2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS) | 978-1-6654-5519-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/FOCS54457.2022.00037

Tight Lipschitz Hardness for Optimizing Mean Field Spin Glasses

Brice Huang
Department of EECS
Massachusetts Institute of Technology
Cambridge, USA
bmhuang@mit.edu

Mark Sellke
Department of Mathematics
Stanford University
Stanford, USA
msellke@stanford.edu

***Abstract*—We study the problem of algorithmically optimizing the Hamiltonian of a spherical or Ising mean field spin glass. The maximum asymptotic value OPT of this random function is characterized by a variational principle known as the Parisi formula, proved first by Talagrand and in more generality by Panchenko. Recently developed approximate message passing algorithms efficiently optimize these functions up to a value ALG given by an extended Parisi formula, which minimizes over a larger space of functional order parameters. These two objectives are equal for spin glasses exhibiting a *no overlap gap* property. However, ALG can be strictly smaller than OPT, and no efficient algorithm producing a value exceeding ALG is known.**

We prove that when all interactions have even degree, no algorithm satisfying an *overlap concentration* property can produce an objective larger than ALG with non-negligible probability. This property holds for all algorithms with suitably Lipschitz dependence on the random disorder coefficients of the objective. It encompasses natural formulations of gradient descent, approximate message passing, and Langevin dynamics run for bounded time and in particular includes the algorithms achieving ALG mentioned above. To prove this result, we substantially generalize the overlap gap property framework introduced by Gamarnik and Sudan to arbitrary ultrametric forbidden structures of solutions.

***Index Terms*—non-convex optimization, statistical physics, spin glass, overlap gap property**

I. Introduction

In a *random optimization problem*, one sets out to optimize an objective function generated from random data. The computational complexity of these problems is not well understood, due to the fact that they are often both non-convex and high-dimensional. Optimizing non-convex functions in high dimensions is well-known to be computationally intractable in the worst case; however, worst-case lower bounds rely on highly structured hard instances, and in average-case settings the picture is far less clear.

In this paper we obtain a sharp computational threshold for a natural class of random optimization problems, namely the Hamiltonians of mean-field spin glasses. These functions have been studied since [67] as models for disordered magnetic systems. From a mathematical point of view, they are simply polynomials or power series in many variables with independent and identically distributed coefficients. Moreover as discussed below they are closely related to random combinatorial optimization problems such as k-SAT and MaxCut. Our main result is a lower bound against a natural class of **stable** algorithms which exactly matches the best known algorithms for this problem.

Our problem is defined as follows. For each $p \in 2\mathbb{N}$, let $\mathbf{G}^{(p)} \in \left(\mathbb{R}^N\right)^{\otimes p}$ be an independent p-tensor with i.i.d. $\mathcal{N}(0,1)$ entries. Let $h \geq 0$ and set $\boldsymbol{h} = (h, \dots, h) \in \mathbb{R}^N$. Fix a sequence $(\gamma_p)_{p\in 2\mathbb{N}}$ with $\gamma_p \geq 0$ and $\sum_{p\in 2\mathbb{N}} 2^p \gamma_p^2 < \infty$.[1] The mixed even p-spin Hamiltonian H_N is defined by

$$H_N(\boldsymbol{\sigma}) = \langle \boldsymbol{h}, \boldsymbol{\sigma}\rangle + \widetilde{H}_N(\boldsymbol{\sigma}), \quad \text{where}$$
$$\widetilde{H}_N(\boldsymbol{\sigma}) = \sum_{p\in 2\mathbb{N}} \frac{\gamma_p}{N^{(p-1)/2}} \langle \mathbf{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p}\rangle. \tag{I.1}$$

We consider inputs $\boldsymbol{\sigma}$ in the sphere $S_N = \{\boldsymbol{\sigma} \in \mathbb{R}^N : \|\boldsymbol{\sigma}\|_2^2 = N\}$ or the cube $\Sigma_N = \{-1, 1\}^N$. These define, respectively, the *spherical* and *Ising* mixed p-spin glass

Brice Huang was supported by an NSF graduate research fellowship, a Siebel scholarship, NSF awards DMS-2022448 and CCF-1940205, and NSF TRIPODS award 1740751. Mark Sellke was supported by an NSF graduate research fellowship, the William R. and Sara Hart Kimball Stanford graduate fellowship, and NSF award CCF-2006489. Part of this work was conducted while the authors were visiting the Simons Institute for the Theory of Computing.

[1] Not much generality is lost by assuming the sequence (γ_p) is eventually zero, in which case H_N is a polynomial in $\boldsymbol{\sigma}$.

978-1-6654-5519-0/22/$31.00 ©2022 IEEE
DOI 10.1109/FOCS54457.2022.00037

models. The coefficients γ_p are customarily encoded in the *mixture function* $\xi(x) = \sum_{p\in 2\mathbb{N}} \gamma_p^2 x^p$. Note that $\widetilde{H}_N$ is equivalently described as the Gaussian process with covariance

$$\mathbb{E}\,\widetilde{H}_N(\boldsymbol{\sigma}^1)\widetilde{H}_N(\boldsymbol{\sigma}^2) = N\xi(\langle \boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2\rangle/N)$$

while the term $\langle \boldsymbol{h}, \boldsymbol{\sigma}\rangle$ represents an external field.

Our purpose is to shed light on a discrepancy between the in-probability limiting maximum values

$$\mathsf{OPT}^{\mathrm{Sp}}_{\xi,h} = \operatorname*{p\text{-}lim}_{N\to\infty} \frac{1}{N} \max_{\boldsymbol{\sigma}\in S_N} H_N(\boldsymbol{\sigma}),$$
$$\mathsf{OPT}^{\mathrm{Is}}_{\xi,h} = \operatorname*{p\text{-}lim}_{N\to\infty} \frac{1}{N} \max_{\boldsymbol{\sigma}\in \Sigma_N} H_N(\boldsymbol{\sigma})$$

and the maximum *efficiently computable* values of H_N over the same sets. We will write $\mathsf{OPT}^{\mathrm{Sp}} = \mathsf{OPT}^{\mathrm{Sp}}_{\xi,h}$ and $\mathsf{OPT}^{\mathrm{Is}} = \mathsf{OPT}^{\mathrm{Is}}_{\xi,h}$ when ξ, h are clear from context.

The investigation of average-case computational thresholds, or information-computation gaps, has a substantial history in computer science. These thresholds have been studied in random optimization problems such as random instances of k-SAT and MaxCut, as well as in signal detection problems such as planted clique and sparse PCA.

