Optimizing mean field spin glasses with external field

Mark Sellke[*]

Abstract

We consider the Hamiltonians of mean-field spin glasses, which are certain random functions H_N defined on high-dimensional cubes or spheres in $\mathbb{R}^N$. The asymptotic maximum values of these functions were famously obtained by Talagrand and later by Panchenko and by Chen. The landscape of approximate maxima of H_N is described by various forms of replica symmetry breaking exhibiting a broad range of behaviors. We study the problem of efficiently computing an approximate maximizer of H_N.

We give a two-phase message passing algorithm to approximately maximize H_N when a no overlap gap condition holds. This generalizes the recent works [Sub21, Mon19, AMS21] by allowing a non-trivial external field. For even Ising spin glasses with constant external field, our algorithm succeeds exactly when existing methods fail to rule out approximate maximization for a wide class of algorithms. Moreover we give a branching variant of our algorithm which constructs a full ultrametric tree of approximate maxima.

Contents

[*]Department of Statistics, Harvard University, Cambridge, MA, USA. E-mail: msellke@fas.harvard.edu

1 Introduction

Optimizing non-convex functions in high dimensions is well-known to be computationally intractible in general. In this paper we study the optimization of a natural class of *random* non-convex functions, namely the Hamiltonians of mean-field spin glasses. These functions H_N are defined on either the cube $\Sigma_N = \{-1, 1\}^N$ or the sphere $\mathbb{S}^{N-1}(\sqrt{N})$ of radius $\sqrt{N}$ and have been studied since [SK75] as models for the behavior of disordered magnetic systems.

The distribution of an N-dimensional mean-field spin glass Hamiltonian H_N is described by an exponentially decaying sequence $(c_p)_{p \geq 2}$ of non-negative real numbers as well as an external field probability distribution $\mathcal{L}_h$ on $\mathbb{R}$ with finite second moment. Given these data, one samples $h_1, \ldots, h_N \sim \mathcal{L}_h$ and standard Gaussians $g_{i_1, \ldots, i_p} \sim \mathsf{N}(0, 1)$ and then defines $H_N : \mathbb{R}^N \to \mathbb{R}$ by

$$H_N(\boldsymbol{x}) = \sum_i h_i x_i + \widetilde{H}_N(\boldsymbol{x}),$$

$$\widetilde{H}_N(\boldsymbol{x}) = \sum_{p=2}^{\infty} \frac{c_p}{N^{(p-1)/2}} \sum_{i_1, \ldots, i_p = 1}^{N} g_{i_1, \ldots, i_p} x_{i_1} \ldots x_{i_p}.$$

The distribution of the non-linear part $\widetilde{H}_N$ is characterized by the mixture function $\xi(z) = \sum_{p \geq 2} c_p^2 z^p$ – there are no issues of convergence for $|z| \leq 1 + \eta$ thanks to the exponential decay assumption. We assume throughout that ξ is not the zero function so that we study a genuine spin glass. $\widetilde{H}_N$ is then a centered Gaussian process with covariance

$$\mathbb{E}\big[\widetilde{H}_N(\boldsymbol{x}_1)\widetilde{H}_N(\boldsymbol{x}_2)\big] = N\xi\left(\frac{\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle}{N}\right).$$

Spin glasses were introduced to model the magnetic properties of diluted materials and have been studied in statistical physics and probability since the seminal work [SK75]. In this context, the object of study is the Gibbs measure $\frac{e^{\beta H_N(\boldsymbol{x})} d\mu(\boldsymbol{x})}{Z_{N,\beta}}$ where $\beta > 0$ is the inverse-temperature, $\mu(\boldsymbol{x})$ is a fixed reference measure and $Z_{N,\beta}$ is a random normalizing constant known as the partition function. The most common choice is to take $\mu(\cdot)$ the uniform measure on $\boldsymbol{\Sigma}_N = \{-1, 1\}^N$, and another canonical choice is the uniform measure on $\mathbb{S}^{N-1}(\sqrt{N})$. These two choices define *Ising* and *spherical* spin glasses. The quantity of primary interest is the free energy

$$F_N(\beta) = \log \mathbb{E}^{\boldsymbol{x} \sim \mu}\big[e^{\beta H_N(\boldsymbol{x})}\big].$$

The in-probability normalized limit $F(\beta) = \text{p-lim}_{N \to \infty} \frac{F_N(\beta)}{N}$ of the free energy at temperature β is famously given by an infinite-dimensional variational problem known as the Parisi formula (or the Cristanti-Sommers formula in the spherical case) as we review in the next section. These free energies are well-concentrated and taking a second limit $\lim_{\beta \to \infty} \frac{F(\beta)}{\beta}$ yields the asymptotic ground state energies

$$GS(\xi, \mathcal{L}_h) = \underset{N \to \infty}{\text{p-lim}} \max_{\boldsymbol{x} \in \boldsymbol{\Sigma}_N} \frac{H_N(\boldsymbol{x})}{N},$$

$$GS_{\mathsf{sph}}(\xi, \mathcal{L}_h) = \underset{N \to \infty}{\text{p-lim}} \max_{\boldsymbol{x} \in \mathbb{S}^{N-1}(\sqrt{N})} \frac{H_N(\boldsymbol{x})}{N}.$$

From the point of view of optimization, spin glass Hamiltonians serve as natural examples of highly non-convex functions. Indeed, the landscape of H_N can exhibit quite complicated behavior. For instance H_N may have exponentially many near-maxima on Σ_N [Cha09, DEZ+15, CHL18]. The structure of these near-maxima is highly nontrivial; the Gibbs measures on Σ_N are approximate ultrametrics in a certain sense, at least in the so-called generic models [Jag17, CS21]. Moreover spherical spin glasses typically have exponentially many *local* maxima and saddle points, which are natural barriers to gradient descent and similar optimization algorithms [ABA13, Auf13, Sub17, AMMN19]. The utility of a rich model of random functions is made clear by a comparison to the theory of high-dimensional non-convex optimization in the worst-case setting. In the black-box model of optimization based on querying function values, gradients, and Hessians, approximately optimizing an unknown non-convex function in high-dimension efficiently is trivially impossible and substantial effort has gone towards the more modest task of finding a local optimum or stationary point [CDHS17, AAZB+17, CDHS18, CDHS19, JNG+21]. Even for quadratic polynomials in N variables, it is quasi-NP hard to reach within a factor $\log(N)^\varepsilon$ of the optimum [ABE+05]. For polynomials of degree $p \geq 3$ on the sphere, [BBH+12] proves that even an approximation ratio $e^{(\log N)^\varepsilon}$ is computationally infeasible to obtain.

Despite the worst-case obstructions just outlined, a series of recent works have found great success in approximately maximizing certain spin glass Hamiltonians. By *approximate maximization* we always mean maximization up to a factor $(1 + \varepsilon)$, where $\varepsilon > 0$ is an arbitrarily small positive constant; we similarly refer to a point $\boldsymbol{x} \in \Sigma_N$ or $\boldsymbol{x} \in \mathbb{S}^{N-1}(\sqrt{N})$ achieving such a nearly optimal value as an *approximate maximizer* (where the small constant ε is implicit). Subag showed in [Sub21] how to approximately maximize spherical spin glasses by using top eigenvectors of the Hessian $\nabla^2 H_N$. Subsequently [Mon19, AMS21] developed a message passing algorithm with similar guarantees for the Ising case. These works all operate under an assumption of no overlap gap, a condition which is expected (known in the spherical setting) to hold for some but not all models $(\xi, \mathcal{L}_h)$ – otherwise they achieve an explicit, suboptimal energy value. Such a no overlap gap assumption is expected to be necessary to find approximate maxima efficiently. Indeed, the works [AJ18, GJ21, GJW20] rule out various algorithms for optimizing spin glasses when an overlap gap holds. Variants of the overlap gap property have been shown to rule out $(1 + \varepsilon)$-approximation by certain classes of algorithms for random optimization problems on sparse graphs [MMZ05, ACORT11, GS14, RV17, GS17, CGP+19, Wei22]. Overlap gaps have also been proposed as evidence of computational hardness for a range of statistical tasks including planted clique, planted dense submatrix, sparse regression, and sparse principal component analysis [GZ17, GL18, GJS21, GZ19, AWZ20]. We discuss overlap gaps more in Subsection 1.2 and Section 6.

The aforementioned algorithms in [Sub21, Mon19, AMS21] are all restricted to settings with no external field, i.e. with $h_i = 0$ for all i. Our main purpose is to generalize these results to allow for an external field. We focus primarily on the Ising case and explain in Section 5 how to handle the easier spherical models. Our main algorithm consists of two stages of message passing. The first stage is inspired by the work [Bol14] which constructs solutions to the TAP equations for the SK model at high temperature. We construct approximate solutions to the generalized TAP equations of [Sub18, CPS21a, CPS21b], which heuristically amounts to locating the root of the ultrametric tree of approximate maxima. The second stage is similar to [Mon19, AMS21] and uses incremental approximate message passing to descend the ultrametric tree by simulating the SDE corresponding to a candidate solution for the Parisi variational problem. A related two-stage message passing algorithm was recently introduced in our

joint work with A.E. Alaoui on the negative spherical perceptron [AS22].

While the primary goal in this line of work is to construct a single approximate maximizer, Subag beautifully observed in [Sub21, Remark 6] that an extension of his Hessian-based construction for spherical models produces approximate maximizers arranged into a completely arbitary ultrametric space obeying an obvious diameter upper bound. The overlap gap property essentially states that distances between approximate maximizers cannot take certain values, and so this is a sort of constructive converse result. In Section 4 we give a branching version of our main algorithm, following a suggestion of [AM20], which constructs an arbitrary ultrametric space of approximate maximizers in the Ising case (again subject to a diameter upper bound).

1.1 Optimizing Ising spin glasses

To state our results we require the Parisi formula for the ground state of a mean field Ising spin glass as given in [AC17]. Let $\mathscr{U}$ be the function space

$$\mathscr{U} = \left\{ \gamma : [0,1) \to [0,\infty) : \gamma \text{ is non-decreasing}, \int_0^1 \gamma(t)\mathrm{d}t < \infty \right\}.$$

The functions γ are meant to correspond to cumulative distribution functions – for finite β the corresponding Parisi formula requires $\gamma(1) = 1$, but this constraint disappears in renormalizing to obtain a zero-temperature limit. For $\gamma \in \mathscr{U}$ we take $\Phi_\gamma(t,x) : [0,1] \times \mathbb{R} \to \mathbb{R}$ to be the solution of the following Parisi PDE:

$$\partial_t \Phi_\gamma(t,x) + \frac{1}{2}\xi''(t)\big(\partial_{xx}\Phi_\gamma(t,x) + \gamma(t)\big(\partial_x\Phi_\gamma(t,x)\big)^2\big) = 0,$$
$$\Phi_\gamma(1,x) = |x|.$$

This PDE is known to be well-posed, see Proposition 2.6. Intimately related to the above PDE is the stochastic differential equation

$$\mathrm{d}X_t = \xi''(t)\gamma(t)\partial_x\Phi_\gamma(t,X_t)\mathrm{d}t + \sqrt{\xi''(t)}\mathrm{d}B_t, \quad X_0 \sim \mathcal{L}_h. \tag{1.1}$$

which we call the Parisi SDE. The Parisi functional $\mathsf{P} : \mathscr{U} \to \mathbb{R}$ with external field distribution $\mathcal{L}_h$ is given by:

$$\mathsf{P}_{\xi,\mathcal{L}_h}(\gamma) = \mathbb{E}^{h \sim \mathcal{L}_h}\big[\Phi_\gamma(0,h)\big] - \frac{1}{2}\int_0^1 t\xi''(t)\gamma(t)\mathrm{d}t. \tag{1.2}$$

The Parisi formula for the ground state energy is as follows.

Theorem 1.1 ([Tal06b, Pan14, AC17, CHL18])**.**

$$GS(\xi,\mathcal{L}_h) = \inf_{\gamma \in \mathscr{U}} \mathsf{P}_{\xi,\mathcal{L}_h}(\gamma).$$

Moreover the minimum is attained at a unique $\gamma_*^{\mathscr{U}} \in \mathscr{U}$.

Through the paper, $\gamma_*^{\mathscr{U}}$ will always refer to the minimizer of Theorem 1.1. We now turn to algorithms. In [Mon19], Montanari introduced the class of *incremental approximate message passing* (IAMP) algorithms to optimize the SK model. These are a special form of the well-studied approximate message passing (AMP) algorithms, reviewed in Subsection 2.1. The work [AMS21] showed that the maximum asympototic value of H_N achievable by IAMP algorithms is given by the minimizer of P, assuming it exists, over a larger class of non-monotone functions, when $\mathcal{L}_h = \delta_0$ so there is no external field. This

larger class is:

$$\mathscr{L} = \left\{ \gamma : [0,1) \to [0,\infty) : \gamma \text{ is right-continuous}, \|\xi'' \cdot \gamma\|_{TV[0,t]} < \infty \forall t \in [0,1), \right.$$
$$\left. \int_0^1 \xi''(t)\gamma(t)\mathrm{d}t < \infty \right\}.$$

Here $TV[0,t]$ denotes the total variation norm

$$\|f\|_{TV[0,t]} \equiv \sup_n \sup_{0 \le t_0 < t_1 < \cdots < t_k \le t} \sum_{i=1}^k |f(t_i) - f(t_{i-1})|.$$

The Parisi PDE (1.2) and associated SDE extend also to $\mathscr{L}$. We denote by $\gamma_*^{\mathscr{L}} \in \mathscr{L}$ the minimizer of P over $\mathscr{L}$, assuming that it exists. Note that uniqueness always holds by Lemma 1.6 below. We remark that [AMS21] does not include the right-continuity constraint in defining $\mathscr{L}$, however this constraint only serves to eliminate ambiguities between γ_1, γ_2 differing on sets of measure 0. In fact [AMS21] assumes $\gamma \in \mathscr{L}$ is right-continuous from Lemma 6.9 onward.

Theorem 1.2 ([AMS21, Theorem 3]). *With $\mathcal{L}_h = \delta_0$, suppose $\inf_{\gamma \in \mathscr{L}} \mathsf{P}(\gamma)$ is achieved at $\gamma_*^{\mathscr{L}} \in \mathscr{L}$. Then for any $\varepsilon > 0$ there exists an efficient AMP algorithm which outputs $\boldsymbol{\sigma} \in \Sigma_N$ satisfying*

$$\frac{H_N(\boldsymbol{\sigma})}{N} \ge \mathsf{P}(\gamma_*^{\mathscr{L}}) - \varepsilon \tag{1.3}$$

with probability tending to 1 as $N \to \infty$.

We clarify our use of the word "efficient" in Subsection 2.1 – in short, it means that $O_\varepsilon(1)$ evaluations of $\nabla \widetilde{H}_N$ and first/second partial derivatives of $\Phi_{\gamma_*^{\mathscr{L}}}$ are required. In general, minimizing over the larger space $\mathscr{L}$ instead of $\mathscr{U}$ may decrease the infimum value of P, so that IAMP algorithms fail to approximately maximize H_N. However if $\gamma_*^{\mathscr{U}}$ is strictly increasing, then the infima are equal.

Corollary 1.3 ([AMS21, Corollary 2.2]). *With $\mathcal{L}_h = \delta_0$, suppose that $\gamma_*^{\mathscr{U}}$ is strictly increasing on $[0,1)$. Then $\gamma_*^{\mathscr{U}} = \gamma_*^{\mathscr{L}}$. Consequently for any $\varepsilon > 0$ there is an efficient AMP algorithm which outputs $\boldsymbol{\sigma} \in \Sigma_N$ satisfying*

$$\frac{H_N(\boldsymbol{\sigma})}{N} \ge GS(\xi, \mathcal{L}_h) - \varepsilon \tag{1.4}$$

with probability tending to 1 as $N \to \infty$.

We define the support $\mathrm{supp}(\gamma)$ of $\gamma \in \mathscr{L}$ to be the closure in $[0,1)$ of $S(\gamma) \equiv \{x \in [0,1) : \gamma(x) > 0\}$. Note that this is not the same as the support of the signed measure with CDF γ.

We now present our new results when there is a non-trivial external field distribution $\mathcal{L}_h \ne \delta_0$. The following proposition shows that this forces $\gamma_*^{\mathscr{U}}(t) = 0$ in a neighborhood of $t = 0$, hence Corollary 1.3 cannot apply. The proof is exactly the same as [Pan, Lemma A.19] (which is the same result for positive temperature).

Proposition 1.4. *We have $0 \in \mathrm{supp}(\gamma_*^{\mathscr{U}})$ if and only if $\mathcal{L}_h = \delta_0$.*

Despite this, we will show that approximate maximization is still possible with an external field if $\gamma_*^{\mathscr{U}}$ is strictly increasing on $[\underline{q}, 1)$ for $\underline{q} = \inf(\mathrm{supp}(\gamma_*^{\mathscr{U}}))$. If this condition holds, we give a two-phase approximate message passing algorithm which first locates a suitable point $\boldsymbol{m}_{\underline{\ell}}$ with L^2 norm $\|\boldsymbol{m}_{\underline{\ell}}\| \approx \sqrt{\underline{q}N}$, and then proceeds as in the no-external-field case. The relevant condition is precisely defined as follows.

Definition 1.5. *For $\gamma_* \in \mathscr{L}$, let $\underline{q} = \inf(\mathrm{supp}(\gamma_*))$. We say γ_* is $\underline{q}$-optimizable if, with X_t given by (1.1):*

$$\mathbb{E}\big[\partial_x \Phi_{\gamma_*}(t, X_t)^2\big] = t, \quad t \in [\underline{q}, 1). \tag{1.5}$$

*We say $\gamma_* \in \mathscr{L}$ is optimizable if it is q-optimizable for $\underline{q} = \inf(\mathrm{supp}(\gamma_*))$. We say that $(\xi, \mathcal{L}_h)$ is optimizable, or equivalently that the **no overlap gap** property holds for $(\xi, \mathcal{L}_h)$, if the function $\gamma_*^{\mathscr{U}}$ is optimizable.*

In [Mon19], the widely believed conjecture that (in our language) the Sherrington-Kirkpatrick model with $\xi(x) = x^2/2$ is optimizable was assumed. Our preliminary numerical simulations suggest that the SK model remains optimizable with any constant external field $\mathcal{L}_h = \delta_h$. However even without external field, proving this conjecture rigorously seems to be very difficult.

For $\underline{q} \in [0, 1)$, let $\mathscr{L}_{\underline{q}} = \{\gamma \in \mathscr{L} : \inf(\mathrm{supp}(\gamma)) \geq \underline{q}\}$ consist of functions in $\mathscr{L}$ vanishing on $[0, \underline{q})$. The next lemma shows optimizability is equivalent to minimizing P over either $\mathscr{L}$ or $\mathscr{L}_{\underline{q}}$. It is related to results in [AC15, JT16, AMS21] which show that $\gamma_*^{\mathscr{U}}$ and $\gamma_*^{\mathscr{L}}$ satisfy (1.5), in the former case when t is a point of increase for $\gamma_*^{\mathscr{U}}$. The proof is given in Section 7.

Lemma 1.6. *For $\gamma_* \in \mathscr{L}$ and $\underline{q} = \inf(\mathrm{supp}(\gamma_*))$, the following are equivalent:*

1. *γ_* is optimizable.*
2. *$\mathsf{P}(\gamma_*) = \inf_{\gamma \in \mathscr{L}} \mathsf{P}(\gamma)$.*
3. *$\mathsf{P}(\gamma_*) = \inf_{\gamma \in \mathscr{L}_{\underline{q}}} \mathsf{P}(\gamma)$.*

Moreover if a minimizer exists in either variational problem just above, then it is unique.

Lemma 1.6 implies that any optimizable γ_* is in fact the unique minimizer $\gamma_*^{\mathscr{L}} \in \mathscr{L}$ of the Parisi functional. However throughout much of the paper we will use γ_* to denote general optimizable function without making use of this result. We made this choice because while Lemma 1.6 is important to make sense of our results, it is not necessary for proving e.g. Theorem 1.7 below. We now state our main results.

Theorem 1.7. *Suppose $\gamma_* \in \mathscr{L}$ is optimizable. Then for any $\varepsilon > 0$ there exists an efficient AMP algorithm which outputs $\boldsymbol{\sigma} \in \Sigma_N$ such that*

$$\frac{H_N(\boldsymbol{\sigma})}{N} \geq \mathsf{P}(\gamma_*) - \varepsilon$$

with probability tending to 1 as $N \to \infty$.

Lemma 1.8. *If $\gamma_*^{\mathscr{U}}$ strictly increases on $[\underline{q}, 1)$ for $\underline{q} = \inf(\mathrm{supp}(\gamma_*^{\mathscr{U}}))$, then no overlap gap holds, i.e. $\gamma_*^{\mathscr{U}}$ is optimizable.*

Corollary 1.9. *Suppose no overlap gap holds. Then for any $\varepsilon > 0$ an efficient AMP algorithm outputs $\boldsymbol{\sigma} \in \Sigma_N$ satisfying*

$$\frac{H_N(\boldsymbol{\sigma})}{N} \geq GS(\xi, \mathcal{L}_h) - \varepsilon$$

with probability tending to 1 as $N \to \infty$.

Remark 1.10. Unlike for $\mathscr{U}$ the infimum $\inf_{\gamma \in \mathscr{L}} \mathsf{P}(\gamma)$ need not be achieved, i.e. an optimizable γ_* need not exist. For instance, one has $\xi''(0) = 0$ whenever $c_2 = 0$. On the other hand if γ is optimizable, Corollary 7.1 and Lemma 7.6 (with $\underline{q} = 0$) yield

$$\int_0^t \xi''(s)\mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*}(s, X_s)^2\big]\mathrm{d}s = \mathbb{E}\big[\partial_x \Phi_{\gamma_*}(t, X_t)^2\big] \geq t, \quad t \geq 0.$$

In light of Lemma 7.2 the integrand on the left-hand side is $O(\xi''(s)) = o(1)$ so the above cannot hold for small t. Hence if $c_2 = 0$ there exists no optimizable γ_*. We conjecture

that conversely a minimizing $\gamma_*^{\mathscr{L}} \in \mathscr{L}$ exists whenever $c_2 > 0$, but we do not have a proof.

Remark 1.11. By the symmetry of $\widetilde{H}_N$, the external field can also be a deterministic vector $\boldsymbol{h} = (h_1, \ldots, h_N)$. As long as the empirical distribution of the values $(h_i)_{i \in [N]}$ is close to $\mathcal{L}_h$ in W_2 distance and the external field is independent of $\widetilde{H}_N$, exactly the same results hold – see [AMS21, Appendix A, Theorem 6] for the corresponding AMP statement in this slightly more general setting.

1.2 Optimizability and no overlap gap

In contrast to Corollary 1.9, the paper [GJ21] rules out approximate maximization using AMP for pure p-spin models without external field based on an overlap gap property whenever $p \geq 4$ is even. In formulating this result, [GJ21] defines an AMP algorithm to be any iteration of the form given in Subsection 2.1 with Lipschitz non-linearities $f_0, f_1, \ldots, f_\ell$ which outputs $\boldsymbol{\sigma} = \max(-1, \min(1, f_\ell(\boldsymbol{z}^0, \ldots, \boldsymbol{z}^\ell)))$, the final iterate f_ℓ projected into $[-1, 1]^N$. Here ℓ is a large constant which cannot grow with N. (In [AMS21] and the present work, the final iterate is rounded to be in $\boldsymbol{\Sigma}_N$ but this step does not change the asymptotic value of H_N and is essentially irrelevant – see for instance Equation (3.14).) In fact for a broad class of models, their main result based on the overlap gap property applies exactly when $\gamma_*^{\mathscr{U}}$ is not optimizable. This justifies our definition of $(\xi, \mathcal{L}_h)$ as having "no overlap gap" if and only if it is optimizable.

Proposition 1.12. *Suppose $\gamma_*^{\mathscr{U}}$ is not optimizable, where ξ is an even polynomial and the external field $\mathcal{L}_h = \delta_h$ is constant. Then there is $\varepsilon > 0$ such that for any AMP algorithm with random output $\boldsymbol{\sigma}$,*

$$\mathbb{P}\left[\frac{H_N(\boldsymbol{\sigma})}{N} \leq GS_{\xi,h} - \varepsilon\right] \geq 1 - e^{-\Omega(N)}.$$

The proof of Proposition 1.12 is identical to that of [GJ21, Theorem 3.3] and we give an outline in Section 6. Taken together, Corollary 1.9 and Proposition 1.12 exactly characterize the mean-field Ising spin glasses for which approximate maximization is possible by AMP, at least when ξ is even and the external field is constant. We remark that similar lower bounds were studied for the class of constant-degree polynomial algorithms in [GJW20]. These results also extend to any non-optimizable Ising spin glass with even ξ and constant h, ruling out approximate maximization algorithms in a slightly weaker sense. Constant-degree polynomials encompass AMP in most cases by approximating each non-linearity f_ℓ by a polynomial in a suitable sense, see e.g. [Mon19, Theorem 6].

We conclude this subsection with a brief discussion on our terminology. Our definition of optimizability is closely related to "full" or "continuous" replica symmetry breaking. For example, the definitions of full RSB used in [Mon19, Sub21] essentially coincide with 0-optimizability. However these terms seem to be slightly ambiguous, as they can also refer to functions $\gamma_*^{\mathscr{U}}$ which are strictly increasing on **any** nontrivial interval instead of being piece-wise constant as in finite RSB. For example, the physics paper [CKP+14] describes "the case where the function $\Delta(x)$ is allowed to have a continuous part: this can be thought as an appropriate limit of the k-RSB construction when $k \to \infty$ and is therefore called 'fullRSB' or '∞-RSB' ". Adding to the potential confusion, [ACZ20] uses the term "infinite step" RSB to refer to functions $\gamma_*^{\mathscr{U}}$ with infinity many points of increase, possibly at a discrete set. We therefore use "no overlap gap" as an unambiguous term for the condition that $\gamma_*^{\mathscr{U}}$ is optimizable, while keeping in mind that it closely is implied via Lemma 1.8 by a strong, specific form of full RSB.

1.3 Branching IAMP and spherical spin glasses

Under no overlap gap, one expects that any finite ultrametric space of diameter at most $\sqrt{2(1-\underline{q})}$ (with size independent of N) can be realized by approximate maximizers of H_N. In fact a modification of our $\underline{q}$-IAMP algorithm is capable of explicitly producing such realizations. In Section 4 we give a *branching $\underline{q}$-IAMP* algorithm which for any finite ultrametric space X and optimizable γ_* constructs points $(\boldsymbol{\sigma}_x)_{x \in X}$ such that $\frac{H_N(\boldsymbol{\sigma}_x)}{N} \simeq \mathsf{P}(\gamma_*)$ and $\frac{\|\boldsymbol{\sigma}_x - \boldsymbol{\sigma}_y\|_2}{\sqrt{N}} \simeq d_X(x,y)$ for each $x, y \in X$. Recall that an ultrametric space X is a metric space which satisfies the ultrametric triangle inequality

$$d_X(x,y) \leq \max\big(d_X(x,z), d_X(y,z)\big), \quad \forall\, x, y, z \in X.$$

Moreover any finite ultrametric can be canonically identified with the leaf set of a rooted tree, see e.g. [BD98].

The idea is to occasionally reset the IAMP part of the algorithm with external randomness. A similar strategy was proposed but not analyzed in [AM20].

Theorem 1.13. *Let $\gamma_* \in \mathscr{L}$ be optimizable, and fix a finite ultrametric space (X, d_X) with diameter at most $\sqrt{2(1-\underline{q})}$ as well as $\varepsilon > 0$. Then an efficient AMP algorithm constructs points $\{\boldsymbol{\sigma}_x | x \in X\}$ in $\boldsymbol{\Sigma}_N$ satisfying*

$$\frac{H_N(\boldsymbol{\sigma}_x)}{N} \geq \mathsf{P}(\gamma_*) - \varepsilon, \quad x \in X,$$

$$\frac{\|\boldsymbol{\sigma}_x - \boldsymbol{\sigma}_y\|}{\sqrt{N}} \in \big[d_X(x,y) - \varepsilon, d_X(x,y) + \varepsilon\big], \quad x, y \in X$$

with probability tending to 1 as $N \to \infty$.

