Computational Hardness in Random Optimization Problems from the Overlap Gap Property

by

Brice Huang
B.S., Mathematics and Electrical Engineering and Computer Science
Massachusetts Institute of Technology, 2019

Submitted to the MIT Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Electrical Engineering and Computer Science

at the

Massachusetts Institute of Technology

February 2022

Signature of Author .....
Department of Electrical Engineering and Computer Science
January 26, 2022

Certified by .....
Guy Bresler
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by .....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Dedicated to the memory of my father, Kevin Xiongfei Huang.

Acknowledgements

I am grateful to my advisor Guy Bresler for his patient and generous mentorship over the past two years. He has been a great collaborator, and his guidance has been invaluable in teaching me how to research.

Matthew Brennan was like an older brother to me in my first year. I miss him and am forever grateful to him.

This thesis is based on joint works [BH21, HS21] with Guy Bresler and Mark Sellke. I am grateful for their collaboration. I thank Wei-Kuo Chen, David Gamarnik, Andrea Montanari, Cristopher Moore, Dmitry Panchenko, Mehtaab Sawhney, Yi Sun, and Alex Wein for helpful conversations and comments. I also thank the Simons Institute at Berkeley, where I spent two stimulating semesters.

Finally, I am grateful to my family for their unending support. Mom, Dad, and Catherine, thank you for being there for me every step of the way.

Abstract

We study the limits of efficient algorithms in random optimization problems. In these problems, we are given a random objective function and our goal is to find an input achieving a large output. These problems often exhibit *information-computation gaps*, where the maximum objective that exists is larger than the maximum objective that known efficient algorithms can find. Our goal is to find rigorous evidence of computational hardness in the hard regime.

We focus on the problems of random k-SAT and mean-field spin glasses. Our results are:

- It is known that random k-SAT has a satisfying assignment with high probability up to clause density $2^k \log 2 - \frac{1}{2}(1 + \log 2) + o_k(1)$, while the best known algorithm (`Fix`) finds a satisfying assignment up to clause density $(1 - o_k(1))2^k \log k/k$. We prove that *low degree polynomial* algorithms cannot find a satisfying assignment above clause density $(1 + o_k(1))\kappa^* 2^k \log k/k$, for a universal constant $\kappa^* \approx 4.911$. Low degree polynomial algorithms encompass `Fix`, message passing algorithms including Belief and Survey Propagation guided decimation, and local algorithms on the factor graph. This is the first hardness result against any class of algorithms within a constant factor of the clause density achieved by `Fix`.

- The maximum asymptotic value OPT of the Hamiltonian H_N/N of a spherical or Ising mixed p-spin glass is given by the celebrated Parisi formula. Recently developed approximate message passing algorithms efficiently optimize H_N/N up to a value ALG given by an extended Parisi formula, which minimizes over a larger space of non-monotone functional order parameters. These two objectives coincide for spin glasses exhibiting a no overlap gap property, but are generically not equal. We prove that for mixed even p-spin models, no algorithm satisfying an *overlap concentration* property can produce an objective larger than ALG. This property holds for all algorithms with suitably Lipschitz dependence on the disorder coefficients of H_N, including natural formulations of gradient descent, approximate message passing, and Langevin dynamics run for bounded time. In particular, this includes the algorithms achieving ALG.

We prove these results by extending the overlap gap property (OGP) framework of Gamarnik and Sudan to *multi-OGPs*, which consider forbidden constellations containing several solutions. Our results for random k-SAT are proved by a multi-OGP that generalizes the ladder constellation introduced by Wein. Our results for spin glasses are proved by a new multi-OGP, the *branching OGP*, that uses an arbitrarily complex ultrametric constellation of solutions.

Contents

Chapter 1

Introduction

1.1 Random Optimization Problems

The subject of this thesis is algorithmic limits in random optimization problems. In a random optimization problem, we are given a random objective function and our goal is to find an input achieving a large output. Such problems are ubiquitous at the intersection of probability and computation. Examples of random optimization problems include:

- Random constraint satisfaction problems (rCSPs): random (MAX-)k-SAT, q-coloring, maxcut, maximum independent set;
- Mean-field spin glasses;
- Perceptron models;
- Number Partitioning.

Random optimization problems are also a rich source of examples of non-convex optimizaton problems.

There are two natural questions one can ask about any random optimization problem. First, what is the largest objective that exists with high probability? Second, what is the largest objective that can be found with high probability by an efficient algorithm? We refer to these values as the *existential* and *algorithmic* limits.

In a large number of random optimization problems, our understanding of these two questions can be described by the following state of affairs. The existential limit is very well understood, and its in-probability limiting value is either known exactly or within precise estimates. However, much less is known about the algorithmic limit. The best known polynomial-time algorithm stalls at an objective well below the existential limit; this barrier has defied all attempts to devise better algorithms, and is believed to be fundamental. This phenomenon is known as an *information-computation gap*.

Understanding the limits of efficient algorithms gives insight into the basic computational limits of optimization. This leads us to our guiding question:

Can we characterize and find rigorous evidence for the limits of efficient algorithms in random optimization problems? Moreover, can we establish algorithmic limits across many problems in a unified way?

We will study this question through a somewhat surprising connection between the limits of efficient algorithms and the geometry of the solution space. It was observed in the influential work [ACO08] that in many problems, the failure of known efficient algorithms coincides with a clustering phase transition, beyond which metastable states proliferate. The popular wisdom is that beyond this threshold, the optimization landscape becomes very rugged, rendering efficient algorithms unable to navigate this landscape. It would thus be of interest to rigorously show that efficient algorithms fail beyond this transition.

Our central tool for showing computational hardness is the *Overlap Gap Property* (OGP). Introduced by Gamarnik and Sudan in [GS14], the OGP technique formalizes clustering as the absence of a constellation of

solutions with a prescribed geometric structure and leverages this fact to show hardness for powerful classes of *stable* algorithms. The notion of stable depends on the problem under consideration, but in all cases includes the state-of-the-art algorithm for the problem and powerful computatonal frameworks. Thus, OGP promises to give a unified geometric explanation of hardness across many random optimization problems.

The choice of the forbidden constellation in OGP is crucial to the success of the OGP argument. The design and analysis of this constellation is one of the main technical barriers to using OGP to show computational hardness. By introducing new tools for this task, our work substantially generalize the reach of OGP methodology. In Section 1.4, we give a detailed overview of the OGP program and our contributions.

We focus our attention on two of the oldest studied random optimization problems, random k-SAT and mean-field spin glasses, which we now introduce.

1.2 Random k-SAT

The k-SAT problem occupies a central role in complexity theory as the first and canonical NP-complete problem [Coo71]. Its average-case analogue, random k-SAT, likewise has a central role in average-case computational complexity. In this problem, we are given a k-CNF formula with M clauses and N variables whose kM literals are sampled uniformly and i.i.d. from the $2N$ possible literals;[1] see [Ach09] for a survey. There are two fundamental questions for random k-SAT. First, at what scalings of (k, N, M) are there satisfying assignments? Second, when can they be found by efficient algorithms? Note that random k-SAT can be identified with the random optimization problem of minimizing the number of violated clauses, and these questions correspond to when the existential and algorithmic limits are zero.[2]

Early work showed that for fixed k the interesting regime of random k-SAT is when $M = \Theta(N)$, and that the problem's qualitative behavior in the large-N limit depends on the *clause density* $\alpha = M/N$. Namely, [FP83] showed that if $\alpha \geq 2^k \log 2$, random k-SAT is unsatisfiable with high probability; on the positive side, [MTF90] showed that if $\alpha < 2^k/k$, a simple algorithm finds a satisfying assignment with nontrivial probability, and [CR92] improved the guarantee to with high probability. We henceforth work in the double limit where $N \to \infty$ at constant $\alpha = \alpha(k)$, and then $k \to \infty$.

As we tune α with k fixed, we encounter phase transitions separating one qualitative behavior from another. Two phase transitions are of primary interest to us: the *satisfiability threshold*, below which random k-SAT admits a satisfying assignment with high probability, and the *algorithmic threshold*, below which a polynomial-time algorithm produces a satisfying assignment with high probability.

The satisfiability threshold is well understood. [KKKS98] showed that random k-SAT is unsatisfiable with high probability at clause density $2^k \log 2 - \frac{1}{2}(\log 2 + 1) + o_k(1)$, where $o_k(1)$ denotes a term limiting to 0 as $k \to \infty$. Conversely, [COP16] showed that random k-SAT is satisfiable with high probability at clause density $2^k \log 2 - \frac{1}{2}(\log 2 + 1) - o_k(1)$. For large k, the landmark result of Ding, Sly, and Sun [DSS15] precisely identified the satisfiability threshold $\alpha_s(k)$ within this range, proving that with high probability, random k-SAT is satisfiable when $\alpha < \alpha_s(k)$ and unsatisfiable when $\alpha > \alpha_s(k)$.

We study the algorithmic threshold, which is much less understood. The best polynomial-time algorithm known, the `Fix` algorithm of Coja-Oghlan [CO10], finds a satisfying assignment with high probability at clause density $(1 - o_k(1))2^k \log k/k$, nearly a factor of k below the satisfiability threshold. A body of evidence has emerged to suggest that this is the correct threshold, but rigorous results that efficient algorithms fail beyond this threshold have been scarce.

In the early 2000s, statistical physicists developed a rich but non-rigorous theory describing the solution geometry of random k-SAT, among other random constraint satisfaction problems [KMRT+07]. This theory predicts several phase transitions in random k-SAT's solution geometry, which we now summarize; see [KMRT+07, Figure 2] for an illustration. At low clause density, the space of satisfying assignments is one large cluster. When the clause density reaches the *uniqueness threshold*, disconnected solution clusters appear but the main cluster contains all but an exponentially small fraction of solutions. At the *clustering threshold*, the solution space shatters into an exponentially large number of clusters, each with an exponentially small

[1] In a variant of this definition, the M clauses are chosen uniformly and without replacement among all $2^k\binom{N}{k}$ clauses with k distinct, non-complementary literals. This definition behaves identically to ours in the large-N limit, and all properties of random k-SAT we show in this chapter apply equally to this model.

[2] The question of determining the existential and algorithmic minimum numbers of violated clauses is the MAX-k-SAT problem, which is interesting in its own right. In this thesis, we focus on the question of satisfying assignments.

fraction of solutions. Additional clauses cause these clusters to shrink until at the *condensation threshold*, the solution space is dominated by a few clusters of strongly varying sizes. Finally, beyond the *satisfiability threshold* there are no satisfying assignments. Many of these predictions have since been proven rigorously: the prediction of the satisfiability threshold was confirmed in [DSS15], and the physics prediction of the condensation threshold in random regular NAE-k-SAT was recently confirmed in [NSS20].

[KMRT+07] predicted that Markov Chain Monte Carlo (MCMC) algorithms succeed up to the clustering threshold and no more. Since then, this threshold has emerged as the predicted limit of *all* efficient algorithms, and structural phenomena in the clustered regime have been rigorously established that (still non-rigorously) suggest algorithmic hardness. [ACO08] showed that clustering occurs at clause density $(1 + o_k(1))2^k \log k/k$, confirming the prediction of [KMRT+07]. They showed that at this clause density, long-range correlations appear in random k-SAT's solution space, in the following sense. Say variable x_i of a satisfying assignment $x \in \{\mathtt{T}, \mathtt{F}\}^N$ is *frozen* if any satisfying assignment y with $x_i \neq y_i$ is at Hamming distance $\Omega(N)$ from x. Then, in all but an $o(1)$ fraction of satisfying assignments, all but an $o_k(1)$ fraction of bits are frozen with high probability. This suggests that above this clause density, local search is unlikely to succeed, and any algorithmic solution to random k-SAT must use a qualitatively different approach.

The rigorous evidence for the algorithmic threshold consists of exhibiting algorithms on one side and producing bounds against specific algorithms or restricted computational models on the other side. There is a long history of work on heuristic algorithms for k-SAT. The oldest heuristic is the *Davis-Putnam-Logemann-Loveland* (DPLL) algorithm [DP60, DLL61], a backtracking based search algorithm which still forms the basis for many modern SAT solvers. Other heuristics that have emerged include the *pure literal* rule [GPB82]; *unit clause propagation* [MTF90]; *shortest clause* [CR92, FS96]; *walksat* [Pap91, COFF+09]; and *Belief* and *Survey Propagation guided decimation* [MRTS07, BMZ05]. However, there is no evidence, rigorous or non-rigorous, that any of these algorithms succeed beyond clause density $O_k(2^k/k)$. (See [CO10, Table 1] for a review of these algorithms' performances.) The breakthrough result [CO10] produced the algorithm `Fix`, which provably finds a satisfying assignment with high probability up to clause density $(1 - o_k(1))2^k \log k/k$. This is the best algorithm to date, and the above physics evidence suggests that this clause density is optimal up to lower order terms.

The earliest rigorous hardness result is [LMS98], which proved that the pure literal rule does not solve random 3-SAT above clause density approximately 1.63. [AS00] generalized this result, showing that so-called *myopic algorithms* cannot solve random 3-SAT above clause density approximately 3.26. (The random 3-SAT satisfiability threshold is conjectured to be about 4.26 [MPZ02].)

For large k, the earliest hardness result is [ABM04], which showed that DPLL type algorithms require exponential running time beyond clause density $O_k(2^k/k)$. Note that this threshold is *smaller* than the clause density $(1 - o_k(1))2^k \log k/k$ where `Fix` succeeds; thus DPLL algorithms are provably suboptimal. Gamarnik and Sudan [GS17] showed that *balanced sequential local algorithms*, which include Belief and Survey Propagation guided decimation (with constant or mildly growing number of message passing rounds) cannot solve random NAE-k-SAT at clause density $(1 + o_k(1))2^{k-1} \log^2 k/k$. The quantity 2^{k-1} is the NAE-k-SAT analogue of 2^k for k-SAT. Remaining negative results are bounds against specific algorithms. [Het16] proved that Survey Propagation guided decimation (without restriction on the number of rounds) fails at clause density $(1+o_k(1))2^k \log k/k$, and [COHH17] proved that walksat fails at clause density $O_k(2^k \log^2 k/k)$. Table 1.1 summarizes these results. To date, all negative results either differ from the conjectured threshold $(1 + o_k(1))2^k \log k/k$ by a factor growing in k or are tailored to a specific algorithm.

We will show that *low degree polynomial* algorithms do not solve random k-SAT above clause density $(1 + o_k(1))\kappa^* 2^k \log k/k$ for a universal constant $\kappa^* \approx 4.911$. Low degree polynomials encompass many of the above algorithms, including `Fix`, Belief and Survey Propagation guided decimation, and local and sequential local algorithms on the factor graph. This is the first hardness result for any class of algorithms within a constant factor of the conjectured algorithmic threshold.

Low degree polynomial algorithms have recently emerged as a prominent class in average case complexity and statistical inference. As outlined in [GJW20, Appendix A], this class contains many popular and powerful frameworks, including spectral methods, local algorithms on graphs, and (approximate) message passing [DMM09, BM11, JM13, Mon19, AMS21, Sel21]. In addition, a recent flurry of work has shown that for many problems in high-dimensional statistics, including planted clique, sparse PCA, community detection, and tensor PCA, low degree polynomials are as powerful as the best polynomial-time algorithms known [HS17, HKP+17, Hop18, BKW20, KWB19, DKWB20, CHK+20, BB20, LZ20, SW20, BBK+21, BBH+21].

Reference	Algorithm or algorithm class	Clause density
[ABM04]	DPLL algorithms	$O_k(2^k/k)$
[GS17]	Balanced sequential local algorithms (NAE-k-SAT)	$(1+o_k(1))2^{k-1}\log^2 k/k$
[Het16]	Survey Propagation guided decimation	$(1+o_k(1))2^k \log k/k$
[COHH17]	Walksat	$O_k(2^k \log^2 k/k)$
This work	Low degree polynomials	$(1+o_k(1))\kappa^* 2^k \log k/k$

Table 1.1: Algorithmic hardness results for random k-SAT with large k. The conjectured algorithmic threshold is $(1+o_k(1))2^k \log k/k$.

Thus, showing that low degree polynomial algorithms fail at some threshold provides evidence that all polynomial-time algorithms fail at that threshold.

Our result gives strong evidence that the algorithmic threshold is within a constant factor of $2^k \log k/k$. Because our techniques link clustering to hardness, we believe the true algorithmic threshold is $(1+o_k(1))2^k \log k/k$, matching `Fix` and the onset of clustering; we leave the question of closing this constant factor gap as an important open problem.

1.3 Optimization of Mean-Field Spin Glasses

The Sherrington-Kirkpatrick model was introduced in [SK75] as a mean-field model for spin glasses. The mixed p-spin model generalizes this model to interactions involving more than two spins, and 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 $\sum_{p\in 2\mathbb{N}} 2^p \gamma_p^2 < \infty$. The mixed even p-spin Hamiltonian H_N is

$$H_N(\boldsymbol{\sigma}) = \langle \boldsymbol{h}, \boldsymbol{\sigma}\rangle + \widetilde{H}_N(\boldsymbol{\sigma}), \quad \text{where} \tag{1.3.1}$$

$$\widetilde{H}_N(\boldsymbol{\sigma}) = \sum_{p\in 2\mathbb{N}} \frac{\gamma_p}{N^{(p-1)/2}} \langle \mathbf{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p}\rangle. \tag{1.3.2}$$

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

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

The Hamiltonian H_N defines a random objective function, and optimizing this function is a random optimization problem. Let

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

denote the existential limits of this problem. We 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 values $\mathsf{OPT}^{\mathrm{Sp}}$ and $\mathsf{OPT}^{\mathrm{Is}}$ are given by the celebrated Parisi formula [Par79] which was proved for even models by [Tal06b, Tal06a] and in more generality by [Pan14]. 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 3.1), the result can be expressed in the following form due to Auffinger and Chen [AC17b].

Define the function space

$$\mathscr{U} = \left\{\zeta : [0,1) \to \mathbb{R}_{\geq 0} : \zeta \text{ is right-continuous and nondecreasing}, \int_0^1 \zeta(t)\, \mathrm{d}t < \infty\right\}. \tag{1.3.3}$$

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

$$\partial_t \Phi_\zeta(t,x) + \frac{1}{2}\xi''(t)\left(\partial_{xx}\Phi_\zeta(t,x) + \zeta(t)(\partial_x \Phi_\zeta(t,x))^2\right) = 0 \tag{1.3.4}$$

$$\Phi_\zeta(1,x) = |x|. \tag{1.3.5}$$

Existence and uniqueness properties for this PDE are recalled in Subsection 3.5.1. The Parisi functional $\mathsf{P}^{\mathrm{Is}} = \mathsf{P}^{\mathrm{Is}}_{\xi,h} : \mathscr{U} \to \mathbb{R}$ is given by

$$\mathsf{P}^{\mathrm{Is}}(\zeta) = \Phi_\zeta(0,h) - \frac{1}{2}\int_0^1 t\xi''(t)\zeta(t)\,\mathrm{d}t. \tag{1.3.6}$$

Theorem 1.3.1 ([AC17b, Theorem 1]). *The following identity holds.*

$$\mathsf{OPT}^{\mathrm{Is}} = \inf_{\zeta \in \mathscr{U}} \mathsf{P}^{\mathrm{Is}}(\zeta). \tag{1.3.7}$$

The infimum over $\zeta \in \mathscr{U}$ is achieved at a unique $\zeta_* \in \mathscr{U}$ as shown in [AC17b, CHL18], which can be obtained as an appropriately renormalized $\beta \to \infty$ 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 $e^{\beta H_N}/Z_N(\beta)$, which 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 [Sub21] and continued in [Mon19, AMS21, Sel21]. 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 $\|f\|_{d_{\mathrm{TV}}(J)}$ denote the total variation of f on J, expressed as the supremum over partitions:

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

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

$$\mathscr{L} = \left\{ \zeta : [0,1) \to \mathbb{R}_{\geq 0} : \zeta \text{ right-continuous}, \|\xi'' \cdot \zeta\|_{d_{\mathrm{TV}}[0,t]} < \infty \text{ for all } t \in [0,1), \int_0^1 \xi''(t)\zeta(t)\,\mathrm{d}t < \infty \right\}. \tag{1.3.8}$$

It turns out (see Subsection 3.5.1) that the definition of P^{Is} above extends from $\mathscr{U}$ to $\mathscr{L}$. Therefore we may define $\mathsf{ALG}^{\mathrm{Is}} = \mathsf{ALG}^{\mathrm{Is}}_{\xi,h}$ by

$$\mathsf{ALG}^{\mathrm{Is}} = \inf_{\zeta \in \mathscr{L}} \mathsf{P}^{\mathrm{Is}}(\zeta). \tag{1.3.9}$$

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 (1.3.9) is attained by some $\zeta \in \mathscr{U}$, and otherwise $\mathsf{ALG}^{\mathrm{Is}} < \mathsf{OPT}^{\mathrm{Is}}$.

Theorem 1.3.2 ([AMS21, Sel21]). *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 $\mathcal{A} : \mathscr{H}_N \to C_N$ such that*

$$\mathbb{P}[H_N(\mathcal{A}(H_N))/N \geq \mathsf{ALG}^{\mathrm{Is}} - \varepsilon] \geq 1 - o(1), \quad c = c(\varepsilon) > 0.$$

All of the algorithms in [Sub21, Mon19, AMS21, Sel21] 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_*}. AMP offers a great deal of flexibility, and the idea introduced in [Mon19] 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 [AMS21] 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 (1.3.7) — it implies that ζ_* is not strictly increasing, see [Sel21] for a more precise condition called "optimizability" therein. As explained in [Sel21, 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 is aligned with results on the 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 [Sub21] and coincides with the onset of exponentially many bounded index critical points, as established in [ABAČ13, Sub17]. 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.

This belief is also aligned with existing hardness results. For pure spherical and Ising p-spin glasses where $h = 0$ and $p \geq 4$ is even, $\mathsf{ALG} < \mathsf{OPT}$ always holds. In such models, [GJW20] showed that low degree polynomials cannot achieve some objective $\mathsf{OPT} - \varepsilon$ strictly smaller than OPT, extending a similar hardness result of [GJ21] for approximate message passing. [GJW21] extended the conclusions of [GJW20] to Boolean circuits of depth less than $\frac{\log n}{2 \log \log n}$. As pointed out in [Sel21, Section 6], these results extend in the Ising case to any mixed even model where $\mathsf{ALG}^{\mathrm{Is}} < \mathsf{OPT}^{\mathrm{Is}}$.

Our results will show that a class of suitably Lipschitz algorithms in the disorder coefficients, which encompasses natural formulations of gradient descent and AMP, cannot surpass ALG. To this end, we define the following distance on the space $\mathscr{H}_N$ 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 $\boldsymbol{g}(H_N)$. We equip $\mathscr{H}_N$ with the (possibly infinite) distance

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

Let $B_N = \{x \in \mathbb{R}^N : \sum_{i=1}^N x_i^2 \leq N\}$ 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 $O(1)$-Lipschitz function $\mathcal{A} : \mathscr{H}_N \to C_N$ can surpass the asymptotic value $\mathsf{ALG}^{\mathrm{Is}}$. (And similarly in the spherical case for $\mathcal{A} : \mathscr{H}_N \to B_N$ and an analogous $\mathsf{ALG}^{\mathrm{Sp}}$.)

Theorem 1.3.3. *Let $\tau, \varepsilon > 0$ be constants. For N sufficiently large, any τ-Lipschitz $\mathcal{A} : \mathscr{H}_N \to C_N$ satisfies*

$$\mathbb{P}\left[H_N(\mathcal{A}(H_N))/N \geq \mathsf{ALG}^{\mathrm{Is}} + \varepsilon\right] \leq \exp(-cN), \quad c = c(\xi, h, \varepsilon, \tau) > 0.$$

Note that the Lipschitz condition $\|\mathcal{A}(H_N) - \mathcal{A}(H'_N)\|_2 \leq \tau \|H_N - H'_N\|_2$ holds vacuously when the latter distance is infinite.

The algorithms of [Mon19, AMS21, Sel21] are $O(1)$-Lipschitz in the sense above[3]. While the approach of [Sub21] is not Lipschitz, its performance is captured by AMP as explained in [AMS21, Remark 2.2].[4] Hence in tandem with these constructive results, Theorem 1.3.3 identifies the exact asymptotic value achievable by Lipschitz functions $\mathcal{A} : \mathscr{H}_N \to C_N$ (assuming the existence of a minimizer $\zeta_* \in \mathscr{L}$ as required in Theorem 1.3.2). 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 1.3.3 is best possible up to the value of c, being achieved even for the constant function $\mathcal{A}(H_N) = (1, 1, \ldots, 1)$.

Abstractly, the assumption that $\mathcal{A}$ is Lipschitz is geometrically natural and brings us near the well-studied setting of Lipschitz selection [Shv84, PY95, Shv02, FS18]. Here one is given a metrized family $\mathcal{S}$ of subsets inside a metric space X. The goal is to find a function $f : \mathcal{S} \to X$ with the selection property that $f(S) \in S$ for all $S \in \mathcal{S}$, and such that f has a small Lipschitz constant. Indeed a Lipschitz function $\mathcal{A} : \mathscr{H}_N \to C_N$ achieving energy E is almost the same as a Lipschitz selector for the level sets

$$S_E(H_N) = \{\boldsymbol{\sigma} \in C_N : H_N(\boldsymbol{\sigma})/N \geq E\}$$

metrized by the norm on $\mathscr{H}_N$ defined above (and leaving aside the fact that $S_E(H_N)$ may not determine H_N). Of course we can only hope for $\mathcal{A}(H_N) \in S_E(H_N)$ to hold with high probability, since $S_E(H_N)$ is empty with small but positive probability.

[3] 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 $\mathcal{A} : \mathscr{H}_N \to C_N$ with the same performance.

[4] We also outline a similar impossibility result for a family of variants of Subag's algorithm in Subsection 3.2.7.

In particular, many natural optimization algorithms satisfy the Lipschitz property above on a set $K_N \subseteq \mathscr{H}_N$ of inputs with $1 - \exp(-\Omega(N))$ probability; this suffices just as well for Theorem 1.3.3 (see Subsection 3.7.1). As explained in Section 3.7, 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 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.

1.3.1 Further Background

We now describe some other results on algorithmically optimizing spin glass Hamiltonians. First, in the worst case over the disorder $\mathbf{G}^{(p)}$, achieving any constant approximation ratio to the true optimum value is known to be quasi-NP hard even for degree 2 polynomials [ABE+05, BBH+12]. 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 [GK21b]. 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 suboptimal local optima and other finite-index saddle points at a given energy level. These are understood to serve as barriers to efficient optimization, and a non-rigorous study was undertaken in [CLR03, CLR05, Par06] followed by a great deal of recent progress in [ABAČ13, ABA13, Sub17, McK21, Kiv21, SZ21]. Notably because the true maximum value of H_N is nothing but its largest critical value, the first moment results of [ABAČ13] combined with the second moment results of [Sub17] gave an alternate self-contained proof of the Parisi formula for the ground state in pure spherical models. In a related spirit, [Cha09, DEZ15, CS17, CHL18] have shown that mixed even p-spin Hamiltonians typically contain exponentially many well-separated near-global maxima.

Other works such as [CK94, BCKM98, BADG06, BAGJ20] 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 [EKZ21, AJK+21] showing that the Glauber dynamics for the Sherrington-Kirkpatrick model mix rapidly at high temperature.

1.4 The Overlap Gap Property as a Barrier to Algorithms

The study of solution geometry as a barrier to algorithmic hardness began in the seminal work [ACO08], which showed the presence of a clustering phase transition in random k-SAT and q-coloring that coincides with the failure of the best known efficient algorithm. Moreover, [ACO08] conjectured that this is a general phenomenon: in a general random CSP exhibiting a clustering transition, polynomial-time algorithms succeed up to this transition and no further. [COE15] showed the presence of a similar clustering transition for the problem of maximum independent set on a sparse random graph, which also coincides with the failure

of the best known polynomial-time algorithm. This intuition is believed to hold in significant generality in random optimization problems.[5]

In the past several years, a line of work [GS14, RV17, GS17, CGPR19, GJ21, GJW20, Wei20, GK21a, GJW21] on the Overlap Gap Property (OGP) has made substantial progress on rigorously linking solution geometry clustering in random optimization problems to algorithmic hardness. A survey can be found in [Gam21]. Initiated by Gamarnik and Sudan in [GS14], the OGP technique can be summarized by the following two steps.

1. Show that if a suitably stable algorithm can construct one solution at the desired objective (or clause density), it can in fact construct a constellation of such solutions with a specified geometric structure.

2. Separately, show that with high probability, this constellation does not occur in the solution space. By contradiction, this implies that no stable algorithm can reach this objective.

In the original version of the OGP [GS14], the forbidden constellation was a pair of solutions with medium overlap, hence the name "overlap gap."

Over many problems, a pattern has emerged where the classic OGP argument shows the failure of stable algorithms above an intermediate objective, smaller than the existential threshold but larger than the algorithmic threshold. Intuitively, this is because the classic OGP is a stronger condition than clustering: in the clustered regime, it is possible to have a small minority of solution pairs with medium overlap, while the classic OGP requires there to be zero such pairs. On the other side, the condition that there does not exist two solutions with medium overlap is a weaker condition than the condition that no solution exists at all, so the onset of OGP occurs below the existential threshold. To tighten the objective where stable algorithms are proven to fail, subsequent works have considered "multi-OGPs," which use forbidden structures involving several solutions. Multi-OGPs improve on the classic OGP if the structure becomes forbidden at a lower objective than the classic OGP.

The first Multi-OGP was introduced in [RV17], which showed that local algorithms (also called *factors of i.i.d.*) cannot attain any objective asymptotically larger than the believed algorithmic limit. Using a different multi-OGP, [Wei20] extended this hardness result to the more powerful class of low degree polynomials. These results are an important success story of multi-OGPs, showing that this technique has the potential to show computational hardness at objectives matching the best algorithms.

OGPs have also been used to study random k-SAT and spin glass optimization, the two problems under consideration in this thesis. For random k-SAT, [DMMZ08] showed that the classic OGP occurs at clause density $(1+o_k(1))\frac{1}{2}2^k \log 2$. (This work predates the OGP literature and was understood as a confirmation of clustering; the rigorous connection to algorithms was realized later.) The aforementioned result of [GS17] against balanced sequential local algorithms at clause density $(1+o_k(1))2^{k-1}\log^2 k/k$ is proved by a multi-OGP. For spin glasses, the aforementioned results of [GJ21, GJW20, GJW21] against AMP, low degree polynomials, and boolean circuits at objective $\mathsf{OPT}-\varepsilon$ are proved by a classic OGP.

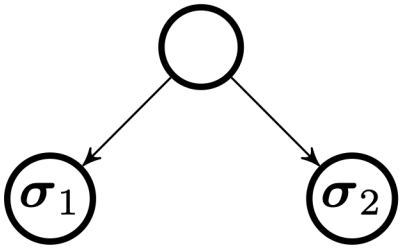
One of the main challenges in a multi-OGP argument is the design of the forbidden structure. Whereas in the classic OGP, there is essentially one possible forbidden structure, with multi-OGPs we have significantly more freedom. If our goal is to show hardness at or close to a problem's algorithmic limit, choosing the correct forbidden structure is crucial; a suboptimal forbidden structure will not be forbidden at the objective or clause density we desire, causing the argument to fail. Moreover, as these structures become more complex, it becomes necessary to develop new techniques to prove that these structures do not occur. By choosing forbidden structures that better capture these problems' solution landscapes and introducing tools for their analysis, we improve the above hardness results to clause density $(1+o_k(1))\kappa^* 2^k \log k/k$ for random k-SAT and to the exact value of ALG for mean-field spin glasses.

The design of our multi-OGPs is a significant departure from previous work. Previous OGPs and multi-OGPs all use one of the following three forbidden structures, see Figure 1.1.

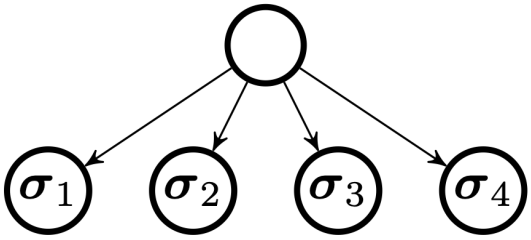
- Classic OGP: two solutions with medium overlap [GS14, CGPR19, GJ21, GJW20, GJW21].

- Star OGP: several solutions with approximately the same pairwise overlap [RV17, GS17, GK21a].

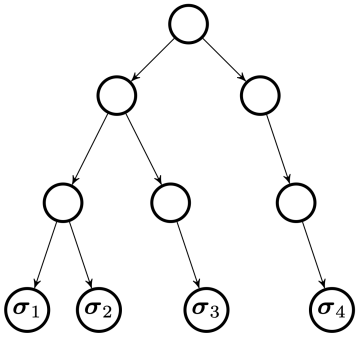
[5] Perceptron is a notable exception; see Chapter 4 for a discussion. We will see that OGP appears to accurately characterize hardness, even though the clustering heuristic is wrong.



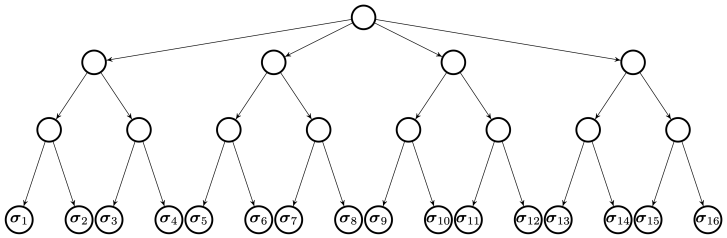
(a) **Classic OGP**: σ_1, σ_2 have medium overlap.



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



(c) **Ladder OGP**: many solutions σ_i, medium "multi-overlaps" with $\{\sigma_1, \ldots, \sigma_{i-1}\}$.



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

Figure 1.1: Schematics of forbidden structures in OGP arguments.

- Ladder OGP: several solutions, where the i-th solution ($i \geq 2$) has medium "multi-overlap" with the first $i-1$ solutions, for a problem-specific notion of multi-overlap of one solution with several solutions [Wei20].

For random k-SAT, our multi-OGP is a generalization of the ladder multi-OGP introduced in [Wei20] for the maximum independent set problem. Like in [Wei20], the analysis of the forbidden structure is by a first moment analysis: we express the log first moment of the forbidden structure as a free entropy, and our goal is to find a structure making this quantity negative. However, adapting this technique to random k-SAT requires new ideas. In maximum independent set, the independence of the Erdős-Rényi graph's edges makes the analysis of the forbidden structure tractable by the principle of deferred decisions. In contrast, the forbidden structure in random k-SAT has complex dependencies which make a tight analysis difficult. It is a priori unclear what forbidden structure to choose.

We overcome these challenges by an appropriate understanding of the energy landscape of random k-SAT. We introduce a notion of overlap profile of several assignments and define our multi-OGP's forbidden structure in terms of this profile. This is a key innovation, as all previous (multi-)OGPs have not required the overlap profile's full power. We then perform a free entropy analysis tailored to random k-SAT to show the forbidden structure does not occur. A technical overview of our proof can be found in Section 2.2.

For mean-field spin glass optimization, we introduce a new forbidden structure consisting of 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 establish this branching OGP as follows. Using a version of the Guerra-Talagrand interpolation, which we take to zero temperature, we derive an upper bound for the maximum average energy of configurations

arranged into the desired structure. 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 a symmetric branching tree, 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 piecewise-constant function that depends on the tree. By making the tree branch rapidly, the function κ can be arranged to decrease as rapidly as desired. As a result, the functions $\kappa\zeta$ are dense in the space $\mathscr{L}$. Thus, we may choose a tree and ζ such that $\mathsf{P}(\kappa\zeta)$ is arbitrarily close to ALG.