A. $\mathsf{OPT}, \mathsf{ALG}$, *and the Parisi Functional*

The values $\mathsf{OPT}^{\mathrm{Sp}}$ and $\mathsf{OPT}^{\mathrm{Is}}$ are given by the celebrated Parisi formula [60] which was proved for even models by [71], [72] and in more generality by [58]. While most often stated as a formula for the limiting free energy at inverse temperature β, the asymptotic maximum can be recovered as a $\beta\to\infty$ limit of the Parisi formula. Restricting for concreteness to the Ising case (we will state the analogous result for the spherical case in Section II), the result can be expressed in the following form due to Auffinger and Chen [13]. Define the function space

$$\mathscr{U} = \left\{ \begin{array}{l} \zeta : [0,1) \to \mathbb{R}_{\geq 0} \text{ right-continuous} \\ \text{and nondecreasing}, \int_0^1 \zeta(t)\mathrm{d}t < \infty \end{array} \right\}. \quad \text{(I.2)}$$

For $\zeta\in\mathscr{U}$, define $\Phi_\zeta : [0,1]\times\mathbb{R}\to\mathbb{R}$ to be the solution of the *Parisi PDE*

$$\partial_t \Phi_\zeta(t,x) = -\frac{\xi''(t)}{2}\left(\partial_{xx}\Phi_\zeta(t,x) + \zeta(t)(\partial_x\Phi_\zeta(t,x))^2\right)$$

with terminal condition $\Phi_\zeta(1,x) = |x|$. Existence and uniqueness properties for this PDE are established in [13], [48]. The Parisi functional $\mathsf{P}^{\mathrm{Is}} = \mathsf{P}^{\mathrm{Is}}_{\xi,h} : \mathscr{U}\to\mathbb{R}$ is given by

$$\mathsf{P}^{\mathrm{Is}}(\zeta) = \Phi_\zeta(0,h) - \frac{1}{2}\int_0^1 t\xi''(t)\zeta(t)\mathrm{d}t. \quad \text{(I.3)}$$

Theorem 1 ([13, Theorem 1])**.** *The following identity holds.*

$$\mathsf{OPT}^{\mathrm{Is}} = \inf_{\zeta\in\mathscr{U}} \mathsf{P}^{\mathrm{Is}}(\zeta). \quad \text{(I.4)}$$

The infimum over $\zeta\in\mathscr{U}$ is achieved at a unique $\zeta_*\in\mathscr{U}$ as shown in [13], [26], which can be obtained as an appropriately renormalized zero-temperature limit of the corresponding minimizers in the positive temperature Parisi formula. These positive temperature minimizers roughly correspond to cumulative distribution functions for the overlap $\langle \boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2\rangle/N$ of two replicas $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2$ sampled from the Gibbs measure $e^{\beta H_N}/Z_N(\beta)$; this is why the functions ζ considered in the Parisi formula are nondecreasing.

Efficient algorithms to find an input $\boldsymbol{\sigma}$ achieving a large objective have recently emerged in a line of work initiated by [69] and continued in [7], [55], [66]. The main results of these works in the Ising case can be described as follows. For a function $f:\mathbb{R}\to\mathbb{R}$ and interval J, let $\|f\|_{\mathrm{TV}(J)}$ denote the total variation of f on J, expressed as the supremum over partitions:

$$\|f\|_{\mathrm{TV}(J)} = \sup_n \sup_{t_0<t_1<\cdots<t_n, t_i\in J} \sum_{i=1}^n |f(t_i) - f(t_{i-1})|.$$

Let $\mathscr{L}\supseteq\mathscr{U}$ denote the set of functions given by

$$\mathscr{L} = \left\{ \begin{array}{l} \zeta : [0,1) \to \mathbb{R}_{\geq 0} \text{ right-continuous}, \\ \|\xi''\cdot\zeta\|_{\mathrm{TV}[0,t]} < \infty \ \forall t\in[0,1), \\ \int_0^1 \xi''(t)\zeta(t)\mathrm{d}t < \infty \end{array} \right\}. \quad \text{(I.5)}$$

Intuitively, one may just think of $\mathscr{L}$ as containing of all reasonable and possibly non-increasing functions $\zeta : [0,1)\to\mathbb{R}_{\geq 0}$. The above definition of P^{Is} extends from $\mathscr{U}$ to $\mathscr{L}$. Therefore we may define $\mathsf{ALG}^{\mathrm{Is}} = \mathsf{ALG}^{\mathrm{Is}}_{\xi,h}$ by

$$\mathsf{ALG}^{\mathrm{Is}} = \inf_{\zeta\in\mathscr{L}} \mathsf{P}^{\mathrm{Is}}(\zeta). \quad \text{(I.6)}$$

Note that $\mathsf{ALG}^{\mathrm{Is}} \leq \mathsf{OPT}^{\mathrm{Is}}$ trivially holds. We have $\mathsf{ALG}^{\mathrm{Is}} = \mathsf{OPT}^{\mathrm{Is}}$ if the infimum in (I.6) is attained by some $\zeta\in\mathscr{U}$, and otherwise $\mathsf{ALG}^{\mathrm{Is}} < \mathsf{OPT}^{\mathrm{Is}}$.

Theorem 2 ([7], [66])**.** *Assume there exists $\zeta_*\in\mathscr{L}$ such that $\mathsf{P}^{\mathrm{Is}}(\zeta_*) = \mathsf{ALG}^{\mathrm{Is}}$. Then for any $\varepsilon > 0$, there exists an efficient algorithm $\mathcal{A} : \mathcal{H}_N \to C_N$ such that for some $c = c(\varepsilon) > 0$,*

$$\mathbb{P}[H_N(\mathcal{A}(H_N))/N \geq \mathsf{ALG}^{\mathrm{Is}} - \varepsilon] \geq 1 - o(1).$$

All of the algorithms in [7], [55], [66], [69] are computationally efficient. The latter three works use a class of iterative algorithms known as approximate message passing (AMP). In particular they require only a constant number of queries of $\nabla H_N(\cdot)$; this results in computation time linear in the description length of H_N when ξ is a polynomial, assuming oracle access to ζ_* and the function Φ_{ζ_*} (which only need to be computed once for each (ξ, h)). AMP offers a great deal of flexibility, and the idea introduced in [55] was to use it to encode a stochastic control problem which is in some sense dual to the Parisi formula. Based on this idea it was shown in [7] that no AMP algorithm of this powerful but specific form can achieve asymptotic value $\mathsf{ALG}^{\mathrm{Is}} + \varepsilon$ in the case $h = 0$. The non-equality $\mathsf{ALG}^{\mathrm{Is}} < \mathsf{OPT}^{\mathrm{Is}}$ also has a natural interpretation in terms of the optimizer ζ_* of (I.4). Namely, it implies that ζ_* is not strictly increasing; see [66] for a more precise condition called "optimizability" therein. As explained in [66, Section 6], in the case of even Ising spin glasses this non-equality exactly coincides with the presence of an *overlap gap property* (discussed below) associated with forms of algorithmic hardness. It is therefore natural to conjecture that the aforementioned AMP algorithms achieve the best asymptotic energy possible for efficient algorithms.

This belief was also aligned with results on the "critical point complexity" of pure spherical spin glasses with $\xi(x) = x^p$ and $h = 0$. In this case, the analogous value $\mathsf{ALG}^{\mathrm{Sp}}$ is the one obtained by [69] and coincides with the onset of exponentially many bounded index critical points, as established in [11], [68]. In this case almost all local optima have energy value $\mathsf{ALG}^{\mathrm{Sp}} \pm o(1)$ with high probability, which suggests from another direction that exceeding the energy $\mathsf{ALG}^{\mathrm{Sp}}$ might be computationally intractable. On the other hand, this threshold (see [19]) does not coincide with $\mathsf{ALG}^{\mathrm{Sp}}$ beyond the pure case.

