The Annals of Probability
2021, Vol. 49, No. 6, 2922–2960
https://doi.org/10.1214/21-AOP1519

OPTIMIZATION OF MEAN-FIELD SPIN GLASSES

BY AHMED EL ALAOUI[1], ANDREA MONTANARI[2] AND MARK SELLKE[3]

[1]*Department of Statistics and Data Science, Cornell University, elalaoui@cornell.edu*
[2]*Department of Electrical Engineering and Department of Statistics, Stanford University, montanar@stanford.edu*
[3]*Department of Mathematics, Stanford University, msellke@stanford.edu*

Mean-field spin glasses are families of random energy functions (Hamiltonians) on high-dimensional product spaces. In this paper, we consider the case of Ising mixed p-spin models,; namely, Hamiltonians $H_N : \Sigma_N \to \mathbb{R}$ on the Hamming hypercube $\Sigma_N = \{\pm 1\}^N$, which are defined by the property that $\{H_N(\boldsymbol{\sigma})\}_{\boldsymbol{\sigma}\in\Sigma_N}$ is a centered Gaussian process with covariance $\mathbb{E}\{H_N(\boldsymbol{\sigma}_1)H_N(\boldsymbol{\sigma}_2)\}$ depending only on the scalar product $\langle\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\rangle$.

The asymptotic value of the optimum $\max_{\boldsymbol{\sigma}\in\Sigma_N} H_N(\boldsymbol{\sigma})$ was characterized in terms of a variational principle known as the Parisi formula, first proved by Talagrand and, in a more general setting, by Panchenko. The structure of superlevel sets is extremely rich and has been studied by a number of authors. Here, we ask whether a near optimal configuration $\boldsymbol{\sigma}$ can be computed in polynomial time.

We develop a message passing algorithm whose complexity per-iteration is of the same order as the complexity of evaluating the gradient of H_N, and characterize the typical energy value it achieves. When the p-spin model H_N satisfies a certain no-overlap gap assumption, for any $\varepsilon > 0$, the algorithm outputs $\boldsymbol{\sigma} \in \Sigma_N$ such that $H_N(\boldsymbol{\sigma}) \geq (1-\varepsilon)\max_{\boldsymbol{\sigma}'} H_N(\boldsymbol{\sigma}')$, with high probability. The number of iterations is bounded in N and depends uniquely on ε. More generally, regardless of whether the no-overlap gap assumption holds, the energy achieved is given by an extended variational principle, which generalizes the Parisi formula.

1. Introduction.

Let $\boldsymbol{W}^{(k)} \in (\mathbb{R}^N)^{\otimes k}$, $k \geq 2$, be a standard symmetric Gaussian tensor of order k with entries $\boldsymbol{W}^{(k)} \equiv (W^{(k)}_{i_1,\dots,i_k})_{1\leq i_1,\dots,i_k\leq N}$. Namely, if $\{G^{(k)}_{i_1,\dots,i_k} : k \geq 2, 1 \leq i_1,\dots,i_k \leq N\}$ is a collection of i.i.d. standard normal $\mathsf{N}(0,1)$ random variables, we set $\boldsymbol{W}^{(k)} \equiv N^{-(k-1)/2}\sum_{\pi\in S_k}\boldsymbol{G}^{(k)}_\pi$ where the sum is over the group of permutations of k objects, and $\boldsymbol{G}^{(k)}_\pi$ is obtained by permuting the indices of $\boldsymbol{G}^{(k)}$ according to π.

We consider the problem of optimizing a polynomial with coefficients given by the tensors $\boldsymbol{W}^{(k)}$ over the hypercube $\Sigma_N = \{-1,+1\}^N$:

$$\mathsf{OPT}_N = \frac{1}{N}\max\{H_N(\boldsymbol{\sigma}) : \boldsymbol{\sigma} \in \Sigma_N\}, \tag{1.1}$$

$$H_N(\boldsymbol{\sigma}) = \sum_{k=2}^{\infty}\frac{c_k}{k!}\langle\boldsymbol{W}^{(k)},\boldsymbol{\sigma}^{\otimes k}\rangle, \quad \langle\boldsymbol{W}^{(k)},\boldsymbol{\sigma}^{\otimes k}\rangle \equiv \sum_{1\leq i_1,\dots,i_k\leq N} W_{i_1,\dots,i_k}\sigma_{i_1}\cdots\sigma_{i_k}. \tag{1.2}$$

The parameters $(c_k)_{k\geq 2}$ are customarily encoded in the function $\xi(x) \equiv \sum_{k\geq 2} c_k^2 x^k$, which we henceforth call the *mixture* of the model. We will assume throughout that $\xi(1+\varepsilon) < \infty$ for some $\varepsilon > 0$. This implies $|c_k| \leq c_*\alpha^k$ for some $c_* > 0$, $\alpha \in (0,1)$, so that the sum defining H_N is almost surely finite. (In fact, there is very little loss of generality in assuming $c_k = 0$ for all k larger than some absolute constant $k_{\max}$.) We call such a model an *Ising* mixed p-spin model. When the combinatorial constraint $\boldsymbol{\sigma} \in \Sigma_N$ is relaxed to the ℓ_2 norm constraint

Received March 2020; revised February 2021.
MSC2020 subject classifications. Primary 68Q87, 82C44; secondary 60K35.
Key words and phrases. Spin glasses, finding ground states, the Parisi formula, optimization algorithms.

$\|\boldsymbol{\sigma}\|^2 = N$, we obtain the related *spherical* mixed p-spin model. If $c_k = 0$ for all $k \neq p$, the model is called a *pure* p-spin model.

We would like to develop an algorithm that accepts as input the tensors $(\boldsymbol{W}^{(k)})_{k\geq 2}$ and returns a vector $\boldsymbol{\sigma}^{\mathrm{alg}} \in \Sigma_N$ such that, with high probability, $H_N(\boldsymbol{\sigma}^{\mathrm{alg}})/N \geq \rho \cdot \mathsf{OPT}_N$ for an approximation factor $\rho \in [0, 1]$ as close to one as possible. From a worst case point of view, this objective is hopeless: achieving any $\rho > 1/(\log N)^c$ (for c a small constant) is NP-hard already in the case of quadratic polynomials [2]. For higher-order polynomials, the task is known to be even more difficult. For instance, [6] proves that obtaining $\rho > \exp(-(\log N)^c)$ is hard already for the spherical pure p-spin model whenever $p \geq 3$.

Worst-case hardness results do not have direct implications on random instances, as described above. However, standard optimization methods based on semidefinite programming (SDP) relaxations appear to fail on such random instances. These methods typically produce an efficiently computable upper bound on OPT_N. For a spherical pure k-spin model with $k \geq 3$, [11] shows that a level-k sum-of-squares relaxation produces an upper bound that is polynomially larger than OPT_N: $\mathsf{SOS}_N(k) \gtrsim N^{(k-2)/4} \cdot \mathsf{OPT}_N$. In contrast, significant progress has been achieved recently for *search* algorithms, that is, algorithms that produce a feasible solution $\boldsymbol{\sigma}^{\mathrm{alg}}$ but not a certificate of (near-)optimality. In particular, Subag [45] developed an algorithm for the spherical mixed p-spin model, and proved that it achieves any approximation factor $\rho = (1 - \varepsilon)$, $\varepsilon > 0$, provided $t \mapsto \xi''(t)^{-1/2}$ is concave. In [37], one of the authors developed an algorithm for the Sherrington–Kirkpatrick model, which corresponds to the quadratic case ($c_2 = 1$ and $c_k = 0$ for $k \geq 3$), with $\boldsymbol{\sigma} \in \Sigma_N$. Under a widely believed conjecture about the so-called Parisi formula, the algorithm of [37] also achieves a $(1 - \varepsilon)$-approximation for any $\varepsilon > 0$.

The main result of this paper is a characterization of the optimal value achieved by a natural class of low-complexity message passing algorithms that generalize the approach of [37]. As special cases, we recover the results of [45] and [37]. For a given approximation error $\varepsilon > 0$, the algorithm complexity is of the same order as evaluating the gradient $\nabla H_N(\boldsymbol{x})$ at a constant number $C(\varepsilon)$ of points. Its output $\boldsymbol{\sigma}^{\mathrm{alg}}$ satisfying $H_N(\boldsymbol{\sigma}^{\mathrm{alg}})/N \geq (1-\varepsilon)\cdot\mathsf{OPT}_N$ with high probability whenever the corresponding Parisi formula satisfies a certain "no-overlap gap" condition. Even more interestingly, we characterize the optimal value achieved by a natural class of message passing algorithms in terms of an extended variational principle, which generalizes the Parisi formula. This points at a possible general picture for the optimal approximation ratio in ensembles of random optimization problems.

The random energy function H_N has been studied for over 40 years in statistical physics and probability theory, and is known as the Hamiltonian of the mixed p-spin model [33, 38, 44, 47]. With the above definitions, it is easy to see that $\{H(\boldsymbol{\sigma})\}_{\boldsymbol{\sigma}\in\Sigma_N}$ is a centered Gaussian process on the hypercube, with covariance

$$\mathbb{E}\big[H_N(\boldsymbol{\sigma})H_N(\boldsymbol{\sigma}')\big] = N\xi\big(\langle\boldsymbol{\sigma}, \boldsymbol{\sigma}'\rangle/N\big). \tag{1.3}$$

The asymptotic value of OPT_N was first derived by physicists using the nonrigorous replica method [40] and subsequently established by Talagrand [46] and Panchenko [38, 39]. This asymptotic value is characterized in terms of a variational principle known as the "Parisi formula." While the Parisi formula allows computation of the asymptotic free energy associated to the Hamiltonian H_N, it can be specialized to the zero temperature case to determine the asymptotics of OPT_N. The resulting characterization was established by Auffinger and Chen in [5] and it is useful to recall it for the reader's convenience.

Let $\mathscr{U}$ be the following subset of functions $\gamma : [0, 1) \to \mathbb{R}_{\geq 0}$:

$$\mathscr{U} \equiv \left\{\gamma : [0, 1) \to \mathbb{R}_{\geq 0} : \gamma \text{ nondecreasing }, \int_0^1 \gamma(t)\,\mathrm{d}t < \infty\right\}. \tag{1.4}$$

For $\gamma \in \mathscr{U}$, let $\Phi^\gamma : [0,1] \times \mathbb{R} \to \mathbb{R}$ be the solution of the following PDE, known as *the Parisi PDE*, with terminal condition at $t = 1$:

$$\begin{aligned} &\partial_t \Phi^\gamma(t,x) + \frac{1}{2}\xi''(t)\big(\partial_x^2 \Phi^\gamma(t,x) + \gamma(t)\big(\partial_x \Phi^\gamma(t,x)\big)^2\big) = 0, \\ &\Phi^\gamma(1,x) = |x|. \end{aligned} \tag{1.5}$$

We refer to Section 2.2 and Section 6 for a construction of solutions of this PDE.

The Parisi functional $\mathsf{P} : \mathscr{U} \to \mathbb{R}$ is then defined by

$$\mathsf{P}(\gamma) \equiv \Phi^\gamma(0,0) - \frac{1}{2}\int_0^1 t\xi''(t)\gamma(t)\,\mathrm{d}t. \tag{1.6}$$

THEOREM 1 ([5]). *The following limit holds almost surely*:

$$\lim_{N\to\infty} \mathsf{OPT}_N = \inf_{\gamma \in \mathscr{U}} \mathsf{P}(\gamma). \tag{1.7}$$

The optimization problem on the right-hand side of the last formula is achieved at a unique function $\gamma_P \in \mathscr{U}$ [5, 16], which has a physical interpretation [33]. Consider the (random) Boltzmann distribution $p_\beta(\boldsymbol{\sigma}) \propto \exp\{\beta H_N(\boldsymbol{\sigma})\}$ at temperature $1/\beta$, and let $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \sim p_\beta$ be two independent samples from this distribution, that is, $(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \sim \mathbb{E}p_\beta^{\otimes 2}$. Then $\beta^{-1}\gamma_P(t)$ is the asymptotic probability of the event $\{(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) : |\langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2\rangle|/N \le t\}$ (when the limit $\beta \to \infty$ is taken *after* $N \to \infty$.) Given this interpretation, the nondecreasing constraint in the definition of $\mathscr{U}$ is very natural: it follows from γ_P being the limit of a sequence of cumulative distribution functions (rescaled by the factor β).

As mentioned above, in this paper we describe and analyze a class of algorithms that aim at finding near-optima, that is, configurations $\boldsymbol{\sigma}^{\mathrm{alg}} \in \Sigma_N$ with $H_N(\boldsymbol{\sigma}^{\mathrm{alg}})/N$ as close as possible to OPT_N (or to its asymptotic value $\inf_{\gamma\in\mathscr{U}} \mathsf{P}(\gamma)$). Our main results can be summarized as follows:

1. If the infimum in the Parisi formula is achieved at γ_P, which is strictly increasing over the interval $[0,1)$, then we provide an efficient algorithm that returns a $(1-\varepsilon)$-optimizer. This condition corresponds to the "no-overlap gap" scenario mentioned above: roughly speaking, the asymptotic distribution of the overlap $|\langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2\rangle|/N$ has no gap in its support.

2. More generally, we introduce a new extended variational principle, which prescribes to minimize the Parisi functional P over a larger space $\mathscr{L}$ of functions γ, which are *not* necessarily monotone. We present an algorithm that achieves $H(\boldsymbol{\sigma}^{\mathrm{alg}})/N \ge (1-\varepsilon)\inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma)$, provided the infimum on the right-hand side is achieved at some $\gamma_* \in \mathscr{L}$. Since $\mathscr{U} \subseteq \mathscr{L}$, this value is of course no larger than the value of the global optimum.

Moreover, under the "no-overlap gap" scenario, we have $\inf_{\gamma\in\mathscr{U}} \mathsf{P}(\gamma) = \inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma)$ and, therefore, we recover the result at the previous point.

3. We show, by a duality argument, that no algorithm in a natural class of class of message passing algorithms that we introduce can overcome the value $\inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma)$. This appears to be an interesting computational threshold, whose importance warrants further exploration.

1.1. *Further background.* Understanding the average case hardness of random computational problems is an outstanding challenge with numerous ramifications. The use of spin glass concepts in this context has a long history, which is impossible to review here. A few pointers include [30, 32–34, 36]. Spin glass theory allows one to derive a detailed picture of the structure of superlevel sets of random optimization problems, or the corresponding Boltzmann distribution $p_\beta(\boldsymbol{\sigma}) \propto \exp\{\beta H_N(\boldsymbol{\sigma})\}$. A central challenge in this area is to understand the connection between this picture and computational tractability. Which features of the energy landscape H_N are connected to intractability?

Of course, the answer depends on the precise formulation of the question. In this paper, we consider the specific problem of achieving the best approximation factor ρ so that a polynomial-time algorithm can output a feasible solution $\boldsymbol{\sigma}^{\rm alg}$ such that $H_N(\boldsymbol{\sigma}^{\rm alg})/N \geq \rho \mathsf{OPT}_N$ with high probability. This question was addressed in the physics literature from at least two points of view:

- Significant effort has been devoted to computing the number (and energy) of local optima that are separated by large energy barriers: the energy of the most numerous such local optima is sometimes used as a proxy for the algorithmic threshold. The exponential growth-rate of the number of such optima is computed using nonrigorous methods in [17, 18, 41].
- An equally large amount of work was devoted to the study of Glauber or Langevin dynamics, which can be interpreted as greedy optimization algorithms. In particular, [13, 20] and follow-up work study the $N \to \infty$ asymptotics of these dynamics, for a fixed time horizon.

These two approaches produced an impressive amount of (mostly nonrigorous) information. Despite these advances, no clear picture has been put forward for the optimum approximation factor ρ (the "algorithmic threshold"), except in particularly simple cases, such as the pure p-spin spherical model. We refer to [23] for a recent illustration of the outstanding challenges.

Over the last two years, significant progress was achieved on this question. Apart from [37, 45] mentioned above, Addario-Berry and Maillard [1] studied this question within the generalized random energy model, which can be viewed as a stylized model for the energy landscape of mean field spin glasses. They prove that a variant of greedy search achieves a $(1-\varepsilon)$-approximation of OPT_N under a suitable variant of the no-overlap gap assumption.

In a different direction, Gamarnik and coauthors showed in several examples that the existence of an overlap gap (defined appropriately) rules out a $(1-\varepsilon)$-approximation for local algorithms in related random optimization problems on sparse graphs [15, 25, 26]. Furthermore, the recent paper [24] proves that approximate message passing algorithms (of the type studied in this paper) cannot achieve a $(1-\varepsilon)$-approximation of the optimum in pure p-spin Ising models, under the assumption that these exhibit an overlap gap. However, [24] does not characterize optimal approximation ratio, which we instead do here, as a special case of our results.

Finally, two recent papers [31, 35] study degree-4 sum-of-squares relaxations for the Sherrington–Kirkpatrick model, and show that they fail at producing an upper bound on OPT_N tighter than what is produced by simple spectral methods. In conjunction with [37], these results suggest that—in the context of spin glass problems—computing a certifiable upper bound on OPT_N is fundamentally harder than searching for an approximate optimizer.

Our approach is based on the construction and analysis of a class of approximate message passing (AMP) algorithms. Following [37], we refer to this family of algorithms as incremental approximate message passing (IAMP). AMP algorithms admit an exact asymptotic characterization in terms of a limiting Gaussian process, which is known as state evolution. This characterization was first established rigorously by Bolthausen [12] for a special case, and subsequently generalized in several papers [7, 8, 10, 29]. Here, we will follow the proof scheme of [10] to generalize state evolution to the case of (mixed) tensors.

1.2. *Notation*. We will typically use lower-case for scalars (e.g., $x, y, \dots$), bold lower-case for vectors (e.g., $\boldsymbol{x}, \boldsymbol{y}, \dots$), and bold upper case for matrices (e.g., $\boldsymbol{X}, \boldsymbol{Y}, \dots$). The ordinary scalar product in $\mathbb{R}^d$ is denoted by $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \sum_{i\le d} x_i y_i$, and the corresponding norm by $\|\boldsymbol{x}\| = \langle \boldsymbol{x}, \boldsymbol{x}\rangle^{1/2}$. Given two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^N$, we will often consider the normalized scalar product $\langle \boldsymbol{a}, \boldsymbol{b}\rangle_N = \sum_{i\le N} a_i b_i/N$, and the norm $\|\boldsymbol{a}\|_N = \langle \boldsymbol{a}, \boldsymbol{a}\rangle_N^{1/2}$. There will be no confusion between this and ℓ_p norms, which will be rarely used in $\mathbb{R}^d$.

We will use standard notation for functional spaces, in particular, spaces of continuously differentiable functions (e.g., $C^k(\Omega)$, $C_c^k(\Omega)$, and so on), and spaces of integrable functions (e.g., $L^p(\Omega)$). We refer, for instance, to [22] for definitions.

Given a sequence of random variables $(Y_n)_{n\ge 1}$, and Y_∞, we write $Y_n \xrightarrow{p} Y_\infty$, or $\text{p-lim}_{n\to\infty} Y_n = Y_\infty$ if Y_n converges in probability to Y_∞.

For a function $f : \mathbb{R} \to \mathbb{R}$, we denote by $\|f\|_{\mathrm{TV}[a,b]}$ the total variation of f on the interval $[a, b]$:

$$\|f\|_{\mathrm{TV}[a,b]} \equiv \sup_n \sup_{a\le t_0<t_1<\cdots<t_n\le b} \sum_{i=1}^n \bigl|f(t_i) - f(t_{i-1})\bigr|. \tag{1.8}$$

(The supremum is taken over all partitions of the interval $[a, b]$.)

We say that a function $\psi : \mathbb{R}^d \to \mathbb{R}$ is *pseudo-Lipschitz* if there exists a constant $L < \infty$ such that, for every $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$,

$$\bigl|\psi(\boldsymbol{x}) - \psi(\boldsymbol{y})\bigr| \le L\bigl(1 + \|\boldsymbol{x}\| + \|\boldsymbol{y}\|\bigr)\|\boldsymbol{x} - \boldsymbol{y}\|.$$

Throughout the paper, we write that an event holds *with high probability*, if its probability converges to one as $N \to \infty$. We use C to denote various constants, whose value can change from line to line.

2. Achievability.

2.1. *Value achieved by message passing algorithms*. We characterize the value achieved by a class of message-passing algorithms, presented in Section 3. This class is parametrized by two functions $u, v : [0, 1] \times \mathbb{R} \to \mathbb{R}$, and the value they achieve is given in Theorem 2, under the assumptions spelled out below.

DEFINITION 2.1. We say that a function $f : [a, b] \times \mathbb{R} \to \mathbb{R}$ has bounded *strong* total variation if there exists $C < \infty$ such that

$$\sup_n \sup_{a\le t_0<\cdots<t_n\le b} \sup_{x_1,\dots,x_n\in\mathbb{R}} \sum_{i=1}^n \bigl|f(t_i, x_i) - f(t_{i-1}, x_i)\bigr| \le C. \tag{2.1}$$

(The supremum is over all partitions (t_i) of the interval $[a, b]$ and all sequences (x_i) in $\mathbb{R}$.)

ASSUMPTION 1. Let $u, v : [0, 1] \times \mathbb{R} \to \mathbb{R}$ be two measurable functions, with u nonvanishing, and assume that the following holds for some constant $C < \infty$:

(A1) u and v are uniformly bounded: $\sup_{t,x} |u(t, x)| \vee |v(t, x)| \le C$.

(A2) u and v are Lipschitz continuous in space, with uniform (in time) Lipschitz constant: $|u(t, x_1) - u(t, x_2)| \vee |v(t, x_1) - v(t, x_2)| \le C|x_1 - x_2|$ for all $x_1, x_2 \in \mathbb{R}$ and $t \in [0, 1]$.

(A3) $u(\cdot, x)$ is continuous for all $x \in \mathbb{R}$.

(A4) u and v have bounded strong total variation.