Roughly speaking, we show that an overlap concentrated $\mathcal{A}$ allows the construction of an arbitrary ultrametric constellation of outputs. Consequently, if $\mathcal{A}$ outputs points with energy at least $\mathsf{ALG} + \varepsilon$, then $\mathcal{A}$ run on the appropriate family of ultrametrically correlated Hamiltonians will output the forbidden structure above, a contradiction. Some additional complications are created by the fact that $\mathbb{E}[\mathcal{A}(H_N)]$ 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.

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. More precisely, consider a spherical model ξ without external field; we restrict to this case for convenience. Consider a fixed ultrametric overlap structure of inputs, whose corresponding rooted tree (cf. Subsection 3.6.2) does not contain a full depth-D binary tree. We prove that if $\mathsf{ALG}^{\text{Sp}} < \mathsf{OPT}^{\text{Sp}}$, with high probability there exists a constellation of inputs with this overlap structure where each input achieves energy at least $\mathsf{ALG}^{\text{Sp}} + \varepsilon_{\xi,D}$, for a constant $\varepsilon_{\xi,D} > 0$ depending only on ξ, D.

Remark 1.4.1. Our proof of the branching OGP is the first to leverage powerful tools from probability to prove the non-existence of the forbidden structure. All prior OGPs and multi-OGPs are established by a first moment analysis. We regard this connection with probability as an important methodological contribution. Guerra's interpolation allows us to prove a hardness result that is tight in the strong sense of characterizing the exact point ALG where hardness occurs. This is the first hardness result in any random optimization problem that is tight in this sense; 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 double limit of $N \to \infty$ followed by $d \to \infty$.

Remark 1.4.2. The significance of ultrametricity in mean-field spin glasses began with [Par79] and has played an enormous role in guiding the mathematical understanding of the low temperature regime in works such as [Rue87, Pan13a, Jag17, CS21]. Ultrametricity also appears naturally in the context of optimization algorithms. Indeed in [Sub21, Remark 6], [AM20, Section 3.4] and [Sel21, 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 based on the branching OGP can be viewed as establishing a sharp converse — the existence of essentially arbitrary ultrametric configurations at a given energy level is *equivalent* to achievability by Lipschitz $\mathcal{A}$.

Remark 1.4.3. Since the algorithm of Subag in [Sub21] uses the top eigenvector of the Hessian $\nabla^2 H_N(\boldsymbol{x})$ for various $\boldsymbol{x} \in B_N$, it is not Lipschitz in H_N in the sense required by Theorem 1.3.3. However a different branching OGP argument shows that a stylized class of algorithms which includes a natural variant of Subag's approach is also incapable of achieving energy $\mathsf{ALG} + \varepsilon$. This argument uses only a single Hamiltonian, constructing a branching tree structure using the internal randomness of the algorithm. In this sense, it bears resemblance to the original OGP analysis of [GS14]. An outline is given in Subsection 3.2.7.

Chapter 2

Random k-SAT

This chapter studies algorithmic hardness in the random k-SAT problem and is structured as follows.

- Section 2.1 states our main results. Theorem 2.1.6 shows hardness for low degree polynomials, while Theorem 2.1.13 shows hardness for the more restricted class of *local algorithms* with a stronger guarantee on the success probability. Theorem 2.1.14 gives a converse, that both computation classes can solve random k-SAT at clause densities where algorithms are known to succeed.
- Section 2.2 gives a technical overview of the design and analysis of the ladder multi-OGP used in our hardness proofs.
- Sections 2.3 through 2.5 are devoted to the proof of Theorem 2.1.6. Section 2.3 sets up the OGP argument and proves Theorem 2.1.6 assuming Propositions 2.3.7(a,c), that outputs of the low degree polynomial are stable and that the main multi-OGP holds. Sections 2.4 and 2.5 prove these propositions.
- Section 2.6 proves Theorem 2.1.13.
- Section 2.7 shows that a class of algorithms we call *local memory algorithms*, which include `Fix` and sequential local algorithms, can be simulated by local algorithms and low degree polynomials.
- Using these simulation results, Section 2.8 proves Theorem 2.1.14.

Notation and Preliminaries

For all positive integers N, $[N]$ denotes the set $\{1, \dots, N\}$. For two assignments $x, y \in \{\mathtt{T}, \mathtt{F}\}^N$, let $\Delta(x, y) = \frac{1}{N}|\{i \in [N] : x_i \neq y_i\}|$ denote the normalized Hamming distance. We occasionally consider assignments $x, y \in \{\mathtt{T}, \mathtt{F}, \mathtt{err}\}^N$ which allow an error symbol; for such assignments the definition of Δ extends verbatim.

Throughout, log denotes the natural logarithm. The binary entropy function $H : [0, 1] \to [0, \log 2]$ is $H(x) = -x \log x - (1 - x) \log(1 - x)$. We often use the basic inequality $H(x) \leq x \log \frac{e}{x}$. We also overload notation and denote by $H(\cdot)$ the entropy of certain distributions. These will be defined where first used.

We reiterate that we work in the double limit $N \to \infty$ followed by $k \to \infty$. The notations $O(\cdot), \Omega(\cdot), o(\cdot), \omega(\cdot)$ indicate asymptotic behavior in N, suppressing any dependence on k. With a tilde (e.g. $\widetilde{O}(\cdot)$) these notations also suppress polylog(N) factors. When subscripted with k, these notations indicate asymptotic behavior in k of a quantity independent of N.

2.1 Results

Throughout this chapter, $\mathcal{V} = \{x_1, \dots, x_N\}$ denotes a set of propositional variables. The set of corresponding literals, consisting of the variables in $\mathcal{V}$ and their negations, is $\mathcal{L} = \{x_1, \dots, x_N, \bar{x}_1, \dots, \bar{x}_N\}$. Let $\Omega_k(N, M)$ denote the set of all k-CNF formulas over $\mathcal{V}$ with M clauses. We allow literals to appear multiple times in a clause and clauses to appear multiple times in a formula. We treat each $\Phi \in \Omega_k(N, M)$ as an ordered

M-tuple of clauses, each of which is an ordered k-tuple of literals. Let Φ_i ($i \in [M]$) denote the ith clause of Φ and $\Phi_{i,j}$ ($j \in [k]$) denote the jth literal of Φ_i. The central object of this chapter is the following distribution.

Definition 2.1.1 (Random k-SAT)**.** The *random k-SAT* distribution $\Phi_k(N, M)$ is the law of a uniformly random sample from $\Omega_k(N, M)$. Equivalently, we can sample $\Phi \sim \Phi_k(N, M)$ by sampling the literals $\Phi_{i,j}$ i.i.d. from $\mathrm{unif}(\mathcal{L})$.

We now define the constant κ^* in our hardness results. Define the function $\iota : (1, +\infty) \to \mathbb{R}$ by

$$\iota(\beta) = \frac{\beta}{1 - \beta e^{-(\beta-1)}}.$$

One easily checks that ι is strictly convex, with $\iota(\beta) \to +\infty$ when $\beta \to 1^+$ or $\beta \to +\infty$. Let $\kappa^* = \min \iota(\beta) \approx 4.911$. The minimum is attained at $\beta^* \approx 3.513$, the unique solution to $\beta^2 e^{-(\beta-1)} = 1$ in $(1, +\infty)$.

2.1.1 Computational Hardness for Low Degree Polynomials

We study the class of low degree polynomial algorithms, defined as follows. This is the same computational model considered in [GJW20, Wei20].

Definition 2.1.2 (Low degree polynomial)**.** A *degree-D polynomial* is a function $f : \mathbb{R}^n \to \mathbb{R}^N$ of the form

$$f(x) = (f_1(x), \dots, f_N(x)),$$

where each $f_i : \mathbb{R}^n \to \mathbb{R}$ is a multivariate polynomial (in the ordinary sense) with real coefficients of degree at most D. A *random degree-D polynomial* is defined similarly, except the coefficients are random (but independent of the input x). Formally, for an arbitrary probability space $(\Omega, \mathbb{P}_\omega)$, a random degree-$D$ polynomial is a function $f : \mathbb{R}^n \times \Omega \to \mathbb{R}^N$ such that for each $\omega \in \Omega$, $f(\cdot, \omega)$ is a degree-D polynomial.

Remark 2.1.3. We will see in Lemma 2.3.1 that randomness does not increase the power of the class of low degree polynomials. Informally, this is because if a random polynomial $f : \mathbb{R}^n \times \Omega \to \mathbb{R}^N$ solves random k-SAT in the appropriate sense, there exists a seed ω such that the deterministic polynomial $f(\cdot, \omega)$ also solves random k-SAT.

We now define how to encode a k-SAT formula as an input to a low degree polynomial. Define an arbitrary total order on $\mathcal{L}$. We encode each $\Phi \in \Omega_k(N, M)$ as a "one-hot" vector of indicators $\Phi_{i,j,s}$ ($i \in [M], j \in [k], s \in [2N]$) that $\Phi_{i,j}$ is the sth element of $\mathcal{L}$. This encoding is an element of $\{0, 1\}^n$, where $n = m \cdot k \cdot 2N$. Slightly abusing notation, we identify Φ with this encoding.

Next, we define how to interpret the output of a low degree polynomial as a Boolean assignment. We introduce the symbol `err` and define the function $\texttt{round} : \mathbb{R} \to \{\texttt{T}, \texttt{F}, \texttt{err}\}$ by

$$\texttt{round}(x) = \begin{cases} \texttt{T} & x \geq 1, \\ \texttt{F} & x \leq -1, \\ \texttt{err} & \text{otherwise.} \end{cases}$$

When applied to a real-valued vector, `round` is applied coordinate-wise. Thus, outputs of the polynomial that are at least 1 represent true, outputs that are at most -1 represent false, and outputs in the interval $(-1, 1)$ are errors. It is important to exclude $(-1, 1)$ so that a small change in the polynomial output cannot induce a large change in (the valid outputs of) the assignment.

In the following definition, we relax the notion of satisfying assignment in two ways: we allow the algorithm to make mistakes in a small fraction η of positions (including all `err` outputs and possibly others), and after repairing these mistakes we allow a small fraction ν of clauses to not be satisfied.

Definition 2.1.4 ((η, ν)-satisfy)**.** Let $\eta, \nu \in [0, 1]$. An assignment $x \in \{\texttt{T}, \texttt{F}\}^N$ *ν-satisfies* $\Phi \in \Omega_k(N, M)$ if it satisfies at least $(1 - \nu)M$ clauses of Φ. Moreover, $x \in \{\texttt{T}, \texttt{F}, \texttt{err}\}^N$ *(η, ν)-satisfies* Φ if there exists $y \in \{\texttt{T}, \texttt{F}\}^N$ such that $\Delta(x, y) \leq \eta$ and y ν-satisfies Φ.

We remark that any x with more than ηN entries equal to $\texttt{err}$ does not (η, ν)-satisfy Φ. We will show that for small η, ν independent of N, a low degree polynomial cannot produce a satisfying assignment for random k-SAT even in this relaxed sense. Formally, we will show hardness for the following notion of solve.

Definition 2.1.5 ($(\delta, \gamma, \eta, \nu)$-solve)**.** Let $\delta, \eta, \nu \in [0, 1]$ and $\gamma \geq 1$. A random polynomial $f : \mathbb{R}^n \times \Omega \to \mathbb{R}^N$ *$(\delta, \gamma, \eta, \nu)$-solves* $\Phi_k(N, M)$ if the following conditions hold.

(a) $\mathbb{P}_{\Phi,\omega}\left[(\texttt{round} \circ f)(\Phi, \omega) \text{ } (\eta, \nu)\text{-satisfies } \Phi\right] \geq 1 - \delta$.

(b) $\mathbb{E}_{\Phi,\omega}\left[\|f(\Phi, \omega)\|_2^2\right] \leq \gamma N$.

Here, δ is the algorithm's failure probability and γ is a normalization parameter. We think of γ as a large constant; condition (b) is necessary because otherwise we can scale f to make the condition that valid outputs of f are outside the interval $(-1, 1)$ meaningless.

The following theorem is our main result, that no low degree polynomial can solve random k-SAT at clause density $\kappa 2^k \log k/k$ for any $\kappa > \kappa^*$.

Theorem 2.1.6 (Hardness for low degree polynomials)**.** *Fix $\kappa > \kappa^*$. Let $\alpha = \kappa 2^k \log k/k$ and $M = \lfloor \alpha N \rfloor$. There exists $k^* = k^*(\kappa) > 0$ such that for any $k \geq k^*$, there exists $N^* > 0$, $\eta = \Omega_k(k^{-1})$, $\nu = \frac{1}{k^2 2^k}$, and $C_1, C_2 > 0$ (depending on κ, k) such that the following holds. If $N \geq N^*$, $\gamma \geq 1$, $1 \leq D \leq \frac{C_1 N}{\gamma \log N}$ and*

$$\delta \leq \exp\left(-C_2 \gamma D \log N\right),$$

then there is no random degree-D polynomial that $(\delta, \gamma, \eta, \nu)$-solves $\Phi_k(N, M)$.

The only property of low degree polynomials we use is their smoothness, in the sense of Proposition 2.5.2. Thus Theorem 2.1.6 applies to any algorithm satisfying the conclusion of this proposition.

Note that Theorem 2.1.6 only rules out algorithms succeeding with quite large probability. This is a limitation of our methods, shared by all results leveraging OGP to show hardness for low degree polynomials [GJW20, Wei20]. Our converse achievability result, Theorem 2.1.14, will show that at clause densities where efficient algorithms solving random k-SAT exist, they can be simulated by low degree polynomials and succeed with probability larger than that forbidden by Theorem 2.1.6. We will also see in Theorem 2.1.13 that local algorithms, a more restricted computation class that nonetheless simulates $\texttt{Fix}$, as well as Belief and Survey Propagation Guided Decimation, cannot solve random k-SAT with even very small probability.

The constant κ^* can likely be optimized further. However, without further conceptual insights our methods stall at a value of κ^* strictly larger than 1, lower bounded by approximately 1.716. Thus further ideas are needed to close the constant factor gap between our hardness results and the best algorithms. See Section 2.9 for a discussion of these points. Despite this barrier, we believe the algorithmic phase transition for low degree polynomials does occur at clause density $(1 + o_k(1))2^k \log k/k$, matching the physics prediction and positive results. This is formalized in the following conjecture, which we leave as an open problem.

Conjecture 2.1.7. *Theorem 2.1.6 (and Theorem 2.1.13) holds for all $\kappa > 1$.*

2.1.2 Computational Hardness for Local Algorithms

We now consider local algorithms on the factor graph. We first define the factor graph of a k-SAT instance.

Definition 2.1.8 (Factor graph)**.** The factor graph of $\Phi \in \Omega_k(N, M)$ is a signed bipartite graph (G, ρ), where $G = (\text{Va}_G, \text{Cl}_G, E_G)$ is a bipartite graph with left-vertices Va_G, right-vertices Cl_G, and edges E_G, and $\rho : E_G \to \{\texttt{T}, \texttt{F}\}$ associates each edge with a polarity. Here, $\text{Va}_G = \{v_1, \ldots, v_N\}$ and $\text{Cl}_G = \{c_1, \ldots, c_M\}$ represent the variables and clauses of Φ. Every literal x_j or $\bar{x}_j$ in clause Φ_i corresponds to an edge $e = (v_j, c_i) \in E_G$. Edge e has polarity $\rho_G(e) = \texttt{T}$ if the literal is x_j and $\rho_G(e) = \texttt{F}$ if the literal is $\bar{x}_j$.

To define local algorithms, we first introduce formalism for rooted graphs. Let $(\Omega, \mathbb{P}_\omega)$ be an arbitrary probability space.

Definition 2.1.9 (Rooted decorated bipartite graph)**.** A *decorated bipartite graph* is a tuple (G, ρ, φ). Here $G = (\mathrm{Va}_G, \mathrm{Cl}_G, E_G)$ is a bipartite graph and $V_G = \mathrm{Va}_G \cup \mathrm{Cl}_G$. Moreover, ρ, φ are maps $\rho : E_G \to \{\mathtt{T}, \mathtt{F}\}$ and $\varphi : V_G \cup E_G \to \Omega$. A *rooted decorated bipartite graph* is a tuple (v, G, ρ, φ), where (G, ρ, φ) is a decorated bipartite graph and $v \in V_G$.

Let Λ denote the set of rooted decorated bipartite graphs. Two such graphs are isomorphic of there exists a bijection between them preserving $v, \mathrm{Va}_G, \mathrm{Cl}_G, E_G, \rho, \varphi$.

Definition 2.1.10 (r-neighborhood)**.** Let $(v, G, \rho, \varphi) \in \Lambda$ and $r \in \mathbb{N}$. Define the r-neighborhood $N_r(v, G) = (v, G')$, where $\mathrm{Va}_{G'} \subseteq \mathrm{Va}_G$, $\mathrm{Cl}_{G'} \subseteq \mathrm{Cl}_G$ are the sets of vertices reachable from v by a path of length at most r and $E_{G'}$ is the set of edges on those paths. Further, define $N_r(v, G, \rho, \varphi) = (v, G', \rho', \varphi') \in \Lambda$, where $(v, G') = N_r(v, G)$ and $\rho' = \rho\big|_{G'}$, $\varphi' = \varphi\big|_{G'}$ are the restrictions of ρ, φ to G'.

Definition 2.1.11 (r-local function)**.** A function f with domain Λ is *r-local* if the value of $f(v, G, \rho, \varphi)$ depends only on the isomorphism class of $N_r(v, G, \rho, \varphi)$.

In other words, a local function has access to the topology of the r-neighborhood, the vertex and edge decorations, and the location of the root, but not the identities of the vertices and edges.

Definition 2.1.12 (r-local algorithm)**.** Let f be an r-local function with codomain $\{\mathtt{T}, \mathtt{F}\}$. The r-local algorithm based on f, denoted $\mathcal{A}_f$, runs as follows on input $\Phi \in \Omega_k(N, M)$ with factor graph (G, ρ).

(1) Sample $\varphi \sim (\Omega, \mathbb{P}_\omega)^{\otimes(V_G \cup E_G)}$ (i.e. each output of $\varphi : V_G \cup E_G \to \Omega$ is sampled i.i.d. from $(\Omega, \mathbb{P}_\omega)$) independently of Φ.

(2) For each $v = v_i \in \mathrm{Va}_G$, set $x_i = f(v, G, \rho, \varphi)$.

(3) Output $(x_1, \ldots, x_N) \in \{\mathtt{T}, \mathtt{F}\}^N$.

We now state our hardness result for local algorithms.

Theorem 2.1.13 (Hardness for local algorithms)**.** *Fix $\kappa > \kappa^*$. Let $\alpha = \kappa 2^k \log k/k$ and $M = \lfloor \alpha N \rfloor$. There exists $k^* = k^*(\kappa) > 0$ such that for any $k \geq k^*$, there exists $\eta = \Omega_k(k^{-2})$ (depending on κ, k) and $\nu = \frac{1}{k^2 2^k}$ such that the following holds. For all $r \in \mathbb{N}$, there exists $N^* > 0$ (depending on κ, k, r) such that if $n \geq N^*$, then for any r-local algorithm $\mathcal{A}$ with output in $\{\mathtt{T}, \mathtt{F}\}^N$,*

$$\mathbb{P}\left[\mathcal{A}(\Phi) \ (\eta, \nu)\text{-satisfies } \Phi\right] \leq \exp(-\widetilde{\Omega}(N^{1/3})).$$

The probability is over the randomness of $\Phi \sim \Phi_k(N, M)$ and the (independent) internal randomness of $\mathcal{A}$.

This theorem rules out a much smaller success probability than Theorem 2.1.6 because our OGP argument in this setting can leverage concentration properties of local algorithms, which are considerably stronger than stability properties of low degree polynomials.

2.1.3 Achievability Results

The following result shows that local algorithms and *constant* degree polynomials solve random k-SAT at clause density $(1 - \varepsilon)2^k \log k/k$ for any $\varepsilon > 0$. This gives a lower bound on the algorithmic phase transition within a constant factor and provides a converse to Theorems 2.1.6 and 2.1.13.

Theorem 2.1.14. *Fix $\varepsilon > 0$. Let $\alpha = (1 - \varepsilon)2^k \log k/k$ and $M = \lfloor \alpha N \rfloor$. There exists $k^* = k^*(\varepsilon) > 0$ such that for any $k \geq k^*$ and $\eta > k^{-12}$, there exist $N^*, r, D, \gamma > 0$ and a sequence $\delta(N) = o(1)$ (dependent on ε, k, η) such that the following holds for all $N \geq N^*$.*

(a) There exists an r-local algorithm $\mathcal{A}$ such that

$$\mathbb{P}\left[\mathcal{A}(\Phi) \ (\eta, 0)\text{-satisfies } \Phi\right] \geq 1 - \delta(N).$$

(b) There exists a (deterministic) degree-D polynomial that $(\delta(N), \gamma, \eta, 0)$-solves $\Phi_k(N, M)$.

There also exists a sequence $\nu(n) = o(1)$ (dependent on ε, k, η) such that the following holds for all $N \geq N^*$.

(c) There exists an r-local algorithm $\mathcal{A}$ such that

$$\mathbb{P}\left[\mathcal{A}(\Phi) \text{ } (\eta, \nu(N))\text{-satisfies } \Phi\right] \geq 1 - \exp(-\widetilde{\Omega}(N^{1/3})).$$

(d) There exists a (deterministic) degree-D polynomial that $(\exp(-\widetilde{\Omega}(N^{1/5})), \gamma, \eta, \nu(N))$-solves $\Phi_k(N, M)$.

We prove this theorem by simulating the first phase of `Fix` by a local algorithm and any local algorithm by a constant degree polynomial. We can arrange both simulations to be accurate within an arbitrarily small constant (i.e. independent of N, arbitrarily small in k) normalized Hamming distance, with failure probability $\exp(-\Omega(N^{1/3}))$. The requirement $\eta > k^{-12}$ arises because the first phase of `Fix` produces an assignment within normalized Hamming distance k^{-12} of a satisfying assignment, which is repaired by the rest of `Fix`. We believe that it is possible to simulate the rest of `Fix` by a local algorithm, which would show Theorem 2.1.14 for any $\eta > 0$; we do not attempt this improvement. Note that k^{-12} is well within the range of η ruled out by our hardness results.

In fact, we will show that local algorithms simulate any *local memory algorithm*. In this generalization of local algorithms, the algorithm makes its local decisions in series (in a random vertex order), and each decision can leave information on the vertices it accesses, which future decisions can see. This class includes the first phase of `Fix` and the sequential local algorithms considered in [GS17]. Recall that the latter class includes Belief and Survey Propagation Guided Decimation.

In parts (c,d), where the goal is to satisfy all but an $o(1)$ fraction of clauses, Theorem 2.1.14 gives algorithms with success probability $1 - \exp(-\widetilde{\Omega}(N^{1/3}))$ and $1 - \exp(-\widetilde{\Omega}(N^{1/5}))$. This is within the range ruled out by even Theorem 2.1.6. Of course, if the goal is to satisfy *all* clauses as in parts (a,b), we cannot ensure such a high success probability because $\Phi \sim \Phi_k(N, M)$ is unsatisfiable with probability $1/\text{poly}(N)$ – for example, if the first 2^k clauses each contain variables $x_1, \ldots, x_k$ with all 2^k possible polarities.

2.2 Technical Overview of Ladder Multi-OGP

In this section, we outline the construction and analysis of the multi-OGP from which we will derive our hardness results. We introduce the ladder multi-OGP from [Wei20], which was used to prove computational hardness for maximum independent set. We discuss the challenges of extending this technique beyond maximum independent set and how we overcome these challenges for random k-SAT.

2.2.1 Ladder OGP in Maximum Independent Set

In the problem of maximum independent set, we are given a sample $G \sim G(N, d/N)$ of a sparse Erdős-Rényi graph and our task is to find a large independent set. We work in the double limit where $N \to \infty$, and then $d \to \infty$. The largest independent set of this graph has asymptotic size $(1 + o_d(1))\frac{2\log d}{d}N$ [Fri90, BGT10]. However, the best known polynomial-time algorithm [Kar76] only finds an independent set of size $(1 - o_d(1))\frac{\log d}{d}N$, half the optimum. It is believed that no polynomial-time algorithm can find an asymptotically larger independent set. Using the ladder multi-OGP, [Wei20] proves that for any $\varepsilon > 0$, low degree polynomials cannot find an independent set of size $(1 + \varepsilon)\frac{\log d}{d}N$ with a probability similar to that ruled out by Theorem 2.1.6.

Both our argument and the argument in [Wei20] use an "ensemble" version of OGP, where we require the forbidden overlap structure to not occur even among the solutions of a polynomial-sized family of problem instances. To simplify the discussion, we omit this detail and consider a forbidden structure consisting of several independent sets in the same problem instance. We will see (cf. Remark 2.4.4) that this argument can be extended to the ensemble with minor adjustments.

At a high level, the ladder OGP chains together many small negative free entropy contributions to force a free entropy to be negative. We consider the normalized log first moment

$$\frac{1}{N}\log \mathop{\mathbb{E}}_{G\sim G(N,d/N)} \#\left(\begin{array}{c} (S^{(1)}, \ldots, S^{(L)}) : S^{(1)}, \ldots, S^{(L)} \text{ are independent} \\ \text{sets of } G \text{ of size } (1+\varepsilon)\frac{\log d}{d}N \text{ satisfying } P \end{array} \right), \tag{2.2.1}$$

where P is a set of conditions on how $S^{(1)}, \dots, S^{(L)}$ overlap. The structure inside the expectation in (2.2.1) is the forbidden structure we wish to rule out. The log first moment (2.2.1) can be thought of as an annealed free entropy density of the uniform model over copies of this structure; we henceforth refer to (2.2.1) as a free entropy. If (2.2.1) is negative, then this structure does not occur with high probability and the multi-OGP occurs.

The key idea in [Wei20] is to set $P = P_2 \cap P_3 \cap \cdots \cap P_L$, where P_ℓ is a condition on how $S^{(\ell)}$ overlaps with $S^{(1)}, \dots, S^{(\ell-1)}$, such that the following occurs for all $2 \le \ell \le L$.

(1) Let $\mathcal{E}_\ell$ denote (2.2.1) with $(S^{(1)}, \dots, S^{(\ell)})$ in place of $(S^{(1)}, \dots, S^{(L)})$ and $P_2 \cap \cdots \cap P_\ell$ in place of P. Then, $\mathcal{E}_\ell$ is smaller than $\mathcal{E}_{\ell-1}$ by an amount bounded away from 0. Informally, P_ℓ requires $S^{(\ell)}$ to overlap with its predecessors in a way that contributes a small negative free entropy to (2.2.1).

(2) For any fixed $S^{(1)}, \dots, S^{(\ell-1)}$, if $S^{(\ell)}$ starts at $S^{(\ell-1)}$, evolves by small steps, and eventually evolves far away from all of $S^{(1)}, \dots, S^{(\ell-1)}$, then at some point along this evolution the condition P_ℓ occurs. Informally, P_ℓ defines a moat that a stably evolving $S^{(\ell)}$ must cross.

Due to condition (1), if we set L large enough, (2.2.1) becomes negative, and the structure in (2.2.1) is forbidden with high probability. Suppose a low degree polynomial can find a size $(1+\varepsilon)\frac{\log d}{d}N$ independent set with large enough probability. Because the outputs of a low degree polynomial on a sequence of correlated problem instances is (with nontrivial probability) a stable sequence, condition (2) allows us to find a subsequence of L outputs forming the forbidden structure. Namely, we take $S^{(1)}$ to be the first output in the sequence, and for $\ell \ge 2$ we take $S^{(\ell)}$ to be the first output after $S^{(\ell-1)}$ such that P_ℓ holds. This derives the desired contradiction.

The main technical challenge is to design the P_ℓ such that both (1) and (2) hold. To do this, one must construct a moat topologically disconnecting a high-dimensional space such that, for all values of $S^{(\ell)}$ in the moat, the free entropy decrease in condition (1) occurs. The requirement that the moat topologically disconnects the space gives us little control, and therein lies the difficulty.

[Wei20, Proposition 2.3] carries out this approach elegantly by defining P_ℓ as the condition that

$$\left| S^{(\ell)} \setminus \left(S^{(1)} \cup \cdots \cup S^{(\ell-1)} \right) \right| \in \left[\frac{\varepsilon \log d}{4d} N, \frac{\varepsilon \log d}{2d} N \right]$$

and proving that the free entropy decrease in condition (1) occurs.

Extending this technique beyond maximum independent set requires new ideas. In maximum independent set, due to the independence of the edges of $G \sim G(N, d/N)$, the expectation in (2.2.1) is essentially controlled by the total number of non-edges in the union $S^{(1)} \cup \cdots \cup S^{(L)}$, for $S^{(1)}, \dots, S^{(L)}$ with overlap structure satisfying P. This fact makes the analysis of (2.2.1) tractable and shows in the relative simplicity of the moats P_ℓ, which only consider $S^{(\ell)}$'s non-intersection with the union of its predecessors.

In random k-SAT and other problems, the corresponding free entropy is highly dependent and more tools are needed to carry out this technique. We develop these tools for random k-SAT. The forbidden structure we devise will take into account more fine-grained overlap information than previous work.

2.2.2 Ladder Multi-OGP for Random k-SAT and Our Contributions

This paper shows a ladder multi-OGP for random k-SAT at clause density $(1 + o_k(1))\kappa^* 2^k \log k/k$. We leverage this OGP to show our hardness results.

Prior to this work, Gamarnik and Sudan [GS17] used a (non-ensemble) multi-OGP to prove that balanced sequential local algorithms do not solve random NAE-k-SAT beyond clause density $(1+o_k(1))2^{k-1}\log^2 k/k$. They required the algorithm to be *balanced*: on any input, each of the algorithm's output bits must be unbiased over the algorithm's internal randomness. To show that a successful algorithm implies the existence of the forbidden overlap structure, they construct an interpolation over several runs of the algorithm on the same input, with partially resampled internal randomness. Their proof requires balance to ensure that two fully independent runs give outputs that are far apart, so that by continuity they can extract the forbidden structure from this interpolation. Due to this requirement, their result required the symmetry provided by the NAE variant of random k-SAT. We improve on this result in three ways:

(1) We improve the threshold clause density by a logarithmic factor, to $(1+o_k(1))\kappa^* 2^k \log k/k$.

(2) We generalize the algorithm class from balanced sequential local algorithms to local and low degree algorithms. Recall that both of these computation classes simulate sequential local algorithms, even without the balance requirement.

(3) We show hardness for random k-SAT instead of NAE-k-SAT. A simple adaptation of our argument shows hardness of random NAE-k-SAT at clause density $(1+o_k(1))\kappa^* 2^{k-1}\log k/k$.

Improvements due to ensemble OGP. We consider an ensemble OGP, where the random variable resampled in the interpolation is the k-SAT instance instead of the algorithm's internal randomness. The ensemble interpolation allows us to show hardness for local and low degree algorithms. It also obviates the requirement of balance, so we no longer require the additional symmetry provided by NAE-k-SAT. This achieves improvements (2) and (3).

A tighter free entropy analysis. Crucially, we conduct a tighter free entropy analysis to achieve improvement (1). In contrast to previous work, our forbidden structure considers all 2^k ways $k+1$ satisfying assignments $y^{(0)},\dots,y^{(k)}$ can agree or disagree. We formalize such an agreement pattern as an *overlap profile* π. We will introduce this formally in Subsection 2.3.3. We will see in Lemma 2.4.1 that the analogue of the free entropy (2.2.1) for random k-SAT at clause density α is

$$\log 2 + \max_{\pi\in P}\left[H(\pi) - \frac{\alpha}{2^k}\mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)}\left|\left\{y^{(\ell)}[I] : 0\le \ell\le k\right\}\right|\right]. \tag{2.2.2}$$

Here $y^{(0)},\dots,y^{(k)}$ have overlap profile π, P is a collection of overlap constraints, and $\pi\in P$ denotes the set of overlap profiles π consistent with P. Moreover, $y^{(\ell)}[I]$ is the bit string obtained by indexing $y^{(\ell)}$ in positions I, namely $(y^{(\ell)}_{I_1},\dots,y^{(\ell)}_{I_k})$. The positive term $H(\pi)$ is the *overlap entropy* of π, which arises because $\log 2 + H(\pi)$ is the exponential rate of the number of assignment sequences $y^{(0)},\dots,y^{(k)}$ with overlap profile π. The negative term captures the log likelihood that a random formula is satisfied by all of $y^{(0)},\dots,y^{(k)}$. We think of these two terms as entropy and energy terms, respectively. We will choose P such that the magnitude of the energy term exceeds the entropy term by more than $\log 2$, which causes (2.2.2) to be negative. This implies the absence (except with exponentially small probability) of a constellation of satisfying assignments with overlap profile $\pi\in P$.

Similarly to [Wei20], we chain together many small negative free entropies to make (2.2.2) negative. Because the random k-SAT free entropy is dependent and harder to analyze, it is significantly more difficult to identify the correct high-dimensional moats. In the multi-OGP of [GS17], the condition P stipulates that the normalized Hamming distances $\Delta(y^{(i)},y^{(j)})$ of k satisfying assignments are pairwise approximately $\frac{\log k}{k}$. Using this, the energy term in (2.2.2) can be lower bounded by an inclusion-exclusion truncated at level 2. The inclusion-exclusion truncation is not sharp, and consequently this analysis requires the larger clause density $\alpha = (1+o_k(1))2^k\log^2 k/k$ (for NAE-k-SAT, $(1+o_k(1))2^{k-1}\log^2 k/k$) to show that the contribution of each $y^{(\ell)}$ to (2.2.2) is a small negative number. The fact that this natural estimate of (2.2.2) gives a threshold too large by a $\log k$ factor highlights the difficulty of accurately controlling the k-SAT free entropy and the necessity of finding good moats.

We find the correct moats. We set $P = P_1\cap\cdots\cap P_k$, where P_ℓ governs how $y^{(\ell)}$ overlaps with its predecessors $y^{(0)},\dots,y^{(\ell-1)}$. Each P_ℓ defines a moat that a smooth evolution of $y^{(\ell)}$ starting from $y^{(\ell-1)}$ must cross. In order to obtain a fine control over the tradeoff between entropy and energy in (2.2.2), we develop a notion of *conditional overlap entropy* $H(\pi(y^{(\ell)}|y^{(0)},\dots,y^{(\ell-1)}))$, which is the contribution of $y^{(\ell)}$ to the entropy term $H(\pi)$. Informally, this is a measure of the additional diversity that $y^{(\ell)}$ adds to the assignments $y^{(0)},\dots,y^{(\ell-1)}$. For each $\ell\ge 1$, our condition P_ℓ stipulates that

$$H\left(\pi(y^{(\ell)}|y^{(0)},\dots,y^{(\ell-1)})\right)\in\left[\beta_-\frac{\log k}{k},\beta_+\frac{\log k}{k}\right].$$

This choice of forbidden structure *in terms of the conditional overlap entropy* is an important contribution of our work. The choice is motivated by the subsequent energy analysis, which shows a lower bound on the energy contribution of $y^{(\ell)}$ that counterbalances the entropy increase. We next summarize this analysis.

Energy increment bound via decoupling. We can express the energy term (without the prefactor) as

$$\mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)}\left|\left\{y^{(\ell)}[I]:0\le\ell\le k\right\}\right| = \sum_{\sigma\in\{\mathtt{T},\mathtt{F}\}^k} p(\sigma), \quad \text{where} \tag{2.2.3}$$

$$p(\sigma) = \mathop{\mathbb{P}}_{I\sim\mathrm{unif}([N]^k)}\left[\sigma = y^{(\ell)}[I] \text{ for some } 0\le\ell\le k\right].$$

For each σ, $1-p(\sigma)$ is the probability that $\sigma\neq y^{(\ell)}[I]$ for all $0\le\ell\le k$. This can be conditionally expanded as a product of k factors, where the ℓth factor is the probability that $\sigma\neq y^{(\ell)}[I]$ given the values of $y^{(0)}[I],\dots,y^{(\ell-1)}[I]$. We think of (1 minus) this factor as the contribution of $y^{(\ell)}$ to $p(\sigma)$.