In Section 5 we give corresponding results for spherical spin glasses, extending [Sub21] to the case of non-trivial external field. At zero temperature, [CS17, Theorem 1] determines the free energy in spherical spin glasses based on a positive, non-decreasing function $\alpha : [0,1) \to [0, \infty)$ as well as a constant L. (See also [JT17] for related results.) More precisely, they show the asymptotic ground state energy is given by the unique minimizer to the variational problem:

$$GS_{\mathsf{sph}}(\xi, h) = \min_{L, \alpha \in \mathcal{K}} \mathcal{Q}(L, \alpha); \tag{1.6}$$

$$\mathcal{K} = \left\{ (L, \alpha) \in (0, \infty) \times \mathscr{U} \ : \ L > \int_0^1 \alpha(s)\mathrm{d}s \right\};$$

$$2\mathcal{Q}(L, \alpha) = \big(\xi'(1) + h^2\big)L - \int_0^1 \xi''(q)\bigg(\int_0^q \alpha(s)\mathrm{d}s\bigg)\mathrm{d}q + \int_0^1 \frac{\mathrm{d}q}{L - \int_0^q \alpha(s)\mathrm{d}s}.$$

The associated definition of no overlap gap is as follows.

Definition 1.14. *The spherical mixed p-spin model is said* no overlap gap *if for some $\underline{q}_{\mathsf{sph}} \in [0,1)$, the unique minimizing $\alpha \in \mathscr{U}$ in (1.6) is strictly increasing on $[\underline{q}_{\mathsf{sph}}, 1)$ and satisfies $\alpha(q) = 0$ for $q \leq \underline{q}_{\mathsf{sph}}$.*

Unlike the Ising case, we do not formulate a generalized variational principle and only show how to achieve a natural energy value, which coincides with the ground state energy when no overlap gap holds by [CS17, Proposition 2]. We also exactly characterize the spherical models exhibiting no overlap gap, which slightly extends the same result.

Theorem 1.15. *Suppose ξ and $\mathcal{L}_h$ satisfy $\mathbb{E}[h^2] + \xi'(1) < \xi''(1)$, and let $\underline{q}_{\mathsf{sph}} \in (0,1)$ be the unique solution to $\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}}) = \underline{q}_{\mathsf{sph}}\xi''(\underline{q}_{\mathsf{sph}})$. Then the spherical spin glass with*

parameters ξ, $\mathcal{L}_h$ has no overlap gap if and only if $\xi''(q)^{-1/2}$ is concave on $q \in [\underline{q}_{\mathsf{sph}}, 1]$, in which case α is supported on $[\underline{q}_{\mathsf{sph}}, 1]$ and takes the explicit form

$$\alpha(s) = \begin{cases} 0, & s \in [0, \underline{q}_{\mathsf{sph}}) \\ \frac{\xi'''(s)}{2\xi''(s)^{3/2}}, & s \in [\underline{q}_{\mathsf{sph}}, 1]. \end{cases}$$

Moreover the ground-state energy satisfies

$$GS_{\mathsf{sph}}(\xi, \mathcal{L}_h) \geq \underline{q}_{\mathsf{sph}}\sqrt{\xi''(\underline{q}_{\mathsf{sph}})} + \int_{\underline{q}_{\mathsf{sph}}}^1 \sqrt{\xi''(q)}\mathrm{d}q$$

with equality if and only if no overlap gap occurs.

Theorem 1.16. *Suppose ξ and $h \sim \mathcal{L}_h$ satisfy $\mathbb{E}[h^2] + \xi'(1) < \xi''(1)$, and let $\underline{q}_{\mathsf{sph}} \in (0, 1)$ be the unique solution to $\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}}) = \underline{q}_{\mathsf{sph}}\xi''(\underline{q}_{\mathsf{sph}})$. Then there exists an efficient AMP algorithm outputting $\boldsymbol{\sigma} \in \mathbb{S}^{N-1}(\sqrt{N})$ such that*

$$\frac{H_N(\boldsymbol{\sigma})}{N} \simeq \underline{q}_{\mathsf{sph}}\sqrt{\xi''(\underline{q}_{\mathsf{sph}})} + \int_{\underline{q}_{\mathsf{sph}}}^1 \sqrt{\xi''(q)}\mathrm{d}q.$$

If on the other hand $\mathbb{E}[h^2] + \xi'(1) \geq \xi''(1)$, then there is an efficient AMP algorithm outputting $\boldsymbol{\sigma} \in \mathbb{S}^{N-1}(\sqrt{N})$ with

$$\frac{H_N(\boldsymbol{\sigma})}{N} \simeq \sqrt{\mathbb{E}[h^2] + \xi'(1)}.$$

Remark 1.17. If $\mathbb{E}[h^2] + \xi'(1) \geq \xi''(1)$ then the model is replica-symmetric by [CS17, Proposition 1]. When $\mathbb{E}[h^2] + \xi'(1) < \xi''(1)$, the function $f(q) = q\xi''(q) - \xi'(q) - \mathbb{E}[h^2]$ is increasing and satisfies $f(0) < 0 < f(1)$, hence has a unique root $\underline{q}_{\mathsf{sph}} \in (0, 1)$.

Remark 1.18. Subsequently to the present work and in collaboration with Brice Huang, we showed in [HS21] that the algorithms presented in this paper are optimal in some sense. More precisely, it is shown that Proposition 1.12 can be strengthed to say that

$$\mathbb{P}\left[\frac{H_N(\boldsymbol{\sigma})}{N} \leq \inf_{\gamma \in \mathscr{L}} \mathsf{P}(\gamma) + \varepsilon\right] \geq 1 - e^{-\Omega(N)} \tag{1.7}$$

for any $\varepsilon > 0$. Here, as in Proposition 1.12, $\boldsymbol{\sigma} \in [-1, 1]^N$ is the output of an AMP algorithm whose number of iterations is independent of N. This result applies more generally to arbitrary algorithms with suitably Lipschitz dependence on the disorder variables defining H_N. In the spherical case, [HS21] similarly shows that the energy attained in Theorem 1.16 is asymptotically best possible for Lipschitz algorithms; see Proposition 2.2 therein.

The essential idea of [HS21] is to consider general finite ultrametric spaces (with size independent of N) of points $\boldsymbol{\sigma}$ with large energy. They show that for $E > \inf_{\gamma \in \mathscr{L}} \mathsf{P}(\gamma)$, the level sets

$$S_E \equiv \left\{ \boldsymbol{\sigma} \in [-1, 1]^N : \frac{H_N(\boldsymbol{\sigma})}{N} \geq E \right\}$$

do not contain approximately isometric embeddings of sufficiently complicated finite ultrametrics. (Technically proving (1.7) requires a more complicated obstruction involving a correlated family of different Hamiltonians.) Theorem 1.13 is a sharp converse to and was a key inspiration for this result, as it constructs arbitrary finite ultrametric configurations at energy $\mathsf{P}(\gamma_*)$ for optimizable γ_*. See the introduction of [HS21] for further discussion and implications.

2 Technical preliminaries

We will use ordinary lower-case letters for scalars $(m, x, \ldots,)$ and bold lower-case for vectors $(\boldsymbol{m}, \boldsymbol{x})$. Ordinary upper-case letters are used for the state-evolution limits of AMP as in Proposition 2.3 such as $(X_j^\delta, Z_j^\delta, N_j^\delta)$ as well as for continuous-time stochastic processes such as (X_t, Z_t, N_t). We denote limits in probability as $N \to \infty$ by $\text{p-lim}_{N \to \infty}(\cdot)$. We write $x \simeq y$ to indicate that $\text{p-lim}_{N \to \infty}(x - y) = 0$ where x, y are random scalars.

We will use the ordinary inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^N x_i y_i$ as well as the normalized inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_N = \frac{\sum_{i=1}^N x_i y_i}{N}$. Here $\boldsymbol{x} = (x_1, \ldots, x_N) \in \mathbb{R}^N$ and similarly for $\boldsymbol{y}$. Associated with these are the norms $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$ and $\|\boldsymbol{x}\|_N = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle_N}$. We will also use the notation $\langle \boldsymbol{x} \rangle_N = \frac{\sum_{i=1}^N x_i}{N}$. Often, for example in (2.3), we apply a scalar function f to a vector $\boldsymbol{x} \in \mathbb{R}^N$. This will always mean that f is applied entrywise, i.e. $f(x_1, \ldots, x_N) = (f(x_1), \ldots, f(x_N))$. Similarly for a function $f : \mathbb{R}^{\ell+1} \to \mathbb{R}$, we define

$$f(\boldsymbol{x}^0, \boldsymbol{x}^1, \ldots, \boldsymbol{x}^\ell) = (f(x_1^0, x_1^1, \ldots, x_1^\ell), f(x_2^0, x_2^1, \ldots, x_2^\ell), \ldots f(x_N^0, x_N^1, \ldots, x_N^\ell)) \in \mathbb{R}^N.$$
(2.1)

The following useful a priori estimate shows that all derivatives of $\widetilde{H}_N$ have order 1 in the $\|\cdot\|_N$ norm. Note that we do not apply any non-standard normalization in the definitions of gradients, Hessians, etc. We use $\|\cdot\|_{\text{op}}$ to denote the operator norm of a tensor $\boldsymbol{T} \in (\mathbb{R}^N)^{\otimes k}$ of arbitrary order k:

$$\|\boldsymbol{T}\|_{\text{op}} = \sup_{\boldsymbol{x} \in \mathbb{R}^N} \frac{|\langle \boldsymbol{T}, \boldsymbol{x}^{\otimes k} \rangle|}{\|\boldsymbol{x}\|^k}.$$
(2.2)

Proposition 2.1 ([ASZ20, Lemma C.1]). *Fix a mixture function ξ, external field distribution $\mathcal{L}_h$, $k \in \mathbb{Z}^+, \eta \in \mathbb{R}^+$, and assume that the coefficients of ξ decay exponentially. Then for suitable $C = C(\xi, \mathcal{L}_h, k, \eta)$,*

$$\mathbb{P}\Big[\sup_{\|\boldsymbol{x}\| \leq (1+\eta)\sqrt{N}} \|\nabla^k \widetilde{H}_N(\boldsymbol{x})\|_{\text{op}} \leq C N^{1-\frac{k}{2}}\Big] \geq 1 - e^{-\Omega(N)}.$$

2.1 Review of approximate message passing

Here we review the general class of approximate message passing (AMP) algorithms. AMP algorithms are a flexible class of efficient algorithms based on a random matrix or, in our setting, mixed tensor. To specify an AMP algorithm, we fix a probability distribution p_0 on $\mathbb{R}$ with finite second moment and a sequence $f_0, f_1, \ldots$ of Lipschitz functions $f_\ell : \mathbb{R}^{\ell+1} \to \mathbb{R}$, with $f_{-1} = 0$. The functions f_ℓ will often be referred to as *non-linearities*. We begin by taking $\boldsymbol{z}^0 \in \mathbb{R}^N$ to have i.i.d. coordinates $(z_i^0)_{i \in [N]} \sim p_0$. Then we recursively define $\boldsymbol{z}^1, \boldsymbol{z}^2, \ldots$ via

$$\boldsymbol{z}^{\ell+1} = \nabla \widetilde{H}_N(f_\ell(\boldsymbol{z}^0, \ldots, \boldsymbol{z}^\ell)) - \sum_{j=1}^\ell d_{\ell,j} f_{j-1}(\boldsymbol{z}^0, \ldots, \boldsymbol{z}^{j-1}),$$
(2.3)

$$d_{\ell,j} = \xi''(\langle f_\ell(\boldsymbol{z}^0, \ldots, \boldsymbol{z}^\ell), f_{j-1}(\boldsymbol{z}^0, \ldots, \boldsymbol{z}^{j-1}) \rangle_N) \cdot \mathbb{E}\Big[\frac{\partial f_\ell}{\partial Z^j}(Z^0, \ldots, Z^\ell)\Big].$$
(2.4)

Here the non-linearity f_ℓ is applied coordinate-wise as in (2.1). Moreover $Z^0 \sim p_0$ while $(Z^\ell)_{\ell \geq 1}$ is an independent centered Gaussian process with covariance $Q_{\ell,j} = \mathbb{E}[Z^\ell Z^j]$ defined recursively by

$$Q_{\ell+1,j+1} = \xi'(\mathbb{E}[f_\ell(Z^0, \ldots, Z^\ell) f_j(Z^0, \ldots, Z^j)]), \quad \ell, j \geq 0.$$
(2.5)

The key property of AMP, stated below in Proposition 2.3, is that for any ℓ the empirical distribution of the N sequences $(z_i^1, z_i^2, \ldots, z_i^\ell)_{i \in [N]}$ converges in distribution to the law of the Gaussian process $(Z^1, \ldots, Z^\ell)$ as $N \to \infty$. This is called *state evolution*.

Definition 2.2. *For non-negative integers n, m the function $\psi : \mathbb{R}^\ell \to \mathbb{R}$ is* pseudo-Lipschitz *if for some constant L and any $x, y \in \mathbb{R}^\ell$,*

$$\|\psi(x) - \psi(y)\| \le L(1 + \|x\| + \|y\|)\|x - y\|.$$

Proposition 2.3 ([AMS21, Proposition 3.1]). *For any pseudo-Lipschitz $\psi : \mathbb{R}^{\ell+1} \to \mathbb{R}$, the AMP iterates satisfy*

$$\operatorname*{p-lim}_{N \to \infty}\langle\psi(\boldsymbol{z}^0, \ldots, \boldsymbol{z}^\ell)\rangle_N = \mathbb{E}[\psi(Z^0, \ldots, Z^\ell)].$$

The first version of state evolution was given for Gaussian random matrices in [Bol14, BM11]. Since then it has been extended to more general situations in many works including [JM13, BLM15, BMN19, CL21, Fan22]. As state evolution holds for essentially arbitrary non-linearities f_ℓ, it allows a great deal of flexibility in solving problems involving random matrices or tensors.

We remark that [AMS21, Proposition 3.1] is phrased in terms of a random mixed tensor $\boldsymbol{W}$, i.e. a sequence of p-tensors $(\boldsymbol{W}^{(p)} \in (\mathbb{R}^N)^{\otimes p})_{p \ge 2}$ – see Equation (3.2) therein. The two descriptions are equivalent because $\boldsymbol{W}$ is constructed so that $\sum_{p \ge 2} c_p \langle \boldsymbol{W}^{(p)}, \boldsymbol{x}^{\otimes p}\rangle = \widetilde{H}_N(\boldsymbol{x})$. While the tensor language is better suited to proving state evolution, for our purposes it is more convenient to express AMP just in terms of $\widetilde{H}_N$ and $\nabla\widetilde{H}_N$.

Let us finally discuss the efficiency of our AMP algorithms. The algorithms we give are described by parameters $\bar{q}$ and $\underline{\ell}$ and require oracle access to the function $\Phi_{\gamma_*}(t, x)$ and its derivatives. We do not address the complexity of computing $\Phi_{\gamma_*}(t, x)$. However as stated in [Mon19, AMS21] it seems unlikely to present a major obstacle because solving for $\gamma_*^{\mathscr{U}}$ is a convex problem which only must be solved once for each $(\xi, \mathcal{L}_h)$. Moreover [AM20] demonstrates that these algorithms are practical to implement.

In the end, our algorithms output rounded points $\boldsymbol{\sigma}$ with $\boldsymbol{\sigma}_i = \operatorname{sign}(f_{\bar{\ell}}(\boldsymbol{z}_i^0, \ldots, \boldsymbol{z}_i^{\bar{\ell}}))$ for a large value $\bar{\ell} = \bar{\ell}(\bar{q}, \underline{\ell})$. The outputs satisfy

$$\lim_{\bar{q} \to 1} \lim_{\underline{\ell} \to \infty} \operatorname*{p-lim}_{N \to \infty} \frac{H_N(\boldsymbol{\sigma})}{N} = H_*$$

for some asymptotic energy value H_*. To achieve an ε-approximation to the value H_*, the parameters $\bar{q}$ and $\underline{\ell}$ must be sent to 1 and ∞ which requires a diverging number of iterations. In particular let χ denote the complexity of computing $\nabla\widetilde{H}_N$ at a point and let χ_1 denote the complexity of computing a single coordinate of $\nabla\widetilde{H}_N$ at a point. Then the total complexity needed to achieve energy $H_* - \varepsilon$ is $C(\varepsilon)(\chi + N) + N\chi_1$. When ξ is a polynomial this complexity is linear in the size of the input specifying H_N – see the comments following [AMS21, Remark 2.1]. In the statements of our results, we refer to such algorithms as "efficient AMP algorithms".

2.2 Initializing AMP

Here we explain some technical points involved in initializing our AMP algorithms and why they arise. First, we would like to use a random external field h_i which varies from coordinate to coordinate. In the most natural AMP implementation, this requires that the non-linearities f_ℓ correspondingly depend on the coordinate rather than being fixed, which is not allowed in state evolution. Second we would like to use many i.i.d. Gaussian vectors throughout the branching version of the algorithm. However Proposition 2.3 allows only a single initial vector $\boldsymbol{z}^0$ as a source of external randomness independent of H_N. One could prove a suitable generalization of Proposition 2.3, but we instead build these additional vectors into the initialization of the AMP algorithm as a sort of preprocessing phase. To indicate that our constructions here are preparation for the "real algorithm", we reparametrize so the preparatory iterates have negative index.

We begin by taking $p_0 = \mathcal{L}_h$ to be the distribution of the external field itself, and initialize $(\boldsymbol{z}^{-K})_i = h_i \sim \mathcal{L}_h$ for some constant $K \in \mathbb{Z}^+$. We then set $f_{-K}(\boldsymbol{z}^{-K}) = \frac{\boldsymbol{z}^{-K}}{\sqrt{\mathbb{E}^{h \sim \mathcal{L}_h}[h^2]}}$ and $f_{-k}(\boldsymbol{z}^{-K}, \dots, \boldsymbol{z}^{-k}) = \boldsymbol{z}^{-k}$ for $2 \le k \le K$. Finally we set $f_{-1}(\boldsymbol{z}^{-K}, \dots, \boldsymbol{z}^{-1}) = c\boldsymbol{z}^{-1}$ for some constant $c > 0$ which the algorithm is free to choose. (Note that the functions f_{-k} correspond to entry-wise applications of the form in (2.1).) State evolution immediately implies the following

Proposition 2.4. *In the state evolution $N \to \infty$ limit, the empirical distribution of $(z_i^{-K}, \dots, z_i^0)$ (for $i \in [N]$ uniformly random) converges in probability to the law of an independent $(K+1)$-tuple $(Z^{-K}, Z^{-K+1}, \dots, Z^{-1}, Z^0)$ with $Z^{-K} \sim \mathcal{L}_h$, $(Z^{-K+1}, \dots, Z^{-1}) \sim \mathsf{N}(0, I_{K-1})$ i.i.d. standard Gaussian, and $Z^0 \sim \mathsf{N}(0, c^2)$.*

In fact taking $K = 1$ suffices for the main construction in the paper. In Section 4 we require larger values of K for branching IAMP, where the iterates $(\boldsymbol{z}^{-K+1}, \dots, \boldsymbol{z}^{-1})$ serve as proxies for i.i.d. Gaussian vectors.

Remark 2.5. Because the sum defining the Onsager correction term in (2.3) starts at $j = 1$, the effect of the external field h_i on future AMP iterates does not enter into any Onsager correction terms in this paper.

2.3 Properties of the Parisi PDE and SDE

Quite a lot is known about the solution Φ_γ to the Parisi PDE. The next results hold for any $\gamma \in \mathscr{L}$ and are taken from [AMS21]. Similar results for $\gamma \in \mathscr{U}$ appear in [AC15, JT16].

Proposition 2.6 ([AMS21, Lemmas 6.2, 6.4])**.** *For any $\gamma \in \mathscr{L}$, the solution $\Phi_\gamma(t, x)$ to the Parisi PDE is continuous on $[0, 1] \times \mathbb{R}$, convex in x, and further satisfies the following regularity properties for any $\varepsilon > 0$.*

(a) $\partial_x^j \Phi \in L^\infty([0, 1 - \varepsilon]; L^2(\mathbb{R}) \cap L^\infty(\mathbb{R}))$ *for $j \ge 2$.*
(b) $\partial_t \Phi \in L^\infty([0, 1] \times \mathbb{R})$ *and* $\partial_t \partial_x^j \Phi \in L^\infty([0, 1 - \varepsilon]; L^2(\mathbb{R}) \cap L^\infty(\mathbb{R}))$ *for $j \ge 1$.*

Proposition 2.7 ([AMS21, Lemmas 6.2, 6.4])**.** *For any $\gamma \in \mathscr{L}$, Φ_γ satisfies*

$$|\partial_x \Phi_\gamma(t, x)| \le 1$$

for all $(t, x) \in [0, 1] \times \mathbb{R}$.

Proposition 2.8 ([AMS21, Lemma 6.5])**.** *For any $\gamma \in \mathscr{L}$, the Parisi SDE (1.1) has unique strong solution $(X_t)_{t \in [0,1]}$ which is a.s. continuous and satisfies*

$$\partial_x \Phi_\gamma(t, X_t) = \int_0^t \sqrt{\xi''(s)}\, \partial_{xx} \Phi_\gamma(s, X_s)\, \mathrm{d}B_s. \tag{2.6}$$

Finally we give two additional properties for optimizable γ_*, which are proved in Section 7.

Lemma 2.9. *If $\gamma_* \in \mathscr{L}$ is $\underline{q}$-optimizable then it satisfies:*

$$\mathbb{E}\big[\partial_{xx} \Phi_{\gamma_*}(t, X_t)^2\big] = \frac{1}{\xi''(t)}, \quad t \ge \underline{q}, \tag{2.7}$$

$$\mathbb{E}\big[\partial_{xx} \Phi_{\gamma_*}(t, X_t)\big] \ge \int_t^1 \gamma_*(s)\mathrm{d}s, \quad t \in [0, 1]. \tag{2.8}$$

Remark 2.10. We expect (2.8) to hold with equality; if this is true, then our analysis in Subsection 3.3 shows that Theorem 1.7 holds as a two-sided estimate. Conversely, the main result of [HS21] implies such a two-sided estimate when ξ is even; retracing Subsection 3.3 then implies (2.8) is indeed an equality in such cases. However this is unsatisfyingly indirect and it would be interesting to give a direct proof.

3 The main algorithm

In this section we explain our main AMP algorithm and prove Theorem 1.7. Throughout we take $\gamma_* \in \mathscr{L}$ to be $\underline{q}$-optimizable for $\underline{q} = \inf(\mathrm{supp}(\gamma_*)) \in [0,1)$.

3.1 Phase 1: finding the root

Here we give the first phase of the algorithm, which proceeds for a large constant number ℓ of iterations after initialization and approximately converges to a fixed point. The AMP iterates during this first phase are denoted by $(\boldsymbol{w}^k)_{-K \leq k \leq \ell}$. We rely on the function

$$f(x) = \partial_x \Phi_{\gamma_*}(\underline{q}, x)$$

and use non-linearities

$$f_k\big(\boldsymbol{h}, \boldsymbol{w}^{-K+1}, \ldots, \boldsymbol{w}^0, \boldsymbol{w}^1, \ldots, \boldsymbol{w}^k\big) = f\big(\boldsymbol{h} + \boldsymbol{w}^k\big)$$

for all $k \geq 1$. (As a reminder, if f is a scalar function, $f(\boldsymbol{x}^k)$ is evaluated entrywise as explained in (2.1).) Proposition 2.6 implies that each f_k is Lipschitz, so that state evolution applies to the AMP iterates. In the initialization phase we take $c = \sqrt{\xi'(\underline{q})}$ as described in Subsection 2.2, so that the coordinates w_i^0 are asymptotically distributed as centered Gaussians with variance $\xi'(\underline{q})$ in the $N \to \infty$ limit. Moreover we set $\boldsymbol{m}^k = f(\boldsymbol{x}^k)$ where $\boldsymbol{x}^k = \boldsymbol{w}^k + \boldsymbol{h}$. This yields the following iteration.

$$
\begin{aligned}
\boldsymbol{w}^{k+1} &= \nabla \widetilde{H}_N\big(f(\boldsymbol{x}^k)\big) - f\big(\boldsymbol{x}^{k-1}\big)\xi''\big(\langle f(\boldsymbol{x}^k), f(\boldsymbol{x}^{k-1})\rangle_N\big)\langle f'(\boldsymbol{x}^k)\rangle_N \qquad (3.1) \\
&= \nabla \widetilde{H}_N\big(\boldsymbol{m}^k\big) - \boldsymbol{m}^{k-1}\xi''\big(\langle \boldsymbol{m}^k, \boldsymbol{m}^{k-1}\rangle_N\big)\langle \partial_{xx}\Phi_{\gamma_*}(\underline{q}, \boldsymbol{x}^k)\rangle_N, \\
\boldsymbol{x}^{k+1} &= \boldsymbol{w}^{k+1} + \boldsymbol{h} \\
\boldsymbol{m}^k &= f\big(\boldsymbol{x}^k\big) = f_k\big(\boldsymbol{w}^k\big).
\end{aligned}
$$

Lemma 3.1. *For f as defined above, $h \sim \mathcal{L}_h$ and $Z \sim \mathsf{N}(0,1)$ an independent standard Gaussian,*

$$\mathbb{E}^{h,Z}\big[f\big(h + Z\sqrt{\xi'(\underline{q})}\big)^2\big] = \underline{q} \qquad (3.2)$$

$$\mathbb{E}^{h,Z}\big[f'\big(h + Z\sqrt{\xi'(\underline{q})}\big)^2\big] = \frac{1}{\xi''(\underline{q})}. \qquad (3.3)$$

Proof. The identities follow by taking $t = \underline{q}$ in the definition of optimizability as well as Lemma 2.9. Here we use the fact that $X_t \overset{d}{=} X_0 + Z\sqrt{\xi'(t)}$ is a time-changed Brownian motion started from X_0 for $t \leq \underline{q}$. $\qquad \square$

Next with $(Z, Z', Z'') \sim \mathsf{N}(0, I_3)$ independent of $h \sim \mathcal{L}_h$, define for $t \leq \xi'(\underline{q})$ the function

$$\phi(t) = \mathbb{E}^{h,Z,Z',Z''}\big[f\big(h + Z\sqrt{t} + Z'\sqrt{\xi'(\underline{q}) - t}\big)f\big(h + Z\sqrt{t} + Z''\sqrt{\xi'(\underline{q}) - t}\big)\big]. \qquad (3.4)$$

Define also $\psi(t) = \xi'(\phi(t))$. It follows from (3.2) that

$$\phi\big(\xi'(\underline{q})\big) = \underline{q}. \qquad (3.5)$$

Lemma 3.2. *The function ψ is strictly increasing and strictly convex on $[0, \xi'(\underline{q})]$. Moreover*

$$\psi\big(\xi'(\underline{q})\big) = \xi'(\underline{q}), \quad \psi'\big(\xi'(\underline{q})\big) = 1.$$

Finally $\psi(t) > t$ for all $t < \xi'(\underline{q})$.

Proof. Using Gaussian integration by parts as in [Bol14, Lemma 2.2], we find

$$\phi'(t) = \mathbb{E}^{h,Z,Z',Z''}\big[f'\big(h + \sqrt{t}Z + \sqrt{\xi'(\underline{q}) - t}Z'\big)f'\big(h + \sqrt{t}Z + \sqrt{\xi'(\underline{q}) - t}Z''\big)\big]$$
$$= \mathbb{E}^{h,Z}\big[\mathbb{E}^{Z'}\big[f'\big(h + \sqrt{t}Z + \sqrt{\xi'(\underline{q}) - t}Z'\big)\big]^2\big],$$
$$\phi''(t) = \mathbb{E}\big[f''\big(h + \sqrt{t}Z + \sqrt{\xi'(\underline{q}) - t}Z'\big)f''\big(h + \sqrt{t}Z + \sqrt{\xi'(\underline{q}) - t}Z''\big)\big]$$
$$= \mathbb{E}^{h,Z}\big[\mathbb{E}^{Z'}\big[f''\big(h + \sqrt{t}Z + \sqrt{\xi'(\underline{q}) - t}Z'\big)\big]^2\big].$$

These expressions are each strictly positive, as the optimizability of γ_* implies that f', f'' are not identically zero. Therefore ϕ is increasing and convex. Since ξ' is also increasing and convex (being a power series with non-negative coefficients) we conclude the same about their composition ψ. The values $\psi(\xi'(\underline{q})) = \xi'(\underline{q})$ and $\psi'(\xi'(\underline{q})) = 1$ follow from Lemma 3.1 and the chain rule. Finally the last claim follows by strict convexity of ψ and $\psi'(\xi'(\underline{q})) = 1$. $\qquad\square$

Next, let $h, W^{-1}, (W^j, X^j, M^j)_{j \geq 0}$ be the state evolution limit of the coordinates of

$$\big(\boldsymbol{h}, \boldsymbol{w}^{-1}, \boldsymbol{w}^0, \boldsymbol{x}^0, \boldsymbol{m}^0, \ldots, \boldsymbol{w}^k, \boldsymbol{x}^k, \boldsymbol{m}^k\big)$$

as $N \to \infty$. Hence each W^j is a centered Gaussian and $X^j = W^j + h$, $M^{j+1} = f(X^j)$ hold for $j \geq 0$. Define the sequence $(a_0, a_1, \ldots)$ recursively by $a_0 = 0$ and $a_{k+1} = \psi(a_k)$.

