CAPACITY THRESHOLD FOR THE ISING PERCEPTRON

BRICE HUANG

Abstract. We show that the capacity of the Ising perceptron is with high probability upper bounded by the constant $\alpha_\star \approx 0.833$ conjectured by Krauth and Mézard, under the condition that an explicit two-variable function $\mathscr{S}_\star(\lambda_1, \lambda_2)$ is maximized at $(1, 0)$. The earlier work of Ding and Sun [DS18] proves the matching lower bound subject to a similar numerical condition, and together these results give a conditional proof of the conjecture of Krauth and Mézard.

Contents

1. Introduction

The Ising perceptron was introduced in [Wen62, Cov65] as a simple model of a neural network. Mathematically, it is an intersection of a high-dimensional discrete cube with random half-spaces, defined as follows. Fix any $\kappa \in \mathbb{R}$ (our main result is for $\kappa = 0$). For $N \geqslant 1$, let $\Sigma_N = \{\pm 1\}^N$, and let $\boldsymbol{g}^1, \boldsymbol{g}^2, \ldots$ be a sequence of i.i.d. samples from $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_N)$. For $M \geqslant 1$, the Ising perceptron is the random set

$$S_N^M = \left\{ \boldsymbol{x} \in \Sigma_N : \frac{\langle \boldsymbol{g}^a, \boldsymbol{x} \rangle}{\sqrt{N}} \geqslant \kappa \quad \forall 1 \leqslant a \leqslant M \right\}. \tag{1}$$

As explained in [Gar87], S_N^M models the set of configurations of synaptic weights in a single-layer neural network that memorize all M patterns $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^M$. Define the random variable $M_N = M_N(\kappa)$ as the largest M such that $S_N^M \neq \varnothing$. Then, the **capacity** of this model is defined as the ratio M_N/N, and models the maximum number of patterns this network can memorize per synapse.

Krauth and Mézard [KM89] analyzed this model using the (non-rigorous) replica method from statistical physics. They conjectured that as $N \to \infty$, the capacity concentrates around an explicit constant

Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

$\alpha_\star = \alpha_\star(\kappa)$, which is approximately 0.833 for $\kappa = 0$ and is formally defined in Proposition 3.2 below.[1] This was part of a series of works in the statistical physics literature [Gar87, GD88, Gar88, KM89, Méz89] which analyzed various perceptron models using the replica or cavity methods and put forward detailed predictions for their behavior. In particular, [KM89] provided a conjecture for the limiting capacity of the Ising perceptron, while [GD88] gave an analogous conjecture for the spherical perceptron, where the spins $\boldsymbol{x}$ belong to the sphere $\{\boldsymbol{x} \in \mathbb{R}^N : \|\boldsymbol{x}\| = \sqrt{N}\}$ instead of Σ_N.

Ding and Sun [DS18] proved that $\alpha_\star$ is a rigorous lower bound for the capacity, subject to a numerical condition that an explicit univariate function is maximized at 0.

Theorem 1.1. [DS18, Theorem 1.1] *Under Condition 1.2 therein, the following holds for the $\kappa = 0$ Ising perceptron. For any $\alpha < \alpha_\star$, $\liminf_{N\to\infty} \mathbb{P}(M_N/N \geqslant \alpha) > 0$.*

Furthermore, [Xu21, NS23] showed that the capacity has a sharp threshold sequence, thereby improving the positive probability guarantee of Theorem 1.1 to high probability. Our main result is a matching upper bound for the capacity, subject to a similar numerical condition.

Theorem 1.2. *Under Condition 1.3 below, the following holds for the $\kappa = 0$ Ising perceptron. For any $\alpha > \alpha_\star$, $\lim_{N\to\infty} \mathbb{P}(M_N/N \geqslant \alpha) = 0$.*

Condition 1.3. The function $\mathscr{S}_\star(\lambda_1, \lambda_2)$ defined in (8) satisfies $\mathscr{S}_\star(\lambda_1, \lambda_2) \leqslant 0$ for all $\lambda_1, \lambda_2 \in \mathbb{R}$.

See §2.6 for a discussion of this condition. In particular $\mathscr{S}_\star(1, 0) = 0$ is a local maximum, and numerical plots suggest it is the unique global maximum.

Theorem 1.2 is a consequence of the more general Theorem 3.6, which states that $\alpha_\star(\kappa)$ upper bounds the capacity for general κ, under a number of numerical conditions depending on κ. The most complicated of these is Condition 1.3, and we derive Theorem 1.2 by verifying the remaining conditions when $\kappa = 0$. This computer-assisted verification is described in Appendix B and carried out in the attached Python 3 file using `python-flint`, a rigorous library for interval arithmetic.

1.1. **Related work.** For the spherical perceptron, the capacity threshold of [GD88] has been proved rigorously for all $\kappa \geqslant 0$ [ST03, Sto13a]. (See also [Sto13b] for some work on the $\kappa < 0$ case.) These works exploit the fact that the spherical perceptron with $\kappa \geqslant 0$ is a convex optimization problem. The Ising perceptron does not have this property, and our understanding of it is comparatively less complete. The replica heuristic also gives a prediction for the free energy of a positive-temperature version of this model [GD88, KM89], which was verified by [Tal00] at sufficiently high temperature using a rigorous version of the cavity method. The works [KR98, Tal99] showed that for the $\kappa = 0$ perceptron, there exists $\varepsilon > 0$ such that $\varepsilon \leqslant M_N/N \leqslant 1 - \varepsilon$ with high probability. The breakthrough work of Ding and Sun [DS18] showed that $\alpha_\star$ lower bounds the capacity for the $\kappa = 0$ perceptron, conditional on a numerical assumption. Very recently, [AT24] showed that 0.847 is a rigorous upper bound for the capacity in this model. Recent works have also shown the replica-symmetric formula for the free energy at low constraint density in generalized perceptron models [BNSX22], existence of a sharp threshold sequence [Xu21, NS23], and universality in the disorder [NS23]. We also mention the works [AS22, MZZ24] on algorithms for the negative spherical perceptron.

Another recent line of work originating with [APZ19] studied the **symmetric binary perceptron**, where the constraints in (1) are replaced by $|\langle \boldsymbol{g}^a, \boldsymbol{x}\rangle|/\sqrt{N} \leqslant \kappa$. Symmetry makes this model significantly more tractable (see §2.1 for more discussion); a series of remarkable works have established the limiting

[1] [KM89] studied a model with Bernoulli disorder, i.e. where the g_i^a are i.i.d. samples from $\mathrm{unif}(\pm 1)$ rather than $\mathcal{N}(0, 1)$. As [NS23] shows this model's sharp threshold sequence is universal with respect to any subgaussian disorder, we may work with gaussian disorder for convenience.

capacity [PX21, ALS22b], "frozen 1-RSB" structure [PX21], lognormal limit of partition function [ALS22b], and critical window [Alt23, SS23], and shed light on the performance of algorithms [ALS22a, GKPX22, GKPX23, BAKZ23].

1.2. **Notation.** While we introduce other parameters over the course of the proof, unless stated otherwise we send $N \to \infty$ first, treating the remaining parameters as small or large constants. Thus, we use $o_N(1)$ to denote a quantity vanishing with N, while notations like $o_\varepsilon(1)$ denote quantities independent of N tending to zero as the subscripted parameter tends to 0 or ∞ (which will be clear from context). We say an event occurs with high probability if it occurs with probability $1 - o_N(1)$. Further notations will be introduced in §4.1, before the main body of proofs.

Acknowledgements. I would like to thank Mehtaab Sawhney for pointing me to the reference [GZ00], and Will Perkins, Mehtaab Sawhney, Mark Sellke, and Nike Sun for helpful feedback on the manuscript. I am also grateful to Andrea Montanari and Huy Tuan Pham for a collaboration that inspired parts of this work. Thanks to Saba Lepsveridze for a helpful and motivating conversation. This work was supported by a Google PhD Fellowship, NSF CAREER grant DMS-1940092, and the Solomon Buchsbaum Research Fund at MIT.

2. Further background and proof outline

This section contains a technical overview of the paper, and is organized as follows. In §2.1, we review the AMP-conditioned moment method used in [DS18] to prove the capacity lower bound and discuss the main difficulties of proving the upper bound. In §2.2, we outline a new approach based on reducing to a planted model and argue that if three primary inputs (R1), (R2), (R3) hold, then the upper bound reduces to a tractable moment computation. §2.3 discusses the most difficult input (R1), and §2.4 discusses the more straightforward inputs (R2) and (R3). §2.5 discusses related work involving planted models. Finally, §2.6 heuristically carries out the aforementioned moment computation, explains how Condition 1.3 emerges from it, and gives numerical evidence for Condition 1.3 when $\kappa = 0$.

2.1. **AMP-conditioned moment method.** A natural approach to studying the limiting capacity is the moment method. Let $M = \alpha N$, and let $\boldsymbol{G} \in \mathbb{R}^{M \times N}$ have rows $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^M$. Then let $S_N(\boldsymbol{G}) = S_N^M$ (recall (1)) and $Z_N(\boldsymbol{G}) = |S_N(\boldsymbol{G})|$. If $\mathbb{E}[Z_N(\boldsymbol{G})] \ll 1$, then $S_N(\boldsymbol{G})$ is w.h.p. empty, and if $\mathbb{E}[Z_N(\boldsymbol{G})^2]/\mathbb{E}[Z_N(\boldsymbol{G})]^2$ is bounded, then $S_N(\boldsymbol{G})$ is nonempty with positive probability. If these two estimates hold for (respectively) $\alpha = \alpha_\star + \varepsilon$ and $\alpha = \alpha_\star - \varepsilon$, for any $\varepsilon > 0$, this shows the limiting capacity is $\alpha_\star$.

Let $\boldsymbol{m}_\star(\boldsymbol{G}) = \frac{1}{|S_N(\boldsymbol{G})|} \sum_{\boldsymbol{x} \in S_N(\boldsymbol{G})} \boldsymbol{x}$ denote the barycenter of the solution set $S_N(\boldsymbol{G})$. For models where $\boldsymbol{m}_\star(\boldsymbol{G}) = \boldsymbol{0}$, such as the symmetric binary perceptron [APZ19, PX21, ALS22b], this two-moment analysis often suffices to determine the limiting capacity. However, due to the asymmetry of the activation in the present model, $\boldsymbol{m}_\star(\boldsymbol{G})$ is typically macroscopic and random. It is expected that for any $\alpha > 0$, large-deviations events in the location of $\boldsymbol{m}_\star(\boldsymbol{G})$ dominate the first and second moments. Thus $Z_N(\boldsymbol{G})$ is typically exponentially smaller than $\mathbb{E}[Z_N(\boldsymbol{G})]$, and $\mathbb{E}[Z_N(\boldsymbol{G})]^2$ exponentially smaller than $\mathbb{E}[Z_N(\boldsymbol{G})^2]$, which causes the moment method to fail. For example, for the $\kappa = 0$ perceptron, $\frac{1}{N} \log \mathbb{E}[Z_N(\boldsymbol{G})]$ crosses zero at $\alpha = 1$, larger than $\alpha_\star(0) \approx 0.833$.

To overcome this difficulty, [DS18] and [Bol19] (the latter for the Sherrington–Kirkpatrick model) concurrently developed a conditional moment method, in which one conditions on a suitable proxy for $\boldsymbol{m}_\star(\boldsymbol{G})$ before computing moments. The conditioning step effectively recenters spins around $\boldsymbol{m}_\star(\boldsymbol{G})$, after which the moment method can potentially succeed.

The choice of conditioning is motivated by the TAP heuristic [TAP77] from statistical physics, which provides a powerful but non-rigorous framework to study this and other mean-field models. The central object in this framework is a **TAP free energy** $\mathcal{F}_{\rm TAP}(\boldsymbol{m},\boldsymbol{n})$, which is defined in (15) and can be thought of as a mean-field (dense graph) limit of the Bethe free energy of an appropriate message-passing system. It is expected that $\mathcal{F}_{\rm TAP}$ has a unique stationary point $(\boldsymbol{m},\boldsymbol{n}) \in [-1,1]^N \times \mathbb{R}^M$, with the following interpretation: $\boldsymbol{m}$ approximates the barycenter $\boldsymbol{m}_\star(\boldsymbol{G})$ of $S_N(\boldsymbol{G})$, and for each $a \in [M]$, n_a approximates a function of the average slack of the constraint $\langle \boldsymbol{g}^a, \boldsymbol{x}\rangle/\sqrt{N} \geqslant \kappa$ over solutions $\boldsymbol{x} \in S_N(\boldsymbol{G})$.[2] It is also predicted that $\boldsymbol{m}$ and $\boldsymbol{n}$ have specific coordinate profiles: for $(q_\star, \psi_\star)$ defined as the fixed point of a scalar recursion (see Condition 3.1) and $F = F_{1-q_\star}$ as in (13), the prediction is that the coordinates of $\dot{\boldsymbol{h}} = \operatorname{th}^{-1}(\boldsymbol{m})$ and $\hat{\boldsymbol{h}} = F^{-1}(\hat{\boldsymbol{h}})$ have empirical distribution approximating $\mathcal{N}(0,\psi_\star)$ and $\mathcal{N}(0,q_\star)$.[3]

An important fact we will exploit is that for fixed $(\boldsymbol{m},\boldsymbol{n})$, the stationarity condition $\nabla\mathcal{F}_{\rm TAP}(\boldsymbol{m},\boldsymbol{n}) = \boldsymbol{0}$ can be written as two **linear** equations in $\boldsymbol{G}$. These are the **TAP equations**, defined in (16). Using this fact, we can define a **planted model**, which plays an important motivational role in [DS18, Bol19]: we first chooose $(\boldsymbol{m},\boldsymbol{n})$ with aforementioned coordinate profile, and then sample $\boldsymbol{G}$ conditional on $\nabla\mathcal{F}_{\rm TAP}(\boldsymbol{m},\boldsymbol{n}) = \boldsymbol{0}$. (This is different from the more well-known notion of planted model introduced in [AC08], in that we are planting a TAP fixed point rather than a satisfying assignment; see §2.5 for further discussion.)

If we imagine for a moment that $\boldsymbol{G}$ were sampled from this planted model, then the moment method becomes tractable. In this model, the law of $\boldsymbol{G}$ conditional on $(\boldsymbol{m},\boldsymbol{n})$ remains gaussian because the TAP equations are linear in $\boldsymbol{G}$, and the conditional first and second moments of $Z_N(\boldsymbol{G})$ can be computed. They amount to tractable $O(1)$-dimensional optimization problems: for example, computing $\mathbb{E}[Z_N(\boldsymbol{G})|\boldsymbol{m},\boldsymbol{n}]$ amounts to optimizing the exponential-order contribution to the first moment from subsets of Σ_N defined by their inner products with $\boldsymbol{m}$ and $\dot{\boldsymbol{h}}$ (see §2.6 for details). The planted model removes the main difficulty of the macroscopically-fluctuating barycenter, giving the moment method a chance to succeed.

However, this planted model is different from the true model, in which the TAP solution $(\boldsymbol{m},\boldsymbol{n})$ depends on $\boldsymbol{G}$ in a complicated way. It is a priori unclear that these can be rigorously linked, because in the true model both existence and uniqueness of the TAP solution are not known. To carry out this approach, [DS18, Bol19] instead condition on a sequence of **approximate message passing** (AMP) iterates $(\boldsymbol{m}^0,\boldsymbol{n}^0,\dots,\boldsymbol{m}^k,\boldsymbol{n}^k)$ whose dependence on $\boldsymbol{G}$ is explicit. The AMP iteration was introduced in [Bol14, BM11], and is defined (roughly speaking, see (17)) by iterating the TAP equations. Its behavior can be understood through the powerful state evolution description of [Bol14, BM11, JM13, BMN20]: for any k not growing with N, state evolution exactly characterizes the limiting overlap structure of $(\boldsymbol{m}^0,\dots,\boldsymbol{m}^k)$ and $(\boldsymbol{n}^0,\dots,\boldsymbol{n}^k)$. Using this description, it can be shown that the AMP iterates converge to an approximate stationary point of $\mathcal{F}_{\rm TAP}$:

$$\lim_{k_1,k_2\to\infty} \operatorname*{p-lim}_{N\to\infty} N^{-1/2}\|(\boldsymbol{m}^{k_1},\boldsymbol{n}^{k_1}) - (\boldsymbol{m}^{k_2},\boldsymbol{n}^{k_2})\| = \lim_{k\to\infty} \operatorname*{p-lim}_{N\to\infty} N^{-1/2}\|\nabla\mathcal{F}_{\rm TAP}(\boldsymbol{m}^k,\boldsymbol{n}^k)\| = 0. \tag{2}$$

Here p-lim denotes limit in probability. It is in this sense that the AMP iterates are a proxy for $(\boldsymbol{m},\boldsymbol{n})$.

While the main advantages of conditioning on the AMP filtration are explicit dependence on $\boldsymbol{G}$ and state evolution, the main disadvantage is the greater complexity of the resulting moment calculation. Although the law of $\boldsymbol{G}$ conditional on $(\boldsymbol{m}^0,\boldsymbol{n}^0,\dots,\boldsymbol{m}^k,\boldsymbol{n}^k)$ remains gaussian, the conditional first and second moments of $Z_N(\boldsymbol{G})$ are now $O(k)$-dimensional optimization problems, in which one optimizes over subsets

[2] More generally, the statistical physics literature predicts that the Gibbs measure — here, the uniform measure on $S_N(\boldsymbol{G})$ — decomposes as a convex combination of well-concentrated "pure states," whose barycenters each approximate a stationary point of the TAP free energy [MPV87]. The present model is expected to be replica symmetric, meaning the entire Gibbs measure is one pure state.

[3] Here and throughout, nonlinearities such as th^{-1} and F^{-1} are applied coordinate-wise.

of Σ_N defined by their inner products with $\boldsymbol{m}^0, \ldots, \boldsymbol{m}^k$ and related vectors. These problems are not in general tractable. We note that [Bol19, BNSX22] successfully carry out this optimization in their respective settings, but only at sufficiently high temperature or low constraint density.

An important insight of [DS18] is that this approach still gives a tractable proof of the capacity lower bound, because — to show a lower bound for $Z_N(\boldsymbol{G})$ — one may truncate $Z_N(\boldsymbol{G})$ before computing moments. They construct a truncation $\widetilde{Z}_N(\boldsymbol{G})$ of $Z_N(\boldsymbol{G})$, restricting (among other conditions) to $\boldsymbol{x} \in \Sigma_N$ with prescribed inner products with $\boldsymbol{m}^0, \ldots, \boldsymbol{m}^k$. The conditional first moment of $\widetilde{Z}_N(\boldsymbol{G})$ is then explicit, while the conditional second moment becomes a 1-dimensional optimization. [DS18] shows that (under the aforementioned numerical condition) $\mathbb{E}[\widetilde{Z}_N(\boldsymbol{G})^2]/\mathbb{E}[\widetilde{Z}_N(\boldsymbol{G})]^2$ is bounded for any $\alpha < \alpha_\star$, which implies the capacity lower bound.

We mention that [BY22, BNSX22] carry out similar truncated second moment arguments in their respective settings, and the former improves the parameter regime where the method of [Bol19] obtains the replica symmetric free energy lower bound for the Sherrington–Kirkpatrick model.

The main difficulty of the capacity upper bound is that truncation is no longer available. Without it, proving the capacity upper bound within the AMP-conditioned moment method would require solving the above $O(k)$-dimensional optimization problem, which does not appear to be tractable.

2.2. **Approximate contiguity with planted model.** Our proof revisits and justifies the planted model heuristic described above, where we select $(\boldsymbol{m}, \boldsymbol{n})$ with appropriate coordinate profile and generate $\boldsymbol{G}$ conditional on $\nabla \mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m}, \boldsymbol{n}) = \boldsymbol{0}$. We will show that the true model is approximately contiguous to the planted model, in the sense of (3) below. So, rather than conditioning on the AMP filtration, we can condition directly on $(\boldsymbol{m}, \boldsymbol{n})$ after all. The conditional first moment of $Z_N(\boldsymbol{G})$ then reverts to a simple optimization in two, rather than $O(k)$, dimensions. This makes the capacity upper bound tractable.

The idea of passing by contiguity to a model with a planted TAP solution is also used in simultaneous joint work with A. Montanari and H. T. Pham [HMP24], on sampling from the Gibbs measure of a spherical mixed p-spin glass in total variation by an algorithmic implementation of stochastic localization [Eld20, AMS22]. A similar inequality to (3) appears as Proposition 4.4(d) therein. However, these two papers differ in both how this reduction is used, and how it is proved. While [HMP24] develops a reduction similar to (3), its main focus is to compute a high-precision estimate for the mean of a Gibbs measure, and the reduction to a planted model arises as a step in the analysis of this estimator. In the present paper, the reduction (3) is itself the main technical step, but the proof of it is also more challenging. Most notably, a key ingredient in the proof of (3), in both the present paper and [HMP24], is the uniqueness of the TAP fixed point in a certain region, see (R1) below. Whereas this ingredient is available in the spin glass setting of [HMP24] from known results, showing it in our setting requires new ideas, described in detail in §2.3.

We now state the approximate contiguity estimate. For small $\upsilon > 0$, let $\mathcal{S}_\upsilon$ denote the set of $(\boldsymbol{m}, \boldsymbol{n})$ whose coordinate profile is υ-close (in a suitable metric, see (27)) to that predicted by the TAP heuristic. We will show, roughly speaking, that there exists $C = O(1)$ such that for any $\boldsymbol{G}$-measurable event $\mathscr{E}$,

$$\mathbb{P}(\mathscr{E}) \leqslant C \sup_{(\boldsymbol{m},\boldsymbol{n}) \in \mathcal{S}_\upsilon} \mathbb{P}(\mathscr{E} | \nabla \mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m}, \boldsymbol{n}) = \boldsymbol{0})^{1/2} + o_N(1). \tag{3}$$

Remark 2.1. For reasons described below, we actually prove (3) for perturbations $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}, \mathcal{S}_{\varepsilon,\upsilon}$ of $\mathcal{F}_{\mathrm{TAP}}, \mathcal{S}_\upsilon$, and this qualification holds for the entire discussion below, even where not stated. These perturbations are defined in (24) and (27), and the formal version of (3) is given in Lemma 3.8.

We then take $\mathscr{E} = \{S_N(\boldsymbol{G}) \neq \varnothing\}$. The first moment bound will show that (under Condition 1.3) this event has vanishing probability in the planted model for any $\alpha > \alpha_\star$. Then (3) implies the conclusion.

Next, we discuss the proof of (3). The following two central ingredients establish uniqueness and existence of the critical point of $\mathcal{F}_{\text{TAP}}$ within the set $\mathcal{S}_v$, with high probability in the true model.

(R1) The expected number of critical points of $\mathcal{F}_{\text{TAP}}$ in $\mathcal{S}_v$ is $1 + o(1)$.
(R2) With high probability, there exists a critical point of $\mathcal{F}_{\text{TAP}}$ in $\mathcal{S}_v$.

Remark 2.2. Although the TAP perspective predicts $\mathcal{F}_{\text{TAP}}$ has a unique critical point in the full input space, uniqueness in $\mathcal{S}_v$ (and for the perturbed $\mathcal{F}^{\varepsilon}_{\text{TAP}}$) suffices for our proof.

A short argument based on the Kac–Rice formula [Kac48, Ric44] (see [AT09, Theorem 11.2.1] for a textbook treatment) shows that (3) follows from (R1), (R2), and the following additional input, which is a concentration condition on the change of volume term $|\det \nabla^2 \mathcal{F}_{\text{TAP}}(\boldsymbol{m}, \boldsymbol{n})|$ in the Kac–Rice formula. This argument is carried out in the proof of Lemma 3.8, see (33).

(R3) There exists $C' = O(1)$ such that uniformly over $(\boldsymbol{m}, \boldsymbol{n}) \in \mathcal{S}_v$,

$$\mathbb{E}[|\det \nabla^2 \mathcal{F}_{\text{TAP}}(\boldsymbol{m}, \boldsymbol{n})|^2 \big| \nabla \mathcal{F}_{\text{TAP}}(\boldsymbol{m}, \boldsymbol{n}) = \mathbf{0}]^{1/2} \leqslant C' \, \mathbb{E}[|\det \nabla^2 \mathcal{F}_{\text{TAP}}(\boldsymbol{m}, \boldsymbol{n})| \big| \nabla \mathcal{F}_{\text{TAP}}(\boldsymbol{m}, \boldsymbol{n}) = \mathbf{0}].$$

Remark 2.3. Since the probability in (3) is exponentially small, the proof can be carried out with $e^{o(N)}$ in place of C in (3). Consequently, showing (R1) and (R3) with $e^{o(N)}$ in place of $1 + o(1)$, $O(1)$ also suffices.

Input (R2) is proved constructively, by showing that AMP finds a critical point in the following sense.

(R4) There exists $r_k = o_k(1)$ such that with high probability, $\mathcal{F}_{\text{TAP}}$ has a unique critical point in a $r_k\sqrt{N}$-neighborhood of the AMP iterate $(\boldsymbol{m}^k, \boldsymbol{n}^k)$ (which lies in $\mathcal{S}_v$ by state evolution), for each sufficiently large k.

Input (R3) will follow from a classic spectral concentration argument of [GZ00]. We next discuss the proofs of (R1), (R4) and (R3), in that order.

2.3. **Topological trivialization of TAP free energy.** Condition (R1) is the most important input to the proof of (3). It is related to a remarkable line of work pioneered by [Fyo04, ABČ13], on the landscapes of random high-dimensional functions. This line of work has obtained expected critical point counts in a variety of settings, including spherical p-spin glasses [AB13, ABČ13] (see [Sub17, AG20, SZ21, BSZ20, HS23a] for matching second moment estimates in certain cases) spiked tensor models [BMMN19, ABL22], the TAP free energy for $\mathbb{Z}_2$-synchronization [FMM21, CFM23], bipartite spin glasses [Kiv23, McK24], the elastic manifold [BBM24], and generalized linear models [MBB20]. We also refer the reader to earlier non-rigorous work on this topic from the statistical physics literature [BM80, PP95, CLR05].

One phenomenon studied in these works is **topological trivialization** [FL14, Fyo15, BČNS22, HS23b], a phase transition where the number of critical points drops from e^{cN} to $e^{o(N)}$, or often $O(1)$. Proving (R1) amounts to showing **annealed topological trivialization** for $\mathcal{F}^{\varepsilon}_{\text{TAP}}$ on $\mathcal{S}_{\varepsilon,v}$.

The strategy of these works is to calculate the expected number of critical points using the Kac–Rice formula, evaluating the integrand using random matrix theory. Usually, the most complicated term in the integrand is the expected absolute value of the determinant of a random matrix. The most well-understood application is where the landscape is a spherical mixed p-spin glass, in which case this random matrix is a GOE shifted by a scalar multiple of the identity. For this case, an exact formula for this expected absolute determinant is known, see [ABČ13, Lemma 3.3]. This makes the Kac–Rice calculation explicit and tractable. In particular, [Fyo15, BČNS22] use this approach to determine the topologically trivial phase of spherical mixed p-spin glasses, and [HMP24] uses these results to establish (R1) for its application. However, for other models, results on topological trivialization are not as readily available.

It may still be possible to show (R1) for our model in this way, by evaluating the more general random determinant that appears in the Kac–Rice formula. This is the approach taken by [FMM21] which, for

$\mathbb{Z}_2$-synchronization at sufficiently large signal, shows annealed trivialization of suitably low-energy TAP solutions. Their method bounds the random determinant in the Kac–Rice formula using free probability [Voi91]. Furthermore, [BBM22] introduced a general tool for studying random determinants, showing that under mild conditions, their exponential order is the integral of $\log|\lambda|$ against the random matrix's limiting spectral measure. The spectral measure can then be studied using free probability.

Using this approach, one can often express the exponential order of the expected number of critical points as a variational formula, in which one term is an implicitly-defined function arising from free probability [Kiv23, HS23b, BBM24, McK24]. This yields a plausible way to show (R1): if we can show the variational formula for our model has value zero, annealed trivialization follows (in the sense of $e^{o(N)}$ expected critical points, which suffices by Remark 2.3). Recently, [HS23b] showed that this method can be carried out for multi-species spherical spin glasses, and it in fact characterizes the topologically trivial phase. Nonetheless, the variational formula is highly model-dependent — the proof in [HS23b] relies on a detailed understanding of a vector Dyson equation — and it is unclear if this method can be carried out for our model.

We instead show annealed topological trivialization by a different, and arguably more conceptual, approach. We will show that (R1) follows from the following variant of (R4):

(R5) In a model where we plant a stationary point $(\boldsymbol{m},\boldsymbol{n}) \in \mathcal{S}_{\varepsilon,v}$ of $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}$ (i.e. condition on $\nabla\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) = \mathbf{0}$), the same AMP iteration finds $(\boldsymbol{m},\boldsymbol{n})$, in the sense of (R4), with high probability.

This implication is proved in Lemma 4.15. Heuristically, the reason (R5) implies (R1) is that any realization of the disorder where $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}$ has $T > 1$ stationary points in $\mathcal{S}_{\varepsilon,v}$ can arise in T different planted models, and the event in (R5) can hold in only one of these T realizations. If the expected number of critical points is too large, (R5) cannot occur with the stated probability. The input (R5) can be proved by similar methods as (R4), as described in the next subsection. This method yields the first proof of topological trivialization that does not directly evaluate the Kac–Rice formula. We believe this is interesting in its own right.

2.4. **Critical point near late AMP iterates and determinant concentration.** This subsection discusses inputs (R4), (R5), and (R3), in that order. As state evolution ensures $\|\nabla\mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m}^k,\boldsymbol{n}^k)\| = o_k(1)\sqrt{N}$ (recall (2)), (R4) holds if, for example, $\mathcal{F}_{\mathrm{TAP}}$ is C-strongly concave in a neighborhood of late AMP iterates for $C > 0$ independent of k. Recent works in the variational inference literature [CFM23, CFLM23, Cel24] develop tools to establish this local concavity, and using them prove analogs of (R4) in several models.

In our setting, the fact that $\mathcal{F}_{\mathrm{TAP}}$ is **not** strongly concave near late AMP iterates introduces some complications. In fact, $\mathcal{F}_{\mathrm{TAP}}$ is strongly concave in $\boldsymbol{m}$, but convex — and problematically, not strongly convex — in $\boldsymbol{n}$. This issue is one reason we carry out the argument on a perturbation $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}$ of $\mathcal{F}_{\mathrm{TAP}}$, and a similarly perturbed AMP iteration and set $\mathcal{S}_{\varepsilon,v}$. (This perturbation serves several other purposes as well, described in Remark 4.5.) We will show that near late AMP iterates, $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}$ is strongly convex in $\boldsymbol{n}$ and $\mathcal{G}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m}) \equiv \inf_{\boldsymbol{n}} \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})$ is strongly concave, which is enough to imply (R4). Strong convexity of $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}$ in $\boldsymbol{n}$ holds (deterministically) essentially by construction.

Our proof of local strong concavity of $\mathcal{G}^{\varepsilon}_{\mathrm{TAP}}$ uses an idea introduced in [Cel24], to bound the Hessian at a late AMP iterate by applying a gaussian comparison inequality conditionally on the AMP iterates. [Cel24] considers a setting where AMP is performed on disorder $\boldsymbol{W} \sim \mathsf{GOE}(N)$ and the relevant Hessian is of the form $\boldsymbol{A} + \boldsymbol{W}$, where $\boldsymbol{A}$ is a function of a late AMP iterate. He develops a method to upper bound the top eigenvalue of this matrix by applying the Sudakov–Fernique inequality [Sud71, Fer75, Sud79] to the part of $\boldsymbol{W}$ that remains random after observing the AMP iterates. For us, the Hessian takes the form

$$\nabla^2\mathcal{G}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) = \boldsymbol{A}_1 + \frac{1}{N}\boldsymbol{G}^{\top}\boldsymbol{A}_2\boldsymbol{G} + \boldsymbol{\Delta}, \tag{4}$$

where $\boldsymbol{A}_1, \boldsymbol{A}_2$ are functions of $(\boldsymbol{m}, \boldsymbol{n})$, and $\boldsymbol{\Delta}$ is a low-rank term depending on both $\boldsymbol{G}$ and $(\boldsymbol{m}, \boldsymbol{n})$. We can arrange $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}$ so that $\boldsymbol{\Delta}$ does not contribute to the top eigenvalue. However, the post-AMP Sudakov–Fernique inequality does not apply to the remaining part, because — unlike for a GOE matrix — the quadratic form induced by $\boldsymbol{G}^\top \boldsymbol{A}_2 \boldsymbol{G}$ is not a gaussian process. We instead recast the top eigenvalue as a minimax program, via the identity (for $\boldsymbol{A}_2 \prec 0$)

$$\lambda_{\max}\left(\boldsymbol{A}_1 + \frac{1}{N}\boldsymbol{G}^\top \boldsymbol{A}_2 \boldsymbol{G}\right) = \sup_{\|\dot{\boldsymbol{v}}\|=1} \inf_{\widehat{\boldsymbol{v}} \in \mathbb{R}^M} \left\{ \langle \dot{\boldsymbol{v}}, \boldsymbol{A}_1 \dot{\boldsymbol{v}} \rangle - \langle \widehat{\boldsymbol{v}}, \boldsymbol{A}_2^{-1} \widehat{\boldsymbol{v}} \rangle + \frac{2}{\sqrt{N}} \langle \widehat{\boldsymbol{v}}, \boldsymbol{G} \dot{\boldsymbol{v}} \rangle \right\}.$$

This can be bounded by Gordon's inequality [Gor85, Gor88] conditional on the AMP iterates. Interestingly, the bound obtained in this way is sharp, matching a lower bound for the top eigenvalue obtained by free probability (see Remark 6.15).

The input (R5) follows similarly to (R4). We will show that with high probability over the planted model, late AMP iterates are approximate critical points of $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}$, near which $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m}, \cdot)$ is strongly convex and $\mathcal{G}^{\varepsilon}_{\mathrm{TAP}}$ is strongly concave. While the law of the disorder is different under the planted model, it remains gaussian and a similar analysis can be carried out.

We turn to (R3). An argument of [GZ00] implies that if a symmetric $\boldsymbol{X} \in \mathbb{R}^{N \times N}$ has independent (not necessarily centered or identically distributed) entries on and above the diagonal with uniformly bounded log-Sobolev constant, then $\frac{1}{\sqrt{N}}\boldsymbol{X}$ enjoys a strong spectral concentration property: any 1-Lipschitz spectral trace has $O(1)$-scale subgaussian fluctuations. We will see that conditional on $\nabla \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m}, \boldsymbol{n}) = \boldsymbol{0}$, $\det \nabla^2 \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m}, \boldsymbol{n})$ is a nonrandom multiple of $\det \nabla^2 \mathcal{G}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m}, \boldsymbol{n})$, which has form (4). The entries of this matrix are not independent, but we can rewrite it via the classical trick

$$\det\left(\boldsymbol{A}_1 + \frac{1}{N}\boldsymbol{G}^\top \boldsymbol{A}_2 \boldsymbol{G}\right) = \det \boldsymbol{X}, \qquad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{A}_1 & \frac{1}{\sqrt{N}}\boldsymbol{G}^\top \\ \frac{1}{\sqrt{N}}\boldsymbol{G} & -\boldsymbol{A}_2^{-1} \end{bmatrix}. \tag{5}$$

Conditional on $\nabla \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m}, \boldsymbol{n}) = \boldsymbol{0}$, the matrices $\boldsymbol{A}_1, \boldsymbol{A}_2$ are nonrandom while $\boldsymbol{G}$ has a (noncentered) gaussian law. Thus the result of [GZ00] applies to $\boldsymbol{X}$. (A slightly more elaborate version of (5) also accounts for the random low-rank spike $\boldsymbol{\Delta}$ in (4), see (76).)

From the above discussion, conditional on $\nabla \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m}, \boldsymbol{n}) = \boldsymbol{0}$, $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m}, \cdot)$ is strongly convex near $\boldsymbol{n}$ and $\mathcal{G}^{\varepsilon}_{\mathrm{TAP}}$ is w.h.p. strongly concave near $\boldsymbol{m}$. This implies that the spectrum of $\nabla^2 \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m}, \boldsymbol{n})$, and thus $\boldsymbol{X}$, is bounded away from zero, and provides the final ingredient to prove (R3): since $x \mapsto \log|x|$ is $O(1)$-Lipschitz away from zero, $\log|\det \boldsymbol{X}|$ is approximately a $O(1)$-Lipschitz spectral trace, which has $O(1)$-scale subgaussian fluctuations by [GZ00].

Remark 2.4. The fact that this log determinant has $O(1)$-scale fluctuations is only possible because the spectrum is bounded away from zero. For Wigner or Ginibre matrices, two examples of random matrices whose limiting bulk spectrum does include zero, the log determinant is known to have $\Theta(\sqrt{\log N})$ fluctuations [TV12, NV14], which diverges with N.

2.5. **On planted models.** Reducing to a planted model is a powerful tool in the analysis of random functions. This technique was introduced in the seminal work [AC08] and has seen a wide range of applications in the past decade. The underlying idea is to show contiguity of the original model with a planted version, defined as the null model conditioned on having a particular (randomly chosen) solution. If this holds, properties of the null model can be deduced from the planted version, which is often easier to analyze.

A frequent application of this method is to probe the local landscape around a typical solution. This is the original application of [AC08]: contiguity implies that the landscape around a typical solution to the null model can be approximated by the landscape around the planted solution in the planted model. From this, [AC08] shows the existence of a shattering transition in several random constraint satisfaction problems. This approach has since also been used to show "frozen 1RSB" structure in the symmetric binary perceptron [PX21, ALS22b] and shattering in the Gibbs measures of spherical spin glasses [AMS23b]. In a similar spirit, [HMP24] passes to a model with a planted TAP solution to obtain a high-precision estimate of the magnetization of a spherical spin glass.

In other applications, including the present work, the object of interest is not the local landscape, but the planted model is nonetheless simpler to analyze than the null model. Such applications include the RS free energy of random constraint satisfaction problems [BC16, BCH+16, CKPZ17, CEJ+18, CKM20], the 1RSB free energy of random regular NAE-SAT [SSZ22], and the Parisi formula for spherical spin glasses in the RS and zero-temperature 1RSB phases [HS23a]. Passage to a planted model is also a crucial tool in the analysis of sampling algorithms based on stochastic localization [AMS22, AMS23a].

2.6. **First moment in planted model.** In this subsection, we give a heuristic calculation of the first moment of $Z_N(\boldsymbol{G})$ in the planted model. The function $\mathscr{S}_\star(\lambda_1,\lambda_2)$ appearing in Condition 1.3 arises from this calculation, and under this condition the first moment method succeeds. At the end of this subsection, we also give numerical evidence for Condition 1.3 when $\kappa = 0$.

We work at constraint density $\alpha_\star$, setting $M = \lfloor \alpha_\star N \rfloor$ and $\boldsymbol{G}, S_N(\boldsymbol{G}), Z_N(\boldsymbol{G})$ as above with this M. Let $\mathbb{P}_{\mathsf{Pl}}^{\boldsymbol{m},\boldsymbol{n}}$ and $\mathbb{E}_{\mathsf{Pl}}^{\boldsymbol{m},\boldsymbol{n}}$ denote probability and expectation w.r.t. the model conditional on $\nabla \mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) = \boldsymbol{0}$. We will argue that under Condition 1.3, $\mathbb{E}_{\mathsf{Pl}}^{\boldsymbol{m},\boldsymbol{n}} Z_N(\boldsymbol{G}) = e^{o(N)}$. Then, at any constraint density $\alpha > \alpha_\star$, the $(\alpha-\alpha_\star)N$ additional constraints will make this moment exponentially small.

This argument will be made rigorous in §7. Per the above discussion, the rigorous version of this argument will plant a critical point of $\mathcal{F}_{\mathrm{TAP}}^{\varepsilon}$ rather than $\mathcal{F}_{\mathrm{TAP}}$.

We first define the function $\mathscr{S}_\star$. Let $(q_0,\psi_0) = (q_\star(\alpha_\star,\kappa),\psi_\star(\alpha_\star,\kappa))$ be defined by Condition 3.1. As discussed in §2.1, these are the variances of the (gaussian) coordinate empirical measures of $\hat{\boldsymbol{h}}$, $\dot{\boldsymbol{h}}$ predicted by the TAP heuristic, at constraint density $\alpha_\star$. Let $\dot{\boldsymbol{H}} \sim \mathcal{N}(0,\psi_0)$ and $\hat{\boldsymbol{H}} \sim \mathcal{N}(0,q_0)$. These two random variables may be defined on different probability spaces, as all expectations in the below formulas will involve random variables from only one space. Let $\boldsymbol{M} = \mathrm{th}(\dot{\boldsymbol{H}})$ and $\boldsymbol{N} = F_{1-q_0}(\hat{\boldsymbol{H}})$. For any measurable $\boldsymbol{\Lambda} : \mathbb{R} \to [-1,1]$, define

$$\mathsf{ent}(\boldsymbol{\Lambda}) = \mathbb{E}\,\mathcal{H}\left(\frac{1+\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})}{2}\right), \tag{6}$$

where $\mathcal{H}(x) = -x\log x - (1-x)\log(1-x)$ is the binary entropy function. Let Ψ be the complementary gaussian cumulative density function defined in (12). For $s \geqslant 0$, define

$$\mathscr{S}_\star(\boldsymbol{\Lambda},s) = \frac{1}{2}s^2\psi_0 + \mathsf{ent}(\boldsymbol{\Lambda}) + \alpha_\star\,\mathbb{E}\log\Psi\left\{\frac{\kappa - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})]}{q_0}\hat{\boldsymbol{H}} - \frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})]}{\psi_0}\boldsymbol{N}}{\sqrt{1-\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})]^2}{q_0}}} + s\boldsymbol{N}\right\}. \tag{7}$$

Finally, let $\boldsymbol{\Lambda}_{\lambda_1,\lambda_2}(x) = \mathrm{th}(\lambda_1 x + \lambda_2\mathrm{th}(x))$ and define

$$\mathscr{S}_\star(\boldsymbol{\Lambda}) = \inf_{s\geqslant 0}\mathscr{S}_\star(\boldsymbol{\Lambda},s), \qquad \mathscr{S}_\star(\lambda_1,\lambda_2) = \mathscr{S}_\star(\boldsymbol{\Lambda}_{\lambda_1,\lambda_2}). \tag{8}$$

These quantities have the following physical meanings. $\dot{\boldsymbol{H}},\hat{\boldsymbol{H}},\boldsymbol{M},\boldsymbol{N}$ are the coordinate distributions of $\dot{\boldsymbol{h}},\hat{\boldsymbol{h}},\boldsymbol{m},\boldsymbol{n}$. $\boldsymbol{\Lambda}$ specifies a set $\Sigma_N(\boldsymbol{\Lambda}) \subseteq \Sigma_N$ of points $\boldsymbol{x}$ where x_i has "conditional average" $\boldsymbol{\Lambda}(\dot{h}_i)$, in the sense that (informally, see (81))

$$\frac{1}{\#\{i\in[N] : \dot{h}_i \approx \dot{h}\}}\sum_{i\in[N]:\dot{h}_i\approx\dot{h}} x_i \approx \boldsymbol{\Lambda}(\dot{h}), \qquad \forall \dot{h}\in\mathbb{R}. \tag{9}$$

Note that $\mathsf{ent}(\mathbf{\Lambda})$ is the entropy of this set, that is (see Lemma 7.6)

$$\frac{1}{N}\log|\Sigma_N(\mathbf{\Lambda})| \simeq \mathsf{ent}(\mathbf{\Lambda}). \tag{10}$$

Here and throughout, $\simeq$ denotes equality up to additive $o_N(1)$.

Let $Z_N(\boldsymbol{G},\mathbf{\Lambda}) = |S_N(\boldsymbol{G}) \cap \Sigma_N(\mathbf{\Lambda})|$ denote the number of solutions with profile $\mathbf{\Lambda}$. We will see that for all $s \geqslant 0$, $\mathcal{S}_\star(\mathbf{\Lambda}, s)$ upper bounds the exponential order of $\mathbb{E}_{\mathsf{Pl}}^{\boldsymbol{m},\boldsymbol{n}} Z_N(\boldsymbol{G},\mathbf{\Lambda})$. Thus $\mathcal{S}_\star(\mathbf{\Lambda})$ also upper bounds this quantity, and $\mathbb{E}_{\mathsf{Pl}}^{\boldsymbol{m},\boldsymbol{n}} Z_N(\boldsymbol{G})$ is bounded (heuristically) by Laplace's principle:

$$\frac{1}{N}\log \mathbb{E}_{\mathsf{Pl}}^{\boldsymbol{m},\boldsymbol{n}} Z_N(\boldsymbol{G}) \simeq \sup_{\mathbf{\Lambda}} \left\{ \frac{1}{N}\log \mathbb{E}_{\mathsf{Pl}}^{\boldsymbol{m},\boldsymbol{n}} Z_N(\boldsymbol{G},\mathbf{\Lambda}) \right\} \leqslant \sup_{\mathbf{\Lambda}} \mathcal{S}_\star(\mathbf{\Lambda}) + o_N(1).$$

While this supremum is a priori an infinite-dimensional optimization problem, the following observation reduces it to two dimensions. For any a_1, a_2, a Lagrange multipliers calculation (see Lemma 7.10) shows that the maximum of $\mathsf{ent}(\mathbf{\Lambda})$ subject to $\mathbb{E}[\dot{\boldsymbol{H}}\mathbf{\Lambda}(\dot{\boldsymbol{H}})] = a_1$, $\mathbb{E}[\boldsymbol{M}\mathbf{\Lambda}(\dot{\boldsymbol{H}})] = a_2$ is attained by $\mathbf{\Lambda}$ of the form $\mathbf{\Lambda}_{\lambda_1,\lambda_2}$. As the remaining terms in $\mathcal{S}_\star(\mathbf{\Lambda}, s)$ depend on $\mathbf{\Lambda}$ only through $\mathbb{E}[\dot{\boldsymbol{H}}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]$ and $\mathbb{E}[\boldsymbol{M}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]$, we may restrict attention to $\mathbf{\Lambda}$ of this form. Thus

$$\frac{1}{N}\log \mathbb{E}_{\mathsf{Pl}}^{\boldsymbol{m},\boldsymbol{n}} Z_N(\boldsymbol{G}) \leqslant \sup_{\lambda_1,\lambda_2} \mathcal{S}_\star(\lambda_1,\lambda_2) + o_N(1).$$

This implies $\mathbb{E}_{\mathsf{Pl}}^{\boldsymbol{m},\boldsymbol{n}} Z_N(\boldsymbol{G}) = e^{o(N)}$ under Condition 1.3.

We next argue that $\mathcal{S}_\star(\mathbf{\Lambda}, s)$ upper bounds the exponential order of $\mathbb{E}_{\mathsf{Pl}}^{\boldsymbol{m},\boldsymbol{n}} Z_N(\boldsymbol{G},\mathbf{\Lambda})$, as claimed above. Due to (10), it suffices to bound the probability that some $\boldsymbol{x} \in \Sigma_N(\mathbf{\Lambda})$ satisfies all constraints. The planted model has the following law. Let $\dot{\boldsymbol{h}} \in \mathbb{R}^N$, $\hat{\boldsymbol{h}} \in \mathbb{R}^M$ have coordinate distributions approximating $\mathcal{N}(0,\psi_0)$, $\mathcal{N}(0,q_0)$, and let $\boldsymbol{m} = \mathrm{th}(\dot{\boldsymbol{h}})$, $\boldsymbol{n} = F_{1-q_0}(\hat{\boldsymbol{h}})$. A gaussian conditioning calculation (see Corollary 4.18) shows that conditional on $\nabla\mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) = \boldsymbol{0}$,

$$\frac{\boldsymbol{G}}{\sqrt{N}} \overset{d}{=} \frac{\hat{\boldsymbol{h}}\boldsymbol{m}^\top}{Nq_0} + \frac{\boldsymbol{n}\dot{\boldsymbol{h}}^\top}{N\psi_0} + \frac{P_{\boldsymbol{n}}^\perp \widetilde{\boldsymbol{G}} P_{\boldsymbol{m}}^\perp}{\sqrt{N}} + o_N(1).$$

Here $\widetilde{\boldsymbol{G}}$ is an i.i.d. copy of $\boldsymbol{G}$, $P_{\boldsymbol{m}}^\perp$ denotes the projection operator to the orthogonal complement of $\boldsymbol{m}$, and $o_N(1)$ is a matrix of operator norm $o_N(1)$. For any $\boldsymbol{x} \in \Sigma_N(\mathbf{\Lambda})$, we have $\frac{1}{N}\langle \boldsymbol{m}, \boldsymbol{x}\rangle \simeq \mathbb{E}[\boldsymbol{M}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]$ and $\frac{1}{N}\langle \dot{\boldsymbol{h}}, \boldsymbol{x}\rangle \simeq \mathbb{E}[\dot{\boldsymbol{H}}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]$. So,

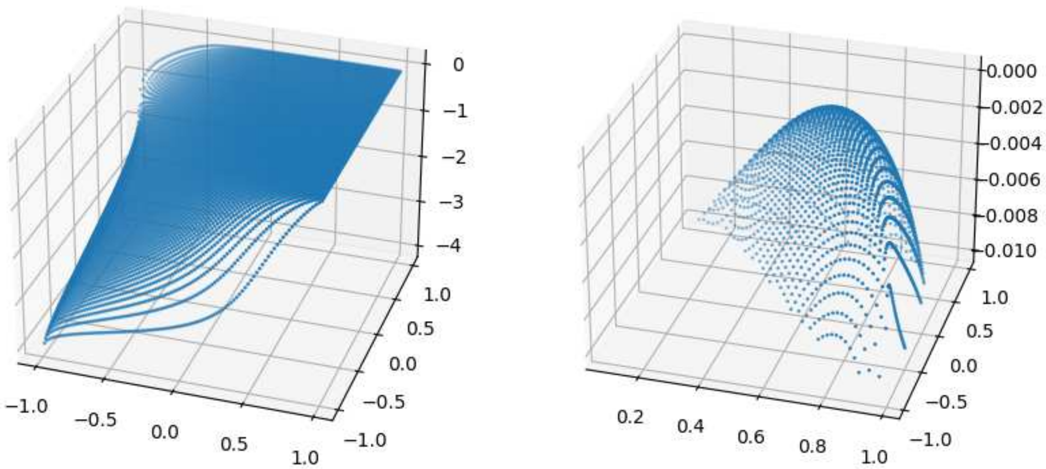
$$\frac{\boldsymbol{G}\boldsymbol{x}}{\sqrt{N}} \overset{d}{=} \frac{\mathbb{E}[\boldsymbol{M}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]}{q_0}\hat{\boldsymbol{h}} + \frac{\mathbb{E}[\dot{\boldsymbol{H}}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]}{\psi_0}\boldsymbol{n} + \sqrt{1 - \frac{\mathbb{E}[\boldsymbol{M}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]^2}{q_0}}\,\widetilde{\boldsymbol{g}} + o(\sqrt{N}),$$

where $\widetilde{\boldsymbol{g}} \sim \mathcal{N}(0, P_{\boldsymbol{n}}^\perp)$ and $o(\sqrt{N})$ denotes a vector of norm $o(\sqrt{N})$. Thus

$$\frac{1}{N}\log \mathbb{P}_{\mathsf{Pl}}^{\boldsymbol{m},\boldsymbol{n}}\left(\frac{\boldsymbol{G}\boldsymbol{x}}{\sqrt{N}} \geqslant \kappa\boldsymbol{1}\right) \simeq \frac{1}{N}\log \mathbb{P}\left\{ \widetilde{\boldsymbol{g}} \geqslant \frac{\kappa\boldsymbol{1} - \frac{\mathbb{E}[\boldsymbol{M}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]}{q_0}\hat{\boldsymbol{h}} - \frac{\mathbb{E}[\dot{\boldsymbol{H}}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]}{\psi_0}\boldsymbol{n}}{\sqrt{1 - \frac{\mathbb{E}[\boldsymbol{M}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]^2}{q_0}}} \right\}. \tag{11}$$

This can be bounded by a change of measure calculation also used in [DS18]. Let $\hat{\boldsymbol{g}} \sim \mathcal{N}(s\boldsymbol{n}, \boldsymbol{I}_N)$ for any $s \geqslant 0$. Note that conditional on $\langle \hat{\boldsymbol{g}}, \boldsymbol{n}\rangle = 0$, we have $\hat{\boldsymbol{g}} =_d \widetilde{\boldsymbol{g}}$. So, if S denotes the event in (11), then

$$\mathbb{P}(\widetilde{\boldsymbol{g}} \in S) \leqslant \frac{\mathbb{P}(\hat{\boldsymbol{g}} \in S)}{\mathbb{P}(\langle \hat{\boldsymbol{g}}, \boldsymbol{n}\rangle \approx 0)} \approx \exp\left(\frac{1}{2}s^2\psi_0 N\right)\mathbb{P}(\hat{\boldsymbol{g}} \in S).$$



(A) $x, y \in [-1,1]^2$ (B) $\overline{\mathscr{S}}_\star(\mathrm{th}^{-1}(x), \mathrm{th}^{-1}(y)) \geqslant -0.01$

FIGURE 1. Plots of $(x,y) \mapsto \overline{\mathscr{S}}_\star(\mathrm{th}^{-1}(x), \mathrm{th}^{-1}(y))$ for $\kappa = 0$. Figure 1a plots over $x, y \in [-1,1]^2$, while Figure 1b restricts to inputs with $\overline{\mathscr{S}}_\star(\mathrm{th}^{-1}(x), \mathrm{th}^{-1}(y)) \geqslant -0.01$. The plots lie below 0, and from Figure 1b it appears the unique maximizer is $(x,y) = (\mathrm{th}(1), 0)$, corresponding to $(\lambda_1, \lambda_2) = (1, 0)$.

Since $\widehat{\boldsymbol{h}}$ has coordinate distribution $\hat{\boldsymbol{H}}$, this implies (see Lemma 7.8 for formal statement) that (11) is bounded by

$$\frac{1}{2}s^2\psi_0 + \alpha_\star \mathbb{E} \log \Psi \left(\frac{\kappa - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})]}{q_0}\hat{\boldsymbol{H}} - \frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})]}{\psi_0}\boldsymbol{N}}{\sqrt{1 - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})]^2}{q_0}}} + s\boldsymbol{N} \right).$$

Combining with (10) shows that $\frac{1}{N} \log \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\mathsf{Pl}} Z_N(\boldsymbol{G}, \boldsymbol{\Lambda}) \leqslant \mathscr{S}_\star(\boldsymbol{\Lambda}, s) + o_N(1)$.

We conclude this subsection with a discussion of Condition 1.3. We expect $\boldsymbol{m}$ to approximate the barycenter of $S_N(\boldsymbol{G})$, and therefore that $\mathscr{S}_\star(\lambda_1, \lambda_2)$ is maximized by $(\lambda_1, \lambda_2) = (1, 0)$, corresponding to $\boldsymbol{\Lambda}_{\lambda_1,\lambda_2}(\dot{\boldsymbol{H}}) = \mathrm{th}(\dot{\boldsymbol{H}}) = \boldsymbol{M}$. Let

$$\overline{\mathscr{S}}_\star(\lambda_1, \lambda_2) = \mathscr{S}_\star(\boldsymbol{\Lambda}_{\lambda_1,\lambda_2}, \sqrt{1-q_0}),$$

which is an upper bound for $\mathscr{S}_\star$.

Lemma 2.5 (Proved in §7)**.** *The following holds.*

(a) The function $\mathscr{S}_\star(\lambda_1, \lambda_2)$ *attains its supremum on* $\mathbb{R}^2$.
(b) $\mathscr{S}_\star(1,0) = \overline{\mathscr{S}}_\star(1,0) = 0$.
(c) $\nabla \mathscr{S}_\star(1,0) = \nabla \overline{\mathscr{S}}_\star(1,0) = 0$.
(d) $\nabla^2 \mathscr{S}_\star(1,0) \preceq \nabla^2 \overline{\mathscr{S}}_\star(1,0)$

Claim 2.6 (Proved in Appendix B)**.** *For* $\kappa = 0$, *there exists* $C > 0$ *such that* $\nabla^2 \overline{\mathscr{S}}_\star(1,0) \preceq -CI$.

Lemma 2.5 is proved for all κ, while Claim 2.6 is verified numerically for $\kappa = 0$ using rigorous interval arithmetic. Together, they imply that for $\kappa = 0$, $(1,0)$ is a local maximum of $\mathscr{S}_\star$ and $\overline{\mathscr{S}}_\star$. In Figure 1, we provide a plot of $\overline{\mathscr{S}}_\star$ for the case $\kappa = 0$. This gives numerical evidence that $\overline{\mathscr{S}}_\star$, and therefore $\mathscr{S}_\star$, has global maximum $(1,0)$.

3. Formal statement of results

In this section we state our main result for general κ, Theorem 3.6. We also reduce Theorem 3.6 to two primary inputs: approximate contiguity with a planted model (Lemma 3.8) and the upper bound for the first moment in the planted model (Proposition 3.9), which are proved in §4–6 and §7.

3.1. Krauth–Mézard threshold.

We first define the threshold $\alpha_\star$ conjectured by [KM89], following the presentation of [DS18]. Define the standard gaussian density and complementary CDF by

$$\varphi(x) = \frac{\exp(-x^2/2)}{(2\pi)^{1/2}}, \qquad \Psi(x) = \int_x^\infty \phi(u)\,\mathrm{d}u. \tag{12}$$

Fix once and for all $\kappa \in \mathbb{R}$. For $q \in [0,1)$, define[4]

$$\mathcal{E}(x) = \frac{\varphi(x)}{\Psi(x)}, \qquad F_{1-q}(x) = \frac{\mathcal{E}}{(1-q)^{1/2}}\left(\frac{\kappa - x}{(1-q)^{1/2}}\right). \tag{13}$$

For $\psi \geqslant 0$ and $Z \sim \mathcal{N}(0,1)$, further define

$$P(\psi) = \mathbb{E}[\mathrm{th}(\psi^{1/2}Z)^2], \qquad R_\alpha(q) = \alpha\,\mathbb{E}[F_{1-q}(q^{1/2}Z)^2],$$

and define the Gardner free energy (or Gardner volume formula) by

$$\mathscr{G}(\alpha, q, \psi) = -\frac{(1-q)\psi}{2} + \mathbb{E}\log(2\mathrm{ch}(\psi^{1/2}Z)) - \alpha\,\mathbb{E}\log\Psi\left(\frac{\kappa - q^{1/2}Z}{(1-q)^{1/2}}\right). \tag{14}$$

The physical meanings of these formulas are best understood in terms of a heuristic derivation of the TAP free energy $\mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})$ and TAP equations, which we explain next. (These quantities will be formally defined in (15), (16).) If we regard $\boldsymbol{G}$ as a complete bipartite factor graph on N variables and M constraints, we can study the perceptron model by the standard **belief propagation** (BP) equations [MM09, Chapter 14]. In the mean-field (dense graph) limit, these equations simplify considerably. First, because the influence of any particular message is small, all the messages emanating from a particular variable $i \in [M]$ (resp. constraint $a \in [M]$) can be consolidated into a single message m_i (resp. n_a). The TAP variables $(\boldsymbol{m},\boldsymbol{n})$ thus represent these consolidated messages. The BP equations then become the TAP equations, and the **Bethe free energy** of this BP system becomes the TAP free energy. See [Méz17] for an example of this derivation in a related model.

Moreover, by central limit theorem considerations, we expect that the coordinates of $\dot{\boldsymbol{h}} = \mathrm{th}^{-1}(\boldsymbol{m})$ and $\widehat{\boldsymbol{h}} = F^{-1}_{1-\|\boldsymbol{m}\|^2/N}(\boldsymbol{n})$ have gaussian empirical measure. Let these gaussians have variance ψ and q, respectively; this is the physical meaning of these parameters. Then the BP consistency relations require that ψ, q satisfy the fixed-point equation $q = P(\psi)$, $\psi = R_\alpha(q)$, and the corresponding Bethe free energy is precisely $\mathscr{G}(\alpha, q, \psi)$. Finally, we expect $\alpha_\star$ to be the constraint density where this Bethe free energy crosses zero. Under the following condition, which was verified in [DS18] for $\kappa = 0$, this heuristic picture can be formalized into a definition of $\alpha_\star$.

Condition 3.1. There exist $0 < \alpha_{\mathrm{lb}} < \alpha_{\mathrm{ub}}$ and $0 < q_{\mathrm{lb}} < q_{\mathrm{ub}} < 1$ (depending on κ) such that the following holds. For any $\alpha \in (\alpha_{\mathrm{lb}}, \alpha_{\mathrm{ub}})$,

$$\sup_{q \in (q_{\mathrm{lb}}, q_{\mathrm{ub}})} (P \circ R_\alpha)'(q) < 1,$$

[4]The function F_{1-q} is denoted F_q in [DS18]. We change this notation to be consistent with the meaning of $F_{\varepsilon,\varrho}$ (18) appearing in our proofs.

and there is a unique $q_\star = q_\star(\alpha, \kappa) \in (q_{\rm lb}, q_{\rm ub})$ such that $q_\star = P(R_\alpha(q_\star))$. Let $\psi_\star = \psi_\star(\alpha,\kappa) = R_\alpha(q_\star)$. For $\alpha \in (\alpha_{\rm lb}, \alpha_{\rm ub})$, the function $\mathscr{G}_\star(\alpha) = \mathscr{G}(\alpha, q_\star(\alpha,\kappa), \psi_\star(\alpha,\kappa))$ is strictly decreasing, with a unique root $\alpha_\star = \alpha_\star(\kappa)$.

Proposition 3.2 ([DS18, Proposition 1.3]). *For $\kappa = 0$, Condition 3.1 holds for $\alpha_{\rm lb} = 0.833078599$, $\alpha_{\rm ub} = 0.833078600$, $q_{\rm lb} = 0.56394907949$, $q_{\rm ub} = 0.56394908030$.*

3.2. **Main result.** Throughout, let $\alpha_\star = \alpha_\star(\kappa)$ and $(q_0, \psi_0) = (q_\star(\alpha_\star,\kappa), \psi_\star(\alpha_\star,\kappa))$ be given by Condition 3.1. We now introduce two more numerical conditions needed for our main result, which will be verified for $\kappa = 0$ in Appendix B using rigorous interval arithmetic. In the below formulas, let $Z \sim \mathcal{N}(0,1)$.

Condition 3.3. We have $\alpha_\star\, \mathbb{E}[\operatorname{th}'(\psi_0^{1/2} Z)^2]\, \mathbb{E}[F'_{1-q_0}(q_0^{1/2}Z)^2] < 1$.

Condition 3.4. Define the functions $m : (-1,+\infty) \to (0,+\infty)$ and $\widehat{f}_0 : \mathbb{R} \to (0,+\infty)$ by

$$m(z) = \mathbb{E}[(z + \operatorname{ch}^2(\psi_0^{1/2}Z))^{-1}],$$

$$\widehat{f}_0(x) = -\frac{F'_{1-q_0}(x)}{1+(1-q_0)F'_{1-q_0}(x)} = \frac{\mathcal{E}'((\kappa - x)/(1-q_0)^{1/2})}{(1-q_0)(1-\mathcal{E}'((\kappa-x)/(1-q_0)^{1/2}))}.$$

(By Lemma 4.21(b) below, $\mathcal{E}'$ has image in $(0,1)$, and thus $\widehat{f}_0(x) > 0$.) Then, for $d_0 = \alpha_\star\, \mathbb{E}[F'_{1-q_0}(q_0^{1/2}Z)]$ and $\lambda : (-1,+\infty) \to \mathbb{R}$ defined by

$$\lambda(z) = z - \alpha_\star\, \mathbb{E}\left[\frac{\widehat{f}_0(q_0^{1/2}Z)}{1+m(z)\widehat{f}_0(q_0^{1/2}Z)}\right] - d_0,$$

we have $\lambda_0 \equiv \inf_{z>-1} \lambda(z) < 0$.