We apply the following estimate to decouple these products into sums. We round any factors in the conditional expansion that are less than $1-\frac{1}{k\log k}$ up to 1. Then, we note that for $0\le\varepsilon_1,\dots,\varepsilon_k\le\frac{1}{k\log k}$,

$$1-(1-\varepsilon_1)(1-\varepsilon_2)\cdots(1-\varepsilon_k)\approx\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_k,$$

up to a $1+o_k(1)$ multiplicative factor. This decouples the contributions of the $y^{(\ell)}$ to the $p(\sigma)$. We can bound the total contribution of $y^{(\ell)}$ to the energy term (2.2.3) by summing the now-decoupled contributions over $\sigma\in\{\mathtt{T},\mathtt{F}\}^k$.

Probabilistic reinterpretation. Miraculously, this sum can be reinterpreted as the success probability of an experiment involving a sum of k i.i.d. random variables, which can be controlled by concentration inequalities. We find that if the contribution of $y^{(\ell)}$ to the entropy term $H(\pi)$ is $\beta\frac{\log k}{k}$, then its contribution to the rescaled energy term (2.2.3) is at least $1-\beta e^{-(\beta-1)}$. This motivates the choice of $\iota(\beta)$ as the (rescaled) ratio of these contributions, and κ^* as the best possible ratio. When $\alpha=\kappa 2^k\log k/k$ for $\kappa>\kappa^*$, the condition P_ℓ requires β to be in a range where the contribution of $y^{(\ell)}$ to the energy term of (2.2.2) exceeds its contribution to the entropy term by at least $\varepsilon\frac{\log k}{k}$, for constant ε depending on κ. Thus the overall contribution of $y^{(\ell)}$ to (2.2.2) is upper bounded by $-\varepsilon\frac{\log k}{k}$. Summed over the $y^{(\ell)}$, this shows that (2.2.2) is negative, establishing the ladder OGP.

This energy analysis via decoupling and probabilistic reinterpretation is original and is another key contribution of our work.

2.3 Proof of Impossibility for Low Degree Polynomials

This section and the next two sections are devoted to proving our main impossibility result, Theorem 2.1.6. Throughout, we fix $\kappa>\kappa^*$. We set $\alpha=\kappa 2^k\log k/k$ and $M=\lfloor\alpha N\rfloor$.

2.3.1 Reduction to Deterministic Low Degree Polynomial

The following lemma shows that randomness does not significantly improve the power of low degree polynomial algorithms.

Lemma 2.3.1. *Suppose there exists a random degree-D polynomial that (δ,γ,η,ν)-solves $\Phi_k(N,M)$. Then, there exists a deterministic degree-D polynomial that $(3\delta,3\gamma,\eta,\nu)$-solves $\Phi_k(N,M)$.*

Proof. Let $f:\mathbb{R}^n\times\Omega\to\mathbb{R}^N$ be a random degree-D polynomial that (δ,γ,η,ν)-solves $\Phi_k(N,M)$. Then,

$$\mathop{\mathbb{E}}_{\omega}\left[\mathop{\mathbb{P}}_{\Phi}[(\mathtt{round}\circ f)(\Phi,\omega)\text{ does not }(\eta,\nu)\text{-satisfy }\Phi]\right]\le\delta \qquad\text{and}\qquad \mathop{\mathbb{E}}_{\omega}\left[\mathop{\mathbb{E}}_{\Phi}\left[\|f(\Phi,\omega)\|_2^2\right]\right]\le\gamma N.$$

By Markov's inequality,

$$\mathop{\mathbb{P}}_{\omega}\left[\mathop{\mathbb{P}}_{\Phi}[(\mathtt{round}\circ f)(\Phi,\omega)\text{ does not }(\eta,\nu)\text{-satisfy }\Phi]\ge 3\delta\right]\le\frac{1}{3} \qquad\text{and}\qquad \mathop{\mathbb{P}}_{\omega}\left[\mathop{\mathbb{E}}_{\Phi}\left[\|f(\Phi,\omega)\|_2^2\right]\ge 3\gamma N\right]\le\frac{1}{3}.$$

So, there exists $\omega\in\Omega$ such that the deterministic polynomial $g(\Phi)=f(\Phi,\omega)$ satisfies

$$\mathop{\mathbb{P}}_{\Phi}[(\mathtt{round}\circ g)(\Phi)\ (\eta,\nu)\text{-satisfies }\Phi]\ge 1-3\delta \qquad\text{and}\qquad \mathop{\mathbb{E}}_{\Phi}\left[\|g(\Phi)\|_2^2\right]\le 3\gamma N.$$

□

By Lemma 2.3.1, it suffices to show hardness for deterministic polynomials. For the rest of this section and Section 2.5, except where stated, $f : \mathbb{R}^n \to \mathbb{R}^N$ is a deterministic degree-D polynomial.

We let $\mathcal{A}(\Phi) = \mathcal{B}((\texttt{round} \circ f)(\Phi), \Phi)$, where $\mathcal{B}(x, \Phi)$ is a deterministic, computationally unbounded subroutine outputting $y \in \{\texttt{T}, \texttt{F}\}^N$ with $\Delta(x, y) \le \eta$. (If x has more than ηN entries equal to $\texttt{err}$, $\mathcal{B}$ outputs "fail.") Informally, $\mathcal{B}$ is a computationally unbounded assistant that repairs an ηN fraction of entries of $(\texttt{round} \circ f)(\Phi)$.

Because f is deterministic, $\mathcal{A}$ is also deterministic. Note that $\texttt{round} \circ f$ outputting a (η, ν)-satisfying assignment of Φ is equivalent to $\mathcal{A}$ outputting a ν-satisfying assignment of Φ. Showing that this does not occur with the required probability will be our task from here on.

2.3.2 The Interpolation Path

We can enumerate the kM literals of a formula $\Phi \in \Omega_k(N, M)$ in lexicographic order:

$$\Phi_{1,1}, \Phi_{1,2}, \ldots, \Phi_{1,k}, \Phi_{2,1}, \ldots, \Phi_{M,k}.$$

For $j \in [kM]$, let $L(j)$ denote the pair (a, b) such that $\Phi_{a,b}$ is the jth literal in this order. That is, $L(j) = (a, b)$ is the unique pair of integers $(a, b) \in [M] \times [k]$ satisfying $k(a-1) + b = j$. We now define a sequence of correlated random k-SAT formulas.

Definition 2.3.2 (Interpolation path)**.** Let $T = k^2 M$. Let $\Phi^{(0)}, \ldots, \Phi^{(T)} \in \Omega_k(N, M)$ be the sequence of k-SAT instances sampled as follows. First, sample $\Phi^{(0)} \sim \Phi_k(N, M)$. For each $1 \le t \le T$, let $\sigma(t) \in [kM]$ be the unique integer such that $t \equiv \sigma(t) \pmod{kM}$. Then, $\Phi^{(t)}$ is obtained from $\Phi^{(t-1)}$ by resampling $\Phi^{(t)}_{L(\sigma(t))}$ from $\mathrm{unif}(\mathcal{L})$. Moreover, for $0 \le t \le T$, let $x^{(t)} = \mathcal{A}(\Phi^{(t)})$.

In other words, we start from a random k-SAT instance and resample the literals one by one in lexicographic order. After we have resampled all the literals we start over, repeating the procedure until each literal has been resampled k times. Note that each $\Phi^{(t)}$ is marginally a sample from $\Phi_k(N, M)$ and that if $|t - t'| \ge kM$, then $\Phi^{(t)} \perp\!\!\!\perp \Phi^{(t')}$. We run our assisted low degree algorithm $\mathcal{A}$ on all these k-SAT instances and collect the outputs as the sequence $x^{(0)}, x^{(1)}, \ldots, x^{(T)} \in \{\texttt{T}, \texttt{F}\}^N$.

2.3.3 Overlap Profiles

We now introduce the overlap profile of an ordered list of assignments. The overlap profile summarizes the bitwise agreement and disagreement pattern of a list of assignments.

Let $\mathcal{P}_2(\ell)$ denote the set of unordered partitions of $\{0, \ldots, \ell - 1\}$ into two (possibly empty) sets. For example,

$$\mathcal{P}_2(3) = \big\{ \{\{0,1,2\}, \emptyset\}, \{\{0,1\}, \{2\}\}, \{\{0,2\}, \{1\}\}, \{\{1,2\}, \{0\}\} \big\}.$$

Note that $|\mathcal{P}_2(\ell)| = 2^{\ell - 1}$.

Definition 2.3.3 (Overlap profile)**.** Let $y^{(0)}, \ldots, y^{(\ell-1)} \in \{\texttt{T}, \texttt{F}\}^N$ be a sequence of assignments. Their *overlap profile* $\pi = \pi(y^{(0)}, \ldots, y^{(\ell-1)})$, is a vector $\pi \in \mathbb{R}^{2^{\ell-1}}$ indexed by unordered pairs $\{S, T\} \in \mathcal{P}_2(\ell)$, where

$$\pi_{S,T} = \frac{1}{N} \Big| i \in [N] : \text{all } \{y_i^{(t)} : t \in S\} \text{ equal one value and all } \{y_i^{(t)} : t \in T\} \text{ equal the other value} \Big|.$$

Example 2.3.4. Let $\ell = 3$. The overlap profile $\pi = \pi(y^{(0)}, y^{(1)}, y^{(2)})$ consists of four entries $\pi_{012,\emptyset}$, $\pi_{01,2}$, $\pi_{02,1}$, and $\pi_{12,0}$, where

$$\pi_{012,\emptyset} = \frac{1}{N} \Big| i \in [N] : y_i^{(0)} = y_i^{(1)} = y_i^{(2)} \Big| \qquad \text{and} \qquad \pi_{01,2} = \frac{1}{N} \Big| i \in [N] : y_i^{(0)} = y_i^{(1)} \neq y_i^{(2)} \Big|,$$

and $\pi_{02,1}, \pi_{12,0}$ are analogous to $\pi_{01,2}$.

We can interpret an overlap profile as a probability distribution: $\pi_{S,T}$ is the probability that in a random position $i \sim \mathrm{unif}([N])$, all $\{y_i^{(t)} : t \in S\}$ equal one value and all $\{y_i^{(t)} : t \in T\}$ equal the other. We naturally define the *overlap entropy* of $y^{(0)}, \dots, y^{(\ell-1)}$ by

$$H\left(\pi(y^{(0)}, \dots, y^{(\ell-1)})\right) = -\sum_{\{S,T\} \in \mathcal{P}_2(\ell)} \pi_{S,T} \log \pi_{S,T}.$$

This is the entropy of the unordered pair of sets $\{S, T\}$ obtained by sampling $i \sim \mathrm{unif}([N])$ and partitioning $\{0, \dots, \ell-1\}$ based on the value of $y_i^{(t)}$.

We also define conditional overlap profiles. Let $\pi = \pi(y^{(0)}, \dots, y^{(\ell-1)})$. For each $\{S, T\} \in \mathcal{P}_2(\ell-1)$ with $\pi_{S,T} > 0$, $\pi_{\cdot|S,T}$ is a probability distribution on the two partitions $\{S \cup \{\ell-1\}, T\}$ and $\{S, T \cup \{\ell-1\}\}$ with

$$\pi_{S\cup\{\ell-1\},T|S,T} = \frac{\pi_{S\cup\{\ell-1\},T}}{\pi_{S\cup\{\ell-1\},T} + \pi_{S,T\cup\{\ell-1\}}} \qquad \text{and} \qquad \pi_{S,T\cup\{\ell-1\}|S,T} = \frac{\pi_{S,T\cup\{\ell-1\}}}{\pi_{S\cup\{\ell-1\},T} + \pi_{S,T\cup\{\ell-1\}}}.$$

(If $\pi_{S\cup\{\ell-1\},T} = \pi_{S,T\cup\{\ell-1\}} = 0$, we define this distribution arbitrarily.) This is the distribution of the agreement pattern of $y^{(0)}, \dots, y^{(\ell-1)}$ on a uniformly random position, conditioned on the agreement pattern of $y^{(0)}, \dots, y^{(\ell-2)}$ in that position being $\{S, T\}$. We denote the resulting collection of distributions, one for each $\{S, T\} \in \mathcal{P}_2(\ell-1)$, by $\pi_{\cdot|\cdot} = \pi(y^{(\ell-1)}|y^{(0)}, \dots, y^{(\ell-2)})$. We analogously define the *conditional overlap entropy*

$$H\left(\pi(y^{(\ell-1)}|y^{(0)}, \dots, y^{(\ell-2)})\right) = \sum_{\{S,T\} \in \mathcal{P}_2(\ell-1)} \pi_{S,T} H(\pi_{\cdot|S,T}).$$

Before proceeding, we collect some properties of overlap profiles which will be useful in the rest of the section. The proofs of these assertions follow readily from the above definitions.

Fact 2.3.5. *Overlap profiles have the following properties.*

(a) There are at most $N^{2^{\ell-1}}$ distinct overlap profiles of ℓ assignments $y^{(0)}, \dots, y^{(\ell-1)} \in \{\mathtt{T}, \mathtt{F}\}^N$.

(b) Overlap entropies satisfy the chain rule

$$H\left(\pi(y^{(0)}, \dots, y^{(\ell-1)})\right) = H\left(\pi(y^{(0)}, \dots, y^{(\ell-2)})\right) + H\left(\pi(y^{(\ell-1)}|y^{(0)}, \dots, y^{(\ell-2)})\right).$$

(c) Repeated assignments do not affect overlap entropies. That is, if $z^{(0)}, \dots, z^{(r-1)}$ are the distinct elements of $y^{(0)}, \dots, y^{(\ell-1)}$, then

$$H\left(\pi(y^{(0)}, \dots, y^{(\ell-1)})\right) = H\left(\pi(z^{(0)}, \dots, z^{(r-1)})\right).$$

If $z^{(0)}, \dots, z^{(r-2)}$ are the distinct elements of $y^{(0)}, \dots, y^{(\ell-2)}$, then

$$H\left(\pi(y^{(\ell-1)}|y^{(0)}, \dots, y^{(\ell-2)})\right) = H\left(\pi(y^{(\ell-1)}|z^{(0)}, \dots, z^{(r-2)})\right).$$

Furthermore, if $y^{(\ell-1)} \in \{y^{(0)}, \dots, y^{(\ell-2)}\}$, then $H\left(\pi(y^{(\ell-1)}|y^{(0)}, \dots, y^{(\ell-2)})\right) = 0$.

2.3.4 Outline of Proof of Impossibility

Recall that $\iota(\beta) = \frac{\beta}{1-\beta e^{-(\beta-1)}}$ is strictly convex with with $\iota(\beta) \to +\infty$ when $\beta \to 1^+$ or $\beta \to +\infty$, and has minimum κ^* attained at β^*. Because $\kappa > \kappa^*$, there exist two solutions $\beta_{\min}, \beta_{\max}$ to $\iota(\beta) = \kappa$, with $\beta_{\min} \in (1, \beta^*)$ and $\beta_{\max} \in (\beta^*, +\infty)$. Set $\beta_- = \frac{\beta_{\min}+\beta^*}{2}$ and $\beta_+ = \frac{\beta_{\max}+\beta^*}{2}$. (This choice is arbitrary; any deterministic $\beta_{\min} < \beta_- < \beta_+ < \beta_{\max}$ will do.) Set $\varepsilon > 0$ such that $\frac{\beta+\varepsilon}{1-\beta e^{-(\beta-1)}} \le \kappa$ for all $\beta \in [\beta_-, \beta_+]$. We emphasize that $\beta_-, \beta_+, \varepsilon$ depend on κ only.

For the rest of this proof, take $\eta = \frac{\beta_+ - \beta_-}{8k}$ and $\nu = \frac{1}{k^2 2^k}$. We next define the events $S_{\text{valid}}, S_{\text{consec}}, S_{\text{indep}}, S_{\text{ogp}}$, which are measurable in the interpolation path $\Phi^{(0)}, \dots, \Phi^{(T)}$ defined in Definition 2.3.2. Define

$$S_{\text{valid}} = \left\{x^{(t)} \text{ } \nu\text{-satisfies } \Phi^{(t)} \text{ for all } 0 \le t \le T\right\}.$$

This is the event that $\mathcal{A}$ succeeds on all $\Phi^{(t)}$. Define

$$S_{\text{consec}} = \left\{ \Delta(x^{(t)}, x^{(t-1)}) \le \frac{\beta_+ - \beta_-}{2k} \text{ for all } 1 \le t \le T \right\}.$$

This is the event that outputs of $\mathcal{A}$ on consecutive $\Phi^{(t)}$ are close in Hamming distance. Define S_{indep} as the event that there do not exist indices $0 \le t_0 \le t_1 \le \cdots \le t_k \le T$ with $t_k \ge t_{k-1} + kM$ and an assignment $y \in \{\mathtt{T}, \mathtt{F}\}^N$ such that

(IND-A) y ν-satisfies $\Phi^{(t_k)}$;

(IND-B) $H\left(\pi(y|x^{(t_0)}, \dots, x^{(t_{k-1})})\right) \le \beta_+ \frac{\log k}{k}$.

This is the event that if t_k is large enough that $\Phi^{(t_k)}$ is independent of $\Phi^{(t_0)}, \dots, \Phi^{(t_{k-1})}$, then all ν-satisfying assignments to $\Phi^{(t_k)}$ have high conditional overlap entropy relative to the outputs of $\mathcal{A}$ on $\Phi^{(t_0)}, \dots, \Phi^{(t_{k-1})}$. Finally, define S_{ogp} as the event that there do not exist indices $0 \le t_0 \le t_1 \le \cdots \le t_k \le T$ and assignments $y^{(0)}, \dots, y^{(k)} \in \{\mathtt{T}, \mathtt{F}\}^N$ such that

(OGP-A) For all $0 \le \ell \le k$, $y^{(\ell)}$ ν-satisfies $\Phi^{(t_\ell)}$;

(OGP-B) For all $1 \le \ell \le k$, $H\left(\pi(y^{(\ell)}|y^{(0)}, \dots, y^{(\ell-1)})\right) \in \left[\beta_- \frac{\log k}{k}, \beta_+ \frac{\log k}{k}\right]$.

S_{ogp} defines the main forbidden structure of our argument. Informally, this forbidden structure consists of $k+1$ assignments, each ν-satisfying possibly different $\Phi^{(t)}$ in the interpolation, such that each assignment has medium conditional overlap entropy relative to its predecessors.

The key ingredients in our proof of Theorem 2.1.6 are the following two propositions. Proposition 2.3.6 shows that these four events do not simultaneously occur, and Proposition 2.3.7 controls their probabilities. These two propositions derive the main contradiction: if a low degree algorithm $(\delta, \gamma, \eta, \nu)$-solves $\Phi_k(N, M)$ for the requisite $(\delta, \gamma, \eta, \nu)$, then Proposition 2.3.6 implies $S_{\text{valid}} \cap S_{\text{consec}} \cap S_{\text{indep}} \cap S_{\text{ogp}} = \emptyset$, while Proposition 2.3.7 and a union bound imply $S_{\text{valid}} \cap S_{\text{consec}} \cap S_{\text{indep}} \cap S_{\text{ogp}} \neq \emptyset$.

Proposition 2.3.6. *For all sufficiently large k, $S_{\text{valid}} \cap S_{\text{consec}} \cap S_{\text{indep}} \cap S_{\text{ogp}} = \emptyset$.*

Proposition 2.3.7. *Suppose f is a deterministic degree-D polynomial that $(\delta, \gamma, \eta, \nu)$-solves $\Phi_k(N, M)$. For all sufficiently large k, the following inequalities hold.*

(a) $\mathbb{P}(S_{\text{valid}} \cap S_{\text{consec}}) \ge (2N)^{-4\gamma D k^2/(\beta_+ - \beta_-)} - (T+1)\delta$.

(b) $\mathbb{P}(S_{\text{indep}}^c) \le \exp(-\Omega(N))$.

(c) $\mathbb{P}(S_{\text{ogp}}^c) \le \exp(-\Omega(N))$.

The remainder of this section and Sections 2.4 and 2.5 will be devoted to proving these propositions. We will prove Proposition 2.3.6 in Subsection 2.3.5 and Proposition 2.3.7(b) in Subsection 2.3.6. We will prove Proposition 2.3.7(c), which establishes the main multi-OGP, in Section 2.4. Finally, we will prove Proposition 2.3.7(a) in Section 2.5. Let us first see how these results imply Theorem 2.1.6.

Proof of Theorem 2.1.6. Assume for sake of contradiction that there exists a (random) degree-D polynomial $f : \mathbb{R}^n \times \Omega \to \mathbb{R}^N$ that $(\delta, \gamma, \eta, \nu)$-solves $\Phi_k(N, M)$. By Lemma 2.3.1, there exists a deterministic degree-D polynomial $g : \mathbb{R}^n \to \mathbb{R}^N$ that $(3\delta, 3\gamma, \eta, \nu)$-solves $\Phi_k(N, M)$. We set $k^* = k^*(\kappa)$ large enough that Propositions 2.3.6 and 2.3.7 both hold. By Proposition 2.3.7 and a union bound,

$$\mathbb{P}(S_{\text{valid}} \cap S_{\text{consec}} \cap S_{\text{indep}} \cap S_{\text{ogp}}) \ge (2N)^{-12\gamma D k^2/(\beta_+ - \beta_-)} - 3(T+1)\delta - \exp(-\Omega(N)).$$

We will show this probability is positive for suitable C_1, C_2. Let $C_2 = 2 + \frac{12k^2}{\beta_+ - \beta_-}$. Recall that $T = k^2 M = k^2 \lfloor \alpha N \rfloor$. If $\delta \le \exp(-C_2 \gamma D \log N)$, then $3(T+1)\delta \le \frac{1}{3}(2N)^{-12\gamma D k^2/(\beta_+ - \beta_-)}$ for sufficiently large N. Note that if $D \le \frac{C_1 N}{\gamma \log N}$, then

$$(2N)^{-12\gamma D k^2/(\beta_+ - \beta_-)} \ge N^{-24\gamma D k^2/(\beta_+ - \beta_-)} \ge \exp\left(-\frac{24 C_1 k^2}{\beta_+ - \beta_-} N\right).$$

Let C_1 be small enough that the right-hand side is asymptotically larger than the $\exp(-\Omega(N))$ term. Thus for sufficiently large N, the $\exp(-\Omega(N))$ term is at most $\frac{1}{3}(2N)^{-12\gamma Dk^2/(\beta_+-\beta_-)}$. Therefore, there exists N^* such that if $N \geq N^*$, then $\mathbb{P}(S_{\text{valid}} \cap S_{\text{consec}} \cap S_{\text{indep}} \cap S_{\text{ogp}}) > 0$. This implies that $S_{\text{valid}} \cap S_{\text{consec}} \cap S_{\text{indep}} \cap S_{\text{ogp}} \neq \emptyset$, contradicting Proposition 2.3.6. □

2.3.5 Constructing the Forbidden Structure from Algorithm Outputs

In this subsection, we will prove Proposition 2.3.6. We will show that if S_{valid}, S_{consec}, and S_{indep} all hold, then we can construct an example of the structure forbidden by S_{ogp}, and therefore all four events cannot hold simultaneously.

We will need the following auxiliary lemma, which shows that a small change of $x \in \{\mathtt{T}, \mathtt{F}\}^N$ in Hamming distance induces only a small change in $H(\pi(x|y^{(0)}, \dots, y^{(\ell-1)}))$. This lemma allows us to convert S_{consec} to a guarantee that consecutive conditional overlap entropies are small. We defer the proof to Subsection 2.3.7.

Lemma 2.3.8. *Let $\ell \in \mathbb{N}$ be arbitrary and let $x, x', y^{(0)}, \dots, y^{(\ell-1)} \in \{\mathtt{T}, \mathtt{F}\}^N$. If $\Delta(x, x') \leq \frac{1}{2}$, then*

$$\left| H\left(\pi(x|y^{(0)}, \dots, y^{(\ell-1)})\right) - H\left(\pi(x'|y^{(0)}, \dots, y^{(\ell-1)})\right) \right| \leq H\left(\Delta(x, x')\right).$$

The $H(\cdot)$ on the right denotes the binary entropy function.

Proof of Proposition 2.3.6. Set k large enough that $\frac{\beta_+-\beta_-}{2k} \leq \frac{1}{2}$ and $H\left(\frac{\beta_+-\beta_-}{2k}\right) \leq (\beta_+ - \beta_-)\frac{\log k}{k}$. The second inequality holds for all sufficiently large k due to the inequality $H(x) \leq x \log \frac{e}{x}$.

Suppose that S_{valid}, S_{consec}, and S_{indep} all hold. For $0 \leq \ell \leq k$, let $y^{(\ell)} = x^{(t_\ell)}$, where $0 \leq t_0 \leq t_1 \leq \cdots \leq t_k \leq T$ are defined as follows. Let $t_0 = 0$. For $1 \leq \ell \leq k$, let t_ℓ be the smallest $t > t_{\ell-1}$ such that $H(x^{(t)}|y^{(0)}, \dots, y^{(\ell-1)}) \in \left[\beta_- \frac{\log k}{k}, \beta_+ \frac{\log k}{k}\right]$. We will show that such t_ℓ exists and satisfies $t_\ell \leq t_{\ell-1} + kM$.

Let $t' = t_{\ell-1} + kM$, and let $I = \{t_{\ell-1}, t_{\ell-1}+1, \dots, t'\}$. For $t \in I$, let $h(t) = H\left(\pi(x^{(t)}|y^{(0)}, \dots, y^{(\ell-1)})\right)$; we will analyze the evolution of $h(t)$ as we increment $t \in I$. By Fact 2.3.5(c), $h(t_{\ell-1}) = 0$.

In the definition of S_{indep}, set $t_k = t'$ and $t_\ell, t_{\ell+1}, \dots, t_{k-1}$ equal to $t_{\ell-1}$. By Fact 2.3.5(c) (which allows us to ignore the duplicated $t_\ell, \dots, t_{k-1}$), $\Phi^{(t')}$ has no ν-satisfying assignment y with $H\left(\pi(y|y^{(0)}, \dots, y^{(\ell-1)})\right) \leq \beta_+ \frac{\log k}{k}$. But because S_{valid} holds, $x^{(t')}$ ν-satisfies $\Phi^{(t')}$. It follows that $h(t') > \beta_+ \frac{\log k}{k}$.

Because S_{consec} holds, we have $\Delta(x^{(t)}, x^{(t-1)}) \leq \frac{\beta_+-\beta_-}{2k}$ for all t. By Lemma 2.3.8,

$$|h(t) - h(t-1)| \leq H\left(\Delta(x^{(t)}, x^{(t-1)})\right) \leq H\left(\frac{\beta_+ - \beta_-}{2k}\right) \leq (\beta_+ - \beta_-)\frac{\log k}{k}. \tag{2.3.1}$$

Since $h(t_{\ell-1}) = 0$ and $h(t') > \beta_+ \frac{\log k}{k}$, (2.3.1) implies the existence of $t \in I$ such that $h(t) \in \left[\beta_- \frac{\log k}{k}, \beta_+ \frac{\log k}{k}\right]$. So, t_ℓ is well defined and satisfies $t_\ell \leq t_{\ell-1} + kM$.

Because the interpolation path has length $T = k^2 M$, and $t_\ell \leq t_{\ell-1} + kM$ for all $1 \leq \ell \leq k$, this procedure sets all of $t_1, \dots, t_k$ before the end of the interpolation. Finally, because S_{valid} holds, y_ℓ ν-satisfies $\Phi^{(t_\ell)}$ for all $0 \leq \ell \leq k$. We have thus constructed the structure forbidden by S_{ogp}. □

2.3.6 Solutions to Independent Instances Contribute Large Overlap Entropy

In this subsection, we will prove Proposition 2.3.7(b). The proof is by a first moment argument.

Proof of Proposition 2.3.7(b). By Markov's inequality, $\mathbb{P}(S_{\text{indep}}^c)$ is upper bounded by the expected number of $(t_0, \dots, t_k, y)$ satisfying $0 \leq t_0 \leq \cdots \leq t_k \leq T$, $t_k \geq t_{k-1} + kM$, and conditions (IND-A) and (IND-B). There are at most $(T+1)^{k+1}$ possible choices of $(t_0, \dots, t_k)$. Because $t_k \geq t_{k-1} + kM$, $\Phi^{(t_k)}$ is independent of $x^{(t_0)}, \dots, x^{(t_{k-1})}$.

Let $P = P(x^{(t_0)}, \dots, x^{(t_{k-1})})$ denote the set of all overlap profiles $\pi = \pi(x^{(t_0)}, \dots, x^{(t_{k-1})}, y)$ over $y \in \{\mathtt{T}, \mathtt{F}\}^N$ with $H(\pi(y|x^{(t_0)}, \dots, x^{(t_{k-1})})) \le \beta_+ \frac{\log k}{k}$. By Fact 2.3.5(a), $|P| \le N^{2^k}$. Thus,

$$\mathbb{P}(S^c_{\text{indep}}) \le (T+1)^{k+1} N^{2^k} \max_{\substack{0 \le t_0 \le \cdots \le t_k \le T \\ t_k \ge t_{k-1} + kM}} \max_{\pi \in P} \mathop{\mathbb{E}}_{\Phi^{(t_k)}} \#\left(y \in \{\mathtt{T}, \mathtt{F}\}^N : y \text{ } \nu\text{-satisfies } \Phi^{(t_k)} \text{ and } \pi(x^{(t_0)}, \dots, x^{(t_{k-1})}, y) = \pi\right)$$

We can evaluate this inner expectation by linearity of expectation. The number of y satisfying that $\pi(x^{(t_0)}, \dots, x^{(t_{k-1})}, y) = \pi$ is

$$\begin{aligned}
\prod_{\{S,T\} \in \mathcal{P}_2(k)} \binom{\pi_{S,T} N}{\pi_{S \cup \{k\},T} N} &= \exp\left(N \sum_{\{S,T\} \in \mathcal{P}_2(k)} \pi_{S,T} H\left(\frac{\pi_{S \cup \{k\},T}}{\pi_{S,T}}\right) + o(N)\right) \\
&= \exp\left(N H\left(\pi(y|x^{(t_0)}, \dots, x^{(t_{k-1})})\right) + o(N)\right) \\
&\le \exp\left(N \beta_+ \frac{\log k}{k} + o(N)\right) . M
\end{aligned}$$

Recall that $\Phi^{(t_k)}$ is independent of $x^{(t_0)}, \dots, x^{(t_{k-1})}$. Because $\sum_{j=0}^{\nu M} \binom{M}{j} \le (M+1)\binom{M}{\nu M}$, the probability that any one of these y ν-satisfies $\Phi^{(t_k)}$ is at most

$$\begin{aligned}
\sum_{S \subseteq [M], |S| \le \nu M} (1-2^{-k})^{M-|S|} &\le (M+1)\binom{M}{\nu M}(1-2^{-k})^{(1-\nu)M} \\
&\le \exp\left(\nu M \log \frac{e}{\nu} - (1-\nu) 2^{-k} M + o(N)\right) \\
&= \exp\left(N\left(-\kappa \frac{\log k}{k} + o_k(1)\right) + o(N)\right) .
\end{aligned}$$

Here we used that $\binom{a}{b} \le \left(\frac{ea}{b}\right)^b$. Thus,

$$\mathbb{P}(S^c_{\text{indep}}) \le \exp\left(-N(\kappa - \beta_+ - o_k(1)) \frac{\log k}{k} + o(N)\right),$$

where the $(T+1)^{k+1} N^{2^k}$ is absorbed in the $o(N)$. Finally, as

$$\beta_+ + \varepsilon \le \frac{\beta_+ + \varepsilon}{1 - \beta_+ e^{-(\beta_+ - 1)}} \le \kappa,$$

we have $\kappa - \beta_+ \ge \varepsilon$. Thus $\mathbb{P}(S^c_{\text{indep}}) = \exp(-\Omega(N))$ for sufficiently large k. □

2.3.7 Small Hamming Distance Implies Small Conditional Overlap Entropy Difference

We now present the deferred proof of Lemma 2.3.8, which shows that a small change in $x \in \{\mathtt{T}, \mathtt{F}\}^N$ causes only a small change in $H(\pi(x|y^{(0)}, \dots, y^{(\ell-1)}))$.

Proof of Lemma 2.3.8. For each partition $\{S, T\} \in \mathcal{P}_2(\ell)$, let

$$I_{S,T} = \left\{ i \in [N] : \text{all } \{y_i^{(t)} : t \in S\} \text{ equal one value and all } \{y_i^{(t)} : t \in T\} \text{ equal the other value} \right\}.$$

Note that $|I_{S,T}| = \pi_{S,T} N$. If $\pi_{S,T} \ne 0$, define

$$\lambda_{S,T} = \frac{1}{|I_{S,T}|} \#\left(i \in I_{S,T} : x_i = \mathtt{T}\right) \qquad \text{and} \qquad \lambda'_{S,T} = \frac{1}{|I_{S,T}|} \#\left(i \in I_{S,T} : x'_i = \mathtt{T}\right).$$

(If $\pi_{S,T} = 0$, we can set these values arbitrarily in $[0,1]$.) On each index set $I_{S,T}$, x and x' differ in at least

$$|I_{S,T}| \cdot |\lambda_{S,T} - \lambda'_{S,T}| = \pi_{S,T}|\lambda_{S,T} - \lambda'_{S,T}|n$$

positions. Thus,

$$\frac{1}{2} \geq \Delta(x,x') \geq \sum_{\{S,T\}\in\mathcal{P}_2(\ell)} \pi_{S,T}|\lambda_{S,T} - \lambda'_{S,T}|.$$

Let $\sum_{\{S,T\}\in\mathcal{P}_2(\ell)} \pi_{S,T}|\lambda_{S,T} - \lambda'_{S,T}| = \mu$. Moreover, from the definition of conditional overlap entropy,

$$H\left(\pi(x|y^{(0)},\dots,y^{(\ell-1)})\right) = \sum_{\{S,T\}\in\mathcal{P}_2(\ell)} \pi_{S,T} H(\lambda_{S,T}),$$

and analogously for x'. Note that $H(\cdot)$ is concave, so $H'(\cdot)$ is decreasing. Thus, for all $[a,b] \in [0,1]$ with $a \geq b$,

$$H(a) - H(b) = \int_a^b H'(x)\ \mathrm{d}x \leq \int_0^{a-b} H'(x)\ \mathrm{d}x = H(a-b).$$

Similarly $H(1-b) - H(1-a) \leq H(a-b)$, whence $|H(a) - H(b)| \leq H(a-b)$. Thus,

$$\begin{aligned}\left|H\left(\pi(x|y^{(0)},\dots,y^{(\ell-1)})\right) - H\left(\pi(x'|y^{(0)},\dots,y^{(\ell-1)})\right)\right| &\leq \sum_{\{S,T\}\in\mathcal{P}_2(\ell)} \pi_{S,T}\left|H(\lambda_{S,T}) - H(\lambda'_{S,T})\right| \\ &\leq \sum_{\{S,T\}\in\mathcal{P}_2(\ell)} \pi_{S,T} H\left(|\lambda_{S,T} - \lambda'_{S,T}|\right).\end{aligned}$$

By concavity of $H(\cdot)$, this last quantity has maximum value $H(\mu)$, attained when all the $|\lambda_{S,T} - \lambda'_{S,T}|$ are equal to μ. Because $H(\cdot)$ is increasing on $[0, \frac{1}{2}]$ and $\mu \leq \Delta(x,x') \leq \frac{1}{2}$, we conclude that

$$\left|H\left(\pi(x|y^{(0)},\dots,y^{(\ell-1)})\right) - H\left(\pi(x'|y^{(0)},\dots,y^{(\ell-1)})\right)\right| \leq H(\mu) \leq H\left(\Delta(x,x')\right).$$

□

2.4 The Multi-OGP

In this section, we will prove Proposition 2.3.7(c), which shows that the forbidden structure in S_{ogp} does not occur with high probability.