Consider the following stochastic differential equation:

$$\mathrm{d}X_t = v(t, X_t)\,\mathrm{d}t + \sqrt{\xi''(t)}\,\mathrm{d}B_t, \quad \text{with } X_0 = 0, \tag{2.2}$$

where $(B_t)_{t\in[0,1]}$ is a standard Brownian motion. Under conditions **(A1)** and **(A2)** (pertaining to v), the above SDE has a unique strong solution, which we denote by $(X_t)_{t\in[0,1]}$ [9]. We define the martingale

$$M_t \equiv \int_0^t \sqrt{\xi''(s)}\,u(s, X_s)\,\mathrm{d}B_s. \tag{2.3}$$

As we will see below, the construction of our algorithm will be such that its evolution is indeed captured (in the $N \to \infty$ limit) by the stochastic process $(X_t)_{t \ge 0}$ and its associated martingale $(M_t)_{t \ge 0}$.

Finally, it is useful to introduce a slight modification of the Hamiltonian (1.1). Namely, we denote by $\tilde{H}_N(\boldsymbol{\sigma})$ the function that is obtained by restricting the sums in $H_N(\boldsymbol{\sigma})$ to sets of distinct indices $i_1, \dots, i_k$. (Notice that $\tilde{H}_N(\boldsymbol{\sigma}) = H_N(\boldsymbol{\sigma}) + o(N)$; cf. Section 5.3.2.)

THEOREM 2. *Let Assumption 1 hold, and further assume that $M_{t_*} \in [-1, 1]$ almost surely and $\mathbb{E}[M_t^2] = t$ for all $t \in [0, t_*]$, for some $t_* \in [0, 1]$.*

Further denote by χ the computational complexity of evaluating $\nabla H_N(\boldsymbol{m})$ at a point $\boldsymbol{m} \in [-1, 1]^N$, and by χ_1 the complexity of evaluating one coordinate of $\nabla \tilde{H}_N(\boldsymbol{m})$ at a point $\boldsymbol{m} \in [-1, +1]^N$.

Then for any $\varepsilon > 0$ there exists a randomized algorithm, with complexity $(C/\varepsilon^2) \cdot (\chi + N) + N\chi_1$, which outputs $\boldsymbol{\sigma}^{\mathrm{alg}} \in \Sigma_N$ such that

$$\frac{1}{N} H_N(\boldsymbol{\sigma}^{\mathrm{alg}}) \ge \int_0^{t_*} \xi''(t) \mathbb{E}[u(t, X_t)] \, \mathrm{d}t - \varepsilon, \tag{2.4}$$

with probability converging to one as $N \to \infty$.

The proof of this theorem is deferred to Section 6.

REMARK 2.1. The stated complexity holds in a simplified model of computation whereby real sums and multiplications have complexity of order one. However, we do not anticipate any difficulty to arise from passing to a finite model.

Typically, computing each gradient has complexity that is linear in the input size, and $N\chi_1$ is of the same order as χ. For instance, if the coefficients c_k vanish for $k > k_{\max}$, it is easy to see that $\chi = O(N^{k_{\max}})$, and $\chi_1 = O(N^{k_{\max}-1})$. As a consequence, the dominant term in the complexity is $(C/\varepsilon^2)\chi$. In words, the algorithm's complexity is of the same order as computing the gradient of the cost function C/ε^2 times. We further note that this constant C depends on the regularity constants in Assumption 1.

REMARK 2.2. A similar result can be established for the *spherical* mixed p-spin model, where the constraint $\boldsymbol{\sigma} \in \Sigma_N$ is replaced by $\|\boldsymbol{\sigma}\|^2 = N$. The same conclusion in Theorem 2 holds while the condition $M_{t_*} \in [-1, 1]$ is no longer required. In this case, the choice of the functions u and v is straightforward. Simply set $u(t, x) = \xi''(t)^{-1/2}$: since this is independent of x, the choice of v is immaterial. The value achieved in this case is

$$\frac{1}{N} H_N(\boldsymbol{\sigma}^{\mathrm{spher}}) \ge \int_0^1 \sqrt{\xi''(t)} \, \mathrm{d}t - \varepsilon, \qquad \|\boldsymbol{\sigma}^{\mathrm{spher}}\|^2 = N. \tag{2.5}$$

In this case, we recover the energy achieved by the algorithm of Subag [45].

2.2. *The extended variational principle.* In the Ising case, a specific choice of the functions u and v leads to the following extended variational principle. For a function $\gamma : [0, 1) \to \mathbb{R}$, we write $\xi''\gamma$ for the pointwise multiplication of ξ'' and γ: $\xi''\gamma(t) = \xi''(t)\gamma(t)$. We consider the functional space

$$\mathscr{L} \equiv \left\{ \gamma : [0, 1) \to \mathbb{R}_{\ge 0} : \|\xi''\gamma\|_{\mathrm{TV}[0,t]} < \infty \ \forall t \in [0, 1), \int_0^1 \xi''\gamma(t) \, \mathrm{d}t < \infty \right\}. \tag{2.6}$$

We endow this space with the weighted L^1 metric

$$\|\gamma_1 - \gamma_2\|_{1,\xi''} \equiv \|\xi''(\gamma_1 - \gamma_2)\|_1 = \int_0^1 \xi''(t) |\gamma_1(t) - \gamma_2(t)| \, \mathrm{d}t, \tag{2.7}$$

hence implicitly identifying γ_1 and γ_2 if they coincide for almost every $t \in [0, 1)$. The notation $\|\cdot\|_{\mathrm{TV}[0,t]}$ for total variation norm is defined in equation (1.8). It follows from the definition that if for $\gamma \in \mathscr{L}$, $\xi''\gamma(t) = \nu([0,t])$ where ν is a signed measure[1] of bounded total variation on intervals $[0, 1-\varepsilon]$, $\varepsilon > 0$.

It is obvious that the space $\mathscr{L}$ is a strict superset of $\mathscr{U}$: most crucially, it includes non-monotone functions. As shown in Section 6, the Parisi functional $\gamma \mapsto \mathsf{P}(\gamma)$ can be defined on this larger space.

THEOREM 3. *Assume that the infimum* $\inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma)$ *is achieved at a function* $\gamma_* \in \mathscr{L}$. *Further denote by* χ *the computational complexity of evaluating* $\nabla H_N(\boldsymbol{m})$ *at a point* $\boldsymbol{m} \in [-1,1]^N$, *and by* χ_1 *the complexity of evaluating one coordinate of* $\nabla \tilde{H}_N(\boldsymbol{m})$ *at a point* $\boldsymbol{m} \in [-1,+1]^N$.

Then for every $\varepsilon > 0$ *there exists an algorithm with complexity at most* $C(\varepsilon)\cdot(\chi + N) + N\chi_1$, *which outputs* $\boldsymbol{\sigma}^{\mathrm{alg}} \in \Sigma_N$ *such that*

$$\frac{1}{N} H_N(\boldsymbol{\sigma}^{\mathrm{alg}}) \geq \inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma) - \varepsilon, \tag{2.8}$$

with probability converging to one as $N \to \infty$.

As an important consequence of Theorem 3, we obtain a $(1-\varepsilon)$-approximation of the optimum whenever $\inf_{\gamma\in\mathscr{U}} \mathsf{P}(\gamma)$ is achieved on a strictly increasing function. For future reference, we introduce the following "no-overlap gap" assumption.

ASSUMPTION 2 (No overlap gap at zero temperature). A mixed p-spin model with mixture ξ is said to satisfy the no-overlap gap assumption at zero-temperature if there exists $\gamma_* \in \mathscr{U}$ strictly increasing in $[0, 1)$ such that $\mathsf{P}(\gamma_*) = \inf_{\gamma\in\mathscr{U}} \mathsf{P}(\gamma)$.

The no-overlap gap assumption is expected to hold for some choices of the mixture ξ but not for others. In particular, it is believed to hold for the Sherrington–Kirkpatrick model, which corresponds to the special case $\xi(t) = t^2$, but not for the pure p-spin model, that is, $\xi(t) = c_p^2 t^p$, $p \geq 3$. It is also expected that no-overlap gap holds for some mixed models, that is, models with $c_p > 0$ for more than one values of p. Evidence toward this expectation is mainly based on heuristic arguments (e.g., this property should hold in a "neighborhood" of the Sherrington–Kirkpatrick model), and on analogy with the spherical models where the variational principle can be solved exactly [5].

If the no-overlap gap assumption holds for ξ with minimizer $\gamma_* \in \mathscr{U}$, γ_* also minimizes the Parisi functional over $\mathscr{L}$. In particular, if no-overlap gap holds for ξ, the IAMP algorithm is able to approximately maximize the corresponding Hamiltonian. This leads to the following result, which is formally proved in Section 6.3.

COROLLARY 2.2. *Assume the no-overlap gap assumption to hold for the mixture* ξ. *Then for every* $\varepsilon > 0$ *there exists an algorithm with the same complexity as in Theorem 3, which outputs* $\boldsymbol{\sigma}^{\mathrm{alg}} \in \Sigma_N$ *such that*

$$\frac{1}{N} H_N(\boldsymbol{\sigma}^{\mathrm{alg}}) \geq \mathsf{OPT}_N - \varepsilon, \tag{2.9}$$

with probability converging to one as $N \to \infty$.

[1] This identification holds possibly apart from a set of values of t of vanishing Lebesgue measure, which will be irrelevant here.

REMARK 2.3. Continuing from Remark 2.2, Theorem 3 has an analogue for the spherical model $\|\boldsymbol{\sigma}\|^2 = N$. In this case, the Parisi functional takes a more explicit form [14, 19]:

$$\mathsf{P}^{\text{spher}}(\gamma) = \frac{1}{2}\int_0^1 \left(\xi''(t)\Gamma(t) + \frac{1}{\Gamma(t)}\right) \mathrm{d}t, \qquad \Gamma(t) \equiv \int_t^1 \gamma(s)\,\mathrm{d}s. \tag{2.10}$$

A simple calculation shows that this is minimized in $\mathscr{L}$ at $\gamma_*(t) = -\frac{\mathrm{d}}{\mathrm{d}t}(\xi''(t)^{-1/2})$. This leads to the optimal value $\mathsf{P}^{\text{spher}}(\gamma_*) = \int_0^1 \sqrt{\xi''(t)}\,\mathrm{d}t$, which we anticipated in Remark 2.2. The condition for the minimizer to be in $\mathscr{U}$, $\gamma_* \in \mathscr{U}$, coincides with the condition that $t \mapsto \xi''(t)^{-1/2}$ is concave. This is the condition for no-overlap gap in the spherical model, and is also the condition under which the algorithm of [45] achieves a $(1-\varepsilon)$-optimum.

REMARK 2.4. As mentioned, the optimal choice of the functions u, v is given in terms of the solution γ_* of the variational problem (2.8). One might wonder how the solution of this problem contributes to the overall complexity of finding near optima of the Hamiltonian H_N. While determining the complexity of computing approximations of γ_* (and, therefore, of the optimal functions u, v) is an open problem, this should not have a major impact on the overall complexity, for two reasons:

(i) The variational problem (2.8) needs to be solved (to a given degree of accuracy) *only once*, that is, not for each realization of the Hamiltonian H_N. For this reason, this complexity is not counted in the statement of Theorem 3.

(ii) For any feasible choice of functions u and v, the value achieved by the corresponding IAMP algorithm is given in equation (2.4). This value depends continuously on u and v (this point is clear from the proof of Theorem 2.) Hence, for any fixed ε, we expect to be able to construct—in a time $C_0(\varepsilon)$, independent of N—functions u, v such that $\int_0^{t_*} \xi''(t)\mathbb{E}[u(t, X_t)]\,\mathrm{d}t \geq \inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma) - \varepsilon$. Using such functions instead of the optimal ones only incurs an additional error ε in equation (2.8), and the extra computational cost can be adsorbed in the constant $C(\varepsilon)$.

Determining the dependence of the factor $C(\varepsilon)$ on ε is an important open problem that is not addressed by Theorem 3.

3. Message passing algorithms.

In this section, we introduce a general class of message passing algorithms that we use to prove Theorem 2 and Theorem 3. These are generalizations of the algorithm introduced in [37] for the Sherrington–Kirkpatrick model $\xi(t) = t^2$.

3.1. *The general iteration.*

For each $\ell \geq 0$, let $f_\ell : \mathbb{R}^{\ell+1} \to \mathbb{R}$ be a real-valued Lipschitz function, and let $f_{-1} \equiv 0$. For a sequence of vectors $\boldsymbol{z}^0, \dots, \boldsymbol{z}^\ell \in \mathbb{R}^N$ we use the notation $f_\ell(\boldsymbol{z}^0, \dots, \boldsymbol{z}^\ell)$ for the vector $(f_\ell(z_i^0, \dots, z_i^\ell))_{1\leq i\leq N}$. For a tensor $\boldsymbol{W} \in (\mathbb{R}^N)^{\otimes p}$ and a vector $\boldsymbol{u} \in \mathbb{R}^N$, we denote by $\boldsymbol{W}\{\boldsymbol{u}\}$ the vector $\boldsymbol{v} \in \mathbb{R}^N$ with coordinates

$$v_i = \frac{1}{(p-1)!} \sum_{1\leq i_1,\dots,i_{p-1}\leq N} W_{i,i_1,\dots,i_{p-1}} u_{i_1} \cdots u_{i_{p-1}}.$$

We let $\langle \boldsymbol{u}\rangle_N \equiv \frac{1}{N}\sum_{i=1}^N u_i$ and $\langle \boldsymbol{u}, \boldsymbol{v}\rangle_N \equiv \frac{1}{N}\sum_{i=1}^N u_i v_i$. We write $\boldsymbol{f}_\ell$ as shorthand for the vector $f_\ell(\boldsymbol{z}^0, \dots, \boldsymbol{z}^\ell) \in \mathbb{R}^N$.

In order to fully specify the message passing algorithm, we need to introduce the following Gaussian process. Quantities defined by this process enter in the construction of the algorithm. Let p_0 be a probability distribution on $\mathbb{R}$ and let $Z^0 \sim p_0$. For each $\ell \in \mathbb{Z}$, let $(Z^1, \dots, Z^\ell)$ be a centered Gaussian vector independent of Z^0 with covariance $Q_{j,k} = \mathbb{E}[Z^j Z^k]$ defined recursively by

$$Q_{j+1,k+1} = \xi'(\mathbb{E}[f_j(Z^0, \dots, Z^j) f_k(Z^0, \dots, Z^k)]), \quad j, k \geq 0. \tag{3.1}$$

The message passing algorithm starts with z^0 with coordinates drawn i.i.d. with distribution p_0 independently of everything else. The general message passing iteration takes the form

$$
\begin{aligned}
z^{\ell+1} &= \sum_{p=2}^{\infty} c_p \boldsymbol{W}^{(p)}\{f_\ell(z^0,\dots,z^\ell)\} - \sum_{j=0}^{\ell} d_{\ell,j} f_{j-1}(z^0,\dots,z^{j-1}), \\
d_{\ell,j} &= \xi''(\mathbb{E}[f_\ell(Z^0,\dots,Z^\ell) f_{j-1}(Z^0,\dots,Z^{j-1}]) \cdot \mathbb{E}\left[\frac{\partial f_\ell}{\partial z^j}(Z^0,\dots,Z^\ell)\right].
\end{aligned} \tag{3.2}
$$

Note that the first term in the update equation is the gradient of H_N at the point $f_\ell(z^0,\dots,z^\ell)$. The joint distribution for the first ℓ iterates of equation (3.2) can be exactly characterized in terms of the previously-defined Gaussian process in the $N \to \infty$ limit.

PROPOSITION 3.1 (State evolution). *Assume that p_0 has finite second moment and let $\psi : \mathbb{R}^{\ell+1} \to \mathbb{R}$ be a pseudo-Lipschitz function as defined in Section 1.2. Then*

$$
\langle \psi(z^0,\dots,z^\ell)\rangle_N \xrightarrow[N\to\infty]{p} \mathbb{E}[\psi(Z^0,\dots,Z^\ell)].
$$

This characterization is known as state evolution [7, 8, 10, 12, 29]. The proof of Proposition 3.1 follows from the same technique introduced in [10], and we present it the supplementary material [21]. We note in passing that a version of this result was announced in [42] without proof; the proof in [21] fills this gap.

3.2. *Choice of the nonlinearities.* We now specify the above general iteration to the optimization problem at hand. We reduce the choice of the sequence of functions f_ℓ to two bivariate functions u and v appearing in the main result of Section 2. Let $u, v : [0,1] \times \mathbb{R} \to \mathbb{R}$ be two functions satisfying Assumption 1. Given $z^0,\dots,z^\ell \in \mathbb{R}$ we consider the finite difference equation

$$
x^{j+1} - x^j = v(j\delta; x^j)\delta + (z^{j+1} - z^j), \quad 0 \le j \le \ell - 1, \text{ with } x^0 = 0, \tag{3.3}
$$

with driving "noise" $z^0,\dots,z^\ell$, drift v and 'step size' $\delta > 0$. This is meant to be a discretization of the SDE (2.2), provided that the sequence $z^0,\dots,z^\ell$ "behaves" like Brownian motion. We further let the discrete analogue of the martingale M_t, equation (2.3), be

$$
m^\ell \equiv m^0 + \sum_{j=0}^{\ell-1} u_j^\delta(x^j)(z^{j+1} - z^j), \quad \text{for } \ell \ge 1 \text{ and } m^0 = \sqrt{\delta}, \tag{3.4}
$$

where $u_j^\delta(x) = a_j u(j\delta; x)$ with a_j a bounded rescaling which will be defined in equation (5.1) below.

Note that x^ℓ is a function of $z^0,\dots,z^\ell$ and so is m^ℓ. We define the nonlinearity f_ℓ as the function mapping $z^0,\dots,z^\ell$ to m^ℓ:

$$
f_\ell : (z^0,\dots,z^\ell) \longmapsto m^\ell \quad \text{as per equation (3.3) and equation (3.4)}. \tag{3.5}
$$

The algorithm is completely specified by a choice of the functions $u, v : [0,1] \times \mathbb{R} \to \mathbb{R}$ and outputs a vector $\boldsymbol{m}^\ell \in \mathbb{R}^N$ after $\ell = \lfloor \delta^{-1} \rfloor$ iterations. For any choice of such functions, Theorem 2 predicts the value achieved by the algorithm (for small δ).[2] Theorem 3 corresponds to a specific choice of these functions. Namely, if γ_* minimizes the Parisi functional

[2]To be precise, this procedure returns a vector $\boldsymbol{m}^{\lfloor \delta^{-1} \rfloor}$ which is close (in ℓ_2 distance) to a vertex of the hypercube $[-1,+1]^N$. Since the final output must be a binary vector, we apply a simple rounding procedure to $\boldsymbol{m}^{\lfloor \delta^{-1} \rfloor}$, which is detailed in Section 5.3.

over $\mathscr{L}$ (i.e., $\mathsf{P}(\gamma_*) = \inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma)$), we let $\Phi^{\gamma_*} : [0, 1] \times \mathbb{R} \to \mathbb{R}$ denote the corresponding solution of the PDE (1.5). We let

$$(3.6)\qquad v(t,x) = \xi''(t)\gamma_*(t)\partial_x \Phi^{\gamma_*}(t,x), \qquad u(t,x) = \partial_x^2 \Phi^{\gamma_*}(t,x), \quad t \in [0, t_*],$$

and extend them as to satisfy the assumptions of Theorem 2 for $t \in (t_*, 1]$. Theorem 3 is proved by letting $t_* = t_*(\varepsilon) \to 1$ as $\varepsilon \to 0$. We prove in the next section that this choice is optimal: no pair of functions satisfying the hypotheses of Theorem 2 can achieve a value larger than $\inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma)$.

4. Optimality and stochastic control. In this section, we show that the value given by the extended variational principle of Section 2.2 is the largest achievable by IAMP algorithms of the form introduced in Section 3.2.

THEOREM 4. *For $u, v : [0, 1] \times \mathbb{R} \to \mathbb{R}$ satisfying Assumption 1, let $\boldsymbol{m}^\ell = f_\ell(\boldsymbol{z}^0, \dots, \boldsymbol{z}^\ell)$ be the output of the message passing algorithm (3.2) with nonlinearity given by (3.5). Then*

$$\lim_{\delta\to 0^+} \text{p-lim}_{N\to\infty} \frac{1}{N} H_N\big(\boldsymbol{m}^{\lfloor \delta^{-1} \rfloor}\big) \le \inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma).$$

(*In the above, the inner limit in $N \to \infty$ is a nonrandom quantity.*)

REMARK 4.1. Let us emphasize that optimality within the broader class of AMP algorithms defined in equation (3.2) is still an open problem. At the same time, the form of IAMP algorithms is fairly constrained by the objective of obtaining a continuous limit of state evolution (as given by the stochastic process of equation (2.2)). This in turn is motivated by the objective of mimicking the structure of the Parisi formula.

The proof of the above theorem is deferred to Section 7. Here, we outline the basic strategy, which formulates the optimality question as a stochastic optimal control problem.

We prove in Proposition 5.4 below that the left-hand side in the inequality of Theorem 4 is equal to

$$(4.1)\qquad \mathscr{E}(u, v) \equiv \int_0^1 \xi''(t)\mathbb{E}\big[u(t, X_t)\big]\,\mathrm{d}t,$$

where (X_t) solves the SDE (2.2). The proof crucially relies on state evolution and the previous choice of the nonlinearities f_ℓ. We next analyze the variational problem consisting in maximizing the objective value (4.1) given the constraints $\mathbb{E}[M_t^2] = t$ for all $t \in [0, 1]$ and $M_1 \in (-1, 1)$ over u and v satisfying Assumption 1. (We recall that $M_t = \int_0^t \sqrt{\xi''(s)}u(s, X_s)\,\mathrm{d}B_s$.)

For $s \le t$, we define the space of admissible controls $D[s, t]$ on the interval $[s, t]$ as the collection of all stochastic processes $(u_r)_{r\in[s,t]}$, which are progressively measurable with respect to the filtration of the Brownian motion $(B_r)_{r\in[s,t]}$ and such that

$$\mathbb{E}\int_s^t \xi''(r)u_r^2\,\mathrm{d}r < +\infty.$$

We are then led to consider the stochastic control problem

$$(4.2)\qquad \begin{aligned} \mathsf{VAL} \equiv \sup_{u\in D[0,1]} \mathbb{E}\Big[\int_0^1 \xi''(s)u_s\,\mathrm{d}s\Big] \\ \text{s.t. } \mathbb{E}\big[\big(M_t^u\big)^2\big] = t \quad \forall t \in [0, 1], \text{ and } M_1^u \in (-1, 1) \text{ a.s.,} \end{aligned}$$

with $M_t^u \equiv \int_0^t \sqrt{\xi''(s)}u_s\,\mathrm{d}B_s$.