The following lemma shows that minimizer of λ exists and is the unique root of a decreasing function, and it suffices to check Condition 3.4 at the value $\lambda(z_0)$.

Lemma 3.5 (Proved in §6). *The function λ is differerentiable with $\lambda'(z) = 1 - \alpha_\star\theta(z)$, where $\theta : (-1,+\infty) \to (0,+\infty)$ is defined by*

$$\theta(z) = \mathbb{E}[(z+\operatorname{ch}^2(\psi_0^{1/2}Z))^{-2}]\, \mathbb{E}\left[\left(\frac{\widehat{f}_0(q_0^{1/2}Z)}{1+m(z)\widehat{f}_0(q_0^{1/2}Z)}\right)^2\right].$$

Moreover θ is continuous and strictly decreasing, with

$$\lim_{z\downarrow -1} \theta(z) = +\infty, \qquad \lim_{z\uparrow+\infty} \theta(z) = 0.$$

In particular θ has a well-defined inverse $\theta^{-1} : (0,+\infty) \to (-1,+\infty)$, and λ is strictly convex on $(-1,+\infty)$ with minimizer $z_0 = \theta^{-1}(\alpha_\star^{-1})$. Thus λ_0 defined in Condition 3.4 satisfies $\lambda_0 = \lambda(z_0)$.

Theorem 3.6 (Main result, general κ). *For any $\kappa \in \mathbb{R}$, under Conditions 1.3, 3.1, 3.3, and 3.4 the following holds. For any $\alpha > \alpha_\star(\kappa)$, we have $\lim_{N\to\infty} \mathbb{P}(M_N(\kappa)/N \geqslant \alpha) = 0$.*

Remark 3.7. The conditions in Theorem 3.6 serve the following purposes.

- Condition 1.3 controls the first moment of the partition function in the planted model.
- Condition 3.1 makes the threshold $\alpha_\star(\kappa)$ well-defined.
- Condition 3.3 ensures that the AMP iterates converge in the sense of (2).
- Condition 3.4 ensures that $\mathcal{G}^{\varepsilon}_{\rm TAP}$ (see §2.4) is locally concave near late AMP iterates.

With the exception of Appendix B, we will assume all conditions in Theorem 3.6 without further notice.

3.3. **Proof of Theorem 3.6.** We will carry out nearly the entire proof at constraint density $\alpha_\star$. Thus, we set $M = \lfloor \alpha_\star N \rfloor$ and define $\boldsymbol{G} \in \mathbb{R}^{M\times N}$ and $Z_N(\boldsymbol{G})$ as above.

The main step of the proof is a reduction to a planted model, formalized by Lemma 3.8 below. Let $\mathbb{P}$ denote the law of $\boldsymbol{G}$ with i.i.d. $\mathcal{N}(0,1)$ entries, and let $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$ be the planted law defined in Definition 4.3. This is the law of $\boldsymbol{G}$ conditional on $\nabla \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) = \boldsymbol{0}$, for a perturbation $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}$ of $\mathcal{F}_{\mathrm{TAP}}$ defined in (24). (These will actually be probability measures over $(\boldsymbol{G},\dot{\boldsymbol{g}},\widehat{\boldsymbol{g}})$ for auxiliary disorder $\dot{\boldsymbol{g}},\widehat{\boldsymbol{g}}$ defined below.) Let $\mathcal{S}_{\varepsilon,\upsilon}$ be a similar perturbation of $\mathcal{S}_\upsilon$ defined in (27).

Lemma 3.8 (Proved in §4–6)**.** *For any $(\boldsymbol{G},\dot{\boldsymbol{g}},\widehat{\boldsymbol{g}})$-measurable event $\mathscr{E}$ and any $\varepsilon,\upsilon > 0$, there exists $C = C(\varepsilon,\upsilon)$ such that*

$$\mathbb{P}(\mathscr{E}) \leqslant C \sup_{(\boldsymbol{m},\boldsymbol{n})\in\mathcal{S}_{\varepsilon,\upsilon}} \mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}(\mathscr{E})^{1/2} + o_N(1).$$

The following proposition controls the first moment of $Z_N(\boldsymbol{G})$ in the planted model, formalizing the heuristic calculation in §2.6. Here $\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$ denotes expectation with respect to $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$.

Proposition 3.9 (Proved in §7)**.** *For any $\delta > 0$, there exists $\varepsilon,\upsilon > 0$ such that*

$$\sup_{(\boldsymbol{m},\boldsymbol{n})\in\mathcal{S}_{\varepsilon,\upsilon}} \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[Z_N(\boldsymbol{G})] \leqslant e^{\delta N}.$$

From these two results, Theorem 3.6 follows by a short argument.

Proposition 3.10. *For any $\delta > 0$,*

$$\mathbb{P}[Z_N(\boldsymbol{G}) \leqslant e^{\delta N}] = 1 - o_N(1).$$

Proof. Let $\mathscr{E} = \{Z_N(\boldsymbol{G}) \leqslant e^{\delta N}\}$. By Lemma 3.8 and Markov's inequality,

$$\mathbb{P}(\mathscr{E}^c) \leqslant C \sup_{(\boldsymbol{m},\boldsymbol{n})\in\mathcal{S}_{\varepsilon,\upsilon}} \mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}(\mathscr{E}^c)^{1/2} + o_N(1) \leqslant Ce^{-\delta N/2} \sup_{(\boldsymbol{m},\boldsymbol{n})\in\mathcal{S}_{\varepsilon,\upsilon}} \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[Z_N(\boldsymbol{G})]^{1/2} + o_N(1).$$

By Proposition 3.9, we may choose ε,υ so this supremum is at most $e^{\delta N/4}$. □

Proof of Theorem 3.6. Let $M_{\mathrm{all}} = \lfloor \alpha N \rfloor$, and let $\boldsymbol{G}_{\mathrm{all}} = \binom{\boldsymbol{G}}{\widehat{\boldsymbol{G}}} \in \mathbb{R}^{M_{\mathrm{all}}\times N}$, where $\widehat{\boldsymbol{G}} \in \mathbb{R}^{(M_{\mathrm{all}}-M)\times N}$ has i.i.d. $\mathcal{N}(0,1)$ entries. Set $\delta < \frac{1}{2}(\alpha - \alpha_\star)\log\frac{1}{\Phi(\kappa)}$. Let $\mathscr{E} = \{Z_N(\boldsymbol{G}) \leqslant e^{\delta N}\}$, which satisfies $\mathbb{P}(\mathscr{E}) = 1 - o_N(1)$ by Proposition 3.10. Then

$$\mathbb{P}(M_N(\kappa)/N \geqslant \alpha) = \mathbb{P}(Z_N(\boldsymbol{G}_{\mathrm{all}}) > 0) \leqslant \mathbb{P}(\mathscr{E}^c) + \mathbb{E}[Z_N(\boldsymbol{G}_{\mathrm{all}})\mathbf{1}\{\mathscr{E}\}].$$

Since the rows of $\widehat{\boldsymbol{G}}$ are i.i.d. samples from $\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_N)$ independent of $\boldsymbol{G}$, for any $\boldsymbol{x} \in \Sigma_N$,

$$\mathbb{E}[Z_N(\boldsymbol{G}_{\mathrm{all}})\mathbf{1}\{\mathscr{E}\}] \leqslant e^{\delta N} \mathbb{P}_{\boldsymbol{g}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_N)}\left(\frac{\langle \boldsymbol{g},\boldsymbol{x}\rangle}{\sqrt{N}} \geqslant \kappa\right)^{M_{\mathrm{all}}-M} = e^{\delta N}\Phi(\kappa)^{M_{\mathrm{all}}-M} = o_N(1).$$ □

3.4. **TAP and AMP formulas.** In this subsection we provide the formulas for the TAP free energy, TAP equations, and AMP iteration mentioned above. The heuristic derivation of the former two were discussed below (14), and the latter is obtained by iterating the TAP equations in a suitable way.

The contents of this subsection play no formal role in the following proofs. We include these formulas for the reader's convenience, to allow a comparison with the ε-perturbed TAP free energy and AMP iteration defined in §4.2 below. (See also (36), (37) for the ε-perturbed TAP equations.) For $(\boldsymbol{m},\boldsymbol{n}) \in \mathbb{R}^N \times \mathbb{R}^M$, let $q(\boldsymbol{m}) = \|\boldsymbol{m}\|^2/N$ and $\psi(\boldsymbol{n}) = \|\boldsymbol{n}\|^2/N$. The TAP free energy for this model is

$$\mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) = \sum_{i=1}^N \mathcal{H}\left(\frac{1+m_i}{2}\right) + \sum_{a=1}^M \log\Psi\left(\frac{\kappa - \frac{\langle \boldsymbol{g}^a,\boldsymbol{m}\rangle}{\sqrt{N}} + (1-q(\boldsymbol{m}))n_a}{(1-q(\boldsymbol{m}))^{1/2}}\right) + \frac{N}{2}(1-q(\boldsymbol{m}))\psi(\boldsymbol{n}). \quad (15)$$

(Recall $\mathcal{H}(x) = -x\log(x) - (1-x)\log(1-x)$ is the binary entropy function.) The TAP equations are the stationarity conditions of $\mathcal{F}_{\mathrm{TAP}}$, and are

$$\boldsymbol{n} = F_{1-q(\boldsymbol{m})}(\widehat{\boldsymbol{h}}) \equiv F_{1-q(\boldsymbol{m})}\left(\frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} - b(\boldsymbol{m})\boldsymbol{n}\right), \qquad \boldsymbol{m} = \mathrm{th}(\dot{\boldsymbol{h}}) \equiv \mathrm{th}\left(\frac{\boldsymbol{G}^\top\boldsymbol{n}}{\sqrt{N}} - d(\boldsymbol{m},\boldsymbol{n})\boldsymbol{m}\right), \tag{16}$$

where

$$b(\boldsymbol{m}) = 1 - q(\boldsymbol{m}), \qquad d(\boldsymbol{m},\boldsymbol{n}) = \frac{1}{N}\sum_{a=1}^{M} F'_{1-q(\boldsymbol{m})}(n_a).$$

Recall that these are the mean-field limit of the BP equations for this model. The terms $b(\boldsymbol{m})\boldsymbol{n}$ and $d(\boldsymbol{m},\boldsymbol{n})\boldsymbol{m}$ compensate for backtracking and are known as the **Onsager correction** terms.

Let q_0, ψ_0 be as in Condition 3.1, and define

$$b_0 = \mathbb{E}[\mathrm{th}'(\psi_0^{1/2}Z)] = 1 - q_0, \qquad d_0 = \alpha_\star\,\mathbb{E}[F'_{1-q_0}(q_0^{1/2}Z)].$$

The AMP iteration associated to $\mathcal{F}_{\mathrm{TAP}}$ is given by $\boldsymbol{n}^{-1} = \boldsymbol{0} \in \mathbb{R}^M$, $\boldsymbol{m}^0 = q_0^{1/2}\boldsymbol{1} \in \mathbb{R}^N$, and

$$\boldsymbol{n}^k = F_{1-q_0}(\widehat{\boldsymbol{h}}^k) = F_{1-q_0}\left(\frac{\boldsymbol{G}\boldsymbol{m}^k}{\sqrt{N}} - b_0\boldsymbol{n}^{k-1}\right), \qquad \boldsymbol{m}^{k+1} = \mathrm{th}(\dot{\boldsymbol{h}}^{k+1}) = \mathrm{th}\left(\frac{\boldsymbol{G}^\top\boldsymbol{n}^k}{\sqrt{N}} - d_0\boldsymbol{m}^k\right). \tag{17}$$

4. Reduction to planted model

In this section we prove the central Lemma 3.8, using inputs from §5–6 as described below. §4.1–4.5 are devoted to this proof. §4.6 derives the law of the planted model $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$, which will be useful for calculations in the rest of the paper. To maintain a smooth presentation, we defer some proofs to §4.7, and routine but technical arguments to Appendix A.

4.1. **Parameter list and notations.** For convenience, we record here the order in which several parameters used in the proof of Lemma 3.8 are set. Each item in this list can be set sufficiently small or large depending on any preceding item.

- ε, size of the perturbation to the AMP iteration and TAP free energy.
- C_{cvx} and C_{bd}, estimates for ρ_ε (defined below, see (22)) and its derivatives.
- η, bound on strong convexity of $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})$ in $\boldsymbol{n}$, and C_{reg}, bound on regularity of $\nabla^2\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}$.
- r_0, radius around late AMP iterates where there is a unique critical point of $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}$.
- υ_0, accuracy of AMP iterate under which there is a unique critical point of $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}$ nearby.
- k, index of AMP iterate $(\boldsymbol{m}^k,\boldsymbol{n}^k)$ with accuracy υ_0.
- υ, tolerance in $\mathcal{S}_{\varepsilon,\upsilon}$.
- υ_1, accuracy of AMP iterate under which, by convex-concavity considerations, the nearby unique critical point lies in $\mathcal{S}_{\varepsilon,\upsilon}$.
- ℓ, index of AMP iterate $(\boldsymbol{m}^\ell,\boldsymbol{n}^\ell)$ with accuracy υ_1.
- N, problem dimension.

This information will be reviewed when these parameters are introduced. Notations such as $o_k(1)$ will denote quantities that tend to zero as the subscripted parameter tends to zero or infinity, which may depend arbitrarily on preceding items in this list but do not depend on subsequent items. We will use the term "absolute constant" to mean a constant depending on none of these parameters (but possibly depending on $\kappa, \alpha_\star, q_0, \psi_0$, which are fixed at the outset). Note that the statement of Lemma 3.8 is monotone in υ, and thus υ can be set small depending on the parameters preceding it in this list.

We also define more notations appearing in the proofs. Throughout, Z, Z', Z'' denote i.i.d. standard gaussians. We use $\mathcal{P}_2(\mathbb{R}^k)$ to denote the space of probability measures on $\mathbb{R}^k$ with bounded second moment and $\mathbb{W}_2$ to denote 2-Wasserstein distance. p-lim denotes limit in probability.

We often consider functions $\mathcal{F} : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}$, with input $(\boldsymbol{m}, \boldsymbol{n}) \in \mathbb{R}^N \times \mathbb{R}^M$. We will write $\nabla_{\boldsymbol{m}} \mathcal{F} \in \mathbb{R}^N$, $\nabla_{\boldsymbol{n}} \mathcal{F} \in \mathbb{R}^M$ for the restriction of $\nabla \mathcal{F}$ to the coordinates corresponding to $\boldsymbol{m}$ and $\boldsymbol{n}$. The Hessian restrictions $\nabla^2_{\boldsymbol{m},\boldsymbol{m}} \mathcal{F} \in \mathbb{R}^{N\times N}$, $\nabla^2_{\boldsymbol{m},\boldsymbol{n}} \mathcal{F} \in \mathbb{R}^{N\times M}$, and $\nabla^2_{\boldsymbol{n},\boldsymbol{n}} \mathcal{F} \in \mathbb{R}^{M\times M}$ are defined similarly. $P_{\boldsymbol{m}} = \boldsymbol{m}\boldsymbol{m}^\top / \|\boldsymbol{m}\|^2 \in \mathbb{R}^{N\times N}$ denotes the projection operator onto the span of $\boldsymbol{m}$, and $P_{\boldsymbol{m}}^\perp = \boldsymbol{I}_N - P_{\boldsymbol{m}}$ denotes the projection operator onto its orthogonal complement.

4.2. **Perturbed nonlinearities, AMP iteration, and TAP free energy.** We next introduce perturbed versions of the AMP iteration (17) and TAP free energy (15). The purpose of the various perturbations is discussed in Remark 4.5 below. Let $\varepsilon > 0$ be small. For $\varrho \geqslant 0$, define

$$\overline{F}_{\varepsilon,\varrho}(x) = \log \mathbb{E}\, \chi_\varepsilon(x + \varrho^{1/2} Z), \qquad \chi_\varepsilon(x) = \exp\left(-\frac{1}{2}\varepsilon x^2\right) \mathbb{P}(x + \varepsilon^{1/2} Z' \geqslant \kappa).$$

Then, define the perturbed nonlinearities

$$\mathrm{th}_\varepsilon(x) = \mathrm{th}(x) + \varepsilon x, \qquad F_{\varepsilon,\varrho}(x) = \overline{F}'_{\varepsilon,\varrho}(x). \tag{18}$$

An elementary calculation shows that explicitly,

$$\begin{aligned}\overline{F}_{\varepsilon,\varrho}(x) &= -\frac{1}{2}\log(1+\varepsilon\varrho) - \frac{\varepsilon x^2}{2(1+\varepsilon\varrho)} + \log \Psi\left(\frac{\kappa(1+\varepsilon\varrho) - x}{\sqrt{(\varrho + \varepsilon(1+\varepsilon\varrho))(1+\varepsilon\varrho)}}\right) \\ F_{\varepsilon,\varrho}(x) &= -\frac{\varepsilon x}{1+\varepsilon\varrho} + \frac{1}{\sqrt{(\varrho + \varepsilon(1+\varepsilon\varrho))(1+\varepsilon\varrho)}} \mathcal{E}\left(\frac{\kappa(1+\varepsilon\varrho) - x}{\sqrt{(\varrho + \varepsilon(1+\varepsilon\varrho))(1+\varepsilon\varrho)}}\right).\end{aligned} \tag{19}$$

Let

$$\varrho_\varepsilon(q,\psi) = \frac{1 - q + \varepsilon - \varepsilon^2(\psi+\varepsilon)}{1 - 2\varepsilon(\psi+\varepsilon)}.$$

Define perturbed variants of the functions $P, R_{\alpha_\star}$ by

$$P^\varepsilon(\psi) = \mathbb{E}[\mathrm{th}_\varepsilon((\psi+\varepsilon)^{1/2} Z)^2], \qquad R^\varepsilon(q,\psi) = \alpha_\star\, \mathbb{E}[F_{\varepsilon,\varrho_\varepsilon(q,\psi)}((q+\varepsilon)^{1/2} Z)^2],$$

and let $\zeta_\varepsilon(\psi) = R^\varepsilon(P^\varepsilon(\psi), \psi)$.

Proposition 4.1 (Proved in Appendix A). *There exists $\iota > 0$ such that for all sufficiently small $\varepsilon > 0$,*

$$\sup_{\psi \in [\psi_0 - \iota, \psi_0 + \iota]} \zeta'_\varepsilon(\psi) < 1,$$

and there is a unique solution $\psi_\varepsilon \in [\psi_0 - \iota, \psi_0 + \iota]$ to $\psi_\varepsilon = \zeta_\varepsilon(\psi_\varepsilon)$. Let $q_\varepsilon = P^\varepsilon(\psi_\varepsilon)$ and $\varrho_\varepsilon = \varrho_\varepsilon(q_\varepsilon, \psi_\varepsilon)$. We further have $(q_\varepsilon, \psi_\varepsilon, \varrho_\varepsilon) \to (q_0, \psi_0, 1 - q_0)$ as $\varepsilon \downarrow 0$.

Lemma 4.2 (Proved in §4.7). *We have $\varrho_\varepsilon = \mathbb{E}[\mathrm{th}'_\varepsilon((\psi_\varepsilon + \varepsilon)^{1/2} Z)]$.*

Let $d_\varepsilon = \alpha_\star\, \mathbb{E}[F'_{\varepsilon,\varrho_\varepsilon}((q_\varepsilon + \varepsilon)^{1/2} Z)]$. Further, let $\dot{\boldsymbol{g}} \sim \mathcal{N}(0, \boldsymbol{I}_N)$, $\widehat{\boldsymbol{g}} \sim \mathcal{N}(0, \boldsymbol{I}_M)$ be independent of $\boldsymbol{G}$. The perturbed AMP iteration is defined by $\boldsymbol{n}^{-1} = \boldsymbol{0} \in \mathbb{R}^M$, $\boldsymbol{m}^0 = q_\varepsilon^{1/2} \boldsymbol{1} \in \mathbb{R}^N$, and

$$\boldsymbol{n}^k = F_{\varepsilon,\varrho_\varepsilon}(\widehat{\boldsymbol{h}}^k) = F_{\varepsilon,\varrho_\varepsilon}\left(\frac{\boldsymbol{G}\boldsymbol{m}^k}{\sqrt{N}} + \varepsilon^{1/2}\widehat{\boldsymbol{g}} - \varrho_\varepsilon \boldsymbol{n}^{k-1}\right), \tag{20}$$

$$\boldsymbol{m}^{k+1} = \mathrm{th}_\varepsilon(\dot{\boldsymbol{h}}^{k+1}) = \mathrm{th}_\varepsilon\left(\frac{\boldsymbol{G}^\top \boldsymbol{n}^k}{\sqrt{N}} + \varepsilon^{1/2}\dot{\boldsymbol{g}} - d_\varepsilon \boldsymbol{m}^k\right). \tag{21}$$

Define the convex function $V_\varepsilon : \mathbb{R} \to \mathbb{R}$ and its dual

$$V_\varepsilon(\dot{h}) = \log(2\mathrm{ch}(\dot{h})) + \frac{1}{2}\varepsilon \dot{h}^2, \qquad V_\varepsilon^*(m) = \inf_{\dot{h}} \left\{ -m\dot{h} + V_\varepsilon(\dot{h}) \right\}.$$

Let $C_{\mathsf{cvx}}, C_{\mathsf{bd}} > 0$ be large in ε. Let $\rho_\varepsilon : \mathbb{R} \to \mathbb{R}$ be an (unspecified) thrice-differentiable function satisfying

$$\rho_\varepsilon(q_\varepsilon) = \varrho_\varepsilon, \qquad \rho_\varepsilon'(q_\varepsilon) = -1, \qquad \rho_\varepsilon''(q_\varepsilon) = C_{\mathsf{cvx}}, \tag{22}$$

such that the image of ρ_ε and its derivatives satisfies

$$\rho_\varepsilon \in [C_{\mathsf{bd}}^{-1}, C_{\mathsf{bd}}], \qquad |\rho_\varepsilon^{(p)}| \leqslant C_{\mathsf{bd}} \text{ for } p \in \{1,2,3\}. \tag{23}$$

(For every C_{cvx}, there exists C_{bd} such that this is possible.) Recall that for $(\boldsymbol{m},\boldsymbol{n}) \in \mathbb{R}^N \times \mathbb{R}^M$, we defined $q(\boldsymbol{m}) = \|\boldsymbol{m}\|^2/N$ and $\psi(\boldsymbol{n}) = \|\boldsymbol{n}\|^2/N$. The perturbed TAP free energy is

$$\mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m},\boldsymbol{n}) = \sum_{i=1}^N V_\varepsilon^*(m_i) + \varepsilon^{1/2}\langle \dot{\boldsymbol{g}}, \boldsymbol{m}\rangle + \sum_{a=1}^M \overline{F}_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}\left(\frac{\langle \boldsymbol{g}^a, \boldsymbol{m}\rangle}{\sqrt{N}} + \varepsilon^{1/2}\widehat{g}_a - \rho_\varepsilon(q(\boldsymbol{m}))n_a \right)$$
$$+ \frac{N}{2}\rho_\varepsilon(q(\boldsymbol{m}))\psi(\boldsymbol{n}). \tag{24}$$

We are now ready to define the planted model.

Definition 4.3. For $(\boldsymbol{m},\boldsymbol{n}) \in \mathbb{R}^N \times \mathbb{R}^M$, let $\mathbb{P}_{\varepsilon,\mathsf{Pl}}^{\boldsymbol{m},\boldsymbol{n}}$ denote the law of $(\boldsymbol{G}, \dot{\boldsymbol{g}}, \widehat{\boldsymbol{g}})$ conditional on $\nabla \mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m},\boldsymbol{n}) = \boldsymbol{0}$, and $\mathbb{E}_{\varepsilon,\mathsf{Pl}}^{\boldsymbol{m},\boldsymbol{n}}$ denote the corresponding expectation. ($\mathbb{P}$ and $\mathbb{E}$ continue to refer to the law of $(\boldsymbol{G}, \dot{\boldsymbol{g}}, \widehat{\boldsymbol{g}})$ with i.i.d. standard gaussian entries.)

Remark 4.4. As shown in Lemma 4.16 below, for any fixed $(\boldsymbol{m},\boldsymbol{n})$, $\nabla \mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m},\boldsymbol{n}) = \boldsymbol{0}$ is equivalent to two linear equations (36), (37) in $(\boldsymbol{G}, \dot{\boldsymbol{g}}, \widehat{\boldsymbol{g}})$, and thus in the planted model $(\boldsymbol{G}, \dot{\boldsymbol{g}}, \widehat{\boldsymbol{g}})$ remains gaussian.

Remark 4.5. The above perturbations serve the following purposes.

- $V_\varepsilon^*(m_i)$ regularizes the term $\mathcal{H}(\frac{1+m_i}{2})$ in the original $\mathcal{F}_{\mathrm{TAP}}$, avoiding the singular behavior of $\mathcal{F}_{\mathrm{TAP}}$ near the boundary of $[-1,1]^N$.
- $\overline{F}_{\varepsilon,\varrho_\varepsilon}$ is chosen so that $\mathcal{F}_{\mathrm{TAP}}^\varepsilon$ is strongly convex in $\boldsymbol{n}$. As a consequence, if we define

  $$\mathcal{G}_{\mathrm{TAP}}(\boldsymbol{m}) = \inf_{\boldsymbol{n}} \mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}), \qquad \mathcal{G}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m}) = \inf_{\boldsymbol{n}} \mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m},\boldsymbol{n}),$$

  then $\mathcal{G}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m})$ also regularizes $\mathcal{G}_{\mathrm{TAP}}(\boldsymbol{m})$, avoiding a singular behavior near the boundary of $\frac{1}{\sqrt{N}}\boldsymbol{G}\boldsymbol{m} \geqslant \kappa$. Indeed, $\mathcal{G}_{\mathrm{TAP}}(\boldsymbol{m}) = -\infty$ if this inequality fails in any coordinate.
- The nonlinearities th_ε and $F_{\varepsilon,\varrho_\varepsilon}$ have Lipschitz inverses, so that Euclidean distances in $(\boldsymbol{m},\boldsymbol{n})$ and $(\dot{\boldsymbol{h}}, \widehat{\boldsymbol{h}})$ are comparable.
- The perturbations $\varepsilon^{1/2}\widehat{\boldsymbol{g}}$ and $\varepsilon^{1/2}\dot{\boldsymbol{g}}$ are for technical convenience, as solutions to the original TAP equation (16) must lie on the codimension-one manifold

  $$\langle \dot{\boldsymbol{h}} + d(\boldsymbol{m},\boldsymbol{n})\boldsymbol{m}, \boldsymbol{m}\rangle = \frac{1}{\sqrt{N}}\langle \boldsymbol{n}, \boldsymbol{G}\boldsymbol{m}\rangle = \langle \boldsymbol{n}, \widehat{\boldsymbol{h}} + b(\boldsymbol{m})\boldsymbol{n}\rangle.$$

  With this perturbation, Kac–Rice arguments can take place on full space.
- We will see in §6 that the Hessian of $\mathcal{G}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m})$ is the sum of an anisotropic sample covariance matrix, a full-rank diagonal matrix, and a low-rank spike (recall (4)). The condition $\rho_\varepsilon''(q_\varepsilon) = C_{\mathsf{cvx}}$ ensures this spike cannot contribute to the top eigenvalue by adding a large negative spike to the Hessian. This simplifies the proof of strong concavity of $\mathcal{G}_{\mathrm{TAP}}^\varepsilon$ near late AMP iterates.

4.3. **Inputs to reduction.** We next state several inputs needed to prove Lemma 3.8. As anticipated in §2.2, the main input is Proposition 4.8, which formalizes criteria (R4) and (R5). First, we record that $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}$ is (deterministically) strongly convex in $\boldsymbol{n}$.

Proposition 4.6 (Proved in §4.7). *There exists $\eta = \eta(\varepsilon, C_{\mathsf{cvx}}, C_{\mathsf{bd}}) > 0$ such that $\nabla^2_{\boldsymbol{n},\boldsymbol{n}} \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) \succeq \eta \boldsymbol{I}_M$ for any $(\boldsymbol{m},\boldsymbol{n}) \in \mathbb{R}^N \times \mathbb{R}^M$.*

We next record a basic regularity estimate. Define

$$\nabla^2_{\diamond} \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) = \nabla^2_{\boldsymbol{m},\boldsymbol{m}} \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) - (\nabla^2_{\boldsymbol{m},\boldsymbol{n}} \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}))(\nabla^2_{\boldsymbol{n},\boldsymbol{n}} \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}))^{-1}(\nabla^2_{\boldsymbol{m},\boldsymbol{n}} \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}))^{\top}. \quad (25)$$

This arises as the Hessian of $\mathcal{G}^{\varepsilon}_{\mathrm{TAP}}$, as shown in Lemma 4.10 below.

Proposition 4.7 (Proved in Appendix A). *For any $D > 0$, there exists $C_{\mathsf{reg}} = C_{\mathsf{reg}}(\varepsilon, C_{\mathsf{cvx}}, C_{\mathsf{bd}}, D)$ such that over both $\mathbb{P}$ and $\mathbb{P}^{\boldsymbol{m}',\boldsymbol{n}'}_{\varepsilon,\mathsf{Pl}}$ for any $\|\boldsymbol{m}'\|^2, \|\boldsymbol{n}'\|^2 \leqslant DN$, with high probability the following holds. For all $\|\boldsymbol{m}\|^2, \|\boldsymbol{n}\|^2 \leqslant DN$, we have $\|\nabla^2 \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})\|_{\mathrm{op}} \leqslant C_{\mathsf{reg}}$.*

For $\dot{\boldsymbol{h}} \in \mathbb{R}^N$, $\widehat{\boldsymbol{h}} \in \mathbb{R}^M$, define the coordinate empirical measures

$$\mu_{\dot{\boldsymbol{h}}} = \frac{1}{N}\sum_{i=1}^N \delta(\dot{h}_i), \qquad \mu_{\widehat{\boldsymbol{h}}} = \frac{1}{M}\sum_{a=1}^M \delta(\widehat{h}_i). \quad (26)$$

In words, these are probability measures on $\mathbb{R}$ with mass $1/N$ on each $\dot{h}_i$ (resp. $1/M$, $\widehat{h}_i$). For $\upsilon > 0$, let

$$\begin{aligned}\mathcal{T}_{\varepsilon,\upsilon} &= \left\{(\dot{\boldsymbol{h}},\widehat{\boldsymbol{h}}) \in \mathbb{R}^N \times \mathbb{R}^M : \mathbb{W}_2(\mu_{\dot{\boldsymbol{h}}}, \mathcal{N}(0,\psi_\varepsilon+\varepsilon)), \mathbb{W}_2(\mu_{\widehat{\boldsymbol{h}}}, \mathcal{N}(0,q_\varepsilon+\varepsilon)) \leqslant \upsilon\right\}, \\ \mathcal{S}_{\varepsilon,\upsilon} &= \left\{(\mathrm{th}_\varepsilon(\dot{\boldsymbol{h}}), F_{\varepsilon,\varrho_\varepsilon}(\widehat{\boldsymbol{h}})) : (\dot{\boldsymbol{h}},\widehat{\boldsymbol{h}}) \in \mathcal{T}_{\varepsilon,\upsilon}\right\}.\end{aligned} \quad (27)$$

Let $(\boldsymbol{m}^k,\boldsymbol{n}^k)$ be as in (20), (21).

Proposition 4.8 (Proved in §5 and §6). *There exist $r_0 > 0$, $k_0 : \mathbb{R}_+ \to \mathbb{N}$, $\upsilon : \mathbb{R}_+ \times \mathbb{N} \to \mathbb{R}_+$, depending on $\varepsilon, C_{\mathsf{cvx}}, C_{\mathsf{bd}}, \eta, C_{\mathsf{reg}}$, and an absolute constant $C_{\mathsf{spec}} > 0$ such that the following holds. For any $\upsilon_0 > 0$ and $k \geqslant k_0(\upsilon_0)$, with high probability under $\mathbb{P}$:*

(a) $(\boldsymbol{m}^k,\boldsymbol{n}^k) \in \mathcal{S}_{\varepsilon,\upsilon_0}$,
(b) $\|\nabla \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m}^k,\boldsymbol{n}^k)\| \leqslant \upsilon_0\sqrt{N}$,
(c) $\nabla^2_{\diamond} \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) \preceq -C_{\mathsf{spec}}\boldsymbol{I}_N$ *for all* $(\boldsymbol{m},\boldsymbol{n})$ *such that* $\|(\boldsymbol{m},\boldsymbol{n}) - (\boldsymbol{m}^k,\boldsymbol{n}^k)\| \leqslant r_0\sqrt{N}$.

Moreover, let $\upsilon = \upsilon(\upsilon_0,k)$. For any $(\boldsymbol{m}',\boldsymbol{n}') \in \mathcal{S}_{\varepsilon,\upsilon}$, with high probability under $\mathbb{P}^{\boldsymbol{m}',\boldsymbol{n}'}_{\varepsilon,\mathsf{Pl}}$, the above three conclusions hold and:

(d) $\|(\boldsymbol{m}^k,\boldsymbol{n}^k) - (\boldsymbol{m}',\boldsymbol{n}')\| \leqslant \upsilon_0\sqrt{N}$.

The following concentration estimate follows by adapting an argument of [GZ00] and provides input (R3).

Lemma 4.9 (Proved in §6). *There exists C depending on $\varepsilon, C_{\mathsf{cvx}}$ such that for sufficiently small υ, uniformly over $(\boldsymbol{m},\boldsymbol{n}) \in \mathcal{S}_{\varepsilon,\upsilon}$,*

$$\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}\left[|\det \nabla^2 \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})|^2\right]^{1/2} \leqslant C\,\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}\left[|\det \nabla^2 \mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})|\right].$$

4.4. **Unique nearby critical point and conditioning lemma.** Lemma 4.11 below provides a criterion under which a function has a unique critical point near a given approximate critical point. Lemma 4.12 is a lemma about conditioning a random function on a random vector with a unique critical point nearby, which is an adaptation of the Kac–Rice formula. This important technical tool also appears as [HMP24, Lemma 3.6], where it is used in conjunction with known results on topological trivialization to condition on the TAP fixed point selected by AMP. Here, we use it with properties of the planted model provided by Proposition 4.8 to prove topological trivialization itself.

Lemma 4.10. *Let $U_1 \subseteq \mathbb{R}^N, U_2 \subseteq \mathbb{R}^M$ be open and convex. Suppose $\mathcal{F} : U_1 \times U_2 \to \mathbb{R}$ is twice differentiable and satisfies $\nabla^2_{\boldsymbol{n},\boldsymbol{n}}\mathcal{F}(\boldsymbol{m},\boldsymbol{n}) \succeq \eta \boldsymbol{I}_M$ for all $(\boldsymbol{m},\boldsymbol{n}) \in U_1 \times U_2$ for some $\eta > 0$, and $\mathcal{G}(\boldsymbol{m}) \equiv \min_{\boldsymbol{n}\in U_2} \mathcal{F}(\boldsymbol{m},\boldsymbol{n})$ exists for all $\boldsymbol{m} \in U_1$. Then $\boldsymbol{n}(\boldsymbol{m}) = \arg\min_{\boldsymbol{n}\in U_2} \mathcal{F}(\boldsymbol{m},\boldsymbol{n})$ is unique and differentiable, with*

$$\nabla \boldsymbol{n}(\boldsymbol{m}) = (\nabla^2_{\boldsymbol{n},\boldsymbol{n}}\mathcal{F}(\boldsymbol{m},\boldsymbol{n}(\boldsymbol{m})))^{-1}(\nabla^2_{\boldsymbol{m},\boldsymbol{n}}\mathcal{F}(\boldsymbol{m},\boldsymbol{n}(\boldsymbol{m})))^\top. \tag{28}$$

Moreover $\mathcal{G}$ is twice differentiable, with

$$\nabla \mathcal{G}(\boldsymbol{m}) = \nabla_{\boldsymbol{m}}\mathcal{F}(\boldsymbol{m},\boldsymbol{n}), \qquad \nabla^2\mathcal{G}(\boldsymbol{m}) = \nabla^2_\diamond\mathcal{F}(\boldsymbol{m},\boldsymbol{n}). \tag{29}$$

Proof. Strong convexity of $\mathcal{F}$ in $\boldsymbol{n}$ implies that $\boldsymbol{n}(\boldsymbol{m})$ is unique, and can be defined as the solution to $\nabla_{\boldsymbol{m}}\mathcal{F}(\boldsymbol{m},\boldsymbol{n}) = \boldsymbol{0}$. Then (28) follows from the implicit function theorem, while (29) follows from (28) and the chain rule. □

Lemma 4.11. *Let $\mathcal{F} : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}$ be twice differentiable and $(\boldsymbol{m}_0,\boldsymbol{n}_0) \in \mathbb{R}^N \times \mathbb{R}^M$. Let $\eta, C_{\rm reg}, \upsilon_0 > 0$, $r_0 = 2\eta^{-1}(1 + C_{\rm reg}\eta^{-1})^2\upsilon_0$, and $U = \mathsf{B}((\boldsymbol{m}_0,\boldsymbol{n}_0), r_0\sqrt{N})$. Suppose that:*

- *(C1) $\|\nabla\mathcal{F}(\boldsymbol{m}_0,\boldsymbol{n}_0)\| \leqslant \upsilon_0\sqrt{N}$,*
- *(C2) $\|\nabla^2\mathcal{F}(\boldsymbol{m},\boldsymbol{n})\|_{\rm op} \leqslant C_{\rm reg}$ for all $(\boldsymbol{m},\boldsymbol{n}) \in U$,*
- *(C3) $\nabla^2_{\boldsymbol{n},\boldsymbol{n}}\mathcal{F}(\boldsymbol{m},\boldsymbol{n}) \succeq \eta\boldsymbol{I}_M$ for all $(\boldsymbol{m},\boldsymbol{n}) \in \mathbb{R}^N \times \mathbb{R}^M$,*
- *(C4) $\nabla^2_\diamond\mathcal{F}(\boldsymbol{m},\boldsymbol{n}) \preceq -\eta\boldsymbol{I}_N$ for all $(\boldsymbol{m},\boldsymbol{n}) \in U$.*

Then, there is a unique $(\boldsymbol{m}_*,\boldsymbol{n}_*) \in U$ such that $\nabla\mathcal{F}(\boldsymbol{m}_*,\boldsymbol{n}_*) = \boldsymbol{0}$. Moreover, for sufficiently small (possibly in N) $\iota > 0$, the image of U under the map $\nabla\mathcal{F}$ contains $\mathsf{B}(\boldsymbol{0},\iota) \subseteq \mathbb{R}^N \times \mathbb{R}^M$ and is one-to-one on this set.

Proof. Let $U_1 = \mathsf{B}(\boldsymbol{m}_0, r_0\sqrt{N}) \subseteq \mathbb{R}^N$ and $U_2 = \mathbb{R}^M$. Item (C3) implies that the hypotheses of Lemma 4.10 hold for $\mathcal{F}^\varepsilon_{\rm TAP}$ with this (U_1,U_2). Thus, for $\boldsymbol{m} \in U_1$, $\boldsymbol{n}(\boldsymbol{m})$ and $\mathcal{G}(\boldsymbol{m})$ from Lemma 4.10 are well-defined, with derivatives given therein. If $(\boldsymbol{m}_*,\boldsymbol{n}_*)$ is a critical point of $\mathcal{F}$, then $\boldsymbol{m}_*$ must be a critical point of $\mathcal{G}$. Item (C4) and equation (29) imply that $\nabla^2\mathcal{G}(\boldsymbol{m}) \preceq -\eta\boldsymbol{I}_N$ for all $\boldsymbol{m} \in U_1$. Thus $\mathcal{G}$ has at most one critical point in U_1, and $\mathcal{F}$ has at most one critical point in $U_1 \times U_2 \supseteq U$.

We now show that such a point exists. By strong concavity of $\mathcal{F}(\boldsymbol{m}_0,\cdot)$ and (C1),

$$\|\boldsymbol{n}_0 - \boldsymbol{n}(\boldsymbol{m}_0)\| \leqslant \eta^{-1}\|\nabla_{\boldsymbol{n}}\mathcal{F}(\boldsymbol{m}_0,\boldsymbol{m}_0)\| \leqslant \eta^{-1}\upsilon_0\sqrt{N}.$$

Because $\|\nabla^2\mathcal{F}(\boldsymbol{m},\boldsymbol{n})\|_{\rm op} \leqslant C_{\rm reg}$, the map $(\boldsymbol{m},\boldsymbol{n}) \mapsto \nabla\mathcal{F}(\boldsymbol{m},\boldsymbol{n})$ is $C_{\rm reg}$-Lipschitz. Thus

$$\|\nabla\mathcal{G}(\boldsymbol{m}_0)\| = \|\nabla\mathcal{F}(\boldsymbol{m}_0,\boldsymbol{n}(\boldsymbol{m}_0))\| \leqslant \|\nabla\mathcal{F}(\boldsymbol{m}_0,\boldsymbol{n}_0)\| + C_{\rm reg}\|\boldsymbol{n}_0 - \boldsymbol{n}(\boldsymbol{m}_0)\| \leqslant (1 + C_{\rm reg}\eta^{-1})\upsilon_0\sqrt{N}.$$

By strong concavity of $\mathcal{G}$, there exists a critical point $\boldsymbol{m}_*$ of $\mathcal{G}$ with

$$\|\boldsymbol{m}_0 - \boldsymbol{m}_*\| \leqslant \eta^{-1}\|\nabla\mathcal{G}(\boldsymbol{m}_0)\| \leqslant \eta^{-1}(1 + C_{\rm reg}\eta^{-1})\upsilon_0\sqrt{N}.$$

Then, with $\boldsymbol{n}_* = \boldsymbol{n}(\boldsymbol{m}_*)$, $(\boldsymbol{m}_*,\boldsymbol{n}_*)$ is a critical point of $\mathcal{F}$. By conditions (C2), (C3) and equation (28), $\boldsymbol{n}(\cdot)$ is $C_{\rm reg}\eta^{-1}$-Lipschitz. So,

$$\|\boldsymbol{n}_0 - \boldsymbol{n}_*\| \leqslant \|\boldsymbol{n}_0 - \boldsymbol{n}(\boldsymbol{m}_0)\| + C_{\rm reg}\eta^{-1}\|\boldsymbol{m}_0 - \boldsymbol{m}_*\| \leqslant \eta^{-1}(1 + C_{\rm reg}\eta^{-1})^2\upsilon_0\sqrt{N}.$$

This shows that $(\boldsymbol{m}_*, \boldsymbol{n}_*) \in U$, proving the first claim, and furthermore $(\boldsymbol{m}_*, \boldsymbol{n}_*)$ lies in the interior of U. To show the second claim, we first prove that any $(\boldsymbol{m}, \boldsymbol{n}) \in U$ such that $\|\nabla \mathcal{F}(\boldsymbol{m}, \boldsymbol{n})\| \leqslant \iota$ lies in a neighborhood of $(\boldsymbol{m}^*, \boldsymbol{n}^*)$. First,

$$\|\boldsymbol{n} - \boldsymbol{n}(\boldsymbol{m})\| \leqslant \eta^{-1} \|\nabla_{\boldsymbol{n}} \mathcal{F}(\boldsymbol{m}, \boldsymbol{n})\| \leqslant \eta^{-1} \iota.$$

Similarly to above, $\|\nabla \mathcal{G}(\boldsymbol{m})\| \leqslant (1 + C_{\mathsf{reg}} \eta^{-1})\iota$, so we conclude

$$\|\boldsymbol{m} - \boldsymbol{m}_*\| \leqslant \eta^{-1}(1 + C_{\mathsf{reg}} \eta^{-1})\iota, \qquad \|\boldsymbol{n} - \boldsymbol{n}_*\| \leqslant \eta^{-1}(1 + C_{\mathsf{reg}} \eta^{-1})^2 \iota.$$

Thus $(\boldsymbol{m}, \boldsymbol{n})$ lies in a neighborhood of $(\boldsymbol{m}_*, \boldsymbol{n}_*)$, which is contained in U because $(\boldsymbol{m}_*, \boldsymbol{n}_*)$ lies in the interior of U. However, by Schur's lemma,

$$\det \nabla^2 \mathcal{F}(\boldsymbol{m}_*, \boldsymbol{n}_*) = \det \nabla^2_{\boldsymbol{n},\boldsymbol{n}} \mathcal{F}(\boldsymbol{m}_*, \boldsymbol{n}_*) \det \nabla^2_{\diamond} \mathcal{F}(\boldsymbol{m}_*, \boldsymbol{n}_*) \neq 0.$$

By the inverse function theorem, $\nabla \mathcal{F}$ is invertible in a neighborhood of $(\boldsymbol{m}_*, \boldsymbol{n}_*)$, mapping it bijectively to a neighborhood of $\mathbf{0}$. This concludes the proof. □

Lemma 4.12. *Let $\mathcal{F} : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}$ be a twice differentiable random function and $(\boldsymbol{m}_0, \boldsymbol{n}_0) \in \mathbb{R}^N \times \mathbb{R}^M$ be a random vector in the same probability space. Let $\eta, C_{\mathsf{reg}}, \upsilon_0, r_0$ be as in Lemma 4.11, and $U = \mathsf{B}((\boldsymbol{m}_0, \boldsymbol{n}_0), r_0\sqrt{N})$ (which is now a random set). Let $D > 0$ be arbitrary and $\mathscr{E}_0$ be the event that (C1) through (C4) hold and $\|\boldsymbol{m}_0\|^2, \|\boldsymbol{n}_0\|^2 \leqslant DN$.*

Let $\varphi_{\nabla \mathcal{F}(\boldsymbol{m},\boldsymbol{n})}$ denote the probability density of $\nabla \mathcal{F}(\boldsymbol{m}, \boldsymbol{n})$ w.r.t. Lebesgue measure on $\mathbb{R}^N \times \mathbb{R}^M$. Suppose $\varphi_{\nabla \mathcal{F}(\boldsymbol{m},\boldsymbol{n})}(\boldsymbol{z})$ is bounded for $(\boldsymbol{m}, \boldsymbol{n}) \in \mathbb{R}^N \times \mathbb{R}^M$ and $\boldsymbol{z}$ in a neighborhood of $\mathbf{0}$, and continuous in $\boldsymbol{z}$ in this neighborhood uniformly over $(\boldsymbol{m}, \boldsymbol{n})$. Then, for any event $\mathscr{E} \subseteq \mathscr{E}_0$ in the same probability space,

$$\mathbb{P}(\mathscr{E}) = \int_{\mathbb{R}^N \times \mathbb{R}^M} \mathbb{E}\left[|\det \nabla^2 \mathcal{F}(\boldsymbol{m}, \boldsymbol{n})| \mathbf{1}\{\mathscr{E} \cap \{(\boldsymbol{m}, \boldsymbol{n}) \in U\}\} \middle| \nabla \mathcal{F}(\boldsymbol{m}, \boldsymbol{n}) = \mathbf{0} \right] \varphi_{\nabla \mathcal{F}(\boldsymbol{m},\boldsymbol{n})}(\mathbf{0}) \, \mathsf{d}(\boldsymbol{m}, \boldsymbol{n}).$$

Proof. On $\mathscr{E}_0$, Lemma 4.11 implies there is a unique critical point $(\boldsymbol{m}_*, \boldsymbol{n}_*)$ of $\mathcal{F}$ in U. Moreover the image of U under $\nabla \mathcal{F}$ contains $\mathsf{B}(\mathbf{0}, \iota)$ for small ι and is one-to-one on this set. By the area formula, on $\mathscr{E}_0$,

$$1 = \frac{1}{|\mathsf{B}(\mathbf{0}, \iota)|} \int_U |\det \nabla^2 \mathcal{F}(\boldsymbol{m}, \boldsymbol{n})| \mathbf{1}\{\|\nabla \mathcal{F}(\boldsymbol{m}, \boldsymbol{n})\| \leqslant \iota\} \, \mathsf{d}(\boldsymbol{m}, \boldsymbol{n}).$$

Since $\mathscr{E} \subseteq \mathscr{E}_0$, multiplying both sides by $\mathbf{1}\{\mathscr{E}\}$ and taking expectations (via Fubini's theorem) yields

$$\begin{aligned}
\mathbb{P}(\mathscr{E}) &= \frac{1}{|\mathsf{B}(\mathbf{0}, \iota)|} \mathbb{E} \int_{\mathbb{R}^N \times \mathbb{R}^M} |\det \nabla^2 \mathcal{F}(\boldsymbol{m}, \boldsymbol{n})| \mathbf{1}\{\|\nabla \mathcal{F}(\boldsymbol{m}, \boldsymbol{n})\| \leqslant \iota\} \mathbf{1}\{\boldsymbol{m} \in U\} \, \mathsf{d}(\boldsymbol{m}, \boldsymbol{n}) \\
&= \int_{\mathbb{R}^N \times \mathbb{R}^M} \mathbb{E}\left[|\det \nabla^2 \mathcal{F}(\boldsymbol{m}, \boldsymbol{n})| \mathbf{1}\{\mathscr{E} \cap \{\boldsymbol{m} \in U\}\} \middle| \|\nabla \mathcal{F}(\boldsymbol{m}, \boldsymbol{n})\| \leqslant \iota \right] \frac{\mathbb{P}\{\|\nabla \mathcal{F}(\boldsymbol{m}, \boldsymbol{n})\| \leqslant \iota\}}{|\mathsf{B}(\mathbf{0}, \iota)|} \, \mathsf{d}(\boldsymbol{m}, \boldsymbol{n}).
\end{aligned}$$

We now take the limit as $\iota \to 0$. On $\mathscr{E}_0$, $|\det \nabla^2 \mathcal{F}(\boldsymbol{m}, \boldsymbol{n})| \leqslant C_{\mathsf{reg}}^{M+N}$. Since $\boldsymbol{m}_0, \boldsymbol{n}_0$ are bounded on $\mathscr{E}_0$, $\mathbf{1}\{\boldsymbol{m} \in U\} = 0$ almost surely for $\boldsymbol{m}$ outside a compact set. Since $\varphi_{\nabla \mathcal{F}(\boldsymbol{m},\boldsymbol{n})}(\boldsymbol{z})$ is bounded and continuous in $\boldsymbol{z}$, $\mathbb{P}\{\|\nabla \mathcal{F}(\boldsymbol{m}, \boldsymbol{n})\| \leqslant \iota\}/|\mathsf{B}(\mathbf{0}, \iota)|$ is bounded, and limits to $\varphi_{\nabla \mathcal{F}(\boldsymbol{m},\boldsymbol{n})}(\boldsymbol{z})$ as $\iota \to 0$. Taking $\iota \to 0$ gives the result by dominated convergence. □

4.5. **Proof of planted reduction.** We are now ready to prove Lemma 3.8. As anticipated in §2.2, Lemma 4.13 deduces (R2) from (R4), and Lemma 4.15 deduces (R1) from (R5). Then, Lemma 3.8 follows readily from the Kac–Rice formula.

Lemma 4.13. *For any $\upsilon > 0$, $\mathcal{S}_{\varepsilon,\upsilon}$ contains a critical point of $\mathcal{F}_{\mathrm{TAP}}^{\varepsilon}$ with high probability under $\mathbb{P}$.*

Proof. Let $\eta = \min(\eta(\varepsilon, C_{\mathsf{cvx}}, C_{\mathsf{bd}}), C_{\mathsf{spec}})$, where these terms are given by Propositions 4.6 and 4.8. Then, let $D = 2\max(q_\varepsilon, \psi_\varepsilon)$ and $C_{\mathsf{reg}} = C_{\mathsf{reg}}(\varepsilon, C_{\mathsf{cvx}}, C_{\mathsf{bd}}, D)$ be given by Proposition 4.7. Let r_0 be given by Proposition 4.8. Let υ_1 be small enough in υ that, with $r_1 = 2\eta^{-1}(1 + C_{\mathsf{reg}}\eta^{-1})^2\upsilon_1$, we have $r_1 \leqslant r_0$ and

$$\bigcup_{(\tilde{\boldsymbol{m}},\tilde{\boldsymbol{n}})\in\mathcal{S}_{\varepsilon,\upsilon_1}} \mathsf{B}((\tilde{\boldsymbol{m}},\tilde{\boldsymbol{n}}), r_1\sqrt{N}) \subseteq \mathcal{S}_{\varepsilon,\upsilon}. \tag{30}$$

(Since $\mathcal{S}_{\varepsilon,\upsilon}$ is the image of a product of two Wasserstein-balls under $(\mathrm{th}_\varepsilon, F_{\varepsilon,\varrho_\varepsilon})$, and $\mathrm{th}_\varepsilon^{-1}, F_{\varepsilon,\varrho_\varepsilon}^{-1}$ have Lipschitz constant depending only on ε, there exists υ_1 such that this holds.) Let $\ell = k_0(\upsilon_1)$ be given by Proposition 4.8. By Propositions 4.7 and 4.8, with high probability under $\mathbb{P}$,

- $\|\nabla^2\mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m},\boldsymbol{n})\|_{\mathrm{op}} \leqslant C_{\mathsf{reg}}$ for all $\|\boldsymbol{m}\|^2, \|\boldsymbol{n}\|^2 \leqslant DN$,
- $(\boldsymbol{m}^\ell, \boldsymbol{n}^\ell) \in \mathcal{S}_{\varepsilon,\upsilon_1}$,
- $\|\nabla\mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m}^\ell,\boldsymbol{n}^\ell)\| \leqslant \upsilon_1\sqrt{N}$,
- $\nabla_\diamond^2\mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m},\boldsymbol{n}) \preceq -C_{\mathsf{spec}}\boldsymbol{I}_N$ for all $\|(\boldsymbol{m},\boldsymbol{n}) - (\boldsymbol{m}^\ell,\boldsymbol{n}^\ell)\| \leqslant r_0\sqrt{N}$.

We now apply Lemma 4.11 with $(\mathcal{F}_{\mathrm{TAP}}^\varepsilon, \boldsymbol{m}^\ell, \boldsymbol{n}^\ell, \upsilon_1, r_1)$ in place of $(\mathcal{F}, \boldsymbol{m}_0, \boldsymbol{n}_0, \upsilon_0, r_0)$. The above points imply that conditions (C1), (C2), (C4) hold, and condition (C3) holds by Proposition 4.6. By Lemma 4.11, $\mathcal{F}_{\mathrm{TAP}}^\varepsilon$ has a critical point in $\mathsf{B}((\boldsymbol{m}^\ell,\boldsymbol{n}^\ell), r_1\sqrt{N})$. This lies in $\mathcal{S}_{\varepsilon,\upsilon}$ by (30). □

The following lemma shows that the condition in Lemma 4.12 regarding $\varphi_{\nabla\mathcal{F}}$ holds for $\mathcal{F} = \mathcal{F}_{\mathrm{TAP}}^\varepsilon$.

Lemma 4.14 (Proved in §4.7). *The density $\varphi_{\nabla\mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m},\boldsymbol{n})}(\boldsymbol{z})$ under $\mathbb{P}$ is bounded for $(\boldsymbol{m},\boldsymbol{n}) \in \mathbb{R}^N \times \mathbb{R}^M$ and $\boldsymbol{z}$ in a neighborhood of $\boldsymbol{0}$, and continuous in $\boldsymbol{z}$ in this neighborhood uniformly over $(\boldsymbol{m},\boldsymbol{n})$.*

Lemma 4.15. *Let Crt_υ denote the set of critical points of $\mathcal{F}_{\mathrm{TAP}}^\varepsilon$ in $\mathcal{S}_{\varepsilon,\upsilon}$. For small $\upsilon > 0$, $\mathbb{E}\,|\mathsf{Crt}_\upsilon| \leqslant 1 + o_N(1)$.*

Proof. By the Kac–Rice formula,

$$\mathbb{E}\,|\mathsf{Crt}_\upsilon| = \int_{\mathcal{S}_{\varepsilon,\upsilon}} \mathbb{E}_{\varepsilon,\mathrm{Pl}}^{\boldsymbol{m},\boldsymbol{n}}\left[|\det\nabla^2\mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m},\boldsymbol{n})|\right]\varphi_{\nabla\mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m},\boldsymbol{n})}(\boldsymbol{0})\,\mathsf{d}(\boldsymbol{m},\boldsymbol{n}). \tag{31}$$

As above, let $\eta = \min(\eta(\varepsilon, C_{\mathsf{cvx}}, C_{\mathsf{bd}}), C_{\mathsf{spec}})$, $D = 2\max(q_\varepsilon, \psi_\varepsilon)$, and $C_{\mathsf{reg}} = C_{\mathsf{reg}}(\varepsilon, C_{\mathsf{cvx}}, C_{\mathsf{bd}}, D)$. Let r_0 be given by Proposition 4.8, and

$$\upsilon_0 = \frac{\eta r_0}{2(1 + C_{\mathsf{reg}}\eta^{-1})^2}.$$

Then set $k = k_0(\upsilon_0)$, where $k_0(\cdot)$ is as in Proposition 4.8. Let $\mathscr{E}$ be the event that:

- $\|\boldsymbol{m}^k\|^2, \|\boldsymbol{n}^k\|^2 \leqslant DN$,
- $\|\nabla^2\mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m},\boldsymbol{n})\|_{\mathrm{op}} \leqslant C_{\mathsf{reg}}$ for all $\|\boldsymbol{m}\|^2, \|\boldsymbol{n}\|^2 \leqslant DN$,
- $\|\nabla\mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m}^k,\boldsymbol{n}^k)\| \leqslant \upsilon_0\sqrt{N}$,
- $\nabla_\diamond^2\mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m},\boldsymbol{n}) \preceq -C_{\mathsf{spec}}\boldsymbol{I}_N$ for all $\|(\boldsymbol{m},\boldsymbol{n}) - (\boldsymbol{m}^k,\boldsymbol{n}^k)\| \leqslant r_0\sqrt{N}$.

We claim that $\mathscr{E} \subseteq \mathscr{E}_0$, where $\mathscr{E}_0$ is the event defined in Lemma 4.12 with $(\mathcal{F}_{\mathrm{TAP}}^\varepsilon, \boldsymbol{m}^k, \boldsymbol{n}^k)$ for $(\mathcal{F}, \boldsymbol{m}_0, \boldsymbol{n}_0)$ (and thus $U = \mathsf{B}((\boldsymbol{m}^k,\boldsymbol{n}^k), r_0\sqrt{N})$). The above points imply conditions (C1), (C2), (C4), and condition (C3) follows from Proposition 4.6. By Lemma 4.14, Lemma 4.12 applies. Thus,

$$1 \geqslant \mathbb{P}(\mathscr{E}) = \int_{\mathbb{R}^N\times\mathbb{R}^M} \mathbb{E}_{\varepsilon,\mathrm{Pl}}^{\boldsymbol{m},\boldsymbol{n}}\left[|\det\nabla^2\mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m},\boldsymbol{n})|\mathbf{1}\{\mathscr{E}\cap\{(\boldsymbol{m},\boldsymbol{n})\in U\}\}\right]\varphi_{\nabla\mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m},\boldsymbol{n})}(\boldsymbol{0})\,\mathsf{d}(\boldsymbol{m},\boldsymbol{n}). \tag{32}$$

Let $\upsilon \leqslant \min(\upsilon(\upsilon_0, k), \upsilon(r_0, k))$, for $\upsilon(\cdot,\cdot)$ as in Proposition 4.8. Define (compare with (31))

$$I_1 = \int_{\mathcal{S}_{\varepsilon,\upsilon}} \mathbb{E}_{\varepsilon,\mathrm{Pl}}^{\boldsymbol{m},\boldsymbol{n}}\left[|\det\nabla^2\mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m},\boldsymbol{n})|\mathbf{1}\{\mathscr{E}\cap\{(\boldsymbol{m},\boldsymbol{n})\in U\}\}\right]\varphi_{\nabla\mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m},\boldsymbol{n})}(\boldsymbol{0})\,\mathsf{d}(\boldsymbol{m},\boldsymbol{n})$$

and $I_2 = \mathbb{E}\,|\mathrm{Crt}_\upsilon| - I_1$. By Propositions 4.7 and 4.8, for any $(\boldsymbol{m},\boldsymbol{n}) \in \mathcal{S}_{\varepsilon,\upsilon}$, we have $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}}(\mathscr{E} \cap \{(\boldsymbol{m},\boldsymbol{n}) \in U\}) \geqslant 1 - \iota$ for some $\iota = o_N(1)$. By Cauchy–Schwarz and Lemma 4.9,

$$
\begin{aligned}
I_2 &= \int_{\mathcal{S}_{\varepsilon,\upsilon}} \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}} \left[|\det \nabla^2 \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})| \mathbf{1}\{(\mathscr{E} \cap \{(\boldsymbol{m},\boldsymbol{n}) \in U\})^c\} \right] \varphi_{\nabla \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})}(\mathbf{0})\, \mathrm{d}(\boldsymbol{m},\boldsymbol{n}) \\
&\leqslant \int_{\mathcal{S}_{\varepsilon,\upsilon}} \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}} \left[|\det \nabla^2 \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})|^2 \right]^{1/2} \mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}} \left[(\mathscr{E} \cap \{(\boldsymbol{m},\boldsymbol{n}) \in U\})^c \right]^{1/2} \varphi_{\nabla \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})}(\mathbf{0})\, \mathrm{d}(\boldsymbol{m},\boldsymbol{n}) \\
&\leqslant C\iota^{1/2} \int_{\mathcal{S}_{\varepsilon,\upsilon}} \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}} \left[|\det \nabla^2 \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})| \right] \varphi_{\nabla \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})}(\mathbf{0})\, \mathrm{d}(\boldsymbol{m},\boldsymbol{n}) \overset{(31)}{=} C\iota^{1/2}\, \mathbb{E}\,|\mathrm{Crt}_\upsilon|.
\end{aligned}
$$

So, $I_1 \geqslant (1 - C\iota^{1/2})\, \mathbb{E}\,|\mathrm{Crt}_\upsilon|$. Since (32) implies $I_1 \leqslant 1$, and $\iota = o_N(1)$, the conclusion follows. □

Proof of Lemma 3.8. Set $\upsilon > 0$ small enough that Lemma 4.15 holds. Let $\mathscr{E}_1$ be the event that $\mathcal{F}^\varepsilon_{\mathrm{TAP}}$ has a critical point in $\mathcal{S}_{\varepsilon,\upsilon}$. By the Kac–Rice formula,

$$
\begin{aligned}
\mathbb{P}(\mathscr{E} \cap \mathscr{E}_1) &\leqslant \mathbb{E}[\mathbf{1}\{\mathscr{E} \cap \mathscr{E}_1\} |\mathrm{Crt}_\upsilon|] \\
&= \int_{\mathcal{S}_{\varepsilon,\upsilon}} \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}} \left[|\det \nabla^2 \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})| \mathbf{1}\{\mathscr{E} \cap \mathscr{E}_1\} \right] \varphi_{\nabla \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})}(\mathbf{0})\, \mathrm{d}(\boldsymbol{m},\boldsymbol{n}).
\end{aligned}
$$

This is bounded by

$$
\begin{aligned}
&\int_{\mathcal{S}_{\varepsilon,\upsilon}} \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}} \left[|\det \nabla^2 \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})|^2 \right]^{1/2} \mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}}(\mathscr{E})^{1/2} \varphi_{\nabla \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})}(\mathbf{0})\, \mathrm{d}(\boldsymbol{m},\boldsymbol{n}) \\
&\leqslant C \sup_{(\boldsymbol{m},\boldsymbol{n}) \in \mathcal{S}_{\varepsilon,\upsilon}} \mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}}(\mathscr{E})^{1/2} \times \int_{\mathcal{S}_{\varepsilon,\upsilon}} \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}} \left[|\det \nabla^2 \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})| \right] \varphi_{\nabla \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})}(\mathbf{0})\, \mathrm{d}(\boldsymbol{m},\boldsymbol{n}) \\
&\leqslant C \sup_{(\boldsymbol{m},\boldsymbol{n}) \in \mathcal{S}_{\varepsilon,\upsilon}} \mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}}(\mathscr{E})^{1/2} \cdot \mathbb{E}\,|\mathrm{Crt}_\upsilon| \overset{Lem.\ 4.15}{\leqslant} (1 + o_N(1)) C \sup_{(\boldsymbol{m},\boldsymbol{n}) \in \mathcal{S}_{\varepsilon,\upsilon}} \mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}}(\mathscr{E})^{1/2}.
\end{aligned}
\tag{33}
$$

The result follows because $\mathbb{P}(\mathscr{E}) \leqslant \mathbb{P}(\mathscr{E} \cap \mathscr{E}_1) + \mathbb{P}(\mathscr{E}_1^c)$, and $\mathbb{P}(\mathscr{E}_1^c) = o_N(1)$ by Lemma 4.13. □

4.6. **Conditional law in planted model.** Having proved the reduction to the planted model $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}}$, we now calculate the law of the disorder in it. This is stated in Lemma 4.17 for general $(\boldsymbol{m},\boldsymbol{n})$, and Corollary 4.18 for $(\boldsymbol{m},\boldsymbol{n}) \in \mathcal{S}_{\varepsilon,\upsilon}$. The following lemma is proved by direct differentiation of $\mathcal{F}^\varepsilon_{\mathrm{TAP}}$.