2.4.1 Proof Outline

We first give a high level overview of the proof, which is by another first moment computation. Throughout this section, for $I \in [N]^k$ and $x \in \{\mathtt{T},\mathtt{F}\}^N$, let $x[I] = (x_{I_1},\dots,x_{I_k})$ be the subsequence of x indexed by I. We begin with the following lemma, which bounds the exponential rate of $\mathbb{P}(S^c_{\text{ogp}})$ in terms of a maximum over overlap profiles. We will prove this lemma in Subsection 2.4.2.

Lemma 2.4.1. *Let P denote the set of overlap profiles $\pi = \pi(y^{(0)},\dots,y^{(k)})$ over $y^{(0)},\dots,y^{(k)} \in \{\mathtt{T},\mathtt{F}\}^N$ satisfying that for all $1 \leq \ell \leq k$, $H\left(\pi(y^{(\ell)}|y^{(0)},\dots,y^{(\ell-1)})\right) \in \left[\beta_- \frac{\log k}{k}, \beta_+ \frac{\log k}{k}\right]$. Then,*

$$\frac{1}{N}\log \mathbb{P}(S^c_{\text{ogp}}) \leq \log 2 + \max_{\pi\in P}\left[H(\pi) - \kappa\frac{\log k}{k} \mathop{\mathbb{E}}_{I\sim\text{unif}([N]^k)}\left|\left\{y^{(\ell)}[I] : 0 \leq \ell \leq k\right\}\right|\right] + o_k(1) + o(1), \quad (2.4.1)$$

where in the expectation, $y^{(0)},\dots,y^{(k)} \in \{\mathtt{T},\mathtt{F}\}^N$ is a sequence of assignments with overlap profile π.

Note that the expectation $\mathbb{E}_{I\sim\mathrm{unif}([N]^k)}\left|\left\{y^{(\ell)}[I]:0\le\ell\le k\right\}\right|$ has the same value for any $y^{(0)},\dots,y^{(k)}\in\{\mathtt{T},\mathtt{F}\}^N$ with overlap profile π. So, the quantity inside the maximum is a function of π.

The negative term in (2.4.1) arises as an upper bound on the exponential rate of the probability that $y^{(0)},\dots,y^{(k)}$ all respectively ν-satisfy $\Phi^{(t_0)},\dots,\Phi^{(t_k)}$, for fixed $y^{(0)},\dots,y^{(k)}$ and $t_0,\dots,t_k$. Let us first argue heuristically that this bounds the exponential rate; we will formalize this reasoning in Lemma 2.4.3 below. We expect this probability to be maximized when $t_0=\cdots=t_k$, because making the t_i different only introduces additional randomness (see Remark 2.4.4). So, let $\Phi^{(t_0)},\dots,\Phi^{(t_k)}$ all equal the same k-SAT instance $\Phi\sim\Phi_k(N,M)$. We also focus on the probability that $y^{(0)},\dots,y^{(k)}$ all *satisfy* Φ; we will see that the relaxation to ν-satisfy only changes the exponential rate by $o_k(1)$. The probability that $y^{(0)},\dots,y^{(k)}$ all satisfy the first clause Φ_1 is $1-2^{-k}\,\mathbb{E}_{I\sim\mathrm{unif}([N]^k)}\left|\left\{y^{(\ell)}[I]:0\le\ell\le k\right\}\right|$, because if Φ_1 contains the variables $x_{I_1},\dots,x_{I_k}$, there are $\left|\left\{y^{(\ell)}[I]:0\le\ell\le k\right\}\right|$ ways to set these variables' polarities in Φ_1 so that one of $y^{(0)},\dots,y^{(k)}$ does not satisfy Φ_1. Then, the probability that $y^{(0)},\dots,y^{(k)}$ all satisfy Φ is upper bounded by

$$\left(1-2^{-k}\mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)}\left|\left\{y^{(\ell)}[I]:0\le\ell\le k\right\}\right|\right)^M\le\exp\left(-\frac{M}{2^k}\mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)}\left|\left\{y^{(\ell)}[I]:0\le\ell\le k\right\}\right|\right),$$

and $\frac{M}{2^k}\approx N\kappa\frac{\log k}{k}$. The second ingredient in the proof of Proposition 2.3.7(c) is the following proposition, which lower bounds the expectation in the negative term of (2.4.1). We will prove this proposition in Subsection 2.4.3. Proving the bound in this proposition is one of the main technical challenges of this work, which we overcome via a surprising probabilistic reformulation of the left-hand expectation.

Proposition 2.4.2. *Let $\beta_1,\dots,\beta_k\in[\beta_-,\beta_+]$, and let $y^{(0)},\dots,y^{(k)}\in\{\mathtt{T},\mathtt{F}\}^N$ be assignments satisfying that $H\left(\pi(y^{(\ell)}|y^{(0)},\dots,y^{(\ell-1)})\right)=\beta_\ell\frac{\log k}{k}$. Then,*

$$\mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)}\left|\left\{y^{(\ell)}[I]:0\le\ell\le k\right\}\right|\ge(1-o_k(1))\sum_{\ell=1}^k\left(1-\beta_\ell e^{-(\beta_\ell-1)}\right).$$

From Lemma 2.4.1 and Proposition 2.4.2, we can see the main ideas of the proof of Proposition 2.3.7(c) and understand the motivation of the definition of S_{ogp}. The ideas are as follows.

We will prove Proposition 2.3.7(c) by showing that the right-hand side of (2.4.1) is negative. For each $\pi\in P$, this quantity can be regarded as a free entropy, with entropy term $\log 2+H(\pi)$ and energy term $\kappa\frac{\log k}{k}\,\mathbb{E}_{I\sim\mathrm{unif}([N]^k)}\left|\left\{y^{(\ell)}[I]:0\le\ell\le k\right\}\right|$. This free entropy exhibits a tradeoff where as the entropy term increases, the assignments $y^{(0)},\dots,y^{(k)}$ become more diverse, and so the energy term increases too. The event S_{ogp} is selected so that for overlap profiles $\pi\in P$, where P is defined in Lemma 2.4.1, the energy term is larger than the entropy term, which makes the free entropy negative. In particular, (due to Fact 2.3.5(b)) we think of $H\left(\pi(y^{(\ell)}|y^{(0)},\dots,y^{(\ell-1)})\right)$ as the amount that $y^{(\ell)}$ contributes to the entropy term. Given this contribution, Proposition 2.4.2 lower bounds the amount that $y^{(\ell)}$ contributes to the energy term. In the definition of S_{ogp}, we require the entropy contribution to be in a medium range $\left[\beta_-\frac{\log k}{k},\beta_+\frac{\log k}{k}\right]$ because (recalling the definition of β_-,β_+) in this range the energy-to-entropy ratio is favorable to the energy term. Specifically, we show that if $y^{(\ell)}$ contributes an entropy in this range, the energy it contributes is at least $\varepsilon\frac{\log k}{k}$ more. Thus each $y^{(\ell)}$ decreases the free entropy by at least $\varepsilon\frac{\log k}{k}$. Together, the k assignments $y^{(1)},\dots,y^{(k)}$ contribute a free entropy decrease of $\varepsilon\log k$, which dominates the starting free entropy of $\log 2$ and makes the overall free entropy negative.

We now prove Proposition 2.3.7(c) given Lemma 2.4.1 and Proposition 2.4.2.

Proof of Proposition 2.3.7(c). Let P be as in Lemma 2.4.1. Let $\pi\in P$, and consider assignments $y^{(0)},\dots,y^{(k)}\in\{\mathtt{T},\mathtt{F}\}^N$ with $\pi(y^{(0)},\dots,y^{(k)})=\pi$. For $1\le\ell\le k$, define β_ℓ by $H\left(\pi(y^{(\ell)}|y^{(0)},\dots,y^{(\ell-1)})\right)=\beta_\ell\frac{\log k}{k}$. Note that the β_ℓ are determined given π and satisfy $\beta_1,\dots,\beta_k\in[\beta_-,\beta_+]$. By Fact 2.3.5(b), $H(\pi)=\frac{\log k}{k}\sum_{\ell=1}^k\beta_\ell$. By Proposition 2.4.2,

$$\begin{aligned}-\kappa\frac{\log k}{k}\mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)}\left|\left\{y^{(\ell)}[I]:0\le\ell\le k\right\}\right|&\le-(1-o_k(1))\frac{\log k}{k}\sum_{\ell=1}^k\kappa\left(1-\beta_\ell e^{-(\beta_\ell-1)}\right)\\&\le-(1-o_k(1))\frac{\log k}{k}\sum_{\ell=1}^k(\beta_\ell+\varepsilon).\end{aligned}$$

The last inequality uses that $\frac{\beta+\varepsilon}{1-\beta e^{-(\beta-1)}} \le \kappa$ for all $\beta \in [\beta_-, \beta_+]$. Therefore,

$$\begin{aligned} H(\pi) - \kappa\frac{\log k}{k} \mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)} \left|\left\{y^{(\ell)}[I] : 0 \le \ell \le k\right\}\right| &\le \frac{\log k}{k}\sum_{\ell=1}^{k}\beta_\ell - (1-o_k(1))\frac{\log k}{k}\sum_{\ell=1}^{k}(\beta_\ell+\varepsilon) \\ &\le o_k(1)\frac{\log k}{k}\sum_{\ell=1}^{k}\beta_\ell - (1-o_k(1))\varepsilon\log k \\ &\le o_k(1)\beta_+\log k - (1-o_k(1))\varepsilon\log k \\ &= -(1-o_k(1))\varepsilon\log k. \end{aligned}$$

This bound holds for an arbitrary $\pi \in P$, and thus for the maximum over $\pi \in P$. By Lemma 2.4.1,

$$\frac{1}{N}\log\mathbb{P}(S^c_{\mathrm{ogp}}) \le \log 2 - (1-o_k(1))\varepsilon\log k + o_k(1) + o(1) < 0$$

for sufficiently large k and N. Thus $\mathbb{P}(S^c_{\mathrm{ogp}}) \le \exp(-\Omega(N))$. □

2.4.2 Bounding the Exponential Rate by a Free Entropy

In this subsection, we will prove Lemma 2.4.1. We begin with the following lemma, which bounds the probability term arising in the first moment upper bound of $\mathbb{P}(S^c_{\mathrm{ogp}})$.

Lemma 2.4.3. *Suppose $y^{(0)}, \dots, y^{(k)} \in \{\mathtt{T},\mathtt{F}\}^N$ is a sequence of assignments and $0 \le t_0 \le t_1 \le \cdots \le t_k \le T$. Then,*

$$\frac{1}{N}\log\mathbb{P}\left[y^{(\ell)} \text{ } \nu\text{-satisfies } \Phi^{(t_\ell)} \text{ for all } 0 \le \ell \le k\right] \le -\kappa\frac{\log k}{k}\mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)}\left|\left\{y^{(\ell)}[I] : 0 \le \ell \le k\right\}\right| + o_k(1) + o(1).$$

Proof. Say a clause index $i \in [M]$ is *interrupted* if for some $0 \le \ell \le k$, t_ℓ satisfies $1 \le \sigma(t_\ell) - (i-1)k \le k-1$, where $\sigma(\cdot)$ is defined in Definition 2.3.2. Informally, i is interrupted if there is some ℓ such that $\Phi^{(t_\ell)}$ is partway through resampling the ith clause. Let U denote the set of interrupted clause indices. Note that each t_ℓ interrupts at most one clause, so $|U| \le k+1$.

Say a clause index $i \in [M]$ is *bad* if some $y^{(\ell)}$ fails to satisfy $\Phi_i^{(t_\ell)}$. Let S denote the set of bad clause indices. If $y^{(\ell)}$ ν-satisfies $\Phi^{(t_\ell)}$ for all $0 \le \ell \le k$, then each $y^{(\ell)}$ fails to satisfy at most νM clauses of $\Phi^{(t_\ell)}$, so $|S| \le (k+1)\nu M$.

We will see that because so few clause indices are interrupted or bad, it does not hurt our analysis to throw them out. We have that

$$\begin{aligned} &\mathbb{P}\left[y^{(\ell)} \text{ satisfies } \Phi^{(t_\ell)} \text{ for all } 0 \le \ell \le k\right] \\ &\le \sum_{S\subseteq[M],|S|\le(k+1)\nu M} \mathbb{P}\left[y^{(\ell)} \text{ satisfies } \Phi_i^{(t_\ell)} \text{ for all } 0 \le \ell \le k,\, i \in [M]\setminus S\right] \\ &\le (M+1)\binom{M}{(k+1)\nu M}\max_{S\subseteq[M],|S|\le(k+1)\nu M}\mathbb{P}\left[y^{(\ell)} \text{ satisfies } \Phi_i^{(t_\ell)} \text{ for all } 0 \le \ell \le k,\, i \in [M]\setminus(S\cup U)\right] \\ &= (M+1)\binom{M}{(k+1)\nu M}\max_{S\subseteq[M],|S|\le(k+1)\nu M}\prod_{i\in[M]\setminus(S\cup U)}\mathbb{P}\left[y^{(\ell)} \text{ satisfies } \Phi_i^{(t_\ell)} \text{ for all } 0 \le \ell \le k\right]. \end{aligned} \tag{2.4.2}$$

The last step uses that over $i \in [M]$, the collections of clauses $\{\Phi_i^{(t)} : 0 \le t \le T\}$ are mutually independent.

We now fix a single $i \in [M]\setminus(S\cup U)$ and analyze the last probability. We exploit the following stochastic property of non-interrupted clauses: if i is not interrupted, then the clauses $\Phi_i^{(t_0)}, \Phi_i^{(t_1)}, \dots, \Phi_i^{(t_k)}$ can be partitioned into equivalence classes, such that all clauses in the same equivalence class are identical and all clauses in different equivalence classes are mutually independent. Formally, for some $1 \le r \le k+1$, there is a surjective map $\tau : \{0,\dots,k\} \to [r]$ (dependent only on the indices $t_0,\dots,t_k$ and i) such that for i.i.d. clauses $C_1,\dots,C_r \sim \Phi_k(N,1)$,

$$\left(\Phi_i^{(t_0)}, \Phi_i^{(t_1)}, \dots, \Phi_i^{(t_k)}\right) =_d \left(C_{\tau(0)}, C_{\tau(1)}, \dots, C_{\tau(k)}\right).$$

For $1 \le s \le r$, let $B_s = \tau^{-1}(s)$ be the set of $\ell \in \{0,\dots,k\}$ such that $\Phi_i^{(t_\ell)}$ corresponds to C_s. Thus $B_1,\dots,B_r$ partition $\{0,\dots,k\}$. Now,

$$\mathbb{P}\left[y^{(\ell)} \text{ satisfies } \Phi_i^{(t_\ell)} \text{ for all } 0 \le \ell \le k\right] = \prod_{s=1}^{r} \mathbb{P}\left[y^{(\ell)} \text{ satisfies } C_s \text{ for all } \ell \in B_s\right]. \tag{2.4.3}$$

Let $I \in [N]^k$ be the indices of the k variables sampled by C_s, so $I \sim \mathrm{unif}([N]^k)$. Given I, there are $\left|\left\{y^{(\ell)}[I] : \ell \in B_s\right\}\right|$ ways to assign polarities to these k variables such that for some $\ell \in B_s$, $y^{(\ell)}$ does not satisfy C_s. Thus, conditioned on I, the probability that $y^{(\ell)}$ satisfies C_s for all $\ell \in B_s$ is $1-2^{-k}\left|\left\{y^{(\ell)}[I] : \ell \in B_s\right\}\right|$. It follows that

$$\begin{aligned}\mathbb{P}\left[y^{(\ell)} \text{ satisfies } C_s \text{ for all } \ell \in B_s\right] &= 1 - 2^{-k} \mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)} \left|\left\{y^{(\ell)}[I] : \ell \in B_s\right\}\right| \\ &\le \exp\left(-2^{-k} \mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)} \left|\left\{y^{(\ell)}[I] : \ell \in B_s\right\}\right|\right).\end{aligned}$$

So, using (2.4.3) and recalling that $B_1,\dots,B_r$ partition $\{0,\dots,k\}$, we have

$$\begin{aligned}\mathbb{P}\left[y^{(\ell)} \text{ satisfies } \Phi_i^{(t_\ell)} \text{ for all } 0 \le \ell \le k\right] &\le \exp\left(-2^{-k} \mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)} \sum_{s=1}^{r} \left|\left\{y^{(\ell)}[I] : \ell \in B_s\right\}\right|\right) \\ &\le \exp\left(-2^{-k} \mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)} \left|\left\{y^{(\ell)}[I] : 0 \le \ell \le k\right\}\right|\right).\end{aligned}$$

Next, we substitute into (2.4.2). Since $|S| \le (k+1)\nu M$, $|U| \le k+1$, and $M = \lfloor \alpha N \rfloor \ge \alpha N - 1$,

$$\begin{aligned}|m \setminus (S \cup U)| &\ge (1-(k+1)\nu)M - (k+1) \\ &\ge (1-(k+1)\nu)\alpha N - k - 2.\end{aligned}$$

So,

$$\begin{aligned}&\mathbb{P}\left[y^{(\ell)} \text{ satisfies } \Phi^{(t_\ell)} \text{ for all } 0 \le \ell \le k\right] \\ &\le (M+1)\binom{M}{(k+1)\nu M} \exp\left(-\frac{(1-(k+1)\nu)\alpha N - k - 2}{2^k} \mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)} \left|\left\{y^{(\ell)}[I] : 0 \le \ell \le k\right\}\right| + o(N)\right).\end{aligned}$$

Thus, using that $\binom{a}{b} \le \left(\frac{ea}{b}\right)^b$,

$$\begin{aligned}&\frac{1}{N}\log \mathbb{P}\left[y^{(\ell)} \text{ satisfies } \Phi^{(t_\ell)} \text{ for all } 0 \le \ell \le k\right] \\ &\le (k+1)\nu\alpha \log\frac{e}{(k+1)\nu} - \frac{(1-(k+1)\nu)\alpha}{2^k} \mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)} \left|\left\{y^{(\ell)}[I] : 0 \le \ell \le k\right\}\right| + o(1) \\ &\le -\frac{\alpha}{2^k} \mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)} \left|\left\{y^{(\ell)}[I] : 0 \le \ell \le k\right\}\right| + o_k(1) + o(1).\end{aligned}$$

The result follows from $\alpha = \kappa 2^k \log k/k$. □

Proof of Lemma 2.4.1. By Markov's inequality, $\mathbb{P}(S_{\mathrm{ogp}}^c)$ is upper bounded by the expected number of $0 \le t_0 \le t_1 \le \cdots \le t_k \le T$ and $(y^{(0)},\dots,y^{(k)})$ satisfying conditions (OGP-A) and (OGP-B). There are at most $(T+1)^{k+1}$ choices of $(t_0,\dots,t_k)$, and (by Fact 2.3.5(a)) $|P| \le N^{2^k}$. By linearity of expectation,

$$\mathbb{P}(S_{\mathrm{ogp}}^c) \le (T+1)^{k+1} N^{2^k} \max_{\substack{0\le t_0\le\cdots\le t_k\le T \\ \pi\in\overline{P}}} \mathbb{E}\left[\#\left(\begin{array}{l}(y^{(0)},\dots,y^{(k)}) \in \{\mathtt{T},\mathtt{F}\}^{N\times(k+1)} : \\ y^{(\ell)} \text{ } \nu\text{-satisfies } \Phi^{(t_\ell)} \text{ for all } 0 \le \ell \le k \\ \text{and } \pi(y^{(0)},\dots,y^{(k)}) = \pi\end{array}\right)\right].$$

Let πN be the scalar product of π, treated as a vector, by N. There are $2^N\binom{N}{\pi N}$ sequences of assignments $(y^{(0)},\dots,y^{(k)})$ with $\pi(y^{(0)},\dots,y^{(k)}) = \pi$: 2^N ways to choose $y^{(0)}$, and then $\binom{N}{\pi N}$ ways to assign the positions

$[N]$ to the partitions of $\{0,\dots,k\}$. Over all of these sequences of assignments, the probability of the event that $y^{(\ell)}$ satisfies $\Phi^{(t_\ell)}$ for all $0 \le \ell \le k$ is uniformly upper bounded by Lemma 2.4.3. By linearity of expectation, the last expectation is upper bounded by

$$2^N \binom{N}{\pi N} \exp\left(-N\left(\kappa \frac{\log k}{k} \mathop{\mathbb{E}}_{I\sim \mathrm{unif}([N]^k)} \left|\left\{y^{(\ell)}[I] : 0 \le \ell \le k\right\}\right| + o_k(1) + o(1)\right)\right).$$

Because $\binom{N}{\pi N} = \exp\left(N(H(\pi) + o(1))\right)$, the result follows. □

Remark 2.4.4. The step in the proof of Lemma 2.4.3 where we lower bound $\sum_{s=1}^{r} \left|\left\{y^{(\ell)}[I] : \ell \in B_s\right\}\right|$ by $\left|\left\{y^{(\ell)}[I] : 0 \le \ell \le k\right\}\right|$ is tight when $t_0,\dots,t_k$ are all equal, because in this case $r = 1$ and $B_1 = \{0,1,\dots,k\}$. Thus the exponential rate of $\mathbb{P}(S_{\mathrm{ogp}}^c)$ is dominated by the case when the t_i are equal. In other words, $\mathbb{P}(S_{\mathrm{ogp}}^c)$ has the same exponential rate as if, in the definition of S_{ogp}, we required all the $y^{(\ell)}$ to ν-satisfy the same $\Phi^{(t)}$. This shows the power of the "ensemble" part of the ensemble multi-OGP: for no cost in the exponential rate, we can generalize the forbidden structure to an ensemble. All ensemble (multi-)OGPs in the literature share and leverage this property, see [GJW20, Wei20].

2.4.3 Lower Bounding the Energy Term

In this subsection, we will prove Proposition 2.4.2. Let $y^{(0)},\dots,y^{(k)}$ and $\beta_1,\dots,\beta_k$ be as in Proposition 2.4.2. Without loss of generality, we can set $y^{(0)} = \mathtt{T}^N$.

To analyze the expectation in Proposition 2.4.2, we introduce the following probabilistic quantities. For $0 \le \ell \le k$ and $\sigma \in \{\mathtt{T},\mathtt{F}\}^k$, define

$$E_\ell(\sigma) = \left\{I \in [N]^k : y^{(\ell')}[I] = \sigma \text{ for some } 0 \le \ell' \le \ell\right\} \qquad \text{and} \qquad p_\ell(\sigma) = \mathop{\mathbb{P}}_{I\sim \mathrm{unif}([N]^k)}\left(E_\ell(\sigma)\right).$$

In other words, $E_\ell(\sigma)$ is the event that σ appears in the set $\left\{y^{(\ell')}[I] : 0 \le \ell' \le \ell\right\}$, and $p_\ell(\sigma)$ is the probability of this event. The probabilities $p_k(\sigma)$ will be relevant to our analysis by the following identity (2.4.4), while the probabilities $p_\ell(\sigma)$ for $\ell < k$ will arise in our inductive analysis below, where we lower bound $p_k(\sigma)$ by peeling off one of $y^{(1)},\dots,y^{(k)}$ at a time. We have that

$$\mathop{\mathbb{E}}_{I\sim \mathrm{unif}([N]^k)} \left|\left\{y^{(\ell)}[I] : 0 \le \ell \le k\right\}\right| = \mathop{\mathbb{E}}_{I\sim \mathrm{unif}([N]^k)} \left[\sum_{\sigma\in\{\mathtt{T},\mathtt{F}\}^k} \mathbb{1}\left\{I \in E_k(\sigma)\right\}\right] = \sum_{\sigma\in\{\mathtt{T},\mathtt{F}\}^k} p_k(\sigma). \tag{2.4.4}$$

To prove Proposition 2.4.2, we will need to lower bound the right-hand side of (2.4.4). This task will require several definitions; to motivate these definitions, we first outline our technique for deriving this lower bound.

Our first step is a conditional expansion. Let $I \sim \mathrm{unif}([N]^k)$. We reveal the k bit strings $y^{(1)}[I],\dots,y^{(k)}[I]$ one by one. (Recall that we fixed $y^{(0)} = \mathtt{T}^N$, so $y^{(0)}[I] = \mathtt{T}^k$ is known.) Conditioned on its predecessors $y^{(1)}[I],\dots,y^{(\ell-1)}[I]$, the distribution of $y^{(\ell)}[I]$ can be described in terms of the conditional overlap profile $\pi(y^{(\ell)}|y^{(0)},\dots,y^{(\ell-1)})$. Then, $1 - p_k(\sigma)$, the probability that σ does not appear in $\left\{y^{(\ell)}[I] : 0 \le \ell \le k\right\}$, can be expanded as a product of k factors: the ℓth factor is the conditional probability that the revealed value of $y^{(\ell)}$ does not equal σ. The ℓth factor of this product can be thought of as (1 minus) the contribution of $y^{(\ell)}$ to $p_k(\sigma)$.

Our second step is to estimate this product by a *sum*, whose ℓth summand is the contribution of $y^{(\ell)}$ to this estimate of $p_k(\sigma)$. The purpose of this estimation is to decouple the contributions of the $y^{(\ell)}$, so that we can analyze the overall contribution of $y^{(\ell)}$ by summing over $\sigma \in \{\mathtt{T},\mathtt{F}\}^k$. We achieve this by truncating the factors in the product at $1 - \frac{1}{k\log k}$; any factor smaller than this gets rounded up to 1. Because $\frac{1}{k\log k} \ll \frac{1}{k}$, we can separate the contributions of $y^{(1)},\dots,y^{(k)}$ to $p_k(\sigma)$ by the estimate

$$1 - (1-\varepsilon_1)(1-\varepsilon_2)\cdots(1-\varepsilon_k) \approx \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_k$$

up to $1 - o_k(1)$ multiplicative error. Propositions 2.4.5 and 2.4.6 below carry out this technique.

Finally, our third step is to collect the (now additive) contributions of each $y^{(\ell)}$ to the estimated $p_k(\sigma)$ over all $\sigma \in \{\mathtt{T},\mathtt{F}\}^k$. Miraculously, we can interpret this sum as a probability of a sum of k i.i.d. random variables, which can be controlled by a Chernoff bound. This step is carried out in Proposition 2.4.7.

Formally, for $0 \le \ell \le k$ and $i \in [N]$, let $y_i^{(\le \ell)} = (y_i^{(1)}, \dots, y_i^{(\ell)})$. Similarly, for $I \in [N]^k$, let $y^{(\le \ell)}[I] = (y^{(1)}[I], \dots, y^{(\ell)}[I])$. Because $y^{(0)} = \mathtt{T}^N$, the overlap profile π determines the distribution of $y_i^{(\le k)}$ over $i \sim \mathrm{unif}([N])$. Namely, for $\xi \in \{\mathtt{T},\mathtt{F}\}^k$,

$$\mathop{\mathbb{P}}_{i\sim\mathrm{unif}([N])}\left[y_i^{(\le k)} = \xi\right] = \pi_{S\cup\{0\},T}$$

where $S = \{\ell \in [k] : \xi_\ell = \mathtt{T}\}$ and $T = \{\ell \in [k] : \xi_\ell = \mathtt{F}\}$. Moreover, the distribution of $y^{(\le k)}[I]$, where $I \sim \mathrm{unif}([N]^k)$, is the product of k i.i.d. copies of this distribution. For $1 \le \ell \le k$, $b \in \{\mathtt{T},\mathtt{F}\}$, and $\xi \in \{\mathtt{T},\mathtt{F}\}^{\ell-1}$, define

$$\phi_\ell(b|\xi) = \mathop{\mathbb{P}}_{i\sim\mathrm{unif}([N])}\left[y_i^{(\ell)} = b | y_i^{(\le \ell-1)} = \xi\right].$$

The probabilities in the aforementioned conditional expansion are products of conditional probabilities $\phi_\ell(b|\xi)$. Namely, the probability that $y^{(\ell)}[I] \ne \sigma$ given $y^{(\le \ell-1)}[I]$ is $1 - \prod_{r=1}^k \phi_\ell(\sigma_r | y_{I_r}^{(\le \ell-1)})$.

For $1 \le \ell \le k$, $\sigma \in \{\mathtt{T},\mathtt{F}\}^k$ and $I \in [N]^k$, further define

$$Q_\ell(\sigma, I) = \left(\prod_{r=1}^k \phi_\ell(\sigma_r | y_{I_r}^{(\le \ell-1)})\right) \mathbb{1}\left\{\prod_{r=1}^k \phi_\ell(\sigma_r | y_{I_r}^{(\le \ell-1)}) \le \frac{1}{k \log k}\right\} \quad \text{and} \quad q_\ell(\sigma) = \mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)}\left[Q_\ell(\sigma, I)\right].$$

Thus, $1 - Q_\ell(\sigma, I)$ is a term in the conditional expansion, truncated at $1 - \frac{1}{k\log k}$ in the aforementioned sense, and $q_\ell(\sigma)$ is its expectation.

For each $\sigma \in \{\mathtt{T},\mathtt{F}\}^k$, the following two propositions lower bound $p_k(\sigma)$ in terms of $q_1(\sigma), \dots, q_k(\sigma)$ by peeling off one of $y^{(1)}, \dots, y^{(k)}$ at a time.

Proposition 2.4.5. *For each $\sigma \in \{\mathtt{T},\mathtt{F}\}^k$ and $1 \le \ell \le k$, we have that*

$$p_\ell(\sigma) \ge \left(1 - \frac{1}{k \log k}\right) p_{\ell-1}(\sigma) + q_\ell(\sigma).$$

Proof. Note that

$$\begin{aligned} 1 - p_\ell(\sigma) &= \mathop{\mathbb{P}}_{I\sim\mathrm{unif}([N]^k)}\left[y^{(\ell')}[I] \ne \sigma \text{ for all } 0 \le \ell' \le \ell\right] \\ &= \mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)}\left[\mathbb{1}\left\{y^{(\ell')}[I] \ne \sigma \text{ for all } 0 \le \ell' \le \ell - 1\right\}\left(1 - \prod_{r=1}^k \phi_\ell(\sigma_r | y_{I_r}^{(\le \ell-1)})\right)\right]. \end{aligned}$$

Here, we use that the event inside the indicator is $y^{(\le \ell-1)}[I]$-measurable, and conditioned on $y^{(\le \ell-1)}[I]$ the probability that $y^{(\ell)}[I] = \sigma$ is $\prod_{r=1}^k \phi_\ell(\sigma_r | y_{I_r}^{(\le \ell-1)})$. Moreover, we have $\prod_{r=1}^k \phi_\ell(\sigma_r | y_{I_r}^{(\le \ell-1)}) \ge Q_\ell(\sigma, I)$ by definition. So,

$$\begin{aligned} 1 - p_\ell(\sigma) &\le \mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)}\left[\left(1 - \mathbb{1}\left\{I \in E_{\ell-1}(\sigma)\right\}\right)\left(1 - Q_\ell(\sigma, I)\right)\right] \\ &\le \mathop{\mathbb{E}}_{I\sim\mathrm{unif}([N]^k)}\left[1 - \left(1 - \frac{1}{k\log k}\right)\mathbb{1}\left\{I \in E_{\ell-1}(\sigma)\right\} - Q_\ell(\sigma, I)\right] \\ &= 1 - \left(1 - \frac{1}{k\log k}\right) p_{\ell-1}(\sigma) - q_\ell(\sigma). \end{aligned}$$

The second-last line uses the fact that $Q_\ell(\sigma, I) \le \frac{1}{k\log k}$ almost surely, and the last line uses the definitions of $p_{\ell-1}(\sigma)$ and $q_\ell(\sigma)$. Rearranging yields the desired bound. □

Proposition 2.4.6. *For each $\sigma \in \{\mathtt{T},\mathtt{F}\}^k$, we have that*

$$p_k(\sigma) \geq \left(1 - \frac{1}{\log k}\right) \sum_{\ell=1}^{k} q_\ell(\sigma).$$

Proof. By iterating Proposition 2.4.5, we get

$$p_k(\sigma) \geq \left(1 - \frac{1}{k \log k}\right)^k p_0(\sigma) + \sum_{\ell=1}^{k} \left(1 - \frac{1}{k \log k}\right)^{k-\ell} q_\ell(\sigma) \geq \left(1 - \frac{1}{k \log k}\right)^k \sum_{\ell=1}^{k} q_\ell(\sigma).$$

The result follows from the bound $\left(1 - \frac{1}{k \log k}\right)^k \geq 1 - \frac{1}{\log k}$, by Bernoulli's inequality. □

Equation (2.4.4) and Proposition 2.4.6 leave the task of lower bounding $\sum_{\sigma \in \{\mathtt{T},\mathtt{F}\}^k} \sum_{\ell=1}^{k} q_\ell(\sigma)$. This is achieved by the following proposition, which reinterprets $\sum_{\sigma \in \{\mathtt{T},\mathtt{F}\}^k} q_\ell(\sigma)$, the total contribution of $y^{(\ell)}$, as a probability.

Proposition 2.4.7. *For each $1 \leq \ell \leq k$, we have that*

$$\sum_{\sigma \in \{\mathtt{T},\mathtt{F}\}^k} q_\ell(\sigma) \geq 1 - \beta_\ell e^{-(\beta_\ell - 1)} - o_k(1).$$

Proof. Using the definition of $q_\ell(\sigma)$, we have

$$\begin{aligned}
\sum_{\sigma \in \{\mathtt{T},\mathtt{F}\}^k} q_\ell(\sigma) &= \mathop{\mathbb{E}}_{I \sim \mathrm{unif}([N]^k)} \left[\sum_{\sigma \in \{\mathtt{T},\mathtt{F}\}^k} \left(\prod_{r=1}^{k} \phi_\ell(\sigma_r | y_{I_r}^{(\leq \ell-1)}) \right) \mathbb{1}\left\{ \prod_{r=1}^{k} \phi_\ell(\sigma_r | y_{I_r}^{(\leq \ell-1)}) \leq \frac{1}{k \log k} \right\} \right] \\
&= \mathop{\mathbb{E}}_{I \sim \mathrm{unif}([N]^k)} \left[\sum_{\sigma \in \{\mathtt{T},\mathtt{F}\}^k} \left(\prod_{r=1}^{k} \phi_\ell(\sigma_r | y_{I_r}^{(\leq \ell-1)}) \right) \mathbb{1}\left\{ -\sum_{r=1}^{k} \log \phi_\ell(\sigma_r | y_{I_r}^{(\leq \ell-1)}) \geq \log k + \log\log k \right\} \right].
\end{aligned}$$

This quantity is the success probability of the following experiment. Sample positive random variables $u_1, \ldots, u_k$ by the following procedure, repeated independently for each $r \in [k]$. Sample $i \in \mathrm{unif}([N])$; this determines the value of $y_i^{(\leq \ell-1)}$. Then, sample $b \in \{\mathtt{T},\mathtt{F}\}$ from the measure $\phi_\ell(\cdot | y_i^{(\leq \ell-1)})$. Finally, set $u_r = -\log \phi_\ell(b | y_{I_r}^{(\leq \ell-1)})$. The experiment succeeds if $\sum_{r=1}^{k} u_r \geq \log k + \log\log k$.