Note that $D[0,1]$ is a larger space of controls than the one arising from the original algorithm; cf. equations (2.2), (2.3). Indeed, for any choice of the drift v, the process $(u(t,X_t))_{t\in[0,1]}$ is in $D[0,1]$, and hence can be encoded in the choice of a stochastic process $(u_t)_{t\in[0,1]} \in D[0,1]$. The proof of Theorem 4 consists in showing $\mathsf{VAL} \le \inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma)$. We achieve this by writing the Lagrangian form of the above constrained optimization problem with respect to the equality constraint $\mathbb{E}[(M_t^u)^2] = t$ for all t. We define the space of piecewise constant, or simple, functions

$$\mathsf{SF}_+ \equiv \left\{ g = \sum_{i=1}^m a_i \mathbb{I}_{[t_{i-1},t_i)} : 0 = t_0 < t_1 < \cdots < t_m = 1, a_i \ge 0, m \in \mathbb{N} \right\}. \tag{4.3}$$

For $\gamma \in \mathsf{SF}_+$, we let $\nu(t) = \int_t^1 \xi''(s)\gamma(s)\,\mathrm{d}s$ and consider the following bivariate function $\mathcal{J}_\gamma : [0,1] \times (-1,1) \to \mathbb{R}$ defined by

$$\begin{aligned} \mathcal{J}_\gamma(t,z) \equiv \sup_{u\in D[t,1]} \mathbb{E}\Big[\int_t^1 \xi''(s) u_s \,\mathrm{d}s + \frac{1}{2}\int_t^1 \nu(s)\big(\xi''(s)u_s^2 - 1\big)\,\mathrm{d}s \Big], \\ \text{s.t. } z + \int_t^1 \sqrt{\xi''(s)} u_s \,\mathrm{d}B_s \in (-1,1) \text{ a.s.}, \end{aligned} \tag{4.4}$$

We claim that the following upper bound holds:

$$\mathsf{VAL} \le \mathcal{J}_\gamma(0,0). \tag{4.5}$$

Indeed, we have by integration by parts

$$\int_0^1 \nu(s)\big(\xi''(s)u_s^2 - 1\big)\,\mathrm{d}s = \int_0^1 \xi''(t)\gamma(t) \Big(\int_0^t \xi''(s) u_s^2 \,\mathrm{d}s - t \Big)\,\mathrm{d}t.$$

Since $\mathbb{E}[(M_t^u)^2] = \mathbb{E}\int_0^t \xi''(s)u_s^2\,\mathrm{d}s$, the second term in the definition of $\mathcal{J}_\gamma(0,0)$ equation (4.4) vanishes for any control (u_s) that satisfies the constraints of the problem (4.2), thus proving equation (4.5). In other words, $\mathcal{J}_\gamma(0,0)$ is the Lagrangian associated to the optimization problem (4.2) with dual variable $\frac{1}{2}\xi''\gamma$.

We are now left with the task of relating $\mathcal{J}_\gamma(0,0)$ to the Parisi functional $\mathsf{P}(\gamma)$:

PROPOSITION 4.1. *For $\gamma \in \mathsf{SF}_+$, $\mathcal{J}_\gamma(0,0) = \mathsf{P}(\gamma)$.*

REMARK 4.2. Although we are ultimately only interested in the value of $\mathcal{J}_\gamma$ at $(0,0)$, we will see shortly that defining it for all arguments (t,z) allows us to define a dynamic programming equation, which can be solved analytically. This is at the heart of the proof of Proposition 4.1.

The bound (4.5) then implies

$$\mathsf{VAL} \le \inf_{\gamma\in\mathsf{SF}_+} \mathsf{P}(\gamma).$$

Since any function in the class $\mathscr{L}$ can be approximated with a piecewise constant function with respect to the modified L^1 metric $\|\cdot\|_{1,\xi''}$, equation (2.7), and $\gamma \mapsto \mathsf{P}(\gamma)$ is continuous in this metric (see Section 6), the above infimum is no larger than $\inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma)$.

We now sketch the first steps in establishing Proposition 4.1, relegating a full proof to Section 7. The value function (4.4) can be (formally) computed by dynamic programming where we search for solutions to the equation

$$\begin{aligned} V(t,z) = \sup_{u\in D[t,\theta]} \mathbb{E}\Big[\int_t^\theta \xi''(s) u_s \,\mathrm{d}s + \frac{1}{2}\int_t^\theta \nu(s)\big(\xi''(s)u_s^2 - 1\big)\,\mathrm{d}s \\ + V\Big(\theta, z + \int_t^\theta \sqrt{\xi''(s)} u_s \,\mathrm{d}B_s\Big) \Big], \end{aligned} \tag{4.6}$$

valid for all $\theta \in [t,1]$ and $z \in (-1,1)$, with terminal condition $V(1,z)=0$ for $|z|<1$. The associated Hamilton–Jacobi–Bellman (HJB) equation, which can be formally obtained from (4.6) by letting $\theta \to t^+$ and applying Itô's formula, is

$$\begin{aligned}&\partial_t V(t,z) + \xi''(t)\sup_{\lambda\in\mathbb{R}}\Big\{\lambda + \frac{\lambda^2}{2}\big(\nu(t)+\partial_z^2 V(t,z)\big)\Big\} - \frac{1}{2}\nu(t)\\ &\quad = 0, \quad (t,z)\in[0,1)\times(-1,1),\\ &V(1,z)=0, \quad z\in(-1,1).\end{aligned} \tag{4.7}$$

Note that it is a priori unclear whether equation (4.6) and equation (4.7) have (classical) solutions and whether they are at all related to equation (4.4): $\mathcal{J}_\gamma$ is not known a priori to be smooth, hence the above derivation is not rigorously justified; it is not even clear that the right-hand side of (4.6) is measurable. To circumvent this issue, we will guess a solution V to (4.7) and use the so-called *verification argument* to certify that the guessed solution is equal to $\mathcal{J}_\gamma$ as defined in equation (4.4). En route, we establish that the optimal control process in the stochastic control problem (4.4) for $t=z=0$ is given by

$$u_t^* = \partial_x^2 \Phi^\gamma(t, X_t),$$

where (X_t) solves the SDE (2.2) with drift $v(t,x)=\xi''(t)\gamma(t)\partial_x\Phi^\gamma(t,x)$ and Φ^γ solves the Parisi PDE. This confirms in hindsight our choice of the functions u and v used in the message passing algorithm, equation (3.6). (See also proof of Theorem 3.)

5. Proof of Theorem 2.

5.1. *The scaling limit.* Consider the message passing iteration (3.2) with nonlinearities f_ℓ given by (3.5) and iterate sequence $(\boldsymbol{z}^0, \boldsymbol{z}^1, \dots)$ starting from $\boldsymbol{z}^0 = \boldsymbol{0}$. We denote by $(\boldsymbol{x}^0, \boldsymbol{x}^1, \dots)$ and $(\boldsymbol{m}^0, \boldsymbol{m}^1, \dots)$ the two auxiliary sequences obtained from the finite difference equation (3.3) and the relation (3.4), respectively. We will define the distributional limit of the message passing iteration for fixed $\delta > 0$ and $N \to \infty$, and indicate these time-discretized variables using a superscript δ. We remind the reader that the nonlinearities f_ℓ act on vectors coordinatewise. It is clear from equation (3.4) that f_ℓ is Lipschitz continuous for each ℓ, with a Lipschitz constant depending on ℓ and C (the uniform bound on u), and therefore the conclusion of Proposition 3.1 about state evolution applies. Let $(Z_\ell^\delta)_{\ell\ge 0}$ be the limit of the sequence $(\boldsymbol{z}^0, \boldsymbol{z}^1, \dots)$. Since u, v are uniformly Lipschitz in x, then $(\boldsymbol{x}^0, \boldsymbol{x}^1, \dots)$ and $(\boldsymbol{m}^0, \boldsymbol{m}^1, \dots)$ converge as well in the sense of Proposition 3.1 to stochastic processes $(X_\ell^\delta)_{\ell\ge 0}$ and $(M_\ell^\delta)_{\ell \ge 0}$, defined respectively via the formulas (3.3) and (3.4) by replacing every occurrence of z^j by Z_j^δ. Define, for all $\ell \ge 0$,

$$q_\ell^\delta \equiv \mathbb{E}\big[(M_\ell^\delta)^2\big].$$

LEMMA 5.1. *The sequence $(Z_\ell^\delta)_{\ell\ge 0}$ is a Gaussian process starting at $Z_0^\delta = 0$. Its increments $\Delta_\ell^\delta \equiv Z_\ell^\delta - Z_{\ell-1}^\delta$ are independent, have zero mean and variance*

$$\begin{aligned}&\mathbb{E}\big[(\Delta_1^\delta)^2\big] = \xi'(\delta),\\ &\mathbb{E}\big[(\Delta_\ell^\delta)^2\big] = \xi'(q_{\ell-1}^\delta) - \xi'(q_{\ell-2}^\delta) \quad \textit{for all } \ell \ge 2.\end{aligned}$$

Furthermore, $(M_\ell^\delta)_{\ell\ge 0}$ is a martingale with respect to the filtration $(\mathcal{F}_\ell = \sigma(Z_0^\delta, \dots, Z_\ell^\delta))_{\ell \ge 0}$, and $M_0^\delta = \sqrt{\delta}$.

PROOF. We proceed by induction. Since $\boldsymbol{z}^0 = \mathbf{0}$ and $\boldsymbol{m}^0 = \sqrt{\delta}\mathbf{1}$, we have $Z_0^\delta = 0$ and $M_0^\delta = \sqrt{\delta}$. We also have for all $j \geq 1$, $\mathbb{E}[Z_1^\delta Z_j^\delta] = \xi'(\mathbb{E}[M_0^\delta M_{j-1}^\delta]) = \xi'(\delta)$. So $\mathbb{E}[\Delta_1^\delta \Delta_2^\delta] = \mathbb{E}[Z_2^\delta Z_1^\delta] - \mathbb{E}[(Z_1^\delta)^2] = 0$, and $\mathbb{E}[(\Delta_1^\delta)^2] = \xi'(\delta)$. Now we assume that the increments $(\Delta_j^\delta)_{j\leq\ell}$ are independent. This implies that $(M_j^\delta)_{j\leq\ell}$ is a martingale. Appealing to the state evolution recursion,

$$\begin{aligned}\mathbb{E}[\Delta_{\ell+1}^\delta \Delta_1^\delta] &= \mathbb{E}[Z_{\ell+1}^\delta Z_1^\delta] - \mathbb{E}[Z_\ell^\delta Z_1^\delta] \\ &= \xi'(\mathbb{E}[M_\ell^\delta M_0^\delta]) - \xi'(\mathbb{E}[M_{\ell-1}^\delta M_0^\delta]) \\ &= 0,\end{aligned}$$

since $M_0^\delta = \sqrt{\delta}$ and $\mathbb{E}[M_\ell^\delta] = \mathbb{E}[M_{\ell-1}^\delta]$. For $2 \leq j \leq \ell$,

$$\begin{aligned}\mathbb{E}[\Delta_{\ell+1}^\delta \Delta_j^\delta] &= \xi'(\mathbb{E}[M_\ell^\delta M_{j-1}^\delta]) - \xi'(\mathbb{E}[M_{\ell-1}^\delta M_{j-1}^\delta]) \\ &\quad - \xi'(\mathbb{E}[M_\ell^\delta M_{j-2}^\delta]) + \xi'(\mathbb{E}[M_{\ell-1}^\delta M_{j-2}^\delta]) \\ &= 0\end{aligned}$$

since $(M_j^\delta)_{j\leq\ell}$ has independent increments. So $\Delta_{\ell+1}^\delta$ is independent from $(\Delta_j^\delta)_{j\leq\ell}$. This ends the induction argument. The variance identity follows straightforwardly. □

We define the functions u^δ by the relations

$$\begin{aligned} u_0^\delta &\equiv \left(\frac{\delta}{\xi'(\delta)}\right)^{1/2}, \\ u_\ell^\delta(x) &\equiv \frac{u(\ell\delta; x)}{\Sigma_\ell^\delta} \quad \text{for all } \ell \geq 1, \\ &\text{with } (\Sigma_\ell^\delta)^2 = \delta^{-1}(\xi'(q_\ell^\delta) - \xi'(q_{\ell-1}^\delta))\mathbb{E}[u(\ell\delta; X_\ell^\delta)^2]. \end{aligned} \tag{5.1}$$

LEMMA 5.2. *Assume u_ℓ^δ takes the form (5.1) for all $\ell \geq 0$. Then $q_\ell^\delta = \mathbb{E}[(M_\ell^\delta)^2] = (\ell + 1)\delta$ for all $\ell \geq 0$.*

PROOF. We proceed by induction over ℓ, the base case being trivial. First, notice that u_ℓ^δ is well defined since $\Sigma_\ell^\delta > 0$ for all ℓ. Indeed, ξ' is strictly increasing and, by the induction hypothesis $q_\delta^\ell > q_\delta^{\ell-1}$, and X_ℓ^δ is a nondegenerate Gaussian, whence $\mathbb{E}[u(\ell\delta; X_\ell^\delta)^2] > 0$ (because by assumption u is nonvanishing). We have $q_0^\delta = \mathbb{E}[(M_0^\delta)^2] = \delta$. Let $\ell \geq 1$. Since Z^δ has independent increments, equation (3.4) implies

$$\begin{aligned}\mathbb{E}[(M_\ell^\delta - M_0^\delta)^2] &= \sum_{j=0}^{\ell-1} \mathbb{E}[u_j^\delta(X_j^\delta)^2] \cdot \mathbb{E}[(\Delta_{j+1}^\delta)^2] \\ &= \mathbb{E}[u_0^\delta(X_0^\delta)^2] \cdot \xi'(\delta) + \sum_{j=1}^{\ell-1} \mathbb{E}[u_j^\delta(X_j^\delta)^2] \cdot (\xi'(q_j^\delta) - \xi'(q_{j-1}^\delta)) \\ &= \delta + (\ell - 1)\delta.\end{aligned}$$

The second line follows from Lemma 5.1, the last line follows from (5.1). The fact that M^δ is a martingale yields the desired result. □

Next, we show that under condition (5.2), $(Z_j^\delta, X_j^\delta, M_j^\delta)_{0\leq j\leq\ell}$ converge to continuous-time stochastic processes $(Z_t, X_t, M_t)_{t\in[0,1]}$ on the interval $[0, 1]$ as $\delta \to 0$, $\ell \to \infty$

and $\ell \le \delta^{-1}$, with $Z_t \equiv \int_0^t \sqrt{\xi''(s)}\,\mathrm{d}B_s$, X_t is the solution to the SDE (2.2) and $M_t \equiv \int_0^t \sqrt{\xi''(s)}u(s, X_s)\,\mathrm{d}B_s$.

PROPOSITION 5.3. *Assume*

$$\mathbb{E}\big[M_t^2\big] = t \quad \textit{for all } t \in [0, 1]. \tag{5.2}$$

Then there exists a coupling between the random variables $\{(Z_\ell^\delta, X_\ell^\delta, M_\ell^\delta)\}_{\ell \ge 0}$ *and the stochastic process* $\{(Z_t, X_t, M_t)\}_{t \ge 0}$ *such that the following holds. There exists* $\delta_0 > 0$ *and a constant* $C > 0$ *such that for all* $\delta \le \delta_0$ *and* $\ell \le \delta^{-1}$,

$$\max_{1 \le j \le \ell} \mathbb{E}\big[\big|X_j^\delta - X_{\delta j}\big|^2\big] \le C\delta, \tag{5.3}$$

$$\max_{1 \le j \le \ell} \mathbb{E}\big[\big|M_j^\delta - M_{\delta j}\big|^2\big] \le C\delta. \tag{5.4}$$

PROOF. Let $(B_t)_{t \in [0,1]}$ be a standard Brownian motion. We couple the increments of Z^δ with (B_t) via the relation

$$Z_\ell^\delta - Z_{\ell-1}^\delta = \int_{\delta(\ell-1)}^{\delta\ell} \sqrt{\xi''(s)}\,\mathrm{d}B_s \quad \text{for all } \ell \ge 1. \tag{5.5}$$

Itô's isometry implies $\mathbb{E}[(Z_\ell^\delta - Z_{\ell-1}^\delta)^2] = \xi'(\delta\ell) - \xi'(\delta(\ell-1))$. By Lemma 5.2, this is in accordance with the characterization of the law of Z^δ obtained in Lemma 5.1. Moreover, we have $Z_\ell^\delta = Z_{\delta\ell}$ for all $\ell \ge 0$. We now show (5.3). Let $\Delta_j^X = X_j^\delta - X_{\delta j}$. Using (2.2) and (3.3), we have

$$\begin{aligned}
\Delta_j^X - \Delta_{j-1}^X &= \int_{(j-1)\delta}^{j\delta} \big(v\big((j-1)\delta; X_j^\delta\big) - v(t; X_t)\big)\,\mathrm{d}t + Z_j^\delta - Z_{j-1}^\delta - \int_{\delta(j-1)}^{\delta j} \sqrt{\xi''(s)}\,\mathrm{d}B_s \\
&= \int_{(j-1)\delta}^{j\delta} \big(v\big((j-1)\delta; X_j^\delta\big) - v(t; X_t)\big)\,\mathrm{d}t \\
&= \int_{(j-1)\delta}^{j\delta} \big(v\big((j-1)\delta; X_j^\delta\big) - v\big((j-1)\delta; X_t\big)\big)\,\mathrm{d}t \\
&\quad + \int_{(j-1)\delta}^{j\delta} \big(v\big((j-1)\delta; X_t\big) - v(t; X_t)\big)\,\mathrm{d}t.
\end{aligned}$$

The first term is the above equation is bounded in absolute value by $C\int_{(j-1)\delta}^{j\delta} |X_j^\delta - X_t|\,\mathrm{d}t$ since v Lipschitz in space uniformly in time. As for the second term,

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

where the last inequality follows from the property of bounded strong total variation of v (see Definition 2.1). Putting to the two bounds together, summing over j, and using $\Delta_0^X = 0$, we have

$$|\Delta_\ell^X| \le \sum_{j=1}^{\ell} |\Delta_j^X - \Delta_{j-1}^X| \le C \sum_{j=1}^{\ell} \int_{(j-1)\delta}^{j\delta} |X_j^\delta - X_t| \, \mathrm{d}t + C\delta.$$

Squaring and taking expectations,

$$\begin{aligned}\mathbb{E}[(\Delta_\ell^X)^2] &\le 2C^2 \mathbb{E}\left(\sum_{j=1}^{\ell} \int_{(j-1)\delta}^{j\delta} |X_j^\delta - X_t| \, \mathrm{d}t \right)^2 + 2C^2\delta^2 \\ &\le 2C^2 \ell\delta \sum_{j=1}^{\ell} \int_{(j-1)\delta}^{j\delta} \mathbb{E}|X_j^\delta - X_t|^2 \, \mathrm{d}t + 2C^2\delta^2.\end{aligned}$$

Furthermore, $\mathbb{E}|X_j^\delta - X_t|^2 \le 2\mathbb{E}|X_j^\delta - X_{\delta j}|^2 + 2\mathbb{E}|X_{\delta j} - X_t|^2$. It is easy to show that $\mathbb{E}|X_t - X_s|^2 \le C|t-s|$ for all t, s. Therefore,

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

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

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

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

Next, we show (5.4). Using the relation (5.5) we have

$$\begin{aligned}\mathbb{E}[(M_\ell^\delta - M_{\delta\ell})^2] &= \mathbb{E}\left[\left(\sum_{j=0}^{\ell-1} u_j^\delta(X_j^\delta)(Z_{j+1}^\delta - Z_j^\delta) - \int_0^{\delta\ell} \sqrt{\xi''(t)} u(t, X_t) \, \mathrm{d}B_t \right)^2 \right] \\ &= \mathbb{E}\left[\left(\sum_{j=0}^{\ell-1} \int_{j\delta}^{(j+1)\delta} (u_j^\delta(X_j^\delta) - u(t, X_t)) \sqrt{\xi''(t)} \, \mathrm{d}B_t \right)^2 \right] \qquad (5.6) \\ &= \sum_{j=0}^{\ell-1} \int_{j\delta}^{(j+1)\delta} \mathbb{E}[(u_j^\delta(X_j^\delta) - u(t, X_t))^2] \xi''(t) \, \mathrm{d}t.\end{aligned}$$

Recall that $u_j^\delta(x) = u(\delta j; x)/\Sigma_j^\delta$ for $j \ge 1$ where Σ_j^δ is given in equation (5.1). Since we have $q_j^\delta = \delta(j+1)$, the formula for Σ_j^δ reduces to

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

Let us first show the bound

$$|(\Sigma_j^\delta)^2 - 1| \le C\sqrt{\delta} \qquad (5.7)$$

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

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

Additionally, since u is Lipschitz in space (and bounded), we use the bound equation (5.3) to obtain

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

Now, since $\mathbb{E}[M_t^2] = t$ for all $t \in [0, 1]$ and $t \mapsto u(t, X_t)$ is a.s. continuous, we have by Lebesgue's differentiation theorem, for all $t \in [0, 1]$,

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

and hence $|(\Sigma_j^\delta)^2 - 1| \le C\sqrt{\delta}$ for δ smaller than some absolute constant. This implies the bound $|u_j^\delta(X_j^\delta) - u(\delta j; X_j^\delta)| \le C|\frac{1}{\Sigma_j^\delta} - 1| \le C\sqrt{\delta}$. Now, going back to equation (5.6), we have

$$\begin{aligned}\mathbb{E}\big[\big(M_\ell^\delta - M_{\delta\ell}\big)^2\big] \le\, & 2\sum_{j=0}^{\ell-1}\int_{j\delta}^{(j+1)\delta} \mathbb{E}\big[\big(u_j^\delta(X_j^\delta) - u\big(\delta j; X_j^\delta\big)\big)^2\big]\xi''(t)\,\mathrm{d}t \\ & + 2\sum_{j=0}^{\ell-1}\int_{j\delta}^{(j+1)\delta} \mathbb{E}\big[\big(u\big(\delta j; X_j^\delta\big) - u(t, X_t)\big)^2\big]\xi''(t)\,\mathrm{d}t\end{aligned}$$

The first term is bounded by $C\ell\delta^2 \le C\delta$. As for the second term,

$$\begin{aligned}&\sum_{j=0}^{\ell-1}\int_{j\delta}^{(j+1)\delta} \mathbb{E}\big[\big(u\big(\delta j; X_j^\delta\big) - u(t, X_t)\big)^2\big]\xi''(t)\,\mathrm{d}t \\ &\quad\le C\sum_{j=0}^{\ell-1}\int_{j\delta}^{(j+1)\delta} \mathbb{E}\big[\big(u\big(\delta j; X_j^\delta\big) - u(\delta j, X_{\delta j})\big)^2\big]\,\mathrm{d}t \\ &\qquad + C\sum_{j=0}^{\ell-1}\int_{j\delta}^{(j+1)\delta} \mathbb{E}\big[\big(u(\delta j; X_{\delta j}) - u(\delta j, X_t)\big)^2\big]\,\mathrm{d}t \\ &\qquad + C\sum_{j=0}^{\ell-1}\int_{j\delta}^{(j+1)\delta} \mathbb{E}\big[\big(u(\delta j; X_t) - u(t, X_t)\big)^2\big]\,\mathrm{d}t \\ &\quad\equiv I + II + III.\end{aligned}$$

Since u is Lipschitz in space, the error bound equation (5.3) implies $I \le C\ell\delta^2$. Further, we have the continuity bound $\mathbb{E}[|X_t - X_s|^2] \le C|t - s|$; therefore, $II \le C\ell\delta^2$. Finally, since u has bounded strong total variation (Definition 2.1) and $\ell\delta \le 1$, it follows that $III \le C\delta$. Putting the pieces together, we obtain

$$\mathbb{E}\big[\big(M_\ell^\delta - M_{\delta\ell}\big)^2\big] \le C\delta,$$

which is the desired bound. □

5.2. *Value achieved by the algorithm.* Throughout this section, we denote by $\langle \boldsymbol{A}, \boldsymbol{B}\rangle_N$ the normalized scalar product between tensors $\boldsymbol{A}, \boldsymbol{B} \in (\mathbb{R}^N)^{\otimes k}$. Namely, $\langle \boldsymbol{A}, \boldsymbol{B}\rangle_N = \sum_{i_1,\dots,i_k\le N} A_{i_1,\dots,i_k}B_{i_1,\dots,i_k}/N$.