It unfortunately seems difficult to establish any limitations on the power of general polynomial-time algorithms to optimize a spin glass Hamiltonian. However one might still hope to characterize the power of natural classes of algorithms that include gradient descent and AMP. To this end, we define the following distance on the space $\mathcal{H}_N$ of Hamiltonians H_N. We identify H_N with its disorder coefficients $(\mathbf{G}^{(p)})_{p \in 2\mathbb{N}}$, which we concatenate (in an arbitrary but fixed order) into an infinite vector $\mathbf{g}(H_N)$. We equip $\mathcal{H}_N$ with the (possibly infinite) distance

$$\|H_N - H'_N\|_2 = \|\mathbf{g}(H_N) - \mathbf{g}(H'_N)\|_2.$$

Let $B_N = \{\boldsymbol{\sigma} \in \mathbb{R}^N : \|\sigma\|_2^2 \le N\}$ and $C_N = [-1, 1]^N$ be the convex hulls of S_N and Σ_N, which we equip with the standard $\|\cdot\|_2$ distance. A consequence of our main result is that no suitably *Lipschitz* function $\mathcal{A} : \mathcal{H}_N \to C_N$ can surpass the asymptotic value $\mathsf{ALG}^{\mathrm{Is}}$. (And similarly in the spherical case for $\mathcal{A} : \mathcal{H}_N \to B_N$ and an analogous $\mathsf{ALG}^{\mathrm{Sp}}$.)

Theorem 3. *Let $\tau, \varepsilon > 0$ be constants. For N sufficiently large, any τ-Lipschitz $\mathcal{A} : \mathcal{H}_N \to C_N$ satisfies, for some $c = c(\xi, h, \varepsilon, \tau) > 0$,*

$$\mathbb{P}\left[H_N(\mathcal{A}(H_N))/N \ge \mathsf{ALG}^{\mathrm{Is}} + \varepsilon\right] \le e^{-cN}.$$

The algorithms of [7], [55], [66] are $O(1)$-Lipschitz in the sense above[2]. While the approach of [69] is not Lipschitz, its performance is captured by AMP as explained in [7, Remark 2.2]. Hence in tandem with these constructive results, Theorem 3 identifies the exact asymptotic value achievable by Lipschitz functions $\mathcal{A} : \mathcal{H}_N \to C_N$ (assuming the existence of a minimizer $\zeta_* \in \mathscr{L}$ as required in Theorem 2). We also give an analogous result for spherical spin glasses, in which there is no question of existence of a minimizer on the algorithmic side. Let us remark that the rate e^{-cN} in Theorem 3 is best possible up to the value of c, being achieved even for the trivial algorithm $\mathcal{A}(H_N) = (1, 1, \ldots, 1)$ which ignores its input entirely.

Many natural optimization algorithms satisfy the Lipschitz property above on a set $K_N \subseteq \mathcal{H}_N$ of inputs with $1 - \exp(-\Omega(N))$ probability; this suffices just as well for Theorem 3 thanks to the Kirszbraun extension theorem. As explained in Section 8 of the full version of the paper, algorithms with this property include the following examples, all run for a constant (i.e. dimension-independent) number of iterations or amount of time.

- Gradient descent and natural variants thereof;
- Approximate message passing;
- More general "higher-order" optimization methods with access to $\nabla^k H_N(\cdot)$ for constant k;
- Langevin dynamics for the Gibbs measure $e^{\beta H_N}$ with suitable reflecting boundary conditions and any positive constant β.

In fact we will not require the full Lipschitz assumption on $\mathcal{A}$, but only a consequence that we call overlap concentration. Roughly speaking, overlap concentration

[2]Technically the algorithms in these papers round their outputs to the discrete set Σ_N at the end, making them discontinuous. Removing the rounding step yields Lipschitz maps $\mathcal{A} : \mathcal{H}_N \to C_N$ with the same performance.

of $\mathcal{A}$ means that given any fixed correlation between the disorder coefficients of H_N^1 and H_N^2, the overlap $\langle \mathcal{A}(H_N^1), \mathcal{A}(H_N^2) \rangle / N$ tightly concentrates around its mean. This property holds automatically for τ-Lipschitz $\mathcal{A}$ thanks to concentration of measure on Gaussian space. It also might plausibly be satisfied for some discontinuous algorithms such as the Glauber dynamics.

B. Algorithms and Lower Bounds in Related Problems

As mentioned previously, mean field spin glasses are just one example of a random optimization problem. In these problems we are given a random objective function and we aim to find an input achieving a large value; examples include random constraint satisfaction problems, such as random k-SAT and MaxCut, as well as random perceptron models [35], [73], [74]. In the large-degree limit, the maxima of the former problems are known to be described by mean-field spin glasses [33], [59] (see [6] for an algorithmic analog). Many other algorithms have been studied for these problems including the Gaussian wave function for MaxCut [52] and a combinatorial algorithm for random k-SAT [28]. Let us also mention a line of work on the algorithmic Lóvasz local lemma for approximate counting and sampling of solutions at low clause density [46], [54], [56].

The aforementioned problems stand in contrast to *signal recovery* problems, such as recovering a unit norm signal $\boldsymbol{v}$ from a noisy matrix observation $A = W + \lambda \boldsymbol{v}^{\otimes 2}$ for $W \sim \mathsf{GOE}(N)$, and the analogous *detection* problems, where we must hypothesis test between this model and a null model where $\boldsymbol{v} = 0$. For detection problems, the so-called low degree heuristic has by now led to several robustly supported predictions for the onset of computational hardness, based on projecting the classical likelihood ratio onto a subspace of low degree polynomials. Indeed, it is often possible to show tight lower bounds on the performance of low degree polynomial testing algorithms (see [51] and references therein). Recovery lower bounds for low degree polynomials can often be similarly shown [65]. However for random optimization problems, no comparable unifying picture has been put forward. We discuss several existing heuristics in the next subsection.

C. The Overlap Gap Property as a Barrier to Algorithms

The main heuristics proposed to understand computational hardness in random optimization problems have focused on geometric properties of the solution space. One version of this connection was proposed in [3], [29] based on a *shattering* phase transition: for suitable random instances of k-SAT, q-coloring, and maximum independent set, beyond a threshold constraint density the solution space breaks into exponentially many small components. Shattering defeats local search heuristics, suggesting that polynomial-time algorithms should not succeed. Other predictions based on the *clustering*, *condensation* [50] and *freezing* [76] transitions have also been suggested. While intuitively appealing, the hypothesis that some form of clustering is responsible for hardness has been shown incorrect in notable examples – see Subsection I-D for more discussion.

In the past several years, a line of work [21], [25], [39]–[41], [43]–[45], [63], [75] on the Overlap Gap Property (OGP) has made substantial progress on rigorously linking solution geometry to hardness. A survey can be found in [38]. Initiated by Gamarnik and Sudan in [44], this line of work links the absence of certain constellations of solutions in the super-level set $S(E) = \{\boldsymbol{\sigma} : H_N(\boldsymbol{\sigma})/N \geq E\}$ – in its original form, a pair of solutions a medium distance apart – with the failure of algorithms with certain stability properties. Roughly, these works proceed by contradiction, arguing that any stable algorithm attaining value E would be able to construct the forbidden constellation. An important difference from the predictions above is that the shattering, clustering, and freezing transitions describe properties of a *typical* solution, while an OGP argument requires the forbidden constellation to *not exist at all*.

