbf{E}\operatorname{sech}^2(\boldsymbol{Y}) > 0$$

$$F''(z) = -2\beta^2\lambda^2 \mathbf{E}[\operatorname{sech}^2(\boldsymbol{Y})\tanh(\boldsymbol{Y})] < 0\,,$$

where the final inequality is Fact 7.2. Plugging this back into (13),

$$h^2 \cdot g'(h) = -\beta^2 h \cdot r' \cdot \mathbf{E}[\boldsymbol{S}^2\boldsymbol{T}^2] + hF'(h) - F(h)\,.$$

Lemma 7.3(ii) and Fact 7.2 imply that the first term is negative. For the other two terms, note that if we set $G(z) = zF'(z) - F(z)$, then because $G'(z) = F''(z) < 0$ and $G(0) = 0$, we have $G(h) < 0$, completing the proof. ☐

Lemma 7.6. *Suppose that $\beta = \lambda > 1$. Then, there exists a fixed point $h > 0$ of the function f_β such that $(\lambda, \lambda^2 h)$ satisfies (AT).*

Proof. Consider the auxiliary function ψ defined by

$$\psi(\gamma) = \lambda^2 \mathbf{E}\left[\tanh^2\left(\gamma + \boldsymbol{g}\sqrt{\gamma}\right)\right]\,.$$

[DAM16, Lemma 6.1] shows that for $\lambda > 1$, ψ has a nonzero stable fixed point $\gamma_\star$. Also, as observed in [DAM16, Eq. (223)], for any $\gamma > 0$ and $k \in \mathbb{Z}_{>0}$,

$$\mathbf{E}\left[\tanh^{2k}\left(\gamma + \boldsymbol{g}\sqrt{\gamma}\right)\right] = \mathbf{E}\left[\tanh^{2k-1}\left(\gamma + \boldsymbol{g}\sqrt{\gamma}\right)\right]\,. \tag{14}$$

Essentially by definition coupled with the above identity, we have that $\frac{\gamma_\star}{\lambda^2}$ is a fixed point of $f_{\lambda,\lambda}$ with $q_{\lambda,\gamma_\star} = \frac{\gamma_\star}{\lambda^2}$.

To see why, let us first introduce the shorthand $\boldsymbol{Y} = \gamma_\star + \boldsymbol{g}\sqrt{\gamma_\star}$. Then the stability of $\gamma_\star$ as a fixed point of ψ implies that

$$\begin{aligned} 1 > \psi'(\gamma_\star) &= \lambda^2 \mathbf{E}\left[2\tanh\left(\boldsymbol{Y}\right) \cdot \operatorname{sech}^2\left(\boldsymbol{Y}\right) \cdot \left(1 + \frac{\boldsymbol{g}}{2\sqrt{\gamma_\star}}\right)\right] \\ &= \lambda^2 \mathbf{E}\left[2\tanh(\boldsymbol{Y})\operatorname{sech}^2(\boldsymbol{Y}) + \operatorname{sech}^4(\boldsymbol{Y}) - 2\operatorname{sech}^2(\boldsymbol{Y})\tanh^2(\boldsymbol{Y})\right]\,, \end{aligned}$$

where the last line is Stein's lemma. We also have by (14) that

$$\mathbf{E}\left[\tanh(\boldsymbol{Y})\operatorname{sech}^2(\boldsymbol{Y})\right] = \mathbf{E}\left[\tanh(\boldsymbol{Y}) - \tanh^3(\boldsymbol{Y})\right] = \mathbf{E}[\tanh^2(\boldsymbol{Y}) - \tanh^4(\boldsymbol{Y})] = \mathbf{E}[\tanh^2(\boldsymbol{Y})\operatorname{sech}^2(\boldsymbol{Y})]\,,$$

so $\lambda^2 \mathbf{E}[\operatorname{sech}^4(\boldsymbol{Y})] < 1$ and $(\lambda, \gamma_\star)$ satisfies (AT), as claimed. ☐

To conclude this section, let us put the pieces together and prove Proposition 7.1.