For $r \in [k]$, let $v_r = \min(u_r, \log k)$. This is a proxy for u_r with an almost sure upper bound, which allows us to control the experiment's failure probability by a Chernoff bound. This failure probability is bounded by

$$\mathbb{P}\left[\sum_{r=1}^{k} u_r < \log k + \log\log k \right] \leq \mathbb{P}\left[\sum_{r=1}^{k} v_r < \log k + \log\log k \right] = \mathbb{P}\left[\sum_{r=1}^{k} \frac{v_r}{\log k} < 1 + \frac{\log\log k}{\log k} \right].$$

Note that the $\frac{v_r}{\log k}$ are i.i.d. random variables in $[0,1]$ almost surely. To bound this last probability by a Chernoff bound, we will lower bound $\mathbb{E}[v_r]$. By the definition of ϕ_ℓ,

$$\begin{aligned}
\mathbb{E}[u_r] &= \mathop{\mathbb{E}}_{i \sim \mathrm{unif}([N])} \mathop{\mathbb{E}}_{b \sim \phi_\ell(\cdot | y_i^{(\leq \ell-1)})} \left[-\log \phi_\ell(b | y_i^{(\leq \ell-1)}) \right] \\
&= H(\pi(y^{(\ell)} | y^{(0)}, \ldots, y^{(\ell-1)})) = \beta_\ell \frac{\log k}{k}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\mathbb{E}[u_r - v_r] &= \mathbb{E}\left[(u_r - \log k) \mathbb{1}\left\{ u_r \geq \log k \right\} \right] \\
&= \mathop{\mathbb{E}}_{i \sim \mathrm{unif}([N])} \left[\sum_{b \in \{\mathtt{T},\mathtt{F}\}} \phi_\ell(b | y_i^{(\leq \ell-1)}) \log \frac{1}{k \phi_\ell(b | y_i^{(\leq \ell-1)})} \mathbb{1}\left\{ \phi_\ell(b | y_i^{(\leq \ell-1)}) \leq \frac{1}{k} \right\} \right].
\end{aligned}$$

For each $i \in [N]$, the quantity inside the last expectation is nonzero for at most one $b \in \{\mathtt{T}, \mathtt{F}\}$ (for $k \geq 3$). Moreover, on the interval $[0, \frac{1}{k}]$, the function $x \mapsto x \log \frac{1}{kx}$ has maximum value $\frac{1}{ek}$, attained at $x = \frac{1}{ek}$. Thus, $\mathbb{E}[u_r - v_r] \leq \frac{1}{ek}$. It follows that $\mathbb{E}[v_r] \geq \beta_\ell \frac{\log k}{k} - \frac{1}{ek}$. So,

$$\mathbb{E}\left[\sum_{r=1}^{k} \frac{v_r}{\log k}\right] \geq \beta_\ell - \frac{1}{e \log k}.$$

Furthermore, $(1 + \frac{\log\log k}{\log k})/(\beta_\ell - \frac{1}{e \log k}) = \frac{1}{\beta_\ell} + o_k(1)$. So, by a Chernoff bound,

$$\mathbb{P}\left[\sum_{r=1}^{k} \frac{v_r}{\log k} < 1 + \frac{\log\log k}{\log k}\right] \leq \left(\frac{e^{-(1-\frac{1}{\beta_\ell} - o_k(1))}}{\left(\frac{1}{\beta_\ell} + o_k(1)\right)^{\frac{1}{\beta_\ell} + o_k(1)}}\right)^{\beta_\ell - o_k(1)} = \beta_\ell e^{-(\beta_\ell - 1)} + o_k(1).$$

Hence,

$$\mathbb{P}\left[\sum_{r=1}^{k} u_r \geq \log k + \log\log k\right] \geq 1 - \beta_\ell e^{-(\beta_\ell - 1)} - o_k(1),$$

as desired. □

We can now combine these propositions to prove Proposition 2.4.2.

Proof of Proposition 2.4.2. By combining (2.4.4), Proposition 2.4.6, and Proposition 2.4.7, we have

$$\begin{aligned}\mathop{\mathbb{E}}_{I \sim \mathrm{unif}([N]^k)} \left|\left\{y^{(\ell)}[I] : 0 \leq \ell \leq k\right\}\right| &\geq \left(1 - \frac{1}{\log k}\right) \sum_{\ell=1}^{k} \sum_{\sigma \in \{\mathtt{T},\mathtt{F}\}^k} q_\ell(\sigma) \\ &\geq (1 - o_k(1)) \sum_{\ell=1}^{k} \left(1 - \beta_\ell e^{-(\beta_\ell - 1)}\right).\end{aligned}$$

□

2.5 Stability of Low Degree Polynomials

In this section, we will prove Proposition 2.3.7(a), which lower bounds the probability that $x^{(t)}$ satisfies $\Phi^{(t)}$ for all $0 \leq t \leq T$ and the sequence $x^{(t)}$ has no large jumps in Hamming distance.

The proof is a mild generalization of the stability analysis in [GJW20, Subsection 4.1] and [Wei20, Subsection 2.3] from a biased Boolean hypercube to a product of discrete uniform measures. Like in these two works, the proof proceeds in two steps. The interpolation path can be modeled as a walk on a product graph whose vertices are the elements of $\Omega_k(N, M)$, where two vertices are adjacent if they differ by one literal. An edge (Φ, Φ') is bad if the output of our polynomial f has a large jump between inputs Φ and Φ'. In the first step, we will use Fourier analysis to upper bound the fraction of bad edges. In the second step, we translate this bound to a lower bound on the probability that our walk encounters no bad edges.

2.5.1 An Upper Bound on the Rate of Bad Steps

We begin by formalizing the notion of c-badness. Recall that $n = M \cdot k \cdot 2N$, and each $\Phi \in \Omega_k(N, M)$ is identified with a vector of indicators in $\{0,1\}^n$, which is the input of a low degree polynomial.

Definition 2.5.1 (c-badness)**.** Let $c > 0$ and let $f : \mathbb{R}^n \to \mathbb{R}^N$ be a deterministic degree-D polynomial. A pair of formulas $(\Phi, \Phi') \in \Omega_k(N, M)^2$ is *c-bad* (with respect to f) if $\|f(\Phi) - f(\Phi')\|_2^2 > c\, \mathbb{E}_{\Phi \sim \Phi_k(N,M)} \|f(\Phi)\|_2^2$.

Recall the interpolation path $\Phi^{(0)}, \Phi^{(1)}, \ldots, \Phi^{(T)}$ defined in Definition 2.3.2. We will prove Proposition 2.3.7(a) via the following proposition, which controls the probability that the output of f does not have a large jump between any pair of consecutive assignments in the interpolation path.

Proposition 2.5.2. *Let $f : \mathbb{R}^n \to \mathbb{R}^N$ be a deterministic degree-D polynomial. With probability at least $(2N)^{-4Dk/c}$, $(\Phi^{(t-1)}, \Phi^{(t)})$ is not c-bad with respect to f for any $1 \le t \le T$.*

We will prove this proposition in Subsection 2.5.3. The objective of this subsection is to prove Proposition 2.5.3 below, which upper bounds the fraction of all possible steps that are bad. To this end, for $1 \le j \le kM$, define $\Phi_k(N, M; j)$ as the measure of a sample $(\Phi, \Phi') \in \Omega_k(N, M)^2$ obtained by sampling $\Phi \sim \Phi_k(N, M)$, and then obtaining Φ' from Φ by resampling the jth lexicographic literal $\Phi'_{L(j)}$ from unif $\left(\mathcal{L} \setminus \{\Phi_{L(j)}\}\right)$. (Recall the definition of $L(j)$ before Definition 2.3.2.) Define

$$\lambda_j = \mathop{\mathbb{P}}_{(\Phi,\Phi') \sim \Phi_k(N,M;j)} \left((\Phi, \Phi') \text{ is } c\text{-bad with respect to } f\right).$$

This is the fraction of pairs of formulas in $\Omega_k(N, M)$, differing in exactly the jth lexicographic literal, that are c-bad with respect to f.

Proposition 2.5.3. *If f is a deterministic degree-D polynomial, then $\sum_{j=1}^{kM} \lambda_j \le \frac{4D}{c}$.*

We recall the following orthogonal decomposition property of functions on product measures, which can be thought of as a generalization of Fourier analysis on the Boolean cube. We will give brief self-contained proofs of the relevant facts; a full discussion can be found in [O'D14, Chapter 8.3]. Let $(\mathcal{X}, \mathbb{P}_X)$ be an arbitrary probability space, and let J be a positive integer. Let $X = (X_1, \dots, X_J) \in \mathcal{X}^J$. For $j \in [J]$, define the operators D_j and E_j as follows. For any function $g : \mathcal{X}^J \to \mathbb{R}$, $\mathsf{E}_j g$ is the function satisfying

$$\mathsf{E}_j g(X) = \mathop{\mathbb{E}}_{X_j \sim (\mathcal{X}, \mathbb{P}_X)} g(X),$$

where in the right-hand side the coordinate X_j is resampled from $(\mathcal{X}, \mathbb{P}_X)$. Let $\mathsf{D}_j g = g - \mathsf{E}_j g$. Note that the operators $\{\mathsf{D}_j, \mathsf{E}_j\}_{j \in [J]}$ commute. For $S \subseteq [J]$, define the functions

$$\hat{g}_S = \prod_{j \in S} \mathsf{D}_j \prod_{j \in [J] \setminus S} \mathsf{E}_j g.$$

Note that $g = \sum_{S \subseteq [J]} \hat{g}_S$. Moreover, $\hat{g}_S$ depends only on the inputs $\{X_j : j \in S\}$. For any j,

$$\mathop{\mathbb{E}}_{X_j \sim (\mathcal{X}, \mathbb{P}_X)} g(X)^2 = \mathop{\mathbb{E}}_{X_j \sim (\mathcal{X}, \mathbb{P}_X)} \left[(\mathsf{D}_j g)(X)^2\right] + (\mathsf{E}_j g)(X)^2,$$

and so by induction

$$\mathop{\mathbb{E}}_{X \sim (\mathcal{X}, \mathbb{P}_X)^{\otimes J}} g(X)^2 = \sum_{S \subseteq [J]} \mathop{\mathbb{E}}_{X \sim (\mathcal{X}, \mathbb{P}_X)^{\otimes J}} \hat{g}_S(X)^2.$$

For $j \in [J]$, define $\mathrm{Var}_j g(X) = \mathbb{E}_{X_j \sim (\mathcal{X}, \mathbb{P}_X)} \left[(\mathsf{D}_j g)(X)^2\right]$. We begin with the following inequality, which can be considered a converse to the Efron-Stein inequality.

Lemma 2.5.4. *Suppose a function $g : \mathcal{X}^J \to \mathbb{R}$ can be written in the form $g(X) = \sum_{i=1}^{I} g_i(X)$, where each $g_i(X)$ depends on at most D coordinates of X. Then,*

$$D \mathop{\mathrm{Var}}_{X \sim (\mathcal{X}, \mathbb{P}_X)^{\otimes J}} g(X) \ge \sum_{j=1}^{J} \mathop{\mathbb{E}}_{X \sim (\mathcal{X}, \mathbb{P}_X)^{\otimes J}} \mathrm{Var}_j g(X).$$

Proof. By the orthogonal expansion above, we have

$$\mathop{\mathrm{Var}}_{X \sim (\mathcal{X}, \mathbb{P}_X)^{\otimes J}} g(X) = \sum_{\substack{S \subseteq [J] \\ S \neq \emptyset}} \mathop{\mathbb{E}}_{X \sim (\mathcal{X}, \mathbb{P}_X)^{\otimes J}} \hat{g}_S(X)^2 \qquad \text{and} \qquad \mathop{\mathbb{E}}_{X \sim (\mathcal{X}, \mathbb{P}_X)^{\otimes J}} \mathrm{Var}_j g(X) = \sum_{\substack{S \subseteq [J] \\ S \ni j}} \mathop{\mathbb{E}}_{X \sim (\mathcal{X}, \mathbb{P}_X)^{\otimes J}} \hat{g}_S(X)^2.$$

We claim that for all $S \subseteq [J]$ with $|S| > D$, we have $\hat{g}_S \equiv 0$. For each $i \in [I]$, we have $\prod_{j \in S} \mathsf{D}_j g_i \equiv 0$, because S contains at least one j such that $g_i(X)$ does not depend on X_j. Thus, $\prod_{j \in S} \mathsf{D}_j g \equiv 0$, and so $\hat{g}_S \equiv 0$, as desired. Hence,

$$\sum_{j=1}^{J} \mathop{\mathbb{E}}_{X \sim (\mathcal{X}, \mathbb{P}_X)^{\otimes J}} \mathrm{Var}_j g(X) = \sum_{\substack{S \subseteq [J] \\ |S| \le D}} |S| \mathop{\mathbb{E}}_{X \sim (\mathcal{X}, \mathbb{P}_X)^{\otimes j}} \hat{g}_S(X)^2 \le D \mathop{\mathrm{Var}}_{X \sim (\mathcal{X}, \mathbb{P}_X)^{\otimes J}} g(X).$$

□

Proof of Proposition 2.5.3. Note that $\Phi_k(N,M)$ is composed of kM i.i.d. literals, and thus can be thought of as the product measure $\mathrm{unif}(\mathcal{L})^{\otimes kM}$. By slight abuse of notation, for $1 \le j \le kM$, we can define D_j and E_j as the above operators with respect to the jth lexicographic literal $\Phi_{L(j)}$ of Φ.

For $1 \le \ell \le N$, let f_ℓ denote the ℓth component of f. By Markov's inequality and the inequality $(a-b)^2 \le 2a^2 + 2b^2$, we have

$$\begin{aligned}
\sum_{j=1}^{kM} \lambda_j &\le \sum_{j=1}^{kM} \frac{\mathbb{E}_{(\Phi,\Phi')\sim\Phi_k(N,M;j)} \|f(\Phi) - f(\Phi')\|_2^2}{c\,\mathbb{E}_{\Phi\sim\Phi_k(N,M)} \|f(\Phi)\|_2^2} \\
&= \frac{\sum_{\ell=1}^N \sum_{j=1}^{kM} \mathbb{E}_{(\Phi,\Phi')\sim\Phi_k(N,M;j)} ((\mathsf{D}_j f_\ell)(\Phi) - (\mathsf{D}_j f_\ell)(\Phi'))^2}{c\sum_{\ell=1}^N \mathbb{E}_{\Phi\sim\Phi_k(N,M)} f_\ell(\Phi)^2} \\
&\le \frac{2\sum_{\ell=1}^N \sum_{j=1}^{kM} \mathbb{E}_{(\Phi,\Phi')\sim\Phi_k(N,M;j)} \left((\mathsf{D}_j f_\ell)(\Phi)^2 + (\mathsf{D}_j f_\ell)(\Phi')^2\right)}{c\sum_{\ell=1}^N \mathbb{E}_{\Phi\sim\Phi_k(N,M)} f_\ell(\Phi)^2} \\
&= \frac{4\sum_{\ell=1}^N \sum_{j=1}^{kM} \mathbb{E}_{\Phi\sim\Phi_k(N,M)} \mathrm{Var}_j f_\ell(\Phi)}{c\sum_{\ell=1}^N \mathbb{E}_{\Phi\sim\Phi_k(N,M)} f_\ell(\Phi)^2}.
\end{aligned}$$

Now, each f_ℓ is a degree-D polynomial in the indicators $\Phi_{i,j,s}$ that $\Phi_{i,j}$ is the sth literal in $\mathcal{L}$. So, each monomial of each f_ℓ depends on at most D literals of Φ. By Lemma 2.5.4,

$$\sum_{j=1}^{kM} \mathop{\mathbb{E}}_{\Phi\sim\Phi_k(N,M)} \mathrm{Var}_j f_\ell(\Phi) \le D \mathop{\mathrm{Var}}_{\Phi\sim\Phi_k(N,M)} f_\ell(\Phi) \le D \mathop{\mathbb{E}}_{\Phi\sim\Phi_k(N,M)} f_\ell(\Phi)^2.$$

So, $\sum_{j=1}^{kM} \lambda_j \le \frac{4D}{c}$. □

2.5.2 Bounding the Probability of no Bad Step

Proposition 2.5.3 bounds the combined rate of c-bad steps. To derive Proposition 2.5.2, we must translate this bound on the rate of c-bad steps to a bound on the probability that interpolation path never takes a c-bad step. To make the ideas in our argument more clear, we abstract to the following graph theoretic problem, which is interesting in its own right.

Let Σ be a set of symbols and J, T be positive integers. Let G be a graph on Σ^J, where two nodes are adjacent if their Hamming distance is exactly 1. Each edge has a *direction* $j \in [J]$, the index on which its endpoints disagree. Let an arbitrary subset of edges be *bad*; for adjacent vertices v, w, let $B(v,w)$ denote the event that the edge (v,w) is bad. For $j \in [J]$, let λ_j denote the fraction of edges in direction j that are bad. Equivalently, $\lambda_j = \mathbb{P}(B(v,w))$, where $v \sim \mathrm{unif}(G)$ and w is obtained from v by resampling w_j from $\mathrm{unif}\,(\Sigma \setminus \{v_j\})$.

Let $\sigma : [T] \to [J]$ be an arbitrary map. Consider the (lazy) random walk $v^{(0)}, v^{(1)}, \ldots, v^{(T)}$ such that $v^{(0)} \sim \mathrm{unif}(G)$ and for $1 \le t \le T$, $v^{(t)}$ is obtained from $v^{(t-1)}$ by resampling $v^{(t)}_{\sigma(t)}$ from $\mathrm{unif}(\Sigma)$.

Lemma 2.5.5. *With probability at least $|\Sigma|^{-\sum_{t=1}^T \lambda_{\sigma(t)}}$, no step of the random walk $v^{(0)}, v^{(1)}, \ldots, v^{(T)}$ traverses a bad edge.*

Note that at each step, the random walk either traverses an edge or does not move; we say that the steps that do not move do not traverse a bad edge. The lemma is sharp, for example, when all the λ_j are 0 or 1: in this case, the random walk does not traverse a bad edge if it does not move at all times t with $\lambda_{\sigma(t)} = 1$.

Proof. For $v \in \Sigma^J$, let $q(v)$ be the probability that the random walk $v^{(0)}, v^{(1)}, \ldots, v^{(T)}$ does not traverse a bad edge, starting from $v^{(0)} = v$. We will prove by induction on T that

$$\mathop{\mathbb{E}}_{v\sim\mathrm{unif}(G)} \log q(v) \ge -\log|\Sigma| \cdot \sum_{t=1}^T \lambda_{\sigma(t)}.$$

The lemma then follows from Jensen's inequality, because $\log \mathbb{E}\, q(v) \ge \mathbb{E} \log q(v)$.

The base case of the claim, $T = 0$, follows trivially. For the inductive step, let $\tilde{q}(v)$ be the probability that the random walk $v^{(1)}, v^{(2)}, \ldots, v^{(T)}$ does not traverse a bad edge, starting from $v^{(1)} = v$. Let $j = \sigma(1)$. Let $v_{\sim j} \in \Sigma^{J-1}$ denote an element of Σ^J with the jth coordinate left blank. For $s \in \Sigma$, let $v_{\sim j}[s] \in \Sigma^J$ denote $v_{\sim j}$ with the jth coordinate set to s.

For now, fix some $v_{\sim j} \in \Sigma^{J-1}$. For $s \in \Sigma$, we have that

$$\begin{aligned} q(v_{\sim j}[s]) &= \sum_{s' \in \Sigma} \frac{1}{|\Sigma|} \mathbb{1}\left\{s' = s \text{ or } B\left(v_{\sim j}[s], v_{\sim j}[s']\right)^c\right\} \tilde{q}(v_{\sim j}[s']) \\ &= \sum_{s' \in \Sigma \setminus \{s\}} \frac{1}{|\Sigma| - 1} \left(\frac{1}{|\Sigma|} \tilde{q}(v_{\sim j}[s]) + \frac{|\Sigma| - 1}{|\Sigma|} \mathbb{1}\left\{B\left(v_{\sim j}[s], v_{\sim j}[s']\right)^c\right\} \tilde{q}(v_{\sim j}[s']) \right). \end{aligned}$$

By Jensen's inequality, this implies

$$\log q(v_{\sim j}[s]) \geq \sum_{s' \in \Sigma \setminus \{s\}} \frac{1}{|\Sigma| - 1} \log \left(\frac{1}{|\Sigma|} \tilde{q}(v_{\sim j}[s]) + \frac{|\Sigma| - 1}{|\Sigma|} \mathbb{1}\left\{B\left(v_{\sim j}[s], v_{\sim j}[s']\right)^c\right\} \tilde{q}(v_{\sim j}[s']) \right).$$

Taking an expectation over $s \sim \text{unif}(\Sigma)$, we have

$$\begin{aligned} \mathop{\mathbb{E}}_{s \sim \text{unif}(\Sigma)} \log q(v_{\sim j}[s]) &\geq \sum_{\substack{s, s' \in \Sigma \\ s \neq s'}} \frac{1}{|\Sigma|(|\Sigma| - 1)} \left[\log \left(\frac{1}{|\Sigma|} \tilde{q}(v_{\sim j}[s]) + \frac{|\Sigma| - 1}{|\Sigma|} \mathbb{1}\left\{B\left(v_{\sim j}[s], v_{\sim j}[s']\right)^c\right\} \tilde{q}(v_{\sim j}[s']) \right) \right] \\ &= \sum_{\substack{s, s' \in \Sigma \\ s \neq s'}} \frac{1}{2|\Sigma|(|\Sigma| - 1)} \xi(v_{\sim j}, s, s'), \end{aligned} \tag{2.5.1}$$

where for $s \neq s'$,

$$\begin{aligned} \xi(v_{\sim j}, s, s') = &\log \left(\frac{1}{|\Sigma|} \tilde{q}(v_{\sim j}[s]) + \frac{|\Sigma| - 1}{|\Sigma|} \mathbb{1}\left\{B\left(v_{\sim j}[s], v_{\sim j}[s']\right)^c\right\} \tilde{q}(v_{\sim j}[s']) \right) \\ &+ \log \left(\frac{1}{|\Sigma|} \tilde{q}(v_{\sim j}[s']) + \frac{|\Sigma| - 1}{|\Sigma|} \mathbb{1}\left\{B\left(v_{\sim j}[s], v_{\sim j}[s']\right)^c\right\} \tilde{q}(v_{\sim j}[s]) \right). \end{aligned}$$

If $B\left(v_{\sim j}[s], v_{\sim j}[s']\right)$ holds, then $\xi(v_{\sim j}, s, s') = \log \tilde{q}(v_{\sim j}[s]) + \log \tilde{q}(v_{\sim j}[s']) - 2 \log |\Sigma|$. Otherwise, by Jensen's inequality we have

$$\log \left(\frac{1}{|\Sigma|} \tilde{q}(v_{\sim j}[s]) + \frac{|\Sigma| - 1}{|\Sigma|} \mathbb{1}\left\{B\left(v_{\sim j}[s], v_{\sim j}[s']\right)^c\right\} \tilde{q}(v_{\sim j}[s']) \right) \geq \frac{1}{|\Sigma|} \log \tilde{q}(v_{\sim j}[s]) + \frac{|\Sigma| - 1}{|\Sigma|} \log \tilde{q}(v_{\sim j}[s'])$$

and similarly for the other term of $\xi(v_{\sim j}, s, s')$. In this case, $\xi(v_{\sim j}, s, s') \geq \log \tilde{q}(v_{\sim j}[s]) + \log \tilde{q}(v_{\sim j}[s'])$. So, in all cases

$$\xi(v_{\sim j}, s, s') \geq \log \tilde{q}(v_{\sim j}[s]) + \log \tilde{q}(v_{\sim j}[s']) - 2\mathbb{1}\left\{B\left(v_{\sim j}[s], v_{\sim j}[s']\right)\right\} \log |\Sigma|.$$

Substituting into (2.5.1), we have

$$\mathop{\mathbb{E}}_{s \sim \text{unif}(\Sigma)} \log q(v_{\sim j}[s]) \geq \mathop{\mathbb{E}}_{s \sim \text{unif}(\Sigma)} \log \tilde{q}(v_{\sim j}[s]) - \log |\Sigma| \cdot \sum_{\substack{s, s' \in \Sigma \\ s \neq s'}} \frac{\mathbb{1}\left\{B\left(v_{\sim j}[s], v_{\sim j}[s']\right)\right\}}{|\Sigma|(|\Sigma| - 1)}.$$

Taking an expectation over $v_{\sim j}$ yields

$$\mathop{\mathbb{E}}_{v \sim \text{unif}(G)} \log q(v) \geq \mathop{\mathbb{E}}_{v \sim \text{unif}(G)} \log \tilde{q}(v) - \log |\Sigma| \cdot \lambda_j.$$

By induction, we have

$$\mathop{\mathbb{E}}_{v \sim \text{unif}(G)} \log \tilde{q}(v) \geq -\log |\Sigma| \cdot \sum_{t=2}^{T} \lambda_{\sigma(t)},$$

and the result follows. □

2.5.3 Completing the Proof of Stability

Proof of Proposition 2.5.2. Our interpolation scheme can be modeled as the random walk in Subsection 2.5.2, with $\Sigma = \mathcal{L}$, $J = kM$, $T = k^2M$, $\sigma(t)$ defined in Definition 2.3.2, and where the bad edges are the c-bad edges. This correspondence is consistent because the steps in the interpolation path where a formula transitions to itself are never c-bad.

Since σ maps to every value in $[kM]$ k times and $|\mathcal{L}| = 2N$, Proposition 2.5.3 and Lemma 2.5.5 imply that the probability of never traversing a c-bad edge is at least $(2N)^{-4Dk/c}$. □

Proof of Proposition 2.3.7(a). Set $c = \frac{\beta_+ - \beta_-}{\gamma k}$. Let $S_{\text{no-bad}}$ be the event that for all $1 \le t \le T$, $(\Phi^{(t-1)}, \Phi^{(t)})$ is not c-bad with respect to f. By Proposition 2.5.2, $\mathbb{P}(S_{\text{no-bad}}) \ge (2N)^{-4Dk^2\gamma/(\beta_+ - \beta_-)}$.

By a union bound, $\mathbb{P}(S_{\text{valid}}) \ge 1 - (T+1)\delta$. Thus, $\mathbb{P}(S_{\text{valid}} \cap S_{\text{no-bad}}) \ge (2N)^{-4Dk^2\gamma/(\beta_+ - \beta_-)} - (T+1)\delta$. We claim that on $S_{\text{valid}} \cap S_{\text{no-bad}}$, the event S_{consec} also occurs.

Suppose for sake of contradiction that $S_{\text{valid}} \cap S_{\text{no-bad}}$ holds and for some $1 \le t \le T$, we have that $\Delta(x^{(t-1)}, x^{(t)}) > \frac{\beta_+ - \beta_-}{2k}$. Because $(\Phi^{(t-1)}, \Phi^{(t)})$ is not c-bad, we have

$$\left\| f(\Phi^{(t-1)}) - f(\Phi^{(t)}) \right\|_2^2 \le c \mathop{\mathbb{E}}_{\Phi \sim \Phi_k(N,M)} \|f(\Phi)\|_2^2 \le c\gamma N = \frac{\beta_+ - \beta_-}{k} N.$$

Let $I = \{i \in [N] : x_i^{(t-1)} \ne x_i^{(t)}\}$, so $|I| > \frac{\beta_+ - \beta_-}{2k} N$. Define

$$B^{(t)} = \left\{ i \in [N] : x_i^{(t)} \ne (\texttt{round} \circ f)(\Phi^{(t)})_i \right\}.$$

Because $x_i^{(t)} = \mathcal{A}(\Phi^{(t)})_i$ and the assistance subroutine $\mathcal{B}$ in $\mathcal{A}$ can edit only an η fraction of bits of the assignment, $|B^{(t)}| \le \eta N = \frac{\beta_+ - \beta_-}{8k} N$. For similarly defined $B^{(t-1)}$, we likewise have $|B^{(t-1)}| \le \frac{\beta_+ - \beta_-}{8k} N$.

Let $J = I \setminus (B^{(t-1)} \cup B^{(t)})$, so $|J| > \frac{\beta_+ - \beta_-}{4k} N$. For all $i \in J$, one of $f_i(\Phi^{(t-1)})$ and $f_i(\Phi^{(t)})$ is at least 1 and the other is at most -1, so $|f_i(\Phi^{(t-1)}) - f_i(\Phi^{(t)})| \ge 2$. So,

$$\left\| f(\Phi^{(t-1)}) - f(\Phi^{(t)}) \right\|_2^2 \ge \sum_{i \in J} |f_i(\Phi^{(t-1)}) - f_i(\Phi^{(t)})|^2 > \frac{\beta_+ - \beta_-}{k} N.$$

This is a contradiction. Therefore $S_{\text{consec}} \supseteq S_{\text{valid}} \cap S_{\text{no-bad}}$, and so

$$\mathbb{P}(S_{\text{valid}} \cap S_{\text{consec}}) \ge \mathbb{P}(S_{\text{valid}} \cap S_{\text{no-bad}}) \ge (2N)^{-4Dk^2\gamma/(\beta_+ - \beta_-)} - (T+1)\delta.$$

□

2.6 Proof of Impossibility for Local Algorithms

This section proves our impossibility result for local algorithms, Theorem 2.1.13. Throughout, fix $\kappa > \kappa^*$ and $r \in \mathbb{N}$. Fix a probability space $(\Omega, \mathbb{P}_\omega)$, and let $\mathcal{A}$ be an r-local algorithm that, on input $\Phi \in \Omega_k(N, M)$ with factor graph (G, ρ), samples internal randomness $\varphi \sim (\Omega, \mathbb{P}_\omega)^{\otimes(V_G \cup E_G)}$.

Let $\varphi_V = \varphi\big|_{V_G}$ and $\varphi_E = \varphi\big|_{E_G}$. We will actually prove Theorem 2.1.13 conditioned on any realization of φ_V. Fix once and for all a realization of φ_V; all probabilities and expectations in this section will implicitly be conditioned on this realization.

2.6.1 A Different Interpolation

Instead of the interpolation path of k^2M problem instances used in the proof of Theorem 2.1.6, we now use an interpolation structured as k paths of length kM originating at a common point. We also couple to this interpolation the internal randomness φ_E of $\mathcal{A}$ run on these problem instances.

Definition 2.6.1 (Interplation structure)**.** Let $T' = kM$. We will sample $\Phi^{(0)}$ and $\Phi^{(\ell,t)} \in \Omega_k(N,M)$ for $1 \le \ell \le k$, $1 \le t \le T'$. Let the factor graphs of these k-SAT instances be $(G^{(0)}, \rho^{(0)})$ and $(G^{(\ell,t)}, \rho^{(\ell,t)})$. We also sample maps $\varphi_E^{(0)} : E_{G^{(0)}} \to \Omega$ and $\varphi_E^{(\ell,t)} : E_{G^{(\ell,t)}} \to \Omega$.

We sample $\Phi^{(0)} \sim \Phi_k(N,M)$ and $\varphi_E^{(0)} \sim (\Omega, \mathbb{P}_\omega)^{\otimes E_{G^{(0)}}}$. For $1 \le \ell \le k$, $1 \le t \le T'$, we obtain $\Phi^{(\ell,t)}$ from $\Phi^{(\ell,t-1)}$ (take $\Phi^{(\ell,0)} = \Phi^{(0)}$ for all ℓ) by resampling $\Phi^{(\ell,t)}_{L(t)}$ from $\mathrm{unif}(\mathcal{L})$. (Recall that $L(t)$ is the tth pair $(a,b) \in [M] \times [k]$ in lexicographic order.)

This resampling deletes an edge e from $G^{(\ell,t-1)}$, adds an edge e' to $G^{(\ell,t)}$ (possibly in the same location), and samples $\rho^{(\ell,t)}(e) \sim \mathrm{unif}(\{\mathtt{T},\mathtt{F}\})$. We obtain $\varphi_E^{(\ell,t)}$ from $\varphi_E^{(\ell,t-1)}$ by deleting the entry for e' and sampling $\varphi_E^{(\ell,t)}(e) \sim (\Omega, \mathbb{P}_\omega)$.

In other words, starting from a random k-SAT instance we sample k interpolation paths, where in each path we resample the literals one by one in the same order. Resampling a literal resamples an edge of the factor graph, and we also resample the output of φ_E on that edge.

Note that each $(\Phi^{(\ell,t)}, \varphi_E^{(\ell,t)})$ is marginally distributed as (Φ, φ_E) where $\Phi \sim \Phi_k(N,M)$ and $\varphi_E \sim (\Omega, \mathbb{P}_\omega)^{\otimes E_G}$, where (G,ρ) is the factor graph of Φ. Moreover, $(\Phi^{(\ell,T')}, \varphi_E^{(\ell,T')})$ is independent of $(\Phi^{(0)}, \varphi_E^{(0)})$ and $(\Phi^{(\ell',t)}, \varphi_E^{(\ell',t)})$ for all $\ell' \ne \ell$.

2.6.2 Selecting Problem Instances Yielding the Forbidden Structure

We set $\beta_-, \beta_+, \varepsilon$ as in Subsection 2.3.4. Recall that these numbers depend only on κ, and $\frac{\beta+\varepsilon}{1-\beta e^{-(\beta-1)}} \le \kappa$ for all $\beta \in [\beta_-, \beta_+]$. Moreover, set $\beta_0 = \frac{\beta_- + \beta_+}{2}$.

In the proof of Theorem 2.1.6, we selected the problem instances $\Phi^{(t_0)}, \Phi^{(t_1)}, \dots, \Phi^{(t_k)}$ where the algorithm outputs form a forbidden structure after observing the entire interpolation. Here, because we can leverage concentration properties of local algorithms (instead of stability properties of low degree polynomials), we know in advance which problem instances to choose. This allows us to immediately restrict our attention to $k+1$ problem instances, instead of the full interpolation structure.

We will choose the problem instances $\Phi^{(0)}$ and $\Phi^{(\ell,t_\ell)}$ for $1 \le \ell \le k$, for indices $t_1, \dots, t_k$ we now determine. Consider random variables (which depend on the t_ℓ)

$$x^{(0)} = \mathcal{A}(\Phi^{(0)}, \varphi_E^{(0)}) \qquad \text{and} \qquad x^{(\ell)} = \mathcal{A}(\Phi^{(\ell,t_\ell)}, \varphi_E^{(\ell,t_\ell)})$$

for $1 \le \ell \le k$. Here, $\mathcal{A}(\Phi, \varphi_E)$ denotes $\mathcal{A}$ run with input Φ and internal randomness φ_E (we suppress the dependence on φ_V, which is fixed).

We inductively define t_ℓ in terms of $t_1, \dots, t_{\ell-1}$ as the smallest number satisfying $1 \le t_\ell \le T'$ and

$$\mathbb{E}\, H\left(\pi(x^{(\ell)} | x^{(0)}, \dots, x^{(\ell-1)})\right) \ge \beta_0 \frac{\log k}{k}.$$

If no such t_ℓ exists, set $t_\ell = T'$ and say ℓ is *deficient*. Note that the t_ℓ are a deterministic function of $\mathcal{A}$ and φ_V. For $1 \le \ell \le k$, define $\Phi^{(\ell)} = \Phi^{(\ell,t_\ell)}$ and $\varphi_E^{(\ell)} = \varphi_E^{(\ell,t_\ell)}$. Let $(G^{(\ell)}, \rho^{(\ell)})$ be the factor graph of $\Phi^{(\ell)}$.

The following lemma follows from the stability of the expected conditional overlap entropy when we resample one literal of $\Phi^{(\ell,t)}$. We defer its proof to Subsection 2.6.6. The exponent $-\frac{1}{2}$ can be replaced by any constant larger than -1.

Lemma 2.6.2. *If $\ell \in [k]$ is not deficient, then*

$$\beta_0 \frac{\log k}{k} \le \mathbb{E}\, H\left(\pi(x^{(\ell)} | x^{(0)}, \dots, x^{(\ell-1)})\right) \le \beta_0 \frac{\log k}{k} + O(N^{-1/2}).$$

2.6.3 Outline of the Proof

For the rest of this proof, take $\eta = \frac{\beta_+ - \beta_-}{32k^2}$ and $\nu = \frac{1}{k^2 2^k}$. We now define events $S_{\mathrm{valid}}, S_{\mathrm{conc}}, S_{\mathrm{indep}}, S_{\mathrm{ogp}}$, which are measurable in the $(\Phi^{(\ell)}, \varphi_E^{(\ell)})$ for $0 \le \ell \le k$.

Let S_{valid} be the event that for all $0 \le \ell \le k$, $x^{(\ell)}$ (η, ν)-satisfies $\Phi^{(\ell)}$. Let S_{conc} be the event that for all $1 \le \ell \le k$,

$$\left| H\left(\pi(x^{(\ell)}|x^{(0)}, \dots, x^{(\ell-1)})\right) - \mathbb{E}\, H\left(\pi(x^{(\ell)}|x^{(0)}, \dots, x^{(\ell-1)})\right)\right| \le \frac{1}{\log N}.$$

For $1 \le \ell \le k$, let S^{ℓ}_{indep} be the event that there does not exist an assignment $y \in \{\mathtt{T}, \mathtt{F}\}^N$ such that

(IND-A) y ν-satisfies $\Phi^{(\ell)}$;

(IND-B) $H\left(\pi(y|x^{(0)}, \dots, x^{(\ell-1)})\right) \le \beta_+ \frac{\log k}{k}$.