PROPOSITION 5.4. *There exist $\delta_0 > 0$ and a constant $C > 0$ such that for all $\delta \le \delta_0$ and $\ell \le \delta^{-1}$,*

$$\left|\text{p-lim}_{N\to\infty}\frac{H_N(\boldsymbol{m}^\ell)}{N} - \int_0^{\ell\delta}\xi''(t)\mathbb{E}\big[u(t, X_t)\big]\,\mathrm{d}t\right| \le C\sqrt{\delta}.$$

PROOF. In order to compute $H_N(\boldsymbol{m}^\ell)$ for large N, we evaluate the differences $H_N(\boldsymbol{m}^k) - H_N(\boldsymbol{m}^{k-1})$ for $1 \le k \le \ell$ and sum them. We have

$$N^{-1}(H_N(\boldsymbol{m}^k) - H_N(\boldsymbol{m}^{k-1})) = \sum_p \frac{c_p}{p!}\langle \boldsymbol{W}^{(p)}, (\boldsymbol{m}^k)^{\otimes p} - (\boldsymbol{m}^{k-1})^{\otimes p}\rangle_N,$$

where the above inner product is of tensors of order p, normalized by N. We want to approximate the term

$$A_p^k \equiv \langle \boldsymbol{W}^{(p)}, (\boldsymbol{m}^k)^{\otimes p} - (\boldsymbol{m}^{k-1})^{\otimes p}\rangle_N$$

with

$$B_p^k \equiv \left\langle \boldsymbol{W}^{(p)}, \frac{p}{2}((\boldsymbol{m}^k)^{\otimes(p-1)} + (\boldsymbol{m}^{k-1})^{\otimes(p-1)}) \otimes (\boldsymbol{m}^k - \boldsymbol{m}^{k-1})\right\rangle_N,$$

which captures the first two the terms in the binomial expansion of A_p^k in $\boldsymbol{m}^k - \boldsymbol{m}^{k-1}$.

The result follows from the next lemma.

LEMMA 5.5. *There exist $\delta_0 > 0$ and a constant $C > 0$ such that for all $\delta \le \delta_0$ and $\ell \le \delta^{-1}$,*

$$\left|\text{p-lim}_{N\to\infty} \sum_{k=1}^{\ell}\sum_{p\ge 2}\frac{c_p}{p!}B_p^k - \int_0^{\ell\delta}\xi''(t)\mathbb{E}[u(t,X_t)]\,\mathrm{d}t\right| \le C\sqrt{\delta}, \tag{5.8}$$

$$\text{and} \quad \left|\sum_{k=1}^{\ell}\sum_{p\ge 3}\frac{c_p}{p!}(A_p^k - B_p^k)\right| \le C\sqrt{\delta}, \tag{5.9}$$

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

Let us first complete the proof of Proposition 5.4. For $\ell \ge 1$, we have

$$\begin{aligned} N^{-1}(H_N(\boldsymbol{m}^\ell) - H_N(\boldsymbol{m}^0)) &= \sum_{k=1}^{\ell} N^{-1}(H_N(\boldsymbol{m}^k) - H_N(\boldsymbol{m}^{k-1})) \\ &= \sum_{k=1}^{\ell}\sum_p \frac{c_p}{p!}B_p^k + \sum_{k=0}^{\ell}\sum_p \frac{c_p}{p!}(A_p^k - B_p^k). \end{aligned}$$

Since $\boldsymbol{m}^0$ is nonrandom, $\text{p-lim}_N H_N(\boldsymbol{m}^0)/N = 0$, and Lemma 5.5 yields the desired result. □

PROOF OF LEMMA 5.5. We prove the two statements separately:

Proof of equation (5.8). We have

$$\begin{aligned} \sum_p \frac{c_p}{p!}B_p^k &= \frac{1}{2}\sum_p c_p\langle \boldsymbol{W}^{(p)}\{\boldsymbol{m}^k\}, \boldsymbol{m}^k - \boldsymbol{m}^{k-1}\rangle_N + \frac{1}{2}\sum_p c_p\langle \boldsymbol{W}^{(p)}\{\boldsymbol{m}^{k-1}\}, \boldsymbol{m}^k - \boldsymbol{m}^{k-1}\rangle_N \\ &\equiv \frac{1}{2}(S_{1,N} + S_{2,N}). \end{aligned}$$

By taking the scalar product of all the terms in iteration (3.2) with $\boldsymbol{m}^k - \boldsymbol{m}^{k-1}$, we see that

$$S_{1,N} = \langle \boldsymbol{z}^{k+1}, \boldsymbol{m}^k - \boldsymbol{m}^{k-1}\rangle_N + \sum_{j=0}^{k} d_{k,j}\langle \boldsymbol{m}^{j-1}, \boldsymbol{m}^k - \boldsymbol{m}^{k-1}\rangle_N,$$

$$S_{2,N} = \langle \boldsymbol{z}^{k}, \boldsymbol{m}^k - \boldsymbol{m}^{k-1}\rangle_N + \sum_{j=0}^{k-1} d_{k-1,j}\langle \boldsymbol{m}^{j-1}, \boldsymbol{m}^k - \boldsymbol{m}^{k-1}\rangle_N.$$

Taking N to infinity and invoking Proposition 3.1, $S_{1,N}$ and $S_{2,N}$ converge in probability to

$$\text{p-lim}_{N\to\infty} S_{1,N} = \mathbb{E}[Z^\delta_{k+1}(M^\delta_k - M^\delta_{k-1})] + \sum_{j=0}^{k} d_{k,j}\mathbb{E}[M^\delta_{j-1}(M^\delta_k - M^\delta_{k-1})],$$

$$\text{p-lim}_{N\to\infty} S_{2,N} = \mathbb{E}[Z^\delta_{k}(M^\delta_k - M^\delta_{k-1})] + \sum_{j=0}^{k-1} d_{k-1,j}\mathbb{E}[M^\delta_{j-1}(M^\delta_k - M^\delta_{k-1})],$$

respectively. Since M^δ is a martingale, the right-most terms in the above expressions vanish. Next, since Z^δ has independent increments, the left-most terms in the above expressions are equal, and we get

$$\begin{aligned}\frac{1}{2}(S_{1,N}+S_{2,N}) &= \mathbb{E}[Z^\delta_k(M^\delta_k - M^\delta_{k-1})]\\ &= \mathbb{E}[(Z^\delta_k - Z^\delta_{k-1})(M^\delta_k - M^\delta_{k-1})]\\ &= \mathbb{E}[u^\delta_{k-1}(X^\delta_{k-1})(Z^\delta_k - Z^\delta_{k-1})^2].\end{aligned}$$

Summing over $k \in \{1,\dots,\ell\}$, we obtain

$$\begin{aligned}\text{p-lim}_{N\to\infty}\sum_{k=1}^{\ell}\sum_p \frac{c_p}{p!}B^k_p &= \sum_{k=1}^{\ell}\mathbb{E}[u^\delta_{k-1}(X^\delta_{k-1})(Z^\delta_k - Z^\delta_{k-1})^2]\\ &= \sqrt{\delta\xi'(\delta)} + \sum_{k=2}^{\ell}\mathbb{E}[u^\delta_{k-1}(X^\delta_{k-1})](\xi'(q^\delta_{k-1}) - \xi'(q^\delta_{k-2}))\\ &= \sqrt{\delta\xi'(\delta)} + \sum_{k=2}^{\ell}\frac{\mathbb{E}[u(\delta(k-1);X^\delta_{k-1})]}{\Sigma^\delta_{k-1}}\big(\xi'(\delta k) - \xi'\big(\delta(k-1)\big)\big).\end{aligned}$$

Since $|\frac{1}{\Sigma^\delta_k} - 1| \le C\sqrt{\delta}$ (this is a consequence of equation (5.7)) and $\xi'(\delta) \le \xi''(1)\delta$, the above is equal to

$$\begin{aligned}&\sum_{k=2}^{\ell}\mathbb{E}\big[u\big(\delta(k-1);X^\delta_{k-1}\big)\big]\big(\xi'(\delta k) - \xi'\big(\delta(k-1)\big)\big) + O(\sqrt{\delta})\\ &= \sum_{k=2}^{\ell}\mathbb{E}\big[u\big(\delta(k-1);X^\delta_{k-1}\big)\big]\xi''\big(\delta(k-1)\big)\delta + O(\sqrt{\delta})\\ &= \int_0^{\ell\delta}\mathbb{E}[u(t,X_t)]\xi''(t)\,\mathrm{d}t + O(\sqrt{\delta}).\end{aligned}$$

The last equality is obtained by invoking the discretization error bound equation (5.3) of Proposition 5.3, and using the regularity properties of u, exactly as done in the proof of equation (5.4).

Proof of equation (5.9). We fix k and write $\boldsymbol{m} = \boldsymbol{m}^{k-1}$, $\boldsymbol{m}' = \boldsymbol{m}^k$ and $\boldsymbol{\alpha} = \boldsymbol{m}' - \boldsymbol{m}$. Since the tensors $\boldsymbol{W}^{(p)}$ are symmetric the approximation error $A^k_p - B^k_p$ is

$$\begin{aligned}A^k_p - B^k_p &= \sum_{j=3}^{p}\binom{p}{j}\langle \boldsymbol{W}^{(p)}, \boldsymbol{m}^{\otimes(p-j)}\otimes\boldsymbol{\alpha}^{\otimes j}\rangle_N\\ &\quad - \sum_{j=2}^{p-1}\binom{p-1}{j}\Big\langle \boldsymbol{W}^{(p)}, \frac{p}{2}\boldsymbol{m}^{\otimes(p-j-1)}\otimes\boldsymbol{\alpha}^{\otimes(j+1)}\Big\rangle_N\end{aligned} \tag{5.10}$$

$$= \sum_{j=3}^{p} \binom{p}{j} (1 - j/2) \langle \boldsymbol{W}^{(p)}, \boldsymbol{m}^{\otimes(p-j)} \otimes \boldsymbol{\alpha}^{\otimes j} \rangle_N.$$

We crudely bound the above inner product as

$$\left| \langle \boldsymbol{W}^{(p)}, \boldsymbol{m}^{\otimes(p-j)} \otimes \boldsymbol{\alpha}^{\otimes j} \rangle_N \right| \le \frac{1}{N} \| \boldsymbol{W}^{(p)} \|_{\mathrm{op}} \cdot \|\boldsymbol{m}\|^{p-j} \cdot \|\boldsymbol{\alpha}\|^{j}.$$

Here, $\|\cdot\|_{\mathrm{op}}$ is the operator (or injective) norm of symmetric tensors in the ℓ_2 norm: for a symmetric tensor $\boldsymbol{T} \in (\mathbb{R}^N)^{\otimes k}$,

$$\|\boldsymbol{T}\|_{\mathrm{op}} \equiv \sup_{\|\boldsymbol{u}_i\| \le 1} \langle \boldsymbol{T}, \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k \rangle = \sup_{\|\boldsymbol{u}\| \le 1} \langle \boldsymbol{T}, \boldsymbol{u}^{\otimes k} \rangle.$$

(The second equality is specific to symmetric tensors [49].) The operator norm of symmetric Gaussian tensors is well understood. In particular, it is known [3, 14] that there exists a p-dependent constant E_p, known as the ground state energy of the spherical p-spin model, such that $\text{p-lim}_{N\to\infty} N^{(p-2)/2} \cdot \|\boldsymbol{W}^{(p)}\|_{\mathrm{op}} = E_p$. A bound on E_p combined with a simple concentration bound [42], Lemma 2, yields

$$\mathbb{P}\big(N^{(p-2)/2} \|\boldsymbol{W}^{(p)}\|_{\mathrm{op}} \ge p!\sqrt{p}\big) \le e^{-Np/8}. \tag{5.11}$$

Furthermore, by Proposition 3.1,

$$\text{p-lim}_{N\to\infty} \|\boldsymbol{m}\|^2/N = \mathbb{E}\big[(M_{k-1}^{\delta})^2\big] = k\delta$$

$$\text{and} \quad \text{p-lim}_{N\to\infty} \|\boldsymbol{\alpha}\|^2/N = \mathbb{E}\big[(M_k^{\delta} - M_{k-1}^{\delta})^2\big] = \delta.$$

Combining the above bounds, and letting $K_p = p!\sqrt{p}$, we get

$$\left| \langle \boldsymbol{W}^{(p)}, \boldsymbol{m}^{\otimes(p-j)} \otimes \boldsymbol{\alpha}^{\otimes j} \rangle_N \right| \le K_p (k\delta)^{(p-j)/2} \delta^{j/2},$$

for all p, with probability tending to one as $N \to \infty$. Bounding $k\delta$ by 1, and plugging back into expression (5.10), we obtain

$$\left| A_p^k - B_p^k \right| \le K_p \sum_{j=3}^{p} \binom{p}{j} |1 - j/2| \delta^{j/2},$$

with probability tending to one as $N \to \infty$. Summing over p and k, we obtain

$$\begin{aligned}
\sum_{k=1}^{\ell} \sum_{p\ge3} \frac{c_p}{p!} \left| A_p^k - B_p^k \right| &\le \ell \sum_{p\ge3} \frac{c_p}{p!} K_p \sum_{j=3}^{p} \binom{p}{j} |1 - j/2| \delta^{j/2} \\
&\le \sum_{p\ge3} \frac{c_p}{p!} K_p \sum_{j=3}^{p} \binom{p}{j} j \delta^{(j-2)/2} \\
&\le \sum_{p\ge3} \frac{c_p}{p!} K_p p^3 \sqrt{\delta} e^{p\sqrt{\delta}} \\
&\le \sqrt{\delta} \sum_{p\ge3} c_p p^4 e^{p\sqrt{\delta}}
\end{aligned}$$

with probability tending to one as $N \to \infty$. By assumption $|c_p| \le c_* \alpha^p$ for some $\alpha < 1$ (since $\xi(t) < \infty$ for some $t > 1$). Therefore, the sum is finite for ε and δ small enough, and the overall upper bound is $C\sqrt{\delta}$. This concludes the proof. □

5.3. *Rounding and proof of Theorem 2*. The algorithm described in the previous section returns a sequence of vectors $\boldsymbol{m}^\ell \in \mathbb{R}^N$. In this section, we describe how to round these in order to construct a feasible solution $\boldsymbol{\sigma}^{\rm alg} \in \{-1,+1\}^N$, and bound the rounding error. We note in passing that the related question of optimizing $H_N(\boldsymbol{m})$ over the solid hypercube $[-1,+1]^N$ turns out to be essentially equivalent to the one studied in [43].

Fix $t_* \in [0,1]$, and let $\ell_* = \lfloor t_*/\delta \rfloor$. The rounding procedure consists in two steps: (i) Threshold the coordinates of $\boldsymbol{m}^{\ell_*}$ to construct a vector $\hat{\boldsymbol{m}} \in [-1,+1]^N$; (ii) Round the entries of $\hat{\boldsymbol{m}}$ in a sequential fashion, to obtain a vector $\boldsymbol{\sigma}^{\rm alg} \in \{-1,+1\}^N$.

5.3.1. *Thresholding*. We define $\hat{\boldsymbol{m}} \in [-1,+1]^N$ by thresholding entrywise $\boldsymbol{m}^{\ell_*}$:

$$\hat{m}_i \equiv \begin{cases} m_i^{\ell_*} & \text{if } |m_i^{\ell_*}| \le 1, \\ \operatorname{sign}(m_i^{\ell_*}) & \text{otherwise,} \end{cases}$$

LEMMA 5.6. *There exist constants $C, \varepsilon_0 > 0$ such that, with high probability*

$$\sup\{\|\nabla H_N(\boldsymbol{x})\|_N : \|\boldsymbol{x}\|_N \le 1+\varepsilon_0\} \le C. \tag{5.12}$$

PROOF. Denoting by $B_N(\varepsilon_0)$ the supremum on the left-hand side of equation (5.12), we have

$$\begin{aligned} B_N(\varepsilon_0) &= \sup_{\|\boldsymbol{y}\|_N \le 1, \|\boldsymbol{x}\|_N \le 1+\varepsilon_0} \langle \boldsymbol{y}, \nabla H_N(\boldsymbol{x})\rangle_N \\ &\le \sup_{\|\boldsymbol{y}\|_N \le 1, \|\boldsymbol{x}\|_N \le 1+\varepsilon_0} \sum_{p\ge 2} \frac{c_p}{p!N} p \langle \boldsymbol{W}^{(p)}, \boldsymbol{x}^{\otimes(p-1)} \otimes \boldsymbol{y}\rangle_N \\ &\le \sum_{p\ge 2} \frac{c_p N^{(p-2)/2}}{p!} p \|\boldsymbol{W}^{(p)}\|_{\rm op} (1+\varepsilon_0)^{p-1} \\ &\overset{(a)}{\le} \sum_{p\ge 2} c_p p^{3/2} (1+\varepsilon_0)^{p-1} \overset{(b)}{\le} C. \end{aligned}$$

Here, the inequality (a) holds by equation (5.11), and (b) since $|c_p| \le c_* \alpha^k$ for some $\alpha < 1$ (recall that $\xi(t) < \infty$ for some $t > 1$). □

LEMMA 5.7. *There exists a constant C such that*

$$\text{p-lim}_{N\to\infty} \left| \frac{1}{N} H_N(\boldsymbol{m}^{\ell_*}) - \frac{1}{N} H_N(\hat{\boldsymbol{m}}) \right| \le C\sqrt{\delta}. \tag{5.13}$$

PROOF. Define the test function $\psi : \mathbb{R} \to \mathbb{R}$, $\psi(x) \equiv \min_{z\in[-1,+1]} (x-z)^2$, that is,

$$\psi(x) = \begin{cases} (|x|-1)^2 & \text{if } |x| > 1, \\ 0 & \text{if } |x| \le 1. \end{cases}$$

Proposition 3.1 implies

$$\text{p-lim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \psi(\boldsymbol{m}_i^{\ell_*}) = \mathbb{E}\psi(M_{\ell_*}^\delta).$$

On the other hand, Proposition 5.3 yields

$$\mathbb{E}\psi(M_{\ell_*}^\delta) \le \mathbb{E}\psi(M_{t_*}) + C\delta \le C\delta,$$

where the second inequality follows because $M_{t_*} \in [-1, +1]$ almost surely. Note that $\|\boldsymbol{m}^{\ell_*} - \hat{\boldsymbol{m}}\|_N^2 = \sum_{i=1}^N \psi(\boldsymbol{m}_i^{\ell_*})/N$ and, therefore, we conclude

$$\text{p-lim}_{N\to\infty} \|\boldsymbol{m}^{\ell_*} - \hat{\boldsymbol{m}}\|_N \le C\sqrt{\delta}. \tag{5.14}$$