Lemma 3.3. *For all non-negative integers $0 \leq j < k$, the following equalities hold:*

$$\mathbb{E}\big[\big(W^j\big)^2\big] = \xi'(\underline{q}) \tag{3.6}$$
$$\mathbb{E}\big[W^j W^k\big] = a_j \tag{3.7}$$
$$\mathbb{E}\big[\big(M^j\big)^2\big] = \underline{q} \tag{3.8}$$
$$\mathbb{E}\big[M^j M^k\big] = \phi(a_j). \tag{3.9}$$

Moreover $(W^j)_{j \geq 0}$ is independent of h.

Proof. We proceed by induction on j, first showing (3.6) and (3.8) together. As a base case, (3.6) holds for $j = 0$ by initialization. For the inductive step, assume first that (3.6) holds for j. Then state evolution and (3.5) yield

$$\mathbb{E}\big[\big(M^j\big)^2\big] = \phi\big(\xi'(\underline{q})\big) = \underline{q}$$

so that (3.6) implies (3.8) for each $j \geq 0$. On the other hand, state evolution directly implies that if (3.8) holds for j then (3.6) holds for $j + 1$. This establishes (3.6) and (3.8) for all $j \geq 0$.

We similarly show (3.7) and (3.9) together by induction, beginning with (3.7) when $j = 0$. By the initialization of Subsection 2.2 it follows that the random variables h, W^{-1}, W^0 are jointly independent. State evolution implies that W^{k-1} is independent of W^{-1} for any $k \geq 0$. Then state evolution yields for any $k \geq 1$:

$$\mathbb{E}\big[W^0 W^k\big] = \xi'\big(\mathbb{E}\big[M^{-1} M^{k-1}\big]\big)$$
$$= \xi'\big(\mathbb{E}\big[W^{-1} f\big(W^{k-1}\big)\big]\big)$$
$$= \xi'(0)$$
$$= 0.$$

Just as above, it follows from state evolution that (3.7) for (j, k) implies (3.9) for (j, k) which in turn implies (3.7) for $(j + 1, k + 1)$. Hence induction on j proves (3.7) and (3.9) for all (j, k). Finally the last independence assertion is immediate from state evolution just because h is the first step in the AMP iteration. $\qquad\square$

Lemma 3.4.

$$\lim_{k \to \infty} a_k = \xi'(\underline{q}).$$

Proof. Since ψ is strictly increasing and maps $[0, \xi'(\underline{q})] \to [0, \xi'(\underline{q})]$, it follows that $(a_k)_{k \geq 0}$ is a strictly increasing sequence with limiting value in $[0, \xi'(\underline{q})]$. Let $a_\infty = \lim_{k \to \infty} a_k$ be this limit. Then continuity implies $\psi(a_\infty) = a_\infty$ which by the last part of Lemma 3.2 implies $a_\infty = \xi'(\underline{q})$. This concludes the proof. $\qquad\square$

We now compute the limiting energy

$$\lim_{k \to \infty} \operatorname*{p-lim}_{N \to \infty} \frac{H_N(\boldsymbol{m}_k)}{N}$$

from the first phase. Since the first phase is similar to many "standard" AMP algorithms, this step is comparable to the computation of their final objective value, for example [DAM17, Lemma 6.3]. This computation is straightforward when $\widetilde{H}_N$ is a homogeneous polynomial of degree p, because one can just rearrange the equation for an AMP iteration to solve for

$$\widetilde{H}_N(\boldsymbol{m}^k) = p^{-1}\langle \boldsymbol{m}^k, \nabla \widetilde{H}_N(\boldsymbol{m}^k)\rangle.$$

However it requires more work in our setting because $\nabla \widetilde{H}_N$ acts differently on terms of different degrees. We circumvent this mismatch by applying state evolution to a t-dependent auxiliary AMP step and integrating in t.

Lemma 3.5. *With X_t the Parisi SDE* (1.1),

$$\lim_{k \to \infty} \operatorname*{p-lim}_{N \to \infty} \frac{H_N(\boldsymbol{m}^k)}{N} = \xi'(\underline{q}) \cdot \mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*}\big(\underline{q}, h + Z\sqrt{\xi'(\underline{q})}\big)\big] + \mathbb{E}\big[h \cdot \partial_x\Phi_{\gamma_*}\big(\underline{q}, h + Z\sqrt{\xi'(\underline{q})}\big)\big]$$

$$= \xi'(\underline{q}) \cdot \mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*}(\underline{q}, X_{\underline{q}})\big] + \mathbb{E}\big[h \cdot \partial_x\Phi_{\gamma_*}(\underline{q}, X_{\underline{q}})\big].$$

Proof. The equivalence of the latter two expressions follows from the fact that $X_{\underline{q}} \sim X_0 + \mathsf{N}(0, \xi'(\underline{q}))$ so we focus on the first equality. Observe that

$$\frac{H_N(\boldsymbol{m}^k)}{N} = \langle \boldsymbol{h}, \boldsymbol{m}^k \rangle_N + \int_0^1 \langle \boldsymbol{m}^k, \nabla \widetilde{H}_N(t\boldsymbol{m}^k)\rangle_N \mathrm{d}t \qquad (3.10)$$

holds for any vector $\boldsymbol{m}^k$ by considering each monomial term of H_N. Our main task now reduces to computing the in-probability limit of the integrand as a function of t. Proposition 2.1 ensures that $t \to \langle \boldsymbol{m}^k, \nabla \widetilde{H}_N(t\boldsymbol{m}^k)\rangle_N$ is Lipschitz assuming $\|\boldsymbol{m}^k\|_N \leq 1 + o(1)$. This holds with high probability for each k as $N \to \infty$ by state evolution and Proposition 2.7, so we may freely interchange the limit in probability with the integral.

To compute the integrand $\langle \boldsymbol{m}^k, \nabla \widetilde{H}_N(t\boldsymbol{m}^k)\rangle_N$ we analyze a modified AMP which agrees with the AMP we have considered so far up to step k, whereupon we replace the non-linearity $f_k(\boldsymbol{h}, \boldsymbol{w}^k) = f(\boldsymbol{w}^k + \boldsymbol{h}) = f(\boldsymbol{x}^k)$ by

$$\tilde{f}_k(\boldsymbol{h}, \boldsymbol{w}^k) \equiv t \cdot f(\boldsymbol{x}^k)$$

for arbitrary $t \in (0, 1)$. We obtain the new iterate

$$\boldsymbol{y}^{k+1}(t) \equiv \nabla \widetilde{H}_N(t\boldsymbol{m}^k) - t\boldsymbol{m}^{k-1}\xi''(t\langle \boldsymbol{m}^k, \boldsymbol{m}^{k-1}\rangle_N)\langle f'(\boldsymbol{x}^k)\rangle_N.$$

Rearranging yields

$$\langle \boldsymbol{m}^k, \nabla \widetilde{H}_N(t\boldsymbol{m}^k)\rangle_N = \langle \boldsymbol{m}^k, \boldsymbol{y}^{k+1}(t)\rangle_N + t\langle \boldsymbol{m}^k, \boldsymbol{m}^{k-1}\rangle_N \xi''(t\langle \boldsymbol{m}^k, \boldsymbol{m}^{k-1}\rangle_N)\langle \boldsymbol{f}'(\boldsymbol{x}^k)\rangle_N$$

$$\simeq \langle \boldsymbol{m}^k, \boldsymbol{y}^{k+1}(t)\rangle_N + ta_{k-1}\xi''(t\phi(a_{k-1}))\langle \boldsymbol{f}'(\boldsymbol{x}^k)\rangle_N.$$

We evaluate the $N \to \infty$ limit in probability of the first term, via the state evolution limits $W^k, X^k, Y^{k+1}(t)$. State evolution directly implies

$$\mathbb{E}\big[W^k Y^{k+1}(t)\big] = \xi'\big(t \cdot \mathbb{E}\big[M^{k-1} M^k\big]\big) = \xi'\big(t\phi(a_{k-1})\big).$$

Since h is independent of (W^k, Y^{k+1}) we use Gaussian integration by parts to derive

$$\begin{aligned}
\mathbb{E}\big[f\big(X^k\big)Y^{k+1}(t)\big] &= \mathbb{E}\big[f\big(h + W^k\big)Y^{k+1}(t)\big] \\
&= \mathbb{E}\big[f'\big(h + W^k\big)\big] \cdot \mathbb{E}\big[W^k Y^{k+1}(t)\big] \\
&= \mathbb{E}\big[f'\big(h + Z\sqrt{\xi'(\underline{q})}\big)\big] \cdot \xi'\big(t\phi(a_{k-1})\big).
\end{aligned}$$

Integrating with respect to t yields

$$\begin{aligned}
&\int_0^1 \big\langle \boldsymbol{m}^k, \nabla \widetilde{H}_N\big(t\boldsymbol{m}^k\big)\big\rangle_N \mathrm{d}t \\
&\simeq \mathbb{E}\big[f'\big(h + Z\sqrt{\xi'(\underline{q})}\big)\big] \cdot \int_0^1 \xi'\big(t\phi(a_{k-1})\big) + t\phi(a_{k-1})\xi''\big(t\phi(a_{k-1})\big) \, \mathrm{d}t \\
&= \mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*}\big(\underline{q}, h + Z\sqrt{\xi'(\underline{q})}\big)\big] \cdot \big[t\xi'\big(t\phi(a_{k-1})\big)\big]\big|_{t=0}^{t=1}
\end{aligned}$$

Finally the first term in (3.10) gives energy contribution

$$\begin{aligned}
\big\langle \boldsymbol{h}, \boldsymbol{m}^k\big\rangle_N &\simeq \mathbb{E}\big[h \cdot f\big(Z\sqrt{\xi'(\underline{q})}\big)\big] \\
&= \mathbb{E}\big[h \cdot \partial_x \Phi_{\gamma_*}\big(\underline{q}, h + Z\sqrt{\xi'(\underline{q})}\big)\big].
\end{aligned}$$

Since $\lim_{k\to\infty} a_{k-1} = \xi'(\underline{q})$ and $\psi(\xi'(\underline{q})) = \xi'(\underline{q})$ combining concludes the proof. $\qquad\square$

3.2 Phase 2: incremental AMP

We now switch to IAMP, which has a more complicated definition. We will begin from the iterates $\boldsymbol{x}^{\underline{\ell}}, \boldsymbol{m}^{\underline{\ell}}$ from phase 1 for a large $\underline{\ell} \in \mathbb{Z}^+$. Our implementation follows that of [AMS21, AS22] and we relegate several proofs to Section 8. First define the functions

$$u(t, x) = \partial_{xx}\Phi_{\gamma_*}(t, x), \quad v(t, x) = \xi''(t)\gamma_*(t)\partial_x \Phi_{\gamma_*}(t, x).$$

Set $\varepsilon_0 = \frac{\underline{q}}{\phi(a_{\underline{\ell}-1})} - 1$ and $\delta = \underline{q}((1 + \varepsilon_0)^2) - 1)$; observe that $\varepsilon_0, \delta \to 0$ as $\underline{\ell} \to \infty$.[1] Define the sequence $(q_\ell^\delta)_{\ell \geq \underline{\ell}}$ by

$$q_\ell^\delta = \underline{q} + (\ell - \underline{\ell})\delta.$$

Fix $\bar{q} \in (\underline{q}, 1)$; the value $\bar{q}$ will be taken close to 1 after sending $\underline{\ell} \to \infty$. In particular we will assume $\delta < 1 - \bar{q}$ holds and set $\bar{\ell} = \min\{\ell \in \mathbb{Z}^+ : q_\ell^\delta \geq \bar{q}\}$. Also define

$$\boldsymbol{n}^{\underline{\ell}} \equiv (1 + \varepsilon_0)\boldsymbol{m}^{\underline{\ell}}.$$

Set $\boldsymbol{z}^{\underline{\ell}} = \boldsymbol{w}^{\underline{\ell}}$. So far, we have defined $(\boldsymbol{x}^{\underline{\ell}}, \boldsymbol{z}^{\underline{\ell}}, \boldsymbol{n}^{\underline{\ell}})$. We turn to inductively defining the triples $(\boldsymbol{x}^\ell, \boldsymbol{z}^\ell, \boldsymbol{n}^\ell)$ for $\underline{\ell} \leq \ell \leq \bar{\ell}$. First, the values $(\boldsymbol{z}^\ell)_{\ell \geq \underline{\ell}}$ are defined as AMP iterates via

$$\begin{aligned}
\boldsymbol{z}^{\ell+1} &= \nabla\widetilde{H}_N\big(f_\ell\big(\boldsymbol{z}^0, \ldots, \boldsymbol{z}^\ell\big)\big) - \sum_{j=0}^\ell d_{\ell,j} f_{j-1}\big(\boldsymbol{z}^0, \ldots, \boldsymbol{z}^{j-1}\big), \\
d_{\ell,j} &= \xi''\big(\mathbb{E}\big[f_\ell\big(Z^0, \ldots, Z^\ell\big) f_{j-1}\big(Z^0, \ldots, Z^{j-1}\big)\big]\big) \cdot \mathbb{E}\bigg[\frac{\partial f_\ell}{\partial z^j}\big(Z^0, \ldots, Z^\ell\big)\bigg].
\end{aligned} \tag{3.11}$$

---

[1] When $\underline{q} = 0$, ε_0 is not defined. In this case we simply take $\delta > 0$ small and begin IAMP at $\boldsymbol{n}^{\underline{\ell}} = 0$ as in [AMS21].

(The non-linearities f_ℓ will be specified below). The sequence $(x^{\ell+1})_{\ell \geq \underline{\ell}}$ is defined by

$$
\begin{aligned}
\boldsymbol{x}^{\ell+1} &\equiv \boldsymbol{x}^{\underline{\ell}} + \sum_{j=\underline{\ell}}^{\ell} v\big(q_j^\delta, \boldsymbol{x}^j\big)\delta + \sum_{j=\underline{\ell}}^{\ell} \big(\boldsymbol{z}^{j+1} - \boldsymbol{z}^j\big) \\
&= \boldsymbol{x}^\ell + v\big(q_\ell^\delta, \boldsymbol{x}^\ell\big)\delta + \big(\boldsymbol{z}^{\ell+1} - \boldsymbol{z}^\ell\big), \quad \underline{\ell} \leq \ell \leq \overline{\ell} - 1.
\end{aligned}
$$

As usual, $v(q_j^\delta, \cdot)$ is applied component-wise so that $v(q_j^\delta, \boldsymbol{x}^j)_i = v(q_j^\delta, x_i^j)$. Next define the scalar function

$$
u_\ell^\delta(x) = \frac{\delta u(q_\ell^\delta, x)}{(\xi'(q_\ell^\delta) - \xi'(q_{\ell-1}^\delta))\mathbb{E}[u(q_\ell^\delta; X_\ell^\delta)^2]}
$$

and consider for $\ell \geq \underline{\ell}$ the recursive definition

$$
\begin{aligned}
\boldsymbol{n}^{\ell+1} &\equiv \boldsymbol{n}^{\underline{\ell}} + \sum_{j=\underline{\ell}}^{\ell} u_j^\delta\big(\boldsymbol{x}^j\big)\big(\boldsymbol{z}^{j+1} - \boldsymbol{z}^j\big) \\
&= \boldsymbol{n}^\ell + u_\ell^\delta\big(\boldsymbol{x}^\ell\big)\big(\boldsymbol{z}^{\ell+1} - \boldsymbol{z}^\ell\big).
\end{aligned}
\tag{3.12}
$$

We define the non-linearity $f_\ell : \mathbb{R}^{\ell+1} \to \mathbb{R}$ to recursively satisfy

$$
f_\ell\big(\boldsymbol{z}^0, \ldots, \boldsymbol{z}^\ell\big) = \boldsymbol{n}^\ell, \quad \ell > \underline{\ell}.
$$

It is not difficult to verify that the equations above form a "closed loop" uniquely determining the sequence $(\boldsymbol{x}^\ell, \boldsymbol{z}^\ell, \boldsymbol{n}^\ell)_{\ell \geq \underline{\ell}}$. Since (x_i^ℓ, n_i^ℓ) is determined by the sequence $(z_i^{\underline{\ell}}, \ldots, z_i^\ell)$ we may define the state evolution limiting random variables $(X_\ell^\delta, N_\ell^\delta, Z_\ell^\delta)_{\ell \geq \underline{\ell}}$. We emphasize that the IAMP just defined is part of the same q-AMP algorithm as the first phase defined in the previous subsection. However the variable naming has changed so that the main iterates are $\boldsymbol{z}^\ell$ for $\ell \geq \underline{\ell}$ rather than $\boldsymbol{w}^\ell$ for $\ell \leq \underline{\ell}$. In particular there is no problem in applying state evolution even though the two AMP phases take different forms.

To complete the algorithm, we output the coordinate-wise sign $\boldsymbol{\sigma} = \text{sign}(\boldsymbol{n}^{\overline{\ell}})$ where

$$
\text{sign}(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x \leq 0. \end{cases}
$$

The key to analyzing the AMP algorithm above is an SDE description in the $\delta \to 0$ limit. Define the filtration

$$
\mathcal{F}_\ell^\delta = \sigma\big(\big(Z_k^\delta, N_k^\delta\big)_{0 \leq k \leq \ell}\big)
\tag{3.13}
$$

for the state evolution limiting process.

Lemma 3.6. *The sequences $(Z_{\underline{\ell}}^\delta, Z_{\underline{\ell}+1}^\delta, \ldots)$ and $(N_{\underline{\ell}}^\delta, N_{\underline{\ell}+1}^\delta, \ldots)$ satisfy for each $\ell \geq \underline{\ell}$:*

$$
\begin{aligned}
\mathbb{E}\big[\big(Z_{\ell+1}^\delta - Z_\ell^\delta\big)Z_j^\delta\big] &= 0, \quad \text{for all } \underline{\ell}+1 \leq j \leq \ell \\
\mathbb{E}\big[\big(Z_{\ell+1}^\delta - Z_\ell^\delta\big)^2 | \mathcal{F}_\ell^\delta\big] &= \xi'\big(q_{\ell+1}^\delta\big) - \xi'\big(q_\ell^\delta\big) \\
\mathbb{E}\big[\big(Z_\ell^\delta\big)^2\big] &= \xi'\big(q_\ell^\delta\big)] \\
\mathbb{E}\big[\big(N_{\ell+1}^\delta - N_\ell^\delta\big) | \mathcal{F}_\ell^\delta\big] &= 0 \\
\mathbb{E}\big[\big(N_{\ell+1}^\delta - N_\ell^\delta\big)^2\big] &= \delta \\
\mathbb{E}\big[\big(N_\ell^\delta\big)^2\big] &= q_{\ell+1}^\delta.
\end{aligned}
$$

From Lemma 3.6 and the fact that $(Z_{\underline{\ell}}^\delta, Z_{\underline{\ell}+1}^\delta, \ldots)$ form a Gaussian process, it follows that there is a coupling with a standard Brownian motion $(B_t)_{t \in [0,1]}$ such that

$\int_0^{q_\ell^\delta} \sqrt{\xi''(t)} \mathrm{d}B_t = Z_\ell^\delta$ for each ℓ. Denote by $(\mathcal{F}_t)_{t \in [0,1]}$ the associated natural filtration. Recall that X_t is defined as the solution to the SDE

$$dX_t = \gamma_*(t) \partial_x \Phi_{\gamma_*}(t, X_t) \mathrm{d}t + \sqrt{\xi''(t)} \mathrm{d}B_t$$

with initialization $X_0 \sim \mathcal{L}_h$. Recalling Proposition 2.8, define processes $(N_t, Z_t)_{t \in [0,1]}$ by

$$N_t \equiv \partial_x \Phi_{\gamma_*}(t, X_t)$$
$$= \partial_x \Phi_{\gamma_*}(\underline{q}, X_{\underline{q}}) + \int_{\underline{q}}^t \sqrt{\xi''(s)} u(s, X_s) dB_s,$$
$$Z_t \equiv \int_0^t \sqrt{\xi''(s)} dB_s.$$

The next lemma states that these continuous-time processes are the $\delta \to 0$ limit of $(X_\ell^\delta, N_\ell^\delta, Z_\ell^\delta)_{\ell \geq \underline{\ell}}$.

Lemma 3.7. *Fix $\bar{q} \in (\underline{q}, 1)$. There exists a coupling between the families of triples $\{(Z_\ell^\delta, X_\ell^\delta, N_\ell^\delta)\}_{\ell \geq 0}$ and $\{(Z_t, X_t, N_t)\}_{t \geq 0}$ such that the following holds for a constant $C > 0$. For large enough $\underline{\ell}$, and every $\ell \geq \underline{\ell}$ with $q_\ell \leq \bar{q}$,*

$$\max_{\underline{\ell} \leq j \leq \ell} \mathbb{E}\left[\left(X_j^\delta - X_{q_j}\right)^2\right] \leq C\delta,$$
$$\max_{\underline{\ell} \leq j \leq \ell} \mathbb{E}\left[\left(N_j^\delta - N_{q_j}\right)^2\right] \leq C\delta.$$

Lemmas 3.6 and 3.7 are proved in Section 8.

3.3 Computing the final energy

In this subsection we establish Theorem 1.7 by showing $\lim_{\bar{q} \to 1} \lim_{\underline{\ell} \to \infty} \text{p-}\lim_{N \to \infty} \frac{H_N(\boldsymbol{\sigma})}{N} = \mathsf{P}(\gamma_*)$. First we show that the replacements $\boldsymbol{m}^\ell \to \boldsymbol{n}^\ell$ and $\boldsymbol{n}^{\overline{\ell}} \to \boldsymbol{\sigma}$ have negligible effect on the asymptotic value attained.

Lemma 3.8.

$$\lim_{\bar{q} \to 1} \lim_{\underline{\ell} \to \infty} \text{p-}\lim_{N \to \infty} \left| \frac{H_N(\boldsymbol{\sigma}) - H_N(\boldsymbol{n}^{\overline{\ell}})}{N} \right| = 0, \tag{3.14}$$

$$\lim_{\underline{\ell} \to \infty} \text{p-}\lim_{N \to \infty} \left| \frac{H_N(\boldsymbol{m}^\ell) - H_N(\boldsymbol{n}^\ell)}{N} \right| = 0. \tag{3.15}$$

Proof. Proposition 2.7 implies that $N_t \in [-1, 1]$ almost surely, while optimizability of γ_* implies that $\mathbb{E}[(N_t)^2] = t$ for $t \in [\underline{q}, \bar{q}]$. It follows that

$$\lim_{\bar{q} \to 1} \lim_{\underline{\ell} \to \infty} \text{p-}\lim_{N \to \infty} \|\boldsymbol{n}^{\overline{\ell}} - \text{sign}(\boldsymbol{n}^{\overline{\ell}})\|_N = \lim_{\bar{q} \to 1} \sqrt{\mathbb{E}\left[\left(N_{\bar{q}} - \text{sign}(N_{\bar{q}})\right)^2\right]}$$
$$= 0.$$

The limit (3.14) follows from Proposition 2.1 with $k = 1$. (3.15) follows similarly as

$$\boldsymbol{n}^\ell - \boldsymbol{m}^\ell = \varepsilon_0 \boldsymbol{m}^\ell$$

and $\varepsilon_0 \to 0$ as $\underline{\ell} \to \infty$ while $\text{p-}\lim_{N \to \infty} \|\boldsymbol{m}^\ell\|_N \leq 1$ thanks to Proposition 2.7. $\square$

In the next lemma, proved in Section 8, we compute the energy gain during IAMP.

Lemma 3.9.

$$\lim_{\bar{q} \to 1} \lim_{\underline{\ell} \to \infty} \text{p-}\lim_{N \to \infty} \frac{H_N(\boldsymbol{n}^{\overline{\ell}}) - H_N(\boldsymbol{n}^\ell)}{N} = \int_{\underline{q}}^1 \xi''(t) \mathbb{E}\left[u(t, X_t)\right] \mathrm{d}t. \tag{3.16}$$

We now put everything together. Recall from Lemma 3.5 that

$$\lim_{\ell\to\infty} \operatorname{p-lim}_{N\to\infty} \frac{H_N(\boldsymbol{m}^\ell)}{N} = \xi'(\underline{q}) \cdot \mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*}(\underline{q}, X_{\underline{q}})\big] + \mathbb{E}\big[h \cdot \partial_x \Phi_{\gamma_*}(\underline{q}, X_{\underline{q}})\big].$$

Proposition 2.8 implies that the process $\partial_x \Phi_{\gamma_*}(t, X_t)$ is a martingale, while Lemma 2.9 states that $\mathbb{E}[u(t, X_t)] = \mathbb{E}[\partial_{xx}\Phi_{\gamma_*}(t, X_t)] \geq \int_t^1 \gamma_*(s)\mathrm{d}s$. Substituting, we find

$$\lim_{\ell\to\infty} \operatorname{p-lim}_{N\to\infty} \frac{H_N(\boldsymbol{m}^\ell)}{N} \geq \xi'(\underline{q}) \int_{\underline{q}}^1 \gamma_*(s)\mathrm{d}s + \mathbb{E}\big[h \cdot \partial_x \Phi_{\gamma_*}(0, h)\big].$$

Using again that $\mathbb{E}[u(t, X_t)] = \int_t^1 \gamma_*(s)\mathrm{d}s$, the right-hand side of (3.16) is

$$\begin{aligned}
\lim_{\bar{q}\to 1} \lim_{\ell\to\infty} \operatorname{p-lim}_{N\to\infty} \frac{H_N(\boldsymbol{n}^{\bar{\ell}}) - H_N(\boldsymbol{n}^\ell)}{N} &\geq \int_{\underline{q}}^1 \xi''(t) \int_t^1 \gamma_*(s)\mathrm{d}s\mathrm{d}t \\
&= \int_{\underline{q}}^1 \int_{\underline{q}}^s \xi''(t)\gamma_*(s)\mathrm{d}t\mathrm{d}s \\
&= \int_{\underline{q}}^1 \big(\xi'(s) - \xi'(\underline{q})\big)\gamma_*(s)\mathrm{d}s \\
&= \int_0^1 \xi'(s)\gamma_*(s)\mathrm{d}s - \xi'(\underline{q}) \int_{\underline{q}}^1 \gamma_*(s)\mathrm{d}s.
\end{aligned}$$

Combining with Lemma 3.8 yields

$$\begin{aligned}
\lim_{\bar{q}\to 1} \lim_{\ell\to\infty} \operatorname{p-lim}_{N\to\infty} \frac{H_N(\boldsymbol{\sigma})}{N} &\geq \lim_{\bar{q}\to 1} \lim_{\ell\to\infty} \operatorname{p-lim}_{N\to\infty} \frac{1}{N} \cdot \big(H_N\big(\operatorname{sign}(\boldsymbol{n}^{\bar{\ell}})\big) - H_N\big(\boldsymbol{n}^{\bar{\ell}}\big) \\
&\quad + H_N\big(\boldsymbol{n}^{\bar{\ell}}\big) - H_N\big(\boldsymbol{n}^\ell\big) + H_N\big(\boldsymbol{n}^\ell\big) - H_N\big(\boldsymbol{m}^\ell\big) + H_N\big(\boldsymbol{m}^\ell\big)\big) \\
&\geq \mathbb{E}^{h\sim\mathcal{L}_h}\big[h \cdot \partial_x \Phi_{\gamma_*}(0, h)\big] + \int_0^1 \xi'(s)\gamma_*(s)\mathrm{d}s. \qquad (3.17)
\end{aligned}$$

Having estimated the limiting energy achieved by our q-AMP algorithm, it remains to verify that the value in Equation (3.17) is $\mathsf{P}_{\xi,h}(\gamma_*)$. Define

$$\Psi_{\gamma_*}(t, x) = \Phi_{\gamma_*}(t, x) - x\partial_x \Phi_{\gamma_*}(t, x)$$

for $(t, x) \in [0, 1] \times \mathbb{R}$. We also use some identities computed in [AMS21].