Lemma 4.16 (Proved in Appendix A)**.** *Let $\boldsymbol{m} \in \mathbb{R}^N$, $\boldsymbol{n} \in \mathbb{R}^M$, and*

$$
\acute{\boldsymbol{h}} = \frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} + \varepsilon^{1/2} \hat{\boldsymbol{g}} - \rho_\varepsilon(q(\boldsymbol{m}))\boldsymbol{n}, \qquad d_\varepsilon(\boldsymbol{m},\boldsymbol{n}) = \frac{1}{N} \sum_{a=1}^M (n_a - F_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\acute{h}_a))^2 + F'_{\rho_\varepsilon(q(\boldsymbol{m}))}(\acute{h}_a).
$$

Then,

$$
\nabla_{\boldsymbol{m}} \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) = -\mathrm{th}_\varepsilon^{-1}(\boldsymbol{m}) + \frac{\boldsymbol{G}^\top F_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\acute{\boldsymbol{h}})}{\sqrt{N}} + \varepsilon^{1/2} \dot{\boldsymbol{g}} + \rho'_\varepsilon(q(\boldsymbol{m})) d_\varepsilon(\boldsymbol{m},\boldsymbol{n}) \boldsymbol{m}, \tag{34}
$$

$$
\nabla_{\boldsymbol{n}} \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) = \rho_\varepsilon(q(\boldsymbol{m})) \left(\boldsymbol{n} - F_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\acute{\boldsymbol{h}}) \right). \tag{35}
$$

In particular $\nabla \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) = \mathbf{0}$ if and only if, with $\dot{\boldsymbol{h}} = \mathrm{th}_\varepsilon^{-1}(\boldsymbol{m})$ and $\hat{\boldsymbol{h}} = F^{-1}_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\boldsymbol{n})$,

$$
\frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} + \varepsilon^{1/2} \hat{\boldsymbol{g}} = \hat{\boldsymbol{h}} + \rho_\varepsilon(q(\boldsymbol{m}))\boldsymbol{n}, \tag{36}
$$

$$
\frac{\boldsymbol{G}^\top \boldsymbol{n}}{\sqrt{N}} + \varepsilon^{1/2} \dot{\boldsymbol{g}} = \dot{\boldsymbol{h}} - \rho'_\varepsilon(q(\boldsymbol{m})) d_\varepsilon(\boldsymbol{m},\boldsymbol{n}) \boldsymbol{m}. \tag{37}
$$

(Note that (36) is equivalent to $\widehat{\boldsymbol{h}} = \acute{\boldsymbol{h}}$.)

Lemma 4.17. *Under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}}$, $\boldsymbol{G}$ has law*

$$\frac{\boldsymbol{G}}{\sqrt{N}} \overset{d}{=} \frac{\widehat{\boldsymbol{h}}\boldsymbol{m}^\top}{N(q(\boldsymbol{m})+\varepsilon)} + \frac{\boldsymbol{n}\dot{\boldsymbol{h}}^\top}{N(\psi(\boldsymbol{n})+\varepsilon)} + \frac{\Delta(\boldsymbol{m},\boldsymbol{n})\boldsymbol{n}\boldsymbol{m}^\top}{N(q(\boldsymbol{m})+\psi(\boldsymbol{n})+\varepsilon)} + \frac{\widetilde{\boldsymbol{G}}}{\sqrt{N}}, \qquad \textit{where} \tag{38}$$

$$\Delta(\boldsymbol{m},\boldsymbol{n}) = \rho_\varepsilon(q(\boldsymbol{m})) - \rho_\varepsilon'(q(\boldsymbol{m}))d_\varepsilon(\boldsymbol{m},\boldsymbol{n}) - \frac{\langle \boldsymbol{n}, \widehat{\boldsymbol{h}}\rangle}{N(q(\boldsymbol{m})+\varepsilon)} - \frac{\langle \boldsymbol{m}, \dot{\boldsymbol{h}}\rangle}{N(\psi(\boldsymbol{n})+\varepsilon)} \tag{39}$$

and where $\widetilde{\boldsymbol{G}}$ has the following law. Let $\dot{\boldsymbol{e}}_1,\dots,\dot{\boldsymbol{e}}_N$ and $\widehat{\boldsymbol{e}}_1,\dots,\widehat{\boldsymbol{e}}_M$ be orthonormal bases of $\mathbb{R}^N$ and $\mathbb{R}^M$ with $\dot{\boldsymbol{e}}_1 = \boldsymbol{m}/\|\boldsymbol{m}\|$ and $\widehat{\boldsymbol{e}}_1 = \boldsymbol{n}/\|\boldsymbol{n}\|$, and abbreviate $\widetilde{\boldsymbol{G}}(i,j) = \langle \widehat{\boldsymbol{e}}_j, \widetilde{\boldsymbol{G}}\dot{\boldsymbol{e}}_i\rangle$. Then the $\widetilde{\boldsymbol{G}}(i,j)$ are independent centered gaussians with variance

$$\mathbb{E}\,\widetilde{\boldsymbol{G}}(i,j)^2 = \begin{cases} \varepsilon/(q(\boldsymbol{m})+\psi(\boldsymbol{n})+\varepsilon) & i=j=1, \\ \varepsilon/(q(\boldsymbol{m})+\varepsilon) & i=1, j\neq 1, \\ \varepsilon/(\psi(\boldsymbol{n})+\varepsilon) & i\neq 1, j=1, \\ 1 & i\neq 1, j\neq 1. \end{cases} \tag{40}$$

Proof. This is a standard gaussian conditioning calculation, which we present briefly. For fixed $\dot{\boldsymbol{v}} \in \mathbb{R}^N$, $\widehat{\boldsymbol{v}} \in \mathbb{R}^M$ and

$$\begin{aligned} \widehat{\boldsymbol{w}} &= \frac{\langle \boldsymbol{m}, \dot{\boldsymbol{v}}\rangle}{N(q(\boldsymbol{m})+\varepsilon)}\widehat{\boldsymbol{v}} - \frac{\langle \boldsymbol{m}, \dot{\boldsymbol{v}}\rangle\langle \boldsymbol{n}, \widehat{\boldsymbol{v}}\rangle}{N^2(q(\boldsymbol{m})+\varepsilon)(q(\boldsymbol{m})+\psi(\boldsymbol{n})+\varepsilon)}\boldsymbol{n}, \\ \dot{\boldsymbol{w}} &= \frac{\langle \boldsymbol{n}, \widehat{\boldsymbol{v}}\rangle}{N(\psi(\boldsymbol{n})+\varepsilon)}\dot{\boldsymbol{v}} - \frac{\langle \boldsymbol{m}, \dot{\boldsymbol{v}}\rangle\langle \boldsymbol{n}, \widehat{\boldsymbol{v}}\rangle}{N^2(\psi(\boldsymbol{n})+\varepsilon)(q(\boldsymbol{m})+\psi(\boldsymbol{n})+\varepsilon)}\boldsymbol{m}, \end{aligned}$$

we may verify the independence

$$\frac{\langle \widehat{\boldsymbol{v}}, \boldsymbol{G}\dot{\boldsymbol{v}}\rangle}{\sqrt{N}} - \left\langle \widehat{\boldsymbol{w}}, \frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} + \varepsilon^{1/2}\widehat{\boldsymbol{g}}\right\rangle - \left\langle \dot{\boldsymbol{w}}, \frac{\boldsymbol{G}^\top\boldsymbol{n}}{\sqrt{N}} + \varepsilon^{1/2}\dot{\boldsymbol{g}}\right\rangle \perp\!\!\!\perp \left\{\frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} + \varepsilon^{1/2}\widehat{\boldsymbol{g}}, \frac{\boldsymbol{G}^\top\boldsymbol{n}}{\sqrt{N}} + \varepsilon^{1/2}\dot{\boldsymbol{g}}\right\}.$$

By Lemma 4.16, $\nabla\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) = \boldsymbol{0}$ if and only if (36) and (37) hold. Let $\widehat{\boldsymbol{u}}, \dot{\boldsymbol{u}}$ denote the right-hand sides of (36), (37), respectively. Then, for all $\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}$,

$$\mathbb{E}\left[\frac{\langle \widehat{\boldsymbol{v}}, \boldsymbol{G}\dot{\boldsymbol{v}}\rangle}{\sqrt{N}}\middle|(36),(37)\right] = \langle \widehat{\boldsymbol{w}}, \widehat{\boldsymbol{u}}\rangle + \langle \dot{\boldsymbol{w}}, \dot{\boldsymbol{u}}\rangle.$$

Expanding shows $\boldsymbol{G}$ has the conditional mean given in (38). The law (40) of $\widetilde{\boldsymbol{G}}$ follows from computing the covariance of the gaussian process

$$(\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}) \mapsto \frac{\langle \widehat{\boldsymbol{v}}, \widetilde{\boldsymbol{G}}\dot{\boldsymbol{v}}\rangle}{\sqrt{N}} \equiv \frac{\langle \widehat{\boldsymbol{v}}, \boldsymbol{G}\dot{\boldsymbol{v}}\rangle}{\sqrt{N}} - \left\langle \widehat{\boldsymbol{w}}, \frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} + \varepsilon^{1/2}\widehat{\boldsymbol{g}}\right\rangle - \left\langle \dot{\boldsymbol{w}}, \frac{\boldsymbol{G}^\top\boldsymbol{n}}{\sqrt{N}} + \varepsilon^{1/2}\dot{\boldsymbol{g}}\right\rangle.$$

(Note that if $\widehat{\boldsymbol{v}} \in \{\widehat{\boldsymbol{e}}_2,\dots,\widehat{\boldsymbol{e}}_M\}$, then $\langle \boldsymbol{n}, \widehat{\boldsymbol{v}}\rangle = 0$ and thus $\dot{\boldsymbol{w}} = \boldsymbol{0}$. Similarly if $\dot{\boldsymbol{v}} \in \{\dot{\boldsymbol{e}}_2,\dots,\dot{\boldsymbol{e}}_N\}$, then $\widehat{\boldsymbol{w}} = \boldsymbol{0}$. So in most cases the above formulas simplify considerably.) □

Corollary 4.18. *If $(\boldsymbol{m},\boldsymbol{n}) \in \mathcal{S}_{\varepsilon,\upsilon}$, then under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}}$, $\boldsymbol{G}$ has law*

$$\frac{\boldsymbol{G}}{\sqrt{N}} \overset{d}{=} \frac{(1+o_\upsilon(1))\widehat{\boldsymbol{h}}\boldsymbol{m}^\top}{N(q_\varepsilon+\varepsilon)} + \frac{(1+o_\upsilon(1))\boldsymbol{n}\dot{\boldsymbol{h}}^\top}{N(\psi_\varepsilon+\varepsilon)} + \frac{o_\upsilon(1)\boldsymbol{n}\boldsymbol{m}^\top}{N} + \frac{\widetilde{\boldsymbol{G}}}{\sqrt{N}}, \tag{41}$$

where $o_\upsilon(1)$ denotes a term vanishing as $\upsilon \to 0$ and $\widetilde{\boldsymbol{G}}$ is as in Lemma 4.17.

This corollary is proved by a standard approximation argument, which we record as Fact 4.20 below.

Definition 4.19. A function $f : \mathbb{R} \to \mathbb{R}$ is $(2, L)$-pseudo-Lipschitz if $|f(x) - f(y)| \leqslant L|x-y|(|x|+|y|+1)$.

Fact 4.20 (Proved in Appendix A). *Suppose $\mu, \mu' \in \mathcal{P}_2(\mathbb{R})$ and let $\mu_2 = \mathbb{E}_{x\sim\mu}[x^2]$. If f is $(2, L)$-pseudo-Lipschitz, then*

$$|\mathbb{E}_\mu[f] - \mathbb{E}_{\mu'}[f]| \leqslant 3L\mathbb{W}_2(\mu, \mu')(\mu_2 + \mathbb{W}_2(\mu, \mu') + 1).$$

Proof of Corollary 4.18. Let $\dot{\boldsymbol{h}} = \mathrm{th}_\varepsilon^{-1}(\boldsymbol{m})$, $\widehat{\boldsymbol{h}} = F_{\varepsilon,\varrho_\varepsilon}^{-1}(\boldsymbol{n})$, so $(\dot{\boldsymbol{h}}, \widehat{\boldsymbol{h}}) \in \mathcal{T}_{\varepsilon,\upsilon}$. Recall $\mu_{\dot{\boldsymbol{h}}}$ defined in (26). Since $\dot{h} \mapsto \mathrm{th}_\varepsilon(\dot{h})^2$ is $(2, O(1))$-pseudo-Lipschitz, by Fact 4.20,

$$|q(\boldsymbol{m}) - q_\varepsilon| = \left|\mathbb{E}_{\dot{h}\sim\mu_{\dot{\boldsymbol{h}}}}[\mathrm{th}_\varepsilon(\dot{h})^2] - \mathbb{E}_{\dot{h}\sim\mathcal{N}(0,\psi_\varepsilon+\varepsilon)}[\mathrm{th}_\varepsilon(\dot{h})^2]\right| = o_\upsilon(1).$$

Similarly $\psi(\boldsymbol{n}) = \psi_\varepsilon + o_\upsilon(1)$ and $d_\varepsilon(\boldsymbol{m}, \boldsymbol{n}) = d_\varepsilon + o_\upsilon(1)$. Also, by gaussian integration by parts and Lemma 4.2,

$$\mathbb{E}_{\dot{h}\sim\mathcal{N}(0,\psi_\varepsilon+\varepsilon)}[\dot{h}\mathrm{th}_\varepsilon(\dot{h})] = (\psi_\varepsilon + \varepsilon)\varrho_\varepsilon.$$

Thus

$$\left|\frac{\langle \boldsymbol{m}, \dot{\boldsymbol{h}}\rangle}{N(\psi_\varepsilon + \varepsilon)} - \varrho_\varepsilon\right| = \left|\mathbb{E}_{\dot{h}\sim\mu_{\dot{\boldsymbol{h}}}}[\dot{h}\mathrm{th}_\varepsilon(\dot{h})] - \mathbb{E}_{\dot{h}\sim\mathcal{N}(0,\psi_\varepsilon+\varepsilon)}[\dot{h}\mathrm{th}_\varepsilon(\dot{h})]\right| = o_\upsilon(1).$$

Similarly $\frac{\langle \boldsymbol{n}, \widehat{\boldsymbol{h}}\rangle}{N(q_\varepsilon+\varepsilon)} = d_\varepsilon + o_\upsilon(1)$. Finally, equation (22) and regularity of $\rho_\varepsilon, \rho'_\varepsilon$ (recall (23)) imply

$$\rho_\varepsilon(q(\boldsymbol{m})) = \varrho_\varepsilon + o_\upsilon(1), \qquad \rho'_\varepsilon(q(\boldsymbol{m})) = -1 + o_\upsilon(1).$$

Combining these estimates shows the conditional mean of $\boldsymbol{G}$ in (38) simplifies to the form (41). In particular note that $\Delta(\boldsymbol{m}, \boldsymbol{n}) = o_\upsilon(1)$. □

4.7. **Deferred proofs.** We now prove various results deferred from the above proof.

Lemma 4.21 ([DS18, Lemma 10.1]). *The function $\mathcal{E}$ satisfies the following for all $x \in \mathbb{R}$.*

(a) $0 \leqslant \mathcal{E}(x) \leqslant |x| + 1$.
(b) $\mathcal{E}'(x) = \mathcal{E}(x)(\mathcal{E}(x) - x) \in (0, 1)$.
(c) $\mathcal{E}''(x) \in (0, 1)$.
(d) $\mathcal{E}^{(3)} \in (-1/2, 13)$.

Proof of Lemma 4.2. We calculate

$$\begin{aligned} q_\varepsilon &= \mathbb{E}[\mathrm{th}_\varepsilon((\psi_\varepsilon + \varepsilon)^{1/2}Z)^2] \\ &= \varepsilon^2(\psi_\varepsilon + \varepsilon) + 2\varepsilon\,\mathbb{E}[(\psi_\varepsilon + \varepsilon)^{1/2}Z\mathrm{th}((\psi_\varepsilon + \varepsilon)^{1/2}Z)] + \mathbb{E}[\mathrm{th}^2((\psi_\varepsilon + \varepsilon)^{1/2}Z)^2] \\ &= \varepsilon^2(\psi_\varepsilon + \varepsilon) + 2\varepsilon(\psi_\varepsilon + \varepsilon)\,\mathbb{E}[1 - \mathrm{th}^2((\psi_\varepsilon + \varepsilon)^{1/2}Z)] + \mathbb{E}[\mathrm{th}^2((\psi_\varepsilon + \varepsilon)^{1/2}Z)^2]. \end{aligned}$$

Thus

$$\mathbb{E}[\mathrm{th}^2((\psi_\varepsilon + \varepsilon)^{1/2}Z)^2] = \frac{q_\varepsilon - 2\varepsilon(\psi_\varepsilon + \varepsilon) - \varepsilon^2(\psi_\varepsilon + \varepsilon)}{1 - 2\varepsilon(\psi_\varepsilon + \varepsilon)},$$

and

$$\mathbb{E}[\mathrm{th}'_\varepsilon((\psi_\varepsilon + \varepsilon)^{1/2}Z)] = 1 + \varepsilon - \mathbb{E}[\mathrm{th}^2((\psi_\varepsilon + \varepsilon)^{1/2}Z)] = \frac{1 - q_\varepsilon + \varepsilon - \varepsilon^2(\psi_\varepsilon + \varepsilon)}{1 - 2\varepsilon(\psi_\varepsilon + \varepsilon)} = \varrho_\varepsilon.$$

□

Differentiating (19) and applying Lemma 4.21(b) shows the following fact.

Fact 4.22. *For $\varepsilon, \varrho > 0$ and any $x \in \mathbb{R}$,*

$$-\frac{1+\varepsilon^2}{\varrho+\varepsilon(1+\varepsilon\varrho)} \leqslant F'_{\varepsilon,\varrho}(x) \leqslant -\frac{\varepsilon}{1+\varepsilon\varrho}.$$

Thus

$$1+\varrho F'_{\varepsilon,\varrho}(x) \geqslant \frac{\varepsilon}{\varrho+\varepsilon(1+\varepsilon\varrho)}. \tag{42}$$

For ϱ in any compact set away from 0, $|F'_{\varepsilon,\varrho}|$, $|F''_{\varepsilon,\varrho}|$ and $|F^{(3)}_{\varepsilon,\varrho}|$ are uniformly bounded independently of ε.

Proof of Proposition 4.6. It is clear that $\nabla^2_{\boldsymbol{n},\boldsymbol{n}}\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})$ is diagonal, so it suffices to check $\partial^2_{n_a}\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) \geqslant \eta$ for all $a \in [M]$. We calculate

$$\begin{aligned}\partial^2_{n_a}\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) &= \rho_\varepsilon(q(\boldsymbol{m}))\left(1+\rho_\varepsilon(q(\boldsymbol{m}))F'_{\varepsilon,\varrho}\left(\frac{\langle \boldsymbol{g}^a,\boldsymbol{m}\rangle}{\sqrt{N}}+\varepsilon^{1/2}\widehat{g}_a-\rho_\varepsilon(q(\boldsymbol{m}))n_a\right)\right)\\ &\overset{(42)}{\geqslant} \frac{\varepsilon\rho_\varepsilon(q(\boldsymbol{m}))}{\rho_\varepsilon(q(\boldsymbol{m}))+\varepsilon(1+\varepsilon\rho_\varepsilon(q(\boldsymbol{m})))}.\end{aligned}$$

Since $\rho_\varepsilon \in [C_{\mathrm{bd}}^{-1}, C_{\mathrm{bd}}]$ the result follows. □

Proof of Lemma 4.14. The function $x \mapsto \rho_\varepsilon(q(\boldsymbol{m}))F_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(x)$ is uniformly Lipschitz over $\boldsymbol{m} \in \mathbb{R}^N$, because $\rho_\varepsilon(q(\boldsymbol{m})) \in [C_{\mathrm{bd}}^{-1}, C_{\mathrm{bd}}]$. Note that $\widehat{\boldsymbol{g}}$ appears in (35) through the term $\varepsilon^{1/2}\widehat{\boldsymbol{g}}$ in $\acute{\boldsymbol{h}}$ and is independent of all other terms apeparing in (35). Thus $\varphi_{\nabla_{\boldsymbol{n}}\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})}(\boldsymbol{z})$ is bounded, and continuous for $\boldsymbol{z}$ in an neighborhood of $\mathbf{0}$, uniformly in $\boldsymbol{m},\boldsymbol{n}$. Similarly, $\dot{\boldsymbol{g}}$ appears in (34), (35) only as the term $\varepsilon^{1/2}\dot{\boldsymbol{g}}$ in (34). This implies the conclusion. □

5. Analysis of AMP

In this section, we prove items (a), (b), and (d) of Proposition 4.8. Item (c) will be proved in §6.

5.1. **Scalar recursions.** For $q \in [0, q_\varepsilon]$, $\psi \in [0, \psi_\varepsilon]$, define

$$\begin{aligned}P_{\mathrm{AMP}}(\psi) &= \mathbb{E}[\mathrm{th}_\varepsilon((\psi+\varepsilon)^{1/2}Z+(\psi_\varepsilon-\psi)^{1/2}Z')\mathrm{th}_\varepsilon((\psi+\varepsilon)^{1/2}Z+(\psi_\varepsilon-\psi)^{1/2}Z'')],\\ R_{\mathrm{AMP}}(q) &= \alpha_\star\,\mathbb{E}[F_{\varepsilon,\varrho_\varepsilon}((q+\varepsilon)^{1/2}Z+(q_\varepsilon-q)^{1/2}Z')F_{\varepsilon,\varrho_\varepsilon}((q+\varepsilon)^{1/2}Z+(q_\varepsilon-q)^{1/2}Z'')],\end{aligned}$$

Define the sequences $(\overline{q}_k)_{k\geqslant 0}$ and $(\overline{\psi}_k)_{k\geqslant 1}$ by $\overline{q}_0 = 0$ and the recursion

$$\overline{\psi}_{k+1} = R_{\mathrm{AMP}}(\overline{q}_k), \qquad \overline{q}_k = P_{\mathrm{AMP}}(\overline{\psi}_k).$$

Lemma 5.1. *The sequences $(\overline{q}_k)_{k\geqslant 0}$, $(\overline{\psi}_k)_{k\geqslant 1}$ are increasing, and for small ε, we have $\overline{q}_k \uparrow q_\varepsilon$ and $\overline{\psi}_k \uparrow \psi_\varepsilon$.*

Proof. Let the functions

$$\widetilde{\mathrm{th}}_\varepsilon(x) = \mathrm{th}_\varepsilon((\psi_\varepsilon+\varepsilon)^{1/2}x), \qquad \widetilde{F}_\varepsilon(x) = F_{\varepsilon,\varrho_\varepsilon}((q_\varepsilon+\varepsilon)^{1/2}x)$$

have Hermite expansions

$$\widetilde{\mathrm{th}}_\varepsilon(x) = \sum_{p\geqslant 0} a_p H_p(x), \qquad \widetilde{F}_\varepsilon(x) = \sum_{p\geqslant 0} b_p H_p(x),$$

where $H_p(x)$ is the p-th Hermite polynomial, normalized to $\mathbb{E}\,H_p(Z)^2 = 1$. Then

$$P_{\mathrm{AMP}}(\psi) = \sum_{p\geqslant 0} a_p^2\left(\frac{\psi+\varepsilon}{\psi_\varepsilon+\varepsilon}\right)^p, \qquad R_{\mathrm{AMP}}(q) = \alpha_\star\sum_{p\geqslant 0} b_p^2\left(\frac{q+\varepsilon}{q_\varepsilon+\varepsilon}\right)^p.$$

So, P_{AMP} and R_{AMP} are increasing and convex. Thus $(\overline{q}_k)_{k\geqslant 0}$, $(\overline{\psi}_k)_{k\geqslant 1}$ are increasing, and their limit is the smallest fixed point of $P_{\mathrm{AMP}}\circ R_{\mathrm{AMP}}$. It remains to show this fixed point is $(q_\varepsilon,\psi_\varepsilon)$. By definition of $q_\varepsilon,\psi_\varepsilon$, $(q_\varepsilon,\psi_\varepsilon)$ is a fixed point. Since $P_{\mathrm{AMP}}\circ R_{\mathrm{AMP}}$ is convex, it suffices to show $(P_{\mathrm{AMP}}\circ R_{\mathrm{AMP}})'(q_\varepsilon)<1$. Note that

$$(P_{\mathrm{AMP}}\circ R_{\mathrm{AMP}})'(q_\varepsilon) = P'_{\mathrm{AMP}}(\psi_\varepsilon)R'_{\mathrm{AMP}}(q_\varepsilon).$$

By gaussian integration by parts,

$$\begin{aligned}
P'_{\mathrm{AMP}}(\psi) &= \mathbb{E}[\mathrm{th}'_\varepsilon((\psi+\varepsilon)^{1/2}Z+(\psi_\varepsilon-\psi)^{1/2}Z')\mathrm{th}'_\varepsilon((\psi+\varepsilon)^{1/2}Z+(\psi_\varepsilon-\psi)^{1/2}Z'')],\\
R'_{\mathrm{AMP}}(q) &= \alpha_\star\,\mathbb{E}[F'_{\varepsilon,\varrho_\varepsilon}((q+\varepsilon)^{1/2}Z+(q_\varepsilon-q)^{1/2}Z')F'_{\varepsilon,\varrho_\varepsilon}((q+\varepsilon)^{1/2}Z+(q_\varepsilon-q)^{1/2}Z'')],
\end{aligned}$$

and in particular

$$P'_{\mathrm{AMP}}(\psi_\varepsilon) = \mathbb{E}[\mathrm{th}'_\varepsilon((\psi_\varepsilon+\varepsilon)^{1/2}Z)^2], \qquad R'_{\mathrm{AMP}}(q_\varepsilon) = \alpha_\star\,\mathbb{E}[F'_{\varepsilon,\varrho_\varepsilon}((q_\varepsilon+\varepsilon)^{1/2}Z)^2].$$

In light of Proposition 4.1, a simple continuity argument shows

$$\mathbb{E}[\mathrm{th}'_\varepsilon((\psi_\varepsilon+\varepsilon)^{1/2}Z)^2] \stackrel{\varepsilon\downarrow 0}{\to} \mathbb{E}[\mathrm{th}'(\psi_0^{1/2}Z)^2], \qquad \mathbb{E}[F'_{\varepsilon,\varrho_\varepsilon}((q_\varepsilon+\varepsilon)^{1/2}Z)^2] \stackrel{\varepsilon\downarrow 0}{\to} \mathbb{E}[F'_{1-q_0}(q_0^{1/2}Z)^2].$$

Thus,

$$\begin{aligned}
(P_{\mathrm{AMP}}\circ R_{\mathrm{AMP}})'(q_\varepsilon) &= \alpha_\star\,\mathbb{E}[\mathrm{th}'_\varepsilon((\psi_\varepsilon+\varepsilon)^{1/2}Z)^2]\,\mathbb{E}[F'_{\varepsilon,\varrho_\varepsilon}((q_\varepsilon+\varepsilon)^{1/2}Z)^2]\\
&\stackrel{\varepsilon\downarrow 0}{\to} \alpha_\star\,\mathbb{E}[\mathrm{th}'(\psi_0^{1/2}Z)^2]\,\mathbb{E}[F'_{1-q_0}(q_0^{1/2}Z)^2] \stackrel{Cond.\ 3.3}{<} 1.
\end{aligned}$$

Thus, $(R_{\mathrm{AMP}}\circ P_{\mathrm{AMP}})'(q_\varepsilon)<1$ for sufficiently small ε. Hence $\overline{q}_k\uparrow q_\varepsilon$ and $\overline{\psi}_k\uparrow\psi_\varepsilon$. □

5.2. **State evolution.** The limiting overlap structure of the AMP iterates in the null model follows directly from the state evolution of [Bol14, BM11, JM13, BMN20]. Define the infinite arrays $(\dot\Sigma_{i,j}: i,j\geqslant 1)$ and $(\widehat\Sigma_{i,j}: i,j\geqslant 0)$ by

$$\dot\Sigma_{i,j} = \begin{cases}\psi_\varepsilon & i=j,\\ \overline{\psi}_{i\wedge j} & i\neq j,\end{cases} \qquad \widehat\Sigma_{i,j} = \begin{cases} q_\varepsilon & i=j,\\ \overline{q}_{i\wedge j} & i\neq j.\end{cases}$$

For any $k\geqslant 0$, let $\dot\Sigma_{\leqslant k}\in\mathbb{R}^{k\times k}$ and $\widehat\Sigma^+_{\leqslant k}\in\mathbb{R}^{(k+1)\times(k+1)}$ denote the sub-arrays indexed by $i,j\leqslant k$.

Proposition 5.2. *For any $k\geqslant 0$, as $N\to\infty$ the empirical coordinate distribution of the AMP iterates converges in $\mathbb{W}_2$ in probability under $\mathbb{P}$, to*

$$\frac{1}{N}\sum_{i=1}^N \delta(\dot h_i^1,\dots,\dot h_i^k) \stackrel{\mathbb{W}_2}{\to} \mathcal{N}(0,\dot\Sigma_{\leqslant k}+\varepsilon\mathbf{1}\mathbf{1}^\top), \qquad \frac{1}{M}\sum_{a=1}^M \delta(\widehat h_a^0,\dots,\widehat h_a^k) \stackrel{\mathbb{W}_2}{\to} \mathcal{N}(0,\widehat\Sigma_{\leqslant k}+\varepsilon\mathbf{1}\mathbf{1}^\top). \tag{43}$$

Proof. The state evolution [BMN20, Theorem 1] implies that (in probability)

$$\frac{1}{N}\sum_{i=1}^N \delta(\dot h_i^1,\dots,\dot h_i^k) \stackrel{\mathbb{W}_2}{\to} \mathcal{N}(0,\dot\Sigma^{(0)}_{\leqslant k}+\varepsilon\mathbf{1}\mathbf{1}^\top), \qquad \frac{1}{M}\sum_{a=1}^M \delta(\widehat h_a^0,\dots,\widehat h_a^k) \stackrel{\mathbb{W}_2}{\to} \mathcal{N}(0,\widehat\Sigma^{(0)}_{\leqslant k}+\varepsilon\mathbf{1}\mathbf{1}^\top).$$

holds for arrays $\dot\Sigma^{(0)}$, $\widehat\Sigma^{(0)}$ defined as follows. As initialization, $\widehat\Sigma^{(0)}_{0,i}=\widehat\Sigma^{(0)}_{i,0}=\widehat\Sigma_{0,i}$ for all $i\geqslant 0$. Then, for $(\widehat H_0,\dots,\widehat H_k)\sim\mathcal{N}(0,\widehat\Sigma^{(0)}_{\leqslant k}+\varepsilon\mathbf{1}\mathbf{1}^\top)$ and $0\leqslant i\leqslant k$, define recursively

$$\dot\Sigma^{(0)}_{k+1,i+1}=\dot\Sigma^{(0)}_{i+1,k+1}=\alpha_\star\,\mathbb{E}[F_{\varepsilon,\varrho_\varepsilon}(\widehat H_i)F_{\varepsilon,\varrho_\varepsilon}(\widehat H_k)].$$

For $(\dot H_0,\dots,\dot H_{k+1})\sim\mathcal{N}(0,\dot\Sigma^{(0)}_{\leqslant k+1}+\varepsilon\mathbf{1}\mathbf{1}^\top)$ and $1\leqslant i\leqslant k+1$, let

$$\widehat\Sigma^{(0)}_{k+1,i}=\widehat\Sigma^{(0)}_{i,k+1}=\mathbb{E}[\mathrm{th}_\varepsilon(\dot H_i)\mathrm{th}_\varepsilon(\dot H_{k+1})].$$

It remains to show $\dot\Sigma^{(0)}, \widehat\Sigma^{(0)}$ coincide with $\dot\Sigma, \widehat\Sigma$. Since $\widehat\Sigma_{0,0} = q_\varepsilon$, induction shows the diagonal entries are

$$\dot\Sigma^{(0)}_{k,k} = \psi_\varepsilon = \dot\Sigma_{k,k}, \qquad \widehat\Sigma^{(0)}_{k,k} = q_\varepsilon = \widehat\Sigma_{k,k}.$$

Then, the above recursion gives $\dot\Sigma^{(0)}_{i+1,j+1} = R_{\mathrm{AMP}}(\widehat\Sigma^{(0)}_{i,j})$, $\widehat\Sigma^{(0)}_{i,j} = P_{\mathrm{AMP}}(\dot\Sigma^{(0)}_{i,j})$. By induction, for $i \neq j$,

$$\dot\Sigma^{(0)}_{i,j} = \overline\psi_{i\wedge j} = \dot\Sigma_{i,j}, \qquad \widehat\Sigma^{(0)}_{i,j} = \overline q_{i\wedge j} = \widehat\Sigma_{i,j}.$$

Thus $\dot\Sigma^{(0)} = \dot\Sigma$ and $\widehat\Sigma^{(0)} = \widehat\Sigma$. □

The following proposition characterizes the limiting overlap structure in the planted model. To conserve notation, we will denote the planted solution by $(\boldsymbol{m},\boldsymbol{n})$, rather than $(\boldsymbol{m}',\boldsymbol{n}')$ as in Proposition 4.8.

Proposition 5.3. *Let $(\boldsymbol{m},\boldsymbol{n}) \in \mathcal{S}_{\varepsilon,o_N(1)}$, $\dot{\boldsymbol{h}} = \mathrm{th}_\varepsilon^{-1}(\boldsymbol{m})$, $\widehat{\boldsymbol{h}} = F^{-1}_{\varepsilon,\varrho_\varepsilon}(\boldsymbol{n})$, and $(\boldsymbol{G},\dot{\boldsymbol{g}},\widehat{\boldsymbol{g}}) \sim \mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$. For any $k \geqslant 0$, as $N \to \infty$ the empirical coordinate distribution of $(\dot{\boldsymbol{h}},\widehat{\boldsymbol{h}})$ and the AMP iterates converges in $\mathbb{W}_2$ in probability under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$, to*

$$\frac{1}{N}\sum_{i=1}^N \delta(\dot h^1_i,\dots,\dot h^k_i,\dot h_i) \stackrel{\mathbb{W}_2}{\to} \mathcal{N}(0,\dot\Sigma_{\leqslant k+1} + \varepsilon \mathbf{1}\mathbf{1}^\top), \quad \frac{1}{M}\sum_{a=1}^M \delta(\widehat h^0_a,\dots,\widehat h^k_a,\widehat h_a) \stackrel{\mathbb{W}_2}{\to} \mathcal{N}(0,\widehat\Sigma_{\leqslant k+1} + \varepsilon \mathbf{1}\mathbf{1}^\top).$$

We prove this proposition by introducing an auxiliary AMP iteration. We fix $\boldsymbol{m},\boldsymbol{n},\dot{\boldsymbol{h}},\widehat{\boldsymbol{h}}$ as in Proposition 5.3. Let $\widetilde{\boldsymbol{G}} \in \mathbb{R}^{M\times N}$ be given by (40) and $\widehat{\boldsymbol{G}} \in \mathbb{R}^{M\times N}$ have i.i.d. $\mathcal{N}(0,1)$ entries, and couple these matrices so that a.s.

$$P^\perp_{\boldsymbol{n}}\widetilde{\boldsymbol{G}}P^\perp_{\boldsymbol{m}} = P^\perp_{\boldsymbol{n}}\widehat{\boldsymbol{G}}P^\perp_{\boldsymbol{m}}, \tag{44}$$

and, with $\overline{\boldsymbol{G}}$ denoting this common value, $\widetilde{\boldsymbol{G}} - \overline{\boldsymbol{G}}$ and $\widehat{\boldsymbol{G}} - \overline{\boldsymbol{G}}$ are independent. Further, let $Z \sim \mathcal{N}(0,1)$, $\dot{\boldsymbol{\xi}} \sim \mathcal{N}(0,\boldsymbol{I}_N)$, $\widehat{\boldsymbol{\xi}} \sim \mathcal{N}(0,\boldsymbol{I}_M)$ be coupled to $\widetilde{\boldsymbol{G}}$ such that

$$\widetilde{\boldsymbol{G}} + \boldsymbol{\Delta} = \overline{\boldsymbol{G}} - \sqrt{\frac{\varepsilon}{q(\boldsymbol{m})+\varepsilon}}\cdot\frac{\widehat{\boldsymbol{\xi}}\boldsymbol{m}^\top}{\|\boldsymbol{m}\|} - \sqrt{\frac{\varepsilon}{\psi(\boldsymbol{n})+\varepsilon}}\cdot\frac{\boldsymbol{n}\dot{\boldsymbol{\xi}}^\top}{\|\boldsymbol{n}\|}, \qquad \text{where} \tag{45}$$

$$\boldsymbol{\Delta} = \sqrt{\frac{\varepsilon}{q(\boldsymbol{m})+\varepsilon} + \frac{\varepsilon}{\psi(\boldsymbol{n})+\varepsilon} - \frac{\varepsilon}{q(\boldsymbol{m})+\psi(\boldsymbol{n})+\varepsilon}}\,\frac{\boldsymbol{n}\boldsymbol{m}^\top}{\|\boldsymbol{n}\|\|\boldsymbol{m}\|}Z. \tag{46}$$

(Such a coupling exists by (40).) The auxiliary AMP iteration has initialization $\boldsymbol{n}^{(1),-1} = \boldsymbol{0}$, $\boldsymbol{m}^{(1),0} = q_\varepsilon^{1/2}\mathbf{1}$, and iteration

$$\boldsymbol{m}^{(1),k} = \mathrm{th}_\varepsilon(\dot{\boldsymbol{h}}^{(1),k}), \qquad \boldsymbol{n}^{(1),k} = F_{\varepsilon,\varrho_\varepsilon}(\widehat{\boldsymbol{h}}^{(1),k}),$$

for $\dot{\boldsymbol{h}}^{(1),k}, \widehat{\boldsymbol{h}}^{(1),k}$ as follows. Let $\overline\psi_0 = 0$, and

$$\widehat{\boldsymbol{h}}^{(1),k} = \frac{1}{\sqrt N}\widehat{\boldsymbol{G}}\left(\boldsymbol{m}^{(1),k} - \frac{\overline q_k}{q_\varepsilon}\boldsymbol{m}\right) + \frac{\sqrt\varepsilon(q_\varepsilon - \overline q_k)}{\sqrt{q_\varepsilon(q_\varepsilon+\varepsilon)}}\widehat{\boldsymbol{\xi}} + \frac{\overline q_k + \varepsilon}{q_\varepsilon+\varepsilon}\widehat{\boldsymbol{h}} - \varrho_\varepsilon\left(\boldsymbol{n}^{(1),k-1} - \frac{\overline\psi_k}{\psi_\varepsilon}\boldsymbol{n}\right) \tag{47}$$

$$\dot{\boldsymbol{h}}^{(1),k+1} = \frac{1}{\sqrt N}\widehat{\boldsymbol{G}}^\top\left(\boldsymbol{n}^{(1),k} - \frac{\overline\psi_{k+1}}{\psi_\varepsilon}\boldsymbol{n}\right) + \frac{\sqrt\varepsilon(\psi_\varepsilon - \overline\psi_{k+1})}{\sqrt{\psi_\varepsilon(\psi_\varepsilon+\varepsilon)}}\dot{\boldsymbol{\xi}} + \frac{\overline\psi_{k+1}+\varepsilon}{\psi_\varepsilon+\varepsilon}\dot{\boldsymbol{h}} - d_\varepsilon\left(\boldsymbol{m}^{(1),k} - \frac{\overline q_k}{q_\varepsilon}\boldsymbol{m}\right).$$

Define augmented arrays $(\dot\Sigma^+_{i,j} : i,j \in \{\diamond, \bowtie\} \cup \mathbb{Z}_{\geqslant 1})$ and $(\widehat\Sigma^+_{i,j} : i,j \in \{\diamond, \bowtie\} \cup \mathbb{Z}_{\geqslant 0})$ by

$$\dot\Sigma^+_{i,j} = \begin{cases} \psi_\varepsilon + \varepsilon & i = j \geqslant 1 \text{ or } i = j = \diamond, \\ \overline{\psi}_j + \varepsilon & i > j \geqslant 1, \\ \overline{\psi}_i + \varepsilon & i \geqslant 1, j = \diamond, \\ \frac{\sqrt{\varepsilon}(\psi_\varepsilon - \overline{\psi}_i)}{\sqrt{\psi_\varepsilon(\psi_\varepsilon+\varepsilon)}} & i \geqslant 1, j = \bowtie, \\ 1 & i = j = \bowtie, \\ 0 & i = \diamond, j = \bowtie, \end{cases} \qquad \widehat\Sigma^+_{i,j} = \begin{cases} q_\varepsilon + \varepsilon & i = j \geqslant 0 \text{ or } i = j = \diamond, \\ \overline{q}_j + \varepsilon & i > j \geqslant 0, \\ \overline{q}_i + \varepsilon & i \geqslant 0, j = \diamond, \\ \frac{\sqrt{\varepsilon}(q_\varepsilon - \overline{q}_i)}{\sqrt{q_\varepsilon(q_\varepsilon+\varepsilon)}} & i \geqslant 0, j = \bowtie, \\ 1 & i = j = \bowtie, \\ 0 & i = \diamond, j = \bowtie, \end{cases}$$

with the remaining entries defined by symmetry over the diagonal. Note that on indices (i,j) where $\{i,j\} \cap \{\diamond, \bowtie\} = \emptyset$, these arrays coincide with $\dot\Sigma + \varepsilon \mathbf{1}\mathbf{1}^\top$ and $\widehat\Sigma + \varepsilon \mathbf{1}\mathbf{1}^\top$. Let $\dot\Sigma^+_{\leqslant k} \in \mathbb{R}^{(k+2)\times(k+2)}$ and $\widehat\Sigma^+_{\leqslant k} \in \mathbb{R}^{(k+3)\times(k+3)}$ denote the sub-arrays indexed by $\{\diamond, \bowtie\}$ and $\{1,\dots,k\}$ (resp. $\{0,\dots,k\}$).

Proposition 5.4 (Proved in Appendix A). *For any $k \geqslant 0$, as $N \to \infty$ we have the convergence in $\mathbb{W}_2$ in probability under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$*

$$\frac{1}{N}\sum_{i=1}^N \delta(\dot h_i, \dot\xi_i, \dot h^{(1),1}_i, \dots, \dot h^{(1),k}_i) \xrightarrow{\mathbb{W}_2} \mathcal{N}(0, \dot\Sigma^+_{\leqslant k}), \quad \frac{1}{M}\sum_{a=1}^M \delta(\widehat h_a, \widehat\xi_a, \widehat h^{(1),0}_a, \dots, \widehat h^{(1),k}_a) \xrightarrow{\mathbb{W}_2} \mathcal{N}(0, \widehat\Sigma^+_{\leqslant k}).$$

This is proved by applying state evolution, analogously to Proposition 5.2. We next show that this AMP iteration approximates the original one, in the following sense.

Proposition 5.5 (Proved in Appendix A). *For any $k \geqslant 0$, as $N \to \infty$ we have $\|\widehat{\boldsymbol{h}}^{(1),k} - \widehat{\boldsymbol{h}}^k\|/\sqrt{N} \to 0$ in probability under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$ and if $k \geqslant 1$, $\|\dot{\boldsymbol{h}}^{(1),k} - \dot{\boldsymbol{h}}^k\|/\sqrt{N} \to 0$ in probability under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$.*

Proof of Proposition 5.3. If we identify index $\diamond$ with $k+1$, the array $\{\dot\Sigma^+_{i,j} : i,j \in \{\diamond\} \cup \{1,\dots,k\}\}$ coincides with $\dot\Sigma_{\leqslant k+1} + \varepsilon \mathbf{1}\mathbf{1}^\top$, and similarly $\{\widehat\Sigma^+_{i,j} : i,j \in \{\diamond\} \cup \{0,\dots,k\}\}$ coincides with $\widehat\Sigma_{\leqslant k+1} + \varepsilon \mathbf{1}\mathbf{1}^\top$. By Proposition 5.4,

$$\frac{1}{N}\sum_{i=1}^N \delta(\dot h^{(1),1}_i, \dots, \dot h^{(1),k}_i, \dot h_i) \xrightarrow{\mathbb{W}_2} \mathcal{N}(0, \dot\Sigma^+_{\leqslant k+1} + \varepsilon \mathbf{1}\mathbf{1}^\top),$$
$$\frac{1}{M}\sum_{a=1}^M \delta(\widehat h^{(1),0}_a, \dots, \widehat h^{(1),k}_a, \widehat h_a) \xrightarrow{\mathbb{W}_2} \mathcal{N}(0, \widehat\Sigma^+_{\leqslant k+1} + \varepsilon \mathbf{1}\mathbf{1}^\top)$$

in probability under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$. Proposition 5.5 implies the conclusion. □

5.3. Completion of the proof.

We separately prove Proposition 4.8 under $\mathbb{P}$ and $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$.

Proof of Proposition 4.8(a)(b), under $\mathbb{P}$. By Proposition 5.2, for *any* k,

$$\mu_{\dot{\boldsymbol{h}}^k} \xrightarrow{\mathbb{W}_2} \mathcal{N}(0, \psi_\varepsilon + \varepsilon), \qquad \mu_{\widehat{\boldsymbol{h}}^k} \xrightarrow{\mathbb{W}_2} \mathcal{N}(0, q_\varepsilon + \varepsilon)$$

in probability. So, with high probability, $(\dot{\boldsymbol{h}}^k, \widehat{\boldsymbol{h}}^k) \in \mathcal{T}_{\varepsilon,v_0}$ and thus item (a) holds. Approximation arguments similar to the proof of Corollary 4.18 using Fact 4.20 yield

$$q(\boldsymbol{m}^k) \to \mathbb{E}[\mathrm{th}_\varepsilon((\psi_\varepsilon + \varepsilon)^{1/2} Z)^2] = q_\varepsilon$$

in probability. Regularity of ρ_ε and its derivatives then implies

$$\rho_\varepsilon(q(\boldsymbol{m}^k)) \to \varrho_\varepsilon, \qquad \rho'_\varepsilon(q(\boldsymbol{m}^k)) \to -1$$

in probability. Proposition 5.2 also implies

$$\lim_{k\to\infty} \operatorname*{p-lim}_{N\to\infty} \frac{1}{N}\|\dot{\boldsymbol{h}}^{k+1} - \dot{\boldsymbol{h}}^{k}\|^2 = \lim_{k\to\infty} \operatorname*{p-lim}_{N\to\infty} \frac{1}{N}\|\widehat{\boldsymbol{h}}^{k+1} - \widehat{\boldsymbol{h}}^{k}\|^2 = 0.$$

Below, let $o_{k,P}(\sqrt{N})$ denote a random vector $\boldsymbol{v}$ such that $\lim_{k\to\infty} \text{p-lim}_{N\to\infty} \frac{1}{\sqrt{N}}\|\boldsymbol{v}\| = 0$, and $o_{k,P}(1)$ denote a random scalar ι such that $\lim_{k\to\infty} \text{p-lim}_{N\to\infty} |\iota| = 0$. Let

$$\acute{\boldsymbol{h}}^{k} = \frac{\boldsymbol{G}\boldsymbol{m}^k}{\sqrt{N}} + \varepsilon^{1/2}\widehat{\boldsymbol{g}} - \rho_\varepsilon(q(\boldsymbol{m}^k))\boldsymbol{n}^k.$$

By Lemma 4.2,

$$\widehat{\boldsymbol{h}}^{k} = \frac{\boldsymbol{G}\boldsymbol{m}^k}{\sqrt{N}} + \varepsilon^{1/2}\widehat{\boldsymbol{g}} - \varrho_\varepsilon \boldsymbol{n}^{k-1}.$$

The above discussion implies $\widehat{\boldsymbol{h}}^{k} - \acute{\boldsymbol{h}}^{k} = o_{k,P}(\sqrt{N})$, and thus $\boldsymbol{n}^k - F_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\acute{\boldsymbol{h}}^{k}) = o_{k,P}(\sqrt{N})$. By (35),

$$\nabla_{\boldsymbol{n}}\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m}^k, \boldsymbol{n}^k) = o_{k,P}(\sqrt{N}).$$

Moreover,

$$d_\varepsilon(\boldsymbol{m}^k, \boldsymbol{n}^k) = \frac{1}{N}\sum_{a=1}^{M} F'_{\varepsilon,\varrho_\varepsilon}(\widehat{h}^k) + o_{k,P}(1) = d_\varepsilon + o_{k,P}(1),$$

for d_ε defined below Lemma 4.2. So

$$\nabla_{\boldsymbol{m}}\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m}^k, \boldsymbol{n}^k) = -\mathrm{th}_\varepsilon^{-1}(\boldsymbol{m}^k) + \frac{\boldsymbol{G}^\top \boldsymbol{n}^k}{\sqrt{N}} + \varepsilon^{1/2}\dot{\boldsymbol{g}} - d_\varepsilon \boldsymbol{m}^k + \left(1 + \frac{\|\boldsymbol{G}\|_{\mathrm{op}}}{\sqrt{N}}\right) o_{k,P}(\sqrt{N}).$$

Since $\|\boldsymbol{G}\|_{\mathrm{op}} \leqslant C\sqrt{N}$ w.h.p.,

$$\begin{aligned}\nabla_{\boldsymbol{m}}\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}(\boldsymbol{m}^k, \boldsymbol{n}^k) &= -\dot{\boldsymbol{h}}^{k} + \frac{\boldsymbol{G}^\top \boldsymbol{n}^k}{\sqrt{N}} + \varepsilon^{1/2}\dot{\boldsymbol{g}} - d_\varepsilon \boldsymbol{m}^k + o_{k,P}(\sqrt{N}) \\ &= \dot{\boldsymbol{h}}^{k+1} - \dot{\boldsymbol{h}}^{k} + o_{k,P}(\sqrt{N}) = o_{k,P}(\sqrt{N}),\end{aligned}$$

proving item (b). □

Proof of Proposition 4.8(a)(b)(d), under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$. Suppose first $(\boldsymbol{m}, \boldsymbol{n}) \in \mathcal{S}_{\varepsilon,o_N(1)}$, and let $\dot{\boldsymbol{h}} = \mathrm{th}_\varepsilon^{-1}(\boldsymbol{m})$, $\widehat{\boldsymbol{h}} = F^{-1}_{\varepsilon,\varrho_\varepsilon}(\boldsymbol{n})$. The above argument, using Proposition 5.3 in place of Proposition 5.2, shows items (a) and (b) hold with high probability under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$. Proposition 5.3 also yields

$$\lim_{k\to\infty} \operatorname*{p-lim}_{N\to\infty} \frac{1}{N}\|\dot{\boldsymbol{h}}^{k} - \dot{\boldsymbol{h}}\|^2 = \lim_{k\to\infty} \operatorname*{p-lim}_{N\to\infty} \frac{1}{N}\|\widehat{\boldsymbol{h}}^{k} - \widehat{\boldsymbol{h}}\|^2 = 0.$$

Thus item (d) holds with high probability under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$. Finally, we show this remains true for $(\boldsymbol{m}, \boldsymbol{n}) \in \mathcal{S}_{\varepsilon,\upsilon}$, for suitably small υ. Let $(\overline{\boldsymbol{m}}, \overline{\boldsymbol{n}}) \in \mathcal{S}_{\varepsilon,o_N(1)}$ be such that $\frac{1}{N}\|\boldsymbol{m} - \overline{\boldsymbol{m}}\|^2, \frac{1}{N}\|\boldsymbol{n} - \overline{\boldsymbol{n}}\|^2 = o_\upsilon(1)$. We will show there is a coupling of $(\boldsymbol{G}, \dot{\boldsymbol{g}}, \widehat{\boldsymbol{g}}) \sim \mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$ and $(\overline{\boldsymbol{G}}, \dot{\overline{\boldsymbol{g}}}, \widehat{\overline{\boldsymbol{g}}}) \sim \mathbb{P}^{\overline{\boldsymbol{m}},\overline{\boldsymbol{n}}}_{\varepsilon,\mathsf{Pl}}$ such that

$$\|\boldsymbol{G} - \overline{\boldsymbol{G}}\|_{\mathrm{op}}, \|\dot{\boldsymbol{g}} - \dot{\overline{\boldsymbol{g}}}\|, \|\widehat{\boldsymbol{g}} - \widehat{\overline{\boldsymbol{g}}}\| \leqslant o_\upsilon(1)\sqrt{N}. \tag{48}$$

If $(\boldsymbol{m}^k, \boldsymbol{n}^k)$ are the AMP iterates under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$ and $(\overline{\boldsymbol{m}}^k, \overline{\boldsymbol{n}}^k)$ are the AMP iterates under $\mathbb{P}^{\overline{\boldsymbol{m}},\overline{\boldsymbol{n}}}_{\varepsilon,\mathsf{Pl}}$, this implies $\|\boldsymbol{m}^k - \overline{\boldsymbol{m}}^k\|, \|\boldsymbol{n}^k - \overline{\boldsymbol{n}}^k\| \leqslant o_\upsilon(1)\sqrt{N}$ (this uses crucially that υ is set small depending on k). This implies (a) and (d) continue to hold, and similar approximation arguments to above show (b) continues to hold.

We now prove (48). Let $\dot{\overline{h}} = \mathrm{th}_\varepsilon^{-1}(\overline{m})$ and $\hat{\overline{h}} = F^{-1}_{\varepsilon,\rho_\varepsilon(q(\overline{m}))}(\overline{n})$. Another approximation argument shows $\|\dot{h} - \dot{\overline{h}}\|, \|\hat{h} - \hat{\overline{h}}\| \leqslant o_v(1)\sqrt{N}$. The conditional means of $G, \overline{G}$ are given by (38), and an approximation argument shows

$$\|\mathbb{E}^{m,n}_{\varepsilon,\mathsf{Pl}}[G] - \mathbb{E}^{\overline{m},\overline{n}}_{\varepsilon,\mathsf{Pl}}[\overline{G}]\|_{\mathrm{op}} \leqslant o_v(1)\sqrt{N}.$$

We couple the random parts $\widetilde{G}, \widetilde{\overline{G}}$ as follows. Let $\dot{e}_1, \hat{e}_1$ (resp. $\dot{\overline{h}}_1, \hat{\overline{h}}_1$) be the the unit vectors parallel to m, n (resp. $\overline{m}, \overline{n}$). Let $\dot{T}, \hat{T}$ be rotation operators on $\mathbb{R}^N, \mathbb{R}^M$ with $\dot{T}\dot{e}_1 = \dot{\overline{h}}_1$ and $\hat{T}\hat{e}_1 = \hat{\overline{h}}_1$. These can be set so $\|\dot{T} - I_N\|_{\mathrm{op}}, \|\hat{T} - I_M\|_{\mathrm{op}} \leqslant o_v(1)$. By (40), we can couple $\widetilde{G}, \widetilde{\overline{G}}$ such that $\widetilde{\overline{G}} = \hat{T}\widetilde{G}\dot{T}^{-1}$. Since, for some absolute constant C, $\|\widetilde{G}\|_{\mathrm{op}} \leqslant C\sqrt{N}$ with high probability, on this event

$$\|\widetilde{G} - \widetilde{\overline{G}}\|_{\mathrm{op}} \leqslant \|\widetilde{G}\|_{\mathrm{op}}(\|\dot{T} - I_N\|_{\mathrm{op}} + \|\hat{T} - I_M\|_{\mathrm{op}}) = o_v(1)\sqrt{N}.$$

Thus $\|G - \overline{G}\|_{\mathrm{op}} \leqslant o_v(1)\sqrt{N}$. The stationary equations (36), (37) then imply $\|\dot{g} - \dot{\overline{g}}\|_{\mathrm{op}}, \|\hat{g} - \hat{\overline{g}}\|_{\mathrm{op}} \leqslant o_v(1)\sqrt{N}$. This proves (48). □

6. Local concavity of perturbed TAP free energy

In this section, we prove Lemmas 3.5 and 4.9 and Proposition 4.8(c).

6.1. Description of spectral gap bound.

We first define a quantity λ_ε, which is a perturbed analog of the value $\lambda_0 = \inf_{z>-1}\lambda(z)$ defined in Condition 3.4. We will see that λ_ε upper bounds the maximum eigenvalue of $\nabla^2_\diamond \mathcal{F}^\varepsilon_{\mathrm{TAP}}$ near late AMP iterates. To define λ_ε, we introduce ε-perturbed variants of quantities appearing in Condition 3.4 and Lemma 3.5. Let

$$\dot{f}_\varepsilon(x) = \frac{\mathrm{ch}^2 x}{1 + \varepsilon\mathrm{ch}^2(x)}, \qquad \hat{f}_\varepsilon(x) = -\frac{F'_{\varepsilon,\varrho_\varepsilon}(x)}{1 + \varrho_\varepsilon F'_{\varepsilon,\varrho_\varepsilon}(x)}.$$

We extend these definitions to $\varepsilon = 0$ by defining $\dot{f}_0(x) = \mathrm{ch}^2(x)$ and $\hat{f}_0$ as in Condition 3.4; this extension will be used solely in Lemma 6.1 and the proof of Lemma 3.5 below.

Note that $\dot{f}_\varepsilon$ and $\hat{f}_\varepsilon$ are positive, the latter because Fact 4.22 implies $F'_{\varepsilon,\varrho_\varepsilon}(x) < 0$ and $1 + \varrho_\varepsilon F'_{\varepsilon,\varrho_\varepsilon}(x) > 0$, and $\dot{f}_\varepsilon(x)$ has minimum $\dot{f}_\varepsilon(0) = \frac{1}{1+\varepsilon}$. The function $\hat{f}_0$ is also positive, as Lemma 4.21(b) implies $F'_{1-q_0}(x) < 0$ and $1 + (1-q_0)F'_{1-q_0}(x) > 0$. In the below, it will be convenient to abbreviate $\widetilde{q}_\varepsilon = q_\varepsilon + \varepsilon$, $\widetilde{\psi}_\varepsilon = \psi_\varepsilon + \varepsilon$.

Lemma 6.1. *For any $\varepsilon \geqslant 0$ (including $\varepsilon = 0$), the functions $m_\varepsilon, \theta_\varepsilon : (-\frac{1}{1+\varepsilon}, +\infty) \to (0, +\infty)$ defined by*

$$m_\varepsilon(z) = \mathbb{E}[(z + \dot{f}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-1}],$$

$$\theta_\varepsilon(z) = \mathbb{E}[(z + \dot{f}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-2}]\,\mathbb{E}\left[\left(\frac{\hat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2} Z)}{1 + m_\varepsilon(z)\hat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2} Z)}\right)^2\right]$$

are continuous and strictly decreasing, with

$$\lim_{z\downarrow -(1+\varepsilon)^{-1}} m_\varepsilon(z) = \lim_{z\downarrow -(1+\varepsilon)^{-1}} \theta_\varepsilon(z) = +\infty, \qquad \lim_{z\uparrow +\infty} m_\varepsilon(z) = \lim_{z\uparrow+\infty}\theta_\varepsilon(z) = 0.$$

In particular θ_ε has a well-defined inverse $\theta_\varepsilon^{-1} : (0, +\infty) \to (-\frac{1}{1+\varepsilon}, +\infty)$.