In many of these problems, the classic OGP shows the failure of stable algorithms above an intermediate value, smaller than the existential maximum but larger than the algorithmic limit. The argument stalls because below this value, $S(E)$ does contain pairs of inputs at each possible distance. To improve the lower bound, subsequent works have considered "multi-OGPs," which consider more complex forbidden structures; this is usually more difficult but often yields much sharper results. Indeed, multi-OGPs have been used to show nearly-tight hardness results for finding maximum independent sets on $G(N, d/N)$ in the limit $N \to \infty$ followed by $d \to \infty$ [63], [75] as well as for random k-SAT [21]. In both cases the threshold is attained by a simple local algorithm, which is shown to be optimal within the larger class of low degree polynomials.

The overlap gap property has been applied previously to the spin glass Hamiltonians we consider. For *pure* spherical and Ising p-spin glasses where $h = 0$ and $p \geq 4$ is even, $\mathsf{ALG} < \mathsf{OPT}$ always holds (recall (I.4), (I.6)). In

such models, [40] showed using a (2-solution) OGP that low degree polynomials cannot achieve some objective strictly smaller than OPT, extending a similar hardness result of [39] for approximate message passing. [41] extended the conclusions of [40] to Boolean circuits of depth less than $\frac{\log n}{2 \log \log n}$. As pointed out in [66, Section 6], these results extend in the Ising case to any mixed even model where $\mathsf{ALG}^{\mathrm{Is}} < \mathsf{OPT}^{\mathrm{Is}}$. In this paper, we will use a multi-OGP to show that overlap concentrated algorithms cannot optimize mixed even spherical or Ising spin glasses to any objective larger than ALG.

The design of our multi-OGP is a significant departure from previous work. Previous OGPs and multi-OGPs all use one of the following three forbidden structures, see Figure 1.

- Classic OGP: two solutions with medium overlap [25], [39]–[41], [44].
- Star OGP: several solutions with approximately the same pairwise overlap [43], [45], [63].
- Ladder OGP: several solutions, where the i-th solution ($i \geq 2$) has medium "multi-overlap" with the first $i-1$ solutions [21], [75].

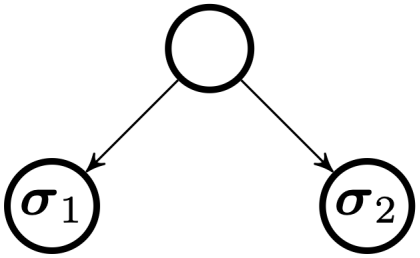
In contrast, the forbidden structure in our multi-OGP is an arbitrarily complicated ultrametric branching tree of solutions. We call this the *Branching OGP*. Informally, the Branching OGP is the condition that for any fixed $\varepsilon > 0$, no constellation of configurations with a certain ultrametric overlap structure has average energy $\mathsf{ALG}+\varepsilon$. The definition involves a family of "ultrametrically correlated" Hamiltonians, with one input in the constellation per Hamiltonian.

We also show that the full strength of the branching OGP is necessary to establish Lipschitz hardness at all objectives above ALG, in the sense that any less complex ultrametric structure fails to be forbidden at an energy bounded away from ALG, via the following theorem. Recall that finite ultrametric spaces can be canonically identified with rooted trees.

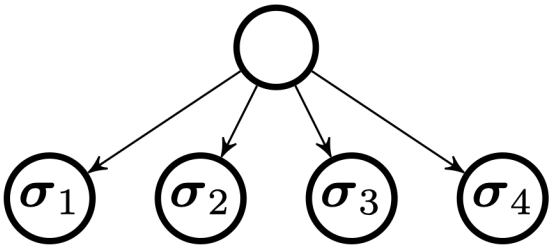
Theorem 4 (Informal)**.** *Consider a spherical model ξ without external field such that $\mathsf{ALG}_{\xi}^{\mathrm{Sp}} < \mathsf{OPT}_{\xi}^{\mathrm{Sp}}$, and a fixed ultrametric overlap structure $\mathbb{T}$ of inputs whose corresponding rooted tree If $\mathbb{T}$ does not contain a complete depth-D binary tree, then with high probability there exists a constellation of inputs $(\boldsymbol{\sigma}(v))_{v \in V(\mathbb{T})}$ with overlap structure $\mathbb{T}$ such that*

$$\min_{v \in V(\mathbb{T})} H_N(\boldsymbol{\sigma}(v)) \geq \mathsf{ALG}^{\mathrm{Sp}} + \varepsilon_{\xi,D}$$

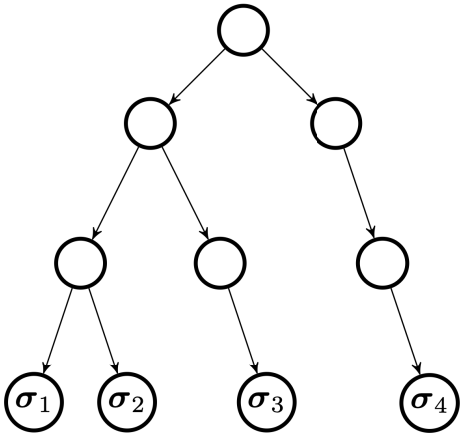
for a constant $\varepsilon_{\xi,D} > 0$ depending only on ξ, D.



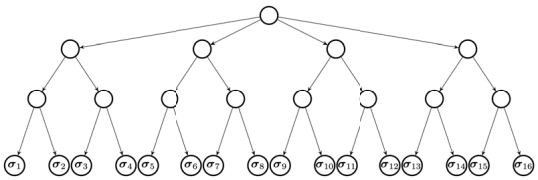
(a) **Classic OGP**: $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$ have medium overlap.



(b) **Star OGP**: many solutions, medium overlaps.



(c) **Ladder OGP**: medium "multi-overlaps" between $\boldsymbol{\sigma}_i$ and $\{\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_{i-1}\}$.



(d) **Branching OGP**: many solutions in an ultrametric tree.

Fig. 1: Schematics of OGP forbidden structures OGP.

More details can be found in Section 7 of the full version of the paper.

Remark I.1. To our knowledge, this is the first hardness result in any random optimization problem that is tight in the strong sense of characterizing the exact point ALG where hardness occurs. The aforementioned hardness results for maximum independent set on $G(N, d/N)$ are tight in the sense of matching the best algorithms within a $1+o_d(1)$ factor in the limit $d \to \infty$, while there is still a constant factor gap for random k-SAT. In fact, prior to this work, all outstanding *predictions* for the algorithmic threshold in any random optimization problem have only matched the best algorithms within a $1+o_d(1)$ factor in the large-degree limit. Consequently we believe that the branching OGP elucidates the fundamental reason for algorithmic hardness and may provide a framework for exact algorithmic thresholds in other random optimization problems.

Remark I.2. The significance of ultrametricity in mean-field spin glasses began with [60] and has played an enormous role in guiding the mathematical understanding of the low temperature regime in works such as [24], [47], [57], [64]. Ultrametricity also appears naturally in the context of optimization algorithms. Indeed in [69, Remark 6], [5, Section 3.4] and [66, Theorem 4] it was realized that the aforementioned algorithms achieving asymptotic energy ALG are capable of more. Namely, they can construct arbitrary ultrametric constellations of solutions (subject to a suitable diameter upper bound), each with energy ALG. Our proof via branching OGP establishes a sharp converse — the existence of essentially arbitrary ultrametric configurations at a given energy level is *equivalent* to achievability by Lipschitz $\mathcal{A}$.