Proposition 3.10 ([AMS21, Lemma 6.13]). *For any $\gamma \in \mathscr{L}$, the following identities hold:*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big[\Phi_\gamma(t, X_t)\big] = \frac{1}{2}\xi''(t)\gamma(t)\mathbb{E}\big[\partial_x \Phi_\gamma(t, X_t)^2\big]$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big[X_t \partial_x \Phi_\gamma(t, X_t)\big] = \xi''(t)\gamma(t)\mathbb{E}\big[\partial_x \Phi_\gamma(t, X_t)^2\big] + \xi''(t)\mathbb{E}\big[\partial_{xx}\Phi_\gamma(t, X_t)\big].$$

Lemma 3.11. *For $h \sim \mathcal{L}_h$, $\bar{q} \geq \underline{q}$, and X_t as in (1.1),*

$$\begin{aligned}
\mathbb{E}\big[\Phi_{\gamma_*}(0, h)\big] = \mathbb{E}\big[h \cdot \partial_x \Phi_{\gamma_*}(0, h)\big] &+ \mathbb{E}\big[\Psi_{\gamma_*}(\bar{q}, X_{\bar{q}})\big] \\
&+ \frac{1}{2}\int_0^{\bar{q}} \xi''(t)\gamma_*(t)\mathbb{E}\big[\partial_x \Phi_{\gamma_*}(t, X_t)^2\big]\mathrm{d}t + \int_0^{\bar{q}} \xi''(t)\mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*}(t, X_t)\big]\mathrm{d}t.
\end{aligned}$$

Proof. We write

$$
\mathbb{E}\big[\Psi_{\gamma_*}(\bar{q}, X_{\bar{q}}) - \Psi_{\gamma_*}(0, X_0)\big] = \int_0^{\bar{q}} \frac{\mathrm{d}}{\mathrm{d}s} \mathbb{E}\big[\Psi_{\gamma_*}(s, X_s)\big]\big|_{s=t} \mathrm{d}t
$$

$$
= \int_0^{\bar{q}} \frac{\mathrm{d}}{\mathrm{d}s} \mathbb{E}\big[\Phi_{\gamma_*}(s, X_s) - X_s \partial_x \Phi_{\gamma_*}(s, X_s)\big]\big|_{s=t} \mathrm{d}t
$$

$$
= -\frac{1}{2} \int_0^{\bar{q}} \xi''(t)\gamma_*(t)\mathbb{E}\big[\partial_x \Phi_{\gamma_*}(t, X_t)^2\big]\mathrm{d}t
$$

$$
- \int_0^{\bar{q}} \xi''(t)\mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*}(t, X_t)\big]\mathrm{d}t.
$$

Rearranging shows:

$$
\mathbb{E}\big[\Phi_{\gamma_*}(0, X_0)\big] = \mathbb{E}\big[X_0 \partial_x \Phi_{\gamma_*}(0, X_0)\big] + \mathbb{E}\big[\Psi_{\gamma_*}(\bar{q}, X_{\bar{q}})\big]
$$

$$
+ \frac{1}{2} \int_0^{\bar{q}} \xi''(t)\gamma_*(t)\mathbb{E}\big[\partial_x \Phi_{\gamma_*}(t, X_t)^2\big]\mathrm{d}t + \int_0^{\bar{q}} \xi''(t)\mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*}(t, X_t)\big]\mathrm{d}t.
$$

As $X_0 = h$ this concludes the proof. $\qquad\square$

Proof of Theorem 1.7. Note that $\gamma_*(t) > 0$ implies $t \geq \underline{q}$ and hence by optimizability

$$
\mathbb{E}\big[\big(\partial_x \Phi_{\gamma_*}(t, X_t)\big)^2\big] = t.
$$

Meanwhile for any $t \in [0, 1]$,

$$
\mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*}(t, X_t)\big] = \int_t^1 \gamma_*(t)\mathrm{d}t.
$$

Therefore

$$
\Phi_{\gamma_*}(0, h) = h \cdot \partial_x \Phi_{\gamma_*}(0, h) + \mathbb{E}\big[\Psi_{\gamma_*}(\bar{q}, X_{\bar{q}})\big] + \frac{1}{2}\int_0^{\bar{q}} \xi''(t)\gamma_*(t)t\,\mathrm{d}t + \int_0^{\bar{q}} \xi''(t)\int_t^1 \gamma_*(s)\mathrm{d}s\,\mathrm{d}t.
$$

Recalling (1.2), we find

$$
\mathsf{P}(\gamma_*) = \mathbb{E}^h\big[\Phi_{\gamma_*}(0, h)\big] - \frac{1}{2}\int_0^1 \xi''(t)\gamma_*(t)t\mathrm{d}t
$$

$$
= \mathbb{E}^h\big[h \cdot \partial_x \Phi_{\gamma_*}(0, h)\big] + \mathbb{E}\big[\Psi_{\gamma_*}(\bar{q}, X_{\bar{q}})\big] + \int_0^{\bar{q}} \xi''(t)\int_t^1 \gamma_*(s)\mathrm{d}s\,\mathrm{d}t + o_{\bar{q}\to 1}(1).
$$

As in [AMS21, Proof of Theorem 3.2] $\lim_{\bar{q}\to 1}\Psi_{\gamma_*}(\bar{q}, x) = 0$ holds uniformly in x. Moreover

$$
\lim_{\bar{q}\to 1}\int_0^{\bar{q}} \xi''(t)\int_t^1 \gamma_*(s)\mathrm{d}s\,\mathrm{d}t = \int_0^1 \xi''(t)\int_t^1 \gamma_*(s)\mathrm{d}s\,\mathrm{d}t
$$

$$
= \int_0^1 \int_0^s \xi''(t)\gamma_*(s)\mathrm{d}t\,\mathrm{d}s
$$

$$
= \int_0^1 \xi'(s)\gamma_*(s)\mathrm{d}s.
$$

Combining the above and comparing with (3.17) yields

$$
\mathsf{P}(\gamma_*) = \mathbb{E}^h\big[h \cdot \partial_x \Phi_{\gamma_*}(0, h)\big] + \int_0^1 \xi'(s)\gamma_*(s)\mathrm{d}s
$$

$$
\leq \lim_{\bar{q}\to 1}\lim_{\ell\to\infty}\operatorname*{p-lim}_{N\to\infty} \frac{H_N(\boldsymbol{\sigma})}{N}.
$$

This completes the proof of Theorem 1.7. $\qquad\square$

4 Constructing many approximate maximizers

Here we explain the modifications needed for branching IAMP and Theorem 1.13. The proofs are a slight extension of those for the main algorithm, and in fact we give many proofs for IAMP directly in this more general setting in Section 8. Let us fix values $Q = (q_1, \ldots, q_m)$ with $\underline{q} \leq q_1 < \cdots < q_m < 1$ and an index $B \in [m]$. To construct a pair of approximate maximizers with overlap q_B we first construct $\boldsymbol{n}^{\underline{\ell}}$ exactly as in Subsection 3.1. For each $i < B$, set $\boldsymbol{g}^{(q_i,1)} = \boldsymbol{g}^{(q_i,2)} = \boldsymbol{z}^{-k_{i,1}} = \boldsymbol{z}^{-k_{i,2}} \in \mathbb{R}^N$ for some $k_{i,1} = k_{i,2} \leq K$ as in Subsection 2.2. For each $B \leq i \leq m$, set $\boldsymbol{g}^{(q_i,1)} = \boldsymbol{z}^{-k_{i,1}}$ and $\boldsymbol{g}^{(q_i,2)} = \boldsymbol{z}^{-k_{i,2}}$ where $k_{i,1} \neq k_{i,2}$. Because the vectors $\boldsymbol{g}^{(q_i,a)}$ are constructed using AMP, we require some additional conditions. We impose the separation condition

$$k_{i,a'} - \ell_{q_i}^\delta > k_{j,a} - \ell_{q_j}^\delta > 0 \tag{4.1}$$

for all $i > j$ and $a, a' \in \{1, 2\}$. (In particular, it implies that $k_{i,a'} \neq k_{j,a}$ for $i \neq j$.) It is not hard to satisfy (4.1) by choosing the values $k_{i,a}$ sequentially in increasing order of i. Finally we insist that $\max_{i,a}(k_{i,a}) + \bar{\ell} + 1 < K$, where $h = \boldsymbol{z}^{-K}$ was the AMP initialization, which is satisfied by choosing K large at the end.

Having fixed this setup, we proceed by defining $\boldsymbol{m}^{k,1} = \boldsymbol{m}^{k,2} = \boldsymbol{m}^k$ for $k \geq 0$ exactly as in the original first phase. Next we generate two sequences of IAMP iterates using (3.12) except at times corresponding to $q_i \in Q$, altogether generating $\boldsymbol{n}^{\ell,a}$ for $\ell > \underline{\ell}$ and $a \in \{1, 2\}$ via:

$$\boldsymbol{n}^{\ell,a} = \begin{cases} \boldsymbol{n}^{\ell-1,a} + \sqrt{\delta}\boldsymbol{g}^{(q_i,a)}, & \ell = \ell_{q_i}^\delta \equiv \underline{\ell} + \lceil (q_i - \underline{q})\delta \rceil + 1 \text{ for some } i \in [m] \\ \boldsymbol{n}^{\ell-1,a} + u_{\ell-1}^\delta(\boldsymbol{x}^{\ell-1,a})(\boldsymbol{z}^{\ell,a} - \boldsymbol{z}^{\ell-1,a}), & \text{else.} \end{cases} \tag{4.2}$$

Recalling Subsection 2.2, this is an AMP algorithm of the required form. The definitions of $\boldsymbol{x}^{\ell,a}, \boldsymbol{z}^{\ell,a}$ are the same as in e.g. (3.11) with superscript a everywhere, though note that now the definition

$$f_{j-1}(\boldsymbol{z}^{-K}, \ldots, \boldsymbol{z}^0, \boldsymbol{z}^{1,a}, \ldots, \boldsymbol{z}^{j-1,a}) = \boldsymbol{n}^{j-1,a}$$

of f_{j-1} has explicit dependence on the negatively indexed variables through $\boldsymbol{g}^{(q_i,a)}$. The following result follows immediately from Lemmas 8.5, 8.6 and readily implies Theorem 1.13.

Lemma 4.1. *For optimizable γ_*,*

$$\lim_{\bar{q} \to 1} \lim_{\underline{\ell} \to \infty} \operatorname*{p-lim}_{N \to \infty} \frac{H_N(\boldsymbol{n}^{\bar{\ell},a})}{N} \geq \mathsf{P}(\gamma_*), \quad a \in \{1, 2\}$$

$$\lim_{\bar{q} \to 1} \lim_{\underline{\ell} \to \infty} \operatorname*{p-lim}_{N \to \infty} \frac{\langle \boldsymbol{n}^{\bar{\ell},1}, \boldsymbol{n}^{\bar{\ell},2} \rangle}{N} = q_B.$$

Next we extend this construction to general finite ultrametric spaces X. Recall that any finite ultrametric space X with all pairwise distances in the set $\{\sqrt{2(1-q_i)}\}_{i \in [m]}$ can be identified with a rooted tree $\mathcal{T}$ whose leaves $\partial\mathcal{T}$ are in bijection with X, and so that $d_X(x_i, x_j) = \sqrt{2(1-q_k)}$ is equivalent to leaves i, j having least common ancestor at depth k.

Given such $\mathcal{T}$, we may assign to each non-root vertex $u \in \mathcal{T}$ a distinct initialization iterate $\boldsymbol{g}^{(u)} = \boldsymbol{z}^{-k_u}$. We require that

1. The k_u are pairwise distinct.
2. Analogously to (4.1),

$$k_{u'} - \ell_{q_i}^\delta > k_u - \ell_{q_j}^\delta$$

if $q_j = \mathrm{depth}(u) < \mathrm{depth}(u') = q_i$.

3. $\max_u(k_u) + \overline{\ell} + 1 < K$.

Again, these conditions are easy to satisfy by choosing k_u in increasing order of depth(u) with ties broken arbitrarily.

Then for each $x \in X$, we first construct $\boldsymbol{m}^{k,x} = \boldsymbol{m}^k$ for $k \geq 0$ exactly as in the original first phase, which does not depend on x. Next, denoting by

$$\text{root} = v_0, v_1, \ldots, v_m = x \in \partial\mathcal{T} = X$$

the root-to-leaf path for x within $\mathcal{T}$, we compute the analog of (4.2):

$$\boldsymbol{n}^{\ell,x} = \begin{cases} \boldsymbol{n}^{\ell-1,x} + \sqrt{\delta}\boldsymbol{g}^{(k_{v_i})}, & \ell = \ell^{\delta}_{q_i} \text{ for some } i \in [m] \\ \boldsymbol{n}^{\ell-1,x} + u^{\delta}_{\ell-1}\big(\boldsymbol{x}^{\ell-1,x}\big)\big(\boldsymbol{z}^{\ell,x} - \boldsymbol{z}^{\ell-1,x}\big), & \text{else.} \end{cases} \tag{4.3}$$

Again, $\boldsymbol{x}^{\ell,x}, \boldsymbol{z}^{\ell,x}$ are defined using the same recursions as before.

Theorem 1.13. *Let $\gamma_* \in \mathscr{L}$ be optimizable, and fix a finite ultrametric space (X, d_X) with diameter at most $\sqrt{2(1 - \underline{q})}$ as well as $\varepsilon > 0$. Then an efficient AMP algorithm constructs points $\{\boldsymbol{\sigma}_x | x \in X\}$ in $\boldsymbol{\Sigma}_N$ satisfying*

$$\frac{H_N(\boldsymbol{\sigma}_x)}{N} \geq \mathsf{P}(\gamma_*) - \varepsilon, \quad x \in X,$$

$$\frac{\|\boldsymbol{\sigma}_x - \boldsymbol{\sigma}_y\|}{\sqrt{N}} \in \big[d_X(x, y) - \varepsilon, d_X(x, y) + \varepsilon\big], \quad x, y \in X$$

with probability tending to 1 as $N \to \infty$.

Proof. It is easy to see that for each distinct $x, y \in X$, the behavior of the pair $\boldsymbol{n}^{\ell,x}, \boldsymbol{n}^{\ell,y}$ in (4.3) is identical to $\boldsymbol{n}^{\ell,1}, \boldsymbol{n}^{\ell,2}$ in (4.2) (e.g. both pairs have the same joint law with H_N). Applying Lemma 4.1 to all such pairs, we find that the iterates $\boldsymbol{n}^{\overline{\ell},x}$ satisfy $\frac{H_N(\boldsymbol{n}^{\overline{\ell},x})}{N} \geq \mathsf{P}(\gamma_*) - \varepsilon$ and $\langle \boldsymbol{n}^{\overline{\ell},x}, \boldsymbol{n}^{\overline{\ell},y} \rangle_N \simeq q_j$ if $d_X(x, y) = \sqrt{2(1 - q_j)}$. The conclusion follows by rounding $\boldsymbol{n}^{\overline{\ell},x} \to \boldsymbol{\sigma}_x \in \boldsymbol{\Sigma}_N$ for each $x \in X$ as in the main algorithm. $\qquad\square$

We remark that our construction differs from the one proposed in [AM20] only because we construct the vectors $\boldsymbol{g}^{(u)}$ using AMP rather than taking them to be literally independent Gaussian vectors. While the latter construction almost certainly works as well, the analysis seems to require a more general version of state evolution.

5 Spherical models

We now consider the case of spherical spin glasses with external field. The law of the Hamiltonian H_N is specified according to the same formula as before depending on $(\xi, \mathcal{L})$, however the state space is the sphere $\mathbb{S}^{N-1}(\sqrt{N})$ instead of the hypercube. The free energy in this case is given by a similar Parisi-type formula, however it turns out to dramatically simplify under no overlap gap so we do not give the general formula. At positive temperature the spherical free energy was computed non-rigorously in [CS92] and rigorously in [Tal06a, Che13], but we rely on [CS17] which directly treats the zero-temperature setting.

Remark 5.1. Due to rotational invariance, for spherical models all that matters about $\mathcal{L}_h$ is the squared L^2 norm $\mathbb{E}^{h \sim \mathcal{L}_h}[h^2]$. In particular unlike the Ising case there is no loss of generality in assuming h is constant. We continue to work with coordinates h_i sampled i.i.d. from $\mathcal{L}_h$ and implicitly use this observation when interpreting the results of [CS17].

Our treatment of spherical models is less detailed and we simply show how to obtain the energy value in Theorem 1.15 which is the ground state in models with $\operatorname{supp}(\gamma_*) = [q, 1]$. In the case that $\mathbb{E}[h^2] + \xi'(1) < \xi''(1)$, we let $\underline{q}_{\mathsf{sph}} \in [0, 1]$ be the unique solution to

$$\underline{q}_{\mathsf{sph}} \xi''(\underline{q}_{\mathsf{sph}}) = \mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}}).$$

When $\mathbb{E}[h^2] + \xi'(1) \geq \xi''(1)$, we simply set $\underline{q}_{\mathsf{sph}} = 1$.

Note that when $h = 0$ almost surely it follows that $\underline{q}_{\mathsf{sph}} = 0$, which is the setting of [Sub21]. Generate initial iterates

$$\left(\boldsymbol{w}_{\mathsf{sph}}^{-K}, \ldots, \boldsymbol{w}_{\mathsf{sph}}^0 \right)$$

as in Subsection 2.2. For non-zero h we take $c = \sqrt{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})}$ so that

$$\|\boldsymbol{w}_{\mathsf{sph}}^0\|_N \simeq \sqrt{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})}.$$

Generate further iterates via the following AMP iteration.

$$\boldsymbol{w}_{\mathsf{sph}}^{k+1} = \nabla \widetilde{H}_N \left(\boldsymbol{m}_{\mathsf{sph}}^k \right) - \boldsymbol{m}_{\mathsf{sph}}^{k-1} \xi'' \left(\langle \boldsymbol{m}_{\mathsf{sph}}^k, \boldsymbol{m}_{\mathsf{sph}}^{k-1} \rangle \right) \sqrt{\frac{\underline{q}_{\mathsf{sph}}}{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})}} \qquad (5.1)$$

$$\boldsymbol{x}_{\mathsf{sph}}^k = \boldsymbol{w}_{\mathsf{sph}}^k + \boldsymbol{h}$$

$$\boldsymbol{m}_{\mathsf{sph}}^k = \boldsymbol{x}_{\mathsf{sph}}^k \cdot \sqrt{\frac{\underline{q}_{\mathsf{sph}}}{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})}}.$$

The next lemma is the spherical analog of Lemmas 3.3, 3.4, 3.5 – the proof is similar to the Ising case and is given in the next subsection.

Lemma 5.2. *Using the AMP of* (5.1)*, the asymptotic overlaps and energies satisfy*

$$\lim_{k,\ell \to \infty} \operatorname*{p-lim}_{N \to \infty} \frac{\langle \boldsymbol{w}_{\mathsf{sph}}^k, \boldsymbol{w}_{\mathsf{sph}}^\ell \rangle}{N} = \xi'(\underline{q}_{\mathsf{sph}}),$$

$$\lim_{k,\ell \to \infty} \operatorname*{p-lim}_{N \to \infty} \frac{\langle \boldsymbol{x}_{\mathsf{sph}}^k, \boldsymbol{x}_{\mathsf{sph}}^\ell \rangle}{N} = \mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}}),$$

$$\lim_{k,\ell \to \infty} \operatorname*{p-lim}_{N \to \infty} \frac{\langle \boldsymbol{m}_{\mathsf{sph}}^k, \boldsymbol{m}_{\mathsf{sph}}^\ell \rangle}{N} = \underline{q}_{\mathsf{sph}},$$

$$\lim_{k,\ell \to \infty} \operatorname*{p-lim}_{N \to \infty} \frac{H_N(\boldsymbol{m}_{\mathsf{sph}}^k)}{N} = \sqrt{\underline{q}_{\mathsf{sph}} \left(\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}}) \right)}. \qquad (5.2)$$

Proof of Theorem 1.16. The latter two parts of Lemma 5.2 directly imply Theorem 1.16 in the case that $\mathbb{E}[h^2] + \xi'(1) \geq \xi''(1)$ (recall $\underline{q}_{\mathsf{sph}} = 1$ in this case). Indeed, it suffices to take

$$\boldsymbol{\sigma}_{\mathsf{sph}} = \frac{\boldsymbol{m}_{\mathsf{sph}}^{\overline{\ell}}}{\|\boldsymbol{m}_{\mathsf{sph}}^{\overline{\ell}}\|_N} \in \mathbb{S}^{N-1}(\sqrt{N}) \qquad (5.3)$$

for a sufficiently large constant $\overline{\ell}$. When $\mathbb{E}[h^2] + \xi'(1) < \xi''(1)$, we can conclude by mimicking the IAMP phase using the simple non-linearities $u(t, x) = u(t) = \xi''(t)^{-1/2}$ and $v(t, x) = 0$ – see also [AMS21, Remark 2.2]. Lemma 3.9 then shows the energy gain from IAMP is

$$\int_{\underline{q}}^1 \xi''(t) u(t) \mathrm{d}t = \int_{\underline{q}}^1 \xi''(t)^{1/2}.$$

As in the Ising case, we may start IAMP from $\boldsymbol{m} = \boldsymbol{m}^k$ for a large constant k. Combining with (5.2) and defining $\boldsymbol{\sigma}_{\mathsf{sph}}$ via (5.3) with $\boldsymbol{m}^{\bar{\ell}}$ an IAMP iterate, we obtain

$$\operatorname*{p-lim}_{N \to \infty} \frac{H_N(\boldsymbol{\sigma}_{\mathsf{sph}})}{N} \geq \underline{q}_{\mathsf{sph}} \xi''(\underline{q}_{\mathsf{sph}})^{1/2} + \int_{\underline{q}_{\mathsf{sph}}}^1 \xi''(\underline{q}_{\mathsf{sph}})^{1/2} \mathrm{d}q.$$

Alternatively to IAMP, in the spherical setting it is possible to use the approach of [Sub21]. Indeed [Sub21, Theorem 4] immediately extends to an algorithm taking in an arbitrary point $\boldsymbol{m}$ with $\|\boldsymbol{m}\|_N \leq 1$ and outputting a point $\boldsymbol{m}_* \in \mathbb{S}^{N-1}(\sqrt{N})$ (which may depend on H_N) satisfying

$$\frac{H_N(\boldsymbol{m}_*) - H_N(\boldsymbol{m})}{N} \geq \int_{\|\boldsymbol{m}\|_N^2}^1 \sqrt{\xi''(q)} \mathrm{d}q - \varepsilon$$

with probability $1 - o_{N \to \infty}(1)$ for any desired $\varepsilon > 0$. Either approach completes the proof of Theorem 1.16. $\qquad\square$

5.1 Proof of Lemma 5.2

For $t \in [0, \xi'(\underline{q}_{\mathsf{sph}})]$, take $h \sim \mathcal{L}_h$ and $(Z, Z', Z'') \sim \mathsf{N}(0, I_3)$ and define the function

$$\begin{aligned}
\phi_{\mathsf{sph}}(t) &= \frac{\underline{q}_{\mathsf{sph}}(\mathbb{E}[h^2] + t)}{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})} \\
&= \frac{\underline{q}_{\mathsf{sph}}}{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})} \\
&\quad \cdot \mathbb{E}^{h,Z,Z',Z''}\left[\left(h + Z\sqrt{t} + Z'\sqrt{\xi'(\underline{q}_{\mathsf{sph}}) - t}\right)\left(h + Z\sqrt{t} + Z''\sqrt{\xi'(\underline{q}_{\mathsf{sph}}) - t}\right)\right]
\end{aligned}$$

so that $\phi_{\mathsf{sph}}(\xi'(\underline{q}_{\mathsf{sph}})) = \underline{q}_{\mathsf{sph}}$. Define $\psi_{\mathsf{sph}}(t) = \xi'(\phi_{\mathsf{sph}}(t))$.

Lemma 5.3. ψ_{sph} *is strictly increasing and convex on* $[0, \xi'(\underline{q}_{\mathsf{sph}})]$ *and*

$$\psi_{\mathsf{sph}}\big(\xi'(\underline{q}_{\mathsf{sph}})\big) = \xi'(\underline{q}_{\mathsf{sph}}), \tag{5.4}$$

$$\psi'_{\mathsf{sph}}\big(\xi'(\underline{q}_{\mathsf{sph}})\big) = 1, \tag{5.5}$$

$$\psi_{\mathsf{sph}}(t) > t, \qquad \forall t < \xi'(\underline{q}_{\mathsf{sph}}). \tag{5.6}$$

Proof. Since ξ' is strictly increasing and convex and ϕ_{sph} is affine and increasing, it follows that ψ_{sph} is strictly increasing and convex. (5.4) is equivalent to the equation $\underline{q}_{\mathsf{sph}} \xi''(\underline{q}_{\mathsf{sph}}) = \mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})$ defining $\underline{q}_{\mathsf{sph}}$. To show (5.5) we use the chain rule to write

$$\psi'_{\mathsf{sph}}\big(\xi'(\underline{q}_{\mathsf{sph}})\big) = \xi''\big(\phi_{\mathsf{sph}}\big(\xi'(\underline{q}_{\mathsf{sph}})\big)\big) \cdot \phi'_{\mathsf{sph}}\big(\xi'(\underline{q}_{\mathsf{sph}})\big) = \xi''(\underline{q}_{\mathsf{sph}}) \cdot \big(\xi''(\underline{q}_{\mathsf{sph}})\big)^{-1} = 1.$$

Equations (5.4) and (5.5) and the convexity of ψ_{sph} just shown imply (5.6) $\qquad\square$

Let $h, W_{\mathsf{sph}}^{-1}, (W_{\mathsf{sph}}^j, X_{\mathsf{sph}}^j, M_{\mathsf{sph}}^j)_{j=0}^k$ be the state evolution limit of the coordinates of

$$\big(\boldsymbol{h}, \boldsymbol{w}_{\mathsf{sph}}^{-1}, \boldsymbol{w}_{\mathsf{sph}}^0, \boldsymbol{x}_{\mathsf{sph}}^0, \boldsymbol{m}_{\mathsf{sph}}^0, \ldots, \boldsymbol{w}_{\mathsf{sph}}^k, \boldsymbol{x}_{\mathsf{sph}}^k, \boldsymbol{m}_{\mathsf{sph}}^k\big)$$

as $N \to \infty$. Define the sequence $(b_0, b_1, \ldots)$ recursively by $b_0 = 0$ and $b_{k+1} = \psi_{\mathsf{sph}}(b_k)$.