Now, by the intermediate value theorem, there exists $s \in [0, 1]$ such that, for $\tilde{\boldsymbol{m}} = (1 - s)\boldsymbol{m}^{\ell_*} + s\hat{\boldsymbol{m}}$,

$$\begin{aligned}\left|\frac{1}{N} H_N(\boldsymbol{m}^{\ell*}) - \frac{1}{N} H_N(\hat{\boldsymbol{m}})\right| &= \frac{1}{N} |\langle \nabla H_N(\tilde{\boldsymbol{m}}), \boldsymbol{m}^{\ell_*} - \hat{\boldsymbol{m}}\rangle_N| \\ &\le \sup_{\|\boldsymbol{x}\|_N \le 1+C\sqrt{\delta}} \|\nabla H_N(\boldsymbol{x})\|_N \cdot \|\boldsymbol{m}^{\ell_*} - \hat{\boldsymbol{m}}\|_N \\ &\le C\sqrt{\delta},\end{aligned}$$

where we used equation (5.14) and Lemma 5.6. □

5.3.2. *Rounding*. We next round $\hat{\boldsymbol{m}} \in [-1, +1]^N$ to $\boldsymbol{\sigma}^{\text{alg}} \in \{-1, +1\}^N$. In order to define the rounding, we introduce the modified Hamiltonian

$$\tilde{H}_N(\boldsymbol{\sigma}) \equiv \sum_{k=2}^{\infty} c_k \sum_{i_1 < \cdots < i_k} W_{i_1,\dots,i_k}^{(k)} \sigma_{i_1} \cdots \sigma_{i_k}.$$

LEMMA 5.8. *There exists a constant $C > 0$ such that, with high probability,*

$$\max_{\boldsymbol{x}\in[-1,1]^N} |H_N(\boldsymbol{x}) - \tilde{H}_N(\boldsymbol{x})| \le C\sqrt{N \log N}. \tag{5.15}$$

PROOF. Note that $\tilde{H}_N(\boldsymbol{x})$ is obtained from $H_N(\boldsymbol{x})$ by restricting the sum in equation (1.1) to terms with distinct indices. As a consequence, $G_N(\boldsymbol{x}) = H_N(\boldsymbol{x}) - \tilde{H}_N(\boldsymbol{x})$ is a Gaussian process independent of $\tilde{H}_N(\boldsymbol{x})$. We therefore have

$$\begin{aligned}\mathbb{E}\{G_N(\boldsymbol{x})^2\} &= \mathbb{E}\{H_N(\boldsymbol{x})^2\} - \mathbb{E}\{\tilde{H}_N(\boldsymbol{x})^2\} \\ &= N\xi(\|\boldsymbol{x}\|_N^2) - \sum_{k=2}^{\infty} c_k^2 \sum_{i_1<\cdots<i_k} \mathbb{E}\{(W_{i_1,\dots,i_k}^{(k)})^2\} x_{i_1}^2 \cdots x_{i_k}^2 \\ &= N \sum_{k=2}^{\infty} c_k^2 \frac{1}{N^k} \sum_{i_1,\dots,i_k \in D^c(N,k)} x_{i_1}^2 \cdots x_{i_k}^2,\end{aligned}$$

where $D^c(N, k)$ is the subset of $[N]^k$ consisting of k-uples that are not distinct. A union bound yields $|D^c(N, k)| \le N^{k-1}k(k - 1)/2$, whence

$$\mathbb{E}\{G_N(\boldsymbol{x})^2\} \le N \sum_{k=2}^{\infty} c_k^2 \frac{|D^c(N,k)|}{N^k} \le \sum_{k=2}^{\infty} c_k^2 k^2 \le C.$$

Note that, with high probability, $\|\nabla G_N(\boldsymbol{x})\| = \|\nabla H_N(\boldsymbol{x})\| + \|\nabla \tilde{H}_N(\boldsymbol{x})\| \le C_*\sqrt{N}$ for all $\boldsymbol{x} \in [-1, +1]^N$ (the bound for $\nabla H_N(\boldsymbol{x})$ is proven in Lemma 5.6, and the one for $\nabla \tilde{H}_N(\boldsymbol{x})$ follows analogously). Let $\mathcal{N}_N(\varepsilon)$ be an ε-net (with respect the ordinary Euclidean distance) of $[-1, 1]^N$. Then, for $\varepsilon < t/(2C_*\sqrt{N})$,

$$\mathbb{P}\Big(\max_{\boldsymbol{x}\in[-1,1]^N} |H_N(\boldsymbol{x}) - \tilde{H}_N(\boldsymbol{x})| \ge t\Big) \le \mathbb{P}\Big(\max_{\boldsymbol{x}\in\mathcal{N}_N(\varepsilon)} |G_N(\boldsymbol{x})| \ge \frac{t}{2}\Big)$$

$$
\begin{aligned}
&+\mathbb{P}\Big(\sup_{\boldsymbol{x}\in[-1,+1]^N}\|\nabla G_N(\boldsymbol{x})\| > C_*\sqrt{N}\Big)\\
&\le 2|\mathcal{N}_N(\varepsilon)|e^{-t^2/2C}+o(1)\\
&\le 2\Big(\frac{\sqrt{N}}{\varepsilon}\Big)^N e^{-t^2/2C}+o(1).
\end{aligned}
$$

The proof is completed by taking $\varepsilon = 1$ and $t = C_0\sqrt{N\log N}$ with C_0 a large enough constant. □

We are now in position to complete our description of the rounding procedure. Notice that $\tilde{H}_N(\boldsymbol{x})$ is linear in each coordinate of $\boldsymbol{x}$. Therefore, viewed as a function of x_i, it is maximized over $[-1,+1]$ at $x_i \in \{-1,+1\}$. We starts from $\hat{\boldsymbol{m}}$ and sequentially maximize $\tilde{H}_N$ over each coordinate.

Explicitly, we can write $\tilde{H}_N(\boldsymbol{x}) = \tilde{H}_N^{(-i)}(\boldsymbol{x}_{-i}) + x_i\Delta_i\tilde{H}_N(\boldsymbol{x}_{-i})$, where $\boldsymbol{x}_{-i} \equiv (x_j)_{j\in[N]\setminus i}$. We then define $\boldsymbol{x}^{(j)}$, $j\in\{0,\dots,N\}$ by letting $\boldsymbol{x}^{(0)} = \hat{\boldsymbol{m}}$ and, for $j\ge 1$,

$$
x_i^{(j)} = \begin{cases} x_i^{(j-1)} & \text{if } i\neq j,\\ \operatorname{sign}(\Delta_i\tilde{H}_N(\boldsymbol{x}_{-i}^{(j)})) & \text{if } i = j.\end{cases}
$$

We then return the last vector $\boldsymbol{\sigma}^{\mathrm{alg}} \equiv \boldsymbol{x}^{(N)}$.

The proof of Theorem 2 is completed by noting that the following inequalities hold with high probability:

$$
\begin{aligned}
\frac{1}{N}H_N(\boldsymbol{\sigma}^{\mathrm{alg}}) &\overset{\text{(a)}}{\ge} \frac{1}{N}\tilde{H}_N(\boldsymbol{\sigma}^{\mathrm{alg}}) - C\sqrt{\frac{\log N}{N}}\\
&\overset{\text{(b)}}{\ge} \frac{1}{N}\tilde{H}_N(\hat{\boldsymbol{m}}) - C\sqrt{\frac{\log N}{N}}\\
&\overset{\text{(c)}}{\ge} \frac{1}{N}H_N(\hat{\boldsymbol{m}}) - 2C\sqrt{\frac{\log N}{N}}\\
&\overset{\text{(d)}}{\ge} \frac{1}{N}H_N(\boldsymbol{m}^{\ell_*}) - C\sqrt{\delta} - 2C\sqrt{\frac{\log N}{N}}.
\end{aligned}
$$

Here, (a) and (c) follow from Lemma 5.8, (b) from the fact that the $\tilde{H}_N$ is nondecreasing along the rounding procedure, and (d) from Lemma 5.7. Finally, the value $H_N(\boldsymbol{m}^{\ell_*})/N$ is lower bounded using Proposition 5.4.

6. Analysis of the variational principle and proof of Theorem 3.

6.1. *Properties of the variational principle.* In this section, we prove several useful properties of the extended variational principle $\inf_{\gamma\in\mathscr{L}}\mathsf{P}(\gamma)$. A first set of properties concerns the solution of the Parisi PDE (1.5) for $\gamma\in\mathscr{L}$. These are mostly generalizations of results obtained in [4, 28] for $\gamma\in\mathscr{U}$ bounded (hence, with finite total variation over $[0,1]$). We will refer to the proofs of [28] whenever they can be adapted without significant changes. In several cases, new arguments are required, for example, in the regularity result of Lemma 6.3, in the first variation formula of Proposition 6.8 and elsewhere. The second set of technical results concerns properties of the minimizers (starting with Lemma 6.9). These are of course entirely new because the minimizer is—in general—outside $\mathscr{U}$.

We consider the function space $\mathscr{L}$ from (2.6), endowed with the weighted L^1 distance $\|\gamma_1-\gamma_2\|_{1,\xi''}=\int_0^1 \xi''(t)|\gamma_1(t)-\gamma_2(t)|\,\mathrm{d}t$. We will write $\gamma_n \xrightarrow{L^1_\xi} \gamma$, whenever $\|\gamma_n-\gamma\|_{1,\xi''}\to 0$ as $n\to\infty$. We recall the space of piecewise constant functions

$$\mathsf{SF}_+ = \left\{ g=\sum_{i=1}^m a_i \mathbb{I}_{[t_{i-1},t_i)} : 0=t_0<t_1<\cdots<t_m=1, a_i\geq 0, m\in\mathbb{N}\right\}. \tag{6.1}$$

We study the PDE (1.5), with a slightly more general initial condition

$$\begin{aligned} &\partial_t \Phi(t,x)+\frac{1}{2}\xi''(t)\big(\partial_x^2\Phi(t,x)+\gamma(t)\big(\partial_x\Phi(t,x)\big)^2\big)=0,\\ &\Phi(1,x)=f_0(x). \end{aligned} \tag{6.2}$$

Throughout we assume f_0 to be convex, continuous, nonnegative, with $f_0(-x)=f_0(x)\geq 0$, and differentiable for $x\neq 0$, with $0\leq f_0'(x)\leq 1$ for all $x>0$. We will write $f_0'(x)$ for the weak derivative of f_0 (the right and left derivatives exist but are potentially different at $x=0$). Associated to the above PDE, we consider the following stochastic differential equation driven by Brownian motion $(B_t)_{t\geq 0}$:

$$\mathrm{d}X_t=\xi''(t)\gamma(t)\partial_x\Phi(t,X_t)\,\mathrm{d}t+\sqrt{\xi''(t)}\,\mathrm{d}B_t, \qquad X_0=0. \tag{6.3}$$

In the following we will also write Φ_x, Φ_{xx} and so on for the partial derivatives of Φ, and Φ^γ whenever we want to emphasize the dependence of Φ on γ. We write $\partial_t^{\pm}\Phi$ for the left and right derivatives of Φ.

We first collect a few properties of $\Phi(t,x)$ when $\gamma\in\mathsf{SF}_+$.

PROPOSITION 6.1.

(a) *For any $\gamma\in\mathsf{SF}_+$ the solution $\Phi:[0,1]\times\mathbb{R}\to\mathbb{R}$ of equation (6.2) exists uniquely in the classical sense and is smooth for $t\in[0,1)$. Namely, for any $j>0$, $\|\partial_x^j\Phi\|_{L^\infty([0,1-\varepsilon)\times\mathbb{R})}\leq C(\gamma,\varepsilon)$, and $\|\partial_t^{\pm}\partial_x^j\Phi\|_{L^\infty([0,1-\varepsilon)\times\mathbb{R})}\leq C(\gamma,\varepsilon)$, with $\partial_t^+\partial_x^j\Phi(t,x)=\partial_t^-\partial_x^j\Phi(t,x)$ whenever t is a continuity point of γ.*

(b) *For any $\gamma\in\mathsf{SF}_+$ the solution Φ of equation (6.2) is such that $x\mapsto\partial_x\Phi(t,\cdot)$ is nondecreasing for all $t\in[0,1]$, with $|\partial_x\Phi(t,x)|\leq 1$ for all $x\in\mathbb{R}$.*

(c) *If $\gamma_1,\gamma_2\in\mathsf{SF}_+$ and Φ^{γ_1}, Φ^{γ_2} are the corresponding solutions, then*

$$\|\Phi^{\gamma_1}-\Phi^{\gamma_2}\|_\infty \leq \|\gamma_1-\gamma_2\|_{1,\xi''}.$$

PROOF. Point (a) follows from the Cole–Hopf representation, which allows us to write an explicit form of the solution for $\gamma\in\mathsf{SF}_+$ [5, 27]. This solution is C^∞ except (possibly) when $t\in\{t_1,\dots,t_{m-1}\}$, the set of discontinuity points of γ. As a consequence of point (a), the SDE (6.3) is well defined, with unique strong solution on $[0,1]$. Further, Φ satisfies the following representation, for $\gamma\in\mathsf{SF}_+$ [28]:

$$\partial_x\Phi(t,x)=\mathbb{E}\big[f_0'(X_1)|X_t=x\big].$$

Since $\|f_0'\|_\infty\leq 1$, this implies $|\partial_x\Phi(t,x)|\leq 1$. The nondecreasing property also follows again by the Cole–Hopf representation.

Finally, point (c) is identical to Lemma 14 in [28] (the assumption that γ is nondecreasing is never used there). □

As a consequence of Proposition 6.1, we can define Φ^γ by continuity for any $\gamma\in\mathscr{L}$. Namely, we construct a sequence $\gamma_n\in\mathsf{SF}_+$, $\gamma_n\xrightarrow{L^1_\xi}\gamma$ and

$$\Phi^\gamma(t,x)=\lim_{n\to\infty}\Phi^{\gamma_n}(t,x).$$

LEMMA 6.2. *For any $\gamma \in \mathscr{L}$, Φ^γ constructed above is such that $\partial_x \Phi^\gamma$ exists in weak sense, is nondecreasing, and $|\partial_x \Phi^\gamma(t,x)| \le 1$ for all $t \in [0,1]$, $x \in \mathbb{R}$. Further, if $\gamma_n \in \mathsf{SF}_+$, $\gamma_n \xrightarrow{L^1_\xi} \gamma$, for any $t \in [0,1]$, we have $\partial_x \Phi^{\gamma_n}(t,x) \to \partial_x \Phi^\gamma(t,x)$ for almost every x.*

Finally, $\Phi = \Phi^\gamma$ is a weak solution of the PDE (6.2). Namely, for any $h \in C_c^\infty((0,1]\times\mathbb{R})$, we have

$$
\begin{aligned}
0 = {} & \int_{(0,1]}\int_{\mathbb{R}} \Big\{ -\Phi \partial_t h + \frac{1}{2}\xi''(t)\big(\Phi \partial_x^2 h + \gamma(t)(\partial_x \Phi)^2 h\big)\Big\}\,\mathrm{d}x\,\mathrm{d}t \\
& + \int_{\mathbb{R}} \Phi(1,x) f_0(x)\,\mathrm{d}x.
\end{aligned}
\tag{6.4}
$$

PROOF. Since $\Phi^\gamma(t,\cdot)$ is the uniform limit of convex 1-Lipschitz functions, it is also convex 1-Lipschitz. Hence its weak derivative exists, is nondecreasing and is bounded as claimed. The claim $\partial_x \Phi^{\gamma_n}(t,x) \to \partial_x \Phi^\gamma(t,x)$ follows by dominated convergence.

In order to show that Φ is a weak solution, let $\Phi^n = \Phi^{\gamma_n}$ for $\gamma_n \in \mathsf{SF}_+$, $\gamma_n \xrightarrow{L^1_\xi} \gamma$ (hence $\|\Phi^n - \Phi\|_\infty \to 0$). Since Φ^n is a classical solution corresponding to γ_n, we have

$$
0 = \int_{(0,1]}\int_{\mathbb{R}} \Big\{ -\Phi^n \partial_t h + \frac{1}{2}\xi''(t)\big(\Phi^n \partial_x^2 h + \gamma_n(t)\big(\partial_x \Phi^n\big)^2 h\big)\Big\}\,\mathrm{d}x\,\mathrm{d}t + \int_{\mathbb{R}} \Phi^n(1,x) f_0(x)\,\mathrm{d}x.
$$

Letting Δ denote the right-hand side of equation (6.4), we have (since $\Phi^n(1,x) = \Phi(1,x)$ is independent of n)

$$
\begin{aligned}
\Delta = {} & \int_{(0,1]}\int_{\mathbb{R}} \Big\{ (\Phi^n - \Phi)\partial_t h - \frac{1}{2}\xi''(t)\big(\Phi^n - \Phi\big)\partial_x^2 h\Big\}\,\mathrm{d}x\,\mathrm{d}t \\
& - \int_{(0,1]}\int_{\mathbb{R}} \frac{1}{2}\xi''(t)\big(\gamma_n(t)\big(\partial_x \Phi^n\big)^2 - \gamma(t)(\partial_x \Phi)^2\big) h\,\mathrm{d}x\,\mathrm{d}t.
\end{aligned}
$$

The first term vanishes as $n \to \infty$ by dominated convergence. For the second term, by the bound on $\partial_x \Phi$, $\partial_x \Phi^n$, we have

$$
\begin{aligned}
|\Delta| \le {} & \frac{1}{2}\int_{(0,1]}\int_{\mathbb{R}} \xi''(t)\big|\gamma_n(t) - \gamma(t)\big||h|\,\mathrm{d}x\,\mathrm{d}t \\
& + \frac{1}{2}\int_{(0,1]}\int_{\mathbb{R}} \xi''\gamma(t)\big|\big(\partial_x \Phi^n\big)^2 - (\partial_x \Phi)^2\big||h|\,\mathrm{d}x\,\mathrm{d}t.
\end{aligned}
$$

The first term vanishes as $n \to \infty$ since $\gamma_n \xrightarrow{L^1_\xi} \gamma$, and the second vanishes by dominated convergence, using the fact that $\|\xi''\gamma\|_1 < \infty$. □

LEMMA 6.3. *For $\gamma \in \mathscr{L}$ and any $t \in [0,1)$, the second derivative $\partial_x^2 \Phi(t,\cdot)$ exists in the weak sense, with $\sup_{0\le t\le 1-\varepsilon} \|\partial_x^2 \Phi(t,\cdot)\|_{L^2(\mathbb{R})} < \infty$ for any $\varepsilon > 0$.*

PROOF. Following [28], it is useful to introduce the the smooth time change $\theta(t) = (\xi'(1) - \xi'(t))/2$, and define $u : [0,\theta_M] \times \mathbb{R}$, $\theta_M = \xi'(1)/2$, via $u(\theta(t),x) = \Phi(t,x)$. By a simple change of variables, u is a weak solution of the PDE

$$
\partial_\theta u - \Delta u = m(\theta) u_x^2, \qquad u(0,x) = f_0(x),
$$

where $m(s) = \gamma(\theta^{-1}(s))$. The desired claim is implied by showing that the partial derivative $\partial_x^2 u$ exists in weak sense and is bounded uniformly over $\theta > \varepsilon$ (for any $\varepsilon > 0$).

Again, as in [28] the fact that u is a weak solution implies the Duhamel principle

$$(6.5)\qquad \begin{aligned} u(\theta) &= G_\theta * f_0 + \int_0^\theta m(s) G_{\theta-s} * u_x(s)^2 \, \mathrm{d}s, \\ G_t(x) &\equiv \frac{1}{\sqrt{4\pi t}} e^{-x^2/4t}. \end{aligned}$$

(Here, $*$ denotes convolution and this equation is to be interpreted in weak sense; namely, for any $h \in C_c^\infty(\mathbb{R})$, $\int h(x)u(\theta,x)\,\mathrm{d}x$ is given by the convolution with h of the right-hand side.) Note that by Lemma 6.2, $x \mapsto u_x(s,x)^2$ is bounded between 0 and 1, nonincreasing in $(-\infty, 0]$, nondecreasing in $[0,\infty)$ and symmetric (the value at $x = 0$ is immaterial). Hence there exists a measure ν_s on $[0,\infty)$, with total mass $\nu_s([0,\infty)) \le 1$, such that

$$u_x(s,x)^2 = \nu_s([0,x))\mathbb{I}_{x>0} + \nu_s([0,-x))\mathbb{I}_{x<0}.$$

We then obtain, from equation (6.5)

$$(6.6)\qquad u_{xx}(\theta) = G'_\theta * f'_0 + \int_0^\theta m(s) \int_{\mathbb{R}_{\ge 0}} \big[G'_{\theta-s}(\cdot - x) + G'_{\theta-s}(\cdot + x)\big] \,\mathrm{d}\nu_s(x)\,\mathrm{d}s.$$

The claim follows by showing that each of the two terms on the right-hand side of equation (6.6) is a well-defined function, bounded in $L^2(\mathbb{R})$. For the first term, notice that f'_0 is bounded and nondecreasing. Hence there exists a measure ω_0 on $\mathbb{R}$ with $\omega_0(\mathbb{R}) \le 2$, such that $G'_\theta * f'_0 = G_\theta * \mathrm{d}\omega_0$, whence

$$\|G'_\theta * f'_0\|_2 = \left\| \int G_\theta(\cdot - x)\,\mathrm{d}\omega_0(x) \right\|_2 \le 2\|G_\theta\|_2 \le \frac{C}{\theta^{1/4}},$$

where the upper bound follows from Jensen's inequality. The second term on the right-hand side of (6.6) can be treated analogously. Denoting it by $w(\theta)$, we have, again by Jensen with $\theta = \theta(1-\varepsilon)$,

$$\begin{aligned} \|w(\theta)\|_2 &\le \int_0^\theta m(s) \int_{\mathbb{R}_{\ge 0}} \|G'_{\theta-s}(\cdot - x) + G'_{\theta-s}(\cdot + x)\|_2 \,\mathrm{d}\nu_s(x)\,\mathrm{d}s \\ &\le C \int_0^\theta m(s) \frac{1}{(\theta - s)^{3/4}} \,\mathrm{d}s \\ &\le C' \int_{1-\varepsilon}^1 \frac{\xi''\gamma(s)}{(\xi'(s) - \xi'(1-\varepsilon))^{3/4}} \,\mathrm{d}s, \end{aligned}$$

where the second inequality follows by $\|G'_t\|_2 \le Ct^{-3/4}$. Decomposing the last integral, we get

$$\begin{aligned} \|w(\theta)\|_2 &\le C' \int_{1-\varepsilon}^{1-\varepsilon/2} \frac{\xi''\gamma(s)}{(\xi'(s) - \xi'(1-\varepsilon))^{3/4}} \,\mathrm{d}s + C' \int_{1-\varepsilon/2}^1 \frac{\xi''\gamma(s)}{(\xi'(s) - \xi'(1-\varepsilon))^{3/4}} \,\mathrm{d}s \\ &\le C'\xi''\gamma(1-\varepsilon/2) \int_{1-\varepsilon}^{1-\varepsilon/2} \frac{1}{(\xi'(s) - \xi'(1-\varepsilon))^{3/4}} \,\mathrm{d}s \\ &\quad + \frac{C'}{(\xi'(1-\varepsilon/2) - \xi'(1-\varepsilon))^{3/4}} \int_{1-\varepsilon/2}^1 \xi''\gamma(s)\,\mathrm{d}s \\ &\le C''\|\xi''\gamma\|_{\mathrm{TV}[0,1-\varepsilon/2]} + C''\varepsilon^{-3/4}\|\xi''\gamma\|_1. \end{aligned}$$

The last expression is bounded by some $C(\varepsilon) < \infty$ since $\gamma \in \mathscr{L}$. □

LEMMA 6.4. *For any $\gamma \in \mathscr{L}$, the solution $\Phi = \Phi^\gamma$ constructed above is continuous on* $[0, 1] \times \mathbb{R}$, *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* $j \geq 2$.
(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* $j \geq 1$.