The aforementioned rigorous results on ultrametricity in [24], [47], [57], [64] state that the Gibbs measure $e^{\beta H_N(\boldsymbol{\sigma})}/Z \ \mathrm{d}\boldsymbol{\sigma}$ is, very roughly speaking, supported on an ultrametric subset S of the cube or sphere. For large β, this Gibbs measure describes the typical near maxima of H_N. However, the pairwise overlaps in S may not cover the entire interval $[0,1]$, which means that S is highly disconnected. By contrast, the ultrametric structures we link with algorithms are forced to branch continuously, which implies that the pairwise overlaps are dense in $[0,1]$. The condition that the Gibbs measure is supported on a continuously branching tree is a strong form of *full replica symmetry breaking*. It was under such a condition that the works [7], [55], [69] gave algorithms achieving the value OPT.

D. On Algorithmic Signatures of Hardness

While it has long been believed that algorithmic hardness in random optimization problems is caused by a change in the solution geometry, the precise geometric phenomenon giving rise to hardness has been the subject of much debate. A popular belief has been that hardness is caused by a clustering transition. Indeed, the influential work [3] shows that in random k-SAT and q-coloring, the maximal constraint density where algorithms succeed coincides (up to leading order in the $k, q \to \infty$ limit) with a *shattering* phase transition. The intuition justifying this belief was that in the shattered regime, the solution geometry becomes rugged, meaning that local search and potentially other algorithms fail. A related conjecture was put forward in [50], that in random CSPs local Markov chains fail above a different *clustering* (or *dynamic RSB*) threshold.

However, for random CSPs with bounded typical degree, it is known that algorithms succeed at constraint densities beyond the clustering transition [4], [76]. Moreover, it was later observed that in random perceptron models, neither clustering nor shattering coincides with hardness! Indeed, [14] empirically demonstrated an algorithm that finds solutions even when (according to physics heuristics) the overall solution space is dominated by well-separated isolated solutions, i.e. clusters of size one; they conjecture that algorithms find rare connected clusters of solutions. As rigorous evidence for this perspective, for the symmetric binary perceptron [2], [62] proved the isolated solutions phenomenon and [1] gives an algorithm to construct a cluster of solutions with macroscopic diameter when the clause density is small. Still, it was not clear even heuristically what type of solution cluster should correspond to computational tractability.

For the spin glass models we consider, our results show that the signature of algorithmic hardness is not clustering properties of typical solutions, but the existence of dense clusters of a particular form. Namely, optimization of H_N to energy E is possible for overlap-concentrated algorithms if and only if the super-level set $S(E)$ contains an "everywhere-branching" ultrametric tree of solutions. We expect that this characterization generalizes to other random optimization problems.

E. Further Background

We now describe some other results on algorithmically optimizing spin glass Hamiltonians. First, in the worst case over the disorder $(\mathbf{G}^{(p)})_{p \in 2\mathbb{N}}$, achieving

any constant approximation ratio to the true optimum value is known to be quasi-NP hard even for degree 2 polynomials [9], [15]. For the Sherrington-Kirkpatrick model with $\xi(t) = t^2/2$ on the cube, it was recently shown to be NP-hard on average to compute the *exact* value of the partition function [42]. Of course, these computational hardness results demand much stronger guarantees than the approximate optimization with high probability that we consider.

Another important line of work, alluded to above, has studied the *complexity* of the landscape of H_N on the sphere, defined as the exponential growth rate for the number of local optima and saddle points of finite-index at a given energy level. These are understood to serve as barriers to efficient optimization, and a non-rigorous study was undertaken in [30], [31], [61] followed by a great deal of recent progress in [10], [11], [49], [53], [68], [70]. Notably because the true maximum value of H_N is nothing but its largest critical value, the first moment results of [11] combined with the second moment results of [68] gave an alternate self-contained proof of the Parisi formula for the ground state in pure spherical models. In a related spirit, [23], [26], [27], [34] have shown that mixed even p-spin Hamiltonians typically contain exponentially many well-separated near-global maxima.

Other works such as [17], [18], [20], [22], [32] have studied natural algorithms such as Langevin and Glauber dynamics on short (independent of N) time scales. These approaches yield (often non-rigorous) predictions for the energy achieved after a fixed amount of time. However these predictions involve complicated systems of differential equations, and to the best of our knowledge it is not known how to cleanly describe the long-time limiting energy achieved. Let us also mention the recent results of [8], [36] showing that the Glauber dynamics for the Sherrington-Kirkpatrick model mix rapidly at high temperature. By contrast the problem of optimization considered in this work is related to the *low* temperature behavior of the model.

II. Main Results

A. Overlap Concentrated Algorithms

For any $p \in [0,1]$, we may construct two correlated copies $H_N^{(1)}, H_N^{(2)}$ of H_N as follows. Construct three i.i.d. Hamiltonians $\widetilde{H}_N^{[0]}, \widetilde{H}_N^{[1]}, \widetilde{H}_N^{[2]}$ with mixture ξ, as in (I.1). For $i = 1, 2$, let

$$H_N^{(i)}(\boldsymbol{\sigma}) = \langle \boldsymbol{h}, \boldsymbol{\sigma} \rangle + \widetilde{H}_N^{(i)}(\boldsymbol{\sigma}), \quad \text{where}$$
$$\widetilde{H}_N^{(i)}(\boldsymbol{\sigma}) = \sqrt{p}\widetilde{H}_N^{[0]}(\boldsymbol{\sigma}) + \sqrt{1-p}\widetilde{H}_N^{[i]}(\boldsymbol{\sigma}).$$

We say the pair of Hamiltonians $H_N^{(1)}, H_N^{(2)}$ is p-correlated. Note that corresponding entries in $\mathbf{g}^{(1)} = \mathbf{g}(H_N^{(1)})$ and $\mathbf{g}^{(2)} = \mathbf{g}(H_N^{(2)})$ are standard Gaussians with covariance p.

We will determine the maximum energy attained by algorithms $\mathcal{A}_N : \mathcal{H}_N \to B_N$ or $\mathcal{A}_N : \mathcal{H}_N \to C_N$ (always assumed to be measurable) obeying the following overlap concentration property.

Definition II.1. *Let $\lambda, \nu > 0$. An algorithm $\mathcal{A} = \mathcal{A}_N$ is (λ, ν) overlap concentrated if for any $p \in [0,1]$ and p-correlated Hamiltonians $H_N^{(1)}, H_N^{(2)}$,*

$$\mathbb{P}\left[\left|R\left(\mathcal{A}(H_N^{(1)}), \mathcal{A}(H_N^{(2)})\right) - \chi_{\mathcal{A}}(p)\right| \geq \lambda\right] \leq \nu;$$
$$\chi_{\mathcal{A}}(p) \equiv \mathbb{E}\, R\left(\mathcal{A}(H_N^{(1)}), \mathcal{A}(H_N^{(2)})\right).$$

B. The Spherical Zero-Temperature Parisi Functional

We introduce a Parisi functional P^{Sp} for the spherical setting, analogous to the Parisi functional P^{Is} for the Ising setting introduced in (I.3). Similarly to Theorem 1, Auffinger and Chen [12], see also [27], characterize the ground state energy of the spherical spin glass by a variational formula in terms of this Parisi functional. Recall the set $\mathscr{U}$ defined in (I.2). Let

$$\mathscr{V}(\xi) = \left\{(B, \zeta) \in \mathbb{R}^+ \times \mathscr{U} : B > \int_0^1 \xi''(t)\zeta(t)\mathrm{d}t\right\}.$$