Proof of Lemma 6.1. Note that $m_\varepsilon(z)$ is clearly decreasing on $(-\frac{1}{1+\varepsilon}, +\infty)$ with $\lim_{z\uparrow+\infty} m_\varepsilon(z) = 0$. To show the other limit, let

$$\dot{g}_\varepsilon(x) = \dot{f}_\varepsilon(x) - \frac{1}{1+\varepsilon} = \frac{\mathrm{sh}^2(x)}{(1+\varepsilon)(1+\varepsilon\mathrm{ch}^2(x))}.$$

For $z = -\frac{1}{1+\varepsilon} + \iota$, with $\iota > 0$ small,

$$m_\varepsilon(z) = \mathbb{E}[(\iota + \dot{g}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-1}] \geqslant \mathbb{E}[\mathbf{1}\{|Z| \leqslant \iota^{1/2}\}(\iota + \dot{g}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-1}] \geqslant \Omega(\iota^{-1/2}).$$

Thus $\lim_{z\downarrow -(1+\varepsilon)^{-1}} m_\varepsilon(z) = +\infty$. We can write $\theta_\varepsilon(z)$ as

$$\theta_\varepsilon(z) = \frac{\mathbb{E}[(z + \dot{f}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-2}]}{\mathbb{E}[(z + \dot{f}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-1}]^2} \mathbb{E}\left[\frac{(m_\varepsilon(z)\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2} Z))^2}{(1 + m_\varepsilon(z)\widehat{f}_\varepsilon \widetilde{q}_\varepsilon^{1/2} Z))^2}\right]. \tag{49}$$

Since $m_\varepsilon(z)$ is decreasing and $\widehat{f}_\varepsilon$ is positive, the second factor of (49) is manifestly decreasing. The z-derivative of the first is

$$\frac{-\mathbb{E}[(z + \dot{f}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-1}]\,\mathbb{E}[(z + \dot{f}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-3}] + \mathbb{E}[(z + \dot{f}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-2}]^2}{\mathbb{E}[(z + \dot{f}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-1}]^3} < 0$$

by Cauchy–Schwarz. Thus θ_ε is decreasing on $(-\frac{1}{1+\varepsilon}, +\infty)$. We now calculate its limits as $z \downarrow -\frac{1}{1+\varepsilon}$ and $z \uparrow +\infty$. Consider first $z = -\frac{1}{1+\varepsilon} + \iota$ for ι small. Then the first factor of (49) is

$$\frac{\mathbb{E}[(\iota + \dot{g}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-2}]}{\mathbb{E}[(\iota + \dot{g}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-1}]^2} \geqslant \frac{\mathbb{E}[\mathbf{1}\{|Z| \leqslant \iota^{1/2}\}(\iota + \dot{g}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-2}]}{\mathbb{E}[\mathbf{1}\{|Z| \leqslant \iota^{1/3}\}(\iota + \dot{g}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-1} + O(\iota^{-2/3})]^2} = \frac{\Omega(\iota^{-3/2})}{O(\iota^{-4/3})},$$

which diverges as $\iota \downarrow 0$. The second factor of (49) tends to 1 in this limit by dominated convergence. Thus $\lim_{z\downarrow -(1+\varepsilon)^{-1}} \theta_\varepsilon(z) = +\infty$. We can write the first factor of (49) as

$$\frac{\mathbb{E}[(1 + z^{-1}\dot{f}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-2}]}{\mathbb{E}[(1 + z^{-1}\dot{f}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z))^{-1}]^2},$$

which tends to 1 as $z \uparrow +\infty$ by dominated convergence. In this limit, the second factor of (49) tends to 0 by dominated convergence, so $\lim_{z\uparrow+\infty} \theta_\varepsilon(z) = 0$. This completes the proof. □

Proof of Lemma 3.5. Note that

$$m'(z) = -\mathbb{E}[(z + \mathrm{ch}^2(\psi_0^{1/2} Z))^{-2}].$$

Thus, differentiating λ yields

$$\lambda'(z) = 1 + \alpha_\star m'(z)\,\mathbb{E}\left[\left(\frac{\widehat{f}_0(q_0^{1/2} Z)}{1 + m(z)\widehat{f}_0(q_0^{1/2} Z)}\right)^2\right] = 1 - \alpha_\star \theta(z).$$

The assertions about θ follow from Lemma 6.1, with $\varepsilon = 0$. Since θ is strictly decreasing on $(-1, +\infty)$, λ' is strictly increasing on this interval, and therefore λ is strictly convex on this interval. Since $\theta^{-1} : (0, +\infty) \to (-1, +\infty)$ is well-defined, we may define $z_0 = \theta^{-1}(\alpha_\star^{-1})$. This point satisfies the stationarity condition $\lambda'(z_0) = 0$ and is thus the unique minimizer of λ on $(-1, +\infty)$. □

Recall from below Lemma 4.2 that $d_\varepsilon = \alpha_\star \mathbb{E}[F'_{\varepsilon,\varrho_\varepsilon}(\widetilde{q}_\varepsilon^{1/2} Z)]$. We now define the threshold λ_ε.

Definition 6.2. Let $z_\varepsilon = \theta_\varepsilon^{-1}(\alpha_\star^{-1})$ and

$$\lambda_\varepsilon \equiv z_\varepsilon - \alpha_\star \mathbb{E}\left[\frac{\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2} Z)}{1 + m_\varepsilon(z_\varepsilon)\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2} Z)}\right] - d_\varepsilon. \tag{50}$$

Lemma 6.3. *As $\varepsilon \downarrow 0$, $\lambda_\varepsilon \to \lambda_0$ (defined in Condition 3.4).*

Proof. By Proposition 4.1, as $\varepsilon \downarrow 0$, $(\widetilde{q}_\varepsilon, \widetilde{\psi}_\varepsilon) \to (q_0, \psi_0)$. Thus, for $\dot{f}_0(x) = \mathrm{ch}^2(x)$, the push-forwards $(\dot{f}_\varepsilon)_\# \mathcal{N}(0, \widetilde{\psi}_\varepsilon)$ and $(\widehat{f}_\varepsilon)_\# \mathcal{N}(0, \widetilde{q}_\varepsilon)$ converge weakly to $(\dot{f}_0)_\# \mathcal{N}(0, \psi_0)$ and $(\widehat{f}_0)_\# \mathcal{N}(0, q_0)$.

For any $z > -1$ and small ε, the integrand of $m_\varepsilon(z)$ is bounded independently of ε, and thus $\lim_{\varepsilon\downarrow 0} m_\varepsilon(z) = m(z)$ by dominated convergence. Similarly, all three integrands in (49) are bounded, so $\lim_{\varepsilon\downarrow 0} \theta_\varepsilon(z) = \theta(z)$. Moreover, one easily checks that on any compact subset of $(-1, +\infty)$, the derivatives of $m_\varepsilon, \theta_\varepsilon$ are bounded independently of ε. Thus $m_\varepsilon \to m$, $\theta_\varepsilon \to \theta$ uniformly on compact subsets of $(-1, +\infty)$.

By Lemma 3.5, $\lim_{z\downarrow -1} \theta(z) = +\infty$, so $z_0 = \theta^{-1}(\alpha_\star^{-1})$ is bounded away from -1. The above uniform convergence then implies $z_\varepsilon \to z_0$ and $m_\varepsilon(z_\varepsilon) \to m(z_0)$. Since the below integrands are bounded,

$$\mathbb{E}\left[\frac{\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2} Z)}{1 + m_\varepsilon(z_\varepsilon)\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2} Z)}\right] \to \mathbb{E}\left[\frac{\widehat{f}_0(q_0^{1/2} Z)}{1 + m(z_0)\widehat{f}_0(q_0^{1/2} Z)}\right].$$

Finally, as $F'_{\varepsilon,\varrho_\varepsilon}$ is bounded (by Fact 4.22) and limits to the bounded function F'_{1-q_0}, we have $d_\varepsilon \to d_0$. □

6.2. **Hessian estimate.** We next prove the following upper bound on $\nabla_\diamond^2 \mathcal{F}_{\mathrm{TAP}}^\varepsilon$.

Lemma 6.4. *Suppose $(\boldsymbol{m}, \boldsymbol{n}) \in \mathcal{S}_{\varepsilon, r_0}$, and $\|\boldsymbol{G}\|_{\mathrm{op}}, \|\widehat{\boldsymbol{g}}\| \leqslant C\sqrt{N}$ for some absolute constant C (i.e. independent of all parameters in §4.1). Let $\dot{\boldsymbol{h}} \in \mathbb{R}^N$, $\acute{\boldsymbol{h}} \in \mathbb{R}^M$ be defined (as in Lemma 4.16) by*

$$\dot{\boldsymbol{h}} = \mathrm{th}_\varepsilon^{-1}(\boldsymbol{m}), \qquad \acute{\boldsymbol{h}} = \frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} + \varepsilon^{1/2}\widehat{\boldsymbol{g}} - \rho_\varepsilon(q(\boldsymbol{m}))\boldsymbol{n},$$

and

$$\boldsymbol{D}_1 = \mathrm{diag}(\dot{f}_\varepsilon(\dot{\boldsymbol{h}})), \qquad \boldsymbol{D}_2 = \mathrm{diag}(\widehat{f}_\varepsilon(\acute{\boldsymbol{h}})).$$

Then,

$$\nabla_\diamond^2 \mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m}, \boldsymbol{n}) \preceq P_{\boldsymbol{m}}^\perp \left(-\boldsymbol{D}_1 - \frac{1}{N}\boldsymbol{G}^\top \boldsymbol{D}_2 \boldsymbol{G} - d_\varepsilon \boldsymbol{I}_N\right) P_{\boldsymbol{m}}^\perp + \frac{\lambda_\varepsilon \boldsymbol{m}\boldsymbol{m}^\top}{\|\boldsymbol{m}\|^2} + (o_{C_{\mathrm{cvx}}}(1) + o_{r_0}(1))\boldsymbol{I}_N.$$

(Recall the meaning of $o_{C_{\mathrm{cvx}}}(1)$, $o_{r_0}(1)$ discussed in §4.1.)

Fact 6.5 (Proved in Appendix A). *Let $\boldsymbol{m} \in \mathbb{R}^N$, $\boldsymbol{n} \in \mathbb{R}^M$, and let $\acute{\boldsymbol{h}}, \dot{\boldsymbol{h}}$ be as above. Let $F = F_{\varepsilon, \rho_\varepsilon(q(\boldsymbol{m}))}$ and*

$$\boldsymbol{D}_3 = \mathrm{diag}\left(F'(\acute{\boldsymbol{h}})\right), \qquad \boldsymbol{D}_4 = \boldsymbol{I}_M + \rho_\varepsilon(q(\boldsymbol{m}))\boldsymbol{D}_3.$$

Then,

$$\begin{aligned}
\nabla_{\boldsymbol{m},\boldsymbol{m}}^2 \mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m}, \boldsymbol{n}) &= -\boldsymbol{D}_1 + \frac{\boldsymbol{G}^\top \boldsymbol{D}_3 \boldsymbol{G}}{N} + \rho'_\varepsilon(q(\boldsymbol{m}))d_\varepsilon(\boldsymbol{m}, \boldsymbol{n})\boldsymbol{I}_N \\
&\quad + \rho'_\varepsilon(q(\boldsymbol{m})) \cdot \frac{\boldsymbol{G}^\top(F''(\acute{\boldsymbol{h}}) + 2\boldsymbol{D}_3(F(\acute{\boldsymbol{h}}) - \boldsymbol{n}))\boldsymbol{m}^\top + \boldsymbol{m}(F''(\acute{\boldsymbol{h}}) + 2\boldsymbol{D}_3(F(\acute{\boldsymbol{h}}) - \boldsymbol{n}))^\top \boldsymbol{G}}{N^{3/2}} \\
&\quad + \left\{\rho''_\varepsilon(q(\boldsymbol{m}))d_\varepsilon(\boldsymbol{m}, \boldsymbol{n}) + \frac{\rho'_\varepsilon(q(\boldsymbol{m}))^2}{N}\sum_{a=1}^M \left(2F'(\acute{h}_a)^2 + F^{(3)}(\acute{h}_a)\right)\right\}\frac{\boldsymbol{m}\boldsymbol{m}^\top}{N} \\
\nabla_{\boldsymbol{m},\boldsymbol{n}}^2 \mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m}, \boldsymbol{n}) &= -\frac{\rho_\varepsilon(q(\boldsymbol{m}))}{\sqrt{N}}\boldsymbol{G}^\top \boldsymbol{D}_3 - \rho'_\varepsilon(q(\boldsymbol{m}))\frac{\boldsymbol{m}(\rho_\varepsilon(q(\boldsymbol{m}))F''(\acute{\boldsymbol{h}}) + 2\boldsymbol{D}_4(F(\acute{\boldsymbol{h}}) - \boldsymbol{n}))^\top}{N} \\
\nabla_{\boldsymbol{n},\boldsymbol{n}}^2 \mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m}, \boldsymbol{n}) &= \rho_\varepsilon(q(\boldsymbol{m}))\boldsymbol{D}_4,
\end{aligned}$$

Furthermore, for

$$\widetilde{\boldsymbol{D}}_2 = -\boldsymbol{D}_3 + \rho_\varepsilon(q(\boldsymbol{m}))\boldsymbol{D}_3^2 \boldsymbol{D}_4^{-1} = \mathrm{diag}\left(-\frac{F'(\acute{\boldsymbol{h}})}{1 + \rho_\varepsilon(q(\boldsymbol{m}))F'(\acute{\boldsymbol{h}})}\right),$$

we have

$$\begin{aligned}\nabla^2_\diamond \mathcal{F}^\varepsilon_{\rm TAP}(\boldsymbol{m},\boldsymbol{n}) &= -\boldsymbol{D}_1 - \frac{\boldsymbol{G}^\top \widetilde{\boldsymbol{D}}_2 \boldsymbol{G}}{N} + \rho'_\varepsilon(q(\boldsymbol{m}))d_\varepsilon(\boldsymbol{m},\boldsymbol{n})\boldsymbol{I}_N \\ &+ \rho'_\varepsilon(q(\boldsymbol{m}))\cdot \frac{\boldsymbol{G}^\top \boldsymbol{D}_4^{-1}F''(\acute{\boldsymbol{h}})\boldsymbol{m}^\top + \boldsymbol{m}F''(\acute{\boldsymbol{h}})^\top \boldsymbol{D}_4^{-1}\boldsymbol{G}}{N^{3/2}} \\ &+ \Bigg\{\rho''_\varepsilon(q(\boldsymbol{m}))d_\varepsilon(\boldsymbol{m},\boldsymbol{n}) + \frac{\rho'_\varepsilon(q(\boldsymbol{m}))^2}{N}\sum_{a=1}^M \Bigg(2F'(\acute{h}_a)^2 + F^{(3)}(\acute{h}_a) \\ &\qquad - \frac{(\rho_\varepsilon(q(\boldsymbol{m}))F''(\acute{h}_a) + 2(F(\acute{h}_a)-n_a)(1+\rho_\varepsilon(q(\boldsymbol{m}))F'(\acute{h}_a)))^2}{\rho_\varepsilon(q(\boldsymbol{m}))(1+\rho_\varepsilon(q(\boldsymbol{m}))F'(\acute{h}_a))}\Bigg)\Bigg\}\frac{\boldsymbol{m}\boldsymbol{m}^\top}{N}.\end{aligned}$$

Lemma 6.6 (Proved in Appendix A)**.** *Suppose $(\boldsymbol{m},\boldsymbol{n}) \in \mathcal{S}_{\varepsilon,r_0}$ and $\|\boldsymbol{G}\|_{\rm op}, \|\widehat{\boldsymbol{g}}\| \leqslant C\sqrt{N}$ for an absolute constant C. The following estimates hold for sufficiently small r_0 (depending on $\varepsilon, C_{\rm cvx}, C_{\rm bd}, \eta$).*

(a) Up to additive $o_{r_0}(1)$ error, $q(\boldsymbol{m}) \approx q_\varepsilon$, $\psi(\boldsymbol{n}) \approx \psi_\varepsilon$, and $d_\varepsilon(\boldsymbol{m},\boldsymbol{n}) \approx d_\varepsilon$, $\rho_\varepsilon(q(\boldsymbol{m})) \approx \varrho_\varepsilon$, $\rho'_\varepsilon(q(\boldsymbol{m})) \approx -1$, $\rho''_\varepsilon(q(\boldsymbol{m})) \approx C_{\rm cvx}$.
(b) $\|\widetilde{\boldsymbol{D}}_2 - \boldsymbol{D}_2\|_{\rm op} = o_{r_0}(1)$.
(c) $\frac{1}{N}\sum_{a=1}^M (2F'(\acute{h}_a)^2 + F^{(3)}(\acute{h}_a))$ is bounded by an absolute constant.
(d) $\frac{1}{\sqrt{N}}\|\boldsymbol{D}_4^{-1}F''(\acute{\boldsymbol{h}})\|$ is bounded, with bound depending only on ε.

Proof of Lemma 6.4. By Fact 6.5 and Lemma 6.6,

$$\nabla^2_\diamond \mathcal{F}^\varepsilon_{\rm TAP}(\boldsymbol{m},\boldsymbol{n}) \preceq -\boldsymbol{D}_1 - \frac{\boldsymbol{G}^\top \widetilde{\boldsymbol{D}}_2 \boldsymbol{G}}{N} - d_\varepsilon \boldsymbol{I}_N + \frac{\boldsymbol{G}^\top \boldsymbol{v}_1\boldsymbol{m}^\top + \boldsymbol{m}\boldsymbol{v}_1^\top \boldsymbol{G}}{N} + (C_{\rm cvx}d_\varepsilon + C_1)\frac{\boldsymbol{m}\boldsymbol{m}^\top}{N} + o_{r_0}(1)\boldsymbol{I}_N,$$

for $C_1 \in \mathbb{R}$, $\boldsymbol{v}_1 \in \mathbb{R}^N$ with $|C_1|$, $\|\boldsymbol{v}_1\|$ bounded depending only on ε. By the assumption on $\|\boldsymbol{G}\|_{\rm op}$, $\frac{1}{\sqrt{N}}\|\boldsymbol{G}^\top \boldsymbol{v}_1\|$ is also bounded depending only on ε. Note that

$$-\boldsymbol{D}_1 \preceq -P^\perp_{\boldsymbol{m}}\boldsymbol{D}_1P^\perp_{\boldsymbol{m}} - (P^\perp_{\boldsymbol{m}}\boldsymbol{D}_1P_{\boldsymbol{m}} + P_{\boldsymbol{m}}\boldsymbol{D}_1P^\perp_{\boldsymbol{m}}) = -P^\perp_{\boldsymbol{m}}\boldsymbol{D}_1P^\perp_{\boldsymbol{m}} - \frac{(P^\perp_{\boldsymbol{m}}\boldsymbol{D}_1\boldsymbol{m})\boldsymbol{m}^\top + \boldsymbol{m}(P^\perp_{\boldsymbol{m}}\boldsymbol{D}_1\boldsymbol{m})}{q(\boldsymbol{m})N}$$

and similarly

$$-\frac{1}{N}\boldsymbol{G}^\top \boldsymbol{D}_2\boldsymbol{G} \preceq -P^\perp_{\boldsymbol{m}}\boldsymbol{G}^\top \boldsymbol{D}_2\boldsymbol{G}P^\perp_{\boldsymbol{m}} - \frac{(P^\perp_{\boldsymbol{m}}\boldsymbol{G}^\top \boldsymbol{D}_2\boldsymbol{G}\boldsymbol{m})\boldsymbol{m}^\top + \boldsymbol{m}(P^\perp_{\boldsymbol{m}}\boldsymbol{G}^\top \boldsymbol{D}_2\boldsymbol{G}\boldsymbol{m})^\top}{q(\boldsymbol{m})N^2}.$$

Moreover $\|\boldsymbol{D}_1\|_{\rm op}, \|\boldsymbol{D}_2\|_{\rm op} \leqslant O(\varepsilon^{-1})$, the latter by (42). So, there exists $C_2 \in \mathbb{R}$, $\boldsymbol{v}_2 \in \mathbb{R}^N$ with $|C_2|$, $\|\boldsymbol{v}_2\|$ bounded depending only on ε, such that

$$\nabla^2_\diamond \mathcal{F}^\varepsilon_{\rm TAP}(\boldsymbol{m},\boldsymbol{n}) \preceq P^\perp_{\boldsymbol{m}}\left(-\boldsymbol{D}_1 - \frac{\boldsymbol{G}^\top \widetilde{\boldsymbol{D}}_2 \boldsymbol{G}}{N}\right)P^\perp_{\boldsymbol{m}} - d_\varepsilon \boldsymbol{I}_N + \frac{\boldsymbol{v}_2\boldsymbol{m}^\top + \boldsymbol{m}\boldsymbol{v}_2^\top}{N^{1/2}} + (C_{\rm cvx}d_\varepsilon + C_2)\frac{\boldsymbol{m}\boldsymbol{m}^\top}{N} + o_{r_0}(1)\boldsymbol{I}_N.$$

Note that $d_\varepsilon < 0$, because $F'_{\varepsilon,\varrho_\varepsilon} < 0$ by Fact 4.22. So, for large $C_{\rm cvx}$,

$$(C_{\rm cvx}d_\varepsilon + C_2)\frac{\boldsymbol{m}\boldsymbol{m}^\top}{N} + \frac{\boldsymbol{v}_2\boldsymbol{m}^\top + \boldsymbol{m}\boldsymbol{v}_2^\top}{N^{1/2}} \preceq \frac{(\lambda_\varepsilon + d_\varepsilon)\boldsymbol{m}\boldsymbol{m}^\top}{\|\boldsymbol{m}\|^2} + \frac{\boldsymbol{v}_2\boldsymbol{v}_2^\top}{C_{\rm cvx}|d_\varepsilon| - C_2 + (\lambda_\varepsilon + d_\varepsilon)/q(\boldsymbol{m})}.$$

The final term has operator norm $o_{C_{\rm cvx}}(1)$. □

6.3. **Null model: post-AMP Gordon's inequality.** We turn to the proof of Proposition 4.8(c), first under the measure $\mathbb{P}$. In light of Lemma 6.4, we define

$$\boldsymbol{R}(\boldsymbol{m},\boldsymbol{n}) = P_{\boldsymbol{m}}^{\perp}\left(-\boldsymbol{D}_1 - \frac{1}{N}\boldsymbol{G}^\top \boldsymbol{D}_2 \boldsymbol{G}\right)P_{\boldsymbol{m}}^{\perp}, \tag{51}$$

where, as in that lemma, $\boldsymbol{D}_1 = \mathrm{diag}(\dot{f}_\varepsilon(\dot{\boldsymbol{h}}))$, $\boldsymbol{D}_2 = \mathrm{diag}(\widehat{f}_\varepsilon(\acute{\boldsymbol{h}}(\boldsymbol{m},\boldsymbol{n},\boldsymbol{G})))$ for $\dot{\boldsymbol{h}} = \mathrm{th}_\varepsilon^{-1}(\boldsymbol{m})$ and

$$\acute{\boldsymbol{h}}(\boldsymbol{m},\boldsymbol{n},\boldsymbol{G}) = \frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} + \varepsilon^{1/2}\widehat{\boldsymbol{g}} - \rho_\varepsilon(q(\boldsymbol{m}))\boldsymbol{n}.$$

Proposition 6.7. *With high probability under $\mathbb{P}$, $\boldsymbol{R}(\boldsymbol{m},\boldsymbol{n}) \preceq (\lambda_\varepsilon + d_\varepsilon + o_{r_0}(1) + o_k(1))P_{\boldsymbol{m}}^{\perp}$ for all $\|(\boldsymbol{m},\boldsymbol{n}) - (\boldsymbol{m}^k,\boldsymbol{n}^k)\| \leqslant r_0\sqrt{N}$.*

For z_ε defined in Definition 6.2, let

$$r_\varepsilon^2 = \mathbb{E}[(z_\varepsilon + \dot{f}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2}Z))^{-2}]^{-1}.$$

Define the AMP iterates $\boldsymbol{m}^0,\boldsymbol{n}^0,\dots,\boldsymbol{m}^k,\boldsymbol{n}^k$ and $\widehat{\boldsymbol{h}}^0,\dot{\boldsymbol{h}}^1,\widehat{\boldsymbol{h}}^1,\dots,\dot{\boldsymbol{h}}^k,\widehat{\boldsymbol{h}}^k$ as in (20), (21), and

$$\mathsf{DATA} = (\dot{\boldsymbol{g}},\dot{\boldsymbol{h}}^1,\dots,\dot{\boldsymbol{h}}^k,\widehat{\boldsymbol{g}},\widehat{\boldsymbol{h}}^0,\dots,\widehat{\boldsymbol{h}}^k).$$

Let $U(r_0) = \{(\boldsymbol{m},\boldsymbol{n}) : \|(\boldsymbol{m},\boldsymbol{n}) - (\boldsymbol{m}^k,\boldsymbol{n}^k)\| \leqslant r_0\sqrt{N}\}$. Let $\acute{\boldsymbol{h}}^k \equiv \acute{\boldsymbol{h}}(\boldsymbol{m}^k,\boldsymbol{n}^k,\boldsymbol{G})$, and note that

$$\acute{\boldsymbol{h}}^k = \widehat{\boldsymbol{h}}^k + \varrho_\varepsilon \boldsymbol{n}^{k-1} - \rho_\varepsilon(q(\boldsymbol{m}^k))\boldsymbol{n}^k \tag{52}$$

is DATA-measurable. Let $U'(r_0) = \{\acute{\boldsymbol{h}} : \|\acute{\boldsymbol{h}} - \acute{\boldsymbol{h}}^k\| \leqslant Cr_0\sqrt{N}\}$, for a suitably large absolute constant C. Since $\|\boldsymbol{G}\|_{\mathrm{op}} = O(\sqrt{N})$ with high probability, on this event $\acute{\boldsymbol{h}}(\boldsymbol{m},\boldsymbol{n},\boldsymbol{G}) \in U'(r_0)$ for all $(\boldsymbol{m},\boldsymbol{n}) \in U(r_0)$.

Below, we will write $\boldsymbol{D}_2(\acute{\boldsymbol{h}}) = \mathrm{diag}(\widehat{f}_\varepsilon(\acute{\boldsymbol{h}}))$ for a varying $\acute{\boldsymbol{h}}$ which is not necessarily $\acute{\boldsymbol{h}}(\boldsymbol{m},\boldsymbol{n},\boldsymbol{G})$. On the other hand $\boldsymbol{D}_1$ always refers to the function of $\boldsymbol{m}$ defined above. The starting point of our proof of Proposition 6.7 is to recast the maximum eigenvalue as a minimax program, as follows:

$$\begin{aligned}&\sup_{(\boldsymbol{m},\boldsymbol{n})\in U(r_0)}\sup_{\substack{\|\dot{\boldsymbol{v}}\|=1\\ \dot{\boldsymbol{v}}\perp\boldsymbol{m}}} \dot{\boldsymbol{v}}^\top\left(-\boldsymbol{D}_1 - \frac{1}{N}\boldsymbol{G}^\top\boldsymbol{D}_2(\acute{\boldsymbol{h}}(\boldsymbol{m},\boldsymbol{n},\boldsymbol{G}))\boldsymbol{G}\right)\dot{\boldsymbol{v}}\\ &= \sup_{(\boldsymbol{m},\boldsymbol{n})\in U(r_0)}\sup_{\substack{\|\dot{\boldsymbol{v}}\|=1\\ \dot{\boldsymbol{v}}\perp\boldsymbol{m}}}\inf_{\widehat{\boldsymbol{v}}\in\mathbb{R}^M}\left\{-\langle\boldsymbol{D}_1\dot{\boldsymbol{v}},\dot{\boldsymbol{v}}\rangle + \langle\boldsymbol{D}_2(\acute{\boldsymbol{h}}(\boldsymbol{m},\boldsymbol{n},\boldsymbol{G}))^{-1}\widehat{\boldsymbol{v}},\widehat{\boldsymbol{v}}\rangle + \frac{2}{\sqrt{N}}\langle\boldsymbol{G}\dot{\boldsymbol{v}},\widehat{\boldsymbol{v}}\rangle\right\}.\end{aligned}$$

Here we used that $\boldsymbol{D}_1,\boldsymbol{D}_2$ are positive definite, by positivity of $\dot{f}_\varepsilon$, $\widehat{f}_\varepsilon$. On the high probability event that $\|\boldsymbol{G}\|_{\mathrm{op}} = O(\sqrt{N})$, this is bounded by

$$\sup_{\substack{(\boldsymbol{m},\boldsymbol{n})\in U(r_0)\\ \acute{\boldsymbol{h}}\in U'(r_0)}}\sup_{\substack{\|\dot{\boldsymbol{v}}\|=1\\ \dot{\boldsymbol{v}}\perp\boldsymbol{m}}}\inf_{\substack{\|\widehat{\boldsymbol{v}}\|=r_\varepsilon\\ \widehat{\boldsymbol{v}}\perp\boldsymbol{n}}}\left\{-\langle\boldsymbol{D}_1\dot{\boldsymbol{v}},\dot{\boldsymbol{v}}\rangle + \langle\boldsymbol{D}_2(\acute{\boldsymbol{h}})^{-1}\widehat{\boldsymbol{v}},\widehat{\boldsymbol{v}}\rangle + \frac{2}{\sqrt{N}}\langle\boldsymbol{G}\dot{\boldsymbol{v}},\widehat{\boldsymbol{v}}\rangle\right\}. \tag{53}$$

We will control (53) by applying Gordon's minimax inequality conditional on the AMP iterates; we explain this next. Let

$$\dot{\mu}_{\mathrm{AMP}} = \frac{1}{N}\sum_{i=1}^N \delta(\varepsilon^{1/2}\dot{g},\dot{h}_i^1,\dots,\dot{h}_i^k), \qquad \widehat{\mu}_{\mathrm{AMP}} = \frac{1}{M}\sum_{a=1}^M \delta(\varepsilon^{1/2}\widehat{g},\widehat{h}_a^0,\dots,\widehat{h}_a^k).$$

Further let $(\dot{\Sigma}^+_{i,j})_{i,j\geqslant 0}$ and $(\widehat{\Sigma}^+_{i,j})_{i,j\geqslant -1}$ be augmented versions of $(\dot{\Sigma}_{i,j})_{i,j\geqslant 1}$, $(\widehat{\Sigma}_{i,j})_{i,j\geqslant 0}$ where we add a row and column of zeros, i.e. $\dot{\Sigma}^+_{0,i} = \dot{\Sigma}^+_{i,0} = \widehat{\Sigma}^+_{-1,i} = \widehat{\Sigma}^+_{i,-1} = 0$.

Lemma 6.8. *For any $\upsilon > 0$, with high probability,*

$$\mathbb{W}_2(\dot{\mu}_{\rm AMP}, \mathcal{N}(0, \dot{\Sigma}^+_{\leqslant k} + \varepsilon \mathbf{1}\mathbf{1}^\top)), \mathbb{W}_2(\widehat{\mu}_{\rm AMP}, \mathcal{N}(0, \widehat{\Sigma}^+_{\leqslant k} + \varepsilon \mathbf{1}\mathbf{1}^\top)) \leqslant \upsilon. \tag{54}$$

Proof. Follows from AMP state evolution, identically to Proposition 5.2. □

We now let υ be sufficiently small depending on r_0, k and condition on a realization of DATA such that (54) holds. (Note that (54) is DATA-measurable.) Define $\bar{\boldsymbol{h}}^i = \dot{\boldsymbol{h}}^i - \varepsilon^{1/2}\dot{\boldsymbol{g}}$, $\breve{\boldsymbol{h}}^i = \widehat{\boldsymbol{h}}^i - \varepsilon^{1/2}\widehat{\boldsymbol{g}}$, and

$$\boldsymbol{M}_{(k)} = (\boldsymbol{m}^0, \dots, \boldsymbol{m}^k) \in \mathbb{R}^{N\times(k+1)}, \qquad \boldsymbol{N}_{(k)} = (\boldsymbol{n}^0, \dots, \boldsymbol{n}^{k-1}) \in \mathbb{R}^{M\times k},$$
$$\bar{\boldsymbol{H}}_{(k)} = (\bar{\boldsymbol{h}}^1, \dots, \bar{\boldsymbol{h}}^k) \in \mathbb{R}^{N\times k}, \qquad \breve{\boldsymbol{H}}_{(k)} = (\breve{\boldsymbol{h}}^0, \dots, \breve{\boldsymbol{h}}^k) \in \mathbb{R}^{M\times(k+1)}.$$

Note that on event (54),

$$\frac{1}{N}\boldsymbol{M}_{(k)}^\top \boldsymbol{M}_{(k)} = \widehat{\Sigma}_{\leqslant k} + o_\upsilon(1), \qquad \frac{1}{N}\boldsymbol{N}_{(k)}^\top \boldsymbol{N}_{(k)} = \dot{\Sigma}_{\leqslant k} + o_\upsilon(1), \tag{55}$$
$$\frac{1}{N}\bar{\boldsymbol{H}}_{(k)}^\top \bar{\boldsymbol{H}}_{(k)} = \dot{\Sigma}_{\leqslant k} + o_\upsilon(1), \qquad \frac{1}{M}\breve{\boldsymbol{H}}_{(k)}^\top \breve{\boldsymbol{H}}_{(k)} = \widehat{\Sigma}_{\leqslant k} + o_\upsilon(1), \tag{56}$$

where $o_\upsilon(1)$ denotes an additive error of operator norm $o_\upsilon(1)$. That is, $\{\boldsymbol{n}^0, \dots, \boldsymbol{n}^{k-1}\}$ and $\{\bar{\boldsymbol{h}}^1, \dots, \bar{\boldsymbol{h}}^k\}$ span k-dimensional subspaces of $\mathbb{R}^M$ and $\mathbb{R}^N$, and the linear mapping between them that sends $\boldsymbol{n}^i$ to $\bar{\boldsymbol{h}}^{i+1}$ is an approximate isometry. The same is true, after scaling by a factor $\alpha_\star$, for $\{\boldsymbol{m}^0, \dots, \boldsymbol{m}^k\}$ and $\{\breve{\boldsymbol{h}}^0, \dots, \breve{\boldsymbol{h}}^k\}$. Define the linear maps

$$\dot{\boldsymbol{T}} = \bar{\boldsymbol{H}}_{(k)}(\boldsymbol{N}_{(k)}^\top \boldsymbol{N}_{(k)})^{-1}\boldsymbol{N}_{(k)}^\top, \qquad \hat{\boldsymbol{T}} = \breve{\boldsymbol{H}}_{(k)}(\boldsymbol{M}_{(k)}^\top \boldsymbol{M}_{(k)})^{-1}\boldsymbol{M}_{(k)}^\top.$$

(The inverses are well-defined because the matrices are full-rank, by (55).) That is, $\dot{\boldsymbol{T}}$ (resp. $\hat{\boldsymbol{T}}$) projects onto the span of $\{\boldsymbol{n}^0, \dots, \boldsymbol{n}^{k-1}\}$ (resp. $\{\boldsymbol{m}^0, \dots, \boldsymbol{m}^k\}$) and then applies the linear map that sends $\boldsymbol{n}^i$ to $\dot{\boldsymbol{h}}^{i+1}$ (resp. $\boldsymbol{m}^i$ to $\widehat{\boldsymbol{h}}^i$).

Lemma 6.9 (Post-AMP Gordon's inequality)**.** *Conditional on any realization of* DATA *satisfying event* (54)*, the following holds. Let $\dot{\boldsymbol{\xi}} \sim \mathcal{N}(0, \boldsymbol{I}_N)$, $\widehat{\boldsymbol{\xi}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_M)$, $Z \sim \mathcal{N}(0,1)$ be independent of everything else and*

$$\dot{\boldsymbol{g}}'_{\rm AMP}(\widehat{\boldsymbol{v}}) = \sqrt{N}\dot{\boldsymbol{T}}\widehat{\boldsymbol{v}} + \|P^\perp_{\boldsymbol{N}_{(k)}}\widehat{\boldsymbol{v}}\| P^\perp_{\boldsymbol{M}_{(k)}}\dot{\boldsymbol{\xi}}, \qquad \widehat{\boldsymbol{g}}'_{\rm AMP}(\dot{\boldsymbol{v}}) = \sqrt{N}\hat{\boldsymbol{T}}\dot{\boldsymbol{v}} + \|P^\perp_{\boldsymbol{M}_{(k)}}\dot{\boldsymbol{v}}\| P^\perp_{\boldsymbol{N}_{(k)}}\widehat{\boldsymbol{\xi}}.$$

For any continuous $f : \mathbb{R}^N \times \mathbb{R}^M \times \mathbb{R}^N \times (\mathbb{R}^M)^2 \times \mathbb{R}^{N\times(k+1)} \times \mathbb{R}^{M\times(k+2)} \to \mathbb{R}$,

$$\sup_{\substack{(\boldsymbol{m},\boldsymbol{n})\in U(r_0)\\ \acute{\boldsymbol{h}}\in U'(r_0)}} \sup_{\substack{\|\dot{\boldsymbol{v}}\|=1\\ \dot{\boldsymbol{v}}\perp \boldsymbol{m}}} \inf_{\substack{\|\widehat{\boldsymbol{v}}\|=r_\varepsilon,\\ \widehat{\boldsymbol{v}}\perp \boldsymbol{n}}} \left\{ f(\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}; \boldsymbol{m}, \boldsymbol{n}, \acute{\boldsymbol{h}}, \mathsf{DATA}) + \frac{2}{\sqrt{N}}\langle \boldsymbol{G}\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}\rangle + \frac{2\|P^\perp_{\boldsymbol{N}_{(k)}}\widehat{\boldsymbol{v}}\|\|P^\perp_{\boldsymbol{M}_{(k)}}\dot{\boldsymbol{v}}\|}{\sqrt{N}} Z \right\}$$

is stochastically dominated by

$$\sup_{\substack{(\boldsymbol{m},\boldsymbol{n})\in U(r_0)\\ \acute{\boldsymbol{h}}\in U'(r_0)}} \sup_{\substack{\|\dot{\boldsymbol{v}}\|=1\\ \dot{\boldsymbol{v}}\perp \boldsymbol{m}^k}} \inf_{\substack{\|\widehat{\boldsymbol{v}}\|=r_\varepsilon,\\ \widehat{\boldsymbol{v}}\perp \boldsymbol{n}^k}} \left\{ f(\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}; \boldsymbol{m}, \boldsymbol{n}, \acute{\boldsymbol{h}}, \mathsf{DATA}) + \frac{2}{\sqrt{N}}\langle \dot{\boldsymbol{v}}, \dot{\boldsymbol{g}}'_{\rm AMP}(\widehat{\boldsymbol{v}})\rangle + \frac{2}{\sqrt{N}}\langle \widehat{\boldsymbol{v}}, \widehat{\boldsymbol{g}}'_{\rm AMP}(\dot{\boldsymbol{v}})\rangle \right\} + o_\upsilon(1).$$

Proof. We will first show that conditional on DATA,

$$\frac{1}{\sqrt{N}}\boldsymbol{G} \overset{d}{=} \dot{\boldsymbol{T}}^\top + \hat{\boldsymbol{T}} + o_\upsilon(1) + \frac{P_{\boldsymbol{N}^\perp_{(k)}}\overline{\boldsymbol{G}}P^\perp_{\boldsymbol{M}_{(k)}}}{\sqrt{N}}, \tag{57}$$

where $o_v(1)$ is a deterministic error of operator norm $o_v(1)$ and $\overline{\boldsymbol{G}}$ is an i.i.d. copy of $\boldsymbol{G}$. Conditioning on DATA amounts to conditioning on the linear relations

$$\frac{1}{\sqrt{N}}\boldsymbol{G}\boldsymbol{m}^i = \breve{\boldsymbol{h}}^i + \varrho_\varepsilon \boldsymbol{n}^{i-1}, \qquad \frac{1}{\sqrt{N}}\boldsymbol{G}^\top \boldsymbol{n}^i = \bar{\boldsymbol{h}}^{i+1} + d_\varepsilon \boldsymbol{m}^i \tag{58}$$

for $0 \leqslant i \leqslant k$ and $0 \leqslant i \leqslant k-1$. So, $P^\perp_{\boldsymbol{N}_{(k)}}\boldsymbol{G}P^\perp_{\boldsymbol{M}_{(k)}}$ is independent of DATA and $\boldsymbol{G} - P^\perp_{\boldsymbol{N}_{(k)}}\boldsymbol{G}P^\perp_{\boldsymbol{M}_{(k)}}$ is DATA-measurable. It suffices to show the latter part is $\dot{\boldsymbol{T}}^\top + \hat{\boldsymbol{T}}$, up to $o_v(1)$ additive operator norm error. Recall from (55) that the condition number of $\frac{1}{N}\boldsymbol{M}_{(k)}^\top \boldsymbol{M}_{(k)}$ and $\frac{1}{N}\boldsymbol{N}_{(k)}^\top \boldsymbol{N}_{(k)}$ is bounded depending on k. So it suffices to show

$$\left\|\frac{1}{\sqrt{N}}\boldsymbol{G}\boldsymbol{M}_{(k)} - (\dot{\boldsymbol{T}}^\top + \hat{\boldsymbol{T}})\boldsymbol{M}_{(k)}\right\|_{\mathrm{op}} = o_v(1)\sqrt{N}, \quad \left\|\frac{1}{\sqrt{N}}\boldsymbol{G}^\top\boldsymbol{N}_{(k)} - (\dot{\boldsymbol{T}} + \hat{\boldsymbol{T}}^\top)\boldsymbol{N}_{(k)}\right\|_{\mathrm{op}} = o_v(1)\sqrt{N}. \tag{59}$$

By (58) and the definition of $\dot{\boldsymbol{T}}, \hat{\boldsymbol{T}}$,

$$\frac{1}{\sqrt{N}}\boldsymbol{G}\boldsymbol{M}_{(k)} = \breve{\boldsymbol{H}}_{(k)} + \varrho_\varepsilon[\boldsymbol{0}, \boldsymbol{N}_{(k)}], \qquad \frac{1}{\sqrt{N}}\boldsymbol{G}^\top\boldsymbol{N}_{(k)} = \bar{\boldsymbol{H}}_{(k)} + d_\varepsilon \boldsymbol{M}_{(k-1)},$$
$$\hat{\boldsymbol{T}}\boldsymbol{M}_{(k)} = \breve{\boldsymbol{H}}_{(k)}, \qquad \dot{\boldsymbol{T}}\boldsymbol{N}_{(k)} = \bar{\boldsymbol{H}}_{(k)}.$$

For all $i, j \geqslant 1$, we have by gaussian integration by parts

$$\begin{aligned}
\frac{1}{N}\langle \bar{\boldsymbol{h}}^i, \boldsymbol{m}^j\rangle &= \frac{1}{N}\langle \bar{\boldsymbol{h}}^i, \mathrm{th}_\varepsilon(\bar{\boldsymbol{h}}^j + \varepsilon^{1/2}\dot{\boldsymbol{g}})\rangle \\
&= \mathbb{E}[(\overline{\psi}_{i\wedge j}^{1/2} Z + (\psi_\varepsilon + \varepsilon - \overline{\psi}_{i\wedge j})Z')\mathrm{th}_\varepsilon(\overline{\psi}_{i\wedge j}^{1/2} Z + (\psi_\varepsilon + \varepsilon - \overline{\psi}_{i\wedge j})^{1/2} Z'')] + o_v(1) \\
&= \varrho_\varepsilon \overline{\psi}_{i\wedge j} + o_v(1).
\end{aligned}$$

Moreover $\frac{1}{N}\langle \bar{\boldsymbol{h}}^i, \boldsymbol{m}^0\rangle = o_v(1)$. Thus,

$$\begin{aligned}
\dot{\boldsymbol{T}}^\top \boldsymbol{M}_{(k)} &= \boldsymbol{N}_{(k)}\left(\frac{1}{N}\boldsymbol{N}_{(k)}^\top \boldsymbol{N}_{(k)}\right)^{-1}\left(\frac{1}{N}\bar{\boldsymbol{H}}_{(k)}^\top \boldsymbol{M}_{(k)}\right) \\
&= \boldsymbol{N}_{(k)}\left(\dot{\Sigma}_{\leqslant k} + o_v(1)\right)^{-1}\left([0, \varrho_\varepsilon \dot{\Sigma}_{\leqslant k}] + o_v(1)\right) = \varrho_\varepsilon[\boldsymbol{0}, \boldsymbol{N}_{(k)}] + o_v(1)\sqrt{N},
\end{aligned}$$

where the errors are all in operator norm. A similar calculation shows

$$\hat{\boldsymbol{T}}^\top \boldsymbol{N}_{(k)} = d_\varepsilon \boldsymbol{M}_{(k-1)} + o_v(1)\sqrt{N}.$$

Combining proves (59) and thus (57). So, conditional on DATA,

$$\frac{1}{\sqrt{N}}\langle \boldsymbol{G}\dot{\boldsymbol{v}}, \hat{\boldsymbol{v}}\rangle \overset{d}{=} \langle \dot{\boldsymbol{v}}, \dot{\boldsymbol{T}}\hat{\boldsymbol{v}}\rangle + \langle \hat{\boldsymbol{v}}, \hat{\boldsymbol{T}}\dot{\boldsymbol{v}}\rangle + o_v(1) + \frac{1}{\sqrt{N}}\langle \overline{\boldsymbol{G}}P^\perp_{\boldsymbol{M}_{(k)}}\dot{\boldsymbol{v}}, P^\perp_{\boldsymbol{N}_{(k)}}\hat{\boldsymbol{v}}\rangle$$

By Gordon's inequality applied to $\overline{\boldsymbol{G}}$,

$$\begin{aligned}
\sup_{\substack{(\boldsymbol{m},\boldsymbol{n})\in U(r_0)\\ \acute{\boldsymbol{h}}\in U'(r_0)}} \sup_{\substack{\|\dot{\boldsymbol{v}}\|=1\\ \dot{\boldsymbol{v}}\perp \boldsymbol{m}}} \inf_{\substack{\|\hat{\boldsymbol{v}}\|=r_\varepsilon,\\ \hat{\boldsymbol{v}}\perp \boldsymbol{n}}} \Bigg\{ & f(\dot{\boldsymbol{v}}, \hat{\boldsymbol{v}}; \boldsymbol{m}, \boldsymbol{n}, \acute{\boldsymbol{h}}, \mathrm{DATA}) + 2\langle \dot{\boldsymbol{v}}, \dot{\boldsymbol{T}}\hat{\boldsymbol{v}}\rangle + 2\langle \hat{\boldsymbol{v}}, \hat{\boldsymbol{T}}\dot{\boldsymbol{v}}\rangle \\
& + \frac{2}{\sqrt{N}}\langle \overline{\boldsymbol{G}}P^\perp_{\boldsymbol{M}_{(k)}}\dot{\boldsymbol{v}}, P^\perp_{\boldsymbol{N}_{(k)}}\hat{\boldsymbol{v}}\rangle + \frac{2\|P^\perp_{\boldsymbol{N}_{(k)}}\hat{\boldsymbol{v}}\|\|P^\perp_{\boldsymbol{M}_{(k)}}\dot{\boldsymbol{v}}\|}{\sqrt{N}} Z\Bigg\}
\end{aligned}$$

is stochastically dominated by

$$\sup_{\substack{(\boldsymbol{m},\boldsymbol{n})\in U(r_0)\\ \acute{\boldsymbol{h}}\in U'(r_0)}} \sup_{\substack{\|\dot{\boldsymbol{v}}\|=1\\ \dot{\boldsymbol{v}}\perp \boldsymbol{m}}} \inf_{\substack{\|\widehat{\boldsymbol{v}}\|=r_\varepsilon,\\ \widehat{\boldsymbol{v}}\perp\boldsymbol{n}}} \Bigg\{ f(\dot{\boldsymbol{v}},\widehat{\boldsymbol{v}};\boldsymbol{m},\boldsymbol{n},\acute{\boldsymbol{h}},\mathsf{DATA}) + 2\langle \dot{\boldsymbol{v}}, \dot{\boldsymbol{T}}\widehat{\boldsymbol{v}}\rangle + 2\langle \widehat{\boldsymbol{v}}, \widehat{\boldsymbol{T}}\dot{\boldsymbol{v}}\rangle$$
$$+ \frac{2\|P^{\perp}_{\boldsymbol{N}_{(k)}}\widehat{\boldsymbol{v}}\|}{\sqrt{N}}\langle \dot{\boldsymbol{v}}, P^{\perp}_{\boldsymbol{M}_{(k)}}\dot{\boldsymbol{\xi}}\rangle + \frac{2\|P^{\perp}_{\boldsymbol{M}_{(k)}}\dot{\boldsymbol{v}}\|}{\sqrt{N}}\langle \widehat{\boldsymbol{v}}, P^{\perp}_{\boldsymbol{N}_{(k)}}\widehat{\boldsymbol{\xi}}\rangle\Bigg\}.$$

The quantity inside the sup-inf is precisely $f(\dot{\boldsymbol{v}},\widehat{\boldsymbol{v}},\mathsf{DATA}) + \frac{2}{\sqrt{N}}\langle \dot{\boldsymbol{v}}, \dot{\boldsymbol{g}}'_{\mathrm{AMP}}(\widehat{\boldsymbol{v}})\rangle + \frac{2}{\sqrt{N}}\langle \widehat{\boldsymbol{v}}, \widehat{\boldsymbol{g}}'_{\mathrm{AMP}}(\dot{\boldsymbol{v}})\rangle$. □

Define

$$\dot{\boldsymbol{g}}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}}) = \sqrt{N}\dot{\boldsymbol{T}}\widehat{\boldsymbol{v}} + \|P^{\perp}_{\boldsymbol{N}_{(k)}}\widehat{\boldsymbol{v}}\|\dot{\boldsymbol{\xi}}, \qquad \widehat{\boldsymbol{g}}_{\mathrm{AMP}}(\dot{\boldsymbol{v}}) = \sqrt{N}\widehat{\boldsymbol{T}}\dot{\boldsymbol{v}} + \|P^{\perp}_{\boldsymbol{M}_{(k)}}\dot{\boldsymbol{v}}\|\widehat{\boldsymbol{\xi}}.$$

Note that

$$\frac{1}{\sqrt{N}}\|\dot{\boldsymbol{g}}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}}) - \dot{\boldsymbol{g}}'_{\mathrm{AMP}}(\widehat{\boldsymbol{v}})\| \leqslant \frac{r_\varepsilon}{\sqrt{N}}\|P_{\boldsymbol{M}_{(k)}}\dot{\boldsymbol{\xi}}\|, \qquad \frac{1}{\sqrt{N}}\|\widehat{\boldsymbol{g}}_{\mathrm{AMP}}(\dot{\boldsymbol{v}}) - \widehat{\boldsymbol{g}}'_{\mathrm{AMP}}(\dot{\boldsymbol{v}})\| \leqslant \frac{1}{\sqrt{N}}\|P_{\boldsymbol{N}_{(k)}}\widehat{\boldsymbol{\xi}}\|,$$

are both bounded by υ with high probability, and similarly $|Z|/\sqrt{N} \leqslant \upsilon$ with high probability. Below, let err denote an error term of order $o_{r_0}(1) + o_k(1) + o_\upsilon(1)$. By (53), Lemma 6.9, and these observations, it suffices to show that with high probability,

$$\sup_{\substack{(\boldsymbol{m},\boldsymbol{n})\in U(r_0)\\ \acute{\boldsymbol{h}}\in U'(r_0)}} \sup_{\substack{\|\dot{\boldsymbol{v}}\|=1\\ \dot{\boldsymbol{v}}\perp \boldsymbol{m}}} \inf_{\substack{\|\widehat{\boldsymbol{v}}\|=r_\varepsilon,\\ \widehat{\boldsymbol{v}}\perp\boldsymbol{n}}} \Bigg\{ -\langle \boldsymbol{D}_1\dot{\boldsymbol{v}},\dot{\boldsymbol{v}}\rangle + \langle \boldsymbol{D}_2(\acute{\boldsymbol{h}})^{-1}\widehat{\boldsymbol{v}},\widehat{\boldsymbol{v}}\rangle$$
$$+ \frac{2}{\sqrt{N}}\langle \dot{\boldsymbol{v}}, \dot{\boldsymbol{g}}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}})\rangle + \frac{2}{\sqrt{N}}\langle \widehat{\boldsymbol{v}}, \widehat{\boldsymbol{g}}_{\mathrm{AMP}}(\dot{\boldsymbol{v}})\rangle\Bigg\} \leqslant \lambda_\varepsilon + d_\varepsilon + \mathsf{err}. \tag{60}$$

Lemma 6.10. *Let*

$$\dot{\mu}'_{\mathrm{AMP}} = \frac{1}{N}\sum_{i=1}^{N}\delta(\dot{\xi}_i, \bar{h}^1_i, \ldots, \bar{h}^k_i), \qquad \widehat{\mu}'_{\mathrm{AMP}} = \frac{1}{M}\sum_{a=1}^{M}\delta(\widehat{\xi}_a, \check{h}^0_a, \ldots, \check{h}^k_a).$$

Conditional on a realization of DATA *such that* (54) *holds, with high probability,*

$$\mathbb{W}_2(\dot{\mu}'_{\mathrm{AMP}}, \mathcal{N}(0,1)\times\mathcal{N}(0,\dot{\Sigma}_{\leqslant k})), \mathbb{W}_2(\widehat{\mu}'_{\mathrm{AMP}}, \mathcal{N}(0,1)\times\mathcal{N}(0,\widehat{\Sigma}_{\leqslant k})) \leqslant 2\upsilon. \tag{61}$$

Proof. Under event (54), the $\mathbb{W}_2$-distance of the marginal of $\dot{\mu}'_{\mathrm{AMP}}$ on all but the first coordinate to $\mathcal{N}(0,\dot{\Sigma}_{\leqslant k})$ is deterministically at most υ. Since $\dot{\boldsymbol{\xi}}$ is independent of DATA, it follows that $\mathbb{W}_2(\dot{\mu}'_{\mathrm{AMP}}, \mathcal{N}(0,1)\times\mathcal{N}(0,\dot{\Sigma}_{\leqslant k})) \leqslant 2\upsilon$ with high probability. The estimate for $\widehat{\mu}'_{\mathrm{AMP}}$ is analogous. □

Fact 6.11 (Proved in Appendix A)**.** *Let $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^3)$, and suppose the marginals of μ have fourth moments. Suppose f_1, f_2, f_3 are L-Lipschitz functions, and f_3 is bounded by L. Then there exists $C = C(\mu, L)$ such that*

$$|\mathbb{E}_{(x,y,z)\sim\mu} f_1(x)f_2(y)f_3(z) - \mathbb{E}_{(x',y',z')\sim\mu'} f_1(x')f_2(y')f_3(z')| \leqslant C\max(\mathbb{W}_2(\mu,\mu'), \mathbb{W}_2(\mu,\mu')^2). \tag{62}$$

Lemma 6.12. *Suppose* (61) *holds. Uniformly over $(\boldsymbol{m},\boldsymbol{n})\in U(r_0)$, $\acute{\boldsymbol{h}}\in U'(r_0)$, $\dot{\boldsymbol{v}}\in\{\|\dot{\boldsymbol{v}}\|=1, \dot{\boldsymbol{v}}\perp\boldsymbol{m}\}$,*

$$\mathbb{W}_2\left(\frac{1}{M}\sum_{a=1}^{M}\delta(\widehat{h}^k_a, \acute{h}_a, n_a, \widehat{g}_{\mathrm{AMP}}(\dot{\boldsymbol{v}})_a), (\widetilde{q}_\varepsilon^{1/2}Z, \widetilde{q}_\varepsilon^{1/2}Z, F_{\varepsilon,\varrho_\varepsilon}(\widetilde{q}_\varepsilon^{1/2}Z), Z')\right) \leqslant \mathsf{err}. \tag{63}$$

Similarly, uniformly over $(\boldsymbol{m},\boldsymbol{n})\in U(r_0)$, $\widehat{\boldsymbol{v}}\in\{\|\widehat{\boldsymbol{v}}\|=r_\varepsilon, \widehat{\boldsymbol{v}}\perp\boldsymbol{n}\}$,

$$\mathbb{W}_2\left(\frac{1}{N}\sum_{i=1}^{N}\delta(\dot{h}^k_i, m_i, \dot{g}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}})_i), (\widetilde{\psi}_\varepsilon^{1/2}Z, \mathrm{th}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2}Z), r_\varepsilon Z')\right) \leqslant \mathsf{err}. \tag{64}$$

Proof. We first show that for any $\dot{\boldsymbol{v}}' \in \{\|\dot{\boldsymbol{v}}'\| = 1, \dot{\boldsymbol{v}}' \perp \boldsymbol{m}\}$,

$$\mathbb{W}_2\left(\frac{1}{M}\sum_{a=1}^{M}\delta(\widehat{h}_a^k, \widehat{g}_{\rm AMP}(\dot{\boldsymbol{v}}')_a), (\widetilde{q}_\varepsilon^{1/2}Z, Z')\right) = o_\upsilon(1). \tag{65}$$

Indeed, let $\dot{\boldsymbol{v}}' = \frac{1}{\sqrt{N}}\boldsymbol{M}_{(k)}\dot{\vec{v}} + P^\perp_{\boldsymbol{M}_{(k)}}\dot{\boldsymbol{v}}'$ for some $\dot{\vec{v}} \in \mathbb{R}^{k+1}$, so that $\widehat{\boldsymbol{g}}_{\rm AMP}(\dot{\boldsymbol{v}}') = \breve{\boldsymbol{H}}_{(k)}\dot{\vec{v}} + \|P^\perp_{\boldsymbol{M}_{(k)}}\dot{\boldsymbol{v}}'\|\widehat{\boldsymbol{\xi}}$. By the approximate isometry (55), (56), since $\frac{1}{\sqrt{N}}\boldsymbol{M}_{(k)}\dot{\vec{v}} \perp \boldsymbol{m}^k$, we have $\frac{1}{N}\langle \breve{\boldsymbol{h}}^k, \breve{\boldsymbol{H}}_{(k)}\dot{\vec{v}}\rangle = o_\upsilon(1)$. (Since υ is small depending on k, we may take it much smaller than the condition number of $\widehat{\Sigma}_{\leqslant k}$.) By this isometry,

$$\mathbb{W}_2\left(\frac{1}{M}\sum_{a=1}^{M}\delta(\widehat{h}_a^k, (\breve{\boldsymbol{H}}_{(k)}\dot{\vec{v}})_a), (\widetilde{q}_\varepsilon^{1/2}Z, \|P_{\boldsymbol{M}_{(k)}}\dot{\boldsymbol{v}}'\|Z')\right) = o_\upsilon(1).$$

Then (61) implies (65). Now consider $(\boldsymbol{m}, \boldsymbol{n}) \in U(r_0)$ and let T be a rotation operator mapping $\boldsymbol{m}/\|\boldsymbol{m}\|$ to $\boldsymbol{m}^k/\|\boldsymbol{m}^k\|$. Note that $\|T - I\|_{\rm op} = o_{r_0}(1)$. Consider any $\dot{\boldsymbol{v}} \in \{\|\dot{\boldsymbol{v}}\| = 1, \dot{\boldsymbol{v}} \perp \boldsymbol{m}\}$, and let $\dot{\boldsymbol{v}}' = T\dot{\boldsymbol{v}}$. Then,

$$\|\widehat{\boldsymbol{g}}_{\rm AMP}(\dot{\boldsymbol{v}}') - \widehat{\boldsymbol{g}}_{\rm AMP}(\dot{\boldsymbol{v}})\| \leqslant (\sqrt{N}\|\hat{\boldsymbol{T}}\|_{\rm op} + \|\widehat{\boldsymbol{\xi}}\|)\|\dot{\boldsymbol{v}}' - \dot{\boldsymbol{v}}\| \leqslant \sqrt{N}(\|\hat{\boldsymbol{T}}\|_{\rm op} + O(1))o_{r_0}(1).$$

Note that

$$\|\hat{\boldsymbol{T}}\|_{\rm op} = \sup_{\dot{\vec{v}}\in\mathbb{R}^{k+1}}\frac{\|\hat{\boldsymbol{T}}\boldsymbol{M}_{(k)}\dot{\vec{v}}\|}{\|\boldsymbol{M}_{(k)}\dot{\vec{v}}\|} = \sup_{\dot{\vec{v}}\in\mathbb{R}^{k+1}}\frac{\|\breve{\boldsymbol{H}}\dot{\vec{v}}\|}{\|\boldsymbol{M}_{(k)}\dot{\vec{v}}\|} = \sup_{\dot{\vec{v}}\in\mathbb{R}^{k+1}}\sqrt{\frac{\langle\frac{1}{N}\breve{\boldsymbol{H}}^\top\breve{\boldsymbol{H}}, \dot{\vec{v}}^{\otimes 2}\rangle}{\langle\frac{1}{N}\boldsymbol{M}^\top\boldsymbol{M}, \dot{\vec{v}}^{\otimes 2}\rangle}}$$

is bounded by an absolute constant by (55), (56). Thus $\|\widehat{\boldsymbol{g}}_{\rm AMP}(\dot{\boldsymbol{v}}') - \widehat{\boldsymbol{g}}_{\rm AMP}(\dot{\boldsymbol{v}})\| \leqslant o_{r_0}(1)\sqrt{N}$. By (52) and definition of $U'(r_0)$,

$$\|\widehat{\boldsymbol{h}}^k - \acute{\boldsymbol{h}}\| \leqslant \|\widehat{\boldsymbol{h}}^k - \acute{\boldsymbol{h}}^k\| + \|\acute{\boldsymbol{h}}^k - \acute{\boldsymbol{h}}\| \leqslant (o_k(1) + o_{r_0}(1))\sqrt{N}. \tag{66}$$

Similarly,

$$\|F_{\varepsilon,\varrho_\varepsilon}(\widehat{\boldsymbol{h}}^k) - \boldsymbol{n}\| = \|\boldsymbol{n}^k - \boldsymbol{n}\| \leqslant o_{r_0}(1)\sqrt{N}. \tag{67}$$

Combining these bounds with (65) proves (63). (64) is proved similarly. □

Proposition 6.13. *If* (61) *holds, uniformly over* $(\boldsymbol{m}, \boldsymbol{n}) \in U(r_0)$, $\acute{\boldsymbol{h}} \in U'(r_0)$, $\dot{\boldsymbol{v}} \in \{\|\dot{\boldsymbol{v}}\| = 1, \dot{\boldsymbol{v}} \perp \boldsymbol{m}\}$,

$$\inf_{\substack{\|\widehat{\boldsymbol{v}}\| = r_\varepsilon, \\ \widehat{\boldsymbol{v}} \perp \boldsymbol{n}}}\langle \boldsymbol{D}_2(\acute{\boldsymbol{h}})^{-1}\widehat{\boldsymbol{v}}, \widehat{\boldsymbol{v}}\rangle + \frac{2}{\sqrt{N}}\langle\widehat{\boldsymbol{v}}, \widehat{\boldsymbol{g}}_{\rm AMP}(\dot{\boldsymbol{v}})\rangle \leqslant -\alpha_\star\mathbb{E}\left[\frac{\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}{1 + m_\varepsilon(z_\varepsilon)\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}\right] - m_\varepsilon(z_\varepsilon)r_\varepsilon^2 + \mathsf{err}.$$

Proof. Let

$$\widehat{\boldsymbol{v}}' = -\frac{1}{\sqrt{N}}\left(\boldsymbol{D}_2(\acute{\boldsymbol{h}})^{-1} + m_\varepsilon(z_\varepsilon)I\right)^{-1}\widehat{\boldsymbol{g}}_{\rm AMP}(\dot{\boldsymbol{v}}).$$

Note the identity

$$\alpha_\star\mathbb{E}\left[\left(\frac{\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}{1 + m_\varepsilon(z_\varepsilon)\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}\right)^2\right] = \frac{\alpha_\star\theta_\varepsilon(z_\varepsilon)}{\mathbb{E}[(z_\varepsilon + \dot{f}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2}Z))^{-2}]} = r_\varepsilon^2. \tag{68}$$

Then,

$$\begin{aligned}\|\widehat{\boldsymbol{v}}'\|^2 &= \frac{1}{N}\widehat{\boldsymbol{g}}_{\mathrm{AMP}}(\dot{\boldsymbol{v}})^\top \left(\widetilde{\boldsymbol{D}}_2(\acute{\boldsymbol{h}})^{-1} + m_\varepsilon(z_\varepsilon)I\right)^{-2} \widehat{\boldsymbol{g}}_{\mathrm{AMP}}(\dot{\boldsymbol{v}}) \\ &= \frac{\alpha_\star}{M}\sum_{a=1}^M \left(\frac{\widehat{f}_\varepsilon(\acute{h}_a)}{1+m_\varepsilon(z_\varepsilon)\widehat{f}_\varepsilon(\acute{h}_a)}\right)^2 \widehat{g}_{\mathrm{AMP}}(\dot{\boldsymbol{v}})_a^2 \\ &= \alpha_\star \mathbb{E}\left[\left(\frac{\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}{1+m_\varepsilon(z_\varepsilon)\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}\right)^2 (Z')^2\right] + \mathsf{err} = r_\varepsilon^2 + \mathsf{err}.\end{aligned}$$