PROOF. Continuity follows since Φ^γ is the uniform limit of continuous functions. Points (a) and (b) follow from the same proof as Lemma 10 in [28], applied to the PDE (6.2) with boundary condition at $t = 1 - \varepsilon$, whereby we use Lemma 6.3 to initiate the bootstrap procedure. □

As a consequence of the stated regularity properties of Φ, we can solve the SDE (6.3).

LEMMA 6.5. *For any $\gamma \in \mathscr{L}$, let $\Phi = \Phi^\gamma$ be the PDE solution defined above. Then the stochastic differential equation* (6.3) *has unique strong solution on* $(X_t)_{t\in[0,1]}$, *which is almost surely continuous. Further, for any* $t \in [0, 1]$,

$$\partial_x \Phi(t, X_t) = \int_0^t \sqrt{\xi''(s)}\partial_x^2 \Phi(s, X_s)\,\mathrm{d}B_s. \tag{6.7}$$

PROOF. Existence and uniqueness for $t \in [0, 1-\varepsilon)$ follow because $\partial_x \Phi(t, \cdot)$ is Lipschitz continuous and $\xi''\gamma$ is bounded on such interval (see, e.g., [9], Chapter 5.) By letting $\varepsilon \downarrow 0$, we obtain existence and uniqueness on $[0, 1)$. Further X_t can be extended at $t = 1$, letting

$$X_1 = \int_0^1 \xi''(t)\gamma(t)\partial_x \Phi(t, X_t)\,\mathrm{d}t + \int_0^1 \sqrt{\xi''(t)}\,\mathrm{d}B_t.$$

It is easy to check that this extension is almost surely continuous at $t = 1$, since

$$|X_1 - X_t| \leq \int_t^1 \xi''\gamma(s)\,\mathrm{d}s + \int_t^1 \sqrt{\xi''(t)}\,\mathrm{d}B_t.$$

The first integral vanishes as $t \to 1$ since $\int_0^1 \xi''\gamma(t)\,\mathrm{d}t < \infty$, while the second vanishes by continuity of the Brownian motion.

Next notice that, since $\Phi_x = \partial_x \Phi$ smooth in space and weakly differentiable in time for $t \in [0, 1)$ by Lemma 6.4, it is a weak solution of

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

More precisely, for any $x \in \mathbb{R}$ and any $h \in C_c^((0, 1))$, we have

$$\int \left\{ h(t)\partial_t \Phi_x(t, x) + \frac{\xi''(t)}{2} h(t)\big(\partial_x^2 \Phi_x(t, x) + \gamma(t)\partial_x\big(\Phi_x(t, x)\big)^2\big) \right\} \mathrm{d}t = 0. \tag{6.8}$$

Equation (6.7) is then obtained by Itô formula (see Proposition 22 in [28])

$$\begin{aligned} \partial_x \Phi(t, X_t) = &\int_0^t \sqrt{\xi''(s)}\partial_x^2 \Phi(s, X_s)\,\mathrm{d}B_s \\ &+ \int_0^t \big(\partial_s \Phi_x(s, X_s) + \frac{1}{2}\xi''(s)\big(\partial_x^2 \Phi_x(s, X_s) + \gamma(s)\partial_x\big(\Phi_x(s, X_s)\big)^2\big)\big\}\,\mathrm{d}s, \end{aligned}$$

The second term vanishes by equation (6.8). □

COROLLARY 6.6. *For any $\gamma \in \mathscr{L}$ and any $0 \le t_1 < t_2 < 1$,*

$$\mathbb{E}\{\partial_x \Phi(t_2, X_{t_2})^2\} - \mathbb{E}\{\partial_x \Phi(t_1, X_{t_1})^2\} = \int_{t_1}^{t_2} \xi''(s)\mathbb{E}\{(\partial_x^2 \Phi(s, X_s))^2\}\,\mathrm{d}s.$$

In particular, $t \mapsto \mathbb{E}\{\partial_x \Phi(t, X_t)^2\}$ is Lipschitz continuous on $[0, 1-\varepsilon)$ for any $\varepsilon > 0$.

PROOF. This follows from Lemma 6.5, using the regularity properties of Lemma 6.4. □

LEMMA 6.7. *For any $\gamma \in \mathscr{L}$, the function $t \mapsto \mathbb{E}\{\partial_x^2 \Phi(t, X_t)^2\}$ is continuous on $[0, 1)$.*

PROOF. The function is continuous by an application of bounded convergence (using the continuity of $t \mapsto X_t$ and the regularity of Lemma 6.4). □

We now compute the first variation of the Parisi functional.

PROPOSITION 6.8. *Let $\gamma \in \mathscr{L}$, and $\delta : [0, 1) \to \mathbb{R}$ be such that $\|\xi''\delta\|_{\mathrm{TV}[0,t]} < \infty$ 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 $\gamma + s\delta \ge 0$ for all $s \in [0, s_0)$ for some positive s_0. Then*

$$\frac{\mathrm{dP}}{\mathrm{d}s}(\gamma + s\delta)|_{s=0+} = \frac{1}{2}\int_0^1 \xi''(t)\delta(t)\big(\mathbb{E}\{\partial_x \Phi(t, X_t)^2\} - t\big)\,\mathrm{d}t. \tag{6.9}$$

(*Here*, $(X_t)_{t\in[0,1]}$ *is the solution of the SDE* (6.3).)

PROOF. Let $\gamma^s \equiv \gamma + s\delta$, $s \in [0, \varepsilon)$, and denote by Φ^s the corresponding solution of the Parisi PDE. Following the proof of Lemma 14 in [28], we get

$$\Phi^s(0, 0) - \Phi^0(0, 0) = \frac{s}{2}\int_0^1 \xi''(t)\delta(t)\mathbb{E}\{\partial_x \Phi^0(t, Y_t^s)^2\}\,\mathrm{d}t, \tag{6.10}$$

where Y_t^s is the solution of the SDE

(6.11)

$$\mathrm{d}Y_t^s = \frac{1}{2}\xi''(t)\gamma^s(t)\big[\partial_x \Phi^0(t, Y_t^s) + \partial_x \Phi^s(t, Y_t^s)\big]\,\mathrm{d}t + \sqrt{\xi''(t)}\,\mathrm{d}B_t, \qquad Y_0^s = 0.$$

We also obtain (by the same argument as in [28], Lemma 14, using Lemma 6.4, and noting that $\delta(t) = 0$ for $t > 1-\varepsilon$ and $\xi''\gamma$ is bounded on $[0, 1-\varepsilon)$)

$$\|\partial_x \Phi^s - \partial_x \Phi^0\|_\infty \le C(\varepsilon, \gamma)\|\xi''\delta\|_1 \cdot s. \tag{6.12}$$

Taking the difference between this equations (6.11) and (6.3), we get for $t \in [0, 1-\varepsilon_0)$,

$$\begin{aligned}
|Y_t^s - X_t| &\le C\int_0^t \xi''(u)|\gamma^s(u) - \gamma(u)|\,\mathrm{d}u + C\int_0^t \xi''\gamma(u)|\partial_x \Phi^0(u, Y_u^s) - \partial_x \Phi^s(u, Y_u^s)|\,\mathrm{d}u \\
&\quad + C\int_0^t \xi''\gamma(u)|\partial_x \Phi^0(u, X_u) - \partial_x \Phi^0(u, Y_u^s)|\,\mathrm{d}u \\
&\le C\|\xi''(\gamma^s - \gamma^0)\|_1 + C(\varepsilon, \gamma)\|\xi''(\gamma^s - \gamma^0)\|_1\|\xi''\gamma\|_1 \\
&\quad + C(\varepsilon_0)\int_0^t \xi''\gamma(u)|Y_u^s - X_u|\,\mathrm{d}u.
\end{aligned}$$

In the second inequality we used equation (6.12), and the fact that $\partial_x^2\Phi$ is bounded for $t \in [0, 1-\varepsilon_0)$, see Lemma 6.4. Since $\xi''\gamma(u) \le \|\xi''\gamma\|_{\mathrm{TV}[0,1-\varepsilon_0]}$ for $u \in [0, 1-\varepsilon_0)$, we finally obtain

$$|Y_t^s - X_t| \le C(\gamma, \varepsilon)s\|\xi''\delta\|_1 + C(\gamma, \varepsilon_0)\int_0^t |Y_u^s - X_u|\,\mathrm{d}u.$$

Therefore, we conclude by Gronwall lemma that

$$\sup_{t\le 1-\varepsilon_0} |Y_t^s - X_t| \le C(\varepsilon, \varepsilon_0, \gamma)\|\xi''\delta\|_1 s$$

Using this in equation (6.10), together with the fact that $\partial_x\Phi^0$ is bounded and Lipschitz, and $\delta(t) = 0$ for $t > 1-\varepsilon$, we get

$$\Phi^s(0,0) - \Phi^0(0,0) = \frac{s}{2}\int_0^1 \xi''(t)\delta(t)\mathbb{E}\{\partial_x\Phi^0(t, X_t)^2\}\,\mathrm{d}t + O(s^2),$$

whence equation (6.9) immediately follows. □

For any $\gamma \in \mathscr{L}$, we have $\|\gamma\|_{\mathrm{TV}[0,t]} < \infty$ for any $t \in [0, 1)$. We can therefore modify γ in (at most) countably many points to obtain a right-continuous function. Since this modification does not change the solution Φ^γ, by Proposition 6.1, we will hereafter assume that any $\gamma \in \mathscr{L}$ is right continuous.

For $\gamma \in \mathscr{L}$, we denote by $S(\gamma) \equiv \{t \in [0, 1) : \gamma(t) > 0\}$ the *support* of γ, and by $\overline{S}(\gamma)$ the closure of $S(\gamma)$ in $[0, 1)$ (in particular, note that $1 \notin \overline{S}(\gamma)$).

LEMMA 6.9. *The support $S(\gamma)$ is a disjoint union of countably many intervals $S(\gamma) = \bigcup_{\alpha\in A} I_\alpha$, where $I_\alpha = (a_\alpha, b_\alpha)$ or $I_\alpha = [a_\alpha, b_\alpha)$, $a_\alpha < b_\alpha$, and A is countable.*

PROOF. If $t_0 \in S(\gamma)$, then by right continuity there exists $\delta > 0$ such that $[t_0, t_0 + \delta) \subseteq S(\gamma)$. This implies immediately the claim. □

COROLLARY 6.10. *Assume $\gamma_* \in \mathscr{L}$ is such that $\mathsf{P}(\gamma_*) = \inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma)$. Then*

$$t \in \overline{S}(\gamma_*) \quad\Rightarrow\quad \mathbb{E}\{\partial_x\Phi^{\gamma_*}(t, X_t)^2\} = t, \tag{6.13}$$

$$t \in [0, 1)\setminus\overline{S}(\gamma_*) \quad\Rightarrow\quad \mathbb{E}\{\partial_x\Phi^{\gamma_*}(t, X_t)^2\} \ge t. \tag{6.14}$$

PROOF. First, consider equation (6.13). For any $0 \le t_1 < t_2 < 1$, set $\delta(t) = \gamma_*(t)\mathbb{I}(t \in [t_1, t_2))$. Clearly, $\gamma_* + s\delta \in \mathscr{L}$ for $s \in (-1, 1)$. By the optimality of γ_*, and using Proposition 6.8, we have

$$0 = \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}s}(\gamma_* + s\delta)|_{s=0} = \frac{1}{2}\int_{t_1}^{t_2} \xi''(t)\gamma_*(t)\big(\mathbb{E}\{\partial_x\Phi^{\gamma_*}(t, X_t)^2\} - t\big)\,\mathrm{d}t$$

Since t_1, t_2 are arbitrary, and $\xi''(t) > 0$ for $t \in (0, 1)$ this implies $\gamma_*(t)(\mathbb{E}\{\partial_x\Phi^{\gamma_*}(t, X_t)^2\} - t) = 0$ for almost every $t \in [0, 1)$. Since $\gamma_*(t)$ is right continuous and $\mathbb{E}\{\partial_x\Phi^{\gamma_*}(t, X_t)^2\}$ is continuous (see Corollary 6.6), it follows that $\gamma_*(t)(\mathbb{E}\{\partial_x\Phi^{\gamma_*}(t, X_t)^2\} - t) = 0$ for every $t \in [0, 1)$. This in turns implies $\mathbb{E}\{\partial_x\Phi^{\gamma_*}(t, X_t)^2\} = t$ for every $t \in S(\gamma_*)$. This can be extended to $t \in \overline{S}(\gamma_*)$ again by continuity of $t \mapsto \mathbb{E}\{\partial_x\Phi^{\gamma_*}(t, X_t)^2\}$.

Next consider equation (6.14). Notice that, by Lemma 6.9, $[0, 1)\setminus\overline{S}(\gamma_*)$ 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 $\gamma_* + s\delta \in \mathscr{L}$ for $s \ge 0$. By Proposition 6.8, we have

$$0 \le \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}s}(\gamma + s\delta)|_{s=0} = \frac{1}{2}\int_{t_1}^{t_2} \xi''(t)\big(\mathbb{E}\{\partial_x\Phi(t, X_t)^2\} - t\big)\,\mathrm{d}t.$$

Since t_1, t_2 are arbitrary, $\xi''(t) > 0$ for $t \in (0, 1)$ and $t \mapsto \mathbb{E}\{\partial_x\Phi(t, X_t)^2\}$ is continuous, this implies $\mathbb{E}\{\partial_x\Phi(t, X_t)^2\} \ge t$ for all $t \in J$, and hence all $t \in [0, 1) \setminus \overline{S}(\gamma_*)$. □

COROLLARY 6.11. *Assume $\gamma_* \in \mathscr{L}$ is such that $\mathsf{P}(\gamma_*) = \inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma)$. Then*

$$t \in \overline{S}(\gamma_*) \quad \Rightarrow \quad \xi''(t)\mathbb{E}\{\partial_x^2\Phi^{\gamma_*}(t, X_t)^2\} = 1.$$

PROOF. Set $\Phi(t, x) = \Phi^{\gamma_*}(t, x)$. By Lemma 6.9, $\overline{S}(\gamma_*)$ is a disjoint union of closed intervals with nonempty interior. Let K be one such intervals. Then, for any $[t_1, t_2] \in K$, we have, by Lemma 6.10,

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

Since t_1, t_2 are arbitrary, we get $\xi''(t)\mathbb{E}\{\partial_x^2\Phi(t, X_t)^2\} = 1$ for almost every $t \in K$. Using Lemma 6.7, we get $\xi''(t)\mathbb{E}\{\partial_x^2\Phi(t, X_t)^2\} = 1$ for every $t \in \overline{S}(\gamma_*)$. □

LEMMA 6.12. *Assume $\gamma \in \mathscr{L}$ to be such that $\gamma(t) = 0$ 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 $M \in \mathbb{R}_{\ge 0}$, there exists $\varepsilon(t_*, M, \gamma) > 0$ such that*

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

PROOF. Since the SDE (6.3) has strong solutions, X_{t_1} is a well-defined random variable taking values in $\mathbb{R}$. Therefore, there exists $C_1 = C_1(\gamma) < \infty$ such that $\mathbb{P}(|X_{t_1}| \le C_1) \ge 1/2$. For $t \in (t_1, 1)$, X_t satisfies $\mathrm{d}X_t = \sqrt{\xi''(t)}\,\mathrm{d}B_t$ and, therefore, the law of X_t is the convolution of a Gaussian (with variance $\theta(t)^2 \equiv \xi'(t) - \xi(t_1) > 0$) with the law of X_{t_1}, and, therefore, has a density. To prove the desired lower bound on the density, let $f_G(x) = \exp(-x^2/2)/\sqrt{2\pi}$ denote the standard Gaussian density. Note that, for any $|x| \le M$,

$$\begin{aligned} p_t(x) &= \mathbb{E}\left\{\frac{1}{\theta(t)} f_G\left(\frac{x - X_{t_1}}{\theta(t)}\right)\right\} \\ &\ge \mathbb{E}\left\{\frac{1}{\theta(t)} f_G\left(\frac{x - X_{t_1}}{\theta(t)}\right)\mathbb{I}_{|X_{t_2}|\le C_1}\right\} \\ &\ge \frac{1}{\theta(t)} f_G\left(\frac{M + C_1}{\theta(t)}\right)\mathbb{P}(|X_{t_1}| \le C_1) \ge \frac{1}{2\theta(t)} f_G\left(\frac{M + C_1}{\theta(t)}\right). \end{aligned}$$

The latter expression is lower bounded by $\varepsilon(t_*, M, \gamma) > 0$ for any $t \in [t_*, 1]$, as claimed. □

LEMMA 6.13. *For any $\gamma \in \mathscr{L}$, let $\Phi = \Phi^\gamma$ be the solution of the Parisi PDE constructed above. Then the following identities hold (as weak derivatives in $[0, 1)$) have*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\{\Phi(t, X_t)\} = \frac{1}{2}\xi''(t)\gamma(t)\mathbb{E}\{\partial_x\Phi(t, X_t)^2\}, \tag{6.15}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\{X_t\partial_x\Phi(t, X_t)\} = \xi''(t)\gamma(t)\mathbb{E}\{\partial_x\Phi(t, X_t)^2\} + \xi''(t)\mathbb{E}\{\partial_x^2\Phi(t, X_t)\}. \tag{6.16}$$

PROOF. We will write $\Phi_t = \partial_t\Phi$, $\Phi_x = \partial_x\Phi$ and $\Phi_{xx} = \partial_x^2\Phi$. For the first identity, using the regularity properties of Lemma 6.4 and Itô's formula, we get

$$\mathrm{d}\Phi(t, X_t) = \Phi_t(t, X_t)\,\mathrm{d}t + \xi''(t)\gamma(t)\Phi_x(t, X_t)^2\,\mathrm{d}t$$

$$+\sqrt{\xi''(t)}\Phi_x(t,X_t)\,\mathrm{d}B_t+\frac{1}{2}\Phi_{xx}(t,X_t)\xi''(t)\,\mathrm{d}t$$
$$=\frac{1}{2}\xi''(t)\gamma(t)\Phi_x(t,X_t)^2\,\mathrm{d}t+\sqrt{\xi''(t)}\Phi_x(t,X_t)\,\mathrm{d}B_t,$$

where the equalities hold after integrating over a test function $h\in C_c^\infty([0,1))$ and in the second step we used the fact that Φ is a weak solution of equation (6.2). The claim (6.15) follows by taking expectations.

We proceed analogously for the second identity. Using Lemma 6.5, and the fact that the $(X_t)_{t\in[0,1)}$ solved the SDE (6.3), we get

$$\begin{aligned}
&\mathrm{d}\big(X_t\Phi_x(t,X_t)\big)\\
&\quad=\Phi_x(t,X_t)\,\mathrm{d}X_t+X_t\,\mathrm{d}\big(\Phi_x(t,X_t)\big)+\xi''(t)\Phi_{xx}(t,X_t)\,\mathrm{d}t\\
&\quad=\xi''(t)\gamma(t)\Phi_x(t,X_t)^2\,\mathrm{d}t+\sqrt{\xi''(t)}\Phi_x(t,X_t)\,\mathrm{d}B_t+\sqrt{\xi''(t)}X_t\Phi_{xx}(t,X_t)\,\mathrm{d}B_t\\
&\qquad+\xi''(t)\Phi_{xx}(t,X_t)\,\mathrm{d}t.
\end{aligned}$$

The claim (6.15) follows again by taking expectations. □

We now show that any minimizer γ_* of the Parisi functional over the extended space $\mathscr{L}$ has full support. Note that this is unrelated to the no-overlap gap property, which concerns solutions γ_* that are nondecreasing, and concerns the points of increase of γ_*.

THEOREM 5. *Consider the case* $f_0(x)=|x|$. *Assume* $\gamma_*\in\mathscr{L}$ *is such that* $\mathsf{P}(\gamma_*)=\inf_{\gamma\in\mathscr{L}}\mathsf{P}(\gamma)$. *Then* $\overline{S}(\gamma_*)=[0,1)$.

PROOF. Throughout this proof, $\Phi(t,x)=\Phi^{\gamma_*}(t,x)$.