Define the spherical Parisi functional $\mathsf{P}^{\mathrm{Sp}} = \mathsf{P}^{\mathrm{Sp}}_{\xi,h} : \mathscr{V}(\xi) \to \mathbb{R}$ by

$$\mathsf{P}^{\mathrm{Sp}}(B, \zeta) = \frac{h^2}{2B_\zeta(0)} + \int_0^1 \left(\frac{\xi''(t)}{2B_\zeta(t)} + \frac{B_\zeta(t)}{2}\right)\mathrm{d}t \tag{II.1}$$

where for $t \in [0,1]$,

$$B_\zeta(t) = B - \int_t^1 \xi''(q)\zeta(q)\mathrm{d}q.$$

Theorem 5 ([12, Theorem 10])**.** *The following identity holds.*

$$\mathsf{OPT}^{\mathrm{Sp}} = \inf_{(B,\zeta)\in\mathscr{V}(\xi)} \mathsf{P}^{\mathrm{Sp}}(B, \zeta). \tag{II.2}$$

The infimum is attained at a unique $(B_*, \zeta_*) \in \mathscr{V}(\xi)$.

C. The Optimal Energy of Overlap Concentrated Algorithms

We defined $\mathsf{ALG}^{\mathrm{Is}}$ in (I.6) by a non-monotone extension of the variational formula in (I.4). We can similarly define $\mathsf{ALG}^{\mathrm{Sp}}$ by a non-monotone extension of (II.2). Recall the set $\mathscr{L}$ defined in (I.5). Let $\mathscr{K}(\xi) \supseteq \mathscr{V}(\xi)$ denote the set

$$\mathscr{K}(\xi) = \left\{ (B,\zeta) \in \mathbb{R}^+ \times \mathscr{L} : B > \int_0^1 \xi''(t)\zeta(t)\mathrm{d}t \right\}.$$

The Parisi functional P^{Sp} can clearly be defined on $\mathscr{K}(\xi)$. We define $\mathsf{ALG}^{\mathrm{Sp}} = \mathsf{ALG}^{\mathrm{Sp}}_{\xi,h}$ by

$$\mathsf{ALG}^{\mathrm{Sp}} = \inf_{(B,\zeta)\in\mathscr{K}(\xi)} \mathsf{P}^{\mathrm{Sp}}(B,\zeta). \tag{II.3}$$

Note that $\mathsf{ALG}^{\mathrm{Sp}} \le \mathsf{OPT}^{\mathrm{Sp}}$ trivially.

We are now ready to state the main result of this work. We show that for any mixed even spherical or Ising spin glass, no overlap concentrated algorithm can attain an energy above the algorithmic thresholds $\mathsf{ALG}^{\mathrm{Sp}}$ and $\mathsf{ALG}^{\mathrm{Is}}$ with nontrivial probability.

Theorem 6 (Main Result)**.** *Consider a mixed even Hamiltonian H_N with model (ξ,h). Let $\mathsf{ALG} = \mathsf{ALG}^{\mathrm{Sp}}$ (resp. $\mathsf{ALG}^{\mathrm{Is}}$). For any $\varepsilon > 0$ there are $\lambda, c, N_0 > 0$ depending only on ξ, h, ε such that the following holds for $N \ge N_0$ and $\nu \in [0,1]$. For any (λ,ν) overlap concentrated $\mathcal{A} = \mathcal{A}_N : \mathcal{H}_N \to B_N$ (resp. C_N),*

$$\mathbb{P}\left[\frac{1}{N} H_N\left(\mathcal{A}(H_N)\right) \ge \mathsf{ALG} + \varepsilon\right] \le e^{-cN} + (\nu/\lambda)^c.$$

Remark II.1. If $\mathcal{A}$ is τ-Lipschitz, (λ,ν) overlap concentration holds with $\nu = \exp(-c_{\lambda,\tau}N)$ by concentration of measure on Gaussian space. . Hence in this case the probability on the right-hand side above is exponentially small in N. The same property holds when $\mathcal{A}$ is τ-Lipschitz on a set of inputs with $1 - \exp(-\Omega(N))$ probability as a consequence of the Kirszbraun extension theorem.

In tandem with Theorem 2 and its spherical analogue Theorem 7 below, Theorem 6 exactly characterizes the maximum energy attained by overlap concentrated algorithms (again with the caveat on the algorithmic side in the Ising case that a minimizer $\gamma_* \in \mathscr{L}$ exists in Theorem 2).

Theorem 7 ([7], [66])**.** *For any $\varepsilon > 0$, there exists an efficient and $O_\varepsilon(1)$-Lipschitz AMP algorithm $\mathcal{A} : \mathcal{H}_N \to B_N$ such that for some $c = c(\varepsilon) > 0$,*

$$\mathbb{P}[H_N(\mathcal{A}(H_N))/N \ge \mathsf{ALG}^{\mathrm{Sp}} - \varepsilon] \ge 1 - e^{-cN}.$$

For spherical spin glasses, the value of $\mathsf{ALG}^{\mathrm{Sp}}$ is explicit and is given by the following proposition.

Proposition II.2. *If $h^2 + \xi'(1) \ge \xi''(1)$, then*

$$\mathsf{ALG}^{\mathrm{Sp}} = (h^2 + \xi'(1))^{1/2},$$

and the infimum in (II.3) is uniquely attained by $B = (h^2+\xi'(1))^{1/2}$, $\zeta = 0$. Otherwise,

$$\mathsf{ALG}^{\mathrm{Sp}} = \widehat{q}\xi''(\widehat{q})^{1/2} + \int_{\widehat{q}}^1 \xi''(q)^{1/2}\mathrm{d}q$$

where $\widehat{q} \in [0,1)$ is the unique number satisfying $h^2 + \xi'(\widehat{q}) = \widehat{q}\xi''(\widehat{q})$. If $h > 0$, the infimum in (II.3) is uniquely attained by $B = \xi''(1)^{1/2}$ and

$$\zeta(q) = \mathbb{I}\{q \ge \widehat{q}\}\frac{\xi'''(q)}{2\xi''(q)^{3/2}} = -\mathbb{I}\{q \ge \widehat{q}\}\frac{\mathrm{d}}{\mathrm{d}q}\xi''(q)^{-1/2}. \tag{II.4}$$

If $h = 0$, the infimum is achieved by $B = \xi''(1)^{1/2}$ and ζ given by (II.4) in the limit as $\widehat{q} \to 0^+$.

Note that $\mathsf{ALG}^{\mathrm{Sp}} = \mathsf{OPT}^{\mathrm{Sp}}$ if and only if the infimum in (II.3) is attained at a pair $(B,\zeta) \in \mathscr{V}(\xi)$. Thus, Proposition II.2 implies that $\mathsf{ALG}^{\mathrm{Sp}} = \mathsf{OPT}^{\mathrm{Sp}}$ if and only if $h^2 + \xi'(1) \ge \xi''(1)$ or $\xi''(q)^{-1/2}$ is concave on $[\widehat{q},1]$. Interestingly, in the case $h^2+\xi'(1) > \xi''(1)$, [16], [37] showed that H_N has "trivial complexity": no critical points on S_N with high probability except for the unique global maximizer and minimizer.