Let S_{ogp} be the event that there does not exist assignments $y^{(0)}, \dots, y^{(k)} \in \{\mathtt{T}, \mathtt{F}\}^N$ such that

(OGP-A) For all $0 \le \ell \le k$, $x^{(\ell)}$ ν-satisfies $\Phi^{(\ell)}$;

(OGP-B) For all $1 \le \ell \le k$, $H\left(\pi(y^{(\ell)}|y^{(0)}, \dots, y^{(\ell-1)})\right) \in \left[\beta_- \frac{\log k}{k}, \beta_+ \frac{\log k}{k}\right]$.

Finally, define

$$p = \mathbb{P}\left[\mathcal{A}(\Phi, \varphi_E) \text{ } (\eta, \nu)\text{-satisfies } \Phi\right],$$

where the probability is over the randomness of $\Phi \sim \Phi_k(N, M)$ and $\varphi_E \sim (\Omega, \mathbb{P}_\omega)^{\otimes E_G}$, where (G, ρ) is the factor graph of Φ. This is the probability upper bounded by Theorem 2.1.13.

We will derive Theorem 2.1.13 from the following two propositions.

Proposition 2.6.3. *For all sufficiently large k and N, the following relations hold.*

(a) If $\ell \in [k]$ is deficient, then $S_{\text{valid}} \cap S_{\text{consec}} \cap S^{\ell}_{\text{indep}} = \emptyset$.

(b) If no $\ell \in [k]$ is deficient, then $S_{\text{valid}} \cap S_{\text{consec}} \cap S_{\text{ogp}} = \emptyset$.

Proposition 2.6.4. *For all sufficiently large k and N, the following inequalities hold.*

(a) $\mathbb{P}(S_{\text{valid}}) \ge p^{k+1}$.

(b) $\mathbb{P}(S_{\text{conc}}) \ge 1 - \exp(-\widetilde{\Omega}(N^{1/3}))$.

(c) If $\ell \in [k]$ is deficient, then $\mathbb{P}(S^{\ell}_{\text{indep}}) \ge 1 - \exp(-\Omega(N))$.

(d) If no $\ell \in [k]$ is deficient, then $\mathbb{P}(S_{\text{ogp}}) \ge 1 - \exp(-\Omega(N))$.

We will prove Proposition 2.6.3 in Subsection 2.6.4, Proposition 2.6.4(a) in Subsection 2.6.5, and Proposition 2.6.4(b) in Subsection 2.6.6. Proposition 2.6.4(c,d) are analogous to Proposition 2.3.7(b,c). The proofs are exactly the same, except we no longer need to union bound over all possible choices of the t_ℓ.

First, let us see how these bounds imply Theorem 2.1.13.

Proof of Theorem 2.1.13. Set k^* such that for all $k \ge k^*$, Propositions 2.6.3 and 2.6.4 both hold.

Suppose some $\ell \in [k]$ is deficient. By Proposition 2.6.3(a), $\mathbb{P}(S_{\text{valid}}) + \mathbb{P}(S_{\text{conc}}) + \mathbb{P}(S^{\ell}_{\text{indep}}) \le 2$. By Proposition 2.6.4, this implies

$$p^{k+1} \le \exp(-\widetilde{\Omega}(N^{1/3})) + \exp(-\Omega(N)),$$

whence $p \le \exp(-\widetilde{\Omega}(N^{1/3}))$. If no $\ell \in [k]$ is deficient, then by Proposition 2.6.3(b), $\mathbb{P}(S_{\text{valid}}) + \mathbb{P}(S_{\text{conc}}) + \mathbb{P}(S_{\text{ogp}}) \le 2$. By Proposition 2.6.4, we get the same conclusion. So, $p \le \exp(-\widetilde{\Omega}(N^{1/3}))$ conditionally on any realization of φ_V. □

2.6.4 Successful Algorithm Outputs Contradict S^{ℓ}_{indep} or S_{ogp}

The following corollary to Lemma 2.3.8 is obvious.

Corollary 2.6.5. *Let $\ell \in \mathbb{N}$ be arbitrary and let $x, x', y^{(0)}, \dots, y^{(\ell-1)} \in \{\mathtt{T}, \mathtt{F}\}^N$. If $\Delta(x, x') \le \frac{1}{2}$, then*

$$\left| H\left(\pi(x, y^{(0)}, \dots, y^{(\ell-1)})\right) - H\left(\pi(x', y^{(0)}, \dots, y^{(\ell-1)})\right) \right| \le H(\Delta(x, x')).$$

Proof. Use Fact 2.3.5(b). □

Proof of Proposition 2.6.3. Suppose S_{valid} and S_{conc} both hold. We will construct an example of the structure forbidden by S^{ℓ}_{indep} or S_{ogp}. Since S_{valid} holds, there exists $y^{(0)}, \dots, y^{(\ell)}$ such that for all $0 \le \ell \le k$, $\Delta(x^{(\ell)}, y^{(\ell)}) \le \eta$ and $y^{(\ell)}$ ν-satisfies $\Phi^{(\ell)}$.

Suppose $\ell \in [k]$ is deficient. Then, $\mathbb{E}\, H\left(\pi(x^{(\ell)} | x^{(0)}, \dots, x^{(\ell-1)})\right) \le \beta_0 \frac{\log k}{k}$ by definition. By Lemma 2.3.8,

$$\left| H\left(\pi(y^{(\ell)} | x^{(0)}, \dots, x^{(\ell-1)})\right) - H\left(\pi(x^{(\ell)} | x^{(0)}, \dots, x^{(\ell-1)})\right) \right| \le H(\eta) \le \frac{\beta_+ - \beta_-}{8} \cdot \frac{\log k}{k^2}$$

for sufficiently large k, using the bound $H(x) \le x \log \frac{e}{x}$. In tandem with S_{conc}, this implies

$$H\left(\pi(y^{(\ell)} | x^{(0)}, \dots, x^{(\ell-1)})\right) \le \beta_0 \frac{\log k}{k} + \frac{\beta_+ - \beta_-}{8} \cdot \frac{\log k}{k^2} + \frac{1}{\log N} \le \beta_+ \frac{\log k}{k}$$

for sufficiently large N. This is an example of the structure forbidden by S^{ℓ}_{indep}. This proves part (a).

Otherwise, suppose no $\ell \in [k]$ is deficient. Writing (by Fact 2.3.5(b))

$$H\left(\pi(x^{(\ell)} | x^{(0)}, \dots, x^{(\ell-1)})\right) = H\left(\pi(x^{(0)}, \dots, x^{(\ell)})\right) - H\left(\pi(x^{(0)}, \dots, x^{(\ell-1)})\right)$$

and applying Corollary 2.6.5 repeatedly, we have, for all $1 \le \ell \le k$,

$$\left| H\left(\pi(x^{(\ell)} | x^{(0)}, \dots, x^{(\ell-1)})\right) - H\left(\pi(y^{(\ell)} | y^{(0)}, \dots, y^{(\ell-1)})\right) \right| \le 2(k+1) H(\eta) \le \frac{\beta_+ - \beta_-}{4} \cdot \frac{\log k}{k}$$

for sufficiently large k. By Lemma 2.6.2 and S_{conc}, this implies

$$\left| H\left(\pi(y^{(\ell)} | y^{(0)}, \dots, y^{(\ell-1)})\right) - \beta_0 \frac{\log k}{k} \right| \le \frac{\beta_+ - \beta_-}{4} \cdot \frac{\log k}{k} + \frac{1}{\log N} + O(N^{-1/2}) \le \frac{\beta_+ - \beta_-}{2} \cdot \frac{\log k}{k}$$

for sufficiently large N. So, $H\left(\pi(y^{(\ell)} | y^{(0)}, \dots, y^{(\ell-1)})\right) \in \left[\beta_- \frac{\log k}{k}, \beta_+ \frac{\log k}{k}\right]$ for all $1 \le \ell \le k$. This is an example of the structure forbidden by S_{ogp}. This proves part (b). □

2.6.5 Lower Bound on the All-Success Probability

Consider the random variable $\Psi = (\Phi, \varphi_E)$, for $\Phi \sim \Phi_k(N, M)$ and $\varphi_E \sim (\Omega, \mathbb{P}_\omega)^{\otimes E_G}$ where (G, ρ) is the factor graph of Φ. In this and the next subsection, the following representation of Ψ as a sequence of kM i.i.d. random variables will be useful. We can reformat $\Psi = (\psi_j)_{1 \le j \le kM}$, where $\psi_j = (\Phi_{L(j)}, \varphi_E(e))$ and e is the edge in G corresponding to $\Phi_{L(j)}$. Each ψ_j is an i.i.d. sample from $\Upsilon = \text{unif}(\mathcal{L}) \times (\Omega, \mathbb{P}_\omega)$.

For $0 \le \ell \le k$, let $\Psi^{(\ell)} = (\Phi^{(\ell)}, \varphi_E^{(\ell)})$, which is marginally distributed as Ψ. We similarly can reformat $\Psi^{(\ell)} = (\psi_j^{(\ell)})_{1 \le j \le kM}$.

Proof of Proposition 2.6.4(a). For Ψ as above, let

$$f(\Psi) = \mathbb{1}\left\{\mathcal{A}(\Phi, \varphi_E) \text{ } (\eta, \nu)\text{-satisfies } \Phi\right\}.$$

Note that $\mathbb{E}\, f(\Psi) = p$ by definition, and

$$\mathbb{P}(S_{\text{valid}}) = \mathbb{E}\left[\prod_{\ell=0}^{k} f(\Psi^{(\ell)})\right].$$

We wish to show this expectation is at least p^{k+1}.

The constituent random variables $\psi_j^{(\ell)}$ of $\Psi^{(\ell)} = (\psi_j^{(\ell)})_{1\le j\le kM}$ have the following stochastic structure. For each $1 \le \ell \le k$, the last $kM - t_\ell$ variables $(\psi_j^{(\ell)})_{t_\ell < j \le kM}$ in $\Psi^{(\ell)}$ are identical to the corresponding variables in $\Psi^{(0)}$, and the first t_ℓ variables $(\psi_j^{(\ell)})_{1\le j\le t_\ell}$ are fresh i.i.d. draws from Υ.

Let $\tau : [k] \to [k]$ be a permutation such that $t_{\tau(1)} \le t_{\tau(2)} \le \cdots \le t_{\tau(k)}$, and let $s_\ell = t_{\tau(\ell)}$. Then, $\Psi^{(0)}, \dots, \Psi^{(k)}$ all share their last $kM - s_k$ variables $\psi_j^{(\ell)}$; all but $\Psi^{(\tau(k))}$ share the next $s_k - s_{k-1}$ variables; all but $\Psi^{(\tau(k))}$ and $\Psi^{(\tau(k-1))}$ share the next $s_{k-1} - s_{k-2}$ variables, and so on.

For $0 \le i \le k$, let ξ_i be a sequence of $s_{i+1} - s_i$ i.i.d. draws from Υ, where $s_0 = 0$ and $s_{k+1} = kM$. Let $\xi_i^0, \xi_i^1, \dots$ be a sequence of i.i.d. copies of ξ_i. By the above discussion, we can generate $\Psi^{(0)}, \dots, \Psi^{(k)}$ by generating $\Psi^{(0)} = (\xi_i^0)_{i=0}^k$, and for $1 \le \ell \le k$, generating $\Psi^{(\tau(\ell))} = (\xi_i^{(\ell-i)_+})_{i=0}^k$. For example, when $k = 3$,

$$\begin{aligned}
\Psi^{(0)} &= (\xi_0^0, \xi_1^0, \xi_2^0, \xi_3^0),\\
\Psi^{(\tau(1))} &= (\xi_0^1, \xi_1^0, \xi_2^0, \xi_3^0),\\
\Psi^{(\tau(2))} &= (\xi_0^2, \xi_1^1, \xi_2^0, \xi_3^0),\\
\Psi^{(\tau(3))} &= (\xi_0^3, \xi_1^2, \xi_2^1, \xi_3^0).
\end{aligned}$$

Let $f(\xi_0, \dots, \xi_k)$ denote $f(\Psi)$, for the Ψ that can be formatted (by the above discussion) as $(\xi_0, \dots, \xi_k)$. Let $f_0 = f$, and for $0 \le d \le k$, define

$$f_{d+1}(\xi_{d+1}, \dots, \xi_k) = \mathop{\mathbb{E}}_{\xi_d} f_d(\xi_d, \dots, \xi_k).$$

Note that f_{k+1} takes no inputs and outputs p. Further, for $0 \le d \le k+1$ define

$$P_d = \mathbb{E}\left[\prod_{\ell=0}^{k} f_d\left((\xi_i^{(\ell-i)_+})_{i=d}^k\right)\right].$$

In particular $P_0 = \mathbb{P}(S_{\text{valid}})$ and $P_{k+1} = p^{k+1}$. To finish the proof we will show that $P_d \ge P_{d+1}$ for all $0 \le d \le k$. By Jensen's inequality,

$$\begin{aligned}
P_d &= \mathbb{E}\left[f_d\left((\xi_i^0)_{i=d}^k\right)^{d+1} \prod_{\ell=d+1}^{k} f_d\left((\xi_i^{(\ell-i)_+})_{i=d}^k\right)\right]\\
&= \mathbb{E}\left[\mathop{\mathbb{E}}_{\xi_d^0}\left[f_d\left((\xi_i^0)_{i=d}^k\right)^{d+1}\right] \prod_{\ell=d+1}^{k} \mathop{\mathbb{E}}_{\xi_d^{\ell-d}}\left[f_d\left((\xi_i^{(\ell-i)_+})_{i=d}^k\right)\right]\right]\\
&\ge \mathbb{E}\left[\mathop{\mathbb{E}}_{\xi_d^0}\left[f_d\left((\xi_i^0)_{i=d}^k\right)\right]^{d+1} \prod_{\ell=d+1}^{k} \mathop{\mathbb{E}}_{\xi_d^{\ell-d}}\left[f_d\left((\xi_i^{(\ell-i)_+})_{i=d}^k\right)\right]\right]\\
&= \mathbb{E}\left[f_{d+1}\left((\xi_i^0)_{i=d+1}^k\right)^{d+1} \prod_{\ell=d+1}^{k} f_{d+1}\left((\xi_i^{(\ell-i)_+})_{i=d+1}^k\right)\right]\\
&= P_{d+1}.
\end{aligned}$$

□

2.6.6 Bounded Differences and Concentration of Local Algorithms

We will use the following variant of McDiarmid's inequality, which allows a bad event on which bounded differences are large.

Lemma 2.6.6 ([Kut02, Theorem 3.3])**.** *Let $m \in \mathbb{N}$. Let $\Omega_1, \dots, \Omega_m$ be probability spaces and $\Omega = \prod_{i=1}^m \Omega_i$. Let $S \subset \Omega$ and $f : \Omega \to \mathbb{R}$ have the following properties.*

(i) If $X, X' \in S$ differ in coordinate i, then $|f(X) - f(X')| \le c_i$.

(ii) If $X, X' \in \Omega$ differ in coordinate i, then $|f(X) - f(X')| \le b_i$.

Then,

$$\mathop{\mathbb{P}}_{X\sim\Omega}[|f(X) - \mathbb{E}\, f(X)| \ge t] \le 2\exp\left(-\frac{t^2}{8\sum_{i=1}^m c_i^2}\right) + 2\,\mathbb{P}(S^c)\sum_{i=1}^m \frac{b_i}{c_i}.$$

The following definition gives the complement of the bad event we will use. The exponent $\frac{1}{3}$ is chosen to minimize the failure probability in Lemma 2.6.6 by balancing the two terms.

Definition 2.6.7. A k-SAT formula $\Phi \in \Omega_k(N, M)$ is *r-locally small* if, for (G, ρ) the factor graph of Φ, $|N_r(v, G)| \le N^{1/3}$ for all $v \in \mathrm{Va}_G$.

Fact 2.6.8. *If $\Phi \sim \Phi_k(N, M)$ and $r \in \mathbb{N}$ is constant, Φ is r-locally small with probability $1 - \exp(-\Omega(N^{1/3}))$.*

Proof. This follows from Lemma 2.7.5 and a union bound. Note that the r-neighborhood of any $v \in \mathrm{Cl}_G$ is contained in the $(r+1)$-neighborhood of any of its neighbors. □

Proof of Proposition 2.6.4(b). We present the argument for $\ell = k$; showing concentration for the other conditional overlap entropies is similar. The random variable $Y = H\left(\pi(x^{(k)}|x^{(0)}, \dots, x^{(k-1)})\right)$ is measurable in $\Psi^{(0)}, \dots, \Psi^{(k)}$. For each $0 \le \ell \le k$, we can write $\Psi^{(\ell)} = (\psi_j^{(\ell)})_{1\le j\le kM}$. The constituent random variables $\psi_j^{(\ell)}$ can be partitioned into equivalence classes, where variables in the same equivalence class are identical and different equivalence classes are mutually independent. Let $\zeta = (\psi_1, \dots, \psi_m)$ contain one representative from each equivalence class. Note that $\psi_1, \dots, \psi_m$ i.i.d. samples from Υ and $m \le k^2 M$.

All the $\Phi^{(\ell)}$ $(G^{(\ell)}, \rho^{(\ell)}, \varphi_E^{(\ell)})$, $x^{(\ell)}$, overlap profiles of the $x^{(\ell)}$, and Y are ζ-measurable. We will use $Y(\zeta)$ to denote the Y given by this realization of ζ, and similarly for the remaining random variables.

Let $S \subseteq \Upsilon^m$ be the event that $\Phi^{(\ell)}(\zeta)$ is r-locally small for all $0 \le \ell \le k$. By a union bound on Fact 2.6.8, $\mathbb{P}(S^c) \le \exp(-\Omega(N^{1/3}))$.

Suppose $\zeta, \zeta' \in S$ differ in only one coordinate ψ_i. For each $0 \le \ell \le k$, the decorated factor graphs $(G^{(\ell)}, \rho^{(\ell)}, \varphi_E^{(\ell)})(\zeta)$ and $(G^{(\ell)}, \rho^{(\ell)}, \varphi_E^{(\ell)})(\zeta')$ differ in at most one edge. Because $\zeta, \zeta' \in S$ and $\mathcal{A}$ is local, $x^{(\ell)}(\zeta)$ and $x^{(\ell)}(\zeta')$ differ in $O(N^{1/3})$ bits. So, corresponding entries in $\pi(x^{(0)}, \dots, x^{(k)})(\zeta)$ and $\pi(x^{(0)}, \dots, x^{(k)})(\zeta')$ differ by $O(N^{-2/3})$. Thus, $|Y(\zeta) - Y(\zeta')| \le O(N^{-2/3}\log N)$.

Moreover, for any $\zeta, \zeta' \in \Upsilon^m$, we have $|Y(\zeta) - Y(\zeta')| \le \log 2$ because the conditional overlap entropy attains values in $[0, \log 2]$. By Lemma 2.6.6,

$$\begin{aligned}\mathbb{P}\left[|Y - \mathbb{E}\, Y| \ge \frac{1}{\log N}\right] &\le 2\exp\left(-\frac{1/\log^2 N}{8k^2 M O((N^{-2/3}\log N)^2)}\right) + \exp(-\Omega(N^{1/3}))O\left(\frac{2k^2 M \log 2}{N^{-2/3}\log N}\right)\\ &\le \exp(-\widetilde{\Omega}(N^{1/3})).\end{aligned}$$

□

Similar ideas prove Lemma 2.6.2.

Proof of Lemma 2.6.2. The lower bound follows from the definition of deficient. For $1 \le t \le T$, let $x^{(\ell,t)} = \mathcal{A}(\Phi^{(\ell,t)}, \varphi_E^{(\ell,t)})$. We will show that

$$\left|\mathbb{E}\, H(\pi(x^{(\ell,t)}|x^{(0)}, \dots, x^{(\ell-1)})) - \mathbb{E}\, H(\pi(x^{(\ell,t-1)}|x^{(0)}, \dots, x^{(\ell-1)}))\right| \le O(N^{-1/2}).$$

Since $\mathbb{E}\, H(\pi(x^{(\ell,t_\ell-1)}|x^{(0)}, \dots, x^{(\ell-1)})) < \beta_0 \frac{\log k}{k}$, the above inequality implies the result.

Let S be the event that $\Phi^{(\ell,t-1)}$ and $\Phi^{(\ell,t)}$ are both r-locally small. By Fact 2.6.8 and a union bound, $\mathbb{P}(S^c) \le \exp(-\Omega(N^{1/3}))$. The decorated factor graphs $(G^{(\ell,t-1)}, \rho^{(\ell,t-1)}, \varphi_E^{(\ell,t-1)})$ and $(G^{(\ell,t)}, \rho^{(\ell,t)}, \varphi_E^{(\ell,t)})$ differ in at most one edge. On the event S, $x^{(\ell,t)}$ and $x^{(\ell,t-1)}$ differ in at most $O(N^{1/3})$ bits, and by Lemma 2.3.8,

$$\left|H(\pi(x^{(\ell,t)}|x^{(0)}, \dots, x^{(\ell-1)})) - H(\pi(x^{(\ell,t-1)}|x^{(0)}, \dots, x^{(\ell-1)}))\right| \le H(\Delta(x^{(\ell,t)}, x^{(\ell,t-1)})) \le O(N^{-2/3}\log N).$$

Moreover, this difference is always at most $\log 2$. Thus

$$\begin{aligned}\left|\mathbb{E}\, H(\pi(x^{(\ell,t)}|x^{(0)},\dots,x^{(\ell-1)})) - \mathbb{E}\, H(\pi(x^{(\ell,t-1)}|x^{(0)},\dots,x^{(\ell-1)}))\right| &\le O(N^{-2/3}\log N) + \mathbb{P}(S^c)\log 2\\ &\le O(N^{-1/2}).\end{aligned}$$

□

2.7 Simulation of Local Memory Algorithms

In this section, we introduce the class of *local memory algorithms*. These algorithms are a natural generalization of local algorithms, which make local decisions in series (in a random vertex order) and allow earlier decisions to leave local information that later decisions can see. This class includes the first phase of `Fix`, as well as the sequential local algorithms considered in [GS17]. We show, somewhat surprisingly, that any local memory algorithm can be simulated by a local algorithm of larger radius. We then show that any local algorithm can be simulated by a constant degree polynomial.

The main results of this section are the following two propositions. Throughout this section, fix arbitrary $\alpha = \alpha(k)$ independent of N and let $M = \lfloor \alpha N \rfloor$.

Proposition 2.7.1 (Local algorithms simulate local memory algorithms)**.** *Suppose $\alpha k, k-1 \ge 2$ and $\eta > 0$. Let $\mathcal{A}$ be an r-local memory algorithm (defined in Definition 2.7.12) with output in $\{\mathtt{T},\mathtt{F}\}^N$. There exists $R \in \mathbb{N}$ depending on α, k, r, η and an R-local algorithm $\mathcal{A}'$ such that, for some coupling of the internal randomnesses of $\mathcal{A}, \mathcal{A}'$,*

$$\mathbb{P}\left[\Delta(\mathcal{A}(\Phi), \mathcal{A}'(\Phi)) \ge \eta\right] \le \exp(-\Omega(N^{1/3})),$$

where the probability is over $\Phi \sim \Phi_k(N, M)$ and the randomnesses of $\mathcal{A}, \mathcal{A}'$.

We parse the outputs of a low degree polynomial with the function $\mathtt{strictRound} : \mathbb{R} \to \{\mathtt{T},\mathtt{F},\mathtt{err}\}$, defined by

$$\mathtt{strictRound}(x) = \begin{cases} \mathtt{T} & x = 1,\\ \mathtt{F} & x = -1,\\ \mathtt{err} & \text{otherwise.}\end{cases}$$

When applied to a real-valued vector, `strictRound` is applied coordinate-wise. Note that this is a more stringent parsing scheme than `round`. Let $n = M \cdot k \cdot 2N$. Recall that each $\Phi \in \Omega_k(N, M)$ can be identified with a vector in $\{0,1\}^n$, as described below Definition 2.1.2.

Proposition 2.7.2 (Low degree polynomials simulate local algorithms)**.** *Suppose $\alpha k, k-1 \ge 2$ and $\eta > 0$. Let $\mathcal{A}$ be an r-local algorithm with output in $\{\mathtt{T},\mathtt{F}\}^N$. There exist $D, \gamma > 0$ depending on α, k, r, η and a (random) degree-D polynomial $f : \mathbb{R}^n \times \Omega \to \mathbb{R}^N$ such that the following holds. Let $\mathcal{A}' = \mathtt{strictRound} \circ f$. For some coupling of the internal randomnesses of $\mathcal{A}$ and f,*

$$\mathbb{P}\left[\Delta(\mathcal{A}(\Phi), \mathcal{A}'(\Phi)) \ge \eta\right] \le \exp(-\Omega(N^{1/3})),$$

where the probability is over $\Phi \sim \Phi_k(N, M)$ and the randomnesses of $\mathcal{A}, \mathcal{A}'$. Moreover, $\mathbb{E}_{\Phi,\omega} \|f(\Phi,\omega)\|_2^2 \le \gamma N$.

Both simulation results incur an error tolerance η independent of N which can be made arbitrarily small in k and fail with probability only $\exp(-\Omega(N^{1/3}))$.

These simulation results imply that our hardness theorems, Theorems 2.1.6 and 2.1.13, apply to any local memory algorithm. We will also use these results in Section 2.8 with the fact that the first phase of `Fix` is a local memory algorithm to show that local algorithms and low degree polynomials solve random k-SAT at clause density $\alpha = (1-\varepsilon)2^k \log k / k$.

This section is structured as follows. In Subsection 2.7.1 we review properties of local algorithms and the k-SAT factor graph. In Subsection 2.7.2 we define local memory algorithms. In Subsection 2.7.3 we prove Proposition 2.7.1, and in Subsection 2.7.4 we prove Proposition 2.7.2. Subsection 2.7.5 contains deferred proofs.

2.7.1 Properties of Local Algorithms and the k-SAT Factor Graph

Throughout this section, fix a probability space $(\Omega, \mathbb{P}_\omega)$. Let $\mathrm{DFG}(N, M, k, (\Omega, \mathbb{P}_\omega))$ denote the law of the decorated random k-SAT factor graph (G, ρ, φ), where $\Phi \sim \Phi_k(N, M)$, (G, ρ) is the factor graph of Φ, and $\varphi \sim (\Omega, \mathbb{P}_\omega)^{\otimes(V_G \cup E_G)}$. We write this as $\mathrm{DFG}(N, M, k)$ when $(\Omega, \mathbb{P}_\omega)$ is unambiguous.

Equivalently, $(G, \rho, \varphi) \sim \mathrm{DFG}(N, M, k)$ can be sampled as follows. $\mathrm{Va}_G = \{v_1, \dots, v_N\}$ and $\mathrm{Cl}_G = \{c_1, \dots, c_M\}$ are fixed. E_G consists of k edges from each $c \in \mathrm{Cl}_G$ to i.i.d. uniformly random vertices in Va_G, and ρ, φ are sampled by $\rho \sim \mathrm{unif}(\{\mathtt{T}, \mathtt{F}\})^{\otimes E_G}$, $\varphi \sim (\Omega, \mathbb{P}_\omega)^{\otimes(V_G \cup E_G)}$.

A (possibly infinite) graph is *locally finite* if every vertex has finite degree. The formalism in Definitions 2.1.9, 2.1.10, and 2.1.11 applies verbatim to locally finite G. The local geometry of a sample from $\mathrm{DFG}(N, M, k)$ can be understood in analogy to the following locally finite tree.

Definition 2.7.3 (Decorated Alternating Galton-Watson Tree)**.** Let $d_1 > 0$, $d_2 \in \mathbb{N}$. Let $\mathrm{DGW}(d_1, d_2, (\Omega, \mathbb{P}_\omega))$ denote the law of the following rooted decorated tree (o, T, ρ, φ). The rooted tree (o, T) is sampled by the following procedure.

- Start with a root vertex o in layer 0.
- For $\ell \geq 1$:
  - If ℓ is even, each vertex in layer ℓ independently spawns $\mathrm{Pois}(d_1)$ children in layer $\ell + 1$.
  - If ℓ is odd, each vertex in layer ℓ spawns d_2 children in layer $\ell + 1$.

  Each non-root vertex is connected to its parent by an edge.

Let Va_T and Cl_T be the sets of even and odd depth vertices of T. Further, let $V_T = \mathrm{Va}_T \cup \mathrm{Cl}_T$ and let E_T be the edge set of T. Sample $\rho \sim \mathrm{unif}(\{\mathtt{T}, \mathtt{F}\})^{\otimes E_T}$ and $\varphi \sim (\Omega, \mathbb{P}_\omega)^{\otimes(V_T \cup E_T)}$.

When $(\Omega, \mathbb{P}_\omega)$ is unambiguous, we write this as $\mathrm{DGW}(d_1, d_2)$. The significance of this tree is that as $N \to \infty$, local neighborhoods $N_r(v, G, \rho, \varphi)$ of a sample $(G, \rho, \varphi) \sim \mathrm{DFG}(N, M, k)$, where $v \in \mathrm{Va}_G$ is fixed, converge weakly to local neighborhoods of the root of $\mathrm{DGW}(\alpha k, k - 1)$. This is analogous to the fact that local neighborhoods of the sparse Erdős-Rényi graph $G(N, d/N)$ converge weakly to local neighborhoods of the root of the Poisson Galton-Watson tree $\mathrm{PGW}(d)$.

We now state several lemmas pertaining to local geometry of samples from $\mathrm{DGW}(d_1, d_2)$ and $\mathrm{DFG}(N, M, k)$. Lemmas 2.7.4 and 2.7.5 control the local neighborhood sizes of the root of $\mathrm{DGW}(d_1, d_2)$ and of a left-vertex in $\mathrm{DFG}(N, M, k)$. Lemma 2.7.6 makes precise the sense in which local neighborhoods of left-vertices of $\mathrm{DFG}(n, m, k)$ converge to local neighborhoods of the root of $\mathrm{DGW}(\alpha k, k - 1)$. Lemma 2.7.7 shows concentration for the sum of a local function. These lemmas are analogous to [BCN20, Lemma 11.1, Lemma 11.2, Lemma 12.4, Proposition 12.3], which give the analogous results with $\mathcal{G}(N, M, k)$ and $\mathrm{DGW}(\alpha k, k - 1)$ replaced by $G(n, d/n)$ and $\mathrm{PGW}(d)$ (and without the decorations ρ, φ, which do not affect the results). We omit their proofs, which are easily adapted from the corresponding proofs of [BCN20].

Lemma 2.7.4. *Let $d_1, d_2 \geq 2$ and $(o, T, \rho, \varphi) \sim \mathrm{DGW}(d_1, d_2)$. There are universal constants $c_0, c_1 > 0$ such that for all $\lambda > 0$,*

$$\mathbb{P}\left[|N_{2r}(o, T)| \leq \lambda (d_1 d_2)^r \text{ for all positive integers } r\right] \geq 1 - c_1 e^{-c_0 \lambda}.$$

Lemma 2.7.5. *Let $\alpha k, k - 1 \geq 2$. Let $(G, \rho, \varphi) \sim \mathrm{DFG}(N, M, k)$, and let $v \in \mathrm{Va}_G$ be fixed. There are universal constants $c_0, c_1 > 0$ such that for all $\lambda > 0$,*

$$\mathbb{P}\left[|N_{2r}(v, G)| \leq \lambda (\alpha k (k - 1))^r \text{ for all positive integers } r\right] \geq 1 - c_1 e^{-c_0 \lambda}.$$

Recall that Λ is the set of (possibly infinite, locally finite) rooted decorated bipartite graphs.

Lemma 2.7.6. *Let $\alpha k, k - 1 \geq 2$. Let $(o, T, \rho, \varphi) \sim \mathrm{DGW}(\alpha k, k - 1)$, $(G, \rho', \varphi') \sim \mathrm{DFG}(N, M, k)$, and let $v \in \mathrm{Va}_G$ be fixed. Let $f : \Lambda \to [-1, 1]$ be a $2r$-local function. There exists $c > 0$ (depending on α, k, r) such that for all N,*

$$|\mathbb{E}\, f(o, T, \rho, \varphi) - \mathbb{E}\, f(v, G, \rho', \varphi')| \leq \frac{c \log N}{N^{1/2}}.$$

Lemma 2.7.7. *Let $\alpha k, k-1 \geq 2$, and let $(G,\rho,\varphi) \sim \mathrm{DFG}(N,M,k)$. Let $f : \Lambda \to [-1,1]$ be a $2r$-local function. There exists $c > 0$ (depending on α, k, r) such that for all $p \geq 2$,*

$$\mathbb{E}\left[\left|\sum_{v\in \mathrm{Va}_G} f(v,G,\rho,\varphi) - \mathbb{E}\sum_{v\in \mathrm{Va}_G} f(v,G,\rho,\varphi)\right|^p\right] \leq \left(cN^{1/2}p^{3/2}\right)^p.$$

We can translate Lemma 2.7.7, into the following tail bound for sums of local functions.

Corollary 2.7.8. *Let $\alpha k, k-1 \geq 2$, and let $(G,\rho,\varphi) \sim \mathrm{DFG}(N,M,k)$. Let $f : \Lambda \to [-1,1]$ be a $2r$-local function. There exists $c > 0$ (depending on α, k, r) such that for all $t \geq cN^{1/2}$,*

$$\mathbb{P}\left[\left|\sum_{v\in \mathrm{Va}_G} f(v,G,\rho,\varphi) - \mathbb{E}\sum_{v\in \mathrm{Va}_G} f(v,G,\rho,\varphi)\right| \geq t\right] \leq \exp\left(-\frac{t^{2/3}}{cN^{1/3}}\right).$$

Proof. Let c be as in Lemma 2.7.7, and suppose $t \geq 2^{3/2}ecN^{1/2}$. Set $p = \left(\frac{t}{ecN^{1/2}}\right)^{2/3} \geq 2$, so by Lemma 2.7.7,

$$\begin{aligned}\mathbb{P}\left[\left|\sum_{v\in \mathrm{Va}_G} f(v,G,\rho,\varphi) - \mathbb{E}\sum_{v\in \mathrm{Va}_G} f(v,G,\rho,\varphi)\right| \geq t\right] &\leq t^{-p}\,\mathbb{E}\left[\left|\sum_{v\in \mathrm{Va}_G} f(v,G,\rho,\varphi) - \mathbb{E}\sum_{v\in \mathrm{Va}_G} f(v,G,\rho,\varphi)\right|^p\right] \\ &\leq \left(\frac{cN^{1/2}p^{3/2}}{t}\right)^p = \exp(-p) \\ &= \exp\left(-\frac{t^{2/3}}{(ec)^{2/3}N^{1/3}}\right).\end{aligned}$$

The result follows by adjusting the constant c. □

2.7.2 Local Memory Algorithms

We now define local memory algorithms. In addition to the usual features of a local algorithm, these algorithms have access to a mutable memory map $\mu : V_G \to \mathbb{Z}_{\geq 0}$, which we think of as an unlimited notepad on each variable. The algorithm processes vertices $v \in V_G$ (both variables and clauses) in a uniformly random order. Each step, the algorithm accesses the r-local neighborhood of a vertex and can overwrite the data written on any vertex in that neighborhood. In the end, each variable $v \in \mathrm{Va}_G$ decides to be true or false depending on the final value $\mu(v)$ on its notepad.

To formalize this algorithm class, we will define memory-augmented versions of Definitions 2.1.9, 2.1.10, and 2.1.11.

Definition 2.7.9 (Rooted memory-augmented decorated bipartite graph)**.** A *memory-augmented decorated bipartite graph* is a tuple (G,ρ,φ,μ), where (G,ρ,φ) is a decorated bipartite graph and μ is a function $\mu : V_G \to \mathbb{Z}_{\geq 0}$. A *rooted memory-augmented decorated bipartite graph* is a tuple (v,G,ρ,φ,μ), where (G,ρ,φ,μ) is a memory-augmented decorated bipartite graph and $v \in V_G$.