Lemma 5.4. *For all non-negative integers* $0 \leq j < k$ *the following equalities hold:*

$$\mathbb{E}\big[(W_{\mathsf{sph}}^j)^2\big] = \xi'(\underline{q}_{\mathsf{sph}})$$

$$\mathbb{E}\big[W_{\mathsf{sph}}^j W_{\mathsf{sph}}^k\big] = b_j$$

$$\mathbb{E}\big[(M_{\mathsf{sph}}^j)^2\big] = \underline{q}_{\mathsf{sph}}$$

$$\mathbb{E}\big[M_{\mathsf{sph}}^j M_{\mathsf{sph}}^k\big] = \phi_{\mathsf{sph}}(b_j).$$

Proof. Follows from state evolution and induction exactly as in Lemma 3.3. □

Lemma 5.5.

$$\lim_{k \to \infty} b_k = \xi'(\underline{q}_{\mathsf{sph}}),$$

$$\lim_{k \to \infty} \phi_{\mathsf{sph}}(b_k) = \underline{q}_{\mathsf{sph}}.$$

Proof. As in the proof of Lemma 3.4, the sequence $b_1, b_2, \ldots$, must converge up to a limit, and this limit must be a fixed point for ψ_{sph}, implying the first claim. The second claim follows by continuity of ϕ_{sph}. □

Lemma 5.6.

$$\lim_{k \to \infty} \operatorname*{p-lim}_{N \to \infty} \frac{H_N(\boldsymbol{m}_{\mathsf{sph}}^k)}{N} = \sqrt{\underline{q}_{\mathsf{sph}}\big(\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})\big)}.$$

Proof. We use again the identity

$$\frac{H_N(\boldsymbol{m}_{\mathsf{sph}}^k)}{N} = \langle \boldsymbol{h}, \boldsymbol{m}_{\mathsf{sph}}^k \rangle_N + \int_0^1 \langle \boldsymbol{m}_{\mathsf{sph}}^k, \nabla \widetilde{H}_N(t\boldsymbol{m}_{\mathsf{sph}}^k) \rangle_N \mathrm{d}t$$

and interchange the limit in probability with the integral. To compute the main term $\operatorname{p-lim}_{N \to \infty} \langle \boldsymbol{m}_{\mathsf{sph}}^k, \nabla \widetilde{H}_N((t\boldsymbol{m}_{\mathsf{sph}}^k) \rangle$ we introduce an auxiliary AMP step

$$\boldsymbol{y}_{\mathsf{sph}}^{k+1} = \nabla \widetilde{H}_N(t\boldsymbol{m}_{\mathsf{sph}}^k) - t\boldsymbol{m}_{\mathsf{sph}}^{k-1} \xi''(t\langle \boldsymbol{m}_{\mathsf{sph}}^k, \boldsymbol{m}_{\mathsf{sph}}^{k-1} \rangle) \sqrt{\frac{\underline{q}_{\mathsf{sph}}}{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})}}.$$

Rearranging yields

$$\langle \boldsymbol{m}_{\mathsf{sph}}^k, \nabla \widetilde{H}_N(t\boldsymbol{m}_{\mathsf{sph}}^k) \rangle_N$$

$$= \langle \boldsymbol{m}_{\mathsf{sph}}^k, \boldsymbol{y}_{\mathsf{sph}}^{k+1} \rangle_N + t\langle \boldsymbol{m}_{\mathsf{sph}}^k, \boldsymbol{m}_{\mathsf{sph}}^{k-1} \rangle_N \xi''(t\langle \boldsymbol{m}_{\mathsf{sph}}^k, \boldsymbol{m}_{\mathsf{sph}}^{k-1} \rangle_N) \sqrt{\frac{\underline{q}_{\mathsf{sph}}}{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})}}$$

$$\simeq \langle \boldsymbol{m}_{\mathsf{sph}}^k, \boldsymbol{y}_{\mathsf{sph}}^{k+1} \rangle_N + tb_{k-1}\xi''(t\phi(b_{k-1})) \sqrt{\frac{\underline{q}_{\mathsf{sph}}}{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})}}.$$

For the first term, Gaussian integration by parts with

$$g(x) = (x + h) \cdot \sqrt{\frac{\underline{q}_{\mathsf{sph}}}{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})}}$$

yields

$$\mathbb{E}[g(X_{\mathsf{sph}}^k)Y^{k+1}] = \mathbb{E}[g'(X_{\mathsf{sph}}^k)] \cdot \mathbb{E}[X_{\mathsf{sph}}^k Y_{\mathsf{sph}}^{k+1}] = \xi'(t\phi_{\mathsf{sph}}(b_{k-1})) \sqrt{\frac{\underline{q}_{\mathsf{sph}}}{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})}}.$$

Integrating with respect to t, we find

$$\int_0^1 \langle \boldsymbol{m}_{\mathsf{sph}}^k, \nabla \widetilde{H}_N(t\boldsymbol{m}_{\mathsf{sph}}^k) \rangle_N \mathrm{d}t$$

$$\simeq \mathbb{E}[g'(Z\sqrt{\xi'(\underline{q}_{\mathsf{sph}})})] \cdot \int_0^1 \xi'(t\phi_{\mathsf{sph}}(b_{k-1})) + t\phi_{\mathsf{sph}}(b_{k-1})\xi''(t\phi_{\mathsf{sph}}(b_{k-1}))$$

$$= [t\xi'(t\phi_{\mathsf{sph}}(b_{k-1}))]\big|_{t=0}^{t=1} \cdot \sqrt{\frac{\underline{q}_{\mathsf{sph}}}{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})}}$$

$$= \psi_{\mathsf{sph}}(b_{k-1}) \sqrt{\frac{\underline{q}_{\mathsf{sph}}}{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})}}.$$

Finally the first term gives energy contribution

$$h\langle \boldsymbol{m}_{\mathsf{sph}}^k\rangle_N \simeq \mathbb{E}\big[h\big(h + Z\sqrt{\xi'(b_{k-1})}\big)\big]\sqrt{\frac{\underline{q}_{\mathsf{sph}}}{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})}}$$

$$= \mathbb{E}\big[h^2\big]\sqrt{\frac{\underline{q}_{\mathsf{sph}}}{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})}}.$$

Since $\lim_{k\to\infty} b_{k-1} = \xi'(\underline{q}_{\mathsf{sph}})$ and $\psi_{\mathsf{sph}}(\xi'(\underline{q}_{\mathsf{sph}})) = \xi'(\underline{q}_{\mathsf{sph}})$ we conclude

$$\lim_{k\to\infty}\operatorname*{p-lim}_{N\to\infty}\frac{H_N(\boldsymbol{m}_{\mathsf{sph}}^k)}{N} = \sqrt{\underline{q}_{\mathsf{sph}}\big(\mathbb{E}\big[h^2\big] + \xi'(\underline{q}_{\mathsf{sph}})\big)}. \qquad \square$$

Proof of Lemma 5.2. The result follows from the preceding lemmas. $\qquad \square$

5.2 Proof of Theorem 1.15

It follows from our algorithm that $GS_{\mathsf{sph}}(\xi,\mathcal{L}_h) \geq \underline{q}_{\mathsf{sph}}\xi''(\underline{q}_{\mathsf{sph}})^{1/2} + \int_{\underline{q}_{\mathsf{sph}}}^1 \xi''(\underline{q}_{\mathsf{sph}})^{1/2}\mathrm{d}q$. We now characterize the models in which equality holds, which coincide with those exhibiting no overlap gap. Moreover we give an alternate proof of the lower bound for $GS(\xi,\mathcal{L}_h)_{\mathsf{sph}}$ which shows that equality holds exactly in no overlap gap models. We thank Wei-Kuo Chen for communicating the latter proof.

Theorem 1.15. *Suppose ξ and $\mathcal{L}_h$ satisfy $\mathbb{E}[h^2] + \xi'(1) < \xi''(1)$, and let $\underline{q}_{\mathsf{sph}} \in (0,1)$ be the unique solution to $\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}}) = \underline{q}_{\mathsf{sph}}\xi''(\underline{q}_{\mathsf{sph}})$. Then the spherical spin glass with parameters $\xi,\mathcal{L}_h$ has no overlap gap if and only if $\xi''(q)^{-1/2}$ is concave on $q \in [\underline{q}_{\mathsf{sph}},1]$, in which case α is supported on $[\underline{q}_{\mathsf{sph}},1]$ and takes the explicit form*

$$\alpha(s) = \begin{cases} 0, & s \in [0,\underline{q}_{\mathsf{sph}}) \\ \frac{\xi'''(s)}{2\xi''(s)^{3/2}}, & s \in [\underline{q}_{\mathsf{sph}},1]. \end{cases}$$

Moreover the ground-state energy satisfies

$$GS_{\mathsf{sph}}(\xi,\mathcal{L}_h) \geq \underline{q}_{\mathsf{sph}}\sqrt{\xi''(\underline{q}_{\mathsf{sph}})} + \int_{\underline{q}_{\mathsf{sph}}}^1 \sqrt{\xi''(q)}\mathrm{d}q$$

with equality if and only if no overlap gap occurs.

Proof. We use the results and notation of [CS17]. If $\xi''(q)^{-1/2}$ is concave on $[\underline{q}_{\mathsf{sph}},1]$ then the proof of Proposition 2 in [CS17] applies verbatim to show that the support of α is $[\underline{q}_{\mathsf{sph}},1]$. In fact it explicitly shows $\alpha(s) = \frac{\xi'''(s)}{2\xi''(s)^{3/2}}$ for $s \in [\underline{q}_{\mathsf{sph}},1]$).

In the other direction, we show that if no overlap gap holds and $\mathbb{E}[h^2] + \xi'(1) < \xi''(1)$, then $\xi''(q)^{-1/2}$ is concave on $[\underline{q}_{\mathsf{sph}},1]$. we use the statement and notation of [CS17, Theorem 2]. Assume α is supported on the interval $[\underline{q}_{\mathsf{sph}},1]$. The last condition in [CS17, Theorem 2] states that $g(u) = \int_u^1 \bar{g}(s)\mathrm{d}s = 0$ for all $u \in [\underline{q}_{\mathsf{sph}},1]$, and therefore $\bar{g}(s) = 0$ for $s \in [\underline{q}_{\mathsf{sph}},1]$, where

$$\bar{g}(s) \equiv \xi'(s) + h^2 - \int_0^s \frac{dq}{(L - \int_0^q \alpha(r)dr)^2}.$$

Setting $s = \underline{q}_{\mathsf{sph}}$ yields $\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}}) = \underline{q}_{\mathsf{sph}}L^{-2}$, i.e. $L = \sqrt{\frac{\underline{q}_{\mathsf{sph}}}{\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}})}}$. Differentiating, all $s \geq \underline{q}_{\mathsf{sph}}$ satisfy

$$\xi''(s) = \frac{1}{(L - \int_{\underline{q}_{\mathsf{sph}}}^s \alpha(r)dr)^2}. \tag{5.7}$$

Taking $s = \underline{q}_{\mathsf{sph}}$ in (5.7) shows $L = \xi''(\underline{q}_{\mathsf{sph}})^{-1/2}$, hence $\mathbb{E}[h^2] + \xi'(\underline{q}_{\mathsf{sph}}) = \underline{q}_{\mathsf{sph}}\xi''(\underline{q}_{\mathsf{sph}})$. Rearranging (5.7) yields

$$L - \int_{\underline{q}_{\mathsf{sph}}}^{s} \alpha(r)dr = \xi''(s)^{-1/2}, \qquad s \geq \underline{q}_{\mathsf{sph}}.$$

As α must be non-decreasing based on [CS17, Equation (9)] it follows that $\xi''(s)^{-1/2}$ is concave on $s \in [\underline{q}_{\mathsf{sph}}, 1]$. This completes the proof of the first equivalence. We turn to the value of $GS_{\mathsf{sph}}(\xi, \mathcal{L}_h)$, first computing

$$\mathbb{E}[h^2] + \xi'(1) = \underline{q}\xi''(\underline{q}_{\mathsf{sph}}) + \int_{\underline{q}_{\mathsf{sph}}}^{1} \xi''(\underline{q}_{\mathsf{sph}})\mathrm{d}q$$

$$= \int_{0}^{\underline{q}_{\mathsf{sph}}} \xi''(\underline{q}_{\mathsf{sph}})\mathrm{d}q + \int_{\underline{q}_{\mathsf{sph}}}^{1} \xi''(\underline{q}_{\mathsf{sph}})\mathrm{d}q.$$

Letting $L > \int_0^1 \alpha(s)\mathrm{d}s$ and let $a(q) = \int_0^q \alpha(s)\mathrm{d}s$, we find

$$2\mathcal{Q}(L, \alpha) = \left(\mathbb{E}[h^2] + \xi'(1)\right)L - \int_0^1 \xi''(q)a(q)\mathrm{d}q + \int_0^1 \frac{\mathrm{d}q}{L - a(q)}$$

$$= \int_0^{\underline{q}_{\mathsf{sph}}} \left(\xi''(\underline{q}_{\mathsf{sph}})L - \xi''(q)a(q)\right)\mathrm{d}q + \int_0^{\underline{q}_{\mathsf{sph}}} \frac{\mathrm{d}q}{L - a(q)}$$

$$+ \int_{\underline{q}_{\mathsf{sph}}}^{1} \left(\xi''(q)\left(L - a(q)\right) + \frac{1}{L - a(q)}\right)\mathrm{d}q.$$

Since ξ'' is increasing, AM-GM shows the second-to-last line is at most

$$\int_0^{\underline{q}_{\mathsf{sph}}} \xi''(\underline{q})\left(L - a(q)\right)\mathrm{d}q + \int_0^{\underline{q}_{\mathsf{sph}}} \frac{\mathrm{d}q}{L - a(q)} \geq 2\int_0^{\underline{q}_{\mathsf{sph}}} \sqrt{\xi''(\underline{q}_{\mathsf{sph}})}\mathrm{d}q = 2\underline{q}\sqrt{\xi''(\underline{q}_{\mathsf{sph}})}, \quad (5.8)$$

and similarly

$$\int_{\underline{q}_{\mathsf{sph}}}^{1} \left(\xi''(q)\left(L - a(q)\right) + \frac{1}{L - a(q)}\right)\mathrm{d}q \geq 2\int_{\underline{q}_{\mathsf{sph}}}^{1} \sqrt{\xi''(q)}\mathrm{d}q. \quad (5.9)$$

Combining, we conclude the lower bound on $GS_{\mathsf{sph}}(\xi, \mathcal{L}_h)$. Moreover for equality to hold in (5.8) and (5.9) we must have

$$\xi''(q)^{-1/2} = \begin{cases} L - a(q), & \forall q \in [0, \underline{q}_{\mathsf{sph}}], \\ L - a(q), & \forall q \in [\underline{q}_{\mathsf{sph}}, 1). \end{cases}$$

The first equality forces $\alpha(s) = 0$ on $[0, \underline{q}_{\mathsf{sph}})$ and $L = \xi''(q_0)^{-1/2}$, while the second equality implies $\alpha(q) = -\frac{\mathrm{d}}{\mathrm{d}s}\xi''(q)^{-1/2}$ for all $q \in [\underline{q}_{\mathsf{sph}}, 1]$. Taken together this means that equality in the GS_{sph} lower bound implies no overlap gap, completing the proof. $\qquad\square$

6 Impossibility of approximate maximization under an overlap gap

Here we explain the modifications of [GJ21, GJW20] needed to establish Proposition 1.12. Throughout this section we assume that $\xi(t) = \sum_{p \in \{2,4,\ldots,2P\}} c_p^2 t^p$ is an even polynomial and that the external field $(h, h, \ldots, h)$ is constant and deterministic. We take $\underline{q} = \inf(\text{supp}(\gamma_*^{\mathscr{U}}))$ and let $H_{N,0}, H_{N,1}$ be i.i.d. copies of H_N and for $t \in [0, 1]$ let $H_{N,t} \equiv \sqrt{t}H_{N,1} + \sqrt{1-t}H_{N,0}$. Moreover we define $\mathcal{H}_N \equiv \{H_{N,t} : t \in [0, 1]\}$.

Definition 6.1. *The model (ξ, h) satisfies the path overlap gap property with parameters $\varepsilon > 0$ and $0 < \nu_1 < \nu_2 < 1$ if the following holds with probability at least $1 - \frac{e^{-KN}}{K}$ for some $K = K(\xi, h) > 0$. For every pair $H_s, H_t \in \mathcal{H}_N$ and every $\boldsymbol{\sigma}_s, \boldsymbol{\sigma}_t$ satisfying*

$$\frac{H_r(\boldsymbol{\sigma}_r)}{N} \geq GS_{\xi,h} - \varepsilon, \qquad r \in \{s, t\}$$

it also holds that

$$|\langle \boldsymbol{\sigma}_s, \boldsymbol{\sigma}_t \rangle_N| \in [0, \nu_1] \cup [\nu_2, 1].$$

Definition 6.2. *The pair (H, H') of Hamiltonians are ν-separated above μ if for any $\boldsymbol{x}, \boldsymbol{y} \in \Sigma_N$ with $H(\boldsymbol{x}) \geq \mu, H'(\boldsymbol{y}) \geq \mu$, it holds that $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle_N| \leq \nu$.*

Lemma 6.3. *For any $t \in (0, 1)$, there exists $q_t = q_t(t, \xi, h) \in [0, \underline{q}]$ such that the following holds. For any $\varepsilon > 0$ there is $\delta(\varepsilon) > 0$ and $K(t, \xi, h, \varepsilon)$ such that with probability at least $1 - \frac{e^{-KN}}{K}$, every pair $\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_t \in \Sigma_N$ satisfying*

$$\frac{H_{N,s}(\boldsymbol{\sigma}_s)}{N} \geq GS_{\xi,h} - \delta, \quad s \in \{0, t\}$$

also satisfies $|\langle \boldsymbol{\sigma}_0, \boldsymbol{\sigma}_t \rangle_N| \in [q_t - \varepsilon, q_t + \varepsilon]$.

Proof. This follows from [CHL18, Proof of Theorem 2], in particular inequality (48) therein implies that for some $\delta(\varepsilon) > 0$:

$$\frac{1}{N} \max_{\substack{\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_t \in \Sigma_N \\ \langle \boldsymbol{\sigma}_0, \boldsymbol{\sigma}_t \rangle_N \notin [q_t - \varepsilon, q_t + \varepsilon]}} H_{N,0}(\boldsymbol{\sigma}_0) + H_{N,t}(\boldsymbol{\sigma}_t) \leq \left(\frac{1}{N} \max_{\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_t \in \Sigma_N} H_{N,0}(\boldsymbol{\sigma}_0) + H_{N,t}(\boldsymbol{\sigma}_t) \right) - \delta. \ \square$$

Lemma 6.4. *If (ξ, h) is not optimizable, then there exist $\varepsilon(\xi, h) > 0$ and $\underline{q} < a < b < 1$ and $K(\xi, h)$ such that with probability $1 - \frac{e^{-KN}}{K}$ the following holds. Every pair $\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1 \in \boldsymbol{\Sigma}_N$ satisfying*

$$\frac{H_N(\boldsymbol{\sigma}_i)}{N} \geq GS_{\xi,h} - \varepsilon, \quad i \in \{0, 1\}$$

also satisfies $|\langle \boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1 \rangle_N| \notin [a, b]$.

Proof. The proof is identical to that of [CGP+19, Theorem 3] using the 2-dimensional Guerra-Talagrand bound. Indeed [CGP+19, Lemma 5.4] exactly establishes that even pure p-spin models are not optimizable, i.e.

$$\mathbb{E}\left[\left(\partial_x \Phi_{\gamma_*}(t, X_t)\right)^2\right] < t \tag{6.1}$$

holds for some $t \in [0, 1)$ where $\gamma_* = \gamma_*^{\mathscr{U}}$. The remainder of the proof (just below [CGP+19, Lemma 5.4]) is fully general and we give an outline below. The point is that by (6.1) must hold in some non-empty open subset of $(0, 1)$, thus in a non-empty interval $[a, b]$. For each $t \in [a, b]$, one considers the Hamiltonian $\frac{H_N(\boldsymbol{\sigma}) + H_N(\boldsymbol{\sigma}')}{2}$ on two-replica configurations $(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ with overlap constraint $\langle \boldsymbol{\sigma}, \boldsymbol{\sigma}' \rangle_N = t$. The free energy of this constrained system can be upper-bounded using an interpolation argument; the relevant Parisi order parameter $\widetilde{\gamma}$ must increase, except that it may decrease by a factor of at most 2 at t, i.e. it only needs to satisfy

$$\lim_{s \uparrow t} \widetilde{\gamma}(s) \leq 2 \lim_{s \downarrow t} \widetilde{\gamma}(s). \tag{6.2}$$

Taking $\widetilde{\gamma} = \gamma_*^{\mathscr{U}}$ recovers the single-replica value. However when (6.1) holds, $\widetilde{\gamma} = \gamma_*^{\mathscr{U}}$ is no longer locally optimal since $\widetilde{\gamma}$ lives in a larger function space due to the relaxation (6.2). Hence the constrained two-replica ground state energy is strictly smaller. This argument can be applied for all $O(N)$ values $t \in [a, b] \cap \mathbb{Z}/N$, yielding the result. $\square$

Lemma 6.5. *If (ξ, h) is not optimizable, then there exists $\varepsilon(\xi, h) > 0$ and $0 \leq \nu_1 < \nu_2 \leq 1$ and $K > 0$ such that with probability at least $1 - \frac{e^{-KN}}{N}$:*

1. *The model (ξ, h) satisfies the path overlap gap property with parameters $(\varepsilon, \nu_1, \nu_2)$.*
2. *$H_{N,0}, H_{N,1}$ are ν_1 separated above $GS_{\xi,h} - \varepsilon$.*

Proof. The proof is identical to [GJ21, Theorem 3.4]. In short, one discretizes $\mathcal{H}_N$ into $\{H_{N,k\delta} : 0 \leq k \leq \delta^{-1}\}$ for some small $\delta > 0$ using Proposition 2.1 and then applies Lemma 6.3 to control the values $H_{N,s}(\boldsymbol{\sigma}_s), H_{N,t}(\boldsymbol{\sigma}_t)$ for $s \neq t$ and Lemma 6.4 to control the cases that $s = t$. Indeed in the proof of [GJ21, Theorem 3.4], the former is accomplished using [CGP$^+$19, Theorem 3] while the latter is accomplished using [CHL18, Theorem 2]. The preceding lemmas exactly generalize the relevant statements to non-optimizable models. $\qquad\square$

Proof of Proposition 1.12. Given Lemma 6.5, the proof is identical to that of [GJ21, Theorem 3.3]. Indeed, that proof does not depend on ξ. The main input is [GJ21, Conjecture 3.2]. This is shown to be implied by the combination of [GJ21, Theorem 3.4] and [GJ21, Conjecture 3.6]. Lemma 6.5 above suitably extends the former, while the latter (for general (ξ, h)) is the main result of our subsequent work [Sel21]. The proof of [GJ21, Theorem 3.3] also uses [GJ21, Theorem 4.2 and 6.1]; these follow from general concentration of measure results on Gaussian space and easily extend to general ξ. $\square$

We remark that Lemma 6.5 is also the only property of pure even p-spin models used in [GJW20, Theorem 2.4] to rule out approximate maximization (in a slightly weaker sense) by constant degree polynomials. Therefore their result also applies under the more general conditions of Proposition 1.12.

7 Proof of Lemmas 1.6, 1.8 and 2.9

We first recall several existing results.

Corollary 7.1 ([AMS21, Corollary 6.6]). *For any $\gamma \in \mathscr{L}$ and any $t \in [0, 1)$,*

$$\mathbb{E}\left[\partial_x \Phi_\gamma(t, X_t)^2\right] = \int_0^t \xi''(s)\, \mathbb{E}\left[\left(\partial_{xx}\Phi_\gamma(s, X_s)\right)^2\right] \mathrm{d}s.$$

Lemma 7.2 ([AMS21, Corollary 6.6 and Lemma 6.7]). *For any $\gamma \in \mathscr{L}$, the values*

$$\mathbb{E}\left[\partial_x \Phi_\gamma(t, X_t)^2\right], \qquad \mathbb{E}\left[\partial_{xx}\Phi_\gamma(t, X_t)^2\right]$$

are continuous functions of $t \in [0, 1)$.

Proposition 7.3 ([AMS21, Proposition 6.8]). *Let $\gamma \in \mathscr{L}$, and $\delta : [0, 1) \to \mathbb{R}$ be such that $\|\xi''\delta\|_{TV[0,t]} < \infty$ for all $t \in [0, 1)$, $\|\xi''\delta\|_1 < \infty$, and $\delta(t) = 0$ for $t \in (1 - \varepsilon, 1]$, $\varepsilon > 0$. Further assume that $\gamma + s\delta \geq 0$ for all $s \in [0, s_0)$ for some positive s_0. Then*

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}s}(\gamma + s\delta)\Big|_{s=0+} = \frac{1}{2} \int_0^1 \xi''(t)\delta(t)\left(\mathbb{E}\left[\partial_x \Phi_\gamma(t, X_t)^2\right] - t\right) \mathrm{d}t. \tag{7.1}$$

Lemma 7.4 ([AMS21, Lemma 6.9]). *The support of $\gamma \in \mathscr{L}_q$ is a disjoint union of countably many intervals $S(\gamma) = \bigcup_{\alpha \in A} I_\alpha$, where $I_\alpha = (a_\alpha, b_\alpha)$ or $I_\alpha = [a_\alpha, b_\alpha)$, $\underline{q} \leq a_\alpha < b_\alpha \leq 1$, and A is countable.*

Lemma 7.5. *The function $\mathsf{P} = \mathsf{P}_{\xi, \mathcal{L}_h}$ is strictly convex on $\mathscr{L}$.*

Proof. The proof is identical to [AC15, Theorem 2] and [CHL18, Lemma 5] which show strict convexity on $\mathscr{U}$. $\qquad\square$

Throughout this section we let $\gamma_*^{\mathscr{L}_{\underline{q}}}$ be the minimizer of P over $\mathscr{L}_{\underline{q}}$, assuming it exists. Note that we will eventually show in Lemma 1.6 that $\gamma_*^{\mathscr{L}_{\underline{q}}} = \gamma_*^{\mathscr{L}}$ if either minimizer exists.

Lemma 7.6. *Assume $\gamma_*^{\mathscr{L}_{\underline{q}}}$ exists. Then*

$$t \in \operatorname{supp}\big(\gamma_*^{\mathscr{L}_{\underline{q}}}\big) \quad \Rightarrow \quad \mathbb{E}\big[\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2\big] = t, \tag{7.2}$$

$$t \geq \underline{q} \quad \Rightarrow \quad \mathbb{E}\big[\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2\big] \geq t. \tag{7.3}$$

Proof. We first show Equation (7.2). For $\underline{q} \leq t_1 < t_2 < 1$ we take $\delta(t) = \gamma_*^{\mathscr{L}_{\underline{q}}}(t) 1_{t \in [t_1, t_2)}$. Clearly $\gamma_*^{\mathscr{L}_{\underline{q}}} + s\delta \in \mathscr{L}_{\underline{q}}$. Since $\gamma_*^{\mathscr{L}_{\underline{q}}}$ minimizes P($\cdot$) over $\mathscr{L}_{\underline{q}}$,

$$0 \leq \frac{\mathrm{dP}}{\mathrm{d}s}\big(\gamma_*^{\mathscr{L}_{\underline{q}}} + s\delta\big)\Big|_{s=0} = \frac{1}{2}\int_{t_1}^{t_2} \xi''(t)\gamma_*^{\mathscr{L}_{\underline{q}}}(t)\big(\mathbb{E}\big[\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2\big] - t\big)\,\mathrm{d}t.$$

Since t_1, t_2 are arbitrary and $\xi''(t) > 0$ for $t \in (0, 1)$, this implies

$$\gamma_*^{\mathscr{L}_{\underline{q}}}(t)\big(\mathbb{E}\big[\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2\big] - t\big) = 0$$

for almost every $t \in [\underline{q}, 1)$. Since $\gamma_*^{\mathscr{L}_{\underline{q}}}(t)$ is right-continuous and $\mathbb{E}[\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2]$ is continuous by Lemma 7.2, it follows that $\gamma_*^{\mathscr{L}_{\underline{q}}}(t)(\mathbb{E}[\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2] - t) = 0$ for every $t \in [\underline{q}, 1)$. This in turns implies $\mathbb{E}[\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2] = t$ for every $t \in S(\gamma_*^{\mathscr{L}_{\underline{q}}})$ by right-continuity of $\gamma_*^{\mathscr{L}_{\underline{q}}}$. This can be extended to all $t \in \operatorname{supp}(\gamma_*^{\mathscr{L}_{\underline{q}}})$ by again using continuity of $t \mapsto \mathbb{E}[\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2]$.