In the important case of the pure p-spin model, with $h = 0$ and $\xi(x) = x^p$ for $p \ge 4$ even,

$$\mathsf{ALG}^{\mathrm{Sp}} = \int_0^1 \xi''(q)^{1/2}\mathrm{d}q = 2\sqrt{\frac{p-1}{p}}.$$

This coincides with the threshold $E_\infty(p)$ identified in [11]. As conjectured in [11] and proved in [68], with high probability an overwhelming majority of local maxima of H_N on S_N have energy value $E_\infty(p) \pm o(1)$. This suggests that it may be computationally intractable to achieve energy at least $E_\infty(p) + \varepsilon$ for any $\varepsilon > 0$; our results confirm this hypothesis for overlap concentrated algorithms.

Remark II.2. Our results generalize with no changes in the proofs to arbitrary external fields $\boldsymbol{h} = (h_1,\dots,h_N)$ which are independent of $\widetilde{H}_N$ — one only needs to replace h^2 by $\frac{\|\boldsymbol{h}\|^2}{N}$ in (II.1) and replace $\Phi_\zeta(0,h)$ by $\frac{1}{N}\sum_{i=1}^N \Phi_\zeta(0,h_i)$ in (I.3). This includes for instance the natural case of Gaussian external field $\boldsymbol{h} \sim \mathcal{N}(0,I_N)$. Here $\mathcal{A}$ can depend arbitrarily on $\boldsymbol{h}$ as long as overlap concentration holds conditionally on $\boldsymbol{h}$.

III. Ideas of the Proof

Our proof has two main pieces, the first of which is relatively conceptual and the second of which is more technical. The first stage is to show that an overlap concentrated $\mathcal{A}$ allows the construction of an arbitrary ultrametric constellation of outputs. To do this, we apply $\mathcal{A}$ to a large constant number of coupled Hamiltonians $H_N^{(u)}$ for $u \in [k]^D$. We think of u as ranging over the k^D leaves of a complete k-ary tree of depth D. The Hamiltonians $H_N^{(u)}$ have a jointly Gaussian structure, and each has the marginal law of H_N. The key property is *ultrametric* correlation between different Hamiltonians. Namely we let $H_N^{(u)}$ and $H_N^{(v)}$ to be $p_{u\wedge v}$-correlated, where $u \wedge v$ is the depth of the least common ancestor of u and v, and

$$0 = p_0 \le p_1 \le \cdots \le p_D = 1$$

is an arbitrary increasing sequence. By Definition II.1, for any overlap concentrated $\mathcal{A}$, there exists χ such that the outputs $\boldsymbol{\sigma}(u) = \mathcal{A}(H_N^{(u)})$ satisfy with high probability

$$\langle \boldsymbol{\sigma}(u), \boldsymbol{\sigma}(v)\rangle / N \approx q_{u\wedge v} = \chi(p_{u\wedge v}), \quad \forall u, v.$$

By the intermediate value theorem, given $\mathcal{A}$ we may choose any sequence

$$\chi(0) = q_0 \le q_1 \le \cdots \le q_D = 1$$

and find $(p_0, \dots, p_D)$ such that $\chi(p_d) = q_d$ for $0 \le d \le D$. Assuming for simplicity that $\chi(0) = 0$, we do this for $q_d = d/D$ with large D.

Consequently, if $\mathcal{A}$ outputs points with energy at least $\mathsf{ALG} + \varepsilon$, then $\mathcal{A}$ run on the above family of ultrametrically correlated Hamiltonians will output the forbidden structure above, a contradiction. Some additional complications are created by the fact that $\chi(0) = \|\mathbb{E}[\mathcal{A}(H_N)]\|^2$ may be arbitrary, and that $\mathcal{A}(H_N)$ may be in the interior of C_N (or in the spherical case, B_N). The former issue requires us to control the maximum average energy of ultrametric constellations of points that all have approximately a fixed overlap with $\mathbb{E}[\mathcal{A}(H_N)]$. We deal with the latter issue by composing $\mathcal{A}$ with an additional phase that grows each output of $\mathcal{A}$ into its own ultrametric tree of points in Σ_N (or S_N), so that the resulting set of points has the forbidden ultrametric structure.

Given the above, the most technical part of the proof is to show that, for any $\varepsilon > 0$, there does not exist a suitable ultrametric constellation of inputs (to an appropriate family of ultrametrically correlated Hamiltonians) with average energy $\mathsf{ALG} + \varepsilon$. Using a version of the Guerra-Talagrand interpolation, which we take to zero temperature, we derive an upper bound for this average energy. This upper bound is a multi-dimensional analogue of the Parisi formula, and depends on an essentially arbitrary increasing function $\zeta : [0,1] \to \mathbb{R}^+$ (which we are free to minimize over). We show that for the above symmetric branching trees, the resulting estimate can be upper bounded by $\mathsf{P}(\kappa\zeta)$. Here P is the Parisi functional P^{Is} or its spherical analogue P^{Sp}, and κ is a decreasing function which is constant on each interval $[q_d, q_{d+1})$. By taking q_d a fine enough discretization of $[0,1]$, the function κ can be arranged to decrease as rapidly as desired. As a result, the functions $\kappa\zeta$ are dense in the space $\mathscr{L}$ of reasonable decreasing functions. Thus, we may choose a tree and ζ such that $\mathsf{P}(\kappa\zeta)$ is arbitrarily close to ALG.

The full version of this paper is available as an online preprint at https://arxiv.org/abs/2110.07847 .

Acknowledgment

We thank Guy Bresler, Wei-Kuo Chen, Andrea Montanari, Cristopher Moore, Dmitry Panchenko, Yi Sun, and Lenka Zdeborová for helpful references and comments.

References

[1] Emmanuel Abbe, Shuangping Li, and Allan Sly. Binary perceptron: efficient algorithms can find solutions in a rare well-connected cluster. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 860–873, 2022.

[2] Emmanuel Abbe, Shuangping Li, and Allan Sly. Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 327–338. IEEE, 2022.

[3] Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *Proceedings of 49th FOCS*, pages 793–802, 2008.

[4] Dimitris Achlioptas and Cristopher Moore. Almost all graphs with average degree 4 are 3-colorable. *Journal of Computer and System Sciences*, 67(2):441–471, 2003.

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

[6] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Local algorithms for maximum cut and minimum bisection on locally treelike regular graphs of large degree. *arXiv preprint arXiv:2111.06813*, 2021.

[7] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *Ann. Probab.*, 49(6):2922–2960, 2021.

[8] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence: optimal mixing of down-up random walks. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1418–1430, 2022.

[9] Sanjeev Arora, Eli Berger, Hazan Elad, Guy Kindler, and Muli Safra. On non-approximability for quadratic programs. In *Proceedings of 46th FOCS*, pages 206–215. IEEE, 2005.

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

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

[12] Antonio Auffinger and Wei-Kuo Chen. On the energy landscape of spherical spin glasses. *Advances in Mathematics*, 330, 02 2017.

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

[14] Carlo Baldassi, Alessandro Ingrosso, Carlo Lucibello, Luca Saglietti, and Riccardo Zecchina. Subdominant dense clusters allow for simple learning and high computational performance in neural networks with discrete synapses. *Physical review letters*, 115(12):128101, 2015.

[15] Boaz Barak, Fernando G.S.L. Brandão, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 307–326. ACM, 2012.