In the last line we used Lemma 6.12 and Fact 6.11, with $f_1(x) = f_2(x) = x$, $f_3(x) = (\frac{\widehat{f}_\varepsilon(x)}{1+m_\varepsilon(z_\varepsilon)\widehat{f}_\varepsilon(x)})^2$. (Note that we have not shown the coordinate empirical measure in (63) has bounded fourth moments, but it suffices for Fact 6.11 that the gaussian approximating it does.) Similarly,

$$\begin{aligned}\frac{1}{\sqrt{N}}\langle \widehat{\boldsymbol{v}}', \boldsymbol{n}\rangle &= -\frac{\alpha_\star}{M}\sum_{a=1}^M \left(\frac{\widehat{f}_\varepsilon(\acute{h}_a)}{1+m_\varepsilon(z_\varepsilon)\widehat{f}_\varepsilon(\acute{h}_a)}\right) n_a \widehat{\boldsymbol{g}}_{\mathrm{AMP}}(\dot{\boldsymbol{v}})_a \\ &= -\alpha_\star \mathbb{E}\left[\left(\frac{\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}{1+m_\varepsilon(z_\varepsilon)\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}\right) F_{\varepsilon,\varrho_\varepsilon}(\widetilde{q}_\varepsilon^{1/2}Z)Z'\right] + \mathsf{err} = \mathsf{err}.\end{aligned}$$

Likewise,

$$\begin{aligned}\langle (\boldsymbol{D}_2(\acute{\boldsymbol{h}})^{-1} + m_\varepsilon(z_\varepsilon)\boldsymbol{I}_M)\widehat{\boldsymbol{v}}', \widehat{\boldsymbol{v}}'\rangle = -\frac{1}{\sqrt{N}}\langle \widehat{\boldsymbol{v}}', \widehat{\boldsymbol{g}}_{\mathrm{AMP}}(\dot{\boldsymbol{v}})\rangle &= \frac{\alpha_\star}{M}\sum_{a=1}^M \left(\frac{\widehat{f}_\varepsilon(\acute{h}_a)}{1+m_\varepsilon(z_\varepsilon)\widehat{f}_\varepsilon(\acute{h}_a)}\right) \widehat{\boldsymbol{g}}_{\mathrm{AMP}}(\dot{\boldsymbol{v}})_a^2 \\ &= \alpha_\star \mathbb{E}\left[\frac{\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}{1+m_\varepsilon(z_\varepsilon)\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}\right] + \mathsf{err}.\end{aligned}$$

From this, it follows that

$$\langle \boldsymbol{D}_2(\acute{\boldsymbol{h}})^{-1}\widehat{\boldsymbol{v}}', \widehat{\boldsymbol{v}}'\rangle + \frac{2}{\sqrt{N}}\langle \widehat{\boldsymbol{v}}', \widehat{\boldsymbol{g}}_{\mathrm{AMP}}(\dot{\boldsymbol{v}})\rangle = -\alpha_\star \mathbb{E}\left[\frac{\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}{1+m_\varepsilon(z_\varepsilon)\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}\right] - m_\varepsilon(z_\varepsilon)r_\varepsilon^2 + \mathsf{err}.$$

By the above estimates on $\|\widehat{\boldsymbol{v}}'\|^2$ and $\frac{1}{\sqrt{N}}\langle \widehat{\boldsymbol{v}}', \boldsymbol{n}\rangle$, we can find $\widehat{\boldsymbol{v}}$ such that $\|\widehat{\boldsymbol{v}}\| = r_\varepsilon$, $\widehat{\boldsymbol{v}} \perp \boldsymbol{n}$, and $\|\widehat{\boldsymbol{v}} - \widehat{\boldsymbol{v}}'\| \leqslant \mathsf{err}$. Since $\boldsymbol{D}_2(\acute{\boldsymbol{h}})^{-1}$ has operator norm bounded independently of r_0, k, v,

$$|\langle \boldsymbol{D}_2(\acute{\boldsymbol{h}})^{-1}\widehat{\boldsymbol{v}}, \widehat{\boldsymbol{v}}\rangle - \langle \boldsymbol{D}_2^{-1}\widehat{\boldsymbol{v}}', \widehat{\boldsymbol{v}}'\rangle| \leqslant 2\|\boldsymbol{D}_2^{-1}(\acute{\boldsymbol{h}})\|_{\mathrm{op}}\|\widehat{\boldsymbol{v}} - \widehat{\boldsymbol{v}}'\| \leqslant \mathsf{err}.$$

By Cauchy–Schwarz,

$$\frac{2}{\sqrt{N}}|\langle \widehat{\boldsymbol{v}}, \widehat{\boldsymbol{g}}_{\mathrm{AMP}}(\dot{\boldsymbol{v}})\rangle - \langle \widehat{\boldsymbol{v}}', \widehat{\boldsymbol{g}}_{\mathrm{AMP}}(\dot{\boldsymbol{v}})\rangle| \leqslant \frac{2}{\sqrt{N}}\|\widehat{\boldsymbol{g}}_{\mathrm{AMP}}(\dot{\boldsymbol{v}})\|\|\widehat{\boldsymbol{v}} - \widehat{\boldsymbol{v}}'\| \leqslant \mathsf{err}.$$

This completes the proof. □

Proposition 6.14. *If* (61) *holds, uniformly over* $(\boldsymbol{m}, \boldsymbol{n}) \in U(r_0)$, $\widehat{\boldsymbol{v}} \in \{\|\widehat{\boldsymbol{v}}\| = r_\varepsilon, \widehat{\boldsymbol{v}} \perp \boldsymbol{n}\}$, *we have*

$$\sup_{\substack{\|\dot{\boldsymbol{v}}\|=1 \\ \dot{\boldsymbol{v}}\perp \boldsymbol{m}}} -\langle \boldsymbol{D}_1\dot{\boldsymbol{v}}, \dot{\boldsymbol{v}}\rangle + \frac{2}{\sqrt{N}}\langle \dot{\boldsymbol{v}}, \dot{\boldsymbol{g}}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}})\rangle \leqslant z_\varepsilon + m_\varepsilon(z_\varepsilon)r_\varepsilon^2 + \mathsf{err}.$$

Proof. Fix any $(\boldsymbol{m}, \boldsymbol{n})$ and $\widehat{\boldsymbol{v}}$ satisfying the stated conditions. We estimate

$$\sup_{\substack{\|\dot{\boldsymbol{v}}\|=1 \\ \dot{\boldsymbol{v}}\perp \boldsymbol{m}}} -\langle \boldsymbol{D}_1\dot{\boldsymbol{v}}, \dot{\boldsymbol{v}}\rangle + \frac{2}{\sqrt{N}}\langle \dot{\boldsymbol{v}}, \dot{\boldsymbol{g}}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}})\rangle \leqslant \sup_{\dot{\boldsymbol{v}}\perp \boldsymbol{m}} -\langle \boldsymbol{D}_1\dot{\boldsymbol{v}}, \dot{\boldsymbol{v}}\rangle + \frac{2}{\sqrt{N}}\langle \dot{\boldsymbol{v}}, \dot{\boldsymbol{g}}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}})\rangle - z_\varepsilon\left(\|\dot{\boldsymbol{v}}\|^2 - 1\right). \quad (69)$$

Note that $-\boldsymbol{D}_1 - z_\varepsilon \boldsymbol{I}_N$ is negative definite, as $z_\varepsilon > -\frac{1}{1+\varepsilon} = \max_{x\in\mathbb{R}}\{-\dot{f}(x)\}$. So, the supremum on the right-hand side of (69) is maximized by $\dot{\boldsymbol{v}}$ solving the stationarity condition (in $\mathrm{span}(\boldsymbol{m})^\perp$):

$$\dot{\boldsymbol{v}} = \frac{1}{\sqrt{N}} P_{\boldsymbol{m}}^\perp (\boldsymbol{D}_1 + z_\varepsilon \boldsymbol{I}_N)^{-1} P_{\boldsymbol{m}}^\perp \dot{\boldsymbol{g}}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}}).$$

Let

$$\dot{\boldsymbol{v}}' = \frac{1}{\sqrt{N}} (\boldsymbol{D}_1 + z_\varepsilon \boldsymbol{I}_N)^{-1} \dot{\boldsymbol{g}}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}}).$$

Note that, by Fact 6.11 and Lemma 6.12,

$$\begin{aligned}
\langle (\boldsymbol{D}_1 + z_\varepsilon \boldsymbol{I}_N)\dot{\boldsymbol{v}}', \dot{\boldsymbol{v}}' \rangle = \frac{1}{\sqrt{N}} \langle \dot{\boldsymbol{v}}', \dot{\boldsymbol{g}}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}}) \rangle &= \frac{1}{N}\sum_{i=1}^N \dot{g}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}})_i^2 (\dot{f}_\varepsilon(\dot{h}_i) + z_\varepsilon)^{-1} \\
&= r_\varepsilon^2 \mathbb{E}\left[(\dot{f}_\varepsilon(\widetilde{\psi}_\varepsilon Z) + z_\varepsilon)^{-1} \right] + \mathsf{err} \\
&= m_\varepsilon(z_\varepsilon) r_\varepsilon^2 + \mathsf{err}.
\end{aligned}$$

Thus

$$-\langle \boldsymbol{D}_1 \dot{\boldsymbol{v}}', \dot{\boldsymbol{v}}' \rangle + \frac{2}{\sqrt{N}} \langle \dot{\boldsymbol{v}}', \dot{\boldsymbol{g}}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}}) \rangle - z_\varepsilon \left(\|\dot{\boldsymbol{v}}'\|^2 - 1 \right) = z_\varepsilon + m_\varepsilon(z_\varepsilon) r_\varepsilon^2 + \mathsf{err}.$$

We now estimate $\|\dot{\boldsymbol{v}} - \dot{\boldsymbol{v}}'\|$. Note that

$$\|\dot{\boldsymbol{v}} - \dot{\boldsymbol{v}}'\| \leqslant \|(\boldsymbol{D}_1 + z_\varepsilon \boldsymbol{I}_N)^{-1}\|_{\mathrm{op}} \|P_{\boldsymbol{m}} \dot{\boldsymbol{g}}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}})\| + \|P_{\boldsymbol{m}} (\boldsymbol{D}_1 + z_\varepsilon \boldsymbol{I}_N)^{-1} \dot{\boldsymbol{g}}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}})\|,$$

and by Fact 6.11 and Lemma 6.12, both terms on the right-hand side are bounded by err. Since $\boldsymbol{D}_1 + z_\varepsilon \boldsymbol{I}_N$ has bounded operator norm,

$$|\langle (\boldsymbol{D}_1 + z_\varepsilon \boldsymbol{I}_N)\dot{\boldsymbol{v}}, \dot{\boldsymbol{v}} \rangle - \langle (\boldsymbol{D}_1 + z_\varepsilon \boldsymbol{I}_N)\dot{\boldsymbol{v}}', \dot{\boldsymbol{v}}' \rangle| \leqslant 2\|\boldsymbol{D}_1 + z_\varepsilon \boldsymbol{I}_N\|_{\mathrm{op}} \|\dot{\boldsymbol{v}} - \dot{\boldsymbol{v}}'\| \leqslant \mathsf{err}.$$

By Cauchy–Schwarz,

$$\frac{2}{\sqrt{N}} |\langle \dot{\boldsymbol{v}}', \dot{\boldsymbol{g}}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}}) \rangle - \langle \dot{\boldsymbol{v}}, \dot{\boldsymbol{g}}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}}) \rangle| \leqslant \frac{2}{\sqrt{N}} \|\dot{\boldsymbol{g}}_{\mathrm{AMP}}(\widehat{\boldsymbol{v}})\| \|\dot{\boldsymbol{v}} - \dot{\boldsymbol{v}}'\| \leqslant \mathsf{err}.$$

Combining completes the proof. □

Proof of Proposition 6.7. By Propositions 6.13 and 6.14, on the high probability event (61), the left-hand side of (60) is bounded by

$$z_\varepsilon - \alpha_\star \mathbb{E}\left[\frac{\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2} Z)}{1 + m_\varepsilon(z_\varepsilon)\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2} Z)} \right] + \mathsf{err} = \lambda_\varepsilon + d_\varepsilon + \mathsf{err}.$$

This proves (60), and by the discussion leading to (60) the proposition follows. □

Proof of Proposition 4.8(c), under $\mathbb{P}$. By Proposition 4.8(a), with high probability, $(\boldsymbol{m}^k, \boldsymbol{n}^k) \in \mathcal{S}_{\varepsilon, \upsilon_0}$. Recall that $\mathrm{th}_\varepsilon, F_{\varepsilon, \varrho_\varepsilon}$ are $O(1)$-Lipschitz, with $O_\varepsilon(1)$-Lipschitz inverses (i.e. Lipschitz constant depending only on ε). On this event, for υ_0 small depending on r_0 and some $C_\varepsilon = O_\varepsilon(1)$,

$$U(r_0) \subseteq \mathcal{S}_{\varepsilon, \upsilon_0 + C_\varepsilon r_0} \subseteq \mathcal{S}_{\varepsilon, 2C_\varepsilon r_0}. \tag{70}$$

Since $\|\boldsymbol{G}\|_{\mathrm{op}}, \|\widehat{\boldsymbol{g}}\| \leqslant C\sqrt{N}$ holds with high probability under $\mathbb{P}$, Lemma 6.4 applies. Applying this lemma with $2C_\varepsilon r_0$ in place of r_0 shows that for all $(\boldsymbol{m}, \boldsymbol{n}) \in U(r_0)$,

$$\nabla_\diamond^2 \mathcal{F}_{\mathrm{TAP}}^\varepsilon(\boldsymbol{m}, \boldsymbol{n}) \preceq \boldsymbol{R}(\boldsymbol{m}, \boldsymbol{n}) + \lambda_\varepsilon P_{\boldsymbol{m}} + (o_{C_{\mathrm{cvx}}}(1) + o_{r_0}(1))\boldsymbol{I}_N.$$

Combined with Proposition 6.7, this gives that with high probability,

$$\nabla_\diamond^2 \mathcal{F}(\boldsymbol{m}, \boldsymbol{n}) \preceq (\lambda_\varepsilon + o_{C_{\mathrm{cvx}}}(1) + o_{r_0}(1) + o_k(1))\boldsymbol{I}_N.$$

By Lemma 6.3,

$$\nabla^2_\diamond \mathcal{F}(\boldsymbol{m},\boldsymbol{n}) \preceq (\lambda_0 + o_\varepsilon(1) + o_{C_{\mathsf{cvx}}}(1) + o_{r_0}(1) + o_k(1))\boldsymbol{I}_N.$$

Under Condition 3.4, $\lambda_0 < 0$. The conclusion follows by setting the parameters so the error term in the last display is bounded by $|\lambda_0|/2$. □

Remark 6.15. The bound $\lambda_\varepsilon + d_\varepsilon$ in Proposition 6.7 is tight. One way to see this is to calculate the upper edge of the limiting spectral measure of

$$\boldsymbol{A} = P^\perp_{\boldsymbol{M}_{(k)}}(-\boldsymbol{D}_1 - \boldsymbol{W})P^\perp_{\boldsymbol{M}_{(k)}}, \qquad \text{where} \qquad \boldsymbol{W} = \frac{1}{N}\boldsymbol{G}^\top P^\perp_{\boldsymbol{N}^{(k)}}\boldsymbol{D}_2 P^\perp_{\boldsymbol{N}^{(k)}}\boldsymbol{G},$$

using free probability [Voi91]. We now outline this calculation. Note that conditional on DATA, $-\boldsymbol{D}_1$ and $-\boldsymbol{W}$ are orthogonally invariant as quadratic forms on $\mathrm{span}(\boldsymbol{m}^0,\ldots,\boldsymbol{m}^k)^\perp$. The inverse Cauchy transform of $-\boldsymbol{D}_1$ is approximated within err by $m_\varepsilon^{-1}(t)$. By e.g. [BS98, Equation 1.2], the inverse Cauchy transform of $-\boldsymbol{W}$ is approximated within err by

$$\frac{1}{t} - \alpha_\star \mathbb{E}\left[\frac{\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}{1 + t\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}\right],$$

Since R-transforms add under free additive convolution, $\boldsymbol{A}$ has limiting inverse Cauchy transform

$$\vartheta_\varepsilon(t) = m_\varepsilon^{-1}(t) - \alpha_\star \mathbb{E}\left[\frac{\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}{1 + t\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}\right].$$

One calculates that

$$\vartheta_\varepsilon'(t) = -\mathbb{E}[(m_\varepsilon^{-1}(t) + \dot{f}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2}Z))^{-2}]^{-1} + \mathbb{E}\left[\left(\frac{\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}{1 + t\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2}Z)}\right)^2\right]$$

has the same sign as $\theta_\varepsilon(m_\varepsilon^{-1}(t)) - \alpha_\star^{-1}$. Thus $\vartheta_\varepsilon(t)$ is decreasing on $(0, m_\varepsilon(z_\varepsilon)]$ and increasing $[m_\varepsilon(z_\varepsilon), +\infty)$. It follows that the limiting spectral measure of $\boldsymbol{A}$ has upper edge $\vartheta_\varepsilon(m_\varepsilon(z_\varepsilon)) = \lambda_\varepsilon + d_\varepsilon$. By the Weyl inequalities the same is true for $\boldsymbol{R}(\boldsymbol{m},\boldsymbol{n})$, so Proposition 6.7 is tight.

6.4. **Planted model.** The proof of Proposition 4.8(c) in the planted model is only simpler, as we will be able to apply Gordon's inequality directly rather than conditional on AMP iterates. The main step is the following proposition. Let υ be sufficiently small depending on r_0, k.

Proposition 6.16. *Suppose $(\boldsymbol{m}',\boldsymbol{n}') \in \mathcal{S}_{\varepsilon,\upsilon}$. With high probability under $\mathbb{P}^{\boldsymbol{m}',\boldsymbol{n}'}_{\varepsilon,\mathsf{Pl}}$, $\boldsymbol{R}(\boldsymbol{m},\boldsymbol{n}) \preceq (\lambda_\varepsilon + d_\varepsilon + \mathrm{err})P^\perp_{\boldsymbol{m}}$ for all $\|(\boldsymbol{m},\boldsymbol{n}) - (\boldsymbol{m}',\boldsymbol{n}')\| \leqslant 2r_0\sqrt{N}$.*

Let $\dot{\boldsymbol{h}}' = \mathrm{th}_\varepsilon^{-1}(\boldsymbol{m}')$, $\widehat{\boldsymbol{h}}' = F^{-1}_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\boldsymbol{n}')$. By Lemma 4.16, under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}'}_{\varepsilon,\mathsf{Pl}}$ we have $\acute{\boldsymbol{h}}(\boldsymbol{m}',\boldsymbol{n}',\boldsymbol{G}) = \widehat{\boldsymbol{h}}'$.

For this subsection, let $U(r_0) = \{(\boldsymbol{m},\boldsymbol{n}) : \|(\boldsymbol{m},\boldsymbol{n}) - (\boldsymbol{m}',\boldsymbol{n}')\| \leqslant 2r_0\sqrt{N}\}$ and $U'(r_0) = \{\acute{\boldsymbol{h}} : \|\acute{\boldsymbol{h}} - \widehat{\boldsymbol{h}}'\| \leqslant Cr_0\sqrt{N}\}$, for suitably large constant C. Identically to the discussion above (53), to prove Proposition 6.16 it suffices to show, with high probability,

$$\sup_{(\boldsymbol{m},\boldsymbol{n})\in U(r_0)} \sup_{\substack{\|\dot{\boldsymbol{v}}\|=1 \\ \dot{\boldsymbol{v}}\perp\boldsymbol{m}}} \inf_{\substack{\|\widehat{\boldsymbol{v}}\|=r_\varepsilon, \\ \widehat{\boldsymbol{v}}\perp\boldsymbol{n}}} \left\{-\langle\boldsymbol{D}_1\dot{\boldsymbol{v}},\dot{\boldsymbol{v}}\rangle + \langle\boldsymbol{D}_2(\acute{\boldsymbol{h}})^{-1}\widehat{\boldsymbol{v}},\widehat{\boldsymbol{v}}\rangle + \frac{2}{\sqrt{N}}\langle\boldsymbol{G}\dot{\boldsymbol{v}},\widehat{\boldsymbol{v}}\rangle\right\} \leqslant \lambda_\varepsilon + d_\varepsilon + \mathrm{err}.$$

Lemma 6.17. *Let* $\dot{\boldsymbol{\xi}}, \dot{\boldsymbol{\xi}}' \sim \mathcal{N}(0, \boldsymbol{I}_N)$, $\widehat{\boldsymbol{\xi}}, \widehat{\boldsymbol{\xi}}' \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_M)$, $Z, Z' \sim \mathcal{N}(0,1)$ *be independent of everything else and*

$$\dot{\boldsymbol{g}}'_{\mathsf{PI}}(\widehat{\boldsymbol{v}}) = \frac{\|P_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\|(\dot{\boldsymbol{h}}' + \varepsilon^{1/2}P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{\xi}}')}{\widetilde{\psi}_\varepsilon^{1/2}} + \|P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\|P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{\xi}}, \quad \widehat{\boldsymbol{g}}'_{\mathsf{PI}}(\dot{\boldsymbol{v}}) = \frac{\|P_{\boldsymbol{m}'}\dot{\boldsymbol{v}}\|(\widehat{\boldsymbol{h}}' + \varepsilon^{1/2}P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{\xi}}')}{\widetilde{q}_\varepsilon^{1/2}} + \|P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{v}}\|P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{\xi}}.$$

For any continuous $f : \mathbb{R}^N \times \mathbb{R}^M \times (\mathbb{R}^N)^2 \times (\mathbb{R}^M)^3 \to \mathbb{R}$,

$$\sup_{\substack{(\boldsymbol{m},\boldsymbol{n})\in U(r_0)\\ \acute{\boldsymbol{h}}\in U'(r_0)}} \sup_{\substack{\|\dot{\boldsymbol{v}}\|=1\\ \dot{\boldsymbol{v}}\perp\boldsymbol{m}}} \inf_{\substack{\|\widehat{\boldsymbol{v}}\|=r_\varepsilon,\\ \widehat{\boldsymbol{v}}\perp\boldsymbol{n}}} \left\{ f(\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}; \boldsymbol{m}', \boldsymbol{m}, \boldsymbol{n}', \boldsymbol{n}, \acute{\boldsymbol{h}}) + \frac{2}{\sqrt{N}}\langle \boldsymbol{G}\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}\rangle + \frac{2\|P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\|\|P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{v}}\|}{\sqrt{N}}Z \right\}$$

is stochastically dominated by

$$\sup_{\substack{(\boldsymbol{m},\boldsymbol{n})\in U(r_0)\\ \acute{\boldsymbol{h}}\in U'(r_0)}} \sup_{\substack{\|\dot{\boldsymbol{v}}\|=1\\ \dot{\boldsymbol{v}}\perp\boldsymbol{m}}} \inf_{\substack{\|\widehat{\boldsymbol{v}}\|=r_\varepsilon,\\ \widehat{\boldsymbol{v}}\perp\boldsymbol{n}}} \left\{ f(\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}; \boldsymbol{m}', \boldsymbol{m}, \boldsymbol{n}', \boldsymbol{n}, \acute{\boldsymbol{h}}) + \frac{2}{\sqrt{N}}\langle \dot{\boldsymbol{v}}, \dot{\boldsymbol{g}}'_{\mathsf{PI}}(\widehat{\boldsymbol{v}})\rangle + \frac{2}{\sqrt{N}}\langle \widehat{\boldsymbol{v}}, \widehat{\boldsymbol{g}}'_{\mathsf{PI}}(\dot{\boldsymbol{v}})\rangle \right.$$
$$\left. + \frac{2\varepsilon^{1/2}\|P_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\|\|P_{\boldsymbol{m}'}\dot{\boldsymbol{v}}\|}{(q_\varepsilon + \psi_\varepsilon + \varepsilon)^{1/2}\sqrt{N}}Z' \right\} + o_v(1).$$

Proof. By Corollary 4.18, the gaussian process $(\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}) \mapsto \frac{1}{\sqrt{N}}\langle \boldsymbol{G}\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}\rangle$ has the form

$$\begin{aligned} \frac{1}{\sqrt{N}}\langle \boldsymbol{G}\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}\rangle &\overset{d}{=} \frac{\langle \dot{\boldsymbol{h}}', \dot{\boldsymbol{v}}\rangle\langle \boldsymbol{n}', \widehat{\boldsymbol{v}}\rangle}{N\widetilde{\psi}_\varepsilon} + \frac{\langle \boldsymbol{m}', \dot{\boldsymbol{v}}\rangle\langle \widehat{\boldsymbol{h}}', \widehat{\boldsymbol{v}}\rangle}{N\widetilde{q}_\varepsilon} + o_v(1) + \frac{1}{\sqrt{N}}\langle \widetilde{\boldsymbol{G}}\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}\rangle \\ &= \frac{\|P_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\|\langle \dot{\boldsymbol{h}}', \dot{\boldsymbol{v}}\rangle}{\widetilde{\psi}_\varepsilon^{1/2}\sqrt{N}} + \frac{\|P_{\boldsymbol{m}'}\dot{\boldsymbol{v}}\|\langle \widehat{\boldsymbol{h}}', \widehat{\boldsymbol{v}}\rangle}{\widetilde{q}_\varepsilon^{1/2}\sqrt{N}} + o_v(1) + \frac{1}{\sqrt{N}}\langle \widetilde{\boldsymbol{G}}\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}\rangle. \end{aligned}$$

Here the $o_v(1)$ is uniform over bounded $\|\dot{\boldsymbol{v}}\|, \|\widehat{\boldsymbol{v}}\|$. Moreover, by (40), the random part $\langle \widetilde{\boldsymbol{G}}\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}\rangle$ expands as

$$\begin{aligned} \langle \widetilde{\boldsymbol{G}}\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}\rangle &= \langle \widetilde{\boldsymbol{G}}P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{v}}, P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\rangle + \langle \widetilde{\boldsymbol{G}}P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{v}}, P_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\rangle + \langle \widetilde{\boldsymbol{G}}P_{\boldsymbol{m}'}\dot{\boldsymbol{v}}, P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\rangle + \langle \widetilde{\boldsymbol{G}}P_{\boldsymbol{m}'}\dot{\boldsymbol{v}}, P_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\rangle \\ &\overset{d}{=} \langle \widetilde{\boldsymbol{G}}P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{v}}, P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\rangle + \frac{\varepsilon^{1/2}}{\widetilde{\psi}_\varepsilon^{1/2}}\|P_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\|\langle P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{\xi}}', \dot{\boldsymbol{v}}\rangle + \frac{\varepsilon^{1/2}}{\widetilde{q}_\varepsilon^{1/2}}\|P_{\boldsymbol{m}'}\dot{\boldsymbol{v}}\|\langle P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{\xi}}', \widehat{\boldsymbol{v}}\rangle + \frac{\varepsilon^{1/2}\|P_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\|\|P_{\boldsymbol{m}'}\dot{\boldsymbol{v}}\|}{(q_\varepsilon + \psi_\varepsilon + \varepsilon)^{1/2}}Z'. \end{aligned}$$

Thus, (as processes)

$$\begin{aligned} \frac{1}{\sqrt{N}}\langle \boldsymbol{G}\dot{\boldsymbol{v}}, \widehat{\boldsymbol{v}}\rangle + \frac{\|P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\|\|P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{v}}\|}{\sqrt{N}}Z &\overset{d}{=} \frac{1}{\sqrt{N}}\langle \widetilde{\boldsymbol{G}}P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{v}}, P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\rangle + \frac{\|P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\|\|P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{v}}\|}{\sqrt{N}}Z \\ &\quad + \frac{\|P_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\|\langle \dot{\boldsymbol{h}}' + \varepsilon^{1/2}P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{\xi}}', \dot{\boldsymbol{v}}\rangle}{\widetilde{\psi}_\varepsilon^{1/2}\sqrt{N}} + \frac{\|P_{\boldsymbol{m}'}\dot{\boldsymbol{v}}\|\langle \widehat{\boldsymbol{h}}' + \varepsilon^{1/2}P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{\xi}}', \widehat{\boldsymbol{v}}\rangle}{\widetilde{q}_\varepsilon^{1/2}\sqrt{N}} \\ &\quad + \frac{\varepsilon^{1/2}\|P_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\|\|P_{\boldsymbol{m}'}\dot{\boldsymbol{v}}\|}{(q_\varepsilon + \psi_\varepsilon + \varepsilon)^{1/2}\sqrt{N}}Z' + o_v(1). \end{aligned}$$

The result now follows by using Gordon's inequality to compare $\frac{1}{\sqrt{N}}\langle \widetilde{\boldsymbol{G}}P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{v}}, P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\rangle + \frac{\|P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\|\|P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{v}}\|}{\sqrt{N}}Z$ to $\frac{1}{\sqrt{N}}\|P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\|\langle \dot{\boldsymbol{v}}, P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{\xi}}\rangle + \frac{1}{\sqrt{N}}\|P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{v}}\|\langle \widehat{\boldsymbol{v}}, P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{\xi}}\rangle$. □

Let

$$\dot{\boldsymbol{g}}_{\mathsf{PI}}(\widehat{\boldsymbol{v}}) = \frac{\|P_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\|(\dot{\boldsymbol{h}}' + \varepsilon^{1/2}\dot{\boldsymbol{\xi}}')}{\widetilde{\psi}_\varepsilon^{1/2}} + \|P^{\perp}_{\boldsymbol{n}'}\widehat{\boldsymbol{v}}\|\dot{\boldsymbol{\xi}}, \qquad \widehat{\boldsymbol{g}}_{\mathsf{PI}}(\dot{\boldsymbol{v}}) = \frac{\|P_{\boldsymbol{m}'}\dot{\boldsymbol{v}}\|(\widehat{\boldsymbol{h}}' + \varepsilon^{1/2}\widehat{\boldsymbol{\xi}}')}{\widetilde{q}_\varepsilon^{1/2}} + \|P^{\perp}_{\boldsymbol{m}'}\dot{\boldsymbol{v}}\|\widehat{\boldsymbol{\xi}}.$$

As argued above (60), with high probability,

$$\frac{1}{\sqrt{N}}|Z|, \frac{1}{\sqrt{N}}|Z'|, \frac{1}{\sqrt{N}} \sup_{\|\widehat{v}\|=r_\varepsilon} \|\dot{g}_{\mathsf{Pl}}(\widehat{v}) - \dot{g}'_{\mathsf{Pl}}(\widehat{v})\|, \frac{1}{\sqrt{N}} \sup_{\|\dot{v}\|=1} \|\widehat{g}_{\mathsf{Pl}}(\dot{v}) - \widehat{g}'_{\mathsf{Pl}}(\dot{v})\| \leqslant \upsilon.$$

So it suffices to show that with high probability,

$$\sup_{\substack{(\boldsymbol{m},\boldsymbol{n})\in U(r_0) \\ \acute{\boldsymbol{h}}\in U'(r_0)}} \sup_{\substack{\|\dot{\boldsymbol{v}}\|=1 \\ \dot{\boldsymbol{v}}\perp \boldsymbol{m}}} \inf_{\substack{\|\widehat{\boldsymbol{v}}\|=r_\varepsilon, \\ \widehat{\boldsymbol{v}}\perp\boldsymbol{n}}} \Bigg\{ - \langle \boldsymbol{D}_1 \dot{\boldsymbol{v}}, \dot{\boldsymbol{v}}\rangle + \langle \boldsymbol{D}_2(\acute{\boldsymbol{h}})^{-1}\widehat{\boldsymbol{v}}, \widehat{\boldsymbol{v}}\rangle + \frac{2}{\sqrt{N}}\langle \dot{\boldsymbol{v}}, \dot{\boldsymbol{g}}_{\mathsf{Pl}}(\widehat{\boldsymbol{v}})\rangle + \frac{2}{\sqrt{N}}\langle \widehat{\boldsymbol{v}}, \widehat{\boldsymbol{g}}_{\mathsf{Pl}}(\dot{\boldsymbol{v}})\rangle \Bigg\} \leqslant \lambda_\varepsilon + d_\varepsilon + \mathsf{err}. \tag{71}$$

Lemma 6.18. *For all $(\boldsymbol{m}', \boldsymbol{n}') \in \mathcal{S}_{\varepsilon,\upsilon}$, the following holds with high probability. Uniformly over $(\boldsymbol{m},\boldsymbol{n}) \in U(r_0)$, $\acute{\boldsymbol{h}} \in U'(r_0)$, $\dot{\boldsymbol{v}} \in \{\|\dot{\boldsymbol{v}}\| = 1, \dot{\boldsymbol{v}} \perp \boldsymbol{m}\}$,*

$$\mathbb{W}_2\left(\frac{1}{M}\sum_{a=1}^{M} \delta(\widehat{h}'_a, \acute{h}_a, n'_a, \widehat{g}_{\mathsf{Pl}}(\dot{\boldsymbol{v}})_a), (\widetilde{q}_\varepsilon^{1/2} Z, \widetilde{q}_\varepsilon^{1/2} Z, F_{\varepsilon,\varrho_\varepsilon}(\widetilde{q}_\varepsilon^{1/2} Z), Z')\right) \leqslant \mathsf{err}. \tag{72}$$

Similarly, uniformly over $(\boldsymbol{m}', \boldsymbol{n}') \in \mathcal{S}_{\varepsilon,\upsilon}$, $(\boldsymbol{m},\boldsymbol{n}) \in U(r_0)$, $\widehat{\boldsymbol{v}} \in \{\|\widehat{\boldsymbol{v}}\| = r_\varepsilon, \widehat{\boldsymbol{v}} \perp \boldsymbol{n}\}$,

$$\mathbb{W}_2\left(\frac{1}{N}\sum_{i=1}^{N} \delta(\dot{h}'_i, m'_i, \dot{g}_{\mathsf{Pl}}(\widehat{\boldsymbol{v}})_i), (\widetilde{\psi}_\varepsilon^{1/2} Z, \mathrm{th}_\varepsilon(\widetilde{\psi}_\varepsilon^{1/2} Z), r_\varepsilon Z')\right) \leqslant \mathsf{err}. \tag{73}$$

Proof. Let $\widehat{\boldsymbol{h}}'' = F^{-1}_{\varepsilon,\varrho_\varepsilon}(\boldsymbol{n}')$. Consider first $\dot{\boldsymbol{v}}' \in \{\|\dot{\boldsymbol{v}}'\| = 1, \dot{\boldsymbol{v}}' \perp \boldsymbol{m}\}$, Then $\widehat{\boldsymbol{g}}_{\mathsf{Pl}}(\dot{\boldsymbol{v}}') = \widehat{\boldsymbol{\xi}}$, so clearly

$$\mathbb{W}_2\left(\frac{1}{M}\sum_{a=1}^{M} \delta(\widehat{h}''_a, \widehat{g}_{\mathsf{Pl}}(\dot{\boldsymbol{v}}')_a), (\widetilde{q}_\varepsilon^{1/2} Z, Z')\right) = o_\upsilon(1).$$

For $(\boldsymbol{m},\boldsymbol{n}) \in U(r_0)$, let T be a rotation operator mapping $\boldsymbol{m}/\|\boldsymbol{m}\|$ to $\boldsymbol{m}'/\|\boldsymbol{m}'\|$. Note that $\|T - I\|_{\mathrm{op}} = o_{r_0}(1)$. Consider any $\dot{\boldsymbol{v}} \in \{\|\dot{\boldsymbol{v}}\| = 1, \dot{\boldsymbol{v}} \perp \boldsymbol{m}\}$, and let $\dot{\boldsymbol{v}}' = T\dot{\boldsymbol{v}}$, so $\|\dot{\boldsymbol{v}} - \dot{\boldsymbol{v}}'\| = o_{r_0}(1)$. Then

$$\|\widehat{\boldsymbol{g}}_{\mathsf{Pl}}(\dot{\boldsymbol{v}}') - \widehat{\boldsymbol{g}}_{\mathsf{Pl}}(\dot{\boldsymbol{v}})\| \leqslant O(1)\left(\|\widehat{\boldsymbol{h}}'\| + \|\widehat{\boldsymbol{\xi}}'\| + \|\widehat{\boldsymbol{\xi}}\|\right)\|\dot{\boldsymbol{v}} - \dot{\boldsymbol{v}}'\|.$$

With high probability over $\widehat{\boldsymbol{\xi}}, \widehat{\boldsymbol{\xi}}'$, this is bounded by $o_{r_0}(1)\sqrt{N}$. Thus

$$\mathbb{W}_2\left(\frac{1}{M}\sum_{a=1}^{M} \delta(\widehat{h}''_a, \widehat{g}_{\mathsf{Pl}}(\dot{\boldsymbol{v}})_a), (\widetilde{q}_\varepsilon^{1/2} Z, Z')\right) = o_{r_0}(1) + o_\upsilon(1). \tag{74}$$

Note that

$$\|\widehat{\boldsymbol{h}}' - \widehat{\boldsymbol{h}}''\| = \|F^{-1}_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\boldsymbol{n}') - F^{-1}_{\varepsilon,\varrho_\varepsilon}(\boldsymbol{n}')\| \leqslant \mathsf{err}\sqrt{N}.$$

Identically to (66) and (67), we can show

$$\|\widehat{\boldsymbol{h}}' - \acute{\boldsymbol{h}}\|, \|F_{\varepsilon,\varrho_\varepsilon}(\widehat{\boldsymbol{h}}'') - \boldsymbol{n}\| \leqslant \mathsf{err}\sqrt{N}.$$

Combined with (74), this proves (72). The proof of (73) is analogous. □

The following two propositions are proved identically to Propositions 6.13 and 6.14, with $\widehat{\boldsymbol{g}}_{\mathsf{Pl}}$, $\dot{\boldsymbol{g}}_{\mathsf{Pl}}$, and Lemma 6.18 playing the roles of $\widehat{\boldsymbol{g}}_{\mathrm{AMP}}$, $\dot{\boldsymbol{g}}_{\mathrm{AMP}}$, and Lemma 6.12.

Proposition 6.19. *For all $(\boldsymbol{m}', \boldsymbol{n}') \in \mathcal{S}_{\varepsilon,\upsilon}$, the following holds with high probability. Uniformly over $(\boldsymbol{m},\boldsymbol{n}) \in U(r_0)$, $\acute{\boldsymbol{h}} \in U'(r_0)$, $\dot{\boldsymbol{v}} \in \{\|\dot{\boldsymbol{v}}\| = 1, \dot{\boldsymbol{v}} \perp \boldsymbol{m}\}$, we have*

$$\inf_{\substack{\|\widehat{\boldsymbol{v}}\|=r_\varepsilon, \\ \widehat{\boldsymbol{v}}\perp\boldsymbol{n}}} \langle \boldsymbol{D}_2(\acute{\boldsymbol{h}})^{-1}\widehat{\boldsymbol{v}}, \widehat{\boldsymbol{v}}\rangle + \frac{2}{\sqrt{N}}\langle \widehat{\boldsymbol{v}}, \widehat{\boldsymbol{g}}_{\mathsf{Pl}}(\dot{\boldsymbol{v}})\rangle \leqslant -\alpha_\star \mathbb{E}\left[\frac{\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2} Z)}{1 + m_\varepsilon(z_\varepsilon)\widehat{f}_\varepsilon(\widetilde{q}_\varepsilon^{1/2} Z)}\right] - m_\varepsilon(z_\varepsilon) r_\varepsilon^2 + \mathsf{err}.$$

Proposition 6.20. *For all $(\boldsymbol{m}',\boldsymbol{n}')\in\mathcal{S}_{\varepsilon,v}$, the following holds with high probability. Uniformly over $(\boldsymbol{m},\boldsymbol{n})\in U(r_0)$, $\widehat{\boldsymbol{v}}\in\{\|\widehat{\boldsymbol{v}}\|=r_\varepsilon,\widehat{\boldsymbol{v}}\perp\boldsymbol{n}\}$, we have*

$$\sup_{\substack{\|\dot{\boldsymbol{v}}\|=1\\ \dot{\boldsymbol{v}}\perp\boldsymbol{m}}} -\langle \boldsymbol{D}_1\dot{\boldsymbol{v}},\dot{\boldsymbol{v}}\rangle + \frac{2}{\sqrt{N}}\langle \dot{\boldsymbol{v}},\dot{\boldsymbol{g}}_{\mathsf{Pl}}(\widehat{\boldsymbol{v}})\rangle \leqslant z_\varepsilon + m_\varepsilon(z_\varepsilon)r_\varepsilon^2 + \mathsf{err}.$$

Proof of Proposition 6.16. Adding Propositions 6.19 and 6.20 shows that (71) holds with high probability. The result follows from the discussion leading to (71). □

Proof of Proposition 4.8(c), under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$. By Proposition 4.8(d), $\|(\boldsymbol{m}^k,\boldsymbol{n}^k)-(\boldsymbol{m},\boldsymbol{n})\| = v_0\sqrt{N}$ with high probability. We set $v_0<r_0$. Since we defined

$$U(r_0)=\{(\boldsymbol{m},\boldsymbol{n}):\|(\boldsymbol{m},\boldsymbol{n})-(\boldsymbol{m}',\boldsymbol{n}')\|\leqslant 2r_0\sqrt{N}\}\supseteq\{(\boldsymbol{m},\boldsymbol{n}):\|(\boldsymbol{m},\boldsymbol{n})-(\boldsymbol{m}^k,\boldsymbol{n}^k)\|\leqslant r_0\sqrt{N}\},$$

the conclusion of Proposition 6.16 holds for all $\|(\boldsymbol{m},\boldsymbol{n})-(\boldsymbol{m}^k,\boldsymbol{n}^k)\|\leqslant r_0\sqrt{N}$. Identically to (70), we have

$$\{(\boldsymbol{m},\boldsymbol{n}):\|(\boldsymbol{m},\boldsymbol{n})-(\boldsymbol{m}^k,\boldsymbol{n}^k)\|\leqslant r_0\sqrt{N}\}\subseteq\mathcal{S}_{\varepsilon,2C_\varepsilon r_0}$$

for some $C_\varepsilon=O_\varepsilon(1)$. Since $\|\boldsymbol{G}\|_{\mathrm{op}},\|\widehat{\boldsymbol{g}}\|\leqslant C\sqrt{N}$ holds with high probability under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$, Lemma 6.4 holds. Applying this lemma (with $2C_\varepsilon r_0$ in place of r_0) gives that for all $\|(\boldsymbol{m},\boldsymbol{n})-(\boldsymbol{m}^k,\boldsymbol{n}^k)\|\leqslant r_0\sqrt{N}$,

$$\begin{aligned}\nabla^2_\diamond\mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n}) &\preceq \boldsymbol{R}(\boldsymbol{m},\boldsymbol{n})+\lambda_\varepsilon P_{\boldsymbol{m}}+(o_{C_{\mathsf{cvx}}}(1)+o_{r_0}(1))\boldsymbol{I}_N\\ &\preceq(\lambda_\varepsilon+o_{C_{\mathsf{cvx}}}(1)+o_{r_0}(1)+o_k(1))\boldsymbol{I}_N\\ &\preceq(\lambda_0+o_\varepsilon(1)+o_{C_{\mathsf{cvx}}}(1)+o_{r_0}(1)+o_k(1))\boldsymbol{I}_N.\end{aligned}$$

Under Condition 3.4, $\lambda_0<0$, and the result follows by setting the error terms small. □

6.5. **Determinant concentration.** In this subsection, we prove Lemma 4.9. We fix some $(\boldsymbol{m},\boldsymbol{n})\in\mathcal{S}_{\varepsilon,v}$ and work under the measure $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$. Define, as in Lemma 4.16,

$$\dot{\boldsymbol{h}}=\mathrm{th}_\varepsilon^{-1}(\boldsymbol{m}),\qquad \widehat{\boldsymbol{h}}=F^{-1}_{\varepsilon,\rho_\varepsilon(\boldsymbol{m})}(\boldsymbol{n}),\qquad \acute{\boldsymbol{h}}=\frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}}+\varepsilon^{1/2}\widehat{\boldsymbol{g}}-\rho_\varepsilon(q(\boldsymbol{m}))\boldsymbol{n}.$$

Recall from Lemma 4.16 that under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$, we have $\acute{\boldsymbol{h}}=\widehat{\boldsymbol{h}}$ deterministically. We computed $\nabla^2\mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})$ in Fact 6.5, and under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$ the matrices $\boldsymbol{D}_1,\widetilde{\boldsymbol{D}}_2,\boldsymbol{D}_3,\boldsymbol{D}_4$ therein are all nonrandom. By Schur's lemma,

$$|\det\nabla^2\mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})|=|\det\nabla^2_{\boldsymbol{n},\boldsymbol{n}}\mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})||\det\nabla^2_\diamond\mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})|, \tag{75}$$

and $\nabla^2_{\boldsymbol{n},\boldsymbol{n}}\mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})$ is nonrandom. By Fact 6.5,

$$\nabla^2_\diamond\mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})=-\boldsymbol{D}_1-\frac{1}{N}\boldsymbol{G}^\top\widetilde{\boldsymbol{D}}_2\boldsymbol{G}+\rho_\varepsilon'(q(\boldsymbol{m}))d_\varepsilon(\boldsymbol{m},\boldsymbol{n})\boldsymbol{I}_N+\frac{C}{N}\boldsymbol{m}\boldsymbol{m}^\top+\frac{1}{N}(\boldsymbol{G}^\top\boldsymbol{v}\boldsymbol{m}^\top+\boldsymbol{m}\boldsymbol{v}^\top\boldsymbol{G})$$

for some nonrandom $C\in\mathbb{R}$, $\boldsymbol{v}\in\mathbb{R}^M$ depending on $(\boldsymbol{m},\boldsymbol{n})$. By Lemma 6.6, $|C|$, $\|\boldsymbol{v}\|$ are uniformly bounded over $(\boldsymbol{m},\boldsymbol{n})\in\mathcal{S}_{\varepsilon,v}$, with bound depending on $\varepsilon,C_{\mathsf{cvx}}$. Define for convenience the nonrandom matrix

$$\boldsymbol{A}=\boldsymbol{D}_1-\rho_\varepsilon'(q(\boldsymbol{m}))d_\varepsilon(\boldsymbol{m},\boldsymbol{n})\boldsymbol{I}_N-\frac{C}{N}\boldsymbol{m}\boldsymbol{m}^\top$$

and note that $\|\boldsymbol{A}\|_{\mathrm{op}}$ is uniformly bounded (depending on $\varepsilon,C_{\mathsf{cvx}}$) over $(\boldsymbol{m},\boldsymbol{n})\in\mathcal{S}_{\varepsilon,v}$. Then let

$$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{A} & \frac{1}{\sqrt{N}}\boldsymbol{m}\boldsymbol{v}^\top & \frac{1}{\sqrt{N}}\boldsymbol{G}^\top\\ \frac{1}{\sqrt{N}}\boldsymbol{v}\boldsymbol{m}^\top & \widetilde{\boldsymbol{D}}_2 & \boldsymbol{I}_M\\ \frac{1}{\sqrt{N}}\boldsymbol{G} & \boldsymbol{I}_M & \boldsymbol{0}\end{bmatrix}\in\mathbb{R}^{(N+2M)\times(N+2M)}. \tag{76}$$

Lemma 6.21. *We have $|\det\nabla^2_\diamond\mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})|=|\det\boldsymbol{X}|$.*

Proof. Let $\boldsymbol{Y} = \begin{bmatrix} \widetilde{\boldsymbol{D}}_2 & \boldsymbol{I}_M \\ \boldsymbol{I}_M & \boldsymbol{0} \end{bmatrix}$. Note that $|\det \boldsymbol{Y}| = 1$ and $\boldsymbol{Y}^{-1} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I}_M \\ \boldsymbol{I}_M & -\widetilde{\boldsymbol{D}}_2 \end{bmatrix}$. By Schur's lemma,

$$|\det \boldsymbol{X}| = \left|\det\left(\boldsymbol{A} - \frac{1}{N}\begin{bmatrix}\boldsymbol{m}\boldsymbol{v}^\top & \boldsymbol{G}^\top\end{bmatrix}\boldsymbol{Y}^{-1}\begin{bmatrix}\boldsymbol{v}\boldsymbol{m}^\top \\ \boldsymbol{G}\end{bmatrix}\right)\right| = |\det \nabla^2_\diamond \mathcal{F}_{\mathrm{TAP}}(\boldsymbol{m}, \boldsymbol{n})|. \qquad \square$$

It therefore suffices to study $|\det \boldsymbol{X}|$. This formulation has the benefit that the only randomness in $\boldsymbol{X}$ is from $\boldsymbol{G}$, and by Lemma 4.17 (in a suitable orthonormal basis) $\boldsymbol{G}$ is a matrix of independent (noncentered) gaussians. This structure will enable us to prove Lemma 4.9 using the spectral concentration results of [GZ00]. Before carrying out this argument, we first prove a preliminary lemma.

Lemma 6.22. *There exists $\tau > 0$ depending on $\varepsilon, C_{\mathsf{cvx}}$ such that, for all $(\boldsymbol{m}, \boldsymbol{n}) \in \mathcal{S}_{\varepsilon,\upsilon}$, $\boldsymbol{X}$ has no eigenvalues in $[-\tau, \tau]$ with high probability under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$.*

Proof. We will show that $\det(z\boldsymbol{I}_{N+2M} - \boldsymbol{X})$ has no zeros in $[-\tau, \tau]$. By Schur's lemma, for any $z \neq 0$,

$$|\det(z\boldsymbol{I}_{2M} - \boldsymbol{Y})| = |\det(z\boldsymbol{I}_M - \widetilde{\boldsymbol{D}}_2)||\det(z\boldsymbol{I}_M - (z\boldsymbol{I}_M - \widetilde{\boldsymbol{D}}_2)^{-1})| = |\det(z(z\boldsymbol{I}_M - \widetilde{\boldsymbol{D}}_2) - \boldsymbol{I}_M)|$$

Let τ_1 be the smallest positive solution to $\tau_1 |\max(\widehat{f}_\varepsilon) + \tau| \leqslant \frac{1}{2}$. Note that τ_1 depends only on ε, and the above determinant is nonzero for any $|z| \leqslant \tau_1$. Further, note that

$$(z\boldsymbol{I}_{2M} - \boldsymbol{Y})^{-1} = \begin{bmatrix} -z(\boldsymbol{I}_M - z(z\boldsymbol{I}_M - \widetilde{\boldsymbol{D}}_2))^{-1} & (\boldsymbol{I}_M - z(z\boldsymbol{I}_M - \widetilde{\boldsymbol{D}}_2))^{-1} \\ (\boldsymbol{I}_M - z(z\boldsymbol{I}_M - \widetilde{\boldsymbol{D}}_2))^{-1} & -(z\boldsymbol{I}_M - \widetilde{\boldsymbol{D}}_2)(\boldsymbol{I}_M - z(z\boldsymbol{I}_M - \widetilde{\boldsymbol{D}}_2))^{-1} \end{bmatrix}.$$

From this, we see that there exists $C_\varepsilon > 0$ such that for all $|z| \leqslant \tau_1$,

$$\|(z\boldsymbol{I}_{2M} - \boldsymbol{Y})^{-1} + \boldsymbol{Y}^{-1}\|_{\mathrm{op}} \leqslant C_\varepsilon |z|.$$

By Schur's lemma, for all $|z| \leqslant \tau_1$,

$$|\det(z\boldsymbol{I}_{N+2M} - \boldsymbol{X})| = |\det(z\boldsymbol{I}_{2M} - \boldsymbol{Y})||\det \boldsymbol{B}(z)|,$$

for

$$\boldsymbol{B}(z) = z\boldsymbol{I}_N - \boldsymbol{A} - \frac{1}{N}\begin{bmatrix}\boldsymbol{m}\boldsymbol{v}^\top & \boldsymbol{G}^\top\end{bmatrix}(z\boldsymbol{I}_{2M} - \boldsymbol{Y})^{-1}\begin{bmatrix}\boldsymbol{v}\boldsymbol{m}^\top \\ \boldsymbol{G}\end{bmatrix}.$$

It follows that for all $|z| \leqslant \tau_1$,

$$\|\boldsymbol{B}(z) - \nabla^2_\diamond \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m}, \boldsymbol{n})\|_{\mathrm{op}} \leqslant |z| + C_\varepsilon |z| \left(\frac{\|\boldsymbol{v}\boldsymbol{m}^\top\|_{\mathrm{op}}}{\sqrt{N}} + \frac{\|\boldsymbol{G}\|_{\mathrm{op}}}{\sqrt{N}}\right)^2.$$

As shown in Proposition 4.8(c), $\nabla^2_\diamond \mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m}, \boldsymbol{n}) \preceq -C_{\mathrm{spec}}\boldsymbol{I}_N$ with high probability under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$. Furthermore, $\frac{\|\boldsymbol{v}\boldsymbol{m}^\top\|_{\mathrm{op}}}{\sqrt{N}} = \frac{1}{\sqrt{N}}\|\boldsymbol{v}\|\|\boldsymbol{m}\|$ is bounded, with bound depending on $\varepsilon, C_{\mathsf{cvx}}$, and with high probability, $\frac{\|\boldsymbol{G}\|_{\mathrm{op}}}{\sqrt{N}}$ is bounded by an absolute constant. It follows that for $|z|$ small enough depending on $\varepsilon, C_{\mathsf{cvx}}$, $\boldsymbol{B}(z) \preceq -C_{\mathrm{spec}}\boldsymbol{I}_N/2$, and thus $|\det \boldsymbol{B}(z)| \neq 0$. $\square$

The core of the proof of Lemma 4.9 is the following spectral concentration inequality, which adapts [GZ00, Theorem 1.1(b)]. For any $f : \mathbb{R} \to \mathbb{R}$, let

$$\mathrm{tr} f(\boldsymbol{X}) = \sum_{i=1}^{N+2M} f(\lambda_i(\boldsymbol{X})),$$

where $\lambda_1(\boldsymbol{X}), \ldots, \lambda_{N+2M}(\boldsymbol{X})$ are the eigenvalues of $\boldsymbol{X}$.

Lemma 6.23. *If f is L-Lipschitz, then for any $t \geqslant 0$,*

$$\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}(|\mathrm{tr} f(\boldsymbol{X}) - \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}} \mathrm{tr} f(\boldsymbol{X})| \geqslant t) \leqslant 2e^{-t^2/8L^2}.$$

Proof. Let $\{\omega_{a,i} : a \in [M], i \in [N]\}$ be i.i.d. standard gaussians, and let $\dot{\boldsymbol{e}}_1, \ldots, \dot{\boldsymbol{e}}_N$ and $\widehat{\boldsymbol{e}}_1, \ldots, \widehat{\boldsymbol{e}}_M$ be orthonormal bases of $\mathbb{R}^N$ and $\mathbb{R}^M$ as in Lemma 4.17. By (40), we can sample $\widetilde{\boldsymbol{G}}$ by

$$\widetilde{\boldsymbol{G}} = \sum_{a=1}^{M}\sum_{i=1}^{N} w_{a,i}\omega_{a,i}\widehat{\boldsymbol{e}}_a\dot{\boldsymbol{e}}_i^\top, \qquad w_{a,i} = \begin{cases} \sqrt{\varepsilon/(q(\boldsymbol{m})+\psi(\boldsymbol{n})+\varepsilon)} & i=j=1, \\ \sqrt{\varepsilon/(q(\boldsymbol{m})+\varepsilon)} & i=1, j\neq 1, \\ \sqrt{\varepsilon/(\psi(\boldsymbol{n})+\varepsilon)} & i\neq 1, j=1, \\ 1 & i\neq 1, j\neq 1. \end{cases}$$

By [GZ00, Lemma 1.2(b)], the map $\{\omega_{a,i} : a \in [M], i \in [N]\} \mapsto \mathrm{tr} f(\boldsymbol{X})$ is $2L$-Lipschitz. The result follows from the gaussian concentration inequality. ◻

Proof of Lemma 4.9. Define $f(x) = \log\max(|x|, \tau)$, which is τ^{-1}-Lipschitz. Lemma 6.23 implies that

$$\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}(|\mathrm{tr} f(\boldsymbol{X}) - \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}\mathrm{tr} f(\boldsymbol{X})| \geqslant t) \leqslant 2e^{-\tau^2 t^2/8}. \tag{77}$$

Let $\widetilde{\det}(\boldsymbol{X}) = \exp \mathrm{tr} f(\boldsymbol{X})$. Also let

$$\mathcal{E}_{\mathsf{spec}}(\boldsymbol{X}) = \{\mathrm{spec}(\boldsymbol{X}) \cap [-\tau,\tau] = \varnothing\},$$

so that $\mathbb{P}(\mathcal{E}_{\mathsf{spec}}) \geqslant 1-\iota$ for some $\iota = o_N(1)$ by Lemma 6.22. Note that $|\det(\boldsymbol{X})| \leqslant \widetilde{\det}(\boldsymbol{X})$ for all $\boldsymbol{X}$, with equality for all $\boldsymbol{X} \in \mathscr{E}_{\mathsf{spec}}$. Thus

$$\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[|\det(\boldsymbol{X})|^2] \leqslant \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[\widetilde{\det}(\boldsymbol{X})^2], \qquad \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[|\det(\boldsymbol{X})|] \geqslant \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[\widetilde{\det}(\boldsymbol{X})\mathbf{1}\{\mathcal{E}_{\mathsf{spec}}\}]. \tag{78}$$

By the concentration (77), there exists C depending on $\varepsilon, C_{\mathsf{cvx}}$ such that

$$\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[\widetilde{\det}(\boldsymbol{X})^2] \leqslant C\exp(2\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}\mathrm{tr} f(\boldsymbol{X})).$$

Furthermore, by Jensen's inequality $\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[\widetilde{\det}(\boldsymbol{X})] \geqslant \exp(\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}\mathrm{tr} f(\boldsymbol{X}))$. Thus,

$$\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[\widetilde{\det}(\boldsymbol{X})^2] \leqslant C\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[\widetilde{\det}(\boldsymbol{X})]^2. \tag{79}$$

By Cauchy–Schwarz,

$$\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[\widetilde{\det}(\boldsymbol{X})\mathbf{1}\{\mathcal{E}^c_{\mathsf{spec}}\}] \leqslant \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[\widetilde{\det}(\boldsymbol{X})^2]^{1/2}\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}(\mathcal{E}^c_{\mathsf{spec}})^{1/2} \leqslant C^{1/2}\iota^{1/2}\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[\widetilde{\det}(\boldsymbol{X})].$$

It follows that

$$\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[\widetilde{\det}(\boldsymbol{X})\mathbf{1}\{\mathcal{E}_{\mathsf{spec}}\}] \geqslant (1 - C^{1/2}\iota^{1/2})\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[\widetilde{\det}(\boldsymbol{X})].$$

Combining with (78), (79) shows that

$$\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[|\det(\boldsymbol{X})|^2]^{1/2} \leqslant C^{1/2}(1 - C^{1/2}\iota^{1/2})^{-1}\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[|\det(\boldsymbol{X})|],$$

which implies the result after adjusting C. ◻

7. First moment in planted model

In this section, we prove Proposition 3.9, bounding the first moment of $Z_N(\boldsymbol{G})$ in the planted model. The proof is structured as follows. In §7.1, we show this moment is bounded by a optimization problem over $\boldsymbol{\Lambda} : \mathbb{R} \to \mathbb{R}$ encoding subsets of Σ_N with a certain coordinate profile (heuristically described in (9)). §7.2 reduces this optimization to two dimensions by showing the maximizer is attained in a two-parameter family. For technical reasons, the functional in this optimization problem is not the $\mathscr{S}_\star$ defined in (8), but a variant $\mathscr{S}_\star^{s_{\max}}$ where s is minimized over $[0, s_{\max}]$ instead of $[0,+\infty)$ see (80). §7.3 and §7.4 show that we recover the optimization of $\mathscr{S}_\star$ when $s_{\max} \to \infty$, completing the proof of Proposition 3.9. §7.5 proves Lemma 2.5, on the local behavior of the first moment functional $\mathscr{S}_\star(\lambda_1, \lambda_2)$ near $(1, 0)$.

7.1. **Reduction to functional optimization.** Recall that (q_0, ψ_0) are given by Condition 3.1. Let $\dot{\boldsymbol{H}} \sim \mathcal{N}(0, \psi_0)$, $\boldsymbol{M} = \operatorname{th}(\dot{\boldsymbol{H}})$, and $\hat{\boldsymbol{H}} \sim \mathcal{N}(0, q_0)$, $\boldsymbol{N} = F_{1-q_0}(\hat{\boldsymbol{H}})$, for F_{1-q_0} given by (13). Let $\mathscr{L} = L^2(\mathbb{R}, \mathcal{N}(0, \psi_0))$ denote the space of measurable functions $\boldsymbol{\Lambda} : \mathbb{R} \to \mathbb{R}$, equipped with the inner product

$$\langle \boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2 \rangle = \mathbb{E}[\boldsymbol{\Lambda}_1(\dot{\boldsymbol{H}})\boldsymbol{\Lambda}_2(\dot{\boldsymbol{H}})]$$

and square-integrable w.r.t. the associated norm. Let $\mathscr{K} \subseteq \mathscr{L}$ denote the set of functions with image in $[-1, 1]$. For $s_{\max} > 0$, define

$$\mathcal{S}_\star^{s_{\max}}(\boldsymbol{\Lambda}) = \inf_{0 \leqslant s \leqslant s_{\max}} \mathcal{S}_\star(\boldsymbol{\Lambda}, s), \tag{80}$$

where $\mathcal{S}_\star : \mathscr{K} \times [0, +\infty) \to \mathbb{R}$ is defined by (7). The following proposition bounds the first moment by the maximum of an optimization problem over functions $\boldsymbol{\Lambda}$, and is the starting point of the proof of Proposition 3.9.

Proposition 7.1. *For any $s_{\max} > 0$, $(\boldsymbol{m}, \boldsymbol{n}) \in \mathcal{S}_{\varepsilon,\upsilon}$, we have $\frac{1}{N} \log \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathrm{Pl}}[Z_N(\boldsymbol{G})] \leqslant \sup_{\boldsymbol{\Lambda} \in \mathscr{K}} \mathcal{S}_\star^{s_{\max}}(\boldsymbol{\Lambda}) + o_{\varepsilon,\upsilon}(1)$.*

Here $o_{\varepsilon,\upsilon}(1)$ denotes a term vanishing as $\varepsilon, \upsilon \to 0$, which can depend on $s_{\max}$; we send $s_{\max} \to \infty$ after $\varepsilon, \upsilon \to 0$ in the end.

Before proving Proposition 7.1, we state a few facts that will be useful below. Lemma 7.2 ensures that the denominator of $\mathcal{S}_\star(\boldsymbol{\Lambda}, s)$ is well-behaved, while Lemmas 7.3 and 7.4 are useful in approximation arguments.