Proof of Proposition 7.1. For $\beta < 1$, note that $(\beta, \beta\lambda h)$ satisfies (AT) for all $h > 0$. Consequently, Lemma 7.5(i) shows that in this regime, it suffices to understand whether $\lim_{h\to 0^+} \frac{f_\beta(h)}{h}$ is greater or less than 1: if it is greater than 1, then 0 is an unstable fixed point and there is a unique fixed point at some positive h which is stable; if it is smaller than 1, then 0 is the unique fixed point and it is stable. Now Lemma 7.5(iii) establishes precisely where 0 is an unstable fixed point, which yields (i) and (ii). Next, (iii) follows from Lemma 7.5(i) and Lemma 7.6. Finally, (iv) is Lemma 7.5(ii). ☐

Acknowledgments

We are very grateful to Brice Huang for numerous discussions about the SK model, and to Hang Du, Brice Huang, and Mark Sellke for showing us the trick used in Lemma 6.4. We thank Ahmed El Alaoui for helpful discussions about the SK model under weak external field. We benefited from countless conversations with Prasad Raghavendra and Sidhanth Mohanty, which inspired this work. A.R. would also like to thank Sam Hopkins and Kuikui Liu for feedback on an earlier draft of this work. We acknowledge the use of ChatGPT for literature search and helping prove several key lemmas.

References

[AKUZ19] Fabrizio Antenucci, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborová. Approximate Survey Propagation for Statistical Inference. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(2):023401, 2019. 6, 11, 13

[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. 3, 7

[BBP05] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. 2005. 2, 5

[BCV25] Thiago Bergamaschi, Chi-Fang Chen, and Umesh Vazirani. A structural theory of quantum metastability: Markov properties and area laws. *arXiv preprint arXiv:2510.08538*, 2025. 7, 15

[BDG06] Gérard Ben Arous, Amir Dembo, and Alice Guionnet. Cugliandolo-Kurchan equations for dynamics of spin-glasses. *Probability theory and related fields*, 136(4):619–660, 2006. 6

[BDH16] Anton Bovier and Frank Den Hollander. *Metastability: a potential-theoretic approach*, volume 351. Springer, 2016. 6

[BGJ20] Gerard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Algorithmic thresholds for tensor PCA. *The Annals of Probability*, 48(4):2052–2087, 2020. 6

[BGP24] Gérard Ben Arous, Cédric Gerbelot, and Vanessa Piccolo. Langevin dynamics for high-dimensional optimization: the case of multi-spiked tensor PCA. *arXiv preprint arXiv:2408.06401*, 2024. 7, 13

[BJ24] Gérard Ben Arous and Aukosh Jagannath. Shattering versus metastability in spin glasses. *Communications on Pure and Applied Mathematics*, 77(1):139–176, 2024. 6

[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. 1, 2, 16

[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. 17

[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

[Che21] Wei-Kuo Chen. On the Almeida-Thouless transition line in the Sherrington-Kirkpatrick model with centered Gaussian external field. *Electronic Communications in Probability*, 26:1–9, 2021. 17

[CK93] Leticia F Cugliandolo and Jorge Kurchan. Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. *Physical Review Letters*, 71(1):173, 1993. 6

[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. 24

[DAM16] Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the binary stochastic block model. pages 185–189, 07 2016. 1, 5, 10, 16, 34

[dAT78] 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. 9

[DGPZ25] Yatin Dandi, David Gamarnik, Francisco Pernice, and Lenka Zdeborová. Sequential dynamics in ising spin glasses. *arXiv preprint arXiv:2506.09877*, 2025. 6, 13

[DKMZ11] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E*, 84:066106, Dec 2011. 1, 2, 6

[DM14] Yash Deshpande and Andrea Montanari. Information-theoretically optimal sparse PCA. In *2014 IEEE International Symposium on Information Theory*, pages 2197–2201. IEEE, 2014. 5

[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. 1

[DW23] Partha S Dey and Qiang Wu. Mean field spin glass models under weak external field. *Communications in Mathematical Physics*, 402(2):1205–1258, 2023. 5, 25, 32

[EAKJ20] Ahmed El Alaoui, Florent Krzakala, and Michael Jordan. Fundamental limits of detection in the spiked wigner model. 2020. 5

[EAMS22] 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. 1

[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. 3

[Gue01] Francesco Guerra. Sum rules for the free energy in the mean field spin glass. *Mathematical Physics in Mathematics and Physics: Quantum and Operator Algebraic Aspects*, 30:161, 2001. 32

[Han07] Albert Hanen. A limit theorem for mean magnetisation in the sherrington-kirkpatrick model with an external field. In *Spin Glasses: Statics and Dynamics: Summer School, Paris 2007*, pages 177–201. Springer, 2007. 24