Next consider Eq. (7.3), where it suffices now to consider $t \in [\underline{q}, 1) \setminus \operatorname{supp}(\gamma_*^{\mathscr{L}_{\underline{q}}})$. By Lemma 7.4, $[\underline{q}, 1) \setminus \operatorname{supp}(\gamma_*^{\mathscr{L}_{\underline{q}}})$ is a disjoint union of open intervals. Let J be such an interval, and consider any $[t_1, t_2] \subseteq J$. Set $\delta(t) = \mathbb{I}(t \in (t_1, t_2])$, and notice that $\gamma_*^{\mathscr{L}_{\underline{q}}} + s\delta \in \mathscr{L}_{\underline{q}}$ for $s \geq 0$. By Proposition 7.3, we have

$$0 \leq \frac{\mathrm{dP}}{\mathrm{d}s}(\gamma + s\delta)\Big|_{s=0} = \frac{1}{2}\int_{t_1}^{t_2} \xi''(t)\big(\mathbb{E}\big[\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2\big] - t\big)\,\mathrm{d}t.$$

Since t_1, t_2 are arbitrary, $\xi''(t) > 0$ for $t \in (0, 1)$ and $t \mapsto \mathbb{E}[\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2]$ is continuous, this implies $\mathbb{E}[\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2] \geq t$ for all $t \in J$, and hence all $t \in [\underline{q}, 1) \setminus \operatorname{supp}(\gamma_*^{\mathscr{L}_{\underline{q}}})$. $\quad\square$

Corollary 7.7. *Assume $\gamma_*^{\mathscr{L}_{\underline{q}}}$ exists. Then*

$$t \in \operatorname{supp}\big(\gamma_*^{\mathscr{L}_{\underline{q}}}\big) \quad \Rightarrow \quad \xi''(t)\mathbb{E}\big[\partial_{xx} \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2\big] = 1.$$

Proof. By Lemma 7.4, $\operatorname{supp}(\gamma_*^{\mathscr{L}_{\underline{q}}})$ is a disjoint union of closed intervals with non-empty interior. Let K be one such interval. Then, for any $[t_1, t_2] \in K$, Corollary 7.1 and Lemma 7.6 imply

$$t_2 - t_1 = \mathbb{E}\big[\partial_x \Phi(t_2, X_{t_2})^2\big] - \mathbb{E}\big[\partial_x \Phi(t_1, X_{t_1})^2\big] = \int_{t_1}^{t_2} \xi''(t)\mathbb{E}\big[\partial_{xx}\Phi(t, X_t)^2\big]\mathrm{d}t.$$

Since t_1, t_2 are arbitrary, $\xi''(t)\mathbb{E}[\partial_{xx}\Phi(t, X_t)^2] = 1$ for almost every $t \in K$. By Lemma 7.2 it follows that $\xi''(t)\mathbb{E}[\partial_{xx}\Phi(t, X_t)^2] = 1$ for all $t \in \operatorname{supp}(\gamma_*^{\mathscr{L}_{\underline{q}}})$. $\quad\square$

Lemma 7.8 ([AMS21, Lemma 6.12]). *Let $\gamma \in \mathscr{L}$ satisfy $\gamma(t) = 0$ for all $t \in (t_1, 1)$, where $t_1 < 1$. Then, for any $t_* \in (t_1, 1)$, the probability distribution of X_{t_*} has a density p_{t_*} with respect to the Lebesgue measure. Further, for any $t_* \in (t_1, 1)$ and any $M \in \mathbb{R}_{\geq 0}$, there exists $\varepsilon(t_*, M, \gamma) > 0$ such that*

$$\inf_{|x| \leq M, t \in [t_*, 1]} p_t(x) \geq \varepsilon(t_*, M, \gamma).$$

Lemma 7.9. *If a minimizer $\gamma_*^{\mathscr{L}_q}$ exists, then $\mathrm{supp}(\gamma_*^{\mathscr{L}_q}) = [\underline{q}, 1)$.*

Proof. By Lemma 7.4, $[\underline{q}, 1) \setminus \mathrm{supp}(\gamma_*^{\mathscr{L}_q})$ is a countable union of disjoint intervals, open in $[\underline{q}, 1)$. First assume that at least one of these intervals is of the form (t_1, t_2) with $\underline{q} \leq t_1 < t_2 < 1$. By Lemma 7.6 and Corollary 7.7 we know that

$$\mathbb{E}\big[\partial_x \Phi_{\gamma_*^{\mathscr{L}_q}}(t_i, X_{t_i})^2\big] = t_i, \quad \xi''(t_i)\mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*^{\mathscr{L}_q}}(t_i, X_{t_i})^2\big] = 1, \quad i \in \{1, 2\}, \tag{7.4}$$

$$\mathbb{E}\big[\partial_x \Phi_{\gamma_*^{\mathscr{L}_q}}(t, X_t)^2\big] \geq t \quad \forall t \in (t_1, t_2). \tag{7.5}$$

Further, for $t \in (t_1, t_2)$, $\Phi_{\gamma_*^{\mathscr{L}_q}}$ solves the PDE

$$\partial_t \Phi_{\gamma_*^{\mathscr{L}_q}}(t, x) + \frac{\xi''(t)}{2}\partial_{xx}\Phi_{\gamma_*^{\mathscr{L}_q}}(t, x) = 0$$

which is simply the heat equation up to a time change. We therefore obtain

$$\Phi_{\gamma_*^{\mathscr{L}_q}}(t, x) = \mathbb{E}^{Z \sim \mathsf{N}(0,1)}\big[\Phi_{\gamma_*^{\mathscr{L}_q}}\big(t_2, x + \sqrt{\xi'(t_2) - \xi'(t)}\, Z\big)\big], \quad \forall t \in (t_1, t_2].$$

Differentiating this equation and using dominated convergence (recall that $\partial_{xx}\Phi_{\gamma_*^{\mathscr{L}_q}}(t_2, x)$ is bounded by Proposition 2.6), we obtain

$$\partial_{xx}\Phi_{\gamma_*^{\mathscr{L}_q}}(t, x) = \mathbb{E}^{Z \sim \mathsf{N}(0,1)}\big[\partial_{xx}\Phi_{\gamma_*^{\mathscr{L}_q}}\big(t_2, x + \sqrt{\xi'(t_2) - \xi'(t)}\, Z\big)\big].$$

Because $\mathrm{d}X_t = \sqrt{\xi''(t)}\, \mathrm{d}B_t$, we can rewrite the last equation as

$$\partial_{xx}\Phi_{\gamma_*^{\mathscr{L}_q}}(t, X_t) = \mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*^{\mathscr{L}_q}}(t_2, X_{t_2})|X_t\big].$$

By Jensen's inequality,

$$\mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*^{\mathscr{L}_q}}(t, X_t)^2\big] \leq \mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*^{\mathscr{L}_q}}(t_2, X_{t_2})^2\big] = \frac{1}{\xi''(t_2)}, \tag{7.6}$$

where in the last step we used Eq. (7.4). Using Corollary 7.1 we get, for $t \in [t_1, t_2]$

$$\mathbb{E}\big[\partial_x \Phi_{\gamma_*^{\mathscr{L}_q}}(t, X_t)^2\big] = \mathbb{E}\big[\partial_x \Phi_{\gamma_*^{\mathscr{L}_q}}(t_1, X_{t_1})^2\big] + \int_{t_1}^t \xi''(s)\mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*^{\mathscr{L}_q}}(s, X_s)^2\big]\, \mathrm{d}s$$

$$\leq t_1 + \int_{t_1}^t \frac{\xi''(s)}{\xi''(t_2)}\, \mathrm{d}s < t,$$

where in the last step we used the fact that $t \mapsto \xi''(t)$ is increasing. The last equation is in contradiction with Eq. (7.5), and therefore $[\underline{q}, 1) \setminus \mathrm{supp}(\gamma_*^{\mathscr{L}_q})$ is either empty or consists of a single interval $(t_1, 1)$.

In order to complete the proof, we need to rule out the case $[\underline{q}, 1) \setminus \mathrm{supp}(\gamma_*^{\mathscr{L}_q}) = (t_1, 1)$. Assume for sake of contradiction that indeed $[\underline{q}, 1) \setminus \mathrm{supp}(\gamma_*^{\mathscr{L}_q}) = (t_1, 1)$. For $t \in (t_1, 1)$, let

$r = r(t) = \xi'(1) - \xi'(t)$, and notice that $r(t)$ is decreasing with $r(t) = \xi''(1)(1-t) + O((1-t)^2)$ as $t \to 1$. By solving the Parisi PDE in the interval $(t_1, 1)$, we find that for all $t \in (t_1, 1)$,

$$\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, x) = \mathbb{E}^{Z \sim \mathsf{N}(0,1)}\left[\text{sign}\left(Z + \frac{x}{\sqrt{r(t)}}\right)\right]$$

and therefore

$$1 - \mathbb{E}\big[\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2\big] = \mathbb{E}\left[Q\left(\frac{X_t}{\sqrt{r(t)}}\right)\right],$$

$$Q(x) \equiv 1 - \mathbb{E}^{Z \sim \mathsf{N}(0,1)}\big[\text{sign}(x+Z)\big]^2.$$

Note that $0 \leq Q(x) \leq 1$ is continuous, with $Q(0) = 1$. Hence, there exists a numerical constant $\delta_0 \in (0, 1)$ such that $Q(x) \geq 1/2$ for $|x| \leq \delta_0$. Therefore, fixing $t_* \in (t_1, 1)$, for any $t \in (t_*, 1)$

$$1 - \mathbb{E}\big[\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2\big] \geq \frac{1}{2}\mathbb{P}\big[|X_t| \leq \delta_0\sqrt{r(t)}\big]$$

$$\overset{(a)}{\geq} \delta_0 \varepsilon(t_*, 1, \gamma)\sqrt{r(t)} \overset{(b)}{\geq} C\sqrt{1-t},$$

where (a) follows by Lemma 7.8 and (b) holds for some $C = C(\gamma) > 0$. We therefore obtain $\mathbb{E}[\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2] \leq 1 - C\sqrt{1-t}$, which contradicts Lemma 7.6 for t close enough to 1. $\qquad\square$

In the next lemma, we show that minimization of P over $\mathscr{L}$ subsumes minimization over $\mathscr{L}_{\underline{q}}$. A priori, one might expect that tuning the value of $\underline{q}$ could lead to many different minima.

Lemma 7.10. *For $\gamma, \widehat{\gamma} \in \mathscr{L}$, define the function $\gamma^{(\varepsilon)} \in \mathscr{L}$ by:*

$$\gamma^{(\varepsilon)}(t) = \begin{cases} \gamma(t), & 0 \leq t < 1 - \varepsilon, \\ \widehat{\gamma}(t), & 1 - \varepsilon \leq t < 1. \end{cases}$$

Then $\lim_{\varepsilon \to 0} \mathsf{P}(\gamma^{(\varepsilon)}) = \mathsf{P}(\gamma)$.

Proof. Using [AMS21, Proposition 6.1(c)] and continuity, we have $\|\mathsf{P}(\gamma) - \mathsf{P}(\gamma^{(\varepsilon)})\| \leq C\|\gamma - \gamma^{(\varepsilon)}\|$ for a constant C depending only on ξ. The right-hand side tends to zero as $\varepsilon \to 0$ by the definition of $\mathscr{L}$. $\qquad\square$

Lemma 7.11. *Suppose $\gamma_*^{\mathscr{L}_{\underline{q}}}$ exists. Then $\gamma_*^{\mathscr{L}} = \gamma_*^{\mathscr{L}_{\underline{q}}}$.*

Proof. Let $f(t) = \mathbb{E}[\partial_x \Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2]$. First we show that $f(t) \geq t$ for all $0 \leq t \leq \underline{q}$. Recall that X_t is simply a time-changed Brownian motion on $0 \leq t \leq \underline{q}$ while $\Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}$ solves the time-changed heat equation on the same time interval, therefore $\partial_{xx}\Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t) = \mathbb{E}^t[\partial_{xx}\Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(\underline{q}, X_{\underline{q}})]$. By Jensen's inequality, it follows that for all $0 \leq t \leq \underline{q}$,

$$\mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(t, X_t)^2\big] \leq \mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*^{\mathscr{L}_{\underline{q}}}}(\underline{q}, X_{\underline{q}})^2\big]$$

$$= \frac{1}{\xi''(\underline{q})}$$

$$\leq \frac{1}{\xi''(t)}.$$

In the last line we used that ξ'' is increasing as ξ is a power series with non-negative coefficients. Next, from Lemma 7.6 and Lemma 7.9 it follows that $f(\underline{q}) = \underline{q}$. In light of

Corollary 7.1, we showed just above that $f'(t) \leq 1$ for $t \leq \underline{q}$. It now follows that $f(t) \geq t$ for all $0 \leq t \leq \underline{q}$.

Next, Proposition 7.3 combined with Corollary 7.1 and Lemma 7.6 imply that $\frac{d}{ds}\mathsf{P}((1-s)\gamma_*^{\mathscr{L}_{\underline{q}}} + s\gamma)|_{s=0+} \geq 0$ for any $\gamma \in \mathscr{L}$ agreeing with $\gamma_*^{\mathscr{L}_{\underline{q}}}$ on $[1-\varepsilon, 1)$ for some $\varepsilon > 0$. Using convexity of P as guaranteed by Lemma 7.5, this implies that

$$\mathsf{P}\big(\gamma_*^{\mathscr{L}_{\underline{q}}}\big) \leq \mathsf{P}(\gamma)$$

for any such γ. Assuming for sake of contradiction that some $\gamma_* \in \mathscr{L}$ satisfies $\mathsf{P}(\gamma_*) < \mathsf{P}(\gamma_*^{\mathscr{L}_{\underline{q}}})$, define $\gamma^{(\varepsilon)} \in \mathscr{L}$ by

$$\gamma^{(\varepsilon)}(t) = \begin{cases} \gamma_*(t), & 0 \leq t < 1 - \varepsilon, \\ \gamma_*^{\mathscr{L}_{\underline{q}}}, & 1 - \varepsilon \leq t < 1. \end{cases}$$

Then Lemma 7.10 implies that $\lim_{\varepsilon \to 0} \mathsf{P}(\gamma^{(\varepsilon)}) = \mathsf{P}(\gamma_*) < \mathsf{P}(\gamma_*^{\mathscr{L}_{\underline{q}}})$. This contradicts the above for small enough ε, completing the proof. □

We now restate and prove Lemmas 1.6 and 1.8.

Lemma 7.12. *For $\gamma_* \in \mathscr{L}$ and $\underline{q} = \inf(\mathrm{supp}(\gamma_*))$, the following are equivalent:*

1. *γ_* is optimizable.*
2. *$\mathsf{P}(\gamma_*) = \inf_{\gamma \in \mathscr{L}} \mathsf{P}(\gamma)$.*
3. *$\mathsf{P}(\gamma_*) = \inf_{\gamma \in \mathscr{L}_{\underline{q}}} \mathsf{P}(\gamma)$.*

Moreover if a minimizer exists in either variational problem just above, then it is unique.

Proof. Lemma 7.5 immediately implies uniqueness of minimizers. The second statement immediately implies the third, while Lemma 7.11 provides the converse result. To show that the first statement implies the third, we observe that Proposition 7.3 immediately yields

$$\frac{d}{ds}\mathsf{P}\big((1-s)\gamma_* + s\gamma\big)\bigg|_{s=0+} = 0$$

for any $\gamma \in \mathscr{L}_{\underline{q}}$ when γ_* is optimizable; this implies the third statement by again invoking Lemma 7.5. It only remains to show that if $\mathsf{P}(\gamma_*) = \inf_{\gamma \in \mathscr{L}} \mathsf{P}(\gamma)$, then γ_* is $\underline{q}$-optimizable, which follows from Lemmas 7.6 and 7.9. □

Lemma 7.13. *If $\gamma_*^{\mathscr{U}}$ strictly increases on $[\underline{q}, 1)$ for $\underline{q} = \inf(\mathrm{supp}(\gamma_*^{\mathscr{U}}))$, then no overlap gap holds, i.e. $\gamma_*^{\mathscr{U}}$ is optimizable.*

Proof. Fix $\underline{q} < t_1 < t_2 < 1$ and define $\delta(t) = [\gamma_*^{\mathscr{U}}(t_1) - \gamma_*^{\mathscr{U}}(t)]\mathbb{I}_{(t_1, t_2)}(t)$. It is easy to see that this satisfies the assumptions of Proposition 7.3 with $s_0 = 1$. Letting $\gamma^s = \gamma_*^{\mathscr{U}} + s\delta$,

$$\frac{d\mathsf{P}}{ds}\big(\gamma^s\big)\bigg|_{s=0+} = -\frac{1}{2}\int_{t_1}^{t_2} \xi''(t)\big(\gamma_*^{\mathscr{U}}(t) - \gamma_*^{\mathscr{U}}(t_1)\big)\big(\mathbb{E}\big[\partial_x \Phi_{\gamma_*^{\mathscr{U}}}(t, X_t)^2\big] - t\big)\,dt.$$

On the other hand, $\gamma^s \in \mathscr{U}$ for $s \in [0,1]$ (since $\gamma_*^{\mathscr{U}}$ is strictly increasing on $[\underline{q}, 1)$), so

$$\int_{t_1}^{t_2} \xi''(t)\big(\gamma_*^{\mathscr{U}}(t) - \gamma_*^{\mathscr{U}}(t_1)\big)\big(\mathbb{E}\big[\partial_x \Phi_{\gamma_*^{\mathscr{U}}}(t, X_t)^2\big] - t\big)\,dt \leq 0,$$

for all $t_1 < t_2$. Since $\gamma_*^{\mathscr{U}}(t) - \gamma_*^{\mathscr{U}}(t_1) > 0$ strictly for all $t > t_1$, this implies

$$\mathbb{E}\big[\partial_x \Phi_{\gamma_*^{\mathscr{U}}}(t, X_t)^2\big] \leq t$$

for almost every t, and therefore for every t. The inequality

$$\mathbb{E}\big[\partial_x \Phi_{\gamma_*^{\mathcal{U}}}(t, X_t)^2\big] \geq t$$

is proved in the same way using $\delta(t) = [\gamma_*^{\mathcal{U}}(t_2) - \gamma_*^{\mathcal{U}}(t)]\mathbb{I}_{(t_1, t_2)}(t)$. $\square$

Finally we turn to Lemma 2.9.

Proof of Lemma 2.9. First, (2.7) is clear given Corollary 7.1. To establish (2.8), we first show that

$$\mathrm{d}\big(\partial_{xx}\Phi_{\gamma_*}(t, X_t)\big) = -\xi''(t)\gamma_*(t)\big(\partial_{xx}\Phi_{\gamma_*}(t, X_t)\big)^2 \mathrm{d}t + \partial_{xxx}\Phi_{\gamma_*}(t, X_t)\sqrt{\xi''(t)}\mathrm{d}B_t. \quad (7.7)$$

Indeed (7.7) follows by using Ito's formula to derive

$$\begin{aligned}
&\mathrm{d}\big(\partial_{xx}\Phi_{\gamma_*}(t, X_t)\big) \\
&= \left(\partial_{txx}\Phi_{\gamma_*}(t, X_t) + \partial_x \Phi_{\gamma_*}(t, X_t)\partial_{xxx}\Phi_{\gamma_*}(t, X_t)\xi''(t)\gamma_*(t) + \frac{\xi''(t)\partial_{xxxx}\Phi_{\gamma_*}(t, X_t)}{2}\right)\mathrm{d}t \\
&\quad + \partial_{xxx}\Phi_{\gamma_*}(t, X_t)\sqrt{\xi''(t)}\mathrm{d}B_t
\end{aligned}$$

and taking the second derivative with respect to x of the Parisi PDE to obtain

$$\begin{aligned}
0 &= \partial_{xx}\left(\partial_t \Phi_{\gamma_*}(t, x) + \frac{\xi''(t)}{2}\big(\partial_{xx}\Phi_{\gamma_*}(t, x) + \gamma_*(t)\big(\partial_x \Phi_{\gamma_*}(t, x)\big)^2\big)\right) \\
&= \partial_{txx}\Phi_{\gamma_*}(t, x) + \frac{\xi''(t)\partial_{xxxx}\Phi_{\gamma_*}(t, x)}{2} + \xi''(t)\gamma_*(t)\big(\big(\partial_{xx}\Phi_{\gamma_*}(t, x)\big)^2 \\
&\quad + \partial_x \Phi_{\gamma_*}(t, x)\partial_{xxx}\Phi_{\gamma_*}(t, x)\big).
\end{aligned}$$

In particular (7.7) implies that for all $t \in [0, 1)$,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big[\partial_{xx}\Phi_{\gamma_*}(t, x)\big] &= -\xi''(t)\gamma_*(t)\mathbb{E}\big[\big(\partial_{xx}\Phi_{\gamma_*}(t, x)\big)^2\big] \\
&= -\gamma_*(t).
\end{aligned}$$

Therefore to show (2.8) it suffices to show $\lim_{t \to 1} \mathbb{E}[\partial_{xx}\Phi_{\gamma_*}(t, X_t)] \geq 0$, but this is clear by convexity of $\Phi_{\gamma_*}(t, \cdot)$. $\square$

8 Incremental AMP proofs

We will prove Lemma 8.3 which generalizes Lemma 3.6 to the setting of branching AMP and describes the limiting Gaussian processes $N_{\ell,a}^\delta, Z_{\ell,a}^\delta$. We recall the setup of Section 4 and in particular continue to use the value $q_B \in (\underline{q}, 1)$ to define the time $\ell_{q_B}^\delta$ at which $Z_{\ell_{q_B}^\delta, 1}^\delta = Z_{\ell_{q_B}^\delta, 2}^\delta$ last holds. For the branching setting we slightly generalize the filtration (3.13) to

$$\mathcal{F}_\ell^\delta = \sigma\big(\big(Z_{k,a}^\delta, N_{k,a}^\delta\big)_{0 \leq k \leq \ell, a \in \{1,2\}}\big).$$

Crucially note that we restrict here to $k \geq 0$, i.e. we do not include the preparatory iterates with negative index. We remark that if we consider all the IAMP iterates $(Z_{\ell,a}^\delta, N_{\ell,a}^\delta)$ together in the linear order given by $(\ell, a) \to 2\ell + a$, then these are iterates of a standard AMP algorithm since each iterate depends only on the previous ones. (This is because in (2.4) the last expectation is zero when f_ℓ does not depend on Z^j, which is the case if f_ℓ and Z^j correspond to different values of a after the two branches separate.) Moreover it is easy to see that the Onsager correction terms are not affected by this rewriting. Therefore we may continue to use state evolution in the natural way even though we do not think of the iterates as actually being totally ordered.

Lemma 8.1. *In branching IAMP, $\mathcal{F}_\ell^\delta$ is jointly independent of the iterates $(Z^{-j})_{J_\ell < j \le K}$ for*

$$J_\ell \equiv \max\big(\{\ell\} \cup \big\{k_{i,a} + \ell - \ell_{q_i}^\delta : \ell_{q_i}^\delta \le \ell, a \in \{1,2\}\big\}\big).$$

Proof. We proceed by induction over ℓ, the base case $\ell = 0$ following from Proposition 2.4. Because the random variables $Z_{k,a}^\ell$ form a Gaussian process it suffices to verify that

$$\mathbb{E}\big[Z_{\ell,a}^\delta Z^{-j}\big] = 0$$

holds whenever $j > J_\ell$. By state evolution,

$$\mathbb{E}\big[Z_{\ell,a}^\delta Z^{-j}\big] = \xi'\big(\big[N_{\ell-1,a}^\delta Z^{-j-1}\big]\big).$$

By definition $N_{\ell-1,a}^\delta$ is $\mathcal{F}_{\ell-1}^\delta$-measurable. Since $\xi'(0) = 0$ it suffices to show that $\mathcal{F}_{\ell-1}^\delta$ is independent of Z^{-j-1}. By the inductive hypothesis, this holds if $j + 1 > J_{\ell-1}$. This in turn follows from the easy-to-verify fact that $J_\ell - 1 \ge J_{\ell-1}$, completing the proof. $\qquad\square$

Corollary 8.2. *Let $G_{q_j,a}^\delta$ be the state evolution limit of $g^{(q_j,a)}$ for each $(j,a) \in [m] \times [2]$. Then the law of $(G_{q_i,1}^\delta, G_{q_i,2}^\delta)$ conditioned on $\mathcal{F}_{\ell_{q_i}^\delta - 1}^\delta$ is $\mathsf{N}(0, I_2)$.*

Proof. Since $k_{i,1} \ne k_{i,2}$ it follows from Proposition 2.4 that $(G_{q_i,1}^\delta, G_{q_i,2}^\delta) \sim \mathsf{N}(0, I_2)$ holds as an unconditional law. Since we chose the values $k_{i,a}$ such that $k_{i,a} - \ell_{q_i}^\delta > k_{j,a'} - \ell_{q_j}^\delta > 0$ for any $i > j$ and $a, a' \in \{1,2\}$, it follows that $k_{i,a} > J_{\ell_{q_i}^\delta - 1}$. Applying Lemma 8.1 now concludes the proof. $\qquad\square$

Lemma 8.3. *The sequences $(Z_{\underline{\ell},a}^\delta, Z_{\underline{\ell}+1,a}^\delta, \dots)$ and $(N_{\underline{\ell},a}^\delta, N_{\underline{\ell}+1,a}^\delta, \dots)$ satisfy for $\ell \ge \underline{\ell}$:*

$$\mathbb{E}\big[\big(Z_{\ell+1,a}^\delta - Z_{\ell,a}^\delta\big)Z_{j,a}^\delta\big] = 0, \quad \text{for all } \underline{\ell}+1 \le j \le \ell \tag{8.1}$$

$$\mathbb{E}\big[\big(Z_{\ell+1,a}^\delta - Z_{\ell,a}^\delta\big)^2 | \mathcal{F}_\ell^\delta\big] = \xi'\big(q_{\ell+1}^\delta\big) - \xi'\big(q_\ell^\delta\big) \tag{8.2}$$

$$\mathbb{E}\big[\big(Z_{\ell+1,1}^\delta - Z_{\ell,1}^\delta\big)\big(Z_{\ell+1,2}^\delta - Z_{\ell,2}^\delta\big) | \mathcal{F}_\ell^\delta\big] = \big(\xi'\big(q_{\ell+1}^\delta\big) - \xi'\big(q_\ell^\delta\big)\big) \cdot \mathbf{1}_{\ell < \ell_{q_B}^\delta} \tag{8.3}$$

$$\mathbb{E}\big[\big(Z_{\ell,a}^\delta\big)^2\big] = \xi'\big(q_\ell^\delta\big) \tag{8.4}$$

$$\mathbb{E}\big[\big(N_{\ell+1,a}^\delta - N_{\ell,a}^\delta\big) | \mathcal{F}_\ell^\delta\big] = 0 \tag{8.5}$$

$$\mathbb{E}\big[\big(N_{\ell+1,a}^\delta - N_{\ell,a}^\delta\big)^2 | \mathcal{F}_\ell^\delta\big] = \delta \tag{8.6}$$

$$\mathbb{E}\big[\big(N_{\ell+1,1}^\delta - N_{\ell,1}^\delta\big)\big(N_{\ell+1,2}^\delta - N_{\ell,2}^\delta\big) | \mathcal{F}_\ell^\delta\big] = \delta \cdot \mathbf{1}_{\ell < \ell_{q_B}^\delta} \tag{8.7}$$

$$\mathbb{E}\big[\big(N_{\ell,a}^\delta\big)^2\big] = q_{\ell+1}^\delta. \tag{8.8}$$

Proof. We recall that $(Z_{\ell,a}^\delta)_{\ell \ge \underline{\ell}, a \in \{1,2\}}$ is a Gaussian process, which means we can ignore the conditioning on $\mathcal{F}_\ell^\delta$ in proving Equation (8.2). First we check that Equations (8.4) and (8.8) hold for $\ell = \underline{\ell}$. For Equation (8.8), by definition of δ:

$$\mathbb{E}\big[\big(N_{\underline{\ell},a}^\delta\big)^2\big] = (1 + \varepsilon_0)^2 \mathbb{E}\big[\big(M^{\underline{\ell}}\big)^2\big] = (1 + \varepsilon_0)^2 \underline{q} = \underline{q} + \delta = q_1^\delta.$$

For Equation (8.4),

$$\mathbb{E}\big[\big(Z_{\underline{\ell},a}^\delta\big)^2\big] = \xi'\big(\mathbb{E}\big[\big(M^{\underline{\ell}-1}\big)^2\big]\big) = \xi'(\underline{q}).$$

Observe now that if Equations (8.1), (8.2), (8.5), (8.6) hold for $\underline{\ell} \le \ell \le k$ then so do Equations (8.4) and (8.8), as

$$\mathbb{E}\big[\big(N_{\ell+1,a}^\delta\big)^2\big] = \mathbb{E}\big[\big(N_{\ell+1,a}^\delta - N_{\ell,a}^\delta\big)^2\big] + 2 \cdot \mathbb{E}\big[\big(N_{\ell+1,a}^\delta - N_{\ell,a}^\delta\big)N_{\ell,a}^\delta\big] + \mathbb{E}\big[\big(N_{\ell,a}^\delta\big)^2\big]$$

and similarly for $\mathbb{E}[(Z_{\ell+1,a}^\delta)^2]$. Therefore to show the six identities (8.1), (8.2), (8.5), (8.6), (8.4) and (8.8) it suffices to check the base case $\ell = \underline{\ell}$ for Equations (8.1), (8.2), (8.5), (8.6) and to perform an inductive step to show these four identities for $\ell = k + 1$, assuming all six of these equations as inductive hypotheses for $\ell \le k$. We turn to doing this, and finally show Equations (8.3), 8.7 at the end.