Lemma 7.2. *There exists $\iota > 0$ such that $\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})]^2 < (1 - \iota)q_0$ for all $\boldsymbol{\Lambda} \in \mathscr{K}$.*

Proof. Since $|\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})| \leqslant 1$, by Cauchy–Schwarz,

$$\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})]^2 \leqslant \mathbb{E}[|\boldsymbol{M}|]^2 < \mathbb{E}[\boldsymbol{M}^2].$$

The inequality is strict because $|\boldsymbol{M}|$ has nonzero variance. Since $\mathbb{E}[\boldsymbol{M}^2] = P(\psi_0) = q_0$ (recall Condition 3.1), the result follows. □

Lemma 7.3. *The function $\log \Psi(x)$ is $(2, 1)$-pseudo-Lipschitz (recall Definition 4.19).*

Proof. Note that $(\log \Psi)'(x) = -\mathcal{E}(x)$. Recall from Lemma 4.21(a) that $0 \leqslant \mathcal{E}(x) \leqslant 1 + |x|$. Thus,

$$|\log \Psi(x) - \log \Psi(y)| = \left| \int_x^y \mathcal{E}(s)\, \mathrm{d}s \right| \leqslant |x - y|(1 + |x| + |y|).$$

□

Lemma 7.4 (Proved in Appendix A)**.** *There exists $C > 0$ such that for all $a_1, a_2, b_1, b_2, c_1, c_2 > 0$,*

$$\left| \mathbb{E} \log \Psi \left\{ \frac{\kappa - a_1 \hat{\boldsymbol{H}} - b_1 \boldsymbol{N}}{c_1} \right\} - \log \Psi \left\{ \frac{\kappa - a_2 \hat{\boldsymbol{H}} - b_2 \boldsymbol{N}}{c_2} \right\} \right|$$
$$\leqslant \frac{C \max(a_1, a_2, b_1, b_2, c_1, c_2, 1)^3}{\min(c_1, c_2)^2} \left(|a_1 - a_2| + |b_1 - b_2| + |c_1 - c_2| \right).$$

We turn to the proof of Proposition 7.1. The main step will be Proposition 7.5 below, where we show the bound in Proposition 7.1 holds for piecewise-constant $\boldsymbol{\Lambda}$ with finitely many parts. This case follows from a direct moment calculation, and Proposition 7.1 follows by approximation.

For any $\vec{r} = (r_1, \ldots, r_{n-1})$ with $-\infty < r_1 < r_2 < \cdots < r_{n-1} < +\infty$, let $\mathscr{K}_{\mathrm{elt}}(\vec{r}) \subseteq \mathscr{K}$ denote the set of right-continuous functions which are constant on each interval $[r_{k-1}, r_k)$, $1 \leqslant k \leqslant n$. Here we take as convention $r_0 = -\infty$, $r_n = +\infty$. Define the quantiles $\vec{p} = (p_0, \ldots, p_n)$ by $p_k = \mathbb{P}(\dot{\boldsymbol{H}} < r_k)$, and let

$$\mathsf{mesh}(\vec{p}) = \min_{1 \leqslant k \leqslant n} (p_k - p_{k-1}).$$

Let $o_{\varepsilon,\upsilon,\vec{p}}(1)$ denote a term vanishing as $\varepsilon, \upsilon, \mathrm{mesh}(\vec{p}) \to 0$, where (like before) this limit is taken after $N \to \infty$ for fixed $s_{\max}$. We will show the following.

Proposition 7.5. *Suppose $s_{\max} > 0$, $(\boldsymbol{m}, \boldsymbol{n}) \in \mathcal{S}_{\varepsilon,\upsilon}$, and $\vec{r} = (r_1, \dots, r_{n-1})$ is as above. We have that $\frac{1}{N} \log \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[Z_N(\boldsymbol{G})] \leqslant \sup_{\boldsymbol{\Lambda} \in \mathscr{K}_{\mathrm{elt}}(\vec{r})} \mathcal{S}^{s_{\max}}_{\star}(\boldsymbol{\Lambda}) + o_{\varepsilon,\upsilon,\vec{p}}(1)$.*

For the rest of this subsection, fix $s_{\max}, \varepsilon, \upsilon, \vec{r}$ and $(\boldsymbol{m}, \boldsymbol{n})$ as in Proposition 7.5. Let $\dot{\boldsymbol{h}} = \mathrm{th}_\varepsilon^{-1}(\boldsymbol{m})$ and $\widehat{\boldsymbol{h}} = F^{-1}_{\varepsilon,\varrho_\varepsilon}(\boldsymbol{n})$, so that $(\dot{\boldsymbol{h}}, \widehat{\boldsymbol{h}}) \in \mathcal{T}_{\varepsilon,\upsilon}$. Fix a partition $[N] = \mathcal{I}_1 \cup \cdots \cup \mathcal{I}_n$ satisfying

$$|\mathcal{I}_k| = \lfloor p_k N \rfloor - \lfloor p_{k-1} N \rfloor, \qquad \forall 1 \leqslant k \leqslant n,$$
$$\max\{\dot{h}_i : i \in \mathcal{I}_k\} \leqslant \min\{\dot{h}_i : i \in \mathcal{I}_{k+1}\}, \qquad \forall 1 \leqslant k \leqslant n-1.$$

(In words, $\mathcal{I}_k$ is the set of coordinates $i \in [N]$ such that the quantile of $\dot{h}_i$ among the entries of $\dot{\boldsymbol{h}}$, breaking ties in an arbitrary but fixed order, lies in $[p_{k-1}, p_k)$.) Then, partition Σ_N into sets

$$\Sigma_N(\vec{a}) = \left\{ \boldsymbol{x} \in \Sigma_N : \sum_{i \in \mathcal{I}_k} x_i = a_k, \ \forall 1 \leqslant k \leqslant n \right\}. \tag{81}$$

indexed by $\vec{a} = (a_1, \dots, a_n) \in \mathbb{Z}^n$. Let $\mathcal{J}$ be the set of $\vec{a}$ such that $\Sigma_N(\vec{a})$ is nonempty, and note that $|\mathcal{J}| \leqslant N^n$. Thus

$$\begin{aligned} \frac{1}{N} \log \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[Z_N(\boldsymbol{G})] &= \frac{1}{N} \log \sum_{\vec{a} \in \mathcal{J}} \sum_{\boldsymbol{x} \in \Sigma_N(\vec{a})} \mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}} \left(\frac{\boldsymbol{G}\boldsymbol{x}}{\sqrt{N}} \geqslant \kappa \right) \\ &= \sup_{\vec{a} \in \mathcal{J}} \left\{ \frac{1}{N} \log |\Sigma_N(\vec{a})| + \sup_{\boldsymbol{x} \in \Sigma_N(\vec{a})} \frac{1}{N} \log \mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}} \left(\frac{\boldsymbol{G}\boldsymbol{x}}{\sqrt{N}} \geqslant \kappa \right) \right\} + o_N(1). \end{aligned} \tag{82}$$

Associate to each $\vec{a} \in \mathcal{J}$ a function $\boldsymbol{\Lambda}^{\vec{a}} \in \mathscr{K}_{\mathrm{elt}}(r_1, \dots, r_{n-1})$ defined by

$$\boldsymbol{\Lambda}^{\vec{a}}(x) = \frac{a_k}{|\mathcal{I}_k|}, \qquad x \in [r_{k-1}, r_k), 1 \leqslant k \leqslant n.$$

Recall the function $\mathsf{ent} : \mathscr{K} \to \mathbb{R}$ defined in (6).

Lemma 7.6. *We have $\frac{1}{N} \log |\Sigma_N(\vec{a})| = \mathsf{ent}(\boldsymbol{\Lambda}^{\vec{a}}) + o_N(1)$ for an error $o_N(1)$ uniform over $\vec{a} \in \mathcal{J}$.*

Proof. By direct counting,

$$|\Sigma_N(\vec{a})| = \prod_{k=1}^n \binom{|\mathcal{I}_k|}{\frac{1}{2}(|\mathcal{I}_k| + a_k)}.$$

Stirling's approximation yields

$$\frac{1}{N} \log |\Sigma_N(\vec{a})| = \sum_{k=1}^n \left\{ (p_k - p_{k-1}) \mathcal{H} \left(\frac{1 + \frac{a_k}{(p_k - p_{k-1})N}}{2} \right) \right\} + o_N(1) = \mathbb{E}\, \mathcal{H} \left(\frac{1 + \boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})}{2} \right) + o_N(1),$$

where the last equality holds because $\mathbb{P}(\dot{\boldsymbol{H}} \in [r_{k-1}, r_k)) = p_k - p_{k-1}$. ◻

Lemma 7.7. *For all $\vec{a} \in \mathcal{J}$ and $\boldsymbol{x} \in \Sigma_N(\vec{a})$,*

$$\frac{1}{N} \langle \dot{\boldsymbol{h}}, \boldsymbol{x} \rangle = \mathbb{E}[\dot{\boldsymbol{H}} \boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})] + o_{\varepsilon,\upsilon,\vec{p}}(1), \qquad \frac{1}{N} \langle \boldsymbol{m}, \boldsymbol{x} \rangle = \mathbb{E}[\boldsymbol{M} \boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})] + o_{\varepsilon,\upsilon,\vec{p}}(1),$$

for error terms $o_{\varepsilon,\upsilon,\vec{p}}(1)$ uniform over $\vec{a}, \boldsymbol{x}$.

Proof. We will only show the proof for $\frac{1}{N}\langle \dot{\boldsymbol{h}}, \boldsymbol{x}\rangle$, as the other estimate is analogous. Let $\boldsymbol{x} \in \Sigma_N(\vec{a})$ be fixed, and let $\boldsymbol{y} \in [-1,1]^N$ be defined by $y_i = \frac{a_k}{|\mathcal{I}_k|}$ for all $i \in \mathcal{I}_k$. We write $(\dot{\boldsymbol{H}}', \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{K})$ for the random variable with value $(\dot{h}_i, x_i, y_i, k)$, where $i \sim \mathsf{unif}([N])$ and $k \in [n]$ is the index of the set $\mathcal{I}_k$ containing i. Recall that $\dot{\boldsymbol{H}} \sim \mathcal{N}(0, \psi_0)$. Note that

$$\mathbb{W}_2(\mathcal{L}(\dot{\boldsymbol{H}}'), \mathcal{L}(\dot{\boldsymbol{H}})) \leqslant \mathbb{W}_2(\mu_{\dot{\boldsymbol{h}}}, \mathcal{N}(0, \psi_\varepsilon + \varepsilon)) + \mathbb{W}_2(\mathcal{N}(0, \psi_\varepsilon + \varepsilon), \mathcal{N}(0, \psi_0)) = o_{\varepsilon,v}(1),$$

where the latter two distances are bounded by definition of $\mathcal{T}_v$ and Proposition 4.1, respectively. We couple $(\dot{\boldsymbol{H}}', \dot{\boldsymbol{H}})$ monotonically (which is the $\mathbb{W}_2$-optimal coupling) and write

$$\frac{1}{N}\langle \dot{\boldsymbol{h}}, \boldsymbol{x}\rangle = \mathbb{E}[\dot{\boldsymbol{H}}'\boldsymbol{X}] = \mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{Y}] + \mathbb{E}[(\dot{\boldsymbol{H}}' - \dot{\boldsymbol{H}})\boldsymbol{X}] + \mathbb{E}[\dot{\boldsymbol{H}}(\boldsymbol{X} - \boldsymbol{Y})].$$

We now estimate each of these terms. Because $(\dot{\boldsymbol{H}}', \dot{\boldsymbol{H}})$ are coupled monotonically, $\boldsymbol{K} = k$ if and only if the quantile of $\dot{\boldsymbol{H}}$ lies in $[p'_{k-1}, p'_k)$, where $p'_k = \frac{1}{N}\lfloor p_k N\rfloor = p_k + O(N^{-1})$. Thus, on an event with probability $1 - O(N^{-1})$, $\boldsymbol{K} = k$ if and only if $\dot{\boldsymbol{H}} \in [r_{k-1}, r_k)$. On this event, $\boldsymbol{Y} = \boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})$. Thus

$$\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{Y}] = \mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})] + o_N(1).$$

Moreover,

$$|\,\mathbb{E}[(\dot{\boldsymbol{H}}' - \dot{\boldsymbol{H}})\boldsymbol{X}]| \leqslant \mathbb{E}[(\dot{\boldsymbol{H}}' - \dot{\boldsymbol{H}})^2]^{1/2} = \mathbb{W}_2(\mathcal{L}(\dot{\boldsymbol{H}}'), \mathcal{L}(\dot{\boldsymbol{H}})) = o_{\varepsilon,v}(1).$$

Finally, note that $\boldsymbol{Y} = \mathbb{E}[\boldsymbol{X}|\boldsymbol{K}]$, so

$$\mathbb{E}[\mathbb{E}[\dot{\boldsymbol{H}}|\boldsymbol{K}](\boldsymbol{X} - \boldsymbol{Y})] = \mathbb{E}[\mathbb{E}[\dot{\boldsymbol{H}}|\boldsymbol{K}]\,\mathbb{E}[\boldsymbol{X} - \boldsymbol{Y}|\boldsymbol{K}]] = 0.$$

Thus

$$|\,\mathbb{E}[\dot{\boldsymbol{H}}(\boldsymbol{X} - \boldsymbol{Y})]| = |\,\mathbb{E}[(\dot{\boldsymbol{H}} - \mathbb{E}[\dot{\boldsymbol{H}}|\boldsymbol{K}])(\boldsymbol{X} - \boldsymbol{Y})]| \leqslant \mathbb{E}[(\dot{\boldsymbol{H}} - \mathbb{E}[\dot{\boldsymbol{H}}|\boldsymbol{K}])^2]^{1/2}.$$

Recall from the above discussion that conditioning on $\boldsymbol{K}$ reveals the interval $[p'_{k-1}, p'_k)$ containing the quantile of $\dot{\boldsymbol{H}}$. It follows that $\mathbb{E}[(\dot{\boldsymbol{H}} - \mathbb{E}[\dot{\boldsymbol{H}}|\boldsymbol{K}])^2] = o_{\varepsilon,v,\vec{p}}(1)$. □

Lemma 7.8. *For all $\vec{a} \in \mathcal{J}$, $\boldsymbol{x} \in \Sigma_N(\vec{a})$, and $s \in [0, s_{\max}]$,*

$$\frac{1}{N}\log \mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}\left(\frac{\boldsymbol{G}\boldsymbol{x}}{\sqrt{N}} \geqslant \kappa\right) \leqslant \frac{1}{2}s^2\psi_0 + \alpha_\star\,\mathbb{E}\log\Psi\left\{\frac{\kappa - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})]}{q_0}\hat{\boldsymbol{H}} - \frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})]}{\psi_0}\boldsymbol{N}}{\sqrt{1 - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})]^2}{q_0}}} + s\boldsymbol{N}\right\} + o_{\varepsilon,v,\vec{p}}(1),$$

where the $o_{\varepsilon,v,\vec{p}}(1)$ is uniform over $\vec{a}, \boldsymbol{x}, s$ (but can depend on $s_{\max}$).

Proof. Let $\widetilde{\boldsymbol{G}}$ be defined in Corollary 4.18. By Corollary 4.18 and Lemma 7.7,

$$\begin{aligned}\frac{\boldsymbol{G}\boldsymbol{x}}{\sqrt{N}} &\overset{d}{=} \left(\frac{(1 + o_{\varepsilon,v}(1))}{q_0}\hat{\boldsymbol{h}} + o_{\varepsilon,v}(1)\boldsymbol{n}\right)\frac{1}{N}\langle \boldsymbol{m}, \boldsymbol{x}\rangle + \frac{(1 + o_{\varepsilon,v}(1))}{\psi_0}\boldsymbol{n}\cdot\frac{1}{N}\langle \dot{\boldsymbol{h}}, \boldsymbol{x}\rangle + \frac{\widetilde{\boldsymbol{G}}\boldsymbol{x}}{\sqrt{N}} \\ &= \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})] + o_{\varepsilon,v,\vec{p}}(1)}{q_0}\hat{\boldsymbol{h}} + \frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})] + o_{\varepsilon,v,\vec{p}}(1)}{\psi_0}\boldsymbol{n} + \frac{\widetilde{\boldsymbol{G}}\boldsymbol{x}}{\sqrt{N}}.\end{aligned}$$

Let $\hat{\boldsymbol{n}} = \boldsymbol{n}/\|\boldsymbol{n}\|$. By inspecting (40), we see that for independent $\widetilde{\boldsymbol{g}} \sim \mathcal{N}(0, P_{\boldsymbol{n}}^{\perp})$ and $Z \sim \mathcal{N}(0,1)$,

$$\frac{\widetilde{\boldsymbol{G}}\boldsymbol{x}}{\sqrt{N}} \overset{d}{=} \left(\frac{\|P_{\boldsymbol{m}}^{\perp}(\boldsymbol{x})\|^2}{N} + o_\varepsilon(1)\right)^{1/2}\widetilde{\boldsymbol{g}} + o_\varepsilon(1)Z\hat{\boldsymbol{n}} = t^{1/2}\widetilde{\boldsymbol{g}} + \iota_1^{1/2}Z\hat{\boldsymbol{n}},$$

where $t = 1 - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})]^2}{q_0} + \iota_2$ and $\iota_1, \iota_2 = o_{\varepsilon,v,\vec{p}}(1)$. For $Z' \sim \mathcal{N}(0,1)$ independent of $\widetilde{\boldsymbol{g}}, Z$, let

$$\hat{\boldsymbol{g}} = \widetilde{\boldsymbol{g}} + Z'\hat{\boldsymbol{n}} + s\boldsymbol{n}$$

so that $\widehat{\boldsymbol{g}} \sim \mathcal{N}(s\boldsymbol{n}, \boldsymbol{I}_N)$. Then, for any measurable $S \subseteq \mathbb{R}^N$,

$$\frac{\mathbb{P}(t^{1/2}\widetilde{\boldsymbol{g}} + \iota_1^{1/2} Z \widehat{\boldsymbol{n}} \in S)}{\mathbb{P}(t^{1/2}\widehat{\boldsymbol{g}} \in S)} \leqslant \sup_{T \subseteq \mathbb{R}} \frac{\mathbb{P}(\iota_1^{1/2} Z \in T)}{\mathbb{P}(st^{1/2}\|\boldsymbol{n}\| + t^{1/2}Z' \in T)}$$
$$\leqslant \sup_{x\in\mathbb{R}} \frac{\iota_1^{-1/2}\exp(-\frac{1}{2\iota_1}x^2)}{t^{-1/2}\exp(-\frac{1}{2t}(x - st^{1/2}\|\boldsymbol{n}\|)^2)} = \sqrt{\frac{t}{\iota_1}} \exp\left(\frac{s^2\|\boldsymbol{n}\|^2}{2(1-\iota_1/t)}\right).$$

Thus,

$$\frac{1}{N}\log \mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}\left(\frac{\boldsymbol{G}\boldsymbol{x}}{\sqrt{N}} \geqslant \kappa\right) \leqslant \frac{s^2\psi(\boldsymbol{n})}{2(1-\iota_1/t)} + o_N(1)$$
$$+ \frac{1}{N}\log \mathbb{P}\left\{\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})] + o_{\varepsilon,\upsilon,\vec{p}}(1)}{q_0}\widehat{\boldsymbol{h}} + \frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})] + o_{\varepsilon,\upsilon,\vec{p}}(1)}{\psi_0}\boldsymbol{n} + t^{1/2}\widehat{\boldsymbol{g}} \geqslant \kappa\right\}. \tag{83}$$

By Lemma 7.2, t is bounded away from 0. Since $\psi(\boldsymbol{n}) = \psi_0 + o_\varepsilon(1)$, we have

$$\frac{s^2\psi(\boldsymbol{n})}{2(1-\iota_1/t)} = (1 + o_{\varepsilon,\upsilon,\vec{p}}(1))\frac{1}{2}s^2\psi_0 = \frac{1}{2}s^2\psi_0 + o_{\varepsilon,\upsilon,\vec{p}}(1).$$

The last estimate holds uniformly over $s \in [0, s_{\max}]$. The last term of (83) equals

$$\frac{1}{N}\sum_{a=1}^M \log\Psi\left\{\frac{\kappa - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})] + o_{\varepsilon,\upsilon,\vec{p}}(1)}{q_0}\widehat{h}_a - \frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})] + o_{\varepsilon,\upsilon,\vec{p}}(1)}{\psi_0}n_a}{\sqrt{1 - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})]^2}{q_0} + o_{\varepsilon,\upsilon,\vec{p}}(1)}} + sn_a\right\} + o_N(1).$$

By Lemma 7.3, $\log\Psi$ is $(2,1)$-pseudo-Lipschitz. By Fact 4.20 and Lemma 7.4 (using again that the denominator is bounded away from 0), the last display equals

$$\alpha_\star \mathbb{E}\log\Psi\left\{\frac{\kappa - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})]}{q_0}\hat{\boldsymbol{H}} - \frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})]}{\psi_0}\boldsymbol{N}}{\sqrt{1 - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}^{\vec{a}}(\dot{\boldsymbol{H}})]^2}{q_0}}} + s\boldsymbol{N}\right\} + o_{\varepsilon,\upsilon,\vec{p}}(1).$$

Combining the above concludes the proof. □

Proof of Proposition 7.5. Follows from equation (82) and Lemmas 7.6 and 7.8. □

Proof of Proposition 7.1. Set $\vec{r}$ such that $\mathsf{mesh}(\vec{p})$ is suitably small depending on (ε, υ). Then

$$\frac{1}{N}\log \mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[Z_N(\boldsymbol{G})] \leqslant \sup_{\boldsymbol{\Lambda}\in\mathscr{K}_{\mathrm{elt}}(\vec{r})} \mathcal{S}_\star^{s_{\max}}(\boldsymbol{\Lambda}) + o_{\varepsilon,\upsilon}(1) \leqslant \sup_{\boldsymbol{\Lambda}\in\mathscr{K}} \mathcal{S}_\star^{s_{\max}}(\boldsymbol{\Lambda}) + o_{\varepsilon,\upsilon}(1).$$

□

7.2. **Reduction to two parameters.** Let $\mathscr{K}_* \subseteq \mathscr{K}$ denote the set of functions of the form $\boldsymbol{\Lambda}_{\lambda_1,\lambda_2}$ defined above (8). Let $\overline{\mathscr{K}}_*$ denote the closure of this set in the topology of $\mathscr{L}$. We next prove the following, which reduces the functional optimization problem in Proposition 7.1 to an optimization over $\overline{\mathscr{K}}_*$.

Proposition 7.9. *For any $s_{\max} > 0$, we have $\sup_{\boldsymbol{\Lambda}\in\mathscr{K}} \mathcal{S}_\star^{s_{\max}}(\boldsymbol{\Lambda}) = \sup_{\boldsymbol{\Lambda}\in\overline{\mathscr{K}}_*} \mathcal{S}_\star^{s_{\max}}(\boldsymbol{\Lambda})$. Similarly, $\sup_{\boldsymbol{\Lambda}\in\mathscr{K}} \mathcal{S}_\star(\boldsymbol{\Lambda}) = \sup_{\boldsymbol{\Lambda}\in\overline{\mathscr{K}}_*} \mathcal{S}_\star(\boldsymbol{\Lambda})$ for $\mathcal{S}_\star(\boldsymbol{\Lambda})$ defined in (8).*

Lemma 7.10. *Let $a_1, a_2 \in \mathbb{R}$ be such that there exists $\mathbf{\Lambda} \in \mathscr{K}$ with $\mathbb{E}[\dot{\mathbf{H}}\mathbf{\Lambda}(\dot{\mathbf{H}})] = a_1$, $\mathbb{E}[\mathbf{M}\mathbf{\Lambda}(\dot{\mathbf{H}})] = a_2$. Then, the concave optimization problem*

$$\textit{maximize}\ \mathsf{ent}(\mathbf{\Lambda}) \qquad \textit{subject to} \quad \mathbf{\Lambda} \in \mathscr{K}, \quad \mathbb{E}[\dot{\mathbf{H}}\mathbf{\Lambda}(\dot{\mathbf{H}}))] = a_1, \quad \mathbb{E}[\mathbf{M}\mathbf{\Lambda}(\dot{\mathbf{H}}))] = a_2$$

has a maximizer in $\overline{\mathscr{K}}_*$.

Proof. Introduce Lagrange multipliers $\lambda_1, \lambda_2 \in \mathbb{R}$. The Lagrangian is

$$L(\mathbf{\Lambda}; \lambda_1, \lambda_2) = \mathbb{E}\left\{\mathcal{H}\left(\frac{1+\mathbf{\Lambda}(\dot{\mathbf{H}})}{2}\right) + \lambda_1 \dot{\mathbf{H}}\mathbf{\Lambda}(\dot{\mathbf{H}}) + \lambda_2 \mathbf{M}\mathbf{\Lambda}(\dot{\mathbf{H}})\right\} - \lambda_1 a_1 - \lambda_2 a_2.$$

The quantity inside the expectation is concave in $\mathbf{\Lambda}(\dot{\mathbf{H}})$, with derivative

$$-\mathrm{th}^{-1}(\mathbf{\Lambda}(\dot{\mathbf{H}})) + \lambda_1 \dot{\mathbf{H}} + \lambda_2 \mathbf{M}.$$

This is pointwise maximized by $\mathbf{\Lambda}(\dot{\mathbf{H}}) = \mathrm{th}(\lambda_1 \dot{\mathbf{H}} + \lambda_2 \mathbf{M})$, i.e. $\mathbf{\Lambda} = \mathbf{\Lambda}_{\lambda_1,\lambda_2}$. □

Proof of Proposition 7.9. Note that $\mathscr{S}_\star^{s_{\max}}(\mathbf{\Lambda})$ is the sum of $\mathsf{ent}(\mathbf{\Lambda})$ and a term depending on $\mathbf{\Lambda}$ only through $\mathbb{E}[\dot{\mathbf{H}}\mathbf{\Lambda}(\dot{\mathbf{H}})]$ and $\mathbb{E}[\mathbf{M}\mathbf{\Lambda}(\dot{\mathbf{H}})]$. Let $\mathbf{\Lambda} \in \mathscr{K}$ be arbitrary. By Lemma 7.10, the maximum of $\mathsf{ent}(\widetilde{\mathbf{\Lambda}})$ subject to $\widetilde{\mathbf{\Lambda}} \in \mathscr{K}$, $\mathbb{E}[\dot{\mathbf{H}}\widetilde{\mathbf{\Lambda}}(\dot{\mathbf{H}})] = \mathbb{E}[\dot{\mathbf{H}}\mathbf{\Lambda}(\dot{\mathbf{H}})]$, $\mathbb{E}[\mathbf{M}\widetilde{\mathbf{\Lambda}}(\dot{\mathbf{H}})] = \mathbb{E}[\mathbf{M}\mathbf{\Lambda}(\dot{\mathbf{H}})]$ is attained by some $\widetilde{\mathbf{\Lambda}} \in \overline{\mathscr{K}}_*$. Thus $\mathscr{S}_\star^{s_{\max}}(\mathbf{\Lambda}) \leqslant \mathscr{S}_\star^{s_{\max}}(\widetilde{\mathbf{\Lambda}})$, which implies the conclusion for $\mathscr{S}^{s_{\max}}$. The proof for $\mathscr{S}_\star$ is identical. □

7.3. **The $s_{\max} \to \infty$ limit.** In this subsection, we prove the following proposition, which shows that the optimization problem derived in Proposition 7.9 has a well-behaved limit when we take $s_{\max} \to \infty$. This allows us to remove the parameter $s_{\max}$, replacing the constrained optimization $\mathscr{S}_\star^{s_{\max}}$ defined in (80) with the $\mathscr{S}_\star$ defined in (8).

Proposition 7.11. *We have $\lim_{s_{\max}\to\infty} \sup_{\mathbf{\Lambda}\in\overline{\mathscr{K}}_*} \mathscr{S}_\star^{s_{\max}}(\mathbf{\Lambda}) = \sup_{\mathbf{\Lambda}\in\overline{\mathscr{K}}_*} \mathscr{S}_\star(\mathbf{\Lambda})$, and moreover $\mathscr{S}_\star$ attains its supremum on $\overline{\mathscr{K}}_*$.*

Lemma 7.12. *The function $\mathscr{S}_\star : \mathscr{K} \times \mathbb{R} \to \mathbb{R}$ (recall (7)) is continuous.*

Proof. Note that $s \mapsto \frac{1}{2}s^2\psi_0$ is manifestly continuous. By concavity of $\mathcal{H}$, $|\mathcal{H}(x) - \mathcal{H}(y)| \leqslant \mathcal{H}(|x-y|)$ for all $x, y \in [0,1]$. By concavity of $x \mapsto \mathcal{H}(\sqrt{x}/2)$ and Jensen's inequality,

$$\begin{aligned}|\mathsf{ent}(\mathbf{\Lambda}) - \mathsf{ent}(\mathbf{\Lambda}')| &\leqslant \mathbb{E}\left|\mathcal{H}\left(\frac{1+\mathbf{\Lambda}(\dot{\mathbf{H}})}{2}\right) - \mathcal{H}\left(\frac{1+\mathbf{\Lambda}'(\dot{\mathbf{H}})}{2}\right)\right| \leqslant \mathbb{E}\left|\mathcal{H}\left(\frac{|\mathbf{\Lambda}(\dot{\mathbf{H}}) - \mathbf{\Lambda}'(\dot{\mathbf{H}})|}{2}\right)\right| \\ &\leqslant \mathcal{H}\left(\frac{\mathbb{E}[|\mathbf{\Lambda}(\dot{\mathbf{H}}) - \mathbf{\Lambda}'(\dot{\mathbf{H}})|^2]^{1/2}}{2}\right) = \mathcal{H}\left(\frac{\|\mathbf{\Lambda} - \mathbf{\Lambda}'\|}{2}\right).\end{aligned}$$

Thus ent is continuous. By Cauchy–Schwarz,

$$|\mathbb{E}[\dot{\mathbf{H}}\mathbf{\Lambda}] - \mathbb{E}[\dot{\mathbf{H}}\mathbf{\Lambda}']| \leqslant \mathbb{E}[\dot{\mathbf{H}}^2]^{1/2}\|\mathbf{\Lambda} - \mathbf{\Lambda}'\| = \psi_0^{1/2}\|\mathbf{\Lambda} - \mathbf{\Lambda}'\|$$

and similarly $|\mathbb{E}[\mathbf{M}\mathbf{\Lambda}] - \mathbb{E}[\mathbf{M}\mathbf{\Lambda}']| \leqslant q_0^{1/2}\|\mathbf{\Lambda} - \mathbf{\Lambda}'\|$. Since the denominator $1 - \frac{\mathbb{E}[\mathbf{M}\mathbf{\Lambda}(\dot{\mathbf{H}})]^2}{q_0}$ is bounded away from 0 by Lemma 7.2, the final term of $\mathscr{S}_\star$ is continuous by Lemma 7.4. Thus $\mathscr{S}_\star$ is continous. □

We will need the following analytical lemma, which is a simple adaptation of Dini's Theorem [Rud76, Theorem 7.13]. We provide a proof for completeness.

Lemma 7.13. *Suppose $f_1, f_2, \ldots : K \to \mathbb{R}$ are a decreasing sequence of continuous functions on a compact space K. Let $f : K \to \mathbb{R} \cup \{-\infty\}$ denote their (not necessarily continuous) pointwise limit, which we assume is not $-\infty$ everywhere. Then $\lim_{n\to\infty} \sup f_n = \sup f$, and furthermore f attains its supremum.*

Proof. Without loss of generality assume $\sup f = 0$. For $\iota > 0$, let $E_n = \{x \in K : f_n(x) < \iota\}$. Then E_n is open and $E_n \subseteq E_{n+1}$. Since the f_n converge pointwise to f, $\cup_n E_n = K$. By compactness of K, $E_n = K$ for some finite n, and thus $\sup f_n < \iota$. As this holds for any ι, $\lim_{n\to\infty} \sup f_n = 0$. Finally, f, as the decreasing limit of (upper-semi)continuous functions, is upper-semicontinuous. Therefore f attains its supremum. □

To apply Lemma 7.13, we verify that $\mathscr{S}_\star$ is not $-\infty$ everywhere by calculating its value at $\mathbf{\Lambda}_{1,0}(x) = \mathrm{th}(x)$ in Lemma 7.15 below. Recalling §2.6, we expect this to be the maximizer of $\mathscr{S}_\star$.

Lemma 7.14. *For any $\mathbf{\Lambda} \in \mathscr{K}$, $s \geqslant 0$, we have $\frac{\partial^2}{\partial s^2}\mathscr{S}_\star(\mathbf{\Lambda}, s) > 0$.*

Proof. Since $(\log \Psi)' = -\mathcal{E}$, we have

$$\frac{\partial^2}{\partial s^2}\mathscr{S}_\star(\mathbf{\Lambda}, s) = \psi_0 - \alpha_\star \mathbb{E}\left\{\mathcal{E}'\left(\frac{\kappa - \frac{\mathbb{E}[\boldsymbol{M}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]}{q_0}\hat{\boldsymbol{H}} - \frac{\mathbb{E}[\dot{\boldsymbol{H}}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]}{\psi_0}\boldsymbol{N}}{\sqrt{1 - \frac{\mathbb{E}[\boldsymbol{M}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]^2}{q_0}}} + s\boldsymbol{N}\right)\boldsymbol{N}^2\right\}$$

$$\overset{Lem.\ 4.21(b)}{>} \psi_0 - \alpha_\star \mathbb{E}[\boldsymbol{N}^2] = 0.$$

□

Lemma 7.15. *We have $\mathscr{S}_\star(\mathbf{\Lambda}_{1,0}) = \mathscr{S}_\star(\mathbf{\Lambda}_{1,0}, \sqrt{1-q_0}) = 0$.*

Proof. Let $\mathbf{\Lambda} = \mathbf{\Lambda}_{1,0}$. Note that $\mathbf{\Lambda}(\dot{\boldsymbol{H}}) = \mathrm{th}(\dot{\boldsymbol{H}}) = \boldsymbol{M}$. Thus $\mathbb{E}[\boldsymbol{M}\mathbf{\Lambda}(\dot{\boldsymbol{H}})] = q_0$ and, by gaussian integration by parts, $\mathbb{E}[\dot{\boldsymbol{H}}\mathbf{\Lambda}(\dot{\boldsymbol{H}})] = (1-q_0)\psi_0$. So

$$\frac{\kappa - \frac{\mathbb{E}[\boldsymbol{M}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]}{q_0}\hat{\boldsymbol{H}} - \frac{\mathbb{E}[\dot{\boldsymbol{H}}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]}{\psi_0}\boldsymbol{N}}{\sqrt{1 - \frac{\mathbb{E}[\boldsymbol{M}\mathbf{\Lambda}(\dot{\boldsymbol{H}})]^2}{q_0}}} + \sqrt{1-q_0}\boldsymbol{N} = \frac{\kappa - \hat{\boldsymbol{H}}}{\sqrt{1-q_0}}.$$

By the identity $\mathcal{H}(\frac{1+\mathrm{th}x}{2}) = \log(2\mathrm{ch}x) - x\mathrm{th}x$,

$$\mathbb{E}\mathcal{H}\left(\frac{1+\mathbf{\Lambda}}{2}\right) = \mathbb{E}\log(2\mathrm{ch}\dot{\boldsymbol{H}}) - \mathbb{E}[\dot{\boldsymbol{H}}\mathbf{\Lambda}] = \mathbb{E}\log(2\mathrm{ch}\dot{\boldsymbol{H}}) - (1-q_0)\psi_0.$$

Thus

$$\mathscr{S}_\star(\mathbf{\Lambda}, \sqrt{1-q_0}) = -\frac{1}{2}(1-q_0)\psi_0 + \mathbb{E}\log(2\mathrm{ch}\dot{\boldsymbol{H}}) + \alpha\,\mathbb{E}\log\Psi\left(\frac{\kappa - \hat{\boldsymbol{H}}}{\sqrt{1-q_0}}\right) = \mathscr{G}(\alpha_\star, q_0, \psi_0),$$

which equals 0 by definition of $\alpha_\star$. Furthermore,

$$\frac{\partial}{\partial s}\mathscr{S}_\star(\mathbf{\Lambda}, s)\big|_{s=\sqrt{1-q_0}} = \sqrt{1-q_0}\psi_0 - \alpha_\star \mathbb{E}\left\{\mathcal{E}\left(\frac{\kappa - \hat{\boldsymbol{H}}}{\sqrt{1-q_0}}\right)\boldsymbol{N}\right\}$$
$$= \sqrt{1-q_0}\left(\psi_0 - \alpha_\star \mathbb{E}[\boldsymbol{N}^2]\right) = 0.$$

By Lemma 7.14, this implies $s = \sqrt{1-q_0}$ minimizes $\mathscr{S}_\star(\mathbf{\Lambda}, s)$, and thus $\mathscr{S}_\star(\mathbf{\Lambda}) = \mathscr{S}_\star(\mathbf{\Lambda}, \sqrt{1-q_0})$. □

Proof of Proposition 7.11. The set $\overline{\mathscr{K}}_*$ is compact in the topology of $\mathscr{L}$. The functions $\mathscr{S}_\star^{s_{\max}} : \overline{\mathscr{K}}_* \to \mathbb{R}$ are continuous by Lemma 7.12 and compactness of $[0, s_{\max}]$. On any sequence of $s_{\max}$ tending to ∞, the sequence of $\mathscr{S}_\star^{s_{\max}}$ is decreasing with pointwise limit $\mathscr{S}_\star$. Since Lemma 7.15 implies $\mathscr{S}_\star$ is not $-\infty$ everywhere, the result follows from Lemma 7.13. □

7.4. **No boundary maximizers and conclusion.** The results proved so far imply that the exponential order of $\mathbb{E}^{m,n}_{\varepsilon,\mathsf{Pl}} Z_N(\boldsymbol{G})$ is bounded up to vanishing error by $\sup_{\boldsymbol{\Lambda}\in\overline{\mathscr{K}}_*} \mathscr{S}_\star(\boldsymbol{\Lambda})$. Condition 1.3 provides a bound on $\sup_{\boldsymbol{\Lambda}\in\mathscr{K}_*} \mathscr{S}_\star(\boldsymbol{\Lambda})$. Since $\mathscr{S}_\star$ (unlike $\mathscr{S}_\star^{S_{\max}}$) is not a priori continuous, to complete the proof we verify in the following proposition that it is not maximized on the boundary.

Proposition 7.16. *The maximum of $\mathscr{S}_\star(\boldsymbol{\Lambda})$ on $\overline{\mathscr{K}}_*$ (which exists by Proposition 7.11) is not attained on $\overline{\mathscr{K}}_*\backslash\mathscr{K}_*$.*

Lemma 7.17. *Let $d_0 = \alpha_\star \mathbb{E}[F'_{1-q_0}(q_0^{1/2} Z)]$, and*

$$\mathscr{O} = \left\{\boldsymbol{\Lambda}\in\mathscr{K} : d_0\,\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})] + \mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})] > \alpha_\star\kappa\right\}.$$

Then, for $\boldsymbol{\Lambda}\in\mathscr{K}$,

$$\lim_{s\to+\infty} \mathscr{S}_\star(\boldsymbol{\Lambda}, s) = \begin{cases} +\infty & \boldsymbol{\Lambda}\in\mathscr{O}, \\ -\infty & \boldsymbol{\Lambda}\notin\mathscr{O}. \end{cases}$$

Proof. A well-known gaussian tail bound gives $\frac{\varphi(x)}{x} < \Psi(x) < \frac{x\varphi(x)}{1+x^2}$ for all $x > 0$. Thus, for large x,

$$\log\Psi(x) = -\frac{1}{2}x^2 - \log x + O(1). \tag{84}$$

Let s be large and define

$$\xi(x) = -\frac{1}{2}x^2 - \mathbf{1}\{s^{1/2}\leqslant x\leqslant s^2\}\log x, \quad \boldsymbol{U} = \frac{\kappa - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})]}{q_0}\hat{\boldsymbol{H}} - \frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})]}{\psi_0}\boldsymbol{N}}{\sqrt{1-\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})]^2}{q_0}}}, \quad \boldsymbol{V} = \boldsymbol{U} + s\boldsymbol{N}.$$

Note that

$$\begin{aligned} |\mathbb{E}\log\Psi(\boldsymbol{V}) - \mathbb{E}\,\xi(\boldsymbol{V})| &\leqslant |\mathbb{E}\,\mathbf{1}\{\boldsymbol{V}\leqslant\log\log s\}(\log\Psi(\boldsymbol{V}) - \xi(\boldsymbol{V}))| \\ &\quad + \left|\mathbb{E}\,\mathbf{1}\{\log\log s\leqslant\boldsymbol{V}\leqslant s^{1/2}\}(\log\Psi(\boldsymbol{V}) - \xi(\boldsymbol{V}))\right| \\ &\quad + \left|\mathbb{E}\,\mathbf{1}\{s^{1/2}\leqslant\boldsymbol{V}\leqslant s^2\}(\log\Psi(\boldsymbol{V}) - \xi(\boldsymbol{V}))\right| \\ &\quad + \left|\mathbb{E}\,\mathbf{1}\{\boldsymbol{V}\geqslant s^2\}(\log\Psi(\boldsymbol{V}) - \xi(\boldsymbol{V}))\right|. \end{aligned}$$

We will show each of these terms is $o(\log s)$. Let $\boldsymbol{V}_+ = \max(\boldsymbol{V}, 0)$, $\boldsymbol{V}_- = -\min(\boldsymbol{V}, 0)$, and let $C > 0$ be a constant varying from line to line. Then,

$$\begin{aligned} &|\mathbb{E}\,\mathbf{1}\{\boldsymbol{V}\leqslant\log\log s\}(\log\Psi(\boldsymbol{V}) - \xi(\boldsymbol{V}))| \\ &\leqslant \mathbb{E}\,\mathbf{1}\{\boldsymbol{V}\leqslant\log\log s\}|\log\Psi(\boldsymbol{V})| + \mathbb{E}\,\mathbf{1}\{\boldsymbol{V}\leqslant\log\log s\}\boldsymbol{V}_+^2 + \mathbb{E}\,\boldsymbol{V}_-^2 \\ &\leqslant C(\log\log s)^2 + \mathbb{E}\,\boldsymbol{U}_-^2 \leqslant C(\log\log s)^2. \end{aligned}$$

In the last line we used that $\boldsymbol{N} > 0$ almost surely, and thus $\boldsymbol{U}_- \geqslant \boldsymbol{V}_-$. By the estimate (84), if $\log\log s\leqslant \boldsymbol{V} < s^{1/2}$, then $|\log\Psi(\boldsymbol{V}) - \xi(\boldsymbol{V})| \leqslant C\log s$. Thus

$$\begin{aligned} \left|\mathbb{E}\,\mathbf{1}\{\log\log s\leqslant\boldsymbol{V} < s^{1/2}\}(\log\Psi(\boldsymbol{V}) - \xi(\boldsymbol{V}))\right| &\leqslant (C\log s)\,\mathbb{P}(\boldsymbol{V}\leqslant s^{1/2}) \\ &\leqslant (C\log s)\left(\mathbb{P}(\boldsymbol{U}\leqslant -s^{1/2}) + \mathbb{P}(s\boldsymbol{N}\leqslant 2s^{1/2})\right) = o(\log s). \end{aligned}$$

The estimate (84) directly implies

$$\left|\mathbb{E}\,\mathbf{1}\{s^{1/2}\leqslant\boldsymbol{V}\leqslant s^2\}(\log\Psi(\boldsymbol{V}) - \xi(\boldsymbol{V}))\right| = O(1).$$

Finally, Lemma 4.21(a) gives $0 \leqslant \mathcal{E}(x) \leqslant |x| + 1$. Thus

$$|\boldsymbol{V}| \leqslant |\boldsymbol{U}| + \frac{s}{\sqrt{1-q_0}}\mathcal{E}\left(\frac{\kappa - \hat{\boldsymbol{H}}}{\sqrt{1-q_0}}\right) \leqslant Cs(|\hat{\boldsymbol{H}}|+1).$$

It follows that for $t \geqslant s^2$, we have $\mathbb{P}(|\boldsymbol{V}| \geqslant t) \leqslant \exp(-t^2/Cs^2)$. So, crudely

$$\begin{aligned}\left|\mathbb{E}\,\mathbf{1}\{\boldsymbol{V} \geqslant s^2\}(\log\Psi(\boldsymbol{V}) - \xi(\boldsymbol{V}))\right| &\leqslant C'\,\mathbb{E}\,\mathbf{1}\{\boldsymbol{V} \geqslant s^2\}\boldsymbol{V}^2 \\ &\leqslant C'\left(s^2\exp(-s^2/C) + \int_{s^2}^{\infty} 2t\exp(-t^2/Cs^2)\,\mathrm{d}t\right) \\ &\leqslant C's^2\exp(-s^2/C).\end{aligned}$$

Thus $|\mathbb{E}\log\Psi(\boldsymbol{V}) - \mathbb{E}\,\xi(\boldsymbol{V})| = o(\log s)$. So,

$$\mathcal{S}_\star(\boldsymbol{\Lambda}, s) = \frac{1}{2}s^2\psi_0 + \alpha_\star\,\mathbb{E}\,\xi(\boldsymbol{V}) + o(\log s).$$

We now evaluate $\alpha_\star\,\mathbb{E}\,\xi(\boldsymbol{V})$. First,

$$\begin{aligned}\frac{1}{2}\alpha_\star\,\mathbb{E}\,\boldsymbol{V}^2 &= \frac{1}{2}\alpha_\star s^2\,\mathbb{E}[\boldsymbol{N}^2] + \alpha_\star s\,\mathbb{E}[\boldsymbol{U}\boldsymbol{N}] + O(1) \\ &= \frac{1}{2}s^2\psi_0 + \frac{s\left(\alpha_\star\kappa - d_0\,\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})] - \mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})]\right)}{\sqrt{1 - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})]^2}{q_0}}} + O(1).\end{aligned}$$

Thus

$$\mathcal{S}_\star(\boldsymbol{\Lambda}, s) = \frac{s\left(d_0\,\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})] + \mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})] - \alpha_\star\kappa\right)}{\sqrt{1 - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}(\dot{\boldsymbol{H}})]^2}{q_0}}} - \mathbb{E}\,\mathbf{1}\{s^{1/2} \leqslant \boldsymbol{V} \leqslant s^2\}\log\boldsymbol{V} + o(\log s).$$

The logarithmic term clearly has magnitude $O(\log s)$. So, $\lim_{s\to+\infty}\mathcal{S}_\star(\boldsymbol{\Lambda}, s) = +\infty$ if $\boldsymbol{\Lambda} \in \mathcal{O}$, and $-\infty$ if $\boldsymbol{\Lambda}$ is in the interior of $\mathcal{K}\backslash\mathcal{O}$. Finally, we have shown above that $\mathbb{P}(\boldsymbol{V} < s^{1/2}), \mathbb{P}(\boldsymbol{V} > s^2) = o_s(1)$, so

$$\mathbb{E}\,\mathbf{1}\{s^{1/2} \leqslant \boldsymbol{V} \leqslant s^2\}\log\boldsymbol{V} \geqslant \frac{1}{2}(1 - o_s(1))\log s.$$

Thus $\lim_{s\to+\infty}\mathcal{S}_\star(\boldsymbol{\Lambda}, s) = -\infty$ for $\boldsymbol{\Lambda}$ on the boundary of $\mathcal{K}\backslash\mathcal{O}$. □

Proof of Proposition 7.16. Suppose for contradiction that $\boldsymbol{\Lambda} \in \overline{\mathcal{K}}_*\backslash\mathcal{K}_*$ maximizes $\mathcal{S}_\star(\boldsymbol{\Lambda})$ in $\overline{\mathcal{K}}_*$. By Proposition 7.9, $\boldsymbol{\Lambda}$ is also a maximizer of $\mathcal{S}_\star(\boldsymbol{\Lambda})$ in $\mathcal{K}$.

By Lemma 7.17, if $\boldsymbol{\Lambda} \notin \mathcal{O}$, then $\mathcal{S}_\star(\boldsymbol{\Lambda}) = -\infty$ is not a maximizer (recall Lemma 7.15). Thus $\boldsymbol{\Lambda} \in \mathcal{O}$. Let $\boldsymbol{\Lambda}^t = (1-t)\boldsymbol{\Lambda}$. Since $\mathcal{O}$ is open, $\boldsymbol{\Lambda}^t \in \mathcal{O}$ for $t \in [0, t_+)$, for sufficiently small t_+.

By Lemma 7.17, for $t \in [0, t_+)$, the infimum of $\mathcal{S}(\boldsymbol{\Lambda}^t, s)$ is attained at some $s(\boldsymbol{\Lambda}^t) \in [0, +\infty)$. Note that

$$\frac{\partial}{\partial s}\mathcal{S}_\star(\boldsymbol{\Lambda}^t, s)\bigg|_{s=0} = -\alpha_\star\,\mathbb{E}\left\{\mathcal{E}\left(\frac{\kappa - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}^t(\dot{\boldsymbol{H}})]}{q_0}\hat{\boldsymbol{H}} - \frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}^t(\dot{\boldsymbol{H}})]}{\psi_0}\boldsymbol{N}}{\sqrt{1 - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}^t(\dot{\boldsymbol{H}})]^2}{q_0}}} + s\boldsymbol{N}\right)\boldsymbol{N}\right\} < 0$$

because $\boldsymbol{N} > 0$ almost surely and the image of $\mathcal{E}$ is positive. Combined with Lemma 7.14, this implies $s(\boldsymbol{\Lambda}^t)$ is the unique solution to $\frac{\partial}{\partial s}\mathcal{S}_\star(\boldsymbol{\Lambda}, s) = 0$, and $s(\boldsymbol{\Lambda}^t) > 0$.

Note that $\frac{\partial}{\partial s}\mathscr{S}_\star(\mathbf{\Lambda}^t, s)$ is differentiable in t, as the denominator $\sqrt{1 - \frac{\mathbb{E}[M\Lambda^t(\dot{H})]^2}{q_0}}$ is bounded away from 0 by Lemma 7.2. By Lemma 7.14 and the implicit function theorem, $s(\mathbf{\Lambda}^t)$ is differentiable in t for all $t \in [0, t_+)$. It follows that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\{\frac{1}{2}s(\mathbf{\Lambda}^t)^2\psi_0 + \alpha_\star \mathbb{E}\log\Psi\left(\frac{\kappa - \frac{\mathbb{E}[M\Lambda^t(\dot{H})]}{q_0}\hat{\boldsymbol{H}} - \frac{\mathbb{E}[\dot{H}\Lambda^t(\dot{H})]}{\psi_0}\boldsymbol{N}}{\sqrt{1 - \frac{\mathbb{E}[M\Lambda^t(\dot{H})]^2}{q_0}}} + s(\mathbf{\Lambda}^t)\boldsymbol{N}\right)\right\}\Bigg|_{t=0}$$

exists and is finite. However, since $\mathbf{\Lambda} \in \overline{\mathscr{K}}_* \backslash \mathscr{K}_*$, we have $\mathbf{\Lambda}(\dot{\boldsymbol{H}}) \in \{-1, 1\}$ $\dot{\boldsymbol{H}}$-almost surely. Thus

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{ent}(\mathbf{\Lambda}^t)\big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(t/2)\big|_{t=0} = +\infty.$$

Hence $\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{S}_\star(\mathbf{\Lambda}^t)\big|_{t=0} = +\infty$, and $\mathbf{\Lambda}$ is not a maximizer of $\mathscr{S}_\star(\mathbf{\Lambda})$ in $\mathscr{K}$. □

Proof of Proposition 3.9. By Propositions 7.1, 7.9, for any $s_{\max} > 0$,

$$\frac{1}{N}\log\mathbb{E}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}[Z_N(\boldsymbol{G})] \leqslant \sup_{\mathbf{\Lambda}\in\mathscr{K}} \mathscr{S}_\star^{s_{\max}}(\mathbf{\Lambda}) + o_{\varepsilon,v}(1) = \sup_{\mathbf{\Lambda}\in\overline{\mathscr{K}}_*} \mathscr{S}_\star^{s_{\max}}(\mathbf{\Lambda}) + o_{\varepsilon,v}(1). \tag{85}$$

By Propositions 7.11 and 7.16 and Condition 1.3,

$$\lim_{s_{\max}\to\infty}\sup_{\mathbf{\Lambda}\in\overline{\mathscr{K}}_*}\mathscr{S}_\star^{s_{\max}}(\mathbf{\Lambda}) = \sup_{\mathbf{\Lambda}\in\overline{\mathscr{K}}_*}\mathscr{S}_\star(\mathbf{\Lambda}) = \sup_{\mathbf{\Lambda}\in\mathscr{K}_*}\mathscr{S}_\star(\mathbf{\Lambda}) = \sup_{\lambda_1,\lambda_2\in\mathbb{R}}\mathscr{S}_\star(\lambda_1,\lambda_2) \leqslant 0.$$

Thus, taking the limit $\varepsilon, v \to 0$ followed by $s_{\max} \to \infty$ in (85) implies the result. □

7.5. **Local analysis of first moment functional at** $(1, 0)$**.** We now prove Lemma 2.5. Note that part (a) follows from Proposition 7.16, and part (b) was already proved in Lemma 7.15. We turn to the proofs of the remaining parts.

Proof of Lemma 2.5(c). Let $\mathscr{S}_\star(\lambda_1, \lambda_2, s) = \mathscr{S}_\star(\mathbf{\Lambda}_{\lambda_1,\lambda_2}, s)$, and let $s(\lambda_1, \lambda_2)$ minimize $\mathscr{S}_\star(\lambda_1, \lambda_2, s)$. Lemma 7.15 shows $s(1, 0) = \sqrt{1 - q_0}$, and the proof of Proposition 7.16 shows that for (λ_1, λ_2) in a neighborhood of $(1, 0)$, $s(\lambda_1, \lambda_2)$ is the unique solution to $\partial_s\mathscr{S}_\star(\lambda_1, \lambda_2, s) = 0$. By Lemma 7.14 and the implicit function theorem, $s(\lambda_1, \lambda_2)$ is differentiable in this neighborhood. So,

$$\begin{aligned}\nabla\mathscr{S}_\star(\lambda_1, \lambda_2) &= \nabla_{\lambda_1,\lambda_2}\mathscr{S}_\star(\lambda_1, \lambda_2, s(\lambda_1, \lambda_2)) + \partial_s\mathscr{S}_\star(\lambda_1, \lambda_2, s(\lambda_1, \lambda_2))\nabla s(\lambda_1, \lambda_2) \\ &= \nabla_{\lambda_1,\lambda_2}\mathscr{S}_\star(\lambda_1, \lambda_2, s(\lambda_1, \lambda_2)),\end{aligned} \tag{86}$$

and in particular $\nabla\mathscr{S}_\star(1, 0) = \nabla\overline{\mathscr{S}}_\star(1, 0)$. To calculate the latter gradient, let $u_1, u_2 \in \mathbb{R}$ be arbitrary and

$$\mathbf{\Delta} \equiv (u_1\partial_{\lambda_1} + u_2\partial_{\lambda_2})\mathbf{\Lambda} = (1 - \mathbf{\Lambda}^2)(u_1\dot{\boldsymbol{H}} + u_2\boldsymbol{M}).$$

Then

$$\begin{aligned}\langle\nabla\overline{\mathscr{S}}_\star(\lambda_1, \lambda_2), (u_1, u_2)\rangle = -\mathbb{E}[\mathrm{th}^{-1}(\mathbf{\Lambda})\mathbf{\Delta}] - \alpha_\star\mathbb{E}\Bigg\{\mathcal{E}\left(\frac{\kappa - \frac{\mathbb{E}[M\Lambda]}{q_0}\hat{\boldsymbol{H}} - \frac{\mathbb{E}[\dot{H}\Lambda]}{\psi_0}\boldsymbol{N}}{\sqrt{1 - \frac{\mathbb{E}[M\Lambda]^2}{q_0}}} + \sqrt{1 - q_0}\boldsymbol{N}\right) \\ \times\left(\frac{-\frac{\mathbb{E}[M\Delta]}{q_0}\hat{\boldsymbol{H}} - \frac{\mathbb{E}[\dot{H}\Delta]}{\psi_0}\boldsymbol{N}}{\sqrt{1 - \frac{\mathbb{E}[M\Lambda]^2}{q_0}}} + \frac{\kappa - \frac{\mathbb{E}[M\Lambda]}{q_0}\hat{\boldsymbol{H}} - \frac{\mathbb{E}[\dot{H}\Lambda]}{\psi_0}\boldsymbol{N}}{\left(1 - \frac{\mathbb{E}[M\Lambda]^2}{q_0}\right)^{3/2}} \cdot \frac{\mathbb{E}[M\Lambda]\,\mathbb{E}[M\Delta]}{q_0}\right)\Bigg\}.\end{aligned} \tag{87}$$

Specializing to $(\lambda_1, \lambda_2) = (1, 0)$,

$$\langle \nabla \overline{\mathcal{S}}_\star(1,0), (u_1, u_2) \rangle$$

$$= -\mathbb{E}[\text{th}^{-1}(\boldsymbol{M})\boldsymbol{\Delta}] - \alpha_\star \mathbb{E}\left\{ \mathcal{E}\left(\frac{\kappa - \hat{\boldsymbol{H}}}{\sqrt{1-q_0}} \right) \left(\frac{-\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]}{q_0}\hat{\boldsymbol{H}} - \frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Delta}]}{\psi_0}\boldsymbol{N}}{\sqrt{1-q_0}} + \frac{\kappa - \hat{\boldsymbol{H}} - (1-q_0)\boldsymbol{N}}{(1-q_0)^{3/2}} \mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}] \right) \right\}$$

$$= -\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Delta}] - \alpha_\star \mathbb{E}\left\{ F_{1-q_0}(\hat{\boldsymbol{H}}) \left(-\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]}{q_0}\hat{\boldsymbol{H}} - \frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Delta}]}{\psi_0}\boldsymbol{N} + \frac{\kappa - \hat{\boldsymbol{H}} - (1-q_0)\boldsymbol{N}}{1-q_0} \mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}] \right) \right\}$$

$$= -\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Delta}] + \frac{\alpha_\star \mathbb{E}[\boldsymbol{N}^2]}{\psi_0} \mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Delta}] + \alpha_\star \left(\frac{\mathbb{E}[\boldsymbol{N}\hat{\boldsymbol{H}}]}{q_0} + \mathbb{E}\left[\boldsymbol{N} \left(\boldsymbol{N} - \frac{\kappa - \hat{\boldsymbol{H}}}{1-q_0} \right) \right] \right) \mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}].$$

The first two terms cancel because $\alpha_\star \mathbb{E}[\boldsymbol{N}^2] = \psi_0$. Finally, note the identity

$$F'_{1-q_0}(x) = -F_{1-q_0}(x) \left(F_{1-q_0}(x) - \frac{x}{1-q_0} \right).$$

By gaussian integration by parts,

$$\mathbb{E}[\boldsymbol{N}\hat{\boldsymbol{H}}] = \mathbb{E}[\hat{\boldsymbol{H}} F_{1-q_0}(\hat{\boldsymbol{H}})] = \mathbb{E}[\hat{\boldsymbol{H}}^2] \mathbb{E}[F'_{1-q_0}(\hat{\boldsymbol{H}})] = -q_0 \mathbb{E}\left[\boldsymbol{N} \left(\boldsymbol{N} - \frac{\kappa - \hat{\boldsymbol{H}}}{1-q_0} \right) \right].$$

It follows that $\langle \nabla \mathcal{S}_\star(1,0), (u_1, u_2) \rangle = 0$. Since u_1, u_2 were arbitrary, $\nabla \mathcal{S}_\star(1,0) = 0$. □

Proof of Lemma 2.5(d). Differentiating (86) and applying the implicit function theorem yields

$$\begin{aligned}
\nabla^2 \mathcal{S}_\star(\lambda_1, \lambda_2) &= \nabla^2_{\lambda_1,\lambda_2} \mathcal{S}_\star(\lambda_1, \lambda_2, s(\lambda_1, \lambda_2)) + \nabla_{\lambda_1,\lambda_2} \partial_s \mathcal{S}_\star(\lambda_1, \lambda_2, s(\lambda_1, \lambda_2)) (\nabla s(\lambda_1, \lambda_2))^\top \\
&= \nabla^2_{\lambda_1,\lambda_2} \mathcal{S}_\star(\lambda_1, \lambda_2, s(\lambda_1, \lambda_2)) - \frac{(\nabla_{\lambda_1,\lambda_2} \partial_s \mathcal{S}_\star(\lambda_1, \lambda_2, s(\lambda_1, \lambda_2)))^{\otimes 2}}{\partial_s^2 \mathcal{S}_\star(\lambda_1, \lambda_2, s(\lambda_1, \lambda_2))} \\
&\preceq \nabla^2_{\lambda_1,\lambda_2} \mathcal{S}_\star(\lambda_1, \lambda_2, s(\lambda_1, \lambda_2)).
\end{aligned}$$

Specializing to $(\lambda_1, \lambda_2) = (1, 0)$ yields the result. □

References

[AB13] Antonio Auffinger and Gérard Ben Arous. Complexity of random smooth functions on the high-dimensional sphere. *Ann. Probab.*, 41(6):4214–4247, 2013.

[ABČ13] Antonio Auffinger, Gérard Ben Arous, and Jiří Černỳ. Random matrices and complexity of spin glasses. *Comm. Pure Appl. Math.*, 66(2):165–201, 2013.

[ABL22] Antonio Auffinger, Gérard Ben Arous, and Zhehua Li. Sharp complexity asymptotics and topological trivialization for the (p, k) spiked tensor model. *J. Math. Phys.*, 63(4):043303, 2022.

[AC08] Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *Proc. 49th FOCS*, pages 793–802. IEEE, 2008.

[AG20] Antonio Auffinger and Julian Gold. The number of saddles of the spherical p-spin model. *arXiv preprint arXiv:2007.09269*, 2020.

[ALS22a] Emmanuel Abbe, Shuangping Li, and Allan Sly. Binary perceptron: efficient algorithms can find solutions in a rare well-connected cluster. In *Proc. 54th STOC*, pages 860–873, 2022.

[ALS22b] Emmanuel Abbe, Shuangping Li, and Allan Sly. Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron. In *Proc. 62nd FOCS*, pages 327–338. IEEE, 2022.

[Alt23] Dylan J Altschuler. Critical window of the symmetric perceptron. *Electron. J. Probab.*, 28:1–28, 2023.

[AMS22] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from the Sherrington-Kirkpatrick Gibbs measure via algorithmic stochastic localization. In *Proc. 63rd FOCS*, pages 323–334. IEEE, 2022.

[AMS23a] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from mean-field Gibbs measures via diffusion processes. *arXiv preprint arXiv:2310.08912*, 2023.

[AMS23b] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Shattering in pure spherical spin glasses. *arXiv preprint arXiv:2307.04659*, 2023.

[APZ19] Benjamin Aubin, Will Perkins, and Lenka Zdeborová. Storage capacity in symmetric binary perceptrons. *J. Phys. A*, 52(29):294003, 2019.

[AS22] Ahmed El Alaoui and Mark Sellke. Algorithmic pure states for the negative spherical perceptron. *J. Stat. Phys.*, 189(2):27, 2022.

[AT09] Robert J Adler and Jonathan E Taylor. *Random Fields and Geometry*. Springer Science & Business Media, 2009.

[AT24] Dylan J Altschuler and Konstantin Tikhomirov. A note on the capacity of the binary perceptron. *arXiv preprint arXiv:2401.15092*, 2024.

[BAKZ23] Damien Barbier, Ahmed El Alaoui, Florent Krzakala, and Lenka Zdeborová. On the atypical solutions of the symmetric binary perceptron. *arXiv preprint arXiv:2310.02850*, 2023.

[BBM22] Gérard Ben Arous, Paul Bourgade, and Benjamin McKenna. Exponential growth of random determinants beyond invariance. *Probab. Math. Phys.*, 3(4):731–789, 2022.

[BBM24] Gérard Ben Arous, Paul Bourgade, and Benjamin McKenna. Landscape complexity beyond invariance and the elastic manifold. *Comm. Pure Appl. Math.*, 77(2):1302–1352, 2024.