Let $\tilde{\Lambda}$ denote the set of rooted memory-augmented decorated bipartite graphs. Two such graphs are isomorphic of there exists a bijection between them preserving $v, \mathrm{Va}_G, \mathrm{Cl}_G, E_G, \rho, \varphi, \mu$.

Definition 2.7.10 (r-neighborhood)**.** Let $(v,G,\rho,\varphi,\mu) \in \tilde{\Lambda}$ and $r \in \mathbb{N}$. Define $N_r(v,G,\rho,\varphi,\mu)$ to be $(v,G',\rho',\varphi',\mu') \in \tilde{\Lambda}$, where $(v,G',\rho',\varphi') = N_r(v,G,\rho,\varphi)$ and $\mu' = \mu\big|_{G'}$ is the restriction of μ to G'.

Definition 2.7.11 (r-local subroutine)**.** An algorithm f with input space $\tilde{\Lambda}$ is an *r-local subroutine* if the execution of $f(v,G,\rho,\varphi)$ depends only on the isomorphism class of $N_r(v,G,\rho,\varphi,\mu) = (v,G',\rho',\varphi',\mu')$, and f interacts with its input by editing the outputs of μ'.

We are now ready to define a local memory algorithm. In the following definition, ψ is an auxiliary random variable on each vertex that determines the order in which vertices are processed.

Definition 2.7.12 (r-local memory algorithm)**.** Let f_1 be an r-local subroutine and $f_2 : \mathbb{Z}_{\geq 0} \to \{\mathtt{T}, \mathtt{F}\}$ be a function. The r-local memory algorithm based on (f_1, f_2), denoted $\mathcal{A}_{f_1,f_2}$, runs as follows on input $\Phi \in \Omega_k(N, M)$ with factor graph (G, ρ).

(1) Initialize $\mu : V_G \to \mathbb{Z}_{\geq 0}$ to the all-0 map. Sample $\varphi \sim (\Omega, \mathbb{P}_\omega)^{\otimes(V_G \cup E_G)}$ and $\psi \sim \mathrm{unif}([0,1])^{\otimes V_G}$.

(2) Loop through vertices $v \in V_G$ (both variables and clauses) in increasing order of $\psi(v)$. For each v, run $f_1(v, G, \rho, \varphi, \mu)$.

(3) Output $x \in \{\mathtt{T}, \mathtt{F}\}^N$ where $x_i = f_2(\mu(v_i))$.

We will see (Fact 2.8.3) that the first phase of `Fix` is in this class. The following variant of the sequential local algorithms in [GS17] is also in this class.

Definition 2.7.13 (Sequential r-local algorithm)**.** Let $f : \Lambda \to [0,1]$ be an r-local function. The sequential r-local algorithm based on f, denoted $\mathcal{B}_f$, runs as follows on input $\Phi \in \Omega_k(N, M)$ with factor graph (G, ρ).

(1) Sample $\varphi \sim (\Omega, \mathbb{P}_\omega)^{\otimes(V_G \cup E_G)}$ and $\psi \sim \mathrm{unif}([0,1])^{\otimes \mathrm{Va}_G}$.

(2) Loop through $v \in \mathrm{Va}_G$ in increasing order of $\psi(v)$. For each $v = v_i$:

  (a) Compute $p = f(v, G, \rho, \varphi)$. Set $x_i = \mathtt{T}$ with probability p, and otherwise $x_i = \mathtt{F}$.

  (b) Simplify Φ by deleting clauses satisfied by this setting of x_i and appearances of x_i in clauses not satisfied by this setting. Furthermore, delete any clauses that become empty (thus not satisfied) as a result of the latter operation.

  (c) Let G' be the corresponding simplification of G, and let $\rho' = \rho\big|_{G'}$ and $\varphi' = \varphi\big|_{G'}$.

  (d) Set $(G, \rho, \varphi) \leftarrow (G', \rho', \varphi')$.

(3) Output $(x_1, \ldots, x_N) \in \{\mathtt{T}, \mathtt{F}\}^N$.

Fact 2.7.14. *For any $r \in \mathbb{N}$, a sequential r-local algorithm can simulated by a $\max(r, 2)$-local memory algorithm.*

Proof. Let $\mathcal{B}_f$ be a sequential r-local algorithm, whose randomness is sampled i.i.d. from $(\Omega, \mathbb{P}_\omega)$. We will construct an r-local memory algorithm $\mathcal{A}_{f_1,f_2}$ simulating $\mathcal{B}_f$.

This algorithm maintains the invariant that for $v = v_i \in \mathrm{Va}_G$, $\mu(v) = 0$ if x_i is not yet set, 1 if x_i is set true, and 2 if x_i is set false. For clause vertices $c \in \mathrm{Cl}_G$, $\mu(c) = 1$ if the clause corresponding to c has been deleted in the simplification, and otherwise $\mu(c) = 0$.

Thus, $\mathcal{A}_{f_1,f_2}$ uses randomness sampled from $(\Omega, \mathbb{P}_\omega) \times \mathrm{unif}([0,1])$. That is, its internal randomness is $\varphi_* = (\varphi, q)$, which is sampled by $\varphi \sim (\Omega, \mathbb{P}_\omega)^{\otimes(V_G \cup E_G)}$ and $q \sim \mathrm{unif}([0,1])^{\otimes(V_G \cup E_G)}$.

The r-local subroutine f_1 runs as follows on input $(v, G, \rho, \varphi_*, \mu)$. If $v \in \mathrm{Cl}_G$, do nothing. Note that the remaining loop over $v \in \mathrm{Va}_G$ runs over these vertices in a uniformly random order, as desired. If $v = v_i \in \mathrm{Va}_G$, let G' be the simplification of G determined by the information recorded in μ, and let $\rho' = \rho\big|_{G'}$, $\varphi' = \varphi\big|_{G'}$. We can simulate the computation of $p = f(v, G', \rho', \varphi')$ because simplification only deletes vertices and edges, so any r-local decision in the simplified factor graph is still r-local in the simulation. We then set $x_i = \mathtt{T}$ if $p < q(v)$, and otherwise $x_i = \mathtt{F}$. We update μ to record this value of x_i and any clause simplifications that result (which is a 2-local operation).

At the end of the algorithm, $\mu(v) \in \{1, 2\}$ for all $v \in \mathrm{Va}_G$. Let $f_2(x) = \mathtt{T}$ if $x = 1$ and $\mathtt{F}$ if $x = 2$. □

Definition 2.7.13 differs slightly from the presentation in [GS17] in the following way. [GS17] studies NAE-k-SAT, in which a clause is satisfied if it contains at least one true and false literal. In partially simplfied formulas of this problem, clauses can exist in four states: "removed," "already contains true," "already contains false," and "contains neither true nor false," and the sequential local algorithms of [GS17] track this information. Of course, we can just as well simulate this by a local memory algorithm by having μ track these clause states.

2.7.3 Local Algorithms Simulate Local Memory Algorithms

In this subsection, we prove Proposition 2.7.1, that any local memory algorithm can be simulated by a local algorithm of larger (but still constant) radius.

The simulation is the natural one: we expand φ to also generate the auxiliary randomness ψ determining the vertex order, and then determine the output at each $v \in \mathrm{Va}_G$ by simulating the local memory algorithm on the R-neighborhood of v. Formally, we expand φ to φ_*, whose outputs are sampled from $(\Omega, \mathbb{P}_\omega) \times \mathrm{unif}([0,1])$. We collect the first coordinates of the outputs into φ and the second coordinates into ψ. (This generates $\psi : V_G \cup E_G \to [0,1]$, and we ignore $\psi\big|_{E_G}$.)

Because sequentiality usually does not create long dependence chains, this simulation will often faithfully capture the local memory algorithm's behavior.

Definition 2.7.15 (R-local simulation)**.** Let $\mathcal{A}_{f_1,f_2}$ be an r-local memory algorithm, with i.i.d. internal randomness from $(\Omega, \mathbb{P}_\omega)$. For $R \in \mathbb{N}$, the R-local simulation of $\mathcal{A}_{f_1,f_2}$ is the R-local algorithm $\mathcal{A}_f$ that runs as follows on input $\Phi \in \Omega_k(N, M)$ with factor graph (G, ρ).

(1) Sample $\varphi_* = (\varphi, \psi)$, where $\varphi \sim (\Omega, \mathbb{P}_\omega)^{\otimes(V_G \cup E_G)}$ and $\psi \sim \mathrm{unif}([0,1])^{\otimes(V_G \cup E_G)}$.

(2) For each $v = v_i \in \mathrm{Va}_G$, set $x_i = f(v, G, \rho, \varphi_*)$. Here $f(v, G, \rho, \varphi_*)$ is the following R-local function.

  (a) Let $N_R(v, G, \rho, \varphi_*) = (v, G', \rho', \varphi'_*)$. Let $\varphi'_* = (\varphi', \psi')$, where $\varphi' = \varphi\big|_{G'}$ and $\psi' = \psi\big|_{G'}$.

  (b) Initialize $\mu : V_{G'} \to \mathbb{Z}_{\geq 0}$ to the all-0 map.

  (c) For $u \in V_{G'}$ in increasing order of $\psi'(u)$, run $f_1(u, G', \rho', \varphi', \mu)$.

  (d) Output $f(v, G, \rho, \varphi_*) = f_2(\mu(v))$.

(3) Output $(x_1, \ldots, x_N)$.

The main idea of the proof of Proposition 2.7.1 is that dependencies caused by sequentiality all arise from the following structure.

Definition 2.7.16 (r-hop ψ-dependence chain)**.** Let G be a locally finite graph and $\psi : V_G \to [0,1]$ be a function. Let $r \in \mathbb{N}$. A sequence $v_1, v_2, \ldots, v_s \in V_G$ is an *r-hop ψ-dependence chain* if consecutive vertices in the sequence are at most distance r apart and $\psi(v_1), \psi(v_2), \ldots, \psi(v_s)$ is decreasing.

We can now define a notion of insulation in terms of these dependence chains. The key point of the following definition is that if in the R-local simulation in Definition 2.7.15, $v \in \mathrm{Va}_G$ is (r, R, ψ)-insulated, then the R-local simulation's output at v must match that of the local memory algorithm run with the same φ, ψ.

Definition 2.7.17 ((r, R, ψ)-insulated)**.** Let G be a locally finite graph and $\psi : V_G \to [0,1]$ be a function. Let $v \in V_G$ and $r, R \in \mathbb{N}$ with $R \geq 2r$. v is *(r, R, ψ)-insulated* if there is no $2r$-hop ψ-dependence chain $v_1, v_2, \ldots, v_s \in V_G$ with $v = v_1$ and $v_s \in N_R(v, G) \setminus N_{R-2r}(v, G)$.

In Definition 2.7.15, if $\Phi \sim \Phi_k(N, M)$, then (G, ρ, φ_*) is a sample from the decorated k-SAT factor graph $\mathrm{DFG}(N, M, k, (\Omega, \mathbb{P}_\omega) \times \mathrm{unif}([0,1]))$. To prove Proposition 2.7.1, it suffices to upper bound the fraction of $v \in \mathrm{Va}_G$ that are not (r, R, ψ)-insulated. To achieve this, we will control the probability that the root of $\mathrm{DGW}(d_1, d_2, (\Omega, \mathbb{P}_\omega) \times \mathrm{unif}([0,1]))$ is not (r, R, ψ)-insulated. Then, because (r, R, ψ)-insulatedness is an R-local property, we can translate this bound to the k-SAT factor graph by the machinery of Lemma 2.7.6 and Corollary 2.7.8.

Proposition 2.7.18. *Let $d_1, d_2 \geq 2$, $r \in \mathbb{N}$, and $\eta \in (0,1)$. Let $(o, T, \rho, \varphi_*) \sim \mathrm{DGW}(d_1, d_2, (\Omega, \mathbb{P}_\omega) \times \mathrm{unif}([0,1]))$, and write $\varphi_* = (\varphi, \psi)$ for $\varphi : V_T \cup E_T \to \Omega$ and $\psi : V_T \cup E_T \to [0,1]$. There exists R dependent on d_1, d_2, r, η such that*

$$\mathbb{P}\left[o \text{ is } (r, R, \psi)\text{-insulated in } T\right] \geq 1 - \eta.$$

The proof of this proposition relies on the following technical lemma, whose proof we defer to Subsection 2.7.5.

Lemma 2.7.19. *Let $d_1, d_2 \geq 2$ and $(o, T, \rho, \varphi) \sim \mathrm{DGW}(d_1, d_2)$. For any $r \in \mathbb{N}$ and $\eta \in (0,1)$, there exist $C, R^* > 0$ depending on d_1, d_2, r, η such that for all integers $R \geq R^*$,*

$$\max_{v \in N_R(o,T)} |N_{2r}(v,T)| \leq \frac{CR}{\log^{(r+1)} R}$$

with probability at least $1-\eta$. Here, $\log^{(r+1)}$ denotes the $(r+1)$th iterate of $\log$.

Proof of Proposition 2.7.18. Let $R \in \mathbb{N}$ be a number we will determine later. Lemma 2.7.19 gives R^* such that if $R \geq R^*$, then the conclusion of Lemma 2.7.19 holds with probability at least $1-\eta/2$. Consider a realization of T, ρ, φ such that this event holds. We will control the probability over ψ that o is not (r, R, ψ)-insulated in T.

If o is not (r, R, ψ)-insulated, there exists a $2r$-hop ψ-dependence chain $o = v_1, v_2, \ldots, v_s \in V_T$ where $v_s \in N_R(o,T) \setminus N_{R-2r}(o,T)$. By taking an initial subsequence, we get a $2r$-hop ψ-dependence chain $o = v_1, v_2, \ldots, v_t \in V_T$ of length $t = \lceil \frac{R}{2r} \rceil$. By Markov's inequality,

$$\mathbb{P}\left[o \text{ is not } (r,R,\psi)\text{-insulated in } T\right] \leq \mathbb{E}\,\#\left(2r\text{-hop } \psi\text{-dependence chains } o = v_1, v_2, \ldots, v_t \in V_T\right).$$

The last expectation is bounded as follows. By Lemma 2.7.19, there are at most $\left(\frac{CR}{\log^{(r+1)} R}\right)^t$ sequences $o = v_1, v_2, \ldots, v_t$ with consecutive vertices at most distance $2r$ apart, and for each one, $\psi(v_1), \psi(v_2), \ldots, \psi(v_t)$ is decreasing with probability $\frac{1}{t!}$. So (using $t! \geq (t/e)^t$) the last expectation is at most

$$\frac{1}{t!}\left(\frac{CR}{\log^{(r+1)} R}\right)^t \leq \left(\frac{eCR}{t \log^{(r+1)} R}\right)^t \leq \left(\frac{2eCr}{\log^{(r+1)} R}\right)^t \leq \eta/2 \tag{2.7.1}$$

for a large enough choice of R. Thus, over the randomness of ψ,

$$\mathbb{P}\left[o \text{ is } (r,R,\psi)\text{-insulated in } T\right] \geq 1 - \eta/2.$$

The result follows by a union bound. □

Unfortunately, due to the last inequality in (2.7.1), the R needed to make this proposition hold is approximately the power tower $\exp^{(r+1)} 2eCr$. This is the R we will need to simulate an r-local memory algorithm by an R-local algorithm. While this R is a constant for any constant r, it would of course be nice to improve this dependence.

Finally, we can prove Proposition 2.7.1.

Proof of Proposition 2.7.1. We let $\mathcal{A}'$ be the R-local simulation of $\mathcal{A}$, for R to be determined. We couple the runs of $\mathcal{A}, \mathcal{A}'$ to use the same φ, ψ. If (G, ρ) is the factor graph of Φ, then

$$\Delta(\mathcal{A}(\Phi), \mathcal{A}'(\Phi)) \leq \frac{1}{N} \sum_{v \in \mathrm{Va}_G} \mathbb{1}\left\{v \text{ is } (r,R,\psi)\text{-insulated in } G\right\}.$$

Recall that $(G, \rho, (\varphi, \psi)) \sim \mathrm{DFG}(N, M, k, (\Omega, \mathbb{P}_\omega) \times \mathrm{unif}([0,1]))$. The last indicator is an R-local function taking values in $[-1,1]$. By Corollary 2.7.8 with $t = \eta N/3$,

$$\frac{1}{N} \sum_{v \in \mathrm{Va}_G} \mathbb{1}\left\{v \text{ is } (r,R,\psi)\text{-insulated in } G\right\} \leq \eta/3 + \mathbb{E}\,\mathbb{1}\left\{v \text{ is } (r,R,\psi)\text{-insulated in } G\right\}$$

with probability $1 - \exp(-\Omega(N^{1/3}))$. Let $(o, T, \rho, (\varphi, \psi)) \sim \mathrm{DGW}(\alpha k, k-1, (\Omega, \mathbb{P}_\omega) \times \mathrm{unif}([0,1]))$. By Lemma 2.7.6,

$$\mathbb{E}\,\mathbb{1}\left\{v \text{ is } (r,R,\psi)\text{-insulated in } G\right\} \leq \frac{c \log N}{N^{1/2}} + \mathbb{E}\,\mathbb{1}\left\{o \text{ is } (r,R,\psi)\text{-insulated in } T\right\}.$$

By Proposition 2.7.18, for sufficiently large R depending on α, k, r, η,

$$\mathbb{E}\,\mathbb{1}\left\{o \text{ is } (r,R,\psi)\text{-insulated in } T\right\} \leq \eta/3.$$

Putting this all together, with probability $1 - \exp(-\Omega(N^{1/3}))$,

$$\Delta(\mathcal{A}(\Phi), \mathcal{A}'(\Phi)) \leq 2\eta/3 + \frac{c \log N}{N^{1/2}} \leq \eta$$

for sufficiently large N. □

2.7.4 Low Degree Polynomials Simulate Local Algorithms

In this subsection, we prove Proposition 2.7.2, that any local algorithm can be simulated by a constant degree polynomial. The proof closely resembles the proof of [Wei20, Theorem 1.4]. The main idea is to construct a low degree polynomial by inclusion-exclusion that simulates the behavior of the local algorithm on any r-neighborhood that is a tree without too many edges. We now define this simulation.

Consider $\Phi \in \Omega_k(N, M)$ with factor graph (G, ρ). Recall that Φ is encoded by indicators $\Phi_{i,j,s}$ ($i \in [M]$, $j \in [k]$, $s \in [2N]$) that $\Phi_{i,j}$ is the sth literal of $\mathcal{L}$. For each $s \in [2N]$, let $v(s) \in [N]$ be the index of the underlying variable of the sth literal of $\mathcal{L}$. Each triple (i, j, s) is naturally associated with the edge $e = (v_{v(s)}, c_i)$ of the factor graph. For a set $S \subseteq [M] \times [k] \times [2N]$, let $e(S)$ be the (multi-)set of edges associated in this manner to triples $(i, j, s) \in S$. For $D \in \mathbb{N}$ and $v \in \mathrm{Va}_G$, let $\mathcal{G}_{v,r,D}$ be the collection of sets $S \subseteq [M] \times [k] \times [2N]$ such that

(a) The bipartite graph $G(S) = (\mathrm{Va}_G, \mathrm{Cl}_G, e(S))$ is a tree in which every non-isolated vertex has a path to v of length at most r. (This includes that $G(S)$ does not have multiple edges.)

(b) $|S| \le D$.

Equivalently, $\mathcal{G}_{v,r,D}$ is the collection of sets of (i, j, s) corresponding to all possible tree shaped r-neighborhoods of v in G of size at most D.

Definition 2.7.20 (Degree-D simulation)**.** Let $\mathcal{A}_g$ be an r-local algorithm, with i.i.d. internal randomness from $(\Omega, \mathbb{P}_\omega)$. For $D \in \mathbb{N}$, the degree-D simulation of $\mathcal{A}_g$ is the random polynomial that runs as follows on input $\Phi \in \Omega_k(N, M)$ with factor graph (G, ρ).

(1) Sample $\varphi \sim (\Omega, \mathbb{P}_\omega)^{\otimes(V_G \cup E_G)}$.

(2) For each $v = v_i \in \mathrm{Va}_G$, set

$$f_i(\Phi, \varphi) = \sum_{S \in \mathcal{G}_{v,r,D}} h(v, G(S), \rho, \varphi) \prod_{(i,j,s) \in S} \Phi_{(i,j,s)}, \tag{2.7.2}$$

where the coefficients $h(v, G(S), \rho, \varphi)$ are given recursively by

$$h(v, G(S), \rho, \varphi) = (\texttt{strictRound}^{-1} \circ g)(v, G(S), \rho\big|_{G(S)}, \varphi\big|_{G(S)}) - \sum_{S' \subsetneq S} h(v, G(S'), \rho, \varphi). \tag{2.7.3}$$

The internal randomness of f is the map φ. It is clear that this f is a degree-D polynomial. We will analyze the performance of the degree-D simulation by analogy to the following local function.

Definition 2.7.21 (D-truncation)**.** If $g : \Lambda \to \{\mathtt{T}, \mathtt{F}\}$ is an r-local function, the *D-truncation* $g_{\le D} : \Lambda \to \{\mathtt{T}, \mathtt{F}, \mathtt{err}\}$ is defined by

$$g_{\le D}(v, G, \rho, \varphi) = \begin{cases} g(v, G, \rho, \varphi) & N_r(v, G) \text{ is a tree and } |N_r(v, G)| \le D, \\ \mathtt{err} & \text{otherwise.} \end{cases}$$

By inclusion-exclusion, (2.7.2) and (2.7.3) immediately imply the following fact.

Fact 2.7.22. *For all $v = v_i \in \mathrm{Va}_G$ where $g_{\le D}(v, G, \rho, \varphi) \neq \mathtt{err}$,*

$$(\texttt{strictRound} \circ f_i)(\Phi, \varphi) = g(v, G, \rho, \varphi) = g_{\le D}(v, G, \rho, \varphi).$$

In other words, when $\mathcal{A}_g$, $\mathcal{A}_{g_{\le D}}$, and the degree-D simulation f of $\mathcal{A}_g$ are run with the same φ, $\texttt{strictRound} \circ f$ correctly simulates any output of $\mathcal{A}_g$ that $\mathcal{A}_{g_{\le D}}$ correctly simulates. Therefore, we can upper bound the fraction of variables where the simulation f fails by bounding the fracton of variables where $\mathcal{A}_{g_{\le D}}$ fails. We achieve this by controlling the corresponding probability in $\mathrm{DGW}(d_1, d_2)$, and then translating this bound to the k-SAT factor graph by the machinery of Lemma 2.7.6 and Corollary 2.7.8.

Lemma 2.7.23. *Suppose $\alpha k, k-1 \geq 2$ and $\eta > 0$. Let $\mathcal{A}_g$ be an r-local algorithm with output in $\{\mathtt{T},\mathtt{F}\}^N$. There exists $D > 0$ depending on α, k, r, η such that if $\mathcal{A}_g$ and $\mathcal{A}_{g_{\leq D}}$ are run with the same φ, then*

$$\mathbb{P}\left[\Delta(\mathcal{A}_g(\Phi), \mathcal{A}_{g_{\leq D}}(\Phi)) \geq \eta\right] \leq \exp(-\Omega(N^{1/3})),$$

where the probability is over the randomness of $\Phi \sim \Phi_k(N, M)$ and φ.

Proof. By Corollary 2.7.8 with $t = \eta N/3$,

$$\begin{aligned}\Delta(\mathcal{A}_g(\Phi), \mathcal{A}_{g_{\leq D}}(\Phi)) &= \frac{1}{N}\sum_{v \in \mathrm{Va}_G} \mathbb{1}\left\{N_r(v,G) \text{ is not a tree or } |N_r(v,G)| > D\right\} \\ &\leq \eta/3 + \mathbb{E}\,\mathbb{1}\left\{N_r(v,G) \text{ is not a tree or } |N_r(v,G)| > D\right\}\end{aligned}$$

with probability $1-\exp(-\Omega(N^{1/3}))$, because the indicator is an r-local function taking values in $[-1,1]$. Let $(o, T, \rho, \varphi) \sim \mathrm{DGW}(\alpha k, k-1)$. By Lemma 2.7.6,

$$\mathbb{E}\,\mathbb{1}\left\{N_r(v,G) \text{ is not a tree or } |N_r(v,G)| > D\right\} \leq \frac{c\log N}{N^{1/2}} + \mathbb{E}\,\mathbb{1}\left\{|N_r(o,T)| > D\right\},$$

where we use that $N_r(o,T)$ is always a tree. By Lemma 2.7.4, we can pick D large enough (depending on α, k, r, η) that

$$\mathbb{E}\,\mathbb{1}\left\{|N_r(o,T)| > D\right\} \leq \eta/3.$$

Putting this all together, with probability $1-\exp(-\Omega(N^{1/3}))$,

$$\Delta(\mathcal{A}_g(\Phi), \mathcal{A}_{g_{\leq D}}(\Phi)) \leq 2\eta/3 + \frac{c\log N}{N^{1/2}} \leq \eta$$

for sufficiently large N. □

We get the second conclusion of Proposition 2.7.2 from the following lemma.

Lemma 2.7.24. *If $\mathcal{A}_g$ is an r-local algorithm and f is its degree-D simulation, then there exists γ depending on α, k, r, D such that*

$$\mathbb{E}\,\|f(\Phi,\varphi)\|_2^2 \leq \gamma N.$$

Proof. We will upper bound each $\mathbb{E}_{\Phi,\varphi}[f_i(\Phi,\varphi)^2]$ by a constant depending only on α, k, r, η. Fix $i \in [N]$. Let $v = v_i \in \mathrm{Va}_G$ and define the random variable $X = |N_r(v,G)|$. In the expansion (2.7.2), the monomial indexed by $S \in \mathcal{G}_{v,r,D}$ is only nonzero if $e(S)$ is a subset of the edges of $N_r(v,G)$. So, the number of nonzero monomials is at most

$$k^D \sum_{d=0}^{D} \binom{X}{d} \leq k^D (X+1)^D.$$

Moreover, by (2.7.3), each of the coefficients $h(v, G(S), \rho, \varphi)$ is upper bounded by a constant a dependent on α, k, r, D. Thus

$$f_i(\Phi,\varphi)^2 \leq a^2 k^{2D} (X+1)^{2D}$$

pointwise, and so

$$\mathbb{E}\left[f_i(\Phi,\varphi)^2\right] \leq a^2 k^{2D}\,\mathbb{E}\left[(X+1)^{2D}\right].$$

Lemma 2.7.5 gives an exponential bound on the tail probability of X. Integration by tails gives the result. □

Proof of Proposition 2.7.2. Set D such that Lemma 2.7.23 holds, and let f be the degree-D simulation of $\mathcal{A}$. We couple $f, \mathcal{A}_g, \mathcal{A}_{g_{\leq D}}$ to all use the same φ. Fact 2.7.22 and Lemma 2.7.23 imply the first conclusion. Since D depends on only α, k, r, η, so does the γ given by Lemma 2.7.24. This implies the second conclusion. □

2.7.5 Deferred Proofs

In this subsection, we give the deferred proof of Lemma 2.7.19. We first prove a sharper version of Lemma 2.7.4 for a *specific* r, where the bound is improved by an r-iterated logarithmic factor.

Lemma 2.7.25. *Let $r \in \mathbb{N}$, $d_1, d_2 \geq 2$ and $(o, T, \rho, \varphi) \sim \mathrm{DGW}(d_1, d_2)$. There exists t^* (depending on r, d_1, d_2) such that for all $t \geq t^*$,*

$$\mathbb{P}\left[|N_{2r}(o,T)| \leq \frac{2t}{\log^{(r)} t}(d_1 d_2)^r\right] \geq 1 - e^{-t}.$$

Proof. For $0 \leq \ell \leq 2r$, let S_ℓ denote the number of vertices in $\mathrm{DGW}(d_1, d_2)$ at depth ℓ. The S_ℓ have the following distribution. First, $S_0 = 1$. For $\ell \geq 1$, S_ℓ is the sum of $S_{\ell-1}$ i.i.d. copies of $\mathrm{Pois}(d_1)$ if ℓ is odd, and $S_\ell = d_2 S_{\ell-1}$ if ℓ is even.

For $n \in [r]$, define the event

$$E_n = \left\{S_{2n-1} \leq \frac{t}{\log^{(n)} t} d_1^n d_2^{n-1}\right\}.$$

This is equivalent to the event that $S_{2n} \leq \frac{t}{\log^{(n)} t} d_1^n d_2^n$. For convenience, also define $E_0 = \{S_0 = 1\}$, which holds almost surely. On $\bigcap_{n=0}^r E_n$, we have

$$|N_{2r}(o,T)| = \sum_{\ell=0}^{2r} S_\ell \leq \frac{t}{\log^{(r)} t} \cdot \frac{d_1^r d_2^{r-1} + d_1^r d_2^r}{1 - (d_1 d_2)^{-1}} \leq \frac{2t}{\log^{(r)} t}(d_1 d_2)^r.$$

So, it remains to show that $\mathbb{P}\left[\bigcap_{n=0}^r E_n\right] \geq 1 - e^{-t}$.

Consider $n \in [r]$; we will upper bound $\mathbb{P}(E_n^c | E_{n-1})$. Let $N = \frac{t}{\log^{(n-1)} t} d_1^{n-1} d_2^{n-1}$ (where $\log^{(0)} t = t$). Conditioned on E_{n-1}, we have $S_{2n-2} \leq N$, so S_{2n-1} is stochastically dominated by $\sum_{j=1}^N \xi_j$, where the ξ_j are i.i.d. samples from $\mathrm{Pois}(d_1)$. By a standard Chernoff bound,

$$\begin{aligned}
\mathbb{P}(E_n^c | E_{n-1}) &\leq \mathbb{P}\left[\sum_{j=1}^N \xi_j \geq \frac{\log^{(n-1)} t}{\log^{(n)} t} d_1 N\right] \\
&\leq \left[\inf_{s>0} \mathbb{E}\exp(s\xi_1) \cdot \exp\left(-\frac{\log^{(n-1)} t}{\log^{(n)} t} s d_1\right)\right]^N \\
&= \left[\inf_{s>0} \exp\left((e^s - 1)d_1 - \frac{\log^{(n-1)} t}{\log^{(n)} t} s d_1\right)\right]^N \\
&= \exp\left(-N d_1 \gamma\left(\frac{\log^{(n-1)} t}{\log^{(n)} t}\right)\right),
\end{aligned}$$

where $\gamma(x) = x \log x - x + 1$. For large enough t,

$$\gamma\left(\frac{\log^{(n-1)} t}{\log^{(n)} t}\right) \geq \frac{3}{4} \frac{\log^{(n-1)} t}{\log^{(n)} t} \log \frac{\log^{(n-1)} t}{\log^{(n)} t} \geq \frac{2}{3} \log^{(n-1)} t,$$

while (as $d_1, d_2 \geq 2$ implies $d_1^n d_2^{n-1} \geq 2^{2n-1} \geq 2n$ for $n \geq 1$)

$$N d_1 = \frac{t}{\log^{(n-1)} t} d_1^n d_2^{n-1} \geq 2n \cdot \frac{t}{\log^{(n-1)} t}.$$

Thus, for large enough t, $\mathbb{P}(E_n^c | E_{n-1}) \leq \exp(-\frac{4}{3} n t)$. So,

$$\mathbb{P}\left[\bigcap_{n=0}^r E_n\right] \geq 1 - \sum_{n=1}^r \mathbb{P}(E_n^c | E_{n-1}) \geq 1 - \frac{\exp(-\frac{4}{3}t)}{1 - \exp(-\frac{4}{3}t)} \geq 1 - e^{-t}$$

for sufficiently large t. □

Proof of Lemma 2.7.19. Set $\lambda > 0$ such that the conclusion of Lemma 2.7.4 holds with probability $1 - \eta/2$. Denote this event S; on this event, $|N_R(o,T)| \le \lambda(d_1 d_2)^{\lceil R/2 \rceil}$ for all $R \in \mathbb{N}$.

For $R \in \mathbb{N}$, let $t(R)$ be the smallest positive integer such that $\lambda(d_1 d_2)^{\lceil R/2 \rceil} e^{-t(R)} \le \eta/2$; note that $t(R) = \Theta(R)$ for d_1, d_2, r, η fixed. Set R^* such that $t(R^*) \ge t^*$ for the t^* in Lemma 2.7.25. Henceforth let $R \ge R^*$ and $t = t(R) \ge t^*$.

For $v \in \mathrm{Va}_T$, let $N_{2r}^{\downarrow}(v,T)$ denote the subset of $N_{2r}(v,T)$ in the descendant subtree of v. Note that the descendant subtree of v has distribution $\mathrm{DGW}(d_1, d_2)$, so $|N_{2r}^{\downarrow}(v,T)| =_d |N_{2r}(o,T)|$. By Lemma 2.7.25, for each $v \in \mathrm{Va}_T$,

$$\mathbb{P}\left[|N_{2r}^{\downarrow}(v,T)| \le \frac{2t}{\log^{(r)} t}(d_1 d_2)^r\right] \ge 1 - e^{-t}.$$

By a union bound,

$$\mathbb{P}\left[S \text{ and } \max_{v \in \mathrm{Va}_T \cap N_R(o,T)} |N_{2r}^{\downarrow}(v,T)| \le \frac{2t}{\log^{(r)} t}(d_1 d_2)^r\right] \ge 1 - \frac{\eta}{2} - \lambda(d_1 d_2)^{\lceil R/2 \rceil} e^{-t} \ge 1 - \eta.$$

Let S' be the event in this probability. Note that for $v \in \mathrm{Va}_T$,

$$N_{2r}(v,T) \subseteq N_{2r}^{\downarrow}(v,T) \cup N_{2r}^{\downarrow}(\mathrm{gr}(v),T) \cup \cdots \cup N_{2r}^{\downarrow}(\mathrm{gr}^r(v),T),$$

where $\mathrm{gr}(v)$ denotes the grandparent of v. Thus, on the event S', we have

$$\max_{v \in \mathrm{Va}_T \cap N_R(o,T)} |N_{2r}(v,T)| \le \frac{2t(r+1)}{\log^{(r)} t}(d_1 d_2)^r.$$

For $v \in \mathrm{Cl}_T \cap N_R(o,T)$, simply note that $N_{2(r-1)}(v,T) \subseteq N_{2r}(\mathrm{pa}(v),T)$, where $\mathrm{pa}(v)$ denotes the parent of v. It follows that on S',

$$\max_{v \in N_R(o,T)} |N_{2(r-1)}(v,T)| \le \frac{2t(r+1)}{\log^{(r)} t}(d_1 d_2)^r \le \frac{CR}{\log^{(r)} R},$$

using that $t = t(R)$ and $t(R) = \Theta(R)$. The result follows by renaming r to $r+1$. □

2.8 Proof of Achievability

Throughout this section, let $\varepsilon > 0$, $\alpha = (1-\varepsilon)2^k \log k/k$, and $M = \lfloor \alpha N \rfloor$. In this section we will prove Theorem 2.1.14, that local algorithms and low degree polynomials can solve random k-SAT at this clause density α.

We will prove this theorem by simulating the first phase of `Fix`, which we denote `Fix1`, by these two computation classes. Parts (a,b) of Theorem 2.1.14 follow immediately from guarantees on `Fix1` in [CO10] and our simulation results, Propositions 2.7.1 and 2.7.2. To prove parts (c,d), we use the fact that `Fix1` is simulated by a local algorithm to argue concentration of the number of clauses satisfied, in order to prove the stronger bound on the failure probability.

This section is structured as follows. In Subsection 2.8.1 we define `Fix1` and introduce its guarantees. This immediately implies Theorem 2.1.14(a,b). In Subsection 2.8.2 we show concentration of the number of clauses satisfied and prove Theorem 2.1.14(c,d).