Base case for Equations (8.1)**,** (8.2)**,** (8.5)**,** (8.6) Note that here, none of the perturbations $g^{(q_i,a)}$ appear yet. We begin with Equation (8.1):

$$
\begin{aligned}
\mathbb{E}\big[\big(Z^\delta_{\underline{\ell}+1,a} - Z^\delta_{\underline{\ell},a}\big)Z^\delta_{\underline{\ell},a}\big] &= \xi'\big(\mathbb{E}\big[N^\delta_{\underline{\ell},a}M^{\ell-1}\big]\big) - \xi'\big(\mathbb{E}\big[M^{\ell-1}M^{\ell-1}\big]\big) \\
&= \xi'\big((1+\varepsilon_0)\mathbb{E}\big[M^\ell M^{\ell-1}\big]\big) - \xi'(\underline{q}) \\
&= \xi'\big((1+\varepsilon_0)\phi(a_{\underline{\ell}-1})\big) - \xi'(\underline{q}) \\
&= \xi'(\underline{q}) - \xi'(\underline{q}) \\
&= 0.
\end{aligned}
$$

This means $\mathbb{E}[Z^\delta_{\underline{\ell}+1,a}|Z^\delta_{\underline{\ell},a}] = Z^\delta_{\underline{\ell},a}$. Hence

$$
\mathbb{E}\big[\big(Z^\delta_{\underline{\ell}+2,a} - Z^\delta_{\underline{\ell}+1,a}\big)Z^\delta_{\underline{\ell}+1,a}\big] = \xi'\big(\mathbb{E}\big[N^\delta_{\underline{\ell}+1}N^\delta_{\underline{\ell},a}\big]\big) - \xi'\big(\mathbb{E}\big[N^\delta_{\underline{\ell},a}N^\delta_{\underline{\ell},a}\big]\big).
$$

To see that the above expression vanishes, it suffices to show that

$$
\mathbb{E}\big[\big(N^\delta_{\underline{\ell}+1,a} - N^\delta_{\underline{\ell},a}\big)N^\delta_{\underline{\ell},a}\big] = 0.
$$

This follows since we just showed $\mathbb{E}[Z^\delta_{\underline{\ell}+1,a}|Z^\delta_{\underline{\ell},a}] = Z^\delta_{\underline{\ell},a}$ and we have

$$
\mathbb{E}\big[\big(N^\delta_{\underline{\ell}+1,a} - N^\delta_{\underline{\ell},a}\big)N^\delta_{\underline{\ell},a}\big] = \mathbb{E}\big[u^\delta_{\underline{\ell},a}\big(X^\delta_{\underline{\ell},a}\big)\big(Z^\delta_{\underline{\ell}+1,a} - Z^\delta_{\underline{\ell},a}\big)\big] = \mathbb{E}\big[u^\delta_{\underline{\ell},a}\big(Z^\delta_{\underline{\ell},a}\big)\big(Z^\delta_{\underline{\ell}+1,a} - Z^\delta_{\underline{\ell},a}\big)\big]
$$

Next we verify the base case for Equation (8.2). Using the base case of Equation (8.1) in the first step we compute:

$$
\begin{aligned}
\mathbb{E}\big[\big(Z^\delta_{\underline{\ell}+1,a} - Z^\delta_{\underline{\ell},a}\big)^2\big] &= \mathbb{E}\big[\big(Z^\delta_{\underline{\ell}+1,a}\big)^2\big] - \mathbb{E}\big[\big(Z^\delta_{\underline{\ell},a}\big)^2\big] \\
&= \xi'\big(\mathbb{E}\big[\big(N^\delta_{\underline{\ell},a}\big)^2\big]\big) - \xi'(\underline{q}) \\
&= \xi'\big((1+\varepsilon_0)^2\underline{q}\big) - \xi'(\underline{q}) \\
&= \xi'(\underline{q} + \delta) - \xi'(\underline{q}) \\
&= \xi'\big(q^\delta_{\underline{\ell}+1}\big) - \xi'(\underline{q}).
\end{aligned}
$$

Continuing, we verify the base case for Equation (8.5). First note that

$$
\begin{aligned}
\mathbb{E}\big[\big(N^\delta_{\underline{\ell}+1,a} - N^\delta_{\underline{\ell},a}\big)|\mathcal{F}^\delta_{\underline{\ell}}\big] &= \mathbb{E}\big[u^\delta_{\underline{\ell}}\big(X^\delta_{\underline{\ell},a}\big)\big(Z^\delta_{\underline{\ell}+1,a} - Z^\delta_{\underline{\ell},a}\big)\,|\,\mathcal{F}^\delta_{\underline{\ell}}\big] \\
&= 0.
\end{aligned}
$$

The last line holds because $X^\delta_{\underline{\ell},a}$ is $\mathcal{F}^\delta_{\underline{\ell}}$-measurable and $\mathbb{E}[Z^\delta_{\underline{\ell}+1,a} - Z^\delta_{\underline{\ell},a}|\mathcal{F}^\delta_{\underline{\ell}}] = 0$ as deduced above. Finally for Equation (8.6) using the martingale property again we obtain:

$$
\begin{aligned}
\mathbb{E}\big[\big(N^\delta_{\underline{\ell}+1,a} - N^\delta_{\underline{\ell},a}\big)^2\big] &= \mathbb{E}\big[\big(u^\delta_{\underline{\ell}}\big(X^\delta_{\underline{\ell},a}\big)\big)^2\big(Z^\delta_{\underline{\ell}+1,a} - Z^\delta_{\underline{\ell},a}\big)^2\big] \\
&= \mathbb{E}\left[\frac{\delta}{\xi'(q^\delta_{\underline{\ell}+1,a}) - \xi'(q^\delta_{\underline{\ell}})}\big(\xi'\big(q^\delta_{\underline{\ell}+1}\big) - \xi'\big(q^\delta_{\underline{\ell}}\big)\big)\right] \\
&= \delta.
\end{aligned}
$$

Here the second line follows from the definition of u^δ_ℓ, and we can multiply the two expectations because $\mathbb{E}[(Z^\delta_{\underline{\ell}+1,a} - Z^\delta_{\underline{\ell},a})^2|\mathcal{F}^\delta_{\underline{\ell},a}]$ is constant while the other term is $\mathcal{F}^\delta_{\underline{\ell},a}$ measurable.

Inductive step We now induct, assuming all 6 identities (8.1), (8.2), (8.5), (8.6), (8.4) and (8.8) up to ℓ and showing Equations (8.1), (8.2), (8.5), (8.6) for $\ell + 1$. We begin with Equation (8.1). Let $\underline{\ell} + 1 \leq j \leq \ell$. State evolution implies

$$
\mathbb{E}\big[\big(Z^\delta_{\ell+1,a} - Z^\delta_{\ell,a}\big)Z^\delta_{j,a}\big] = \xi'\big(\mathbb{E}\big[N^\delta_{\ell,a}N^\delta_{j-1,a}\big]\big) - \xi'\big(\mathbb{E}\big[N^\delta_{\ell-1,a}N^\delta_{j-1,a}\big]\big).
$$

To show this equals 0 we must show

$$\mathbb{E}\big[N_{\ell,a}^\delta N_{j-1,a}^\delta\big] = \mathbb{E}\big[N_{\ell-1,a}^\delta N_{j-1,a}^\delta\big].$$

When $\ell = \ell_{q_i}^\delta$ for some $i \in [m]$ this follows from Corollary 8.2. Assuming $\ell \neq \ell_{q_i}^\delta$ for all i, the difference between the left and right sides is

$$\mathbb{E}\big[u_{\ell-1}^\delta\big(X_{\ell-1,a}^\delta\big)\big(Z_{\ell,a}^\delta - Z_{\ell-1,a}^\delta\big)N_{j-1,a}^\delta\big].$$

Since $N_{j-1,a}^\delta$ is $\mathcal{F}_{\ell-1}^\delta$ measurable and $\mathbb{E}[Z_{\ell,a}^\delta|\mathcal{F}_{\ell-1}^\delta] = Z_{\ell-1,a}^\delta$ holds by inductive hypothesis, we conclude the inductive step for Equation (8.1).

We continue to Equation (8.2). Using Equation (8.1) just proven in the first step we get

$$\begin{aligned}
\mathbb{E}\big[\big(Z_{\ell+1,a}^\delta - Z_{\ell,a}^\delta\big)^2\big] &= \mathbb{E}\big[\big(Z_{\ell+1,a}^\delta\big)^2 - \big(Z_{\ell,a}^\delta\big)^2\big] \\
&= \xi'\big(\mathbb{E}\big[N_{\ell,a}^\delta N_{\ell,a}^\delta\big]\big) - \xi'\big(\mathbb{E}\big[N_{\ell-1,a}^\delta N_{\ell-1,a}^\delta\big]\big) \\
&= \xi'\big(q_{\ell+1}^\delta\big) - \xi'\big(q_\ell^\delta\big)
\end{aligned}$$

Next we show Equation (8.5) continues to hold. If $\ell + 1 = \ell_{q_i}^\delta$ for some $i \in [m]$ again follows from Corollary 8.2. When $\ell + 1 \neq \ell_{q_i}^\delta$ for all i, it follows from the definition of the sequence $N_{\ell,a}^\delta$ and the just proven fact that $(Z_{\ell,a}^\delta)_{\ell \geq \underline{\ell}+1}$ forms a martingale sequence. Finally we show Equation (8.6) continues to hold inductively. Again for $\ell + 1 = \ell_{q_i}^\delta$ it follows from Corollary 8.2, and otherwise by definition

$$\mathbb{E}\big[\big(u_\ell^\delta\big)^2\big] = \frac{\delta}{\xi'(q_\ell^\delta) - \xi'(q_{\ell-1}^\delta)}.$$

Moreover what we showed before implies $\mathbb{E}[(Z_{\ell+1,a}^\delta - Z_{\ell,a}^\delta)^2|\mathcal{F}_\ell^\delta] = \xi'(q_\ell^\delta) - \xi'(q_{\ell-1}^\delta)$. Applying these observations to the identity

$$\mathbb{E}\big[\big(N_{\ell+1,a}^\delta - N_\ell^\delta\big)^2|\mathcal{F}_\ell\big] = \big(u_\ell^\delta\big)^2\mathbb{E}\big[\big(Z_{\ell+1,a}^\delta - Z_{\ell,a}^\delta\big)^2\big]$$

implies Equation (8.6) continues to hold.

Equations (8.3) **and** (8.7)   Finally we consider (8.3) and (8.7). For $\ell < \ell_{q_i}^\delta$ they follow directly from (8.2), (8.6). For $\ell = \ell_{q_i}^\delta$, (8.7) is trivial while (8.3) immediately follows from state evolution. For $\ell > \ell_{q_i}^\delta$, (8.7) follows from the inductive hypothesis and the computation

$$\mathbb{E}\big[\big(N_{\ell+1,1}^\delta - N_{\ell,1}^\delta\big)\big(N_{\ell+1,2}^\delta - N_{\ell,2}^\delta\big)|\mathcal{F}_\ell^\delta\big] = \big(u_\ell^\delta\big)^2\mathbb{E}\big[\big(Z_{\ell+1,1}^\delta - Z_{\ell,1}^\delta\big)\big(Z_{\ell+1,2}^\delta - Z_{\ell,2}^\delta\big)|\mathcal{F}_\ell^\delta\big] = 0.$$

Finally for $\ell > \ell_{q_i}^\delta$, (8.3) follows from the expansion

$$\begin{aligned}
\mathbb{E}\big[\big(Z_{\ell+1,1}^\delta - Z_{\ell,1}^\delta\big)\big(Z_{\ell+1,2}^\delta - Z_{\ell,2}^\delta\big)\big] = {}&\xi'\big(\mathbb{E}\big[N_{\ell,1}^\delta N_{\ell,2}^\delta\big]\big) - \xi'\big(\mathbb{E}\big[N_{\ell-1,1}^\delta N_{\ell,2}^\delta\big]\big) \\
&- \xi'\big(\mathbb{E}\big[N_{\ell-1,1}^\delta N_{\ell,2}^\delta\big]\big) + \xi'\big(\mathbb{E}\big[N_{\ell-1,1}^\delta N_{\ell-1,2}^\delta\big]\big)
\end{aligned}$$

and the fact that all 4 terms on the right hand side are equal thanks to (8.5), (8.7). $\qquad\square$

8.1 Diffusive scaling limit

We begin with the following slight generalization of Lemma 3.7 which allows for the additional perturbation steps of branching IAMP but still considers only a single sample path.

Lemma 8.4. *Fix $\bar{q} \in (q, 1)$ and an index a. There exists a coupling between the families of triples $\{(Z_{\ell,a}^\delta, X_{\ell,a}^\delta, N_{\ell,a}^\delta)\}_{\ell \geq 0}$ and $\{(Z_t, X_t, N_t)\}_{t \geq 0}$ such that the following holds for a constant $C > 0$. For large enough $\underline{\ell}$, and every $\ell \geq \underline{\ell}$ with $q_\ell \leq \bar{q}$,*

$$\max_{\underline{\ell} \leq j \leq \ell} \mathbb{E}\big[\big(X_{j,a}^\delta - X_{q_j}\big)^2\big] \leq C\delta, \tag{8.9}$$

$$\max_{\underline{\ell} \leq j \leq \ell} \mathbb{E}\big[\big(N_{j,a}^\delta - N_{q_j}\big)^2\big] \leq C\delta. \tag{8.10}$$

Proof. We prove the scaling limits for X_ℓ^δ and N_ℓ^δ separately, inducting over ℓ in each proof. We suppress the index a as it is irrelevant.

Scaling limit for X_ℓ^δ We begin by checking the claim for $\ell = \underline{\ell}$. Recalling that $\int_0^{q_{\underline{\ell}+1}} \sqrt{\xi''(t)} \mathrm{d}B_t = Z_{\underline{\ell}+1}$, we have

$$\begin{aligned}
\mathbb{E}\big[\big(X_{\underline{\ell}}^\delta - X_q\big)^2\big] &= \mathbb{E}\bigg[\bigg(Z_{\underline{\ell}}^\delta - \int_0^{q_{\underline{\ell}+1}} \sqrt{\xi''(t)} \mathrm{d}B_t\bigg)^2\bigg] \\
&\leq 2\mathbb{E}\big[\big(Z_{\underline{\ell}}^\delta - Z_{\underline{\ell}+1}^\delta\big)^2\big] + 2\mathbb{E}\bigg[\bigg(\int_q^{q_{\underline{\ell}+1}} \sqrt{\xi''(t)} \mathrm{d}B_t\bigg)^2\bigg] \\
&= 4\big(\xi'(q_{\underline{\ell}+1}) - \xi'(q)\big) \\
&\leq C\delta.
\end{aligned}$$

We continue using a standard self-bounding argument. Let $\ell \geq \underline{\ell} + 1$ such that $q_\ell \leq \bar{q}$. Define $\Delta_\ell^X = X_\ell^\delta - X_{q_\ell}$. Then

$$\begin{aligned}
\Delta_\ell^X - \Delta_{\ell-1}^X &= \int_{q_{\ell-1}^\delta}^{q_\ell^\delta} \big(v\big(q_{\ell-1}^\delta; X_\ell^\delta\big) - v(t; X_t)\big)\mathrm{d}t + Z_\ell^\delta - Z_{\ell-1}^\delta - \int_{q_{\ell-1}^\delta}^{q_\ell^\delta} \sqrt{\xi''(s)} \mathrm{d}B_s \\
&= \int_{q_{\ell-1}^\delta}^{q_\ell^\delta} \big(v\big(q_{\ell-1}^\delta; X_\ell^\delta\big) - v(t; X_t)\big)\mathrm{d}t \\
&= \int_{q_{\ell-1}^\delta}^{q_\ell^\delta} \big(v\big(q_{\ell-1}^\delta; X_\ell^\delta\big) - v\big(q_{\ell-1}^\delta; X_t\big)\big)\mathrm{d}t + \int_{q_{\ell-1}^\delta}^{q_\ell^\delta} \big(v\big(q_{\ell-1}^\delta; X_t\big) - v(t; X_t)\big)\mathrm{d}t.
\end{aligned}$$

The first term just above is at most $C\int_{q_{j-1}^\delta}^{q_j^\delta} |X_j^\delta - X_t|\mathrm{d}t$ since v is Lipschitz in space uniformly for $t \in [0, 1]$. For the second term we estimate

$$\begin{aligned}
&\sum_{k=\underline{\ell}+1}^\ell \int_{q_{k-1}^\delta}^{q_k^\delta} \big|v\big(q_{k-1}^\delta; X_t\big) - v(t; X_t)\big|\mathrm{d}t \\
&\leq \sum_{k=\underline{\ell}+1}^\ell \int_{q_{k-1}^\delta}^{q_k^\delta} \big\{\big|v\big(q_{k-1}^\delta; X_t\big) - v(t; X_t)\big| + \big|v(t; X_t) - v\big(q_k^\delta; X_t\big)\big|\big\}\mathrm{d}t \\
&\leq \delta \sum_{k=\underline{\ell}+1}^\ell \sup_{q_{k-1}^\delta \leq t \leq q_k^\delta} \big\{\big|v\big(q_{k-1}^\delta; X_t\big) - v(t; X_t)\big| + \big|v(t; X_t) - v\big(q_k^\delta; X_t\big)\big|\big\} \\
&\leq \delta \sup_{t_1, \ldots, t_k} \sum_{k=\underline{\ell}+1}^\ell \big\{\big|v\big(q_{k-1}^\delta; X_{t_k}\big) - v(t_k; X_{t_k})\big| + \big|v(t_k; X_{t_k}) - v\big(q_k^\delta; X_{t_k}\big)\big|\big\} \\
&\leq C\delta,
\end{aligned}$$

where the last inequality follows from the strong total variation of v. Combining the bounds and summing over j, we find

$$|\Delta_\ell^X| \le |\Delta_{\underline{\ell}}^X| + \sum_{j=\underline{\ell}+1}^\ell |\Delta_j^X - \Delta_{j-1}^X| \le C \sum_{j=\underline{\ell}+1}^\ell \int_{q_{j-1}^\delta}^{q_j^\delta} |X_j^\delta - X_t| \mathrm{d}t + 2C\delta.$$

Squaring and taking expectations,

$$\mathbb{E}\big[(\Delta_\ell^X)^2\big] \le 2C^2 \mathbb{E}\bigg(\sum_{j=\underline{\ell}+1}^\ell \int_{q_{j-1}^\delta}^{q_j^\delta} |X_j^\delta - X_t| \mathrm{d}t\bigg)^2 + 10C^2\delta^2$$

$$\le 2C^2(\ell - \underline{\ell})\delta \sum_{j=\underline{\ell}+1}^\ell \int_{q_{j-1}^\delta}^{q_j^\delta} \mathbb{E}|X_j^\delta - X_t|^2 \mathrm{d}t + 10C^2\delta^2.$$

Furthermore, $\mathbb{E}|X_j^\delta - X_t|^2 \le 2\mathbb{E}|X_j^\delta - X_{q_j^\delta}|^2 + 2\mathbb{E}|X_{q_j^\delta} - X_t|^2$. It is clear that $\mathbb{E}|X_t - X_s|^2 \le C|t - s|$ for all t, s, as ξ'' is bounded on $[0,1]$. Therefore

$$\mathbb{E}\big[(\Delta_\ell^X)^2\big] \le 4C^2(\ell - \underline{\ell})\delta^2 \sum_{j=\underline{\ell}+1}^\ell \mathbb{E}\big[(\Delta_j^X)^2\big] + 4C^3(\ell - \underline{\ell})\delta \sum_{j=\underline{\ell}+1}^\ell \int_{q_{j-1}^\delta}^{q_j^\delta} \delta \mathrm{d}t + 10C^2\delta^2.$$

The middle term is proportional to $(\ell - \underline{\ell})^2 \delta^3$. Using $(\ell - \underline{\ell})\delta \le 1$ we obtain that for δ smaller than an absolute constant, it holds that

$$\mathbb{E}\big[(\Delta_\ell^X)^2\big] \le C\delta \sum_{j=\underline{\ell}+1}^{\ell-1} \mathbb{E}\big[(\Delta_j^X)^2\big] + C\delta,$$

for a different absolute constant C. This implies $\mathbb{E}[(\Delta_\ell^X)^2] \le C\delta$ as desired.

Scaling limit for N_ℓ^δ   Again we begin by checking that $\ell = \underline{\ell}$. We compute:

$$\mathbb{E}\big[(N_{\underline{\ell}}^\delta - N_q)^2\big] = \mathbb{E}\big[\big((1 + \varepsilon_0)\partial_x \Phi_{\gamma_*}(q, X_{\underline{\ell}}) - \partial_x \Phi_{\gamma_*}(q, X_q)\big)^2\big]$$

$$\le 2\varepsilon_0^2 \mathbb{E}\big[(\partial_x \Phi_{\gamma_*}(q, X_{\underline{\ell}}))^2\big] + 2\mathbb{E}\big[\big(\partial_x \Phi_{\gamma_*}(q, X_{\underline{\ell}}) - \partial_x \Phi_{\gamma_*}(q, X_q)\big)^2\big]$$

$$= C\varepsilon_0^2 + C\mathbb{E}\big[(X_{\underline{\ell}}^\delta - X_q)^2\big]$$

$$\le C\delta.$$

Here we have used again the inequality $(x - z)^2 \le 2(x - y)^2 + 2(y - z)^2$ and the fact that derivatives of Φ_{γ_*} are bounded, as well as $\varepsilon_0^2 \le \delta/q$. At the end we use the bound on $\mathbb{E}[(X_{\underline{\ell}}^\delta - X_q)^2]$ shown in the previous part of this proof. Next we turn to $\ell \ge \underline{\ell} + 1$. We have

$$\big(N_{j+1}^\delta - N_{q_{j+1}^\delta}\big) - \big(N_j^\delta - N_{q_j^\delta}\big) = u_j^\delta(X_j^\delta)\big(Z_{j+1}^\delta - Z_j^\delta\big) - \int_{q_j^\delta}^{q_{j+1}^\delta} \sqrt{\xi''(t)} u(t, X_t) \mathrm{d}B_t$$

$$= \int_{q_j^\delta}^{q_{j+1}^\delta} \sqrt{\xi''(t)}\big(u_j^\delta(X_j^\delta) - u(t, X_t)\big) \mathrm{d}B_t,$$

and so

$$\mathbb{E}\big[(N_\ell^\delta - N_{q_\ell^\delta})^2\big] \le 2 \cdot \mathbb{E}\big[(N_{\underline{\ell}}^\delta - N_q)^2\big] + 2 \cdot \mathbb{E}\Bigg[\bigg(\sum_{j=\underline{\ell}}^{\ell-1} \int_{q_j^\delta}^{q_{j+1}^\delta} \sqrt{\xi''(t)}\big(u_j^\delta(X_j^\delta) - u(t, X_t)\big) \mathrm{d}B_t\bigg)^2\Bigg]$$

$$\le 2C\delta + 2\sum_{j=\underline{\ell}}^{\ell-1} \int_{q_j^\delta}^{q_{j+1}^\delta} \mathbb{E}\big[\big(u_j^\delta(X_j^\delta) - u(t, X_t)\big)^2\big] \xi''(t) \mathrm{d}t. \tag{8.11}$$

Recall that $u_j^\delta(x) = u(q_j^\delta; x)/\Sigma_j^\delta$ for $j \geq 1$ where Σ_j^δ is given by

$$\big(\Sigma_j^\delta\big)^2 = \frac{\xi'(q_{j+1}^\delta) - \xi'(q_j^\delta)}{\delta} \mathbb{E}\big[u\big(q_j^\delta; X_j^\delta\big)^2\big].$$

We first show the bound

$$\big|\big(\Sigma_j^\delta\big)^2 - 1\big| \leq C\sqrt{\delta} \tag{8.12}$$

for δ small enough, which is of independent interest. Since u is bounded and ξ''' is bounded on $[0, 1]$, we have

$$\big|\big(\Sigma_j^\delta\big)^2 - \xi''\big(q_j^\delta\big)\mathbb{E}\big[u\big(q_j^\delta; X_j^\delta\big)^2\big]\big| \leq C\delta.$$

Observe now that

$$\mathbb{E}\big[|X_j^\delta - X_{q_j^\delta}|\big] \leq \sqrt{\delta} + \frac{\mathbb{E}[|X_j^\delta - X_{q_j^\delta}|^2]}{\sqrt{\delta}} \leq C\sqrt{\delta}.$$

Since u is Lipschitz in space and bounded, this implies

$$\big|\big(\Sigma_j^\delta\big)^2 - \xi''\big(q_j^\delta\big)\mathbb{E}\big[u\big(q_j^\delta; X_{q_j^\delta}\big)^2\big]\big| \leq C\sqrt{\delta}.$$

Since $\mathbb{E}[N_t^2] = t$ for all $t \in [0, 1]$ and $t \mapsto u(t, X_t)$ is a.s. continuous, Lebesgue's differentiation theorem implies that for all $t \in [0, 1]$,

$$\xi''(t)\mathbb{E}\big[u(t; X_t)^2\big] = 1,$$

and hence $|(\Sigma_j^\delta)^2 - 1| \leq C\sqrt{\delta}$ for δ smaller than some absolute constant. This implies the bound $|u_j^\delta(X_j^\delta) - u(q_j^\delta; X_j^\delta)| \leq C|\frac{1}{\Sigma_j^\delta} - 1| \leq C\sqrt{\delta}$. Now, going back to Eq. (8.11), we have

$$\mathbb{E}\big[\big(N_\ell^\delta - N_{q_\ell^\delta}\big)^2\big] \leq 2\sum_{j=\underline{\ell}}^{\ell-1} \int_{q_j^\delta}^{q_{j+1}^\delta} \mathbb{E}\big[\big(u_j^\delta\big(X_j^\delta\big) - u\big(q_j^\delta; X_j^\delta\big)\big)^2\big]\xi''(t)\mathrm{d}t$$

$$+ 2\sum_{j=\underline{\ell}}^{\ell-1} \int_{q_j^\delta}^{q_{j+1}^\delta} \mathbb{E}\big[\big(u\big(q_j^\delta; X_j^\delta\big) - u(t, X_t)\big)^2\big]\xi''(t)\mathrm{d}t.$$

From what we just established the first term is at most $C(\ell - \underline{\ell})\delta^2 \leq C\delta$. To estimate the second term we compute:

$$\sum_{j=\underline{\ell}}^{\ell-1} \int_{q_j^\delta}^{q_{j+1}^\delta} \mathbb{E}\big[\big(u\big(q_j^\delta; X_j^\delta\big) - u(t, X_t)\big)^2\big]\xi''(t)\mathrm{d}t$$

$$\leq C\sum_{j=\underline{\ell}}^{\ell-1} \int_{q_j^\delta}^{q_{j+1}^\delta} \mathbb{E}\big[\big(u\big(q_j^\delta; X_j^\delta\big) - u\big(q_j^\delta, X_{q_j^\delta}\big)\big)^2\big]\mathrm{d}t$$

$$+ C\sum_{j=\underline{\ell}}^{\ell-1} \int_{q_j^\delta}^{q_{j+1}^\delta} \mathbb{E}\big[\big(u\big(q_j^\delta; X_{q_j^\delta}\big) - u\big(q_j^\delta, X_t\big)\big)^2\big]\mathrm{d}t$$

$$+ C\sum_{j=\underline{\ell}}^{\ell-1} \int_{q_j^\delta}^{q_{j+1}^\delta} \mathbb{E}\big[\big(u\big(q_j^\delta; X_t\big) - u(t, X_t)\big)^2\big]\mathrm{d}t$$

$$= I + II + III.$$

Since u is Lipschitz in space, we obtain $I \leq C(\ell - \underline{\ell})\delta^2$. From $\mathbb{E}[|X_t - X_s|^2] \leq C|t - s|$, we obtain $II \leq C(\ell - \underline{\ell})\delta^2$. Finally, since u is Lipschitz in time uniformly in space and $(\ell - \underline{\ell})\delta \leq 1$, it follows that $III \leq C\delta$. Altogether we obtain

$$\mathbb{E}\big[\big(N_\ell^\delta - N_{q_\ell^\delta}\big)^2\big] \leq C\delta$$

concluding the proof. $\qquad\square$

We now extend Lemmas 3.7, 8.4 to describe the joint scaling limit of multiple branches, which become independent at the branching time. Let $(B_t^a)_{t\in[0,1],a\in\{1,2\}}$ be standard Brownian motions with $B_t^1 = B_t^2$ for $t \leq q_B$ and with independent increments after time q_B. Couple B_t^a with $(Z_{\ell,a}^\delta)_{\ell\geq 0}$ via

$$Z_{j,a}^\delta = \int_0^{q_j^\delta} \sqrt{\xi''(t)}\mathrm{d}B_t^a$$

and natural filtration $(\mathcal{F}_t)_{t\in[0,1]}$ with $\mathcal{F}_t = \sigma((B_s^1, B_s^2)_{s\leq t})$. We consider for $a \in \{1,2\}$ the SDE

$$dX_t^a = \gamma_*(t)\partial_x\Phi_{\gamma_*}(t, X_t^a)\mathrm{d}t + \sqrt{\xi''(t)}\mathrm{d}B_t^a$$

with initial condition $X_0^a = 0$, and define

$$N_t^a \equiv \partial_x\Phi_{\gamma_*}(q, X_{\underline{q}}^a) + \int_{\underline{q}}^t \sqrt{\xi''(s)}u(s, X_s^a)\mathrm{d}B_s^a = \partial_x\Phi_{\gamma_*}(t, X_t^a),$$

$$Z_t^a \equiv \int_0^t \sqrt{\xi''(s)}\mathrm{d}B_s^a.$$

Lemma 8.5. *Fix $\bar{q} \in (\underline{q}, 1)$. There exists a coupling between the families of triples $\{(Z_{\ell,a}^\delta, X_{\ell,a}^\delta, N_{\ell,a}^\delta)\}_{\ell\geq 0,a\in\{1,2\}}$ and $\{(Z_t^a, X_t^a, N_t^a)\}_{t\geq 0,a\in\{1,2\}}$ such that the following holds. For some $\delta_0 > 0$ and constant $C > 0$, for every $\delta \leq \delta_0$ and $\ell \geq \underline{\ell}$ with $q_\ell \leq \bar{q}$ we have*

$$\max_{\underline{\ell}\leq j\leq\ell} \mathbb{E}\big[(X_{j,a}^\delta - X_{q_j}^a)^2\big] \leq C\delta,$$

$$\max_{\underline{\ell}\leq j\leq\ell} \mathbb{E}\big[(N_{j,a}^\delta - N_{q_j}^a)^2\big] \leq C\delta.$$

Proof. We generate the desired "grand coupling" by starting with (B_t^1, B_t^2) as above, generating (Z_t^1, Z_t^2), and then setting $Z_{j,a}^\delta = Z_{q_j^\delta}^a$ for each $a \in \{1,2\}$ and $j \leq \bar{\ell}$ as in the coupling of Lemma 8.4. It follows from Lemma 8.3 that this results in the correct law for $(Z_{j,a}^\delta)_{j\in\mathbb{N},a\in\{1,2\}}$. Now, all 3 continuous-time functions in the coupling of Lemma 8.4 are determined almost surely by Z_t. Furthermore all 3 discrete-time functions are determined almost surely by the sequence Z_j^δ. Therefore the coupling just constructed between $\{Z_{\ell,a}^\delta\}_{\ell\geq 0,a\in\{1,2\}}$ and $\{Z_t^a\}_{t\geq 0,a\in\{1,2\}}$ automatically extends to a coupling of $\{(Z_{\ell,a}^\delta, X_{\ell,a}^\delta, N_{\ell,a}^\delta)\}_{\ell\geq 0,a\in\{1,2\}}$ and $\{(Z_t^a, X_t^a, N_t^a)\}_{t\geq 0,a\in\{1,2\}}$. Since the two a-marginals of the coupling just constructed both agree with that of Lemma 8.4, the claimed approximation estimates carry over as well, concluding the proof. $\square$

8.2 The energy gain of incremental AMP

Here we prove Lemma 3.9, stated for the branching case.