[BC16] Victor Bapst and Amin Coja-Oghlan. The condensation phase transition in the regular k-SAT model. *Approx. Random. Combin. Optim. Algor. Techniques*, 2016.

[BCH$^+$16] Victor Bapst, Amin Coja-Oghlan, Samuel Hetterich, Felicia Raßmann, and Dan Vilenchik. The condensation phase transition in random graph coloring. *Comm. Math. Phys.*, 341:543–606, 2016.

[BČNS22] David Belius, Jiří Černỳ, Shuta Nakajima, and Marius A Schmidt. Triviality of the geometry of mixed p-spin spherical Hamiltonians with external field. *J. Stat. Phys.*, 186(1):12, 2022.

[BM80] Alan J Bray and Michael A Moore. Metastable states in spin glasses. *J. Phys. C*, 13(19):L469, 1980.

[BM11] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inform. Theory*, 57(2):764–785, 2011.

[BMMN19] Gérard Ben Arous, Song Mei, Andrea Montanari, and Mihai Nica. The landscape of the spiked tensor model. *Comm. Pure Appl. Math.*, 72(11):2282–2330, 2019.

[BMN20] Raphael Berthier, Andrea Montanari, and Phan-Minh Nguyen. State evolution for approximate message passing with non-separable functions. *Inform. Inference*, 9(1):33–79, 2020.

[BNSX22] Erwin Bolthausen, Shuta Nakajima, Nike Sun, and Changji Xu. Gardner formula for Ising perceptron models at small densities. In *Proc. 35th COLT*, pages 1787–1911. PMLR, 2022.

[Bol14] Erwin Bolthausen. An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Comm. Math. Phys.*, 325(1):333–366, 2014.

[Bol19] Erwin Bolthausen. A Morita type proof of the replica-symmetric formula for SK. *Stat. Mech. Class. Disord. Syst.*, page 63, 2019.

[BS98] Zhi-Dong Bai and Jack W Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.*, 26(1):316–345, 1998.

[BSZ20] Gérard Ben Arous, Eliran Subag, and Ofer Zeitouni. Geometry and temperature chaos in mixed spherical spin glasses at low temperature: the perturbative regime. *Comm. Pure Appl. Math.*, 73(8):1732–1828, 2020.

[BY22] Christian Brennecke and Horng-Tzer Yau. The replica symmetric formula for the SK model revisited. *J. Math. Phys.*, 63(7), 2022.

[CEJ$^+$18] Amin Coja-Oghlan, Charilaos Efthymiou, Nor Jaafari, Mihyun Kang, and Tobias Kapetanopoulos. Charting the replica symmetric phase. *Comm. Math. Phys.*, 359:603–698, 2018.

[Cel24] Michael Celentano. Sudakov-Fernique post-AMP, and a new proof of the local convexity of the TAP free energy. *Ann. Probab.*, 2024.

[CFLM23] Michael Celentano, Zhou Fan, Licong Lin, and Song Mei. Mean-field variational inference with the TAP free energy: Geometric and statistical properties in linear models. *arXiv preprint arXiv:2311.08442*, 2023.

[CFM23] Michael Celentano, Zhou Fan, and Song Mei. Local convexity of the TAP free energy and AMP convergence for $\mathbb{Z}_2$-synchronization. *Ann. Stat.*, 51(2):519–546, 2023.

[CKM20] Amin Coja-Oghlan, Tobias Kapetanopoulos, and Noela Müller. The replica symmetric phase of random constraint satisfaction problems. *Combin. Probab. Comp.*, 29(3):346–422, 2020.

[CKPZ17] Amin Coja-Oghlan, Florent Krzakala, Will Perkins, and Lenka Zdeborová. Information-theoretic thresholds from the cavity method. In *Proc. 49th STOC*, pages 146–157, 2017.

[CLR05] Andrea Crisanti, Luca Leuzzi, and Tommaso Rizzo. Complexity in mean-field spin-glass models: Ising p-spin. *Phys. Rev. B*, 71(9):094202, 2005.

[Cov65] Thomas M Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Electron. Comp.*, 3(14):326–334, 1965.

[DS18] Jian Ding and Nike Sun. Capacity lower bound for the Ising perceptron. *arXiv preprint arXiv:1809.07742*, 2018.

[Eld20] Ronen Eldan. Taming correlations through entropy-efficient measure decompositions with applications to mean-field approximation. *Probab. Theory Rel. Fields*, 176(3-4):737–755, 2020.

[Fer75] Xavier Fernique. Regularite des trajectoires des fonctions aleatoires gaussiennes. *Ecole d'Eté Probab. Saint-Flour IV-1974*, pages 1–96, 1975.

[FL14] Yan V Fyodorov and Pierre Le Doussal. Topology trivialization and large deviations for the minimum in the simplest random optimization. *J. Stat. Phys.*, 154(1-2):466–490, 2014.

[FMM21] Zhou Fan, Song Mei, and Andrea Montanari. TAP free energy, spin glasses and variational inference. *Ann. Probab.*, 49(1), 2021.

[Fyo04] Yan V Fyodorov. Complexity of random energy landscapes, glass transition, and absolute value of the spectral determinant of random matrices. *Phys. Rev. Lett.*, 92(24):240601, 2004.

[Fyo15] Yan V Fyodorov. High-dimensional random fields and random matrix theory. *Markov Proc. and Rel. Fields*, 21(3):483–518, 2015.

[Gar87] Elizabeth Gardner. Maximum storage capacity in neural networks. *Europhys. Lett.*, 4(4):481, 1987.

[Gar88] Elizabeth Gardner. The space of interactions in neural network models. *J. Phys. A*, 21(1):257, 1988.

[GD88] Elizabeth Gardner and Bernard Derrida. Optimal storage properties of neural network models. *J. Phys. A*, 21(1):271, 1988.

[GKPX22] David Gamarnik, Eren C Kızıldağ, Will Perkins, and Changji Xu. Algorithms and barriers in the symmetric binary perceptron model. In *Proc. 63rd FOCS*, pages 576–587. IEEE, 2022.

[GKPX23] David Gamarnik, Eren C Kızıldağ, Will Perkins, and Changji Xu. Geometric barriers for stable and online algorithms for discrepancy minimization. In *Proc. 36th COLT*, pages 3231–3263. PMLR, 2023.

[Gor85] Yehoram Gordon. Some inequalities for Gaussian processes and applications. *Israel J. Math.*, 50:265–289, 1985.

[Gor88] Yehoram Gordon. On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$. In *Geomet. Aspects Func. Anal.: Israel Sem. 1986–87*, pages 84–106. Springer, 1988.

[GZ00] Alice Guionnet and Ofer Zeitouni. Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.*, 5:119–136, 2000.

[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.

[HS23a] Brice Huang and Mark Sellke. A constructive proof of the spherical Parisi formula. *arXiv preprint arXiv:2311.15495*, 2023.

[HS23b] Brice Huang and Mark Sellke. Strong topological trivialization of multi-species spherical spin glasses. *arXiv preprint arXiv:2308.09677*, 2023.

[JM13] Adel Javanmard and Andrea Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inform. Inference*, 2(2):115–144, 2013.

[Kac48] Mark Kac. On the average number of real roots of a random algebraic equation (ii). *Proc. London Math. Soc.*, 2(1):390–408, 1948.

[Kiv23] Pax Kivimae. The ground state energy and concentration of complexity in spherical bipartite models. *Comm. Math. Phys.*, 403(1):37–81, 2023.

[KM89] Werner Krauth and Marc Mézard. Storage capacity of memory networks with binary couplings. *J. Physique*, 50(20):3057–3066, 1989.

[KR98] Jeong Han Kim and James R Roche. Covering cubes by random half cubes, with applications to binary neural networks. *J. Comput. Syst. Sci.*, 56(2):223–252, 1998.

[MBB20] Antoine Maillard, Gérard Ben Arous, and Giulio Biroli. Landscape complexity for the empirical risk of generalized linear models. In *Math. Sci. Mach. Learning*, pages 287–327. PMLR, 2020.

[McK24] Benjamin McKenna. Complexity of bipartite spherical spin glasses. *Ann. l'Inst. Henri Poincaré*, 2024.

[Méz89] Marc Mézard. The space of interactions in neural networks: Gardner's computation with the cavity method. *J. Phys. A*, 22(12):2181, 1989.

[Méz17] Marc Mézard. Mean-field message-passing equations in the Hopfield model and its generalizations. *Phys. Rev. E*, 95(2):022117, 2017.

[MM09] Marc Mézard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, 2009.

[MPV87] Marc Mézard, Giorgio Parisi, and Miguel A Virasoro. *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.

[MZZ24] Andrea Montanari, Yiqiao Zhong, and Kangjie Zhou. Tractability from overparametrization: the example of the negative perceptron. *Probab. Theory Rel. Fields*, 2024.

[NS23] Shuta Nakajima and Nike Sun. Sharp threshold sequence and universality for ising perceptron models. In *Proc. 34th SODA*, pages 638–674. SIAM, 2023.

[NV14] Hoi H Nguyen and Van Vu. Random matrices: law of the determinant. *Ann. Probab.*, 42(1):146–167, 2014.

[PP95] Giorgio Parisi and Marc Potters. Mean-field equations for spin models with orthogonal interaction matrices. *J. Phys. A*, 28(18):5267, 1995.

[PX21] Will Perkins and Changji Xu. Frozen 1-RSB structure of the symmetric Ising perceptron. In *Proc. 53rd STOC*, pages 1579–1588, 2021.

[Ric44] Stephen O Rice. Mathematical analysis of random noise. *Bell Syst. Tech. J.*, 23(3):282–332, 1944.

[Rud76] Walter Rudin. *Principles of Mathematical Analysis*, volume 3. McGraw-Hill New York, 1976.

[SS23] Ashwin Sah and Mehtaab Sawhney. Distribution of the threshold for the symmetric perceptron. In *Proc. 64th FOCS*, 2023.

[SSZ22] Allan Sly, Nike Sun, and Yumeng Zhang. The number of solutions for random regular NAE-SAT. *Probab. Theory Rel. Fields*, pages 1–109, 2022.

[ST03] Mariya Shcherbina and Brunello Tirozzi. Rigorous solution of the Gardner problem. *Comm. Math. Phys.*, 234:383–422, 2003.

[Sto13a] Mihailo Stojnic. Another look at the Gardner problem. *arXiv preprint arXiv:1306.3979*, 2013.

[Sto13b] Mihailo Stojnic. Negative spherical perceptron. *arXiv preprint arXiv:1306.3980*, 2013.

[Sub17] Eliran Subag. The complexity of spherical p-spin models – a second moment approach. *Ann. Probab.*, 45(5):3385–3450, 2017.

[Sud71] Vladimir N Sudakov. Gaussian random processes and measures of solid angles in Hilbert space. *Dokl. Akad. Nauk SSSR*, 197(1):43–45, 1971.

[Sud79] Vladimir N Sudakov. *Geometric Problems in the Theory of Infinite-Dimensional Probability Distributions*. Number 141. American Mathematical Soc., 1979.

[SZ21] Eliran Subag and Ofer Zeitouni. Concentration of the complexity of spherical pure p-spin models at arbitrary energies. *J. Math. Phys.*, 62(12):123301, 2021.

[Tal99] Michel Talagrand. Intersecting random half cubes. *Random Struct. Algor.*, 15(3-4):436–449, 1999.

[Tal00] Michel Talagrand. Intersecting random half-spaces: toward the Gardner-Derrida formula. *Ann. Probab.*, 28(2):725–758, 2000.

[TAP77] David J Thouless, Philip W Anderson, and Robert G Palmer. Solution of 'solvable model of a spin glass'. *Philos. Mag.*, 35(3):593–601, 1977.

[TV12] Terence Tao and Van Vu. A central limit theorem for the determinant of a Wigner matrix. *Adv. Math.*, 231(1):74–101, 2012.

[Voi91] Dan Voiculescu. Limit laws for random matrices and free products. *Invent. Math.*, 104(1):201–220, 1991.

[Wen62] James G Wendel. A problem in geometric probability. *Math. Scandinavica*, 11(1):109–111, 1962.

[Xu21] Changji Xu. Sharp threshold for the Ising perceptron model. *Ann. Probab.*, 49(5):2399–2415, 2021.

Appendix A. Deferred proofs

In this appendix, we provide proofs of various results deferred from the paper.

A.1. Well definedness and $\varepsilon \downarrow 0$ limit of $(q_\varepsilon, \psi_\varepsilon, \varrho_\varepsilon)$.

Proof of Proposition 4.1. Let ι_0 be small enough that $[q_0 - 3\iota_0, q_0 + 3\iota_0] \subseteq [0,1]$. Note that $\zeta_0(\psi) = (R_{\alpha_\star} \circ P)(\psi)$. By Condition 3.1, $\zeta_0(\psi_0) = \psi_0$ and

$$\zeta_0'(\psi_0) = R_{\alpha_\star}'(q_0) P'(\psi_0) = (P \circ R_{\alpha_\star})'(q_0) < 1.$$

By continuity of ζ_0 and ζ_0', we can find $\iota > 0$ such that for all $\psi \in [\psi_0 - \iota, \psi_0 + \iota]$, $P(\psi) \in [q_0 - \iota_0, q_0 + \iota_0]$ and $\zeta_0'(\psi) < 1$. Set ι_1 small enough that

$$\zeta_0(\psi_0 - \iota) \geqslant \psi_0 - \iota + 2\iota_1, \qquad \zeta_0(\psi_0 + \iota) \leqslant \psi_0 + \iota - 2\iota_1, \qquad \sup_{\psi \in [\psi_0 - \iota, \psi_0 + \iota]} \zeta_0'(\psi) \leqslant 1 - 2\iota_1.$$

We will show that for sufficiently small ε,

$$\sup_{\psi \in [\psi_0 - \iota, \psi_0 + \iota]} |\zeta_\varepsilon(\psi) - \zeta_0(\psi)|, \sup_{\psi \in [\psi_0 - \iota, \psi_0 + \iota]} |\zeta_\varepsilon'(\psi) - \zeta_0'(\psi)| = o_\varepsilon(1). \tag{88}$$

We first explain why this implies the result. First, (88) implies that for sufficiently small ε,

$$\zeta_\varepsilon(\psi_0 - \iota) \geqslant \psi_0 - \iota + \iota_1, \qquad \zeta_\varepsilon(\psi_0 + \iota) \leqslant \psi_0 + \iota - \iota_1, \qquad \sup_{\psi \in [\psi_0 - \iota, \psi_0 + \iota]} \zeta_\varepsilon'(\psi) \leqslant 1 - \iota_1.$$

This implies that ζ_ε has a unique fixed point ψ_ε in $[\psi_0 - \iota, \psi_0 + \iota]$. Furthermore, it implies $|\zeta_\varepsilon(\psi_0) - \psi_0| = o_\varepsilon(1)$, which combined with the above derivative estimate gives

$$|\psi_\varepsilon - \psi_0| \leqslant |\zeta_\varepsilon(\psi_0) - \psi_0|/\iota_1 = o_\varepsilon(1).$$

Continuity considerations then imply $(q_\varepsilon, \psi_\varepsilon, \varrho_\varepsilon) \to (q_0, \psi_0, 1 - q_0)$ as $\varepsilon \downarrow 0$. We now turn to the proof of (88). Let $\psi \in [\psi_0 - \iota, \psi_0 + \iota]$. Below, $o_\varepsilon(1)$ is an error uniform over ψ. Let $q = P^\varepsilon(\psi)$ and $\widetilde{q} = P(\psi)$. Note that

$$|q - \widetilde{q}| \leqslant \mathbb{E}\left[|(\mathrm{th}((\psi + \varepsilon)^{1/2} Z) + \varepsilon(\psi + \varepsilon)^{1/2} Z)^2 - \mathrm{th}^2(\psi^{1/2} Z)|\right] \leqslant o_\varepsilon(1).$$

Let $\varrho = \varrho_\varepsilon(q, \psi)$, and note that

$$|\varrho - (1 - q)| = o_\varepsilon(1).$$

Thus

$$\varrho \geqslant (1 - \widetilde{q}) - |\widetilde{q} - q| - |\varrho - (1 - q)| \geqslant 2\iota_0 - o_\varepsilon(1) \geqslant \iota_0,$$

so ϱ is bounded away from 0. By Cauchy-Schwarz,

$$\begin{aligned}
&|\zeta_\varepsilon(\psi) - \zeta_0(\psi)| = |R^\varepsilon(q, \psi) - R_{\alpha_\star}(\widetilde{q})| \\
&= \alpha_\star \mathbb{E}\left[|F_{\varepsilon,\varrho}((q + \varepsilon)^{1/2} Z) - F_{1-q_0}(q^{1/2} Z)||F_{\varepsilon,\varrho}((q + \varepsilon)^{1/2} Z) + F_{1-q_0}(q^{1/2} Z)|\right] \\
&\leqslant \alpha_\star \mathbb{E}\left[(F_{\varepsilon,\varrho}((q + \varepsilon)^{1/2} Z) - F_{1-\widetilde{q}}(\widetilde{q}^{1/2} Z))^2\right]^{1/2} \mathbb{E}\left[(F_{\varepsilon,\varrho}((q + \varepsilon)^{1/2} Z) + F_{1-\widetilde{q}}(\widetilde{q}^{1/2} Z))^2\right]^{1/2}.
\end{aligned}$$

Expanding $F_{\varepsilon,\varrho}$ using (19) shows the first expectation is $o_\varepsilon(1)$, while the second is bounded by Lemma 4.21(a). Thus $|\zeta_\varepsilon(\psi) - \zeta_0(\psi)| = o_\varepsilon(1)$ uniformly in $\psi \in [\psi_0 - \iota, \psi_0 + \iota]$. Furthermore,

$$\zeta_\varepsilon'(\psi) = \frac{\partial R^\varepsilon}{\partial q}(q, \psi)(P^\varepsilon)'(\psi) + \frac{\partial R^\varepsilon}{\partial \psi}(q, \psi), \qquad \zeta_0'(\psi) = R_{\alpha_\star}'(\widetilde{q}) P'(\psi).$$

Similar computations to above show

$$\left|\frac{\partial R^\varepsilon}{\partial q}(q, \psi) - R_{\alpha_\star}'(\widetilde{q})\right|, |(P^\varepsilon)'(\psi) - P'(\psi)|, \left|\frac{\partial R^\varepsilon}{\partial \psi}(q, \psi)\right| = o_\varepsilon(1),$$

and thus $|\zeta_\varepsilon'(\psi) - \zeta_0'(\psi)| = o_\varepsilon(1)$ uniformly in ψ. This proves (88). □

A.2. **Approximation for (pseudo)-Lipschitz functions.**

Proof of Fact 4.20. Let (x, y) be a sample from the optimal coupling of (μ, μ'). Then

$$\begin{aligned}|\mathbb{E}_\mu[f] - \mathbb{E}_{\mu'}[f]| &\leqslant \mathbb{E}\,|f(x) - f(y)| \leqslant L\,\mathbb{E}\,[|x-y|(|x|+|y|+1)] \\ &\leqslant L\,\mathbb{E}[|x-y|^2]^{1/2}\,\mathbb{E}[3(|x|^2+|y|^2+1)]^{1/2} \\ &\leqslant L\,\mathbb{E}[|x-y|^2]^{1/2}\,\mathbb{E}[3(3|x|^2+2|x-y|^2+1)]^{1/2} \\ &\leqslant 3L\mathbb{W}_2(\mu,\mu')(\mu_2+\mathbb{W}_2(\mu,\mu')+1),\end{aligned}$$

where we have used the estimate $|y|^2 \leqslant 2|x|^2 + 2|x-y|^2$. □

Proof of Fact 6.11. Couple $(x, y, z) \sim \mu$ and $(x', y', z') \sim \mu'$ in the $\mathbb{W}_2$-optimal way. Then, the left-hand side of (62) is bounded by the sum of:

$$\begin{aligned}\mathbb{E}|f_1(x)||f_2(y)||f_3(z)-f_3(z')| &\leqslant L(\mathbb{E}f_1(x)^4)^{1/4}(\mathbb{E}f_2(y)^4)^{1/4}(\mathbb{E}|z-z'|^2)^{1/2} \\ &\leqslant L(\mathbb{E}f_1(x)^4)^{1/4}(\mathbb{E}f_2(y)^4)^{1/4}\mathbb{W}_2(\mu,\mu'), \\ \mathbb{E}|f_1(x)||f_3(z')||f_2(y)-f_2(y')| &\leqslant L^2(\mathbb{E}f_1(x)^2)^{1/2}(\mathbb{E}|y-y'|^2)^{1/2} \leqslant L^2(\mathbb{E}f_1(x)^2)^{1/2}\mathbb{W}_2(\mu,\mu') \\ \mathbb{E}|f_2(y')||f_3(z')||f_1(x)-f_1(x')| &\leqslant L^2(\mathbb{E}f_2(y')^2)^{1/2}(\mathbb{E}|x-x'|^2)^{1/2} \leqslant L^2(\mathbb{E}f_2(y')^2)^{1/2}\mathbb{W}_2(\mu,\mu').\end{aligned}$$

Finally, by Fact 4.20,

$$\mathbb{E}f_2(y')^2 \leqslant \mathbb{E}f_2(y)^2 + 3\mathbb{W}_2(\mu,\mu')(\mathbb{E}f_2(y)^2 + \mathbb{W}_2(\mu,\mu')+1).$$

Combining gives the conclusion. □

A.3. **Gradient and Hessian formulas for $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}$, and regularity estimates.**

Proof of Lemma 4.16. By standard properties of convex duals,

$$(V_\varepsilon^*)'(m) = -\arg\min_{\dot h}\left\{-m\dot h + V_\varepsilon(\dot h)\right\} = -\mathrm{th}_\varepsilon^{-1}(m).$$

We differentiate the interaction term in $\mathcal{F}^{\varepsilon}_{\mathrm{TAP}}$ by gaussian integration by parts. For each $i \in [N]$, $a \in [M]$,

$$\begin{aligned}&\frac{\partial}{\partial m_i}\overline{F}_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}\left(\frac{\langle \boldsymbol{g}^a, \boldsymbol{m}\rangle}{\sqrt{N}} + \varepsilon^{1/2}\hat{g}_a - \rho_\varepsilon(q(\boldsymbol{m}))n_a\right) \\ &= \frac{\partial}{\partial m_i}\log\mathbb{E}\,\chi_\varepsilon\left(\frac{\langle \boldsymbol{g}^a, \boldsymbol{m}\rangle}{\sqrt{N}} + \varepsilon^{1/2}\hat{g}_a - \rho_\varepsilon(q(\boldsymbol{m}))n_a + \rho_\varepsilon(q(\boldsymbol{m}))^{1/2}Z\right) \\ &= \frac{\mathbb{E}\,\chi_\varepsilon'\left(\frac{\langle \boldsymbol{g}^a, \boldsymbol{m}\rangle}{\sqrt{N}} + \varepsilon^{1/2}\hat{g}_a - \rho_\varepsilon(q(\boldsymbol{m}))n_a + \rho_\varepsilon(q(\boldsymbol{m}))^{1/2}Z\right)\left(\frac{g_i^a}{\sqrt{N}} - \rho_\varepsilon'(q(\boldsymbol{m}))\frac{2m_in_a}{N} + \frac{\rho_\varepsilon'(q(\boldsymbol{m}))}{\rho_\varepsilon(q(\boldsymbol{m}))^{1/2}}\frac{m_i}{N}Z\right)}{\mathbb{E}\,\chi_\varepsilon'\left(\frac{\langle \boldsymbol{g}^a, \boldsymbol{m}\rangle}{\sqrt{N}} + \varepsilon^{1/2}\hat{g}_a - \rho_\varepsilon(q(\boldsymbol{m}))n_a + \rho_\varepsilon(q(\boldsymbol{m}))^{1/2}Z\right)} \\ &= F_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\acute{h}_a)\left(\frac{g_i^a}{\sqrt{N}} - \rho_\varepsilon'(q(\boldsymbol{m}))\frac{2m_in_a}{N}\right) + \frac{\mathbb{E}\,\chi_\varepsilon''(\acute{h}_a + \rho_\varepsilon(q(\boldsymbol{m}))^{1/2}Z)}{\mathbb{E}\,\chi_\varepsilon(\acute{h}_a + \rho_\varepsilon(q(\boldsymbol{m}))^{1/2}Z)}\cdot\frac{\rho_\varepsilon'(q(\boldsymbol{m}))m_i}{N}. \\ &= \frac{g_i^a}{\sqrt{N}}F_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\acute{h}_a) + \frac{\rho_\varepsilon'(q(\boldsymbol{m}))m_i}{N}\left(-2F_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\acute{h}_a)n_a + F_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\acute{h}_a)^2 + F'_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\acute{h}_a)\right).\end{aligned}$$

Thus

$$\frac{\partial}{\partial m_i}\mathcal{F}^{\varepsilon}_{\rm TAP}(\boldsymbol{m},\boldsymbol{n}) = -\mathrm{th}_\varepsilon^{-1}(m_i) + \varepsilon^{1/2}\dot{g}_i + \frac{(\boldsymbol{G}^\top F_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\acute{h}_a))_i}{\sqrt{N}} + \frac{\rho_\varepsilon'(q(\boldsymbol{m}))m_i}{N}\sum_{a=1}^M\left((n_a - F_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\acute{h}_a))^2 + F'_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\acute{h}_a)\right),$$

which implies (34). The formula (35) follows by directly differentiating $\mathcal{F}^{\varepsilon}_{\rm TAP}$. Setting (35) to zero shows that $\nabla_{\boldsymbol{n}}\mathcal{F}^{\varepsilon}_{\rm TAP}(\boldsymbol{m},\boldsymbol{n}) = 0$ if and only if $\acute{\boldsymbol{h}} = \widehat{\boldsymbol{h}}$, which rearranges to (36). This implies $F_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\acute{\boldsymbol{h}}) = \boldsymbol{n}$, so setting (34) to zero yields (37). □

Proof of Fact 6.5. Note that

$$\frac{\partial}{\partial m_i}\mathrm{th}_\varepsilon^{-1}(m_i) = \frac{1}{\mathrm{th}_\varepsilon'(\dot{h}_i)} = \frac{1}{1+\varepsilon - \mathrm{th}^2(\dot{h}_i)} = \frac{\mathrm{ch}^2(\dot{h}_i)}{1+\varepsilon\mathrm{ch}^2(\dot{h}_i)}.$$

The functions $F_{\varepsilon,\varrho}, F'_{\varepsilon,\varrho}$ can be differentated in ϱ as follows. By gaussian integration by parts (or Itô's formula),

$$\frac{\mathrm{d}}{\mathrm{d}\varrho}\mathbb{E}\,\chi_\varepsilon(x+\varrho^{1/2}Z) = \frac{1}{2}\mathbb{E}\,\chi_\varepsilon''(x+\varrho^{1/2}Z),$$

and similarly for χ_ε'. Thus, abbreviating $\chi_{\varepsilon,\varrho}(x) = \mathbb{E}\,\chi_\varepsilon(x+\varrho^{1/2}Z)$,

$$\frac{\mathrm{d}}{\mathrm{d}\varrho}F_{\varepsilon,\varrho}(x) = \frac{\mathrm{d}}{\mathrm{d}\varrho}\frac{\chi_{\varepsilon,\varrho}(x)}{\chi'_{\varepsilon,\varrho}(x)} = \frac{1}{2}\left(\frac{\chi^{(3)}_{\varepsilon,\varrho}(x)}{\chi_{\varepsilon,\varrho}(x)} - \frac{\chi'_{\varepsilon,\varrho}(x)\chi''_{\varepsilon,\varrho}(x)}{\chi_{\varepsilon,\varrho}(x)^2}\right).$$

We also have

$$F'_{\varepsilon,\varrho}(x) = \frac{\chi''_{\varepsilon,\varrho}(x)}{\chi_{\varepsilon,\varrho}(x)} - \frac{(\chi'_{\varepsilon,\varrho}(x))^2}{\chi_{\varepsilon,\varrho}(x)^2}, \qquad F''_{\varepsilon,\varrho}(x) = \frac{\chi'''_{\varepsilon,\varrho}(x)}{\chi_{\varepsilon,\varrho}(x)} - \frac{3(\chi'_{\varepsilon,\varrho}(x))(\chi''_{\varepsilon,\varrho}(x))}{\chi_{\varepsilon,\varrho}(x)^2} + \frac{2(\chi'_{\varepsilon,\varrho}(x))^3}{\chi_{\varepsilon,\varrho}(x)^3}.$$

Thus

$$\frac{\mathrm{d}}{\mathrm{d}\varrho}F_{\varepsilon,\varrho}(x) = \frac{1}{2}\left(2F_{\varepsilon,\varrho}(x)F'_{\varepsilon,\varrho}(x) + F''_{\varepsilon,\varrho}(x)\right).$$

A similar calculation shows

$$\frac{\mathrm{d}}{\mathrm{d}\varrho}F'_{\varepsilon,\varrho}(x) = \frac{1}{2}\left(2F_{\varepsilon,\varrho}(x)F''_{\varepsilon,\varrho}(x) + 2F'_{\varepsilon,\varrho}(x)^2 + F^{(3)}_{\varepsilon,\varrho}(x)\right).$$

The result follows by directly differentiating (34) and (35) using the above formulas. □

Proof of Lemma 6.6. As $(\boldsymbol{m},\boldsymbol{n}) \in \mathcal{S}_{\varepsilon,r_0}$, approximation arguments identical to the proof of Corollary 4.18 show the estimates for $q(\boldsymbol{m}), \psi(\boldsymbol{n}), d_\varepsilon(\boldsymbol{m},\boldsymbol{n})$ in part (a). The regularity estimate (23) of ρ_ε and its derivatives proves the rest of part (a). Differentiating (19) yields

$$F'_{\varepsilon,\varrho}(x) = -\frac{\varepsilon}{1+\varepsilon\varrho} - \frac{1}{(\varrho+\varepsilon(1+\varepsilon\varrho))(1+\varepsilon\varrho)}\mathcal{E}'\left(\frac{\kappa(1+\varepsilon\varrho)-x}{\sqrt{(\varrho+\varepsilon(1+\varepsilon\varrho))(1+\varepsilon\varrho)}},\right).$$

By Lemma 4.21, we see that for ϱ in a neighborhood of ϱ_ε, $\sup_{x\in\mathbb{R}}\left|\frac{\mathrm{d}}{\mathrm{d}\varrho}F'_{\varepsilon,\varrho}(x)\right|$ is bounded by an absolute constant. Note that

$$\sup_{x\in\mathbb{R}}\left|\frac{\mathrm{d}}{\mathrm{d}\varrho}\frac{F'_{\varepsilon,\varrho}(x)}{1+\varrho F'_{\varepsilon,\varrho}(x)}\right| \leqslant \sup_{x\in\mathbb{R}}\left|\frac{F'_{\varepsilon,\varrho}(x)}{(1+\varrho F'_{\varepsilon,\varrho}(x))^2}\right| + \sup_{x\in\mathbb{R}}\left|\frac{1}{(1+\varrho F'_{\varepsilon,\varrho}(x))^2}\right|\cdot\sup_{x\in\mathbb{R}}\left|\frac{\mathrm{d}}{\mathrm{d}\varrho}F'_{\varepsilon,\varrho}(x)\right|. \tag{89}$$

By (42),

$$\frac{1}{1+\varrho F'_{\varepsilon,\varrho}(x)} \geqslant \frac{\varrho+\varepsilon(1+\varepsilon\varrho)}{\varepsilon},$$

which for ϱ in a neighborhood of ϱ_ε is bounded depending only on ε. It follows that (89) is is bounded depending only on ε. So,

$$\|\boldsymbol{D}_2 - \widetilde{\boldsymbol{D}}_2\|_{\mathrm{op}} \leqslant \left| \frac{F'_{\varepsilon,\varrho_\varepsilon}(x)}{1+\varrho_\varepsilon F'_{\varepsilon,\varrho_\varepsilon}(x)} - \frac{F'_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(x)}{1+\rho_\varepsilon(q(\boldsymbol{m}))F'_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(x)} \right| = o_{r_0}(1).$$

This proves part (b). Part (c) follows from Fact 4.22, as (for $\rho_\varepsilon(q(\boldsymbol{m}))$ in a neighborhood of $\varrho_\varepsilon > 0$) the images of $F'_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}$ and $F^{(3)}_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}$ are bounded. Similarly,

$$\frac{1}{\sqrt{N}}\|\boldsymbol{D}_4^{-1}F''(\acute{\boldsymbol{h}})\| \leqslant \|\boldsymbol{D}_4^{-1}\|_{\mathrm{op}}\|F''(\acute{\boldsymbol{h}})\|_\infty \overset{(42)}{\leqslant} \frac{\rho_\varepsilon(q(\boldsymbol{m}))+\varepsilon(1+\varepsilon\rho_\varepsilon(q(\boldsymbol{m})))}{\varepsilon}\|F''(\acute{\boldsymbol{h}})\|_\infty.$$

Since the image of $F''_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}$ is bounded by Fact 4.22, this proves part (d). □

Proof of Proposition 4.7. We will show that the matrices $\nabla^2_{\boldsymbol{m},\boldsymbol{m}}\mathcal{F}^\varepsilon_{\mathrm{TAP}}$, $\nabla^2_{\boldsymbol{m},\boldsymbol{n}}\mathcal{F}^\varepsilon_{\mathrm{TAP}}$, $\nabla^2_{\boldsymbol{n},\boldsymbol{n}}\mathcal{F}^\varepsilon_{\mathrm{TAP}}$ in Fact 6.5 have bounded operator norm (with bound depending on $\varepsilon, C_{\mathrm{cvx}}, C_{\mathrm{bd}}, D$). Throughout this proof, C is a constant depending on $\varepsilon, C_{\mathrm{cvx}}, C_{\mathrm{bd}}, D$, which may change from line to line.

Under $\mathbb{P}$, we have $\|\boldsymbol{G}\|_{\mathrm{op}}, \|\widehat{\boldsymbol{g}}\| \leqslant C\sqrt{N}$ with high probability. Under $\mathbb{P}^{\boldsymbol{m}',\boldsymbol{n}'}_{\varepsilon,\mathrm{Pl}}$, we may write $\boldsymbol{G} = \mathbb{E}^{\boldsymbol{m}',\boldsymbol{n}'}_{\varepsilon,\mathrm{Pl}}\boldsymbol{G} + \widetilde{\boldsymbol{G}}$ for $\widetilde{\boldsymbol{G}}$ as in Lemma 4.17. Then $\|\widetilde{\boldsymbol{G}}\|_{\mathrm{op}} \leqslant C\sqrt{N}$ with high probability, and by Lemma 4.17, $\|\mathbb{E}^{\boldsymbol{m}',\boldsymbol{n}'}_{\varepsilon,\mathrm{Pl}}\boldsymbol{G}\| \leqslant C\sqrt{N}$. On this event, $\|\boldsymbol{G}\|_{\mathrm{op}} \leqslant C\sqrt{N}$. Since $\rho_\varepsilon(q(\boldsymbol{m}')) \in [C_{\mathrm{bd}}^{-1}, C_{\mathrm{bd}}]$, $\widehat{\boldsymbol{h}}' = F^{-1}_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}'))}(\boldsymbol{n})$ satisfies $\|\widehat{\boldsymbol{h}}'\| \leqslant C\sqrt{N}$. Then, (37) implies $\|\widehat{\boldsymbol{g}}\| \leqslant C\sqrt{N}$. So, under both $\mathbb{P}$ and $\mathbb{P}^{\boldsymbol{m}',\boldsymbol{n}'}_{\varepsilon,\mathrm{Pl}}$, we have $\|\boldsymbol{G}\|_{\mathrm{op}}, \|\widehat{\boldsymbol{g}}\| \leqslant C\sqrt{N}$ with high probability. For the remainder of this proof, we assume this event holds.

Consider any $\|\boldsymbol{m}\|^2, \|\boldsymbol{n}\|^2 \leqslant DN$. The above bounds on $\|\boldsymbol{G}\|_{\mathrm{op}}, \|\widehat{\boldsymbol{g}}\|$ imply $\|\acute{\boldsymbol{h}}\| \leqslant C\sqrt{N}$. By (23), $C_{\mathrm{bd}}^{-1} \leqslant \rho_\varepsilon(q(\boldsymbol{m})) \leqslant C_{\mathrm{bd}}$ and $|\rho'_\varepsilon(q(\boldsymbol{m}))|, |\rho''_\varepsilon(q(\boldsymbol{m}))| \leqslant C_{\mathrm{bd}}$. Abbreviate $F = F_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}$ as above. By Fact 4.22,

$$\sup_{x\in\mathbb{R}}|F'(x)|, \sup_{x\in\mathbb{R}}|F''(x)|, \sup_{x\in\mathbb{R}}|F^{(3)}(x)| \leqslant C. \tag{90}$$

Thus F is C-Lipschitz. By (19),

$$F(0) = \frac{1}{\sqrt{(\rho_\varepsilon(q(\boldsymbol{m}))+\varepsilon(1+\varepsilon\rho_\varepsilon(q(\boldsymbol{m}))))(1+\varepsilon\rho_\varepsilon(q(\boldsymbol{m})))}}\mathcal{E}\left(\frac{\kappa\sqrt{1+\varepsilon\rho_\varepsilon(q(\boldsymbol{m}))}}{\sqrt{\rho_\varepsilon(q(\boldsymbol{m}))+\varepsilon(1+\varepsilon\rho_\varepsilon(q(\boldsymbol{m})))}}\right)$$

is bounded, and thus

$$\|F(\acute{\boldsymbol{h}})\| \leqslant \|F(\boldsymbol{0})\| + C\|\acute{\boldsymbol{h}}\| \leqslant C\sqrt{N}.$$

By (90) we also have $\|F'(\acute{\boldsymbol{h}})\|, \|F''(\acute{\boldsymbol{h}})\|, \|F^{(3)}(\acute{\boldsymbol{h}})\| \leqslant C\sqrt{N}$. This also implies $d_\varepsilon(\boldsymbol{m},\boldsymbol{n}) \leqslant C$.

Since $\dot{f}_\varepsilon$ is bounded, $\|\boldsymbol{D}_1\|_{\mathrm{op}} \leqslant C$. Since F' is bounded, $\|\boldsymbol{D}_3\|_{\mathrm{op}}, \|\boldsymbol{D}_4\|_{\mathrm{op}} \leqslant C$. The estimate (42) also implies $\|\widetilde{\boldsymbol{D}}_2\|_{\mathrm{op}}, \|\boldsymbol{D}_4^{-1}\|_{\mathrm{op}} \leqslant C$. Combining these estimates shows $\|\nabla^2_{\boldsymbol{m},\boldsymbol{m}}\mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})\|_{\mathrm{op}}$, $\|\nabla^2_{\boldsymbol{m},\boldsymbol{n}}\mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})\|_{\mathrm{op}}$, $\|\nabla^2_{\boldsymbol{n},\boldsymbol{n}}\mathcal{F}^\varepsilon_{\mathrm{TAP}}(\boldsymbol{m},\boldsymbol{n})\|_{\mathrm{op}} \leqslant C$. □

A.4. **Analysis of AMP iteration in planted model.**

Proof of Proposition 5.4. The state evolution [BMN20, Theorem 1] implies that

$$\frac{1}{N}\sum_{i=1}^N \delta(\dot{h}_i, \dot{\xi}_i, \dot{h}_i^{(1),1}, \ldots, \dot{h}_i^{(1),k}) \xrightarrow{\mathbb{W}_2} \mathcal{N}(0, \dot{\Sigma}^{(1)}_{\leqslant k}), \quad \frac{1}{M}\sum_{a=1}^M \delta(\widehat{h}_a, \widehat{\xi}_a, \widehat{h}_a^{(1),0}, \ldots, \widehat{h}_a^{(1),k}) \xrightarrow{\mathbb{W}_2} \mathcal{N}(0, \widehat{\Sigma}^{(1)}_{\leqslant k}),$$

for the following arrays $\dot\Sigma^{(1)}, \widehat\Sigma^{(1)}$. First, $\widehat\Sigma^{(1)}$ agrees with $\widehat\Sigma^{+}$ on indices (i,j) where $\{(i,j)\}\cap\{\diamond,\bowtie\}\neq\emptyset$, and $\dot\Sigma^{(1)}$ agrees with $\dot\Sigma^{+}$ on (i,j) where $\{(i,j)\}\cap\{\diamond,\bowtie,0\}\neq\emptyset$. The remaining entries are defined by the following recursion. For $(\dot H,\dot\Xi,\dot H_1,\dots,\dot H_k)\sim\mathcal N(0,\dot\Sigma^{(1)}_{\leqslant k})$ and $0\leqslant i\leqslant k$,

$$\widehat\Sigma^{(1)}_{i,k} = \mathbb E\left[\left(\mathrm{th}_\varepsilon(\dot H_i) - \frac{\overline q_i}{q_\varepsilon}\mathrm{th}_\varepsilon(\dot H)\right)\left(\mathrm{th}_\varepsilon(\dot H_k) - \frac{\overline q_k}{q_\varepsilon}\mathrm{th}_\varepsilon(\dot H)\right)\right] + \frac{\varepsilon(q_\varepsilon-\overline q_i)(q_\varepsilon-\overline q_k)}{q_\varepsilon(q_\varepsilon+\varepsilon)} + \frac{(\overline q_i+\varepsilon)(\overline q_k+\varepsilon)}{q_\varepsilon+\varepsilon}. \tag{91}$$

For $(\widehat H,\widehat\Xi,\widehat H_0,\dots,\widehat H_k)\sim\mathcal N(0,\widehat\Sigma^{(1)}_{\leqslant k})$ and $0\leqslant i\leqslant k$, we have

$$\begin{aligned}\dot\Sigma^{(1)}_{i+1,k+1} = \alpha_\star\,\mathbb E\left[\left(F_{\varepsilon,\varrho_\varepsilon}(\widehat H_i) - \frac{\overline\psi_{i+1}}{\psi_\varepsilon}F_{\varepsilon,\varrho_\varepsilon}(\widehat H)\right)\left(F_{\varepsilon,\varrho_\varepsilon}(\widehat H_k) - \frac{\overline\psi_{k+1}}{\psi_\varepsilon}F_{\varepsilon,\varrho_\varepsilon}(\widehat H)\right)\right]\\ + \frac{\varepsilon(\psi_\varepsilon-\overline\psi_{i+1})(\psi_\varepsilon-\overline\psi_{k+1})}{\psi_\varepsilon(\psi_\varepsilon+\varepsilon)} + \frac{(\overline\psi_{i+1}+\varepsilon)(\overline\psi_{k+1}+\varepsilon)}{\psi_\varepsilon+\varepsilon}.\end{aligned} \tag{92}$$

We now verify by induction that $\widehat\Sigma^{(1)}$ and $\dot\Sigma^{(1)}$ coincide with $\widehat\Sigma^{+}$ and $\dot\Sigma^{+}$. Suppose $\dot\Sigma^{(1)}_{\leqslant k}=\dot\Sigma^{+}_{\leqslant k}$. Then,

$$\mathbb E[\mathrm{th}_\varepsilon(\dot H_i)\mathrm{th}_\varepsilon(\dot H_k)] = \dot\Sigma_{i,k}, \qquad \mathbb E[\mathrm{th}_\varepsilon(\dot H_i)\mathrm{th}_\varepsilon(\dot H)] = \overline q_i, \qquad \mathbb E[\mathrm{th}_\varepsilon(\dot H)^2] = q_\varepsilon,$$

so the right-hand side of (91) simplifies as

$$\dot\Sigma_{i,k} - \frac{\overline q_i\overline q_k}{q_\varepsilon} + \frac{\varepsilon(q_\varepsilon-\overline q_i)(q_\varepsilon-\overline q_k)}{q_\varepsilon(q_\varepsilon+\varepsilon)} + \frac{(\overline q_i+\varepsilon)(\overline q_k+\varepsilon)}{q_\varepsilon+\varepsilon} = \dot\Sigma_{i,k}+\varepsilon = \dot\Sigma^{+}_{i,k}.$$

Now, suppose $\widehat\Sigma^{(1)}_{\leqslant k}=\widehat\Sigma^{+}_{\leqslant k}$. Then,

$$\alpha_\star\,\mathbb E[F_{\varepsilon,\varrho_\varepsilon}(\widehat H_i)F_{\varepsilon,\varrho_\varepsilon}(\widehat H_k)] = \widehat\Sigma_{i+1,k+1}, \quad \alpha_\star\,\mathbb E[F_{\varepsilon,\varrho_\varepsilon}(\widehat H_i)F_{\varepsilon,\varrho_\varepsilon}(\widehat H)] = \overline\psi_{i+1}, \quad \alpha_\star\,\mathbb E[F_{\varepsilon,\varrho_\varepsilon}(\widehat H)^2] = \psi_\varepsilon,$$

so the right-hand side of (92) simplifies as

$$\widehat\Sigma_{i+1,k+1} - \frac{\overline\psi_{i+1}\overline\psi_{k+1}}{\psi_\varepsilon} + \frac{\varepsilon(\psi_\varepsilon-\overline\psi_{i+1})(\psi_\varepsilon-\overline\psi_{k+1})}{\psi_\varepsilon(\psi_\varepsilon+\varepsilon)} + \frac{(\overline\psi_{i+1}+\varepsilon)(\overline\psi_{k+1}+\varepsilon)}{\psi_\varepsilon+\varepsilon} = \widehat\Sigma_{i+1,k+1}+\varepsilon = \widehat\Sigma^{+}_{i+1,k+1}.$$

This completes the induction. □

To prove Proposition 5.5, we introduce two additional auxiliary AMP iterations. They are initialized at $\boldsymbol n^{(2),-1}=\boldsymbol n^{(3),-1}=\boldsymbol 0$, $\boldsymbol m^{(2),0}=\boldsymbol m^{(3),0}=q_\varepsilon^{1/2}\boldsymbol 1$, with iteration

$$\boldsymbol m^{(i),k} = \mathrm{th}_\varepsilon(\dot{\boldsymbol h}^{(i),k}), \qquad\qquad \boldsymbol n^{(i),k} = F_{\varepsilon,\varrho_\varepsilon}(\widehat{\boldsymbol h}^{(i),k}),$$

for $i\in\{2,3\}$ and $\dot{\boldsymbol h}^{(i),k},\widehat{\boldsymbol h}^{(i),k}$ as follows. Recall that $\overline{\boldsymbol G}$ is the matrix (44), and $\overline\psi_0=0$. Then,

$$\widehat{\boldsymbol h}^{(2),k} = \frac{1}{\sqrt N}\overline{\boldsymbol G}\left(\boldsymbol m^{(2),k} - \frac{\overline q_k}{q_\varepsilon}\boldsymbol m\right) + \frac{\sqrt\varepsilon(q_\varepsilon-\overline q_k)}{\sqrt{q_\varepsilon(q_\varepsilon+\varepsilon)}}\widehat{\boldsymbol\xi} + \frac{\overline q_k+\varepsilon}{q_\varepsilon+\varepsilon}\widehat{\boldsymbol h} - \varrho_\varepsilon\left(\boldsymbol n^{(2),k-1} - \frac{\overline\psi_k}{\psi_\varepsilon}\boldsymbol n\right) \tag{93}$$

$$\dot{\boldsymbol h}^{(2),k+1} = \frac{1}{\sqrt N}\overline{\boldsymbol G}^\top\left(\boldsymbol n^{(2),k} - \frac{\overline\psi_{k+1}}{\psi_\varepsilon}\boldsymbol n\right) + \frac{\sqrt\varepsilon(\psi_\varepsilon-\psi_{k+1})}{\sqrt{\psi_\varepsilon(\psi_\varepsilon+\varepsilon)}}\dot{\boldsymbol\xi} + \frac{\overline\psi_{k+1}+\varepsilon}{\psi_\varepsilon+\varepsilon}\dot{\boldsymbol h} - d_\varepsilon\left(\boldsymbol m^{(2),k} - \frac{\overline q_k}{q_\varepsilon}\boldsymbol m\right)$$

$$\widehat{\boldsymbol h}^{(3),k} = \frac{1}{\sqrt N}\widetilde{\boldsymbol G}\left(\boldsymbol m^{(3),k} - \boldsymbol m\right) + \frac{\overline q_k+\varepsilon}{q_\varepsilon+\varepsilon}\widehat{\boldsymbol h} - \varrho_\varepsilon\left(\boldsymbol n^{(3),k-1} - \frac{\overline\psi_k + \boldsymbol 1\{k\geqslant 1\}\varepsilon}{\psi_\varepsilon+\varepsilon}\boldsymbol n\right) \tag{94}$$

$$\dot{\boldsymbol h}^{(3),k+1} = \frac{1}{\sqrt N}\widetilde{\boldsymbol G}^\top\left(\boldsymbol n^{(3),k} - \boldsymbol n\right) + \frac{\overline\psi_{k+1}+\varepsilon}{\psi_\varepsilon+\varepsilon}\dot{\boldsymbol h} - d_\varepsilon\left(\boldsymbol m^{(3),k} - \frac{\overline q_k+\varepsilon}{q_\varepsilon+\varepsilon}\boldsymbol m\right).$$

The following proposition shows that all these AMP iterations approximate each other.

Proposition A.1. *For any $k \geqslant 0$, as $N \to \infty$ we have the following convergences in probability under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{PI}}$.*

(a) $\|\widehat{\boldsymbol{h}}^{(1),k} - \widehat{\boldsymbol{h}}^{(2),k}\|/\sqrt{N} \to 0$*, and if* $k \geqslant 1$, $\|\dot{\boldsymbol{h}}^{(1),k} - \dot{\boldsymbol{h}}^{(2),k}\|/\sqrt{N} \to 0$.

(b) $\|\widehat{\boldsymbol{h}}^{(2),k} - \widehat{\boldsymbol{h}}^{(3),k}\|/\sqrt{N} \to 0$*, and if* $k \geqslant 1$, $\|\dot{\boldsymbol{h}}^{(2),k} - \dot{\boldsymbol{h}}^{(3),k}\|/\sqrt{N} \to 0$.

(c) $\|\widehat{\boldsymbol{h}}^{(3),k} - \widehat{\boldsymbol{h}}^{k}\|/\sqrt{N} \to 0$*, and if* $k \geqslant 1$, $\|\dot{\boldsymbol{h}}^{(3),k} - \dot{\boldsymbol{h}}^{k}\|/\sqrt{N} \to 0$.

Proof of Proposition A.1(a). Similarly to (45), we can sample $Z' \sim \mathcal{N}(0,1)$, $\dot{\boldsymbol{\xi}}' \sim \mathcal{N}(0, \boldsymbol{I}_N)$, $\widehat{\boldsymbol{\xi}}' \sim \mathcal{N}(0, \boldsymbol{I}_M)$ coupled to $\widehat{\boldsymbol{G}}$ such that

$$\widehat{\boldsymbol{G}} + \boldsymbol{\Delta}' = \overline{\boldsymbol{G}} - \frac{\widehat{\boldsymbol{\xi}}'\boldsymbol{m}^\top}{\|\boldsymbol{m}\|} - \frac{\boldsymbol{n}(\dot{\boldsymbol{\xi}}')^\top}{\|\boldsymbol{n}\|}, \qquad \boldsymbol{\Delta}' = \frac{\boldsymbol{n}\boldsymbol{m}^\top}{\|\boldsymbol{n}\|\|\boldsymbol{m}\|} Z' \tag{95}$$

Note that $\|\boldsymbol{\Delta}'\|_{\mathrm{op}} = o(\sqrt{N})$ with high probability. Let $\simeq$ denote equality up to additive $o_N(1)$. By Proposition 5.4, for $(\dot{H}, \dot{\Xi}, \dot{H}_1, \ldots, \dot{H}_k) \sim \mathcal{N}(0, \dot{\Sigma}^{(1)}_{\leqslant k})$ and $(\widehat{H}, \widehat{\Xi}, \widehat{H}_0, \ldots, \widehat{H}_k) \sim \mathcal{N}(0, \widehat{\Sigma}^{(1)}_{\leqslant k})$,

$$\frac{1}{N}\langle \boldsymbol{m}, \dot{\boldsymbol{h}}^{(1),k}\rangle \simeq \mathbb{E}[\mathrm{th}_\varepsilon(\dot{H})\dot{H}_k] = \varrho_\varepsilon(\overline{\psi}_k + \varepsilon), \qquad \frac{1}{N}\langle \boldsymbol{n}, \widehat{\boldsymbol{h}}^{(1),k}\rangle \simeq \alpha_\star \mathbb{E}[F_{\varepsilon,\varrho_\varepsilon}(\widehat{H})\widehat{H}_k] = d_\varepsilon(\overline{q}_k + \varepsilon),$$
$$\frac{1}{N}\langle \boldsymbol{m}, \dot{\boldsymbol{h}}\rangle \simeq \mathbb{E}[\mathrm{th}_\varepsilon(\widehat{H})\widehat{H}] = \varrho_\varepsilon(\psi_\varepsilon + \varepsilon), \qquad \frac{1}{N}\langle \boldsymbol{n}, \widehat{\boldsymbol{h}}\rangle \simeq \alpha_\star \mathbb{E}[F_{\varepsilon,\varrho_\varepsilon}(\widehat{H})\widehat{H}] = d_\varepsilon(q_\varepsilon + \varepsilon).$$

Also,

$$\frac{1}{N}\left\langle \boldsymbol{m}, \boldsymbol{m}^{(1),k} - \frac{\overline{q}_k}{q_\varepsilon}\boldsymbol{m}\right\rangle \simeq \overline{q}_k - \frac{\overline{q}_k}{q_\varepsilon}\cdot q_\varepsilon = 0, \quad \frac{1}{N}\left\langle \boldsymbol{n}, \boldsymbol{n}^{(1),k-1} - \frac{\overline{\psi}_k}{\psi_\varepsilon}\boldsymbol{n}\right\rangle \simeq \overline{\psi}_k - \frac{\overline{\psi}_k}{\psi_\varepsilon}\cdot \psi_\varepsilon = 0. \tag{96}$$

Finally $\frac{1}{N}\langle \dot{\boldsymbol{\xi}}, \boldsymbol{m}\rangle \simeq \frac{1}{N}\langle \widehat{\boldsymbol{\xi}}, \boldsymbol{n}\rangle \simeq 0$. Considering the inner product of (47) with $\boldsymbol{n}$ shows

$$0 \simeq \frac{1}{N}\left\langle \boldsymbol{n}, \frac{1}{\sqrt{N}}\widehat{\boldsymbol{G}}\left(\boldsymbol{m}^{(1),k} - \frac{\overline{q}_k}{q_\varepsilon}\boldsymbol{m}\right)\right\rangle.$$

We can expand $\widehat{\boldsymbol{G}}$ using (95). Since $\boldsymbol{n}^\top\overline{\boldsymbol{G}} = \boldsymbol{0}$, $\frac{1}{N}\langle \boldsymbol{n}, \widehat{\boldsymbol{\xi}}'\rangle \simeq 0$ in probability, and $\|\boldsymbol{\Delta}'\|_{\mathrm{op}} = o(\sqrt{N})$,

$$0 \simeq \frac{1}{N}\left\langle \boldsymbol{n}, \frac{1}{\sqrt{N}}\left(\overline{\boldsymbol{G}} - \frac{\widehat{\boldsymbol{\xi}}'\boldsymbol{m}^\top}{\|\boldsymbol{m}\|} - \frac{\boldsymbol{n}(\dot{\boldsymbol{\xi}}')^\top}{\|\boldsymbol{n}\|} - \boldsymbol{\Delta}'\right)\left(\boldsymbol{m}^{(1),k} - \frac{\overline{q}_k}{q_\varepsilon}\boldsymbol{m}\right)\right\rangle \simeq \frac{\|\boldsymbol{n}\|}{N^{3/2}}\left\langle \dot{\boldsymbol{\xi}}', \boldsymbol{m}^{(1),k} - \frac{\overline{q}_k}{q_\varepsilon}\boldsymbol{m}\right\rangle.$$

Thus,

$$\frac{1}{N}\left\langle \dot{\boldsymbol{\xi}}', \boldsymbol{m}^{(1),k} - \frac{\overline{q}_k}{q_\varepsilon}\boldsymbol{m}\right\rangle \simeq 0 \tag{97}$$

in probability for all k. An analogous computation shows

$$\frac{1}{N}\left\langle \widehat{\boldsymbol{\xi}}', \boldsymbol{n}^{(1),k-1} - \frac{\overline{\psi}_k}{\psi_\varepsilon}\boldsymbol{n}\right\rangle \simeq 0.$$

By (95),

$$\frac{1}{\sqrt{N}}(\widehat{\boldsymbol{G}} - \overline{\boldsymbol{G}})\left(\boldsymbol{m}^{(1),k} - \frac{\overline{q}_k}{q_\varepsilon}\boldsymbol{m}\right) = \frac{\widehat{\boldsymbol{\xi}}'}{\sqrt{N}\|\boldsymbol{m}\|}\left\langle \boldsymbol{m}^\top, \boldsymbol{m}^{(1),k} - \frac{\overline{q}_k}{q_\varepsilon}\boldsymbol{m}\right\rangle + \frac{\boldsymbol{n}}{\sqrt{N}\|\boldsymbol{n}\|}\left\langle \dot{\boldsymbol{\xi}}', \boldsymbol{m}^{(1),k} - \frac{\overline{q}_k}{q_\varepsilon}\boldsymbol{m}\right\rangle$$
$$- \frac{1}{\sqrt{N}}\boldsymbol{\Delta}'\left(\boldsymbol{m}^{(1),k} - \frac{\overline{q}_k}{q_\varepsilon}\boldsymbol{m}\right),$$

and this has norm $o(\sqrt{N})$ by (96), (97). Subtracting (47) and (93) yields

$$\begin{aligned}\widehat{\boldsymbol{h}}^{(1),k} - \widehat{\boldsymbol{h}}^{(2),k} &= \frac{1}{\sqrt{N}}(\widehat{\boldsymbol{G}} - \overline{\boldsymbol{G}}) \left(\boldsymbol{m}^{(1),k} - \frac{\overline{q}_k}{q_\varepsilon}\boldsymbol{m}\right) + \frac{1}{\sqrt{N}}\overline{\boldsymbol{G}}(\boldsymbol{m}^{(1),k} - \boldsymbol{m}^{(2),k}) - \varrho_\varepsilon(\boldsymbol{n}^{(1),k-1} - \boldsymbol{n}^{(2),k-1}) \\ &= \frac{1}{\sqrt{N}}\overline{\boldsymbol{G}}(\boldsymbol{m}^{(1),k} - \boldsymbol{m}^{(2),k}) - \varrho_\varepsilon(\boldsymbol{n}^{(1),k-1} - \boldsymbol{n}^{(2),k-1}) + o(\sqrt{N}),\end{aligned}$$

where $o(\sqrt{N})$ denotes a vector with this norm. Analogously,

$$\dot{\boldsymbol{h}}^{(1),k+1} - \dot{\boldsymbol{h}}^{(2),k+1} = \frac{1}{\sqrt{N}}\overline{\boldsymbol{G}}^\top(\boldsymbol{n}^{(1),k} - \boldsymbol{n}^{(2),k}) - d_\varepsilon(\boldsymbol{m}^{(1),k} - \boldsymbol{m}^{(2),k}) + o(\sqrt{N}).$$

On the high probability event that $\|\overline{\boldsymbol{G}}\|_{\mathrm{op}} = O(\sqrt{N})$, we have

$$\begin{aligned}\|\widehat{\boldsymbol{h}}^{(1),k} - \widehat{\boldsymbol{h}}^{(2),k}\| &\leqslant O(1)\|\boldsymbol{m}^{(1),k} - \boldsymbol{m}^{(2),k}\| + \varrho_\varepsilon\|\boldsymbol{n}^{(1),k-1} - \boldsymbol{n}^{(2),k-1}\| + o(\sqrt{N}), \\ \|\dot{\boldsymbol{h}}^{(1),k+1} - \dot{\boldsymbol{h}}^{(2),k+1}\| &\leqslant O(1)\|\boldsymbol{n}^{(1),k} - \boldsymbol{n}^{(2),k}\| + |d_\varepsilon|\|\boldsymbol{m}^{(1),k} - \boldsymbol{m}^{(2),k}\| + o(\sqrt{N}).\end{aligned}$$

The claim now follows by induction on k: $\|\boldsymbol{m}^{(1),0} - \boldsymbol{m}^{(2),0}\| = \|\boldsymbol{n}^{(1),-1} - \boldsymbol{n}^{(2),-1}\| = 0$ by initialization, and because th_ε and $F_{\varepsilon,\varrho_\varepsilon}$ are $O(1)$-Lipschitz,

$$\|\boldsymbol{m}^{(1),k} - \boldsymbol{m}^{(2),k}\| \leqslant O(1)\|\dot{\boldsymbol{h}}^{(1),k} - \dot{\boldsymbol{h}}^{(2),k}\|, \qquad \|\boldsymbol{n}^{(1),k} - \boldsymbol{n}^{(2),k}\| \leqslant O(1)\|\widehat{\boldsymbol{h}}^{(1),k} - \widehat{\boldsymbol{h}}^{(2),k}\|,$$

for all $k \geqslant 1$, $k \geqslant 0$ respectively. □

Proof of Proposition A.1(b). Note that $\boldsymbol{\Delta}$ defined in (46) w.h.p. satisfies $\|\boldsymbol{\Delta}\|_{\mathrm{op}} = o(\sqrt{N})$. We write (94) as

$$\begin{aligned}\widehat{\boldsymbol{h}}^{(3),k} &= \frac{1}{\sqrt{N}}\widetilde{\boldsymbol{G}}(\boldsymbol{m}^{(2),k} - \boldsymbol{m}) + \frac{\overline{q}_k + \varepsilon}{q_\varepsilon + \varepsilon}\widehat{\boldsymbol{h}} - \varrho_\varepsilon\left(\boldsymbol{n}^{(2),k-1} - \frac{\overline{\psi}_k + \mathbf{1}\{k \geqslant 1\}\varepsilon}{\psi_\varepsilon + \varepsilon}\boldsymbol{n}\right) \\ &\quad + \frac{1}{\sqrt{N}}\widetilde{\boldsymbol{G}}(\boldsymbol{m}^{(3),k} - \boldsymbol{m}^{(2),k}) - \varrho_\varepsilon(\boldsymbol{n}^{(3),k-1} - \boldsymbol{n}^{(2),k-1}).\end{aligned}$$

By Proposition A.1(a), $\mathbb{W}_2(\mu_{\dot{\boldsymbol{h}}^{(2),k}}, \mu_{\dot{\boldsymbol{h}}^{(1),k}}) = o_N(1)$. So, Fact 4.20 and Proposition 5.4 imply

$$\begin{aligned}\frac{1}{N}\langle \boldsymbol{m}, \boldsymbol{m}^{(2),k}\rangle &\simeq \frac{1}{N}\langle \boldsymbol{m}, \boldsymbol{m}^{(1),k}\rangle \simeq \overline{q}_k, \\ \frac{1}{N}\langle \dot{\boldsymbol{\xi}}, \boldsymbol{m}^{(2),k}\rangle &\simeq \frac{1}{N}\langle \dot{\boldsymbol{\xi}}, \boldsymbol{m}^{(1),k}\rangle \simeq \mathbf{1}\{k \geqslant 1\}\varrho_\varepsilon\frac{\sqrt{\varepsilon}(\psi_\varepsilon - \overline{\psi}_k)}{\sqrt{\psi_\varepsilon(\psi_\varepsilon + \varepsilon)}}.\end{aligned} \tag{98}$$