[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. 1

[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. 14

[HR04] David C Hoyle and Magnus Rattray. Principal-component-analysis eigenvalue spectra from data with symmetry-breaking structure. *Physical Review E*, 69(2):026124, 2004. 2

[Hub59] John Hubbard. Calculation of partition functions. *Physical Review Letters*, 3(2):77, 1959. 8

[JT17] Aukosh Jagannath and Ian Tobasco. Some Properties of the Phase Diagram for Mixed p-Spin Glasses. *Probability Theory and Related Fields*, 167(3):615–672, 2017. 17

[KX25] Pravesh K Kothari and Jeff Xu. Smooth trade-off for tensor pca via sharp bounds for kikuchi matrices. *arXiv preprint arXiv:2510.03061*, 2025. 6

[LM17] Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. In *Conference on Learning Theory*, pages 1297–1301. PMLR, 2017. 5

[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, 2024. 2, 3, 4, 7, 8, 15

[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. 8

[LSS22] Tengyuan Liang, Subhabrata Sen, and Pragya Sur. High-dimensional asymptotics of langevin dynamics in spiked matrix models. *arXiv preprint arXiv:2204.04476*, 2022. 6

[LST21] Yin Tat Lee, Ruoqi Shen, and Kevin Tian. Structured logconcave sampling with a restricted Gaussian oracle. In *Conference on Learning Theory*, pages 2993–3050. PMLR, 2021. 2

[LW22] Gen Li and Yuting Wei. A non-asymptotic framework for approximate message passing in spiked models. *arXiv preprint arXiv:2208.03313*, 2022. 6

[Mio17] Léo Miolane. Fundamental limits of low-rank matrix estimation: the non-symmetric case. *arXiv preprint arXiv:1702.00473*, 2017. 6

[MPV87] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987. 3

[MR14] Andrea Montanari and Emile Richard. A statistical model for tensor PCA. *Advances in neural information processing systems*, 27, 2014. 6

[MR15] Andrea Montanari and Emile Richard. Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. *IEEE Transactions on Information Theory*, 62(3):1458–1484, 2015. 5

[MV21] Andrea Montanari and Ramji Venkataramanan. Estimation of low-rank matrices via approximate message passing. 2021. 1, 6, 10

[MW23] Andrea Montanari and Yuchen Wu. Posterior sampling in high dimension via diffusion processes. *arXiv preprint arXiv:2304.11449*, 2023. 6

[PWBM18] Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. Optimality and sub-optimality of pca i: Spiked random matrix models. *The Annals of Statistics*, 46(5):2416–2451, 2018. 5

[RSS+25] Dhruv Rohatgi, Abhishek Shetty, Donya Saless, Yuchen Li, Ankur Moitra, Andrej Risteski, and Dylan J Foster. Taming imperfect process verifiers: A sampling perspective on backtracking. *arXiv preprint arXiv:2510.03149*, 2025. 7

[RV18] Cynthia Rush and Ramji Venkataramanan. Finite sample analysis of approximate message passing algorithms. *IEEE Transactions on Information Theory*, 64(11):7264–7286, 2018. 6

[Sel24] Mark Sellke. The threshold energy of low temperature langevin dynamics for pure spherical spin glasses. *Communications on Pure and Applied Mathematics*, 77(11):4065–4099, 2024. 6, 13

[STL20] Ruoqi Shen, Kevin Tian, and Yin Tat Lee. Composite Logconcave Sampling with a Restricted Gaussian Oracle. *arXiv preprint arXiv:2006.05976*, 2020. 2

[Str57] RL Stratonovich. On a method of calculating quantum distribution functions. In *Soviet Physics Doklady*, volume 2, page 416, 1957. 8

[Tal10] Michel Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*, volume 54. Springer Science & Business Media, 2010. 32

[Tal11] Michel Talagrand. *Mean Field Models for Spin Glasses: Volume II: Advanced replica-symmetry and low temperature*, volume 55. Springer Science & Business Media, 2011. 17

[Ton02] Fabio Lucio Toninelli. About the Almeida-Thouless transition line in the Sherrington-Kirkpatrick mean-field spin glass model. *Europhysics letters*, 60(5):764, 2002. 17

[WEM19] Alexander Wein, Ahmed El Alaoui, and Cristopher Moore. The Kikuchi hierarchy and tensor PCA. *Journal of the ACM*, 2019. 6