Lemma 8.6.

$$\lim_{\bar{q}\to 1}\lim_{\underline{\ell}\to\infty}\operatorname*{p-lim}_{N\to\infty}\frac{H_N(\boldsymbol{n}^{\bar{\ell},a}) - H_N(\boldsymbol{n}^{\underline{\ell},a})}{N} = \int_{\underline{q}}^1 \xi''(t)\mathbb{E}[u(t, X_t)]\mathrm{d}t. \tag{8.13}$$

Proof. We give the main part of the proof for the ordinary (non-branching) version of the algorithm and explain at the end why the same arguments apply in the branching case. Recall that $\delta = \delta(\underline{\ell}) \to 0$ as $\underline{\ell} \to \infty$, which we will implicitly use throughout the proof. Observe also that $\langle h, \boldsymbol{n}^{\bar{\ell}} - \boldsymbol{n}^{\underline{\ell}}\rangle_N \simeq 0$ because the values $(N_\ell^\delta)_{\ell\geq\underline{\ell}}$ form a martingale sequence. Therefore it suffices to compute the in-probability limit of $\frac{\widetilde{H}_N(\boldsymbol{n}^{\bar{\ell}}) - \widetilde{H}_N(\boldsymbol{n}^{\underline{\ell}})}{N}$. The key is to write

$$\frac{\widetilde{H}_N(\boldsymbol{n}^{\bar{\ell}}) - \widetilde{H}_N(\boldsymbol{n}^{\underline{\ell}})}{N} = \sum_{\ell=\underline{\ell}}^{\bar{\ell}-1} \frac{\widetilde{H}_N(\boldsymbol{n}^{\ell+1}) - \widetilde{H}_N(\boldsymbol{n}^\ell)}{N}$$

and use a Taylor series approximation of the summand. In particular for $F \in C^3(\mathbb{R})$, applying Taylor's approximation theorem twice yields

$$
\begin{aligned}
F(1) - F(0) &= aF'(0) + \frac{1}{2}F''(0) + O\Big(\sup_{a \in [0,1]} |F'''(a)|\Big) \\
&= F'(0) + \frac{1}{2}\big(F'(a) - F'(0)\big) + O\Big(\sup_{a \in [0,1]} |F'''(a)|\Big) \\
&= \frac{1}{2}\big(F'(1) + F'(0)\big) + O\Big(\sup_{a \in [0,1]} |F'''(a)|\Big).
\end{aligned}
$$

Assuming $\sup_\ell \frac{\|\boldsymbol{n}^\ell\|}{\sqrt{N}} \leq 1 + \eta$, which holds with high probability for any $\eta > 0$ if $\underline{\ell}$ is large enough, we apply this estimate with $F(a) = \widetilde{H}_N((1-a)\boldsymbol{n}^\ell + a\boldsymbol{n}^{\ell+1})$. Recalling (2.2), the result is:

$$
\begin{aligned}
\Big| \widetilde{H}_N\big(\boldsymbol{n}^{\ell+1}\big) - \widetilde{H}_N\big(\boldsymbol{n}^\ell\big) &- \frac{1}{2}\big\langle \nabla \widetilde{H}_N\big(\boldsymbol{n}^\ell\big) + \nabla \widetilde{H}_N\big(\boldsymbol{n}^{\ell+1}\big), \boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell \big\rangle \Big| \\
&\leq O\Big(\sup_{\|\boldsymbol{v}\| \leq (1+\eta)\sqrt{N}} \big\|\nabla^3 \widetilde{H}_N(\boldsymbol{v})\big\|_{\mathrm{op}} \Big) \|\boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell\|^3.
\end{aligned}
$$

Proposition 2.1 implies that

$$
\sup_{|\boldsymbol{v}| \leq (1+\eta)\sqrt{N}} \big\|\nabla^3 \widetilde{H}_N(\boldsymbol{v})\big\|_{\mathrm{op}} \leq O(N^{-1/2})
$$

with high probability. On the other hand $\text{p-}\lim_{N \to \infty} \|\boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell\| = \sqrt{\delta N}$ for each $\underline{\ell} \leq \ell \leq \overline{\ell} - 1$. Summing and recalling that $\overline{\ell} - \underline{\ell} \leq \delta^{-1}$ yields the high-probability estimate

$$
\begin{aligned}
\sum_{\ell=\underline{\ell}}^{\overline{\ell}-1} \Big| \widetilde{H}_N\big(\boldsymbol{n}^{\ell+1}\big) &- \widetilde{H}_N\big(\boldsymbol{n}^\ell\big) - \frac{1}{2}\big\langle \nabla \widetilde{H}_N\big(\boldsymbol{n}^\ell\big) + \nabla \widetilde{H}_N\big(\boldsymbol{n}^{\ell+1}\big), \boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell \big\rangle \Big| \\
&\leq \sum_{\ell=\underline{\ell}}^{\overline{\ell}-1} O\Big(\sup_{\|\boldsymbol{x}\| \leq (1+\eta)\sqrt{N}} \big\|\nabla^3 \widetilde{H}_N(\boldsymbol{x})\big\|_{\mathrm{op}} \Big) \cdot \sup_\ell \big\|\boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell\big\|^3 \\
&\leq O(N\sqrt{\delta}).
\end{aligned}
$$

Because $\underline{\ell} \to \infty$ implies $\delta \to 0$ this term vanishes in the limit, and it remains to show

$$
\lim_{\overline{q} \to 1} \lim_{\underline{\ell} \to \infty} \text{p-}\lim_{N \to \infty} \sum_{\ell=\underline{\ell}}^{\overline{\ell}-1} \frac{1}{2}\big\langle \nabla \widetilde{H}_N\big(\boldsymbol{n}^\ell\big) + \nabla \widetilde{H}_N\big(\boldsymbol{n}^{\ell+1}\big), \boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell \big\rangle_N = \int_q^1 \xi''(t) \mathbb{E}\big[u(t, X_t)\big] \mathrm{d}t.
$$

Next, observe by (3.11) that:

$$
\nabla \widetilde{H}_N\big(\boldsymbol{n}^\ell\big) = \boldsymbol{z}^{\ell+1} + \sum_{j=0}^{\ell} d_{\ell,j} \boldsymbol{n}^{j-1}. \tag{8.14}
$$

Passing to the limiting Gaussian process $(Z_k^\delta)_{k \in \mathbb{Z}^+}$ via state evolution, and ignoring for now the constant number of branching updates,

$$
\text{p-}\lim_{N \to \infty}\big\langle \nabla \widetilde{H}_N\big(\boldsymbol{n}^\ell\big), \boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell \big\rangle_N = \mathbb{E}\big[Z_{\ell+1}^\delta\big(N_{\ell+1}^\delta - N_\ell^\delta\big)\big] + \sum_{j=0}^{\ell} d_{\ell,j} \mathbb{E}\big[N_{j-1}^\delta\big(N_{\ell+1}^\delta - N_\ell^\delta\big)\big],
$$

$$
\text{p-}\lim_{N \to \infty}\big\langle \nabla \widetilde{H}_N\big(\boldsymbol{n}^{\ell+1}\big), \boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell \big\rangle_N = \mathbb{E}\big[Z_{\ell+2}^\delta\big(N_{\ell+1}^\delta - N_\ell^\delta\big)\big] + \sum_{j=0}^{\ell+1} d_{\ell+1,j} \mathbb{E}\big[N_{j-1}^\delta\big(N_{\ell+1}^\delta - N_\ell^\delta\big)\big].
$$

As $(N_k^\delta)_{k \geq \mathbb{Z}^+}$ is a martingale process, it follows that the right-hand expectations all vanish. Similarly it holds that

$$\mathbb{E}\big[Z_{\ell+2}^\delta\big(N_{\ell+1}^\delta - N_\ell^\delta\big)\big] = \mathbb{E}\big[Z_{\ell+1}^\delta\big(N_{\ell+1}^\delta - N_\ell^\delta\big)\big]$$
$$\mathbb{E}\big[Z_\ell^\delta\big(N_{\ell+1}^\delta - N_\ell^\delta\big)\big] = 0.$$

Rewriting and using Lemma 8.3 in the last step,

$$\underset{N \to \infty}{\text{p-lim}} \frac{1}{2}\big\langle \nabla \widetilde{H}_N\big(\boldsymbol{n}^\ell\big) + \nabla \widetilde{H}_N\big(\boldsymbol{n}^{\ell+1}\big), \boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell \big\rangle_N = \mathbb{E}\big[\big(Z_{\ell+1}^\delta - Z_\ell^\delta\big)\big(N_{\ell+1}^\delta - N_\ell^\delta\big)\big]$$
$$= \mathbb{E}\big[u_\ell^\delta\big(X_\ell^\delta\big)\big(Z_{\ell+1}^\delta - Z_\ell^\delta\big)^2\big]$$
$$= \mathbb{E}\big[\mathbb{E}\big[u_\ell^\delta\big(X_\ell^\delta\big)\big(Z_{\ell+1}^\delta - Z_\ell^\delta\big)^2\big|\mathcal{F}_\ell^\delta\big]\big]$$
$$= \big(\xi'\big(q_{\ell+1}^\delta\big) - \xi'\big(q_\ell^\delta\big)\big) \cdot \mathbb{E}\big[u_\ell^\delta\big(X_\ell^\delta\big)\big]$$
$$= \big(\xi'\big(q_{\ell+1}^\delta\big) - \xi'\big(q_\ell^\delta\big)\big) \cdot \frac{\mathbb{E}[u_{q_\ell^\delta}\big(X_\ell^\delta\big)]}{\Sigma_\ell^\delta}.$$

Recalling (8.12), the fact that $u_t(x)$ is uniformly Lipschitz in x for $t \in [0, \bar{q}]$, the fact that $\xi'(q_{\ell+1}^\delta) - \xi'(q_\ell^\delta) = \delta \xi''(q_\ell^\delta) + O(\delta^2)$ and the coupling of Lemma 8.4, it follows that

$$\underset{N \to \infty}{\text{p-lim}} \frac{1}{2}\big\langle \nabla \widetilde{H}_N\big(\boldsymbol{n}^\ell\big) + \nabla \widetilde{H}_N\big(\boldsymbol{n}^{\ell+1}\big), \boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell \big\rangle_N = \delta \xi''\big(q_\ell^\delta\big)\mathbb{E}\big[u_{q_\ell^\delta}(X_{q_\ell^\delta})\big] + O_{\bar{q}}\big(\delta^{3/2}\big).$$

Summing over ℓ and using continuity of $u_t(x)$ in t, it follows that

$$\lim_{\underline{\ell} \to \infty} \underset{N \to \infty}{\text{p-lim}} \frac{\widetilde{H}_N\big(\boldsymbol{n}^{\overline{\ell}}\big) - \widetilde{H}_N\big(\boldsymbol{n}^{\underline{\ell}}\big)}{N} = \int_{\underline{q}}^{\bar{q}} \xi''(t)\mathbb{E}\big[u(t, X_t)\big]\mathrm{d}t.$$

Sending $\bar{q} \to 1$ now concludes the proof when there are no branching steps. Extending the proof to cover branching steps is not difficult and we explain it now. Everything up to (8.14) is still valid, and if the number $|Q|$ of branching steps is m, then the full analysis applies to all but m terms. However the simple uniform bound

$$\widetilde{H}_N\big(\boldsymbol{n}^{\ell+1}\big) - \widetilde{H}_N\big(\boldsymbol{n}^\ell\big) \leq \|\boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell\| \cdot \sup_{\|\boldsymbol{x}\|_N \leq 1 + \frac{\eta}{2}} \|\nabla \widetilde{H}_N(\boldsymbol{x})\|$$
$$\leq O(\sqrt{N\delta}) \cdot O(\sqrt{N})$$
$$\leq O(N\sqrt{\delta})$$

holds with high probability. Here we used Proposition 2.7 to deduce $\|\boldsymbol{n}^\ell\|_N, \|\boldsymbol{n}^{\ell+1}\|_N \leq 1 + o(1)$ with high probability, and also Proposition 2.1 and Equation (8.6). Therefore all telescoping terms, branching or not, uniformly contribute $O(N\sqrt{\delta})$ energy in probability. As a result, even when a constant number of non-branching terms are replaced by branching terms, the same analysis applies up to error $O(N\sqrt{\delta})$, yielding the same asymptotic energy for branching $\underline{q}$-IAMP and completing the proof. $\qquad \square$

References

[AAZB+17]   Naman Agarwal, Zeyuan Allen-Zhu, Brian Bullins, Elad Hazan, and Tengyu Ma. Finding approximate local minima faster than gradient descent. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1195–1199, 2017. MR3678262

[ABA13]   Antonio Auffinger and Gérard Ben Arous. Complexity of random smooth functions on the high-dimensional sphere. *Annals of Probability*, 41(6):4214–4247, 2013. MR3161473

[ABE+05]   Sanjeev Arora, Eli Berger, Hazan Elad, Guy Kindler, and Muli Safra. On non-approximability for quadratic programs. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 206–215. IEEE, 2005.

[AC15]   Antonio Auffinger and Wei-Kuo Chen. The Parisi formula has a unique minimizer. *Communications in Mathematical Physics*, 335(3):1429–1444, 2015. MR3320318

[AC17]   Antonio Auffinger and Wei-Kuo Chen. Parisi formula for the ground state energy in the mixed p-spin model. *The Annals of Probability*, 45(6b):4617–4631, 2017. MR3737919

[ACORT11]   Dimitris Achlioptas, Amin Coja-Oghlan, and Federico Ricci-Tersenghi. On the solution-space geometry of random constraint satisfaction problems. *Random Structures & Algorithms*, 38(3):251–268, 2011. MR2663730

[ACZ20]   Antonio Auffinger, Wei-Kuo Chen, and Qiang Zeng. The SK model is infinite step replica symmetry breaking at zero temperature. *Communications on Pure and Applied Mathematics*, 73(5), 2020. MR4078711

[AJ18]   Gérard Ben Arous and Aukosh Jagannath. Spectral gap estimates in mean field spin glasses. *Communications in Mathematical Physics*, 361(1):1–52, 2018. MR3825934

[AM20]   Ahmed El Alaoui and Andrea Montanari. Algorithmic thresholds in mean field spin glasses. *arXiv preprint arXiv:2009.11481*, 2020.

[AMMN19]   Gerard Ben Arous, Song Mei, Andrea Montanari, and Mihai Nica. The landscape of the spiked tensor model. *Communications on Pure and Applied Mathematics*, 72(11):2282–2330, 2019. MR4011861

[AMS21]   Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *The Annals of Probability*, 49(6):2922–2960, 2021. MR4348682

[AS22]   Ahmed El Alaoui and Mark Sellke. Algorithmic pure states for the negative spherical perceptron. *Journal of Statistical Physics*, 189(2):27, 2022. MR4483884

[ASZ20]   Gérard Ben Arous, Eliran Subag, and Ofer Zeitouni. Geometry and temperature chaos in mixed spherical spin glasses at low temperature: the perturbative regime. *Communications on Pure and Applied Mathematics*, 73(8):1732–1828, 2020. MR4113545

[Auf13]   Antonio Auffinger and Ben Arous, Gérard and Cernỳ, Jirí. Random matrices and complexity of spin glasses. *Communications on Pure and Applied Mathematics*, 66(2):165–201, 2013. MR2999295

[AWZ20]   Gérard Ben Arous, Alexander S Wein, and Ilias Zadik. Free energy wells and overlap gap property in sparse PCA. In *Conference on Learning Theory*, pages 479–482. PMLR, 2020.

[BBH+12]   Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*, pages 307–326. ACM, 2012. MR2961513

[BD98]   Sebastian Böcker and Andreas WM Dress. Recovering symbolically dated, rooted trees from symbolic ultrametrics. *Advances in Mathematics*, 138(1):105–125, 1998. MR1645064

[BLM15]   Mohsen Bayati, Marc Lelarge, and Andrea Montanari. Universality in polytope phase transitions and message passing algorithms. *The Annals of Applied Probability*, 25(2):753–822, 2015. MR3313755

[BM11]   Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. on Inform. Theory*, 57:764–785, 2011. MR2810285

[BMN19]   Raphaël Berthier, Andrea Montanari, and Phan-Minh Nguyen. State evolution for approximate message passing with non-separable functions. *Information and Inference: A Journal of the IMA*, 2019. MR4079177

[Bol14]   Erwin Bolthausen. An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Communications in Mathematical Physics*, 325(1):333–366, 2014. MR3147441

[CDHS17]   Yair Carmon, John C Duchi, Oliver Hinder, and Aaron Sidford. "Convex until proven guilty": Dimension-free acceleration of gradient descent on non-convex functions. In *International Conference on Machine Learning*, pages 654–663. PMLR, 2017.

[CDHS18]   Yair Carmon, John C Duchi, Oliver Hinder, and Aaron Sidford. Accelerated methods for nonconvex optimization. *SIAM Journal on Optimization*, 28(2):1751–1772, 2018. MR3814027

[CDHS19]   Yair Carmon, John C Duchi, Oliver Hinder, and Aaron Sidford. Lower bounds for finding stationary points I. *Mathematical Programming*, pages 1–50, 2019. MR4163541

[CGP$^+$19]   Wei-Kuo Chen, David Gamarnik, Dmitry Panchenko, Mustazee Rahman, et al. Suboptimality of local algorithms for a class of max-cut problems. *The Annals of Probability*, 47(3):1587–1618, 2019. MR3945754

[Cha09]   Sourav Chatterjee. Disorder chaos and multiple valleys in spin glasses. *arXiv preprint arXiv:0907.3381*, 2009. MR2363958

[Che13]   Wei-Kuo Chen. The Aizenman-Sims-Starr scheme and Parisi formula for mixed p-spin spherical models. *Electronic Journal of Probability*, 18, 2013. MR3126577

[CHL18]   Wei-Kuo Chen, Madeline Handschy, and Gilad Lerman. On the energy landscape of the mixed even p-spin model. *Probability Theory and Related Fields*, 171(1-2):53–95, 2018. MR3800830

[CKP$^+$14]   Patrick Charbonneau, Jorge Kurchan, Giorgio Parisi, Pierfrancesco Urbani, and Francesco Zamponi. Exact theory of dense amorphous hard spheres in high dimension. iii. the full replica symmetry breaking solution. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(10):P10009, 2014. MR3278087

[CL21]   Wei-Kuo Chen and Wai-Kit Lam. Universality of approximate message passing algorithms. *Electronic Journal of Probability*, 26:1–44, 2021. MR4235487

[CPS21a]   Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. The generalized TAP free energy. *Communications on Pure and Applied Mathematics*, to appear, 2021. MR4588318

[CPS21b]   Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. The generalized TAP free energy II. *Communications in Mathematical Physics*, 381(1):257–291, 2021. MR4207445

[CS92]   Andrea Crisanti and H-J Sommers. The spherical p-spin interaction spin glass model: the statics. *Zeitschrift für Physik B Condensed Matter*, 87(3):341–354, 1992.

[CS17]   Wei-Kuo Chen and Arnab Sen. Parisi formula, disorder chaos and fluctuation for the ground state energy in the spherical mixed p-spin models. *Communications in Mathematical Physics*, 350(1):129–173, 2017. MR3606472

[CS21]   Sourav Chatterjee and Leila Sloman. Average Gromov hyperbolicity and the Parisi ansatz. *Advances in Mathematics*, 376:107417, 2021. MR4178903

[DAM17]   Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the balanced binary stochastic block model. *Information and Inference: A Journal of the IMA*, 6(2):125–170, 2017. MR3671474

[DEZ$^+$15]   Jian Ding, Ronen Eldan, Alex Zhai, et al. On multiple peaks and moderate deviations for the supremum of a gaussian field. *The Annals of Probability*, 43(6):3468–3493, 2015. MR3433585

[Fan22]   Zhou Fan. Approximate message passing algorithms for rotationally invariant matrices. *The Annals of Statistics*, 50(1):197–224, 2022. MR4382014

[GJ21]   David Gamarnik and Aukosh Jagannath. The overlap gap property and approximate message passing algorithms for p-spin models. *The Annals of Probability*, 49(1):180–205, 2021. MR4203336

[GJS21]   David Gamarnik, Aukosh Jagannath, and Subhabrata Sen. The overlap gap property in principal submatrix recovery. *Probability Theory and Related Fields*, 181(4):757–814, 2021. MR4344133

[GJW20]   David Gamarnik, Aukosh Jagannath, and Alexander S Wein. Low-degree hardness of random optimization problems. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 131–140, 2020. MR4232029

[GL18]   David Gamarnik and Quan Li. Finding a large submatrix of a gaussian random matrix. *Annals of Statistics*, 46(6A):2511–2561, 2018. MR3851747

[GS14]   David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*, pages 369–376. ACM, 2014. MR3359490

[GS17]   David Gamarnik and Madhu Sudan. Performance of sequential local algorithms for the random NAE-K-SAT problem. *SIAM Journal on Computing*, 46(2):590–619, 2017. MR3620150

[GZ17]   David Gamarnik and Ilias Zadik. Sparse high-dimensional linear regression. algorithmic barriers and a local search algorithm. *arXiv preprint arXiv:1711.04952*, 2017.

[GZ19]   David Gamarnik and Ilias Zadik. The landscape of the planted clique problem: Dense subgraphs and the overlap gap property. *arXiv preprint arXiv:1904.07174*, 2019.

[HS21]   Brice Huang and Mark Sellke. Tight Lipschitz Hardness for Optimizing Mean Field Spin Glasses. *arXiv preprint arXiv:2110.07847*, 2021. MR4537213

[Jag17]   Aukosh Jagannath. Approximate ultrametricity for random measures and applications to spin glasses. *Communications on Pure and Applied Mathematics*, 70(4):611–664, 2017. MR3628881

[JM13]   Adel Javanmard and Andrea Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference: A Journal of the IMA*, 2(2):115–144, 2013. MR3311445

[JNG^{+}21]   Chi Jin, Praneeth Netrapalli, Rong Ge, Sham M Kakade, and Michael I Jordan. On nonconvex optimization for machine learning: Gradients, stochasticity, and saddle points. *Journal of the ACM (JACM)*, 68(2):1–29, 2021. MR4267060

[JT16]   Aukosh Jagannath and Ian Tobasco. A dynamic programming approach to the Parisi functional. *Proceedings of the American Mathematical Society*, 144(7):3135–3150, 2016. MR3487243

[JT17]   Aukosh Jagannath and Ian Tobasco. Low temperature asymptotics of spherical mean field spin glasses. *Communications in Mathematical Physics*, 352(3):979–1017, 2017. MR3631397

[MMZ05]   Marc Mézard, Thierry Mora, and Riccardo Zecchina. Clustering of solutions in the random satisfiability problem. *Physical Review Letters*, 94(19):197205, 2005.

[Mon19]   Andrea Montanari. Optimization of the Sherrington-Kirkpatrick Hamiltonian. In *IEEE Symposium on the Foundations of Computer Science, FOCS*, November 2019. MR4228234

[Pan]   Dmitry Panchenko. Properties of the Parisi formula. https://drive.google.com/file/d/0B6JeBUquZ5BwRFpLVjdVd3IwV1E/view.

[Pan14]   Dmitry Panchenko. The Parisi formula for mixed p-spin models. *Annals of Probability*, 42(3):946–958, 2014. MR3189062

[RV17]   Mustazee Rahman and Balint Virag. Local algorithms for independent sets are half-optimal. *The Annals of Probability*, 45(3):1543–1577, 2017. MR3650409

[Sel21]   Mark Sellke. Approximate ground states of hypercube spin glasses are near corners. *Comptes Rendus. Mathématique*, 359(9):1097–1105, 2021. MR4334857

[SK75]   David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35(26):1792, 1975.

[Sub17]   Eliran Subag. The complexity of spherical p-spin models—a second moment approach. *The Annals of Probability*, 45(5):3385–3450, 2017. MR3706746

[Sub18]   Eliran Subag. Free energy landscapes in spherical spin glasses. *arXiv preprint arXiv:1804.10576*, 2018.

[Sub21]   Eliran Subag. Following the ground states of full-RSB spherical spin glasses. *Communications on Pure and Applied Mathematics*, 74(5):1021–1044, 2021. MR4230065

[Tal06a]   Michel Talagrand. Free energy of the spherical mean field model. *Probability Theory and Related Fields*, 134(3):339–382, 2006. MR2226885

[Tal06b] Michel Talagrand. The Parisi formula. *Annals of Mathematics*, pages 221–263, 2006. MR2195134

[Wei22] Alexander S Wein. Optimal low-degree hardness of maximum independent set. *Mathematical Statistics and Learning*, 4(3):221–251, 2022. MR4383734

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4], ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System: https://vtex.lt/services/ejms-peer-review/
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: https://imstat.org/shop/donation/