By Lemma 6.9, $\overline{S}^c(\gamma_*)=[0,1)\setminus\overline{S}(\gamma_*)$ is a countable union of disjoint intervals, open in $[0,1)$. First, assume that at least one of these intervals is of the form (t_1,t_2) with $0<t_1<t_2<1$, or $[t_1=0,t_2)$, $t_2<1$. By Corollary 6.10 and Corollary 6.11, we know that

$$\mathbb{E}\big\{\partial_x\Phi(t_1,X_{t_1})^2\big\}=t_1,\qquad \xi''(t_2)\mathbb{E}\big\{\partial_x^2\Phi(t_2,X_{t_2})^2\big\}=1,\quad i\in\{1,2\},\tag{6.17}$$

$$\mathbb{E}\big\{\partial_x\Phi(t,X_t)^2\big\}\ge t\quad\forall t\in(t_1,t_2).\tag{6.18}$$

(Notice that the first identity in equation (6.17) holds also for $t_1=0$ since $\partial_x\Phi(0,0)=0$ by a symmetry argument.) Further, for $t\in(t_1,t_2)$, Φ solves the PDE $\Phi_t+(\xi''(t)/2)\partial_x^2\Phi=0$, which coincides with the heat equation, apart from a time change. We therefore obtain, for $t\in(t_1,t_2]$

$$\Phi(t,x)=\mathbb{E}\big\{\Phi\big(t_2,x+\sqrt{\xi'(t_2)-\xi'(t)}G\big)\big\},\quad G\sim\mathsf{N}(0,1).$$

Differentiating this equation, and using dominated convergence (thanks to the fact that $\partial_x^2\Phi(t_2,x)$ is bounded by Lemma 6.4), we get $\partial_x^2\Phi(t,x)=\mathbb{E}\{\partial_x^2\Phi(t_2,x+\sqrt{\xi'(t_2)-\xi'(t)}G)\}$. Notice also that the SDE (6.3) reads, for $t\in(t_1,t_2)$, $\mathrm{d}X_t=\sqrt{\xi''(t)}\,\mathrm{d}B_t$ and, therefore, we can rewrite the last equation as

$$\partial_x^2\Phi(t,X_t)=\mathbb{E}\big\{\partial_x^2\Phi(t_2,X_{t_2})|X_t\big\}.$$

By Jensen inequality, we have

$$\mathbb{E}\big\{\partial_x^2\Phi(t,X_t)^2\big\}\le\mathbb{E}\big\{\partial_x^2\Phi(t_2,X_{t_2})^2\big\}=\frac{1}{\xi''(t_2)},\tag{6.19}$$

where in the last step we used equation (6.17). Using Corollary 6.6, we get for $t \in [t_1, t_2]$,

$$\mathbb{E}\{\partial_x \Phi(t, X_t)^2\} = \mathbb{E}\{\partial_x \Phi(t_1, X_{t_1})^2\} + \int_{t_1}^{t} \xi''(s) \mathbb{E}\{\partial_x^2 \Phi(s, X_s)^2\} \, \mathrm{d}s$$
$$\leq t_1 + \int_{t_1}^{t} \frac{\xi''(s)}{\xi''(t_2)} \, \mathrm{d}s < t,$$

where in the last step we used the fact that $t \mapsto \xi''(t)$ is monotone increasing. The last equation is in contradiction with equation (6.17) and, therefore, $\overline{S}^c(\gamma_*)$ can be either empty, or consist of a single interval $(t_1, 1)$.

In order to complete the proof, we need to rule out the case $\overline{S}^c(\gamma_*) = (t_1, 1)$. Assume by contradiction that indeed $\overline{S}^c(\gamma_*) = (t_1, 1)$. For $t \in (t_1, 1)$, let $r = r(t) = \xi'(1) - \xi'(t)$, and notice that $r(t)$ is monotone 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 get $\partial_x \Phi(t, x) = \mathbb{E}\,\mathrm{sign}(G + x/\sqrt{r(t)})$, where $G \sim \mathsf{N}(0, 1)$, whence for $t \in (t_1, 1)$,

$$1 - \mathbb{E}\{\partial_x \Phi(t, X_t)^2\} = \mathbb{E}Q\left(\frac{X_t}{\sqrt{r(t)}}\right),$$
$$Q(x) \equiv 1 - \mathbb{E}\{\mathrm{sign}(x + G)\}^2.$$

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 $t \in (t_*, 1)$

$$1 - \mathbb{E}\{\partial_x \Phi(t, X_t)^2\} \geq \frac{1}{2}\mathbb{P}(|X_t| \leq \delta_0 \sqrt{r(t)})$$
$$\overset{\text{(a)}}{\geq} \delta_0 \varepsilon(t_*, 1, \gamma)\sqrt{r(t)} \overset{\text{(b)}}{\geq} C\sqrt{1-t},$$

where (a) follows by Lemma 6.12 and (b) holds for some $C = C(\gamma) > 0$. We therefore obtain $\mathbb{E}\{\partial_x \Phi(t, X_t)^2\} \leq 1 - C\sqrt{1-t}$, which contradicts Corollary 6.10 for t close enough to 1. □

6.2. *Proof of Theorem* 3. Before passing to the actual proof, we state and prove a simple lemma.

LEMMA 6.14. *Let $g : [a, b] \times \mathbb{R} \to \mathbb{R}$ be bounded and Lipschitz continuous in its first argument, that is, $|g(t_1, x) - g(t_2, x)| \leq L|t_1 - t_2|$ for all $x \in \mathbb{R}$, $t_1, t_2 \in [a, b]$, and $h : [a, b] \to \mathbb{R}$ have bounded total variation. Then $f = gh$ has bounded* strong *total variation.*

PROOF. Fix $a \leq t_0 < \cdots < t_n \leq b$ and $x_1, \ldots, x_n \in \mathbb{R}$. Then

$$\sum_{i=1}^{n} |f(t_i, x_i) - f(t_{i-1}, x_i)|$$
$$= \sum_{i=1}^{n} |h(t_i)g(t_i, x_i) - h(t_{i-1})g(t_{i-1}, x_i)|$$
$$\leq \sum_{i=1}^{n} |h(t_i)||g(t_i, x_i) - g(t_{i-1}, x_i)| + \sum_{i=1}^{n} |h(t_i) - h_(t_{i-1})||g(t_{i-1}, x_i)|$$

$$\le \sum_{i=1}^{n} |h(t_i)| L |t_i - t_{i-1}| + \|g\|_\infty \sum_{i=1}^{n} |h(t_i) - h(t_{i-1})|$$
$$\le L(b-a)\|h\|_\infty + \|g\|_\infty \|h\|_{\mathrm{TV}}.$$

The claim follows since $\|h\|_\infty \le |h(a)| + \|h\|_{\mathrm{TV}} < \infty$. □

PROOF OF THEOREM 3. Let $\gamma \in \mathscr{L}$ be such that $\mathsf{P}(\gamma) = \inf_{\tilde{\gamma}\in\mathscr{L}} \mathsf{P}(\tilde{\gamma})$. We denote by $\Phi(t,x) = \Phi^\gamma(t,x)$ the corresponding solution of the Parisi PDE, as constructed in Section 6, and fix $t_* \in [0,1)$. We apply Theorem 2 whereby u, v are defined as follows for $t \in [0, t_*]$:

$$v(t,x) \equiv \xi''(t)\gamma(t)\partial_x\Phi(t,x), \qquad u(t,x) \equiv \partial_x^2\Phi(t,x). \tag{6.20}$$

For $t \in (t_*, 1]$, we simply set $v(t,x) = v(t_*,x)$, $u(t,x) = u(t_*,x)$. Notice that this choice is immaterial since the algorithm of Theorem 2 never uses $v(t,x)$, $u(t,x)$ for $t > t_*$. We define $(X_t)_{t\in[0,1]}$ by solving the SDE (2.2), which coincides, for $t \in [0,t_*]$ with the SDE (6.3).

We next check that these choices satisfy Assumption 1. Notice that, by construction, it is sufficient to consider $t \in [0, t_*]$.

(A1) v is bounded, since $\|\partial_x\Phi\|_\infty \le 1$ by Lemma 6.2 and, therefore, for $t \in [0,t_*]$, $x \in \mathbb{R}$, $|v(t,x)| \le \|\xi''\gamma\|_{\mathrm{TV}[0,t_*]} < \infty$. Further, u is bounded because $\|\partial_x^2\Phi(t,\cdot)\|_\infty \le C(t_*)$ for almost all $t \le t_*$ (by Lemma 6.4.(a)), and that we can choose a representative of $\partial_x^2\Phi$, which is continuous in time by Lemma 6.4(b).

(A2, 3) v is Lipschitz continuous in space, because $|v(t,x_1) - v(t,x_2)| \le \xi''\gamma(t) \times \|\partial_x^2\Phi(t,\cdot)\|_\infty |x_1 - x_2| \le \|\xi''\gamma\|_{\mathrm{TV}[0,t_*]} C(t_*)|x_1 - x_2| \le C'(t_*)|x_1 - x_2|$ where we used the fact that $\|\partial_x^2\Phi(t,\cdot)\|_\infty \le C(t_*)$ for almost all $t \le t_*$ (by Lemma 6.4(a)), and that we can choose a representative of $\partial_x^2\Phi$, which s continuous in time by Lemma 6.4(b).

Analogously, u is Lipschitz continuous in space, because $|u(t,x_1) - u(t,x_2)| \le \|\partial_x^3\Phi(t,\cdot)\|_\infty |x_1 - x_2|$, and using Lemma 6.4.

(A4) v has bounded strong total variation by applying Lemma 6.14. Indeed $\xi''\gamma$ has bounded total variation on $[0,t_*]$, and $\partial_x\Phi$ is bounded by Lemma 6.2 and Lipschitz by Lemma 6.4 as discussed above.

Further, u has bounded strong total variation because $\partial_x^2\Phi$ is Lipschitz continuous on $[0,t_*] \times \mathbb{R}$, again by Lemma 6.4.

Let us next check the other assumptions in Theorem 2. By Lemma 6.5, we have $M_{t_*} = \partial_x\Phi(t_*, X_{t_*})$ and, therefore, using Lemma 6.2, $|M_{t_*}| \le 1$ almost surely.

Further, $\mathbb{E}[M_t^2] = \mathbb{E}[\partial_x\Phi(t,X_t)^2] = t$ by Corollary 6.10 and Theorem 5.

We are left with the task of computing the value achieved by the algorithm. By Theorem 2, this is given by

$$\mathscr{E}(u,v) = \int_0^{t_*} \xi''(t)\mathbb{E}\big[u(t,X_t)\big]\,\mathrm{d}t = \int_0^{t_*} \xi''(t)\mathbb{E}\big[\partial_x^2\Phi(t,X_t)\big]\,\mathrm{d}t. \tag{6.21}$$

Define $\Psi : [0,1) \times \mathbb{R} \to \mathbb{R}$ by $\Psi(t,x) = \Phi(t,x) - x\partial_x\Phi(t,x)$. By Lemma 6.4, we can assume this to be continuous, and hence $\lim_{t\to 0}\mathbb{E}\Psi(t,X_t) = \mathbb{E}\Psi(0,X_0) = \Phi(0,0)$. We therefore get, using Lemma 6.13,

$$\Phi(0,0) = \mathbb{E}\Psi(t_*, X_{t_*}) + \frac{1}{2}\int_0^{t_*} \xi''(t)\gamma(t)\mathbb{E}\big\{\partial_x\Phi(t,X_t)^2\big\}\,\mathrm{d}t + \int_0^{t_*} \xi''(t)\mathbb{E}\big\{\partial_x^2\Phi(t,X_t)\big\}\,\mathrm{d}t.$$

Comparing this with equation (6.21), we get

$$\begin{aligned}\mathsf{P}(\gamma) - \mathscr{E}(u, v) &= \mathbb{E}\Psi(t_*, X_{t_*}) + \frac{1}{2}\int_0^{t_*} \xi''(t)\gamma(t)\big(\mathbb{E}\{\partial_x \Phi(t, X_t)^2\} - t\big)\,\mathrm{d}t \\ &= \mathbb{E}\Psi(t_*, X_{t_*}),\end{aligned}$$

where in the second step we used Corollary 6.10 and Theorem 5.

The proof is completed by showing that we obtain $\mathsf{P}(\gamma) - \mathscr{E}(u, v) = \mathbb{E}\Psi(t_*, X_{t_*}) \le \varepsilon$ by taking t_* close enough to one. In order to show this, recall that $\Phi(t, \cdot)$ is convex, so $\Phi(t, x) - x\partial_x \Phi(t, x) \le \Phi(t, 0)$. Moreover, $|\partial_x \Phi(t, x)| \le 1$. Whence

$$\Phi(t, 0) - |x| \le \Psi(t, x) \le \Phi(t, 0).$$

Notice that $\Phi(t, 0) \to 0$ as $t \to 1$ (because Φ is continuous on $[0, 1] \times \mathbb{R}$, and $\Phi(1, x) = |x|$) and, therefore,

$$\limsup_{t_*\to 1} \mathbb{E}\Psi(t_*, X_{t_*}) = \limsup_{t_*\to 1} \mathbb{E}\{\Psi(t_*, X_{t_*})\} - \Phi(t, 0) \le 0. \qquad \square$$

6.3. *Proof of Corollary* 2.2. The key tool is provided by the following lemma, which is a variant of Corollary 6.10, and of results from earlier literature (the difference being that we focus on the zero-temperature case).

LEMMA 6.15. *Assume the no-overlap gap assumption to hold for the mixture* ξ; *namely, there exists* $\gamma_* \in \mathscr{U}$ *strictly increasing in* $[0, 1)$ *such that* $\mathsf{P}(\gamma_*) = \inf_{\gamma\in\mathscr{U}} \mathsf{P}(\gamma)$. *Then, for any* $t \in [0, 1)$,

$$\mathbb{E}\{\partial_x \Phi^{\gamma_*}(t, X_t)^2\} = t. \tag{6.22}$$

PROOF. Fix $0 < t_1 < t_2 < 1$, and define $\delta(t) = [\gamma_*(t_1) - \gamma_*(t)]\mathbb{I}_{(t_1,t_2)}(t)$. It is easy to see that this satisfies the assumptions of Proposition 6.8, with $s_0 = 1$, whence letting $\gamma^s = \gamma_* + s\delta$,

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}s}(\gamma^s)|_{s=0+} = -\frac{1}{2}\int_{t_1}^{t_2} \xi''(t)\big(\gamma_*(t) - \gamma_*(t_1)\big)\big(\mathbb{E}\{\partial_x \Phi(t, X_t)^2\} - t\big)\,\mathrm{d}t.$$

(Here, $\Phi = \Phi^{\gamma_*}$.) On the other hand, $\gamma^s \in \mathscr{U}$ for $s \in [0, 1]$ (since γ_* is strictly increasing), whence

$$\int_{t_1}^{t_2} \xi''(t)\big(\gamma_*(t) - \gamma_*(t_1)\big)\big(\mathbb{E}\{\partial_x \Phi(t, X_t)^2\} - t\big)\,\mathrm{d}t \le 0.$$

for all $t_1 < t_2$. Since $\gamma_*(t) - \gamma_*(t_1) > 0$ strictly for all $t > t_1$, this implies $\mathbb{E}\{\partial_x \Phi(t, X_t)^2\} - t \le 0$ for almost every t and, therefore, for every t by Lemma 6.7.

The $\mathbb{E}\{\partial_x \Phi(t, X_t)^2\} - t \ge 0$ is proved in the same way, by using $\delta(t) = [\gamma_*(t_2) - \gamma_*(t)]\mathbb{I}_{(t_1,t_2)}(t)$. $\square$

Let γ_* be a strictly increasing minimizer of $\mathsf{P}(\cdot)$ in $\mathscr{U}$; namely, $\mathsf{P}(\gamma_*) = \inf_{\gamma\in\mathscr{U}} \mathsf{P}(\gamma)$. We claim that γ_* minimizes $\mathsf{P}(\cdot)$ over the larger space $\mathscr{L}$, that is, $\mathsf{P}(\gamma_*) = \inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma)$, thus proving the corollary.

By the last lemma, γ_* verifies the stationarity condition (6.22). Since $\mathsf{P} : \mathscr{L} \to \mathbb{R}$ is convex (this follows by exactly the same proof as [28], Theorem 20), the function $s \mapsto \mathsf{P}((1 - s)\gamma_* + s\gamma)$ is convex over the interval $[0, 1]$ for any $\gamma \in \mathscr{L}$, whence

$$\begin{aligned}\mathsf{P}(\gamma) - \mathsf{P}(\gamma_*) &\ge \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}s}(\gamma_* + s(\gamma - \gamma_*))|_{s=0} \\ &= \frac{1}{2}\int_0^1 \xi''(t)\big(\gamma(t) - \gamma_*(t)\big)\big(\mathbb{E}\{\partial_x \Phi^{\gamma_*}(t, X_t)^2\} - t\big)\,\mathrm{d}t = 0.\end{aligned}$$

We thus conclude that γ_* minimizes P over $\mathscr{L}$.

7. Proof of Theorem 4.

7.1. *A candidate solution.* We produce a solution to HJB (4.7) via a change of variables by taking the Legendre transform of the solution to the Parisi PDE (1.5), which we redisplay here:

$$\partial_t \Phi^\gamma(t,x) + \frac{1}{2}\xi''(t)\big(\partial_x^2 \Phi^\gamma(t,x) + \gamma(t)\big(\partial_x \Phi^\gamma(t,x)\big)^2\big) = 0, \quad (t,x) \in [0,1) \times \mathbb{R},$$
$$\Phi^\gamma(1,x) = |x|, \quad x \in \mathbb{R}. \tag{7.1}$$

We remind the reader that it suffices to consider $\gamma \in \mathsf{SF}_+$ by the approximation argument in Section 4. Since γ is piecewise constant, the PDE (7.1) can be solved via the Cole–Hopf transform and the solution is highly regular in space as shown in Proposition 6.1. We define (the negative of) the Legendre transform of Φ^γ as

$$\Phi^{*,\gamma}(t,z) \equiv \inf_{x\in\mathbb{R}} \{\Phi^\gamma(t,x) - xz\},$$

and define a candidate solution to HJB as

$$V(t,z) \equiv \Phi^{*,\gamma}(t,z) - \frac{1}{2}\nu(t)z^2 - \frac{1}{2}\int_t^1 \nu(s)\,\mathrm{d}s, \tag{7.2}$$

where we recall that $\nu(t) = \int_t^1 \xi''(s)\gamma(s)\,\mathrm{d}s$.

PROPOSITION 7.1. *For all* $(t,z) \in [0,1] \times (-1,1)$, $\mathcal{J}_\gamma(t,z) = V(t,z)$, *where* $\mathcal{J}_\gamma$ *is defined in* (4.4).

In particular, the value at $(0,0)$ is

$$\begin{aligned}\mathcal{J}_\gamma(0,0) &= \inf_x \Phi^\gamma(0,x) - \frac{1}{2}\int_0^1 \nu(s)\,\mathrm{d}s \\ &= \Phi^\gamma(0,0) - \frac{1}{2}\int_0^1 s\xi''(s)\gamma(s)\,\mathrm{d}s = \mathsf{P}(\gamma).\end{aligned}$$

The second equality follows since $\Phi^\gamma(t,\cdot)$ is convex and even. This proves Proposition 4.1.

7.2. *Verification.* We dedicate this section to the proof of Proposition 7.1. We collect in the next lemma the regularity properties of Φ^γ which will be used in what follows.

LEMMA 7.2. *For* $\gamma \in \mathsf{SF}_+$, *we have the following*:

(a) $\partial_x^j \Phi^\gamma \in C([0,1) \times \mathbb{R})$ *for all* $j \geq 0$.

(b) $\partial_t \partial_x^j \Phi^\gamma \in C([a,b) \times \mathbb{R})$ *for all* $j \geq 0$ *and for any interval* $[a,b)$ *on which* γ *is constant*.

Further, for all $t \in [0,1)$:

(c) *The range of the map* $x \mapsto \partial_x \Phi^\gamma(t,x)$ *is the open interval* $(-1,1)$. *In particular,* $|\partial_x \Phi^\gamma| < 1$.

(d) $\partial_x \Phi^\gamma(t,\cdot)$ *is strictly increasing*.

(e) *For all* $x \in \mathbb{R}$, $0 < \partial_x^2 \Phi^\gamma(t',x) \leq C(t,\gamma)$ *for all* $t' \in [0,t]$ *and some constant* $C(t,\gamma) < \infty$.

PROOF. Set $\Phi = \Phi^\gamma$. All of these claims can be proved by direct calculus using the explicit expression for the Cole–Hopf solution. Given $\gamma(t) = \sum_{i=1}^m \gamma_i \mathbb{I}_{[t_{i-1},t_i)}$, $0 = t_0 < t_1 < \cdots < t_m = 1$, we let $r(t) = \xi'(1) - \xi'(t)$. The Cole–Hopf solution is then constructed recursively as follows. For each $i \in \{1, \dots, m\}$ and each $t \in [t_{i-1}, t_i)$, let

$$\Phi(t,x) = \frac{1}{\gamma_i} \log \mathbb{E} \exp\{\gamma_i \Phi(t_i, x + \sqrt{r(t) - r(t_i)} G)\} G \sim \mathsf{N}(0,1), \tag{7.3}$$

(with $\Phi(t,x) = |x|$). Claims (a), (b) follow by standard properties of convolutions (they are also a special case of Lemma 6.4).

Claim (c), (d), (e) can be proved by differentiating (7.3). For $t \in [t_{i-1}, t_i)$, define $\mathsf{P}_{t,x}$ to the probability distribution with density

$$\mathsf{p}_{t,x}(x') \equiv \frac{1}{\mathbb{E}\{e^{\gamma_i \Phi(t_i, x + \sqrt{r(t_i) - r(t)} G)}\}} \exp\left\{-\frac{(x'-x)^2}{2(r(t) - r(t_i))} + \gamma_i \Phi(t_i, x')\right\}.$$

Let $\mathsf{E}_{t,x}$, and $\mathsf{Var}_{t,x}$ denote expectation and variance with respect to this density. Consider first $t \in [t_{m-1}, t_m = 1)$,

$$\partial_x \Phi(t,x) = \mathsf{E}_{t,x} \operatorname{sign}(X),$$

$$\partial_x^2 \Phi(t,x) = 2\mathsf{p}_{t,x}(0) + \gamma_m\{1 - \mathsf{E}_{t,x}(\operatorname{sign}(X))^2\}.$$

The last expression yields $0 < \partial_x^2 \Phi(t,x) < C(t_*, \gamma)$ for all $t < t_* < 1$ (notice indeed that $p_{t,x}(0)$ is bounded and nonnegative for all $t < t_*$), which is claim (e). In particular, this implies that $x \mapsto \partial_x \Phi(t,x)$ is strictly increasing (claim (d)). Further, $|\partial_x \Phi(t,x)| < 1$, because $p_{t,x}$ is strictly positive mass on $(-\infty, 0)$ and on $(0, +\infty)$. Finally, $\lim_{x\to\pm\infty} \partial_x \Phi(t,x) = \pm 1$ because $\mathsf{P}_{t,x}((-\infty, a]) \to 0$ for all $a \in \mathbb{R}$ as $x \to +\infty$, $\mathsf{P}_{t,x}([a, +\infty]) \to 0$ for all $a \in \mathbb{R}$ as $x \to -\infty$.