[16] David Belius, Jiří Černý, Shuta Nakajima, and Marius Schmidt. Triviality of the geometry of mixed p-spin spherical hamiltonians with external field. *arXiv preprint arXiv:2104.06345*, 2021.

[17] Gérard Ben Arous, Amir Dembo, and Alice Guionnet. Cugliandolo-Kurchan equations for dynamics of spin-glasses. *Probability Theory and Related Fields*, 136(4):619–660, 2006.

[18] Gérard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Bounding flows for spherical spin glass dynamics. *Communications in Mathematical Physics*, 373(3):1011–1048, 2020.

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

[20] Jean-Philippe Bouchaud, Leticia F Cugliandolo, Jorge Kurchan, and Marc Mézard. Out of equilibrium dynamics in spin-glasses and other glassy systems. *Spin glasses and random fields*, pages 161–223, 1998.

[21] Guy Bresler and Brice Huang. The Algorithmic Phase Transition of Random k-SAT for Low Degree Polynomials. In *Proceedings of 62nd FOCS*, pages 298–309, 2021.

[22] Michael Celentano, Chen Cheng, and Andrea Montanari. The high-dimensional asymptotics of first order methods with random data. *arXiv preprint arXiv:2112.07572*, 2021.

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

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

[25] Wei-Kuo Chen, David Gamarnik, Dmitry Panchenko, and Mustazee Rahman. Suboptimality of local algorithms for a class of max-cut problems. *The Annals of Probability*, 47(3):1587–1618, 2019.

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

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

[28] Amin Coja-Oghlan. A better algorithm for random k-SAT. *SIAM Journal on Computing*, 39(7):2823–2864, 2010.

[29] Amin Coja-Oghlan and Charilaos Efthymiou. On independent sets in random graphs. *Random Structures & Algorithms*, 47(3):436–486, 2015.

[30] Andrea Crisanti, Luca Leuzzi, and Tommaso Rizzo. The complexity of the spherical p-spin spin glass model, revisited. *The European Physical Journal B-Condensed Matter and Complex Systems*, 36(1):129–136, 2003.

[31] Andrea Crisanti, Luca Leuzzi, and Tommaso Rizzo. Complexity in mean-field spin-glass models: Ising p-spin. *Physical Review B*, 71(9):094202, 2005.

[32] Leticia F. Cugliandolo and Jorge Kurchan. On the out-of-equilibrium relaxation of the Sherrington-Kirkpatrick model. *Journal of Physics A: Mathematical and General*, 27(17):5749, 1994.

[33] Amir Dembo, Andrea Montanari, and Subhabrata Sen. Extremal cuts of sparse random graphs. *The Annals of Probability*, 45(2):1190–1217, 2017.

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

[35] Jian Ding and Nike Sun. Capacity lower bound for the ising perceptron. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 816–827, 2019.

[36] Ronen Eldan, Frederic Koehler, and Ofer Zeitouni. A spectral condition for spectral gap: fast mixing in high-temperature ising models. *Probability Theory and Related Fields*, pages 1–17, 2021.

[37] YV Fyodorov. High-dimensional random fields and random matrix theory. *Markov Processes and Related Fields*, 21(3):483–518, 2015.

[38] David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41), 2021.

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

[40] David Gamarnik, Aukosh Jagannath, and Alexander S. Wein. Low-degree hardness of random optimization problems. In *Proceedings of 61st FOCS*, pages 131–140. IEEE, 2020.

[41] David Gamarnik, Aukosh Jagannath, and Alexander S. Wein. Circuit lower bounds for the p-spin optimization problem. *arXiv preprint arXiv:2109.01342*, 2021.

[42] David Gamarnik and Eren C. Kızıldağ. Computing the partition function of the Sherrington–Kirkpatrick model is hard on average. *The Annals of Applied Probability*, 31(3):1474–1504, 2021.

[43] David Gamarnik and Eren C. Kızıldağ. Algorithmic obstructions in the random number partitioning problem. *arXiv preprint arXiv:2103.01369*, 2021.

[44] David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 369–376. ACM, 2014.

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

[46] Kun He, Kewen Wu, and Kuan Yang. Improved Bounds for Sampling Solutions of Random CNF Formulas. *arXiv preprint arXiv:2207.11892*, 2022.

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

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

[49] Pax Kivimae. The ground state energy and concentration of complexity in spherical bipartite models. *arXiv preprint arXiv:2107.13138*, 2021.

[50] Florent Krzakala, Andrea Montanari, Federico Ricci-Tersenghi, Guilhem Semerjian, and Lenka Zdeborová. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 104(25):10318–10323, 2007.

[51] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*, pages 1–50. Springer, 2022.

[52] Russell Lyons. Factors of IID on trees. *Combinatorics, Probability and Computing*, 26(2):285–300, 2017.

[53] Benjamin McKenna. Complexity of bipartite spherical spin glasses. *arXiv preprint arXiv:2105.05043*, 2021.

[54] Ankur Moitra. Approximate counting, the Lovász local lemma, and inference in graphical models. *Journal of the ACM (JACM)*, 66(2):1–25, 2019.

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

[56] Robin A Moser. A constructive proof of the Lovász local lemma. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 343–350, 2009.

[57] Dmitry Panchenko. The Parisi ultrametricity conjecture. *Annals of Mathematics*, pages 383–393, 2013.

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

[59] Dmitry Panchenko. On the k-sat model with large number of clauses. *Random Structures & Algorithms*, 52(3):536–542, 2018.

[60] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754, 1979.

[61] Giorgio Parisi. Computing the number of metastable states in infinite-range models. *arXiv preprint arXiv:cond-mat/0602349*, 2006.

[62] Will Perkins and Changji Xu. Frozen 1-rsb structure of the symmetric ising perceptron. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1579–1588, 2021.

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

[64] David Ruelle. A mathematical reformulation of Derrida's REM and GREM. *Communications in Mathematical Physics*, 108(2):225–239, 1987.

[65] Tselil Schramm and Alexander S Wein. Computational barriers to estimation from low-degree polynomials. *The Annals of Statistics*, 50(3):1833–1858, 2022.

[66] Mark Sellke. Optimizing mean field spin glasses with external field. *arXiv preprint arXiv:2105.03506*, 2021.

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

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

[69] Eliran Subag. Following the Ground States of Full-RSB Spherical Spin Glasses. *Communications on Pure and Applied Mathematics*, 74(5):1021–1044, 2021.

[70] Eliran Subag and Ofer Zeitouni. Concentration of the complexity of spherical pure p-spin models at arbitrary energies. *Journal of Mathematical Physics*, 62(12):123301, 2021.

[71] Michel Talagrand. Free energy of the spherical mean field model. *Probability Theory and Related Fields*, 134:339–382, 03 2006.

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

[73] Michel Talagrand. *Mean Field Models for Spin Glasses. Volume I: Basic Examples*, volume 54. Springer Science & Business Media, 2011.

[74] Michel Talagrand. *Mean Field Models for Spin Glasses. Volume II: Advanced Replica-Symmetry and Low Temperature*, volume 55. Springer Science & Business Media, 2011.

[75] Alexander S Wein. Optimal low-degree hardness of maximum independent set. *Mathematical Statistics and Learning*, 2022.

[76] Lenka Zdeborová and Florent Krząkała. Phase transitions in the coloring of random graphs. *Physical Review E*, 76(3):031131, 2007.