By (45),

$$\begin{aligned}\frac{1}{\sqrt{N}}\widetilde{\boldsymbol{G}}(\boldsymbol{m}^{(2),k} - \boldsymbol{m}) &= \frac{1}{\sqrt{N}}\left(\overline{\boldsymbol{G}} - \sqrt{\frac{\varepsilon}{q(\boldsymbol{m}) + \varepsilon}} \cdot \frac{\widehat{\boldsymbol{\xi}}\boldsymbol{m}^\top}{\|\boldsymbol{m}\|} - \sqrt{\frac{\varepsilon}{\psi(\boldsymbol{n}) + \varepsilon}} \cdot \frac{\boldsymbol{n}\dot{\boldsymbol{\xi}}^\top}{\|\boldsymbol{n}\|} - \boldsymbol{\Delta}\right)(\boldsymbol{m}^{(2),k} - \boldsymbol{m}) \\ &= \frac{1}{\sqrt{N}}\overline{\boldsymbol{G}}(\boldsymbol{m}^{(2),k} - \boldsymbol{m}) + \frac{\sqrt{\varepsilon}(q_\varepsilon - \overline{q}_k)}{\sqrt{q_\varepsilon(q_\varepsilon + \varepsilon)}}\widehat{\boldsymbol{\xi}} - \mathbf{1}\{k \geqslant 1\}\varrho_\varepsilon\frac{\varepsilon(\psi_\varepsilon - \overline{\psi}_k)}{\psi_\varepsilon(\psi_\varepsilon + \varepsilon)}\boldsymbol{n} + o(\sqrt{N}).\end{aligned}$$

Since $\overline{\boldsymbol{G}}\boldsymbol{m} = \boldsymbol{0}$, we have $\overline{\boldsymbol{G}}(\boldsymbol{m}^{(2),k} - \boldsymbol{m}) = \overline{\boldsymbol{G}}(\boldsymbol{m}^{(2),k} - \frac{\overline{q}_k}{q_\varepsilon}\boldsymbol{m})$. Moreover,

$$\frac{\overline{\psi}_k + \mathbf{1}\{k \geqslant 1\}\varepsilon}{\psi_\varepsilon + \varepsilon} - \mathbf{1}\{k \geqslant 1\}\frac{\varepsilon(\psi_\varepsilon - \overline{\psi}_k)}{\psi_\varepsilon(\psi_\varepsilon + \varepsilon)} = \frac{\overline{\psi}_k}{\psi_\varepsilon}.$$

Combining the above and comparing with (93) shows

$$\widehat{\boldsymbol{h}}^{(3),k} = \widehat{\boldsymbol{h}}^{(2),k} + \frac{1}{\sqrt{N}}\widetilde{\boldsymbol{G}}(\boldsymbol{m}^{(3),k} - \boldsymbol{m}^{(2),k}) - \varrho_\varepsilon(\boldsymbol{n}^{(3),k-1} - \boldsymbol{n}^{(2),k-1}) + o(\sqrt{N}).$$

Similarly,

$$\dot{\boldsymbol{h}}^{(3),k+1} = \widehat{\boldsymbol{h}}^{(2),k+1} + \frac{1}{\sqrt{N}}\widetilde{\boldsymbol{G}}^\top(\boldsymbol{n}^{(3),k} - \boldsymbol{n}^{(2),k}) - d_\varepsilon(\boldsymbol{m}^{(3),k} - \boldsymbol{m}^{(2),k}) + o(\sqrt{N}).$$

On the high-probability event that $\|\widetilde{\boldsymbol{G}}\|_{\mathrm{op}} = O(\sqrt{N})$, this implies

$$\|\widehat{\boldsymbol{h}}^{(3),k} - \widehat{\boldsymbol{h}}^{(2),k}\| \leqslant O(1)\|\boldsymbol{m}^{(3),k} - \boldsymbol{m}^{(2),k}\| + \varrho_\varepsilon\|\boldsymbol{n}^{(3),k-1} - \boldsymbol{n}^{(2),k-1}\| + o(\sqrt{N}),$$
$$\|\dot{\boldsymbol{h}}^{(3),k+1} - \widehat{\boldsymbol{h}}^{(2),k+1}\| \leqslant O(1)\|\boldsymbol{n}^{(3),k} - \boldsymbol{n}^{(2),k}\| + |d_\varepsilon|\|\boldsymbol{m}^{(3),k} - \boldsymbol{m}^{(2),k}\| + o(\sqrt{N}).$$

The result follows by induction on k, like above. □

Proof of Proposition A.1(c). By Corollary 4.18, we have

$$\begin{aligned}\frac{\boldsymbol{G}}{\sqrt{N}} &\overset{d}{=} \frac{(1+o_N(1))\widehat{\boldsymbol{h}}\boldsymbol{m}^\top}{N(q_\varepsilon+\varepsilon)} + \frac{(1+o_N(1))\boldsymbol{n}\dot{\boldsymbol{h}}^\top}{N(\psi_\varepsilon+\varepsilon)} + \frac{o_N(1)\boldsymbol{n}\boldsymbol{m}^\top}{N} + \frac{\widetilde{\boldsymbol{G}}}{\sqrt{N}} \\ &= \frac{\widehat{\boldsymbol{h}}\boldsymbol{m}^\top}{N(q_\varepsilon+\varepsilon)} + \frac{\boldsymbol{n}\dot{\boldsymbol{h}}^\top}{N(\psi_\varepsilon+\varepsilon)} + \frac{\widetilde{\boldsymbol{G}}}{\sqrt{N}} + o_N(1), \end{aligned} \tag{99}$$

for $\widetilde{\boldsymbol{G}}$ as above and $o_N(1)$ a matrix with this operator norm. Since $q(\boldsymbol{m}) \simeq q_\varepsilon$, $\psi(\boldsymbol{n}) \simeq \psi_\varepsilon$, and under $\mathbb{P}^{\boldsymbol{m},\boldsymbol{n}}_{\varepsilon,\mathsf{Pl}}$ we have a.s. $\acute{\boldsymbol{h}} = F^{-1}_{\varepsilon,\rho_\varepsilon(q(\boldsymbol{m}))}(\boldsymbol{n})$, the following terms appearing in (36), (37) satisfy

$$\rho_\varepsilon(q(\boldsymbol{m})) \simeq \varrho_\varepsilon, \qquad \rho'_\varepsilon(q(\boldsymbol{m})) \simeq -1, \qquad d_\varepsilon(\boldsymbol{m},\boldsymbol{n}) \simeq d_\varepsilon.$$

Combining the AMP iteration (20) with (37) yields

$$\begin{aligned}\widehat{\boldsymbol{h}}^k &= \frac{1}{\sqrt{N}}\boldsymbol{G}(\boldsymbol{m}^k - \boldsymbol{m}) + \widehat{\boldsymbol{h}} + \varrho_\varepsilon(\boldsymbol{n} - \boldsymbol{n}^{k-1}) \\ &= \frac{1}{\sqrt{N}}\boldsymbol{G}(\boldsymbol{m}^{(3),k} - \boldsymbol{m}) + \widehat{\boldsymbol{h}} - \varrho_\varepsilon(\boldsymbol{n}^{(3),k-1} - \boldsymbol{n}) + \frac{1}{\sqrt{N}}\boldsymbol{G}(\boldsymbol{m}^k - \boldsymbol{m}^{(3),k}) - \varrho_\varepsilon(\boldsymbol{n}^{k-1} - \boldsymbol{n}^{(3),k-1}).\end{aligned}$$

By Proposition A.1(a)(b), $\mathbb{W}_2(\mu_{\dot{\boldsymbol{h}}^{(3),k}}, \mu_{\dot{\boldsymbol{h}}^{(1),k}}) = o_N(1)$. So, Fact 4.20 and Proposition 5.4 imply $\frac{1}{N}\langle \boldsymbol{m}, \boldsymbol{m}^{(3),k}\rangle \simeq \overline{q}_k$ (similarly to (98)) and

$$\frac{1}{N}\langle \dot{\boldsymbol{h}}, \boldsymbol{m}^{(3),k}\rangle \simeq \frac{1}{N}\langle \dot{\boldsymbol{h}}, \boldsymbol{m}^{(1),k}\rangle \simeq (\overline{\psi}_k + \mathbf{1}\{k \geqslant 1\}\varepsilon)\varrho_\varepsilon.$$

Expanding $\boldsymbol{G}$ using (99) then yields

$$\begin{aligned}\widehat{\boldsymbol{h}}^k &= \frac{1}{\sqrt{N}}\widetilde{\boldsymbol{G}}(\boldsymbol{m}^{(3),k} - \boldsymbol{m}) + \frac{\overline{q}_k + \varepsilon}{q_\varepsilon + \varepsilon}\widehat{\boldsymbol{h}} - \varrho_\varepsilon\left(\boldsymbol{n}^{(3),k-1} - \frac{\overline{\psi}_k + \varepsilon}{\psi_\varepsilon + \varepsilon}\boldsymbol{n}\right) \\ &\quad + \frac{1}{\sqrt{N}}\boldsymbol{G}(\boldsymbol{m}^k - \boldsymbol{m}^{(3),k}) - \varrho_\varepsilon(\boldsymbol{n}^{k-1} - \boldsymbol{n}^{(3),k-1}) + o(\sqrt{N}) \\ &= \widehat{\boldsymbol{h}}^{(3),k} + \frac{1}{\sqrt{N}}\boldsymbol{G}(\boldsymbol{m}^k - \boldsymbol{m}^{(3),k}) - \varrho_\varepsilon(\boldsymbol{n}^{k-1} - \boldsymbol{n}^{(3),k-1}) + o(\sqrt{N}).\end{aligned}$$

Analogously,

$$\dot{\boldsymbol{h}}^{k+1} = \dot{\boldsymbol{h}}^{(3),k+1} + \frac{1}{\sqrt{N}}\boldsymbol{G}^\top(\boldsymbol{n}^k - \boldsymbol{n}^{(3),k}) - d_\varepsilon(\boldsymbol{m}^{k-1} - \boldsymbol{m}^{(3),k-1}) + o(\sqrt{N}).$$

So, on the high probability event that $\|\boldsymbol{G}\|_{\rm op} = O(\sqrt{N})$,

$$\|\widehat{\boldsymbol{h}}^k - \widehat{\boldsymbol{h}}^{(3),k}\| = O(1)\|\boldsymbol{m}^k - \boldsymbol{m}^{(3),k}\| + \varrho_\varepsilon\|\boldsymbol{n}^{k-1} - \boldsymbol{n}^{(3),k-1}\| + o(\sqrt{N}),$$
$$\|\dot{\boldsymbol{h}}^{k+1} - \dot{\boldsymbol{h}}^{(3),k+1}\| = O(1)\|\boldsymbol{n}^k - \boldsymbol{n}^{(3),k}\| + |d_\varepsilon|\|\boldsymbol{m}^k - \boldsymbol{m}^{(3),k}\| + o(\sqrt{N}).$$

The result follows by induction on k, like above. □

Proof of Proposition 5.5. Immediate from Proposition A.1. □

A.5. **Continuity of first moment functional term.**

Proof of Lemma 7.4. Let C denote an absolute constant, which may change from line by line. By Lemma 7.3, $\log \Psi$ is $(2,1)$-pseudo-Lipschitz. By Cauchy–Schwarz (similarly to the proof of Fact 4.20),

$$\left|\mathbb{E}\log\Psi\left\{\frac{\kappa - a_1\hat{\boldsymbol{H}} - b_1\boldsymbol{N}}{c_1}\right\} - \log\Psi\left\{\frac{\kappa - a_2\hat{\boldsymbol{H}} - b_2\boldsymbol{N}}{c_2}\right\}\right| \leqslant C\sqrt{T_1T_2},$$

where

$$\begin{aligned} T_1 &= \mathbb{E}\left[\left(\frac{\kappa - a_1\hat{\boldsymbol{H}} - b_1\boldsymbol{N}}{c_1} - \frac{\kappa - a_2\hat{\boldsymbol{H}} - b_2\boldsymbol{N}}{c_2}\right)^2\right] \\ &\leqslant C\left(\frac{\max(a_1,a_2,b_1,b_2,c_1,c_2,1)(|a_1-a_2|+|b_1-b_2|+|c_1-c_2|)}{\min(c_1,c_2)^2}\right)^2 \end{aligned}$$

and

$$T_2 = \mathbb{E}\left[\left(\frac{\kappa - a_1\hat{\boldsymbol{H}} - b_1\boldsymbol{N}}{c_1}\right)^2 + \left(\frac{\kappa - a_2\hat{\boldsymbol{H}} - b_2\boldsymbol{N}}{c_2}\right)^2 + 1\right] \leqslant C\left(\frac{\max(a_1,a_2,b_1,b_2,c_1,c_2,1)}{\min(c_1,c_2)}\right)^4.$$

□

Appendix B. Verification of numerical conditions for $\kappa = 0$

In this appendix, we use rigorous interval arithmetic (implemented in the attached Python 3 file using `python-flint`) to verify the conditions in Theorem 3.6, other than Condition 1.3, at $\kappa = 0$. This proves Theorem 1.2. We also verify Claim 2.6 using interval arithmetic.

Throughout this section we take $\kappa = 0$, $\alpha_\star = \alpha_\star(0)$, $q_0 = q_\star(\alpha_\star, 0)$, and $\psi_0 = \psi_\star(\alpha_\star, 0)$. We will use Claims to denote statements whose proofs require interval arithmetic.

B.1. **Numerical estimates of parameters and special functions.** By [DS18, §7], the following are lower and upper bounds for $\alpha_\star, q_0, \psi_0$:

$$\begin{aligned} \alpha_{\rm lb} &= 0.833078599, & q_{\rm lb} &= 0.56394907949, & \psi_{\rm lb} &= 2.5763513100, \\ \alpha_{\rm ub} &= 0.833078600, & q_{\rm ub} &= 0.56394908030, & \psi_{\rm ub} &= 2.5763513224. \end{aligned}$$

Let $\gamma_0 = \frac{q_0}{1-q_0}$, $\gamma_{\rm lb} = \frac{q_{\rm lb}}{1-q_{\rm lb}}$ and $\gamma_{\rm ub} = \frac{q_{\rm ub}}{1-q_{\rm ub}}$. Note that Condition 3.4 only requires us to exhibit a value of $z > -1$ such that $\lambda(z) < 0$. In the verification below we will use the value

$$\widehat{z} = -0.669316.$$

For $k \in \{2,4\}$, define

$$p_k(\psi) = \mathbb{E}[\mathrm{th}(\psi^{1/2}Z)^k], \qquad r_k(\gamma) = \mathbb{E}[\mathcal{E}(\gamma^{1/2}Z)^k].$$

Note that the fixed-point condition in Condition 3.1 defining (q_0, ψ_0) implies (for $\kappa = 0$)

$$p_2(\psi_0) = q_0, \qquad r_2(\gamma_0) = \frac{(1-q_0)\psi_0}{\alpha_\star}. \tag{100}$$

Let

$$m(z,\psi) = \mathbb{E}[(z + \mathrm{ch}^2(\psi^{1/2}Z))^{-1}]. \tag{101}$$

Finally, define

$$g(m,q,\gamma) = \mathbb{E}\left\{\frac{\mathcal{E}'(\gamma^{1/2}Z)}{(1-q)(1-\mathcal{E}'(\gamma^{1/2}Z)) + m\mathcal{E}'(\gamma^{1/2}Z)}\right\}. \tag{102}$$

We now collect the main estimates in the verification whose proofs require computer assistance. The proofs of these claims are deferred to §B.4, with computer-assisted parts carried out in the attached Python file.

Claim B.1. *We have* $p_4(\psi_0) \in [p_{4,\mathrm{lb}}, p_{4,\mathrm{ub}}] \equiv [0.4405902310, 0.4405902320]$.

Claim B.2. *We have* $r_4(\gamma_0) \in [r_{4,\mathrm{lb}}, r_{4,\mathrm{ub}}] \equiv [5.297, 5.317]$.

Claim B.3. *We have* $m(\hat{z}) \leqslant m_{\mathrm{ub}} \equiv 0.9309695$, *where* $m(z) = m(z,\psi_0)$ *is defined in Condition 3.4.*

Claim B.4. *We have* $g(m(\hat{z}), q_0, \gamma_0) \geqslant g_{\mathrm{lb}} \equiv 0.7739$.

We conclude this preparatory subsection with a few useful lemmas. First, we reduce several integrals that will appear below to the functions p_2, p_4, r_2, r_4.

Lemma B.5. *The following identities hold.*

$$\mathfrak{t}(\psi) \equiv \mathbb{E}[\mathrm{th}'(\psi_0^{1/2}Z)^2] = 1 - 2p_2(\psi) + p_4(\psi), \tag{103}$$

$$\mathfrak{s}_1(\gamma) \equiv \mathbb{E}\left\{\mathcal{E}'(\gamma^{1/2}Z)\right\} = \frac{r_2(\gamma)}{1+\gamma}, \tag{104}$$

$$\mathfrak{s}_2(\gamma) \equiv \mathbb{E}\left\{\mathcal{E}(\gamma^{1/2}Z)^2\mathcal{E}'(\gamma^{1/2}Z)\right\} = \frac{r_4(\gamma)}{1+3\gamma}, \tag{105}$$

$$\mathfrak{s}_3(\gamma) \equiv \mathbb{E}\left\{\gamma^{1/2}Z\mathcal{E}(\gamma^{1/2}Z)\mathcal{E}'(\gamma^{1/2}Z)\right\} = -\frac{\gamma}{1+2\gamma}r_2(\gamma) + \frac{3\gamma}{(1+2\gamma)(1+3\gamma)}r_4(\gamma), \tag{106}$$

$$\mathfrak{s}_4(\gamma) \equiv \mathbb{E}\left\{(\gamma^{1/2}Z)^2\mathcal{E}'(\gamma^{1/2}Z)\right\} = -\frac{\gamma(4\gamma^2+\gamma-1)}{(1+\gamma)^2(1+2\gamma)}r_2(\gamma) + \frac{6\gamma^2}{(1+\gamma)(1+2\gamma)(1+3\gamma)}r_4(\gamma), \tag{107}$$

$$\mathfrak{s}_5(\gamma) \equiv \mathbb{E}\left\{\mathcal{E}'(\gamma^{1/2}Z)^2\right\} = \frac{\gamma}{1+2\gamma}r_2(\gamma) + \frac{1-\gamma}{(1+2\gamma)(1+3\gamma)}r_4(\gamma). \tag{108}$$

Proof. Equation (103) follows directly from the identity

$$\mathrm{th}'(x)^2 = (1-\mathrm{th}^2(x))^2 = 1 - 2\mathrm{th}^2(x) + \mathrm{th}^4(x).$$

For the remaining parts, we apply the identity $\mathcal{E}'(x) = \mathcal{E}(x)(\mathcal{E}(x) - x)$ (Lemma 4.21(b)) and integrate by parts. First,

$$\begin{aligned}\mathfrak{s}_1(\gamma) &= \mathbb{E}\left\{\mathcal{E}(\gamma^{1/2}Z)^2\right\} - \mathbb{E}\left\{\mathcal{E}(\gamma^{1/2}Z)\gamma^{1/2}Z\right\} \\ &= \mathbb{E}\left\{\mathcal{E}(\gamma^{1/2}Z)^2\right\} - \gamma\,\mathbb{E}\left\{\mathcal{E}'(\gamma^{1/2}Z)\right\} = r_2(\gamma) - \gamma\mathfrak{s}_1(\gamma),\end{aligned}$$

which proves (104). Similarly,

$$\mathfrak{s}_2(\gamma) = \mathbb{E}\left\{\mathcal{E}(\gamma^{1/2}Z)^4\right\} - \mathbb{E}\left\{\mathcal{E}(\gamma^{1/2}Z)^3\gamma^{1/2}Z\right\} = r_4(\gamma) - 3\gamma\mathfrak{s}_2(\gamma),$$

which proves (105). Then,

$$\begin{aligned}\mathfrak{s}_3(\gamma) &= \mathbb{E}\left\{\gamma^{1/2}Z\mathcal{E}(\gamma^{1/2}Z)^3\right\} - \mathbb{E}\left\{(\gamma^{1/2}Z)^2\mathcal{E}(\gamma^{1/2}Z)^2\right\} \\ &= 3\gamma\,\mathbb{E}\left\{\mathcal{E}(\gamma^{1/2}Z)^2\mathcal{E}'(\gamma^{1/2}Z)\right\} - \gamma\,\mathbb{E}\left\{\mathcal{E}(\gamma^{1/2}Z)^2\right\} - 2\gamma\,\mathbb{E}\left\{(\gamma^{1/2}Z)\mathcal{E}(\gamma^{1/2}Z)\mathcal{E}'(\gamma^{1/2}Z)\right\} \\ &= 3\gamma\mathfrak{s}_2(\gamma) - \gamma r_2(\gamma) - 2\gamma\mathfrak{s}_3(\gamma).\end{aligned}$$

Rearranging proves (106). Further,

$$\begin{aligned}\mathfrak{s}_4(\gamma) &= \mathbb{E}\left\{(\gamma^{1/2}Z)^2\mathcal{E}(\gamma^{1/2}Z)^2\right\} - \mathbb{E}\left\{(\gamma^{1/2}Z)^3\mathcal{E}(\gamma^{1/2}Z)\right\} \\ &= \gamma\,\mathbb{E}\left\{\mathcal{E}(\gamma^{1/2}Z)^2\right\} + 2\gamma\,\mathbb{E}\left\{(\gamma^{1/2}Z)\mathcal{E}(\gamma^{1/2}Z)\mathcal{E}'(\gamma^{1/2}Z)\right\} \\ &\quad - 2\gamma\,\mathbb{E}\left\{(\gamma^{1/2}Z)\mathcal{E}(\gamma^{1/2}Z)\right\} - \gamma\,\mathbb{E}\left\{(\gamma^{1/2}Z)^2\mathcal{E}'(\gamma^{1/2}Z)\right\}.\end{aligned}$$

Integrating by parts again yields

$$\mathbb{E}\left\{(\gamma^{1/2}Z)\mathcal{E}(\gamma^{1/2}Z)\right\} = \gamma\,\mathbb{E}\left\{\mathcal{E}'(\gamma^{1/2}Z)\right\} = \gamma\mathfrak{s}_1(\gamma).$$

So

$$\mathfrak{s}_4(\gamma) = \gamma r_2(\gamma) + 2\gamma\mathfrak{s}_3(\gamma) - 2\gamma^2\mathfrak{s}_1(\gamma) - \gamma\mathfrak{s}_4(\gamma).$$

Rearranging proves (107). Finally,

$$\begin{aligned}\mathfrak{s}_5(\gamma) &= \mathbb{E}\left\{\mathcal{E}(\gamma^{1/2}Z)^4\right\} - 2\,\mathbb{E}\left\{(\gamma^{1/2}Z)\mathcal{E}(\gamma^{1/2}Z)^3\right\} + \mathbb{E}\left\{(\gamma^{1/2}Z)^2\mathcal{E}(\gamma^{1/2}Z)^2\right\} \\ &= \mathbb{E}\left\{\mathcal{E}(\gamma^{1/2}Z)^4\right\} - 6\gamma\,\mathbb{E}\left\{\mathcal{E}(\gamma^{1/2}Z)^2\mathcal{E}'(\gamma^{1/2}Z)\right\} + \gamma\,\mathbb{E}\left\{\mathcal{E}(\gamma^{1/2}Z)^2\right\} \\ &\quad + 2\gamma\,\mathbb{E}\left\{(\gamma^{1/2}Z)\mathcal{E}(\gamma^{1/2}Z)\mathcal{E}'(\gamma^{1/2}Z)\right\} \\ &= r_4(\gamma) - 6\gamma\mathfrak{s}_2(\gamma) + \gamma r_2(\gamma) + 2\gamma\mathfrak{s}_3(\gamma).\end{aligned}$$

Rearranging proves (108). □

Recall from Condition 3.4 that $d_0 = \alpha_\star\,\mathbb{E}[F'_{1-q_0}(q_0^{1/2}Z)]$. As a consequence of (100) and (104), we have

$$d_0 = -\frac{\alpha_\star}{1-q_0}\mathfrak{s}_1(\gamma_0) = -\frac{\alpha_\star}{1-q_0}\cdot\frac{r_2(\gamma_0)}{1+\gamma_0} = -(1-q_0)\psi_0, \tag{109}$$

where we have used that $(1-q_0)(1+\gamma_0) = 1$.

Lemma B.6. *The functions p_4 and r_4 are increasing. Moreover, for any $z > -1$, and m defined in* (101)*, the function $\psi \mapsto m(z,\psi)$ is decreasing.*

Proof. The function p_4 is increasing simply because the maps $\psi \mapsto \mathrm{th}(\psi^{1/2}x)^4$ are pointwise increasing for all $x \in \mathbb{R}$. Similarly, since the maps $\psi \mapsto (z + \mathrm{ch}^2(\psi^{1/2}x))^{-1}$ are pointwise increasing for all $x \in \mathbb{R}$, $z > -1$, the function $\psi \mapsto m(z,\psi)$ is decreasing. Finally,

$$r_4'(\gamma) = \mathbb{E}\left\{6\mathcal{E}(\gamma^{1/2}Z)^2\mathcal{E}'(\gamma^{1/2}Z)^2 + 2\mathcal{E}(\gamma^{1/2}Z)^3\mathcal{E}''(\gamma^{1/2}Z)\right\} \geqslant 0,$$

as Lemma 4.21(c) implies $\mathcal{E}'' > 0$. Thus r_4 is increasing. □

B.2. **Verification of numerical conditions in Theorem 3.6.** Condition 3.1 was proved in [DS18, Proposition 1.3] (recorded as Proposition 3.2). We now verify Conditions 3.3 and 3.4 by proving the following.

Claim B.7. *Condition 3.3 holds for* $\kappa = 0$, *with* $\alpha_\star \mathbb{E}[\mathrm{th}'(\psi_0^{1/2}Z)^2]\,\mathbb{E}[F'_{1-q_0}(q_0^{1/2}Z)^2] \leqslant a_{\mathrm{ub}} \equiv 0.5446$.

Proof. We calculate:

$$\begin{aligned}
\alpha_\star \mathbb{E}[\mathrm{th}'(\psi_0^{1/2}Z)^2]\,\mathbb{E}[F'_{1-q_0}(q_0^{1/2}Z)^2] &= \frac{\alpha_\star}{(1-q_0)^2}\mathfrak{t}(\psi_0)\mathfrak{s}_5(\gamma_0) \\
&\overset{\text{Lem. B.5}}{=} \frac{\alpha_\star}{(1-q_0)^2}(1-2p_2(\psi)+p_4(\psi))\left(\frac{\gamma_0}{1+2\gamma_0}r_2(\gamma_0)+\frac{1-\gamma_0}{(1+2\gamma_0)(1+3\gamma_0)}r_4(\gamma_0)\right) \\
&\overset{(100)}{=} \frac{\alpha_\star}{(1-q_0)^2}(1-2q_0+p_4(\psi))\left(\frac{\gamma_0}{1+2\gamma_0}\cdot\frac{(1-q_0)\psi_0}{\alpha_\star}+\frac{1-\gamma_0}{(1+2\gamma_0)(1+3\gamma_0)}r_4(\gamma_0)\right) \\
&= (1-2q_0+p_4(\psi))\left(\frac{\gamma_0\psi_0}{1+q_0}+\frac{\alpha_\star(1-\gamma_0)}{(1+q_0)(1+2q_0)}r_4(\gamma_0)\right) \\
&\leqslant (1-2q_{\mathrm{lb}}+p_{4,\mathrm{ub}})\left(\frac{\gamma_{\mathrm{ub}}\psi_{\mathrm{ub}}}{1+q_{\mathrm{lb}}}+\frac{\alpha_{\mathrm{lb}}(1-\gamma_{\mathrm{lb}})}{(1+q_{\mathrm{ub}})(1+2q_{\mathrm{ub}})}r_{4,\mathrm{lb}}\right) \overset{(*)}{\leqslant} a_{\mathrm{ub}}.
\end{aligned}$$

The estimate $(*)$ is verified in the attached Python file. We note that this is a simple arithmetic comparison, as all terms are explicitly defined decimal numbers. □

Claim B.8. *Condition 3.4 holds for* $\kappa = 0$, *with* $\lambda(\hat{z}) \leqslant \lambda_{\mathrm{ub}} \equiv -0.1906$.

Proof. Note that for g defined in (102),

$$\begin{aligned}
\lambda(\hat{z}) &= \hat{z} - \alpha_\star g(m(\hat{z}), q_0, \gamma_0) - d_0 \overset{(109)}{=} \hat{z} - \alpha_\star g(m(\hat{z}), q_0, \gamma_0) + (1-q_0)\psi_0 \\
&\leqslant \hat{z} - \alpha_{\mathrm{lb}} g_{\mathrm{lb}} + (1-q_{\mathrm{lb}})\psi_{\mathrm{ub}} \overset{(*)}{\leqslant} \lambda_{\mathrm{ub}}.
\end{aligned}$$

The step $(*)$ is verified in the attached Python file, and is a simple arithmetic comparison of explicitly defined decimal numbers. □

Proof of Theorem 1.2. Follows from Theorem 3.6, Proposition 3.2, and Claims B.7 and B.8. □

B.3. **Local maximality of first moment functional at** $(1, 0)$**.** We next verify Claim 2.6.

Lemma B.9. *For* $\kappa = 0$, *we have*

$$\begin{aligned}
\langle \nabla^2 \overline{\mathcal{S}}_\star(1,0), (u_1,u_2)^{\otimes 2}\rangle = &-\mathbb{E}[(1-\boldsymbol{M}^2)(u_1\dot{\boldsymbol{H}}+u_2\boldsymbol{M})^2] + C_1\,\mathbb{E}[(1-\boldsymbol{M}^2)(u_1\dot{\boldsymbol{H}}+u_2\boldsymbol{M})\dot{\boldsymbol{H}}]^2 \\
&+ C_2\,\mathbb{E}[(1-\boldsymbol{M}^2)(u_1\dot{\boldsymbol{H}}+u_2\boldsymbol{M})\boldsymbol{M}]\,\mathbb{E}[(1-\boldsymbol{M}^2)(u_1\dot{\boldsymbol{H}}+u_2\boldsymbol{M})\dot{\boldsymbol{H}}] \\
&+ C_3\,\mathbb{E}[(1-\boldsymbol{M}^2)(u_1\dot{\boldsymbol{H}}+u_2\boldsymbol{M})\boldsymbol{M}]^2,
\end{aligned}$$

where

$$C_1 = \frac{\alpha_\star}{\psi_0^2}\mathbb{E}\left\{F'_{1-q_0}(\hat{\boldsymbol{H}})\boldsymbol{N}^2\right\}, \qquad C_2 = \frac{2\alpha_\star}{\psi_0}\mathbb{E}\left\{F'_{1-q_0}(\hat{\boldsymbol{H}})\left(\frac{1}{q_0(1-q_0)}\hat{\boldsymbol{H}}+\boldsymbol{N}\right)\boldsymbol{N}\right\}+\frac{2}{1-q_0},$$

$$C_3 = \alpha_\star\,\mathbb{E}\left\{F'_{1-q_0}(\hat{\boldsymbol{H}})\left(\frac{1}{q_0(1-q_0)}\hat{\boldsymbol{H}}+\boldsymbol{N}\right)^2\right\}+\frac{\psi_0}{q_0}.$$

Proof. Analogously to the proof of Lemma 2.5(c), define $\boldsymbol{\Delta}_2 = (u_1\partial_{\lambda_1}+u_2\partial_{\lambda_2})^2\boldsymbol{\Lambda}$. Also abbreviate

$$\boldsymbol{V} = \frac{\kappa - \frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}]}{q_0}\hat{\boldsymbol{H}} - \frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}]}{\psi_0}\boldsymbol{N}}{\sqrt{1-\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}]^2}{q_0}}} + \sqrt{1-q_0}\boldsymbol{N}.$$

We differentiate (87) to obtain

$$\begin{aligned}
&\langle \overline{\mathcal{S}}_\star(\lambda_1,\lambda_2),(u_1,u_2)^{\otimes 2}\rangle = -\mathbb{E}[(u_1\dot{\boldsymbol{H}}+u_2\boldsymbol{M})\boldsymbol{\Delta}] \\
&-\alpha_\star\,\mathbb{E}\left\{\mathcal{E}'(\boldsymbol{V})\left(\frac{-\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]}{q_0}\hat{\boldsymbol{H}}-\frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Delta}]}{\psi_0}\boldsymbol{N}}{\sqrt{1-\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}]^2}{q_0}}}+\frac{\kappa-\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}]}{q_0}\hat{\boldsymbol{H}}-\frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}]}{\psi_0}\boldsymbol{N}}{\left(1-\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}]^2}{q_0}\right)^{3/2}}\cdot\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}]\,\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]}{q_0}\right)^2\right\} \\
&-\alpha_\star\,\mathbb{E}\left\{\mathcal{E}(\boldsymbol{V})\left(\frac{-\frac{2\,\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]}{q_0}\hat{\boldsymbol{H}}-\frac{2\,\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Delta}]}{\psi_0}\boldsymbol{N}}{\left(1-\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}]^2}{q_0}\right)^{3/2}}\cdot\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}]\,\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]}{q_0}\right.\right. \\
&\left.\left.+\frac{\kappa-\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}]}{q_0}\hat{\boldsymbol{H}}-\frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}]}{\psi_0}\boldsymbol{N}}{\left(1-\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}]^2}{q_0}\right)^{5/2}}\cdot\frac{3\,\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}]^2\,\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]^2}{q_0^2}+\frac{\kappa-\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}]}{q_0}\hat{\boldsymbol{H}}-\frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Lambda}]}{\psi_0}\boldsymbol{N}}{\left(1-\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Lambda}]^2}{q_0}\right)^{3/2}}\cdot\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]^2}{q_0}\right)\right\}+f(\boldsymbol{\Delta}_2),
\end{aligned}$$

where $f(\boldsymbol{\Delta}_2)$ is (87) with $\boldsymbol{\Delta}$ replaced by $\boldsymbol{\Delta}_2$. We now specialize to $(\lambda_1,\lambda_2)=(1,0)$. As argued in the proof of Lemma 2.5(c), at $(\lambda_1,\lambda_2)=(1,0)$ we have $f(\boldsymbol{\Delta}_2)=0$. So,

$$\begin{aligned}
&\langle \overline{\mathcal{S}}_\star(1,0),(u_1,u_2)^{\otimes 2}\rangle = -\mathbb{E}[(u_1\dot{\boldsymbol{H}}+u_2\boldsymbol{M})\boldsymbol{\Delta}] \\
&+\alpha_\star\,\mathbb{E}\left\{F'_{1-q_0}(\hat{\boldsymbol{H}})\left(-\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]}{q_0}\hat{\boldsymbol{H}}-\frac{\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Delta}]}{\psi_0}\boldsymbol{N}+\frac{\kappa-\hat{\boldsymbol{H}}-(1-q_0)\boldsymbol{N}}{1-q_0}\cdot\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]\right)^2\right\} \\
&-\alpha_\star\,\mathbb{E}\left\{F_{1-q_0}(\hat{\boldsymbol{H}})\left(\frac{-\frac{2\,\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]}{q_0}\hat{\boldsymbol{H}}-\frac{2\,\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Delta}]}{\psi_0}\boldsymbol{N}}{1-q_0}\cdot\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]\right.\right. \\
&\left.\left.+\frac{\kappa-\hat{\boldsymbol{H}}-(1-q_0)\boldsymbol{N}}{(1-q_0)^2}\cdot 3\,\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]^2+\frac{\kappa-\hat{\boldsymbol{H}}-(1-q_0)\boldsymbol{N}}{1-q_0}\cdot\frac{\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]^2}{q_0}\right)\right\}.
\end{aligned}$$

Specializing further to $\kappa=0$ (which was not used up to here),

$$\begin{aligned}
&\langle \overline{\mathcal{S}}_\star(1,0),(u_1,u_2)^{\otimes 2}\rangle \\
&=-\mathbb{E}[(u_1\dot{\boldsymbol{H}}+u_2\boldsymbol{M})\boldsymbol{\Delta}]+\alpha_\star\,\mathbb{E}\left\{F'_{1-q_0}(\hat{\boldsymbol{H}})\left(\left(\frac{1}{q_0(1-q_0)}\hat{\boldsymbol{H}}+\boldsymbol{N}\right)\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]+\frac{\boldsymbol{N}}{\psi_0}\,\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Delta}]\right)^2\right\} \\
&+\alpha_\star\,\mathbb{E}\left\{\boldsymbol{N}\left(\left(\frac{3}{q_0(1-q_0)^2}\hat{\boldsymbol{H}}+\frac{1+2q_0}{q_0(1-q_0)}\boldsymbol{N}\right)\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]^2+\frac{2}{\psi_0(1-q_0)}\boldsymbol{N}\,\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]\,\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Delta}]\right)\right\}
\end{aligned}$$

Finally, as $\alpha_\star\,\mathbb{E}[\boldsymbol{N}\hat{\boldsymbol{H}}]=q_0d_0=-q_0(1-q_0)\psi_0$ (by (109)) and $\alpha_\star\,\mathbb{E}[\boldsymbol{N}^2]=\psi_0$, the last term simplifies to

$$\frac{\psi_0}{q_0}\,\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]^2+\frac{2}{1-q_0}\,\mathbb{E}[\boldsymbol{M}\boldsymbol{\Delta}]\,\mathbb{E}[\dot{\boldsymbol{H}}\boldsymbol{\Delta}].$$

Expanding $\boldsymbol{\Delta}=(1-\boldsymbol{M}^2)(u_1\dot{\boldsymbol{H}}+u_2\boldsymbol{M})$ concludes the proof. □

Claim B.10. *The following estimates hold.*

(a) $C_1\in[C_{1,\mathrm{lb}},C_{1,\mathrm{ub}}]\equiv[-0.7193,-0.7165]$.
(b) $C_2\in[C_{2,\mathrm{lb}},C_{2,\mathrm{ub}}]\equiv[5.0439,5.0568]$.
(c) $C_3\in[C_{3,\mathrm{lb}},C_{3,\mathrm{ub}}]\equiv[1.1345,1.1526]$.

Proof. We compute using Lemma B.5 and (100):

$$C_1 = \frac{\alpha_\star}{\psi_0^2} \cdot \frac{-\mathfrak{s}_2(\gamma_0)}{(1-q_0)^2} = -\frac{\alpha_\star r_4(\gamma_0)}{\psi_0^2(1-q_0)^2(1+3\gamma_0)} = -\frac{\alpha_\star r_4(\gamma_0)}{\psi_0^2(1-q_0)(1+2q_0)},$$

$$\begin{aligned} C_2 &= \frac{2\alpha_\star}{\psi_0(1-q_0)^2}\left(-\mathfrak{s}_2(\gamma_0) + \frac{\mathfrak{s}_3(\gamma_0)}{q_0}\right) + \frac{2}{1-q_0} \\ &= \frac{2\alpha_\star}{\psi_0(1-q_0)^2}\left(\frac{(2-q_0)(1-q_0)r_4(\gamma_0)}{(1+q_0)(1+2q_0)} - \frac{r_2(\gamma_0)}{1+q_0}\right) + \frac{2}{1-q_0} = \frac{2(2-q_0)\alpha_\star r_4(\gamma_0)}{\psi_0(1-q_0^2)(1+2q_0)} + \frac{2q_0}{1-q_0^2}, \\ C_3 &= -\frac{\alpha_\star}{(1-q_0)^2}\left(\mathfrak{s}_2(\gamma_0) - \frac{2\mathfrak{s}_3(\gamma_0)}{q_0} + \frac{\mathfrak{s}_4(\gamma_0)}{q_0^2}\right) + \frac{\psi_0}{q_0} \\ &= -\frac{\alpha_\star}{(1-q_0)^2}\left(\frac{1-q_0}{1+2q_0} r_4(\gamma_0) - \frac{2q_0-1}{q_0} r_2(\gamma_0)\right) + \frac{\psi_0}{q_0} = -\frac{\alpha_\star r_4(\gamma_0)}{(1-q_0)(1+2q_0)} + \frac{\psi_0}{1-q_0}. \end{aligned}$$

So

$$C_{1,\mathrm{lb}} \overset{(*)}{\leqslant} -\frac{\alpha_{\mathrm{ub}} r_{4,\mathrm{ub}}}{\psi_{\mathrm{lb}}^2(1-q_{\mathrm{ub}})(1+2q_{\mathrm{lb}})} \leqslant C_1 \leqslant -\frac{\alpha_{\mathrm{lb}} r_{4,\mathrm{lb}}}{\psi_{\mathrm{ub}}^2(1-q_{\mathrm{lb}})(1+2q_{\mathrm{ub}})} \overset{(*)}{\leqslant} C_{1,\mathrm{ub}},$$

$$C_{2,\mathrm{lb}} \overset{(*)}{\leqslant} \frac{2(2-q_{\mathrm{ub}})\alpha_{\mathrm{lb}} r_{4,\mathrm{lb}}}{\psi_{\mathrm{ub}}(1-q_{\mathrm{lb}}^2)(1+2q_{\mathrm{ub}})} + \frac{2q_{\mathrm{lb}}}{1-q_{\mathrm{lb}}^2} \leqslant C_2 \leqslant \frac{2(2-q_{\mathrm{lb}})\alpha_{\mathrm{ub}} r_{4,\mathrm{ub}}}{\psi_{\mathrm{lb}}(1-q_{\mathrm{ub}}^2)(1+2q_{\mathrm{lb}})} + \frac{2q_{\mathrm{ub}}}{1-q_{\mathrm{ub}}^2} \overset{(*)}{\leqslant} C_{2,\mathrm{ub}},$$

$$C_{3,\mathrm{lb}} \overset{(*)}{\leqslant} -\frac{\alpha_{\mathrm{ub}} r_{4,\mathrm{ub}}}{(1-q_{\mathrm{ub}})(1+2q_{\mathrm{lb}})} + \frac{\psi_{\mathrm{lb}}}{1-q_{\mathrm{lb}}} \leqslant C_3 \leqslant -\frac{\alpha_{\mathrm{lb}} r_{4,\mathrm{lb}}}{(1-q_{\mathrm{lb}})(1+2q_{\mathrm{ub}})} + \frac{\psi_{\mathrm{ub}}}{1-q_{\mathrm{ub}}} \overset{(*)}{\leqslant} C_{3,\mathrm{ub}}.$$

The steps marked $(*)$ are verified in the attached Python file, and are simple arithmetic comparisons of explicitly defined decimal numbers. □

Claim B.11. *Define* $I_1 = \mathbb{E}[(1-\boldsymbol{M}^2)\dot{\boldsymbol{H}}^2]$, $I_2 = \mathbb{E}[(1-\boldsymbol{M}^2)\dot{\boldsymbol{H}}\boldsymbol{M}]$, $I_3 = \mathbb{E}[(1-\boldsymbol{M}^2)\boldsymbol{M}^2]$. *Then,*

(a) $I_1 \in [I_{1,\mathrm{lb}}, I_{1,\mathrm{ub}}] \equiv [0.24759912, 0.24759923]$.
(b) $I_2 \in [I_{2,\mathrm{lb}}, I_{2,\mathrm{ub}}] \equiv [0.16997315, 0.16997318]$.
(c) $I_3 \in [I_{3,\mathrm{lb}}, I_{3,\mathrm{ub}}] \equiv [0.12335884, 0.12335885]$.

Proof. By repeated integration by parts and (100):

$$I_1 = \psi_0(1-q_0) - 2\psi_0^2(1-4q_0+3p_4(\psi_0)), \quad I_2 = \psi_0(1-4q_0+3p_4(\psi_0)), \quad I_3 = q_0 - p_4(\psi_0).$$

Thus

$$I_{1,\mathrm{lb}} \overset{(*)}{\leqslant} \psi_{\mathrm{lb}}(1-q_{\mathrm{ub}}) - 2\psi_{\mathrm{ub}}^2(1-4q_{\mathrm{lb}}+3p_{4,\mathrm{ub}}) \leqslant I_1 \leqslant \psi_{\mathrm{ub}}(1-q_{\mathrm{lb}}) - 2\psi_{\mathrm{lb}}^2(1-4q_{\mathrm{ub}}+3p_{4,\mathrm{lb}}) \overset{(*)}{\leqslant} I_{1,\mathrm{ub}},$$

$$I_{2,\mathrm{lb}} \overset{(*)}{\leqslant} \psi_{\mathrm{lb}}(1-4q_{\mathrm{ub}}+3p_{4,\mathrm{lb}}) \leqslant I_2 \leqslant \psi_{\mathrm{ub}}(1-4q_{\mathrm{lb}}+3p_{4,\mathrm{ub}}) \overset{(*)}{\leqslant} I_{2,\mathrm{ub}},$$

$$I_{3,\mathrm{lb}} \overset{(*)}{\leqslant} q_{\mathrm{lb}} - p_{4,\mathrm{ub}} \leqslant I_3 \leqslant q_{\mathrm{ub}} - p_{4,\mathrm{lb}} \overset{(*)}{\leqslant} I_{3,\mathrm{ub}}.$$

The steps marked $(*)$ are verified in the attached Python file, and are simple arithmetic comparisons of explicitly defined decimal numbers. □

Claim B.12. *Let* $M = \nabla^2\overline{\mathcal{S}}_\star(1,0)$. *The following estimates hold.*

(a) $M_{1,1} \leqslant M_{1,1,\mathrm{ub}} \equiv -0.045408$.
(b) $M_{2,2} \leqslant M_{2,2,\mathrm{ub}} \equiv -0.020490$.
(c) $M_{1,2} \in [M_{1,2,\mathrm{lb}}, M_{1,2,\mathrm{ub}}] \equiv [-0.025685, -0.026567]$.
(d) $\det(M) \geqslant M_{\det,\mathrm{lb}} \equiv 0.0002246$.

Proof. By Lemma B.9,

$$M_{1,1} = -I_1 + C_1 I_1^2 + C_2 I_1 I_2 + C_3 I_2^2,$$
$$M_{1,2} = -I_2 + C_1 I_1 I_2 + \frac{1}{2} C_2 (I_2^2 + I_1 I_3) + C_3 I_2 I_3,$$
$$M_{2,2} = -I_3 + C_1 I_2^2 + C_2 I_2 I_3 + C_3 I_3^2.$$

Estimating with Claims B.10 and B.11, we find

$$M_{1,1} \leqslant -I_{1,\mathrm{lb}} + C_{1,\mathrm{ub}} I_{1,\mathrm{lb}}^2 + C_{2,\mathrm{ub}} I_{1,\mathrm{ub}} I_{2,\mathrm{ub}} + C_{3,\mathrm{ub}} I_{2,\mathrm{ub}}^2 \overset{(*)}{\leqslant} M_{1,1,\mathrm{ub}},$$
$$M_{2,2} \leqslant -I_{3,\mathrm{lb}} + C_{1,\mathrm{ub}} I_{2,\mathrm{lb}}^2 + C_{2,\mathrm{ub}} I_{2,\mathrm{ub}} I_{3,\mathrm{ub}} + C_{3,\mathrm{ub}} I_{3,\mathrm{ub}}^2 \overset{(*)}{\leqslant} M_{2,2,\mathrm{ub}},$$
$$M_{1,2} \leqslant -I_{2,\mathrm{lb}} + C_{1,\mathrm{ub}} I_{1,\mathrm{lb}} I_{2,\mathrm{lb}} + \frac{1}{2} C_{2,\mathrm{ub}} (I_{2,\mathrm{ub}}^2 + I_{1,\mathrm{ub}} I_{3,\mathrm{ub}}) + C_{3,\mathrm{ub}} I_{2,\mathrm{ub}} I_{3,\mathrm{ub}} \overset{(*)}{\leqslant} M_{1,2,\mathrm{ub}},$$
$$M_{1,2} \geqslant -I_{2,\mathrm{ub}} + C_{1,\mathrm{lb}} I_{1,\mathrm{ub}} I_{2,\mathrm{ub}} + \frac{1}{2} C_{2,\mathrm{lb}} (I_{2,\mathrm{lb}}^2 + I_{1,\mathrm{lb}} I_{3,\mathrm{lb}}) + C_{3,\mathrm{lb}} I_{2,\mathrm{lb}} I_{3,\mathrm{lb}} \overset{(*)}{\geqslant} M_{1,2,\mathrm{lb}}.$$

The steps marked $(*)$ are verified in the attached Python file, and are simple arithmetic comparisons of explicitly defined decimal numbers. This proves parts (a), (b), and (c). Finally,

$$\det(M) = M_{1,1} M_{2,2} - M_{1,2}^2 \geqslant M_{1,1,\mathrm{ub}} M_{2,2,\mathrm{ub}} - M_{1,2,\mathrm{lb}}^2 \overset{(*)}{\geqslant} M_{\det,\mathrm{lb}},$$

where the step $(*)$ is verified in the attached Python file. This proves part (d). □

Proof of Claim 2.6. Follows from Claim B.12, which implies $M_{1,1}, M_{2,2} < 0$ and $\det(M) > 0$. □

B.4. **Interval arithmetic estimates.** We now describe the computer-assisted proofs of Claims B.1, B.2, B.3, and B.4. We begin with the more straightforward Claims B.1 and B.3.

Proof of Claim B.1. We first show the upper bound. Set $L = 10$. Since th^4 takes values in $[0, 1]$,

$$p_4(\psi_0) \overset{\text{Lem. B.6}}{\leqslant} p_4(\psi_{\mathrm{ub}}) \leqslant \mathbb{E}[\mathrm{th}^4(\psi_{\mathrm{ub}}^{1/2} Z) \mathbf{1}\{|Z| \leqslant L\}] + \mathbb{P}[|Z| \geqslant L]$$
$$\leqslant \int_{-L}^{L} \mathrm{th}^4(\psi_{\mathrm{ub}}^{1/2} x) \varphi(x) \, \mathrm{d}x + 2e^{-L^2/2} \overset{(*)}{\leqslant} p_{4,\mathrm{ub}},$$

where the step $(*)$ is verified in the attached Python file. Similarly,

$$p_4(\psi_0) \overset{\text{Lem. B.6}}{\geqslant} p_4(\psi_{\mathrm{lb}}) \geqslant \mathbb{E}[\mathrm{th}^4(\psi_{\mathrm{lb}}^{1/2} Z) \mathbf{1}\{|Z| \leqslant L\}] = \int_{-L}^{L} \mathrm{th}^4(\psi_{\mathrm{lb}}^{1/2} x) \varphi(x) \, \mathrm{d}x \overset{(*)}{\geqslant} p_{4,\mathrm{lb}},$$

where the step $(*)$ is verified in the attached Python file. □

Proof of Claim B.3. Let $L = 10$. Note that for any $x \in \mathbb{R}$, $(\widehat{z} + \mathrm{ch}^2(x))^{-1} \leqslant (1 + \widehat{z})^{-1}$. Then,

$$m(\widehat{z}) = m(\widehat{z}, \psi_0) \overset{\text{Lem. B.6}}{\leqslant} m(\widehat{z}, \psi_{\mathrm{lb}}) \leqslant \mathbb{E}[(\widehat{z} + \mathrm{ch}^2(\psi_{\mathrm{lb}}^{1/2} Z))^{-1} \mathbf{1}\{|Z| \leqslant L\}] + (1 + \widehat{z})^{-1} \, \mathbb{P}[|Z| \geqslant L]$$
$$\leqslant \int_{-L}^{L} (\widehat{z} + \mathrm{ch}^2(\psi_{\mathrm{lb}}^{1/2} x))^{-1} \varphi(x) \, \mathrm{d}x + 2(1 + \widehat{z})^{-1} e^{-L^2/2} \overset{(*)}{\leqslant} m_{\mathrm{ub}},$$

where the step $(*)$ is verified in the attached Python file. □

Claims B.2 and B.4 will involve integrating functions that involve $\mathcal{E}$ against the gaussian measure. This is more challenging because $\mathcal{E}$ is itself defined in terms of an integral, which makes these claims less amenable to numerical integration. We take a cruder approach of discretizing these integrals into small intervals, and bounding the integral on each small interval using monotonicity properties of $\mathcal{E}$ and $\mathcal{E}'$.

Proof of Claim B.2. Let $L = 8$, $\delta = 10^{-3}$, and $J = L/\delta$. For integer $j \in [-J, J]$, let $x_j = j\delta$. Then,

$$r_4(\gamma_0) \overset{Lem.\ B.6}{\leqslant} r_4(\gamma_{\rm ub}) = \sum_{j=-J}^{J-1} \mathbb{E}\left\{\mathcal{E}(\gamma_{\rm ub}^{1/2} Z)^4 \mathbf{1}\{Z \in [x_j, x_{j+1}]\}\right\} + \mathbb{E}\left\{\mathcal{E}(\gamma_{\rm ub}^{1/2} Z)^4 \mathbf{1}\{|Z| \geqslant L\}\right\}.$$

These terms can be bounded as follows. Since $\mathcal{E}$ is nonnegative and increasing (by Lemma 4.21(a)(b)),

$$\mathbb{E}\left\{\mathcal{E}(\gamma_{\rm ub}^{1/2} Z)^4 \mathbf{1}\{Z \in [x_j, x_{j+1}]\}\right\} \leqslant \mathcal{E}(\gamma_{\rm ub}^{1/2} x_{j+1})^4\, \mathbb{P}[Z \in [x_j, x_{j+1}]],$$

and this probability is bounded above by $\delta\varphi(x_{j+1})$ if $j \leqslant -1$ and $\delta\varphi(x_j)$ if $j \geqslant 0$. We estimate the tail term using Cauchy–Schwarz:

$$\mathbb{E}\left\{\mathcal{E}(\gamma_{\rm ub}^{1/2} Z)^4 \mathbf{1}\{|Z| \geqslant L\}\right\} \leqslant \mathbb{E}\left\{\mathcal{E}(\gamma_{\rm ub}^{1/2} Z)^8\right\}^{1/2} \mathbb{P}[|Z| \geqslant L]^{1/2}.$$

The probability is bounded by $2e^{-L^2/2}$. For the remaining expectation, recall from Lemma 4.21(a) that $0 \leqslant \mathcal{E}(x) \leqslant |x| + 1$. So,

$$\mathbb{E}\left\{\mathcal{E}(\gamma_{\rm ub}^{1/2} Z)^8\right\} \leqslant \mathbb{E}\left\{(1 + \gamma_{\rm ub}^{1/2}|Z|)^8\right\} \leqslant 2^7\, \mathbb{E}\left\{1 + \gamma_{\rm ub}^4 Z^8\right\} = 2^7(1 + 105\gamma_{\rm ub}^2).$$

Combining these estimates yields

$$\mathbb{E}\left\{\mathcal{E}(\gamma_{\rm ub}^{1/2} Z)^4 \mathbf{1}\{|Z| \geqslant L\}\right\} \leqslant 2^4(1 + 105\gamma_{\rm ub}^2)^{1/2} e^{-L^2/4} \leqslant 2^4(1 + 11\gamma_{\rm ub}) e^{-L^2/4}.$$

All in all,

$$r_4(\gamma_0) \leqslant \delta \sum_{j=-J}^{-1} \mathcal{E}(\gamma_{\rm ub}^{1/2} x_{j+1})^4 \varphi(x_{j+1}) + \delta \sum_{j=0}^{J-1} \mathcal{E}(\gamma_{\rm ub}^{1/2} x_{j+1})^4 \varphi(x_j) + 2^4(1 + 11\gamma_{\rm ub}) e^{-L^2/4} \overset{(*)}{\leqslant} r_{4,\rm ub},$$

where the step $(*)$ is verified in the attached Python file. (See Remark B.13 below for how the function $\mathcal{E}$ is evaluated numerically). For the lower bound, we similarly have

$$\begin{aligned} r_4(\gamma_0) \overset{Lem.\ B.6}{\geqslant} r_4(\gamma_{\rm lb}) &= \sum_{j=-J}^{J-1} \mathbb{E}\left\{\mathcal{E}(\gamma_{\rm lb}^{1/2} Z)^4 \mathbf{1}\{Z \in [x_j, x_{j+1}]\}\right\} \\ &\geqslant \delta \sum_{j=-J}^{-1} \mathcal{E}(\gamma_{\rm lb}^{1/2} x_j)^4 \varphi(x_j) + \delta \sum_{j=0}^{J-1} \mathcal{E}(\gamma_{\rm lb}^{1/2} x_j)^4 \varphi(x_{j+1}) \overset{(*)}{\geqslant} r_{4,\rm lb}, \end{aligned}$$

where the step $(*)$ is verified in the attached Python file. ☐

Remark B.13. The above computer-assisted proof requires evaluating the function $\mathcal{E}(x) = \varphi(x)/\Psi(x)$, where $\Psi(x) = \mathbb{P}[Z \geqslant x]$ is itself an integral. We evaluate this as follows. Note that the inputs x on which we numerically evaluate $\mathcal{E}$ are bounded above by $\gamma_{\rm ub}^{1/2} L \leqslant 10$. Define $L_+ = 12$. We estimate

$$\mathcal{E}(x)^{-1} = \int_x^{L_+} \frac{\varphi(y)}{\varphi(x)}\, \mathrm{d}y + \frac{\mathbb{P}[Z \geqslant L_+]}{\varphi(x)} \leqslant \int_x^{L_+} e^{-(y^2 - x^2)/2}\, \mathrm{d}y + \frac{e^{-(L_+^2 - x^2)/2}}{\sqrt{2\pi}}$$

and

$$\mathcal{E}(x)^{-1} \geqslant \int_x^{L_+} \frac{\varphi(y)}{\varphi(x)}\, \mathrm{d}y = \int_x^{L_+} e^{-(y^2 - x^2)/2}\, \mathrm{d}y.$$

The remaining integral can be rigorously bounded by numerical integration, and for $x \leqslant 10$ the term $e^{-(L_+^2 - x^2)/2}/\sqrt{2\pi}$ will contribute an error that is multiplicatively small.

Finally, we turn to Claim B.4. By Lemma 4.21(b), $\mathcal{E}'$ takes values in $(0,1)$. Thus the function g defined in (102) is decreasing in m and increasing in q, and

$$g(m(\widehat{z}), q_0, \gamma_0) \geqslant g(m_{\mathrm{ub}}, q_{\mathrm{lb}}, \gamma_0). \tag{110}$$

However, g is not clearly monotone in γ, so we instead control the derivative of g in γ.

Lemma B.14. *Let $\widetilde{g}(\gamma) = g(m_{\mathrm{ub}}, q_{\mathrm{lb}}, \gamma)$. Then, for all $\gamma \geqslant 0$, $|\widetilde{g}'(\gamma)| \leqslant 20$.*

Proof. We write $\widetilde{g}(\gamma) = \mathbb{E}[\widehat{g}(\gamma^{1/2}Z)]$, where

$$\widehat{g}(x) = \frac{\mathcal{E}'(x)}{(1-q_{\mathrm{lb}})(1-\mathcal{E}'(x)) + m_{\mathrm{ub}}\mathcal{E}'(x)}. \tag{111}$$

A straightforward calculation shows that

$$\widehat{g}''(x) = \frac{(1-q_{\mathrm{lb}})\mathcal{E}^{(3)}(x)}{((1-q_{\mathrm{lb}})(1-\mathcal{E}'(x)) + m_{\mathrm{ub}}\mathcal{E}'(x))^2} - \frac{2(1-q_{\mathrm{lb}})(m_{\mathrm{ub}}+q_{\mathrm{lb}}-1)\mathcal{E}''(x)^2}{(((1-q_{\mathrm{lb}})(1-\mathcal{E}'(x)) + m_{\mathrm{ub}}\mathcal{E}'(x))^2)^3}.$$

Since $\mathcal{E}'(x) \in (0,1)$ by Lemma 4.21(b),

$$(1-q_{\mathrm{lb}})(1-\mathcal{E}'(x)) + m_{\mathrm{ub}}\mathcal{E}'(x) \geqslant \min(1-q_{\mathrm{lb}}, m_{\mathrm{ub}}) = 1-q_{\mathrm{lb}}.$$

Lemma 4.21(c)(d) yields $|\mathcal{E}''(x)| \leqslant 1$, $|\mathcal{E}^{(3)}(x)| \leqslant 13$. Thus

$$|\widehat{g}''(x)| \leqslant \frac{13}{1-q_{\mathrm{lb}}} + \frac{2(m_{\mathrm{ub}}+q_{\mathrm{lb}}-1)}{(1-q_{\mathrm{lb}})^2} \leqslant 40,$$

where the final estimate follows from the simple bounds $q_{\mathrm{lb}} \leqslant 3/5$, $m_{\mathrm{ub}} \leqslant 1$. Finally, a gaussian integration by parts calculation yields

$$\widetilde{g}'(\gamma) = \frac{1}{2}\mathbb{E}[\widehat{g}''(\gamma^{1/2}Z)],$$

which implies the result. □

Proof of Claim B.4. In light of (110) and Lemma B.14, we will estimate

$$g(m(\widehat{z}), q_0, \gamma_0) \geqslant g(m_{\mathrm{ub}}, q_{\mathrm{lb}}, \gamma_{\mathrm{lb}}) - 20|\gamma_{\mathrm{ub}} - \gamma_{\mathrm{lb}}|.$$

We will estimate $g(m_{\mathrm{ub}}, q_{\mathrm{lb}}, \gamma_{\mathrm{lb}})$ by discretization, like in the proof of CLaim B.2. Let $L = 8$, $\delta = 10^{-3}$, and $J = L/\delta$. For integer $j \in [-J, J]$, let $x_j = j\delta$.

Note that $\widehat{g}(x)$ defined in (111) takes positive values, and is an increasing function of $\mathcal{E}'(x)$. Moreover, by Lemma 4.21(c), $\mathcal{E}'(x)$ is an increasing function of x. Thus $\widehat{g}(x)$ is an increasing function of x. Hence,

$$\begin{aligned} g(m_{\mathrm{ub}}, q_{\mathrm{lb}}, \gamma_{\mathrm{lb}}) = \mathbb{E}[\widehat{g}(\gamma_{\mathrm{lb}}^{1/2}Z)] &\geqslant \sum_{j=-J}^{J-1} \mathbb{E}[\widehat{g}(\gamma_{\mathrm{lb}}^{1/2}Z)\mathbf{1}\{Z \in [x_j, x_{j+1}]\}] \\ &\geqslant \delta \sum_{j=-J}^{-1} \widehat{g}(\gamma_{\mathrm{lb}}^{1/2}x_j)\varphi(x_j) + \delta \sum_{j=0}^{J-1} \widehat{g}(\gamma_{\mathrm{lb}}^{1/2}x_j)\varphi(x_{j+1}). \end{aligned}$$

Combining the above,

$$g(m(\widehat{z}), q_0, \gamma_0) \geqslant \delta \sum_{j=-J}^{-1} \widehat{g}(\gamma_{\mathrm{lb}}^{1/2}x_j)\varphi(x_j) + \delta \sum_{j=0}^{J-1} \widehat{g}(\gamma_{\mathrm{lb}}^{1/2}x_j)\varphi(x_{j+1}) - 20|\gamma_{\mathrm{ub}} - \gamma_{\mathrm{lb}}| \overset{(*)}{\geqslant} g_{\mathrm{lb}},$$

where the step $(*)$ is verified in the attached Python file. We numerically evaluate $\widehat{g}$ using the identity $\mathcal{E}'(x) = \mathcal{E}(x)(\mathcal{E}(x) - x)$ (Lemma 4.21(b)), evaluating $\mathcal{E}$ as in Remark B.13. □