2.8.1 Review of Fix

At clause density $\alpha = (1-\varepsilon)2^k \log k/k$, `Fix` produces a (exactly) satisfying assignment with high probability. At a high level, `Fix` runs in three phases. In the first phase, it produces a almost-satisfying assignment. In the second phase, it modifies this assignment in a small fraction of variables, at most k^{-12} with high probability, to "don't know." This is done in such a way that the remaining problem of assigning truth values to the "don't know" variables is equivalent to a very subcritical random 3-SAT instance. The third phase solves the remaining problem with a maxflow algorithm.

We will only show that local algorithms and low degree polynomials simulate the first phase `Fix1`. Because the the rest of `Fix` changes at most a k^{-12} fraction of variables with high probability, simulating `Fix1` within normalized Hamming distance $\eta' > 0$ simulates `Fix` within error $\eta = k^{-12} + \eta'$. This is why Theorem 2.1.14 requires $\eta > k^{-12}$. Let us record the guarantees on `Fix1` proved in [CO10].

Theorem 2.8.1 (Implicit in [CO10, Section 3])**.** *Let* `Fix1` *be defined in Algorithm 2.8.2 below. Then,*

$$\mathbb{P}\left[\texttt{Fix1}(\Phi) \text{ } (k^{-12}, 0)\text{-satisfies } \Phi\right] \geq 1 - o(1).$$

The probability is over $\Phi \sim \Phi_k(N, M)$ *and the (independent) internal randomness of* `Fix1`*.*

We now define `Fix1`. This phase starts from the all-true assignment $x = \texttt{T}^N$ and selects a set of indices $Z \subseteq [N]$ such that if $\{x_i : i \in Z\}$ are set false, most clauses are satisfied. To do this, it scans through the clauses of the input formula Φ. When it encounters an all-negative clause that does not contain any variable from Z, it tries to find a true variable x_i from this clause that when made false does not create more unsatisfied clauses. It adds this i to Z. Formalizing this idea, we say x_i (for $i \in [N] \setminus Z$) is *Z-safe* if, when we set $\{x_{i'} : i' \in [N] \setminus Z\}$ to true and $\{x_{i'} : i' \in Z\}$ to false, x_i is not the sole true literal in any clause.

Algorithm 2.8.2 (Fix, Phase 1; [CO10])**.** On input $\Phi \in \Omega_k(N, M)$, `Fix1` runs as follows.

(1) Set $Z = \emptyset$.

(2) Relabel the clauses $\{\Phi_i : i \in [M]\}$ in a uniformly random order. Also, for each $i \in [M]$, relabel the literals $\{\Phi_{i,j} : j \in [k]\}$ in a uniformly random order.

(3) For $i \in [M]$ in increasing order:

  (a) If Φ_i is all-negative and contains no variable from $\{x_i : i \in Z\}$:

    (i) If there is $1 \leq j < \lceil k/2 \rceil$ such that the underlying variable of $\Phi_{i,j}$ is Z-safe, pick the smallest such j and add the underlying variable of $\Phi_{i,j}$ to Z.

    (ii) Otherwise, add the underlying variable of $\Phi_{i,\lceil k/2 \rceil}$ to Z.

(4) Output $x \in \{\texttt{T}, \texttt{F}\}^N$ where $x_i = \texttt{F}$ if $i \in Z$ and otherwise $x_i = \texttt{T}$.

The presentation of `Fix1` in [CO10] does not rerandomize the clause and literal orders, but of course this makes no difference. We add this rerandomization so that the algorithm is a local memory algorithm in the sense we define. For technical reasons having to do with the analysis in [CO10], `Fix1` only considers flipping variables $\Phi_{i,j}$ where $j \leq \lceil k/2 \rceil$.

Fact 2.8.3. `Fix1` *is a 3-local memory algorithm.*

Proof. We will construct a 3-local subroutine f_1 and a function $f_2 : \mathbb{Z}_{\geq 0} \to \{\texttt{T}, \texttt{F}\}$ such that $\mathcal{A}_{f_1, f_2}$ simulates `Fix1`. Let (G, ρ) be the factor graph of Φ, and let the i.i.d. randomness of φ be sampled from $\mathrm{unif}([0, 1])$.

We will maintain the invariant that for each $v = v_i \in \mathrm{Va}_G$, $\mu(v) = 1$ if $i \in Z$, and otherwise $\mu(v) = 0$.

The subroutine f_1 runs as follows on $(v, G, \rho, \varphi, \mu)$. If $v \in \mathrm{Va}_G$, do nothing. The remaining loop over $v \in \mathrm{Cl}_G$ runs over these vertices in a uniformly random order, as desired. If $v = c_i \in \mathrm{Cl}_G$, f_1 orders the edges e incident to c_i in increasing order of $\varphi(e)$. It runs the logic inside the for loop of `Fix1`, with the corresponding literals $\{\Phi_{i,j} : j \in [k]\}$ relabeled in this order, and records the outcome on μ. Note that the literals $\{\Phi_{i,j} : j \in [k]\}$ are relabeled in a uniformly random order, and that the logic inside the for loop is 3-local. □

The proof of Theorem 2.1.14(a,b) follows immediately from Fact 2.8.3 and our simulation results.

Proof of Theorem 2.1.14(a,b). Since $\eta > k^{-12}$, we can find $\eta' > 0$ such that $\eta = k^{-12} + 2\eta'$. Theorem 2.8.1, Fact 2.8.3 and Proposition 2.7.1 give $r > 0$ and an r-local algorithm $\mathcal{A}$ such that

$$\mathbb{P}\left[\mathcal{A}(\Phi) \text{ } (k^{-12} + \eta', 0)\text{-satisfies } \Phi\right] \geq 1 - \delta(N) \tag{2.8.1}$$

for $\delta(N) = o(1) + \exp(-\Omega(N^{1/3})) = o(1)$. Since $k^{-12} + \eta' < \eta$, this proves part (a).

Proposition 2.7.2 gives D, γ and a random degree-D polynomial that $(\delta(N), \gamma, \eta, 0)$-solves $\Phi_k(N, M)$, for $\delta(N)$ with a larger $\exp(-\Omega(N^{1/3}))$ term. Here we use that $\eta = k^{-12} + 2\eta'$. Finally, Lemma 2.3.1 gives a deterministic degree-D polynomial that $(3\delta(N), 3\gamma, \eta, 0)$-solves $\Phi_k(N, M)$. This proves part (b). □

2.8.2 Concentration of Clauses Satisfied

For $x \in \{\mathtt{T},\mathtt{F}\}^N$, $\Phi \in \Omega_k(N,M)$, and $\eta \in (0,1)$, define the objective

$$\mathrm{Sat}_\eta(x,\Phi) = \max_{y \in \{\mathtt{T},\mathtt{F}\}^N : \Delta(x,y) \le \eta} (\#\text{ clauses of } \Phi \text{ satisfied by } y)\,. \tag{2.8.2}$$

To prove Theorem 2.1.14(c), we will show that the objective attained by any local algorithm concentrates. For an assignment $x \in \{\mathtt{T},\mathtt{F}\}^N$ and a partial assignment $z \in \{\mathtt{T},\mathtt{F}\}^B$, where $B \subseteq [N]$, it will be useful to define the replacement operator $\mathrm{Rep}(x,z)$ by

$$\mathrm{Rep}(x,z) = y \in \{\mathtt{T},\mathtt{F}\}^N \qquad \text{where} \qquad y_i = \begin{cases} x_i & i \notin B, \\ z_i & i \in B. \end{cases} \tag{2.8.3}$$

Proposition 2.8.4. *Let $\mathcal{A}$ be an r-local algorithm with internal randomness φ. Let $\eta \in (0,1)$ and $Y = \mathrm{Sat}_\eta(\mathcal{A}(\Phi,\varphi),\Phi)$. Then,*

$$\mathbb{P}\left[|Y - \mathbb{E}\,Y| \ge \frac{M}{\log N}\right] \le \exp(-\widetilde{\Omega}(N^{1/3})).$$

Proof. We will show Y has bounded differences with high probability, which implies concentration by Lemma 2.6.6.

Throughout this section, we will write $\mathcal{A}(\Phi,\varphi)$ for $\mathcal{A}$ run with input Φ and internal randomness φ. Define $z = z(\Phi,\varphi) \in \{\mathtt{T},\mathtt{F}\}^B$ as the partial assignment maximizing

$$\mathrm{Sat}_0\left(\mathrm{Rep}(\mathcal{A}(\Phi,\varphi),z),\Phi\right),$$

where the maximization is over all $B \subseteq [N]$ with $|B| \le \eta N$. We break ties arbitrarily but deterministically. By definition of Y, $\mathrm{Rep}(\mathcal{A}(\Phi,\varphi),z)$ satisfies Y clauses of Φ.

Let (G,ρ) be the factor graph of Φ. Let the vertex sets of G be $\mathrm{Va} = \{v_1,\dots,v_N\}$, $\mathrm{Cl} = \{c_1,\dots,c_M\}$, and $V = \mathrm{Va} \cup \mathrm{Cl}$, which are fixed across all realizations of G.

All the above random variables are (Φ,φ)-measurable. We can reformat (Φ,φ) into $kM+M+N$ independent parts $\zeta = (\psi_1,\dots,\psi_{kM+M+N})$: for $1 \le j \le kM$, $\psi_j = (\Phi_{L(j)},\varphi(e))$ where e is the edge in G corresponding to $\Phi_{L(j)}$, and for $kM+1 \le j \le kM+M+N$, $\psi_j = \varphi(v)$ ranges over $v \in V$. We will henceforth write $Y(\zeta)$ to denote the Y corresponding to this realization of ζ, and similarly for other random variables, which are all ζ-measurable.

Let S denote the set of ζ such that $\Phi(\zeta)$ is $(r+1)$-locally small (recall Definition 2.6.7). By Fact 2.6.8, $\mathbb{P}(S^c) \le \exp(-\Omega(N^{1/3}))$. Suppose $\zeta,\zeta' \in S$ differ in only one coordinate. We will upper bound $|Y(\zeta)-Y(\zeta')|$. For now, assume the differing coordinate is ψ_j for $1 \le j \le kM$; thus the factor graphs $G(\zeta), G(\zeta')$ differ in one edge. Let $c \in \mathrm{Cl}$ be the common endpoint of this edge.

Assume without loss of generality that $Y(\zeta) \ge Y(\zeta')$. Then,

$$\begin{aligned} Y(\zeta) - Y(\zeta') &= \mathrm{Sat}_0(\mathrm{Rep}(\mathcal{A}(\zeta),z(\zeta)),\Phi(\zeta)) - \mathrm{Sat}_0(\mathrm{Rep}(\mathcal{A}(\zeta'),z(\zeta')),\Phi(\zeta')) \\ &\le \mathrm{Sat}_0(\mathrm{Rep}(\mathcal{A}(\zeta),z(\zeta)),\Phi(\zeta)) - \mathrm{Sat}_0(\mathrm{Rep}(\mathcal{A}(\zeta'),z(\zeta)),\Phi(\zeta')) \end{aligned}$$

The last inequality holds because $z(\zeta')$ maximizes the number of clauses of $\Phi(\zeta')$ satisfied by $\mathrm{Rep}(\mathcal{A}(\zeta'),z(\zeta'))$.

Note that $\mathcal{A}(\zeta)$ and $\mathcal{A}(\zeta')$ only differ in coordinates $i \in [N]$ where $v_i \in N_r(c,G(\zeta)) \cup N_r(c,G(\zeta'))$. Thus $\mathrm{Rep}(\mathcal{A}(\zeta),z(\zeta))$ and $\mathrm{Rep}(\mathcal{A}(\zeta'),z(\zeta))$ differ in only these coordinates. So, if

$$\mathbb{1}\left\{\mathrm{Rep}(\mathcal{A}(\zeta),z(\zeta))\right\} \text{ satisfies clause } \Phi(\zeta)_i\} \neq \mathbb{1}\left\{\mathrm{Rep}(\mathcal{A}(\zeta'),z(\zeta))\right\} \text{ satisfies clause } \Phi(\zeta')_i\},$$

then $c_i \in N_{r+1}(c,G(\zeta)) \cup N_{r+1}(c,G(\zeta'))$. Because $\zeta,\zeta' \in S$, this implies $|Y(\zeta)-Y(\zeta')| \le O(N^{1/3})$.

We can analogously show the same bounded difference inequality when $\zeta,\zeta' \in S$ differ in coordinate ψ_j for $kM+1 \le j \le kM+M+N$, corresponding to a vertex of the factor graphs. Moreover, for all ζ,ζ', clearly $|Y(\zeta)-Y(\zeta')| \le M$. By Lemma 2.6.6,

$$\begin{aligned} \mathbb{P}\left[|Y-\mathbb{E}\,Y| \ge \frac{M}{\log N}\right] &\le 2\exp\left(\frac{m^2/\log^2 N}{8(kM+M+N)O(N^{2/3})}\right) + \exp(-\Omega(N^{1/3}))O\left(\frac{2(kM+M+N)m}{N^{1/3}}\right) \\ &\le \exp(-\widetilde{\Omega}(N^{1/3})). \end{aligned}$$

□

Proof of Theorem 2.1.14(c). Equation (2.8.1) gives an r-local algorithm $\mathcal{A}$ such that $\mathcal{A}(\Phi,\varphi)$ $(\eta,0)$-satisfies Φ with probability $1-o(1)$. If $Y=\mathrm{Sat}_\eta(\mathcal{A}(\Phi,\varphi),\Phi)$, this implies

$$\mathbb{E}\,Y=(1-o(1))M.$$

Proposition 2.8.4 proves the result with $\nu(N)=\frac{1}{\log N}+o(1)=o(1)$. □

Recall that the proof of Proposition 2.7.2 simulates a local algorithm by its D-truncation, which can be implemented by a low degree polynomial. We will prove Theorem 2.1.14(d) by showing a concentration result analogous to Proposition 2.8.4 for D-truncations of local algorithms.

To formulate this result, we first extend the definition (2.8.2) of Sat_η to allow $x\in\{\mathtt{T},\mathtt{F},\mathtt{err}\}^N$. Note that the y in the maximum of (2.8.2) must differ from x in all positions where the entry of x is $\mathtt{err}$. We define $\mathrm{Sat}_\eta(x,\Phi)=0$ if x has more than ηN entries equal to $\mathtt{err}$. (In particular, $\mathrm{Sat}_0(x,\Phi)=0$ if x has any entry equal to $\mathtt{err}$.) We similarly extend the definition (2.8.3) of Rep to allow $x\in\{\mathtt{T},\mathtt{F},\mathtt{err}\}^N$.

Proposition 2.8.5. *Let $\mathcal{A}_g$ be an r-local algorithm with internal randomness φ, where $g:\Lambda\to\{\mathtt{T},\mathtt{F}\}$ is an r-local function, and let $g_{\le D}$ be the D-truncation of g. Let $\eta\in(0,1)$. Let D be large enough that*

$$\mathbb{P}\left[\Delta(\mathcal{A}_g(\Phi,\varphi),\mathcal{A}_{g_{\le D}}(\Phi,\varphi))\ge\eta\right]\le\exp(-\Omega(N^{1/3})). \tag{2.8.4}$$

(Such D exists by Lemma 2.7.23.) Let $Y=\mathrm{Sat}_\eta(\mathcal{A}_{g_{\le D}}(\Phi,\varphi),\Phi)$. Then,

$$\mathbb{P}\left[|Y-\mathbb{E}\,Y|\ge\frac{M}{\log N}\right]\le\exp(-\widetilde{\Omega}(N^{1/5})).$$

Proof. We will again show Y has bounded differences with high probability and use Lemma 2.6.6. Define $z=z(\Phi,\varphi)\in\{\mathtt{T},\mathtt{F}\}^B$ as the partial assignment maximizing

$$\mathrm{Sat}_0(\mathrm{Rep}(\mathcal{A}_{g_{\le D}}(\Phi,\varphi),z),\Phi).$$

Let (G,ρ) be the factor graph of Φ, with vertex sets $\mathrm{Va}=\{v_1,\dots,v_N\}$, $\mathrm{Cl}=\{c_1,\dots,c_M\}$, and $V=\mathrm{Va}\cup\mathrm{Cl}$.

Define ζ as in the proof of Proposition 2.8.4. Let S denote the set of ζ such that:

(i) $\Delta(\mathcal{A}_g(\zeta),\mathcal{A}_{g_{\le D}}(\zeta))\le\eta$ and

(ii) For all $v\in V_G$, $|N_{r+1}(v,G)|\le N^{1/5}$.

By the assumption (2.8.4) and Lemma 2.7.5, $\mathbb{P}(S^c)\le\exp(-\Omega(N^{1/5}))$. Note that for $\zeta\in S$, (i) implies that $\mathrm{Rep}(\mathcal{A}_{g_{\le D}}(\zeta),z(\zeta))$ has no $\mathtt{err}$ symbols.

Suppose $\zeta,\zeta'\in S$ differ in only one coordinate. We will upper bound $|Y(\zeta)-Y(\zeta')|$. Assume the differing coordinate is ψ_j for some $1\le j\le kM$. (The case $kM+1\le j\le kM+M+N$ is analogous.) Then, the factor graphs $G(\zeta),G(\zeta')$ differ in one edge. Let $c\in\mathrm{Cl}$ be the common endpoint of this edge.

Let $U=N_r(c,G(\zeta))\cup N_r(c,G(\zeta'))$. Because $\zeta,\zeta'\in S$, (ii) implies $|U|\le 2N^{1/5}$. Note that $\mathcal{A}_{g_{\le D}}(\zeta)$ and $\mathcal{A}_{g_{\le D}}(\zeta')$ only differ in coordinates $i\in[N]$ where $v_i\in U$.

Assume without loss of generality that $Y(\zeta)\ge Y(\zeta')$. Unlike in the proof of Proposition 2.8.4, the estimate

$$\mathrm{Sat}_0(\mathrm{Rep}(\mathcal{A}_{g_{\le D}}(\zeta'),z(\zeta')))\ge\mathrm{Sat}_0(\mathrm{Rep}(\mathcal{A}_{g_{\le D}}(\zeta'),z(\zeta)))$$

is not helpful because the right-hand side is 0 when $\mathrm{Rep}(\mathcal{A}_{g_{\le D}}(\zeta'),z(\zeta))$ has $\mathtt{err}$ symbols. Instead we note that, because $\mathcal{A}_{g_{\le D}}(\zeta)$ and $\mathcal{A}_{g_{\le D}}(\zeta')$ differ in at most $|U|$ positions, there exists z' differing from $z(\zeta)$ in at most $2|U|$ positions ($|U|$ entries in $z(\zeta)$ not in z' and vice versa) such that $\mathrm{Rep}(\mathcal{A}_{g_{\le D}}(\zeta'),z')$ has no $\mathtt{err}$ symbols. We use the estimate

$$\begin{aligned}Y(\zeta)-Y(\zeta')&=\mathrm{Sat}_0(\mathrm{Rep}(\mathcal{A}_{g_{\le D}}(\zeta),z(\zeta)),\Phi(\zeta))-\mathrm{Sat}_0(\mathrm{Rep}(\mathcal{A}_{g_{\le D}}(\zeta'),z(\zeta')),\Phi(\zeta'))\\&\le\mathrm{Sat}_0(\mathrm{Rep}(\mathcal{A}_{g_{\le D}}(\zeta),z(\zeta)),\Phi(\zeta))-\mathrm{Sat}_0(\mathrm{Rep}(\mathcal{A}_{g_{\le D}}(\zeta'),z'),\Phi(\zeta')).\end{aligned}$$

Now, if

$$\mathbb{1}\left\{\mathrm{Rep}(\mathcal{A}_{g_{\le D}}(\zeta),z(\zeta)))\text{ satisfies clause }\Phi(\zeta)_i\right\}\ne\mathbb{1}\left\{\mathrm{Rep}(\mathcal{A}_{g_{\le D}}(\zeta'),z'))\text{ satisfies clause }\Phi(\zeta')_i\right\},$$

either $c_i \in N_{r+1}(c, G(\zeta)) \cup N_{r+1}(c, G(\zeta'))$ or c_i is adjacent to one of the (at most) $2|U|$ variables where $z(\zeta)$ and z' disagree. By definition of S, there are at most $2N^{1/5}$ clauses in the former case, and $2|U|N^{1/5}$ clauses in the latter case. Thus $|Y(\zeta) - Y(\zeta')| \le O(N^{2/5})$.

For general ζ, ζ', we have $|Y(\zeta) - Y(\zeta')| \le M$. By Lemma 2.6.6,

$$\begin{aligned}\mathbb{P}\left[|Y - \mathbb{E}\,Y| \ge \frac{M}{\log N}\right] &\le 2\exp\left(\frac{M^2/\log^2 N}{8(kM+M+N)O((N^{2/5})^2)}\right) + \exp(-\Omega(N^{1/5}))O\left(\frac{2(kM+M+N)m}{N^{2/5}}\right)\\ &\le \exp(-\widetilde{\Omega}(N^{1/5})).\end{aligned}$$

□

Proof of Theorem 2.1.14(d). Set $\eta' > 0$ such that $\eta = k^{-12} + 2\eta'$. Let $\mathcal{A} = \mathcal{A}_g$ be the r-local algorithm achieving (2.8.1). By Lemma 2.7.23, there exists D dependent on ε, k, η such that

$$\mathbb{P}\left[\Delta(\mathcal{A}_g(\Phi,\varphi), \mathcal{A}_{g_{\le D}}(\Phi,\varphi)) \ge \eta'\right] \le \exp(-\Omega(N^{1/3})).$$

With (2.8.1), this implies

$$\mathbb{P}\left[\mathcal{A}_{g_{\le D}}(\Phi,\varphi) \text{ } (\eta,0)\text{-satisfies } \Phi\right] \ge 1 - o(1).$$

Thus,

$$\mathbb{E}\,\mathrm{Sat}_\eta(\mathcal{A}_{g_{\le D}}(\Phi,\varphi),\Phi) = (1-o(1))M.$$

For $\nu(N) = o(1) + \frac{1}{\log N} = o(1)$, Proposition 2.8.5 implies that

$$\mathbb{P}\left[\mathrm{Sat}_\eta(\mathcal{A}_{g_{\le D}}(\Phi,\varphi),\Phi) \ge (1-\nu(N))M\right] \ge 1 - \exp(-\widetilde{\Omega}(N^{1/5})).$$

Fact 2.7.22 implies that the degree-D simulation f of $\mathcal{A}_g$ satisfies

$$\mathbb{P}\left[\mathrm{Sat}_\eta((\texttt{strictRound} \circ f)(\Phi,\varphi),\Phi) \ge (1-\nu(N))M\right] \ge 1 - \exp(-\widetilde{\Omega}(N^{1/5})).$$

In other words,

$$\mathbb{P}\left[(\texttt{strictRound} \circ f)(\Phi,\varphi) \text{ } (\eta,\nu(N))\text{-satisfies } \Phi\right] \ge 1 - \exp(-\widetilde{\Omega}(N^{1/5})).$$

Lemma 2.7.24 gives γ such that

$$\mathbb{E}\,\|f(\Phi,\varphi)\|_2^2 \le \gamma N.$$

Thus, f is a degree-D polynomial that $(\exp(-\widetilde{\Omega}(N^{1/5})), \gamma, \eta, \nu(N))$-solves $\Phi_k(N,M)$. Finally, Lemma 2.3.1 gives a deterministic degree-D polynomial that $(\exp(-\widetilde{\Omega}(N^{1/5})), 3\gamma, \eta, \nu(N))$-solves $\Phi_k(N,M)$. □

2.9 Appendix: On Improving the Constant κ^*

In this section, we discuss how the constant κ^* in Theorem 2.1.6 can be improved. We define a constant κ^{**} as the solution to a maximin problem. We will show that $\kappa^{**} \le \kappa^*$ and sketch how our proof of Theorem 2.1.6 can be lightly modified to improve the constant κ^* to κ^{**}. We heuristically argue that $\kappa^{**} < \kappa^*$, so that this modification is an improvement. We also prove that κ^{**} is bounded below by a constant larger than 1, approximately 1.716. Further ideas will be needed to prove Theorem 2.1.6 for any κ smaller than κ^{**}. Because κ^{**} remains bounded away from 1, and we believe 1 is the optimal constant in Theorem 2.1.6, we did not attempt to rigorously evaluate κ^{**} or optimize κ^*.

2.9.1 A Maximin Problem

Let (Ξ, P_ξ) be an arbitrary probability space and let $\mathcal{Q}$ be the space of functions $q : \Xi \to [0,1]$. These are abstractions of quantities in the proof of Proposition 2.4.7: $\xi \sim (\Xi, P_\xi)$ is an abstraction of the random variables $y_I^{(\le \ell-1)}$ where $I \sim \mathrm{unif}([N])$, and $q(\xi)$ is an abstraction of $\phi_\ell(\mathtt{T}|y_I^{(\le \ell-1)})$. We equip $\mathcal{Q}$ with the metric $d(q,q') = \mathbb{E}_\xi\,|q(\xi) - q'(\xi)|$.

For $q \in \mathcal{Q}$, let $D(q)$ be the law of u sampled by the following experiment. First, sample $\xi \sim (\Xi, P_\xi)$. Then, set $u = -\log q(\xi)$ with probability $q(\xi)$, and otherwise set $u = -\log(1-q(\xi))$. Clearly $\mathbb{E}_{u\sim D(q)}\, u = \mathbb{E}_\xi\, H(q(\xi))$. Define

$$F(q) = \frac{1}{\log k} \cdot \frac{\log 2 + k\,\mathbb{E}_\xi\, H(q(\xi))}{\mathbb{P}_{(u_1,\dots,u_k)\sim D(q)^{\otimes k}}\left[\sum_{i=1}^k u_i \geq \log k + \log\log k\right]}.$$

Let $\mathcal{P}$ be the set of functions $p : \Xi \times [0,1] \to [0,1]$, such that $p(\xi, 0) \in \{0,1\}$ and $p(\xi,1) = \frac{1}{2}$ for all $\xi \in \Xi$, and $p(\cdot, s)$ (which, for fixed $s \in [0,1]$, is an element of $\mathcal{Q}$) is continuous in s with respect to the topology of $\mathcal{Q}$. Consider the maximin problem

$$\kappa^{**} = \limsup_{k\to\infty} \max_{p\in\mathcal{P}} \min_{s\in[0,1]} F(p(\cdot,s)). \tag{2.9.1}$$

This has the following geometric interpretation: κ^{**} is the smallest constant such that the sub-level set $\{q \in \mathcal{Q} : F(q) \leq \kappa^{**}\}$ topologically disconnects the functions $q \equiv 0$ and $q \equiv \frac{1}{2}$ in $\mathcal{Q}$. (Note that $\mathcal{Q}$ is symmetric under replacing $q(\xi)$ with $1-q(\xi)$ for any subset of the $\xi \in \Xi$, and $F(q) = F(q')$ for any q, q' related by such a symmetry. Thus, equivalently κ^{**} is the smallest constant such that this sub-level set disconnects the function $q \equiv \frac{1}{2}$ from any $q \in \mathcal{Q}$ with $q(\xi) \in \{0,1\}$ for all $\xi \in \Xi$.)

First, we show that κ^* is an upper bound on the solution to this maximin problem.

Proposition 2.9.1. *We have that $\kappa^* \geq \kappa^{**}$.*

Proof. Fix some $p \in \mathcal{P}$. By continuity of $p(\cdot, s)$ in s, we can set $s \in [0,1]$ such that $\mathbb{E}_\xi\, H(p(\xi,s)) = \beta^* \frac{\log k}{k}$. As in the proof of Proposition 2.4.7, we apply a Chernoff bound on the random variables $\frac{\min(u_i, \log k)}{\log k}$ to show that, for any $\beta > 1$ and $q : \Omega \to [0,1]$ with $\mathbb{E}_\xi\, H(q(\xi)) = \beta\frac{\log k}{k}$, we have

$$\mathbb{P}_{(u_1,\dots,u_k)\sim D(q)^{\otimes k}}\left[\sum_{i=1}^k u_i < \log k + \log\log k\right] \leq \beta e^{-(\beta-1)} + o_k(1). \tag{2.9.2}$$

In particular, for the s we chose,

$$F(p(\cdot,s)) \leq \frac{\frac{2}{\log k} + \beta^*}{1 - \beta^* e^{-(\beta^*-1)} - o_k(1)} \to \iota(\beta^*) = \kappa^*.$$

□

Next, we sketch how the proof of Theorem 2.1.6 can be improved to replace κ^* with κ^{**}. The proof of Theorem 2.1.13 can be modified similarly.

Proposition 2.9.2. *Theorem 2.1.6 holds for all $\kappa > \kappa^{**}$.*

Proof Sketch. Identically to the original proof of Theorem 2.1.6, we define the interpolation path $\Phi^{(0)}, \dots, \Phi^{(T)}$ and set $x^{(t)} = \mathcal{A}(\Phi^{(t)})$ for $0 \leq t \leq T$. We define S_{valid} as before. S_{consec} and S_{indep} are analogous to before: S_{consec} is the event that consecutive $x^{(t)}$ are close in Hamming distance, and S_{indep} is the event that if $0 \leq t_0 \leq t_1 \leq \cdots \leq t_k \leq T$ and $t_k \geq t_{k-1} + kM$, then any ν-satisfying assignment to $\Phi^{(t_k)}$ has large conditional overlap entropy relative to $x^{(t_0)}, \dots, x^{(t_{k-1})}$. We change the parameters quantifying "close" and "large conditional overlap" slightly so that the below proof succeeds; we omit the details. Lower bounds on $\mathbb{P}(S_{\text{valid}} \cap S_{\text{consec}})$ and $\mathbb{P}(S_{\text{indep}})$ can be proved analogously to Proposition 2.3.7(a,b).

The interesting change will be in the definition of S_{ogp}. For $1 \leq \ell \leq k$, the conditional overlap profile $\pi(y^{(\ell)} | y^{(0)}, \dots, y^{(\ell-1)})$ determines the conditional probabilities[1]

$$\phi_\ell(b|\xi) = \mathbb{P}_{i\sim\text{unif}([N])}\left[y_i^{(\ell)} \oplus y_i^{(0)} = b | (y_i^{(1)} \oplus y_i^{(0)}, \dots, y_i^{(\ell-1)} \oplus y_i^{(0)}) = \xi\right],$$

where $\oplus$ denotes $\texttt{XOR}$. Let (Ξ, P_ξ) be the sample space of $(y_i^{(1)} \oplus y_i^{(0)}, \dots, y_i^{(\ell-1)} \oplus y_i^{(0)})$, and let $q(\xi) = \phi_\ell(\texttt{T}|\xi)$. Let $\varepsilon > 0$ satisfy $\kappa - \varepsilon > \kappa^{**}$. S_{ogp} is now the event that there does not exist $0 \leq t_0 \leq t_1 \leq \cdots \leq t_k \leq T$ and assignments $y^{(0)}, \dots, y^{(k)} \in \{\texttt{T}, \texttt{F}\}^N$ such that

[1] this is a rewriting of the argument in Subsection 2.4.3. The $\texttt{XOR}$s arise because we no longer assume $y^{(0)} = \texttt{T}^N$.

(OGP-A) For all $0 \le \ell \le k$, $y^{(\ell)}$ ν-satisfies $\Phi^{(t_\ell)}$;

(OGP-B) For all $1 \le \ell \le k$, the conditional overlap profile $H\left(\pi(y^{(\ell)}|y^{(0)}, \dots, y^{(\ell-1)})\right)$ satisfies that $F(q) \le \kappa - \varepsilon$ for the q defined above.

The key point is that our proof that $\mathbb{P}(S_{\text{ogp}}^c) \le \exp(-\Omega(N))$ requires precisely these properties. Using the argument in Section 2.4, we readily prove $\mathbb{P}(S_{\text{ogp}}^c) \le \exp(-\Omega(N))$.

By a union bound, this gives a positive lower bound on $\mathbb{P}(S_{\text{valid}} \cap S_{\text{consec}} \cap S_{\text{indep}} \cap S_{\text{ogp}})$, so $S_{\text{valid}} \cap S_{\text{consec}} \cap S_{\text{indep}} \cap S_{\text{ogp}} \neq \emptyset$. We will show (analogously to Proposition 2.3.6) that $S_{\text{valid}} \cap S_{\text{consec}} \cap S_{\text{indep}} \cap S_{\text{ogp}} = \emptyset$, yielding a contradiction. When $S_{\text{valid}}, S_{\text{consec}}, S_{\text{indep}}$ simultaneously hold, we will construct an example of the structure forbidden by S_{ogp}.

We will set $y^{(\ell)} = x^{(t_\ell)}$ for all $0 \le \ell \le k$, for a sequence $0 \le t_0 \le t_1 \le \cdots \le t_k \le T$ we now construct. We set $t_0 = 0$. For $1 \le \ell \le k$ we set t_ℓ to be the smallest $t > t_{\ell-1}$ such that (OGP-B) holds for $y^{(\ell)} = x^{(t)}$. We now sketch why t_ℓ exists and satisfies $t_\ell \le t_{\ell-1} + kM$. Note that this ensures all the t_ℓ are well defined because $T = k^2 M$.

By S_{consec}, $x^{(t)}$ evolves by small steps. Thus, for fixed $y^{(0)}, \dots, y^{(\ell-1)}$ and varying $y^{(\ell)} = x^{(t)}$ (varying as we increment t), the q defined above moves by small steps in $\mathcal{Q}$. Let q_t denote this q at time t. By Fact 2.3.5(c) $q_{t_{\ell-1}}(\xi) \in \{0, 1\}$ for all $\xi \in \Xi$. S_{indep} ensures that $q_{t_{\ell-1}+kM}$ is far from $q_{t_{\ell-1}}$ in $\mathcal{Q}$. The evolution of q_t from $t = t_{\ell-1}$ to $t = t_{\ell-1} + kM$ can be modeled essentially by a continuous path, and the definition of the maximin κ^{**} implies that for some t in this range, $F(q_t) \approx \kappa^{**} \le \kappa - \varepsilon$. (Although q_t does not necessarily evolve to the all-$\frac{1}{2}$ function, S_{indep} implies that it ends far from where it started, and we can show that over this evolution we already encounter q_t such that $F(q_t)$ is near the maximin value.) This shows the existence of t_ℓ with $t_\ell \le t_{\ell-1} + kM$.

Since, by S_{valid}, each $y^{(\ell)}$ ν-satisfies $\Phi^{(t_\ell)}$, we have constructed an example of the structure forbidden by S_{ogp}. This gives the desired contradiction. □

2.9.2 Suboptimality of κ^*

We believe that $\kappa^* > \kappa^{**}$ due to the following heuristic argument. The Chernoff bound (2.9.2) is tighest when most of the mass of the random variables $\frac{\min(u_i, \log k)}{\log k}$ is near 0 or 1. When this occurs, most of the the mass of u_i is near 0 or $\log k$. Then, the event that $\sum_{i=1}^k u_i \ge \log k + \log\log k$ is the event that one or two of the u_i attains a value near $\log k$. This is a tail probability in a non-asymptotic regime – approximately, the probability that a Poisson random variable is larger than 1 or 2 – so the Chernoff bound will not get the correct probability.

2.9.3 Proof that $\kappa^{**} > 1$

In this subsection, we will show that κ^{**} is bounded below by a constant larger than 1, approximately 1.716. Thus our methods cannot improve the constant κ^* in Theorems 2.1.6 and 2.1.13 to 1.

We will first show a weaker lower bound on κ^{**}. Define $\psi_1 : (0, +\infty) \to \mathbb{R}$ by

$$\psi_1(\lambda) = \frac{\lambda/2}{1 - (1+\lambda)e^{-\lambda}},$$

and let $\psi_1^* = \min_{\lambda > 0} \psi_1(\lambda) \approx 1.675$.

Proposition 2.9.3. *We have $\kappa^{**} \ge \psi_1^*$.*

Proof. We will prove this proposition by constructing a suitable function family $p \in \mathcal{P}$.

Let $\Xi = [0, 1]$ equipped with the uniform measure. Let $p : \Xi \times [0, 1] \to [0, 1]$ be defined by

$$p(\xi, s) = \begin{cases} \min(s, \frac{1}{