Next, for $t \in [t_{i-1}, t_i)$, $i < m$, we have

$$\partial_x \Phi(t,x) = \mathsf{E}_{t,x} \partial_x \Phi(t_i, X),$$

$$\partial_x^2 \Phi(t,x) = \mathsf{E}_{t,x} \partial_x^2 \Phi(t_i, X) + \gamma_i \operatorname{Var}_{t,x}(\partial_x \Phi(t_i, X)),$$

Claims (c)–(e) are proved buy induction using arguments similar to the above. In particular, if $0 < \partial_x^2 \Phi(t_i, x) < C_{i+1}$ the last equation implies $0 < \partial_x^2 \Phi(t,x) < C_{i+1} + \gamma_i$ for $t \in [t_{i-1}, t_i)$. □

LEMMA 7.3. *For $\gamma \in \mathsf{SF}_+$, the function V defined in equation (7.2) is a solution to the HJB equation (4.7) on $[0,1] \times (-1,1)$.*

PROOF. First, since $\nu(1) = 0$, it is clear that V satisfies the terminal condition $V(1,z) = 0$ for $|z| < 1$. Next, let $t < 1$. Since $\Phi^\gamma(t, \cdot)$ is twice continuously differentiable and strictly convex, there exists a continuous strictly increasing map $z \in (-1,1) \mapsto x_t^*(z)$ defined as the unique root x of the equation $\partial_x \Phi^\gamma(t,x) = z$. Furthermore, the envelope theorem implies that $\partial_z \Phi_\gamma^*(t,z) = -x_t^*(z)$ and $\partial_z^2 \Phi_\gamma^*(t,z) = -1/\partial_x^2 \Phi^\gamma(t, x_t^*(z))$ for all $z \in (-1,1)$.

Exploiting equation (7.2), we have

$$\partial_t V(t,z) = \partial_t \Phi^\gamma(t, x_t^*(z)) + \frac{1}{2}\xi''(t)\gamma(t) z^2 + \frac{1}{2}\nu(t),$$

$$\partial_z^2 V(t,z) = -\frac{1}{\partial_x^2 \Phi^\gamma(t, x_t^*(z))} - \nu(t).$$

Given that Φ^γ satisfies the Parisi PDE, we have for all $z \in (-1, 1)$

$$\partial_t V(t,z) - \frac{1}{2}\xi''(t)\gamma(t)z^2 - \frac{1}{2}\nu(t) + \frac{\xi''(t)}{2}\left(\gamma(t)z^2 - \frac{1}{\partial_z^2 V(t,z) + \nu(t)}\right) = 0.$$

Simplifying the quadratic term in z, we obtain

$$\partial_t V(t,z) - \frac{1}{2}\nu(t) - \frac{\xi''(t)}{2(\partial_z^2 V(t,z) + \nu(t))} = 0.$$

Since $\partial_x^2 \Phi^\gamma > 0$, we have $\partial_z^2 V(t,z) + \nu(t) < 0$ hence

$$\sup_{\lambda \in \mathbb{R}} \left\{\lambda + \frac{\lambda^2}{2}\left(\nu(t) + \partial_z^2 V(t,z)\right)\right\} = -\frac{1}{2(\partial_z^2 V(t,z) + \nu(t))}.$$

Therefore, V is a solution to HJB (4.7) on $[0, 1) \times (-1, 1)$ with the right-terminal condition at $t = 1$, for any function $\gamma \in \mathsf{SF}_+$. $\square$

PROOF OF PROPOSITION 7.1. We closely follow the proof of Theorem 4.1 in the textbook [48]. We recall the expression of $\mathcal{J}_\gamma$:

$$\tag{7.4} \begin{aligned} \mathcal{J}_\gamma(t,z) \equiv \sup_{u \in D[t,1]} \mathbb{E}\left[\int_t^1 \xi''(s)u_s \,\mathrm{d}s + \frac{1}{2}\int_t^1 \nu(s)\left(\xi''(s)u_s^2 - 1\right)\mathrm{d}s\right], \\ \text{s.t. } z + \int_t^1 \sqrt{\xi''(s)}u_s \,\mathrm{d}B_s \in (-1,1) \text{ a.s.,} \end{aligned}$$

where $\nu(t) \equiv \int_t^1 \xi''(s)\gamma(s)\,\mathrm{d}s$.

Let us first prove the bound $V \geq \mathcal{J}_\gamma$. Lemma 7.2 implies that $V \in C^{1,2}([a,b) \times (-1,1))$ whenever γ is constant on $[a, b)$.

We momentarily assume that γ is constant on $[0, 1]$. Let $(t, z) \in [0, 1) \times (-1, 1)$, and let $(u_s)_{s\geq t} \in D[t, 1]$. Consider the process M^u defined by $\mathrm{d}M_s^u = \sqrt{\xi''(s)}u_s \,\mathrm{d}B_s$, $s \geq t$ with initial condition $M_t^u = z$, and recall that $M_1^u = z + \int_t^1 \sqrt{\xi''(s)}u_s \,\mathrm{d}B_s \in (-1, 1)$ a.s. Since $(M_s^u)_{s\geq t}$ is a martingale (w.r.t. the filtration of Brownian motion $\mathcal{F}_t$), we have $M_t^u = \mathbb{E}[M_1^u|\mathcal{F}_t]$ and, therefore, $M_s^u \in (-1, 1)$ for all $s \in [t, 1]$ a.s.

By Itô's formula we have for $t \leq \theta < 1$,

$$\tag{7.5} \begin{aligned} &\mathbb{E}_{t,z}\left[V\left(\theta, M_\theta^u\right)\right] - V(t,z) \\ &= \mathbb{E}_{t,z}\int_t^\theta \left(\partial_z V\left(s, M_s^u\right) + \frac{1}{2}\xi''(s)u_s^2 \partial_z^2 V\left(s, M_s^u\right)\right)\mathrm{d}s \\ &\leq \mathbb{E}_{t,z}\int_t^\theta \left(\partial_t V\left(s, M_s^u\right) + \xi''(s)\sup_{u\in\mathbb{R}}\left\{u + \frac{u^2}{2}\left(\nu(s) + \partial_z^2 V\left(s, M_s^u\right)\right)\right\}\right)\mathrm{d}s \\ &\quad - \mathbb{E}_{t,z}\int_t^\theta \left(\xi''(s)u_s + \frac{1}{2}\xi''(s)\nu(s)u_s^2\right)\mathrm{d}s \\ &= \mathbb{E}_{t,z}\int_t^\theta \left(\frac{1}{2}\nu(s) - \xi''(s)u_s - \frac{1}{2}\xi''(s)\nu(s)u_s^2\right)\mathrm{d}s. \end{aligned}$$

The first inequality follows by taking a supremum over $u_s \in \mathbb{R}$, and the last equality follows since V is a solution to HJB (4.7) as shown in Lemma 7.3.

Next, we have $\mathbb{E}[V(\theta, M_\theta^u)] \to 0$ as $\theta \to 1$. Indeed notice that M^u is continuous, $M_1^u \in (-1, 1)$ almost surely and $V(\theta, x)$ is continuous on $[0, 1] \times (-1, 1)$. Therefore, for $W_\theta \equiv V(\theta, M_\theta^u)$, we have $W_\theta \to W_1 = 0$ almost surely as $\theta \to 1$. Further, we claim that W_θ

is bounded, whence the claim $\mathbb{E}[W_\theta] = \mathbb{E}[V(\theta, M^u_\theta)] \to \mathbb{E}[W_1] = 0$ follows by dominated convergence. In order to show that W_θ is bounded, note that $\Phi^\gamma(t,x) \geq |x|$ for $t \in [0,1]$ by equation (7.3) and Jensen inequality. This implies that $0 \leq \Phi^*_\gamma(t,z) \leq \Phi^\gamma(t,0)$ and, therefore, $V(\theta,z)$ bounded in $[0,1] \times (-1,1)$.

Since $u \in L^1 \cap L^2$, we obtain

$$V(t,z) \geq \mathbb{E}_{t,z} \int_t^1 \left(\frac{1}{2}\nu(s)\big(\xi''(s)u_s^2 - 1\big) + \xi''(s)u_s\right) \mathrm{d}s,$$

for all processes $u \in D[t,1]$ satisfying $M^u_1 \in (-1,1)$ a.s. Therefore, $V(t,z) \geq \mathcal{J}_\gamma(t,z)$.

Returning to the general case, if γ has $0 < t_1 < \cdots < t_m < 1$ points of discontinuity then Itô's formula and the above argument can be applied inside every interval $[t_i, \theta_i]$ with $\theta_i < t_{i+1}$. Letting $\theta_i \to t_{i+1}$ and applying the dominated convergence theorem, then summing over i, the left-hand side in equation (7.5) telescopes and we obtain the desired result.

Now we show the converse bound. Fix $(t,z) \in [0,1) \times (-1,1)$ and consider the control process

$$u^*_s = \partial_x^2 \Phi^\gamma(s, X_s) \quad \text{for } s \in [t,1), \quad \text{and} \quad u^*_1 = 0,$$

where $(X_s)_{s \geq t}$ solves the SDE

$$\mathrm{d}X_s = \xi''(s)\gamma(s)\partial_x \Phi^\gamma(s,X_s)\,\mathrm{d}s + \sqrt{\xi''(s)}\,\mathrm{d}B_s,$$

with initial condition $X_t = x$. This is the same SDE as in equation (2.2) with drift $v(t,x) = \xi''(t)\gamma(t)\partial_x\Phi^\gamma(t,x)$, which is bounded and Lipschitz in space for $\gamma \in \mathsf{SF}_+$, therefore, a strong solution exists. Further, since $\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}[\partial_x\Phi^\gamma(s,X_s)^2] = \xi''(s)\mathbb{E}[\partial_x^2\Phi^\gamma(s,X_s)^2]$ (Corollary 6.6) and $|\partial_x\Phi^\gamma| \leq 1$ then u^* is an admissible control on $[t,1]$: $u^* \in D[t,1]$.

Legendre duality implies that u^*_s can also be written as

$$u^*_s = -\frac{1}{(\partial_z^2 V(s, M^*_s) + \nu(s))}, \quad \text{with } M^*_s = \partial_x\Phi^\gamma(s,X_s).$$

Since Φ^γ is a solution to the Parisi PDE, an application of Itô's formula reveals that M^* is a martingale, which is represented by the stochastic integral

$$\mathrm{d}M^*_s = \sqrt{\xi''(s)}\partial_x^2\Phi^\gamma(s,X_s)\,\mathrm{d}B_s = \sqrt{\xi''(t)}u^*_s\,\mathrm{d}B_s,$$

with initial condition $M^*_t = \partial_x\Phi^\gamma(t,x)$. Further, observe that $|M^*_1| \leq 1$ a.s. and that by surjectivity of $\partial_x\Phi^\gamma(t,\cdot)$, we can choose x such that $M^*_t = z$. We repeat the above execution of Itô's formula with M^* and u^* replacing M^u and u, respectively. We see that the crucial step (7.5) holds with equality, as u^*_s achieves the supremum displayed inside the integral. Hence equality $V(t,z) = \mathcal{J}_\gamma(t,z)$, and this conclude our proof. □

Funding. This work was partially supported by NSF Grants CCF-1714305, IIS-1741162 and ONR N00014-18-1-2729.

SUPPLEMENTARY MATERIAL

Supplement to "Optimization of mean-field spin glasses" (DOI: 10.1214/21-AOP1519SUPP; .pdf). Contains omitted proofs.

REFERENCES

[1] ADDARIO-BERRY, L. and MAILLARD, P. (2019). The algorithmic hardness threshold for continuous random energy models. *Math. Stat. Learn.* **2** 77–101. MR4073148 https://doi.org/10.4171/msl/12

[2] ARORA, S., BERGER, E., ELAD, H., KINDLER, G. and SAFRA, M. (2005). On non-approximability for quadratic programs. In *FOCS* 2005. 46*th Annual IEEE Symposium on*. *Foundations of Computer Science*, 2005 206–215. IEEE, Los Alamitos.

[3] AUFFINGER, A., BEN AROUS, G. and ČERNÝ, J. (2013). Random matrices and complexity of spin glasses. *Comm. Pure Appl. Math.* **66** 165–201. MR2999295 https://doi.org/10.1002/cpa.21422

[4] AUFFINGER, A. and CHEN, W.-K. (2015). The Parisi formula has a unique minimizer. *Comm. Math. Phys.* **335** 1429–1444. MR3320318 https://doi.org/10.1007/s00220-014-2254-z

[5] AUFFINGER, A. and CHEN, W.-K. (2017). Parisi formula for the ground state energy in the mixed p-spin model. *Ann. Probab.* **45** 4617–4631. MR3737919 https://doi.org/10.1214/16-AOP1173

[6] BARAK, B., BRANDÃO, F. G. S. L., HARROW, A. W., KELNER, J., STEURER, D. and ZHOU, Y. (2012). Hypercontractivity, sum-of-squares proofs, and their applications [extended abstract]. In *STOC'*12—*Proceedings of the* 2012 *ACM Symposium on Theory of Computing* 307–326. ACM, New York. MR2961513 https://doi.org/10.1145/2213977.2214006

[7] BAYATI, M., LELARGE, M. and MONTANARI, A. (2015). Universality in polytope phase transitions and message passing algorithms. *Ann. Appl. Probab.* **25** 753–822. MR3313755 https://doi.org/10.1214/14-AAP1010

[8] BAYATI, M. and MONTANARI, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theory* **57** 764–785. MR2810285 https://doi.org/10.1109/TIT.2010.2094817

[9] BERNT, O. (2013). *Stochastic Differential Equations*: *An Introduction with Applications*. Springer, Berlin.

[10] BERTHIER, R., MONTANARI, A. and NGUYEN, P.-M. (2020). State evolution for approximate message passing with non-separable functions. *Inf. Inference* **9** 33–79. MR4079177 https://doi.org/10.1093/imaiai/iay021

[11] BHATTIPROLU, V., GURUSWAMI, V. and LEE, E. (2017). Sum-of-squares certificates for maxima of random tensors on the sphere. In *Approximation*, *Randomization*, *and Combinatorial Optimization*. *Algorithms and Techniques*. *LIPIcs*. *Leibniz Int*. *Proc*. *Inform*. **81** 31. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR3695598

[12] BOLTHAUSEN, E. (2014). An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Comm. Math. Phys.* **325** 333–366. MR3147441 https://doi.org/10.1007/s00220-013-1862-3

[13] BOUCHAUD, J.-P., CUGLIANDOLO, L. F., KURCHAN, J. and MÉZARD, M. (1998). Out of equilibrium dynamics in spin-glasses and other glassy systems. In *Spin Glasses and Random Fields* 161–223.

[14] CHEN, W.-K. (2013). The Aizenman–Sims–Starr scheme and Parisi formula for mixed p-spin spherical models. *Electron. J. Probab.* **18** 94. MR3126577 https://doi.org/10.1214/EJP.v18-2580

[15] CHEN, W.-K., GAMARNIK, D., PANCHENKO, D. and RAHMAN, M. (2019). Suboptimality of local algorithms for a class of max-cut problems. *Ann. Probab.* **47** 1587–1618. MR3945754 https://doi.org/10.1214/18-AOP1291

[16] CHEN, W.-K., HANDSCHY, M. and LERMAN, G. (2018). On the energy landscape of the mixed even p-spin model. *Probab. Theory Related Fields* **171** 53–95. MR3800830 https://doi.org/10.1007/s00440-017-0773-1

[17] CRISANTI, A., LEUZZI, L. and RIZZO, T. (2003). The complexity of the spherical p-spin spin glass model, revisited. *Eur. Phys. J. B* **36** 129–136.

[18] CRISANTI, A., LEUZZI, L. and RIZZO, T. (2005). Complexity in mean-field spin-glass models: Ising p-spin. *Phys. Rev. B* **71** 094202.

[19] CRISANTI, A. and SOMMERS, H.-J. (1992). The sphericalp-spin interaction spin glass model: The statics. *Z. Phys. B*, *Condens. Matter* **87** 341–354.

[20] CUGLIANDOLO, L. F. and KURCHAN, J. (1994). On the out-of-equilibrium relaxation of the Sherrington–Kirkpatrick model. *J. Phys. A* **27** 5749–5772. MR1304208

[21] EL ALAOUI, A., MONTANARI, A. and SELLKE, M. (2021). Supplement to "Optimization of mean-field spin glasses": Proof of state evolution. https://doi.org/10.1214/21-AOP1519SUPP.

[22] EVANS, L. C. (2009). *Partial Differential Equations*. Springer, Berlin.

[23] FOLENA, G., FRANZ, S. and RICCI-TERSENGHI, F. (2019). Memories from the ergodic phase: The awkward dynamics of spherical mixed p-spin models. Available at arXiv:1903.01421.

[24] GAMARNIK, D. and JAGANNATH, A. (2021). The overlap gap property and approximate message passing algorithms for p-spin models. *Ann. Probab.* **49** 180–205. MR4203336 https://doi.org/10.1214/20-AOP1448

[25] GAMARNIK, D. and SUDAN, M. (2014). Limits of local algorithms over sparse random graphs [extended abstract]. In *ITCS'14—Proceedings of the* 2014 *Conference on Innovations in Theoretical Computer Science* 369–375. ACM, New York. MR3359490

[26] GAMARNIK, D. and SUDAN, M. (2017). Performance of sequential local algorithms for the random NAE-K-SAT problem. *SIAM J. Comput*. **46** 590–619. MR3620150 https://doi.org/10.1137/140989728

[27] GUERRA, F. (2001). Sum rules for the free energy in the mean field spin glass model. *Fields Inst*. *Commun*. **30**.

[28] JAGANNATH, A. and TOBASCO, I. (2016). A dynamic programming approach to the Parisi functional. *Proc*. *Amer. Math*. *Soc*. **144** 3135–3150. MR3487243 https://doi.org/10.1090/proc/12968

[29] JAVANMARD, A. and MONTANARI, A. (2013). State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference* **2** 115–144. MR3311445 https://doi.org/10.1093/imaiai/iat004

[30] KRZĄKAŁA, F., MONTANARI, A., RICCI-TERSENGHI, F., SEMERJIAN, G. and ZDEBOROVÁ, L. (2007). Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **104** 10318–10323. MR2317690 https://doi.org/10.1073/pnas.0703685104

[31] KUNISKY, D. and BANDEIRA, A. S. (2019). A tight degree 4 sum-of-squares lower bound for the Sherrington–Kirkpatrick Hamiltonian. Available at arXiv:1907.11686.

[32] MÉZARD, M. and MONTANARI, A. (2009). *Information*, *Physics*, *and Computation*. *Oxford Graduate Texts*. Oxford Univ. Press, Oxford. MR2518205 https://doi.org/10.1093/acprof:oso/9780198570837.001.0001

[33] MÉZARD, M., PARISI, G. and VIRASORO, M. A. (1987). *Spin Glass Theory and Beyond*. *World Scientific Lecture Notes in Physics* **9**. World Scientific, Teaneck, NJ. MR1026102

[34] MÉZARD, M., PARISI, G. and ZECCHINA, R. (2002). Analytic and algorithmic solution of random satisfiability problems. *Science* **297** 812–815.

[35] MOHANTY, S., RAGHAVENDRA, P. and XU, J. (2019). Lifting sum-of-squares lower bounds: Degree-2 to degree-4. Available at arXiv:1911.01411.

[36] MONASSON, R., ZECCHINA, R., KIRKPATRICK, S., SELMAN, B. and TROYANSKY, L. (1999). Determining computational complexity from characteristic "phase transitions". *Nature* **400** 133–137. MR1704845 https://doi.org/10.1038/22055

[37] MONTANARI, A. (2019). *Optimization of the Sherrington–Kirkpatrick Hamiltonian*, *IEEE Symposium on the Foundations of Computer Science*, *FOCS*.

[38] PANCHENKO, D. (2013). *The Sherrington–Kirkpatrick Model*. Springer, Berlin.

[39] PANCHENKO, D. (2013). The Parisi ultrametricity conjecture. *Ann*. *of Math*. (2) **177** 383–393. MR2999044 https://doi.org/10.4007/annals.2013.177.1.8

[40] PARISI, G. (1979). Infinite number of order parameters for spin-glasses. *Phys*. *Rev*. *Lett*. **43** 1754.

[41] PARISI, G. (2006). Computing the number of metastable states in infinite-range models. In *Mathematical Statistical Physics* 295–329. Elsevier, Amsterdam. MR2581888 https://doi.org/10.1016/S0924-8099(06)80044-2

[42] RICHARD, E. and MONTANARI, A. (2014). A statistical model for tensor pca. *Adv. Neural Inf. Process*. *Syst*. 2897–2905.

[43] SELLKE, M. (2020). Approximate ground states of hypercube spin glasses are near corners. Available at arXiv:2009.09316.

[44] SHERRINGTON, D. and KIRKPATRICK, S. (1975). Solvable model of a spin-glass. *Phys*. *Rev*. *Lett*. **35** 1792.

[45] SUBAG, E. (2021). Following the ground states of full-RSB spherical spin glasses. *Comm*. *Pure Appl*. *Math*. **74** 1021–1044. MR4230065 https://doi.org/10.1002/cpa.21922

[46] TALAGRAND, M. (2006). The Parisi formula. *Ann*. *of Math*. (2) **163** 221–263. MR2195134 https://doi.org/10.4007/annals.2006.163.221

[47] TALAGRAND, M. (2010). *Mean Field Models for Spin Glasses*: *Volume I*. Springer, Berlin.

[48] TOUZI, N. (2012). *Optimal Stochastic Control*, *Stochastic Target Problems*, *and Backward SDE*. *Fields Institute Monographs* **29**. Springer, New York.

[49] ZHANG, X., LING, C. and QI, L. (2012). The best rank-1 approximation of a symmetric tensor and related spherical optimization problems. *SIAM J. Matrix Anal*. *Appl*. **33** 806–821. MR3023452 https://doi.org/10.1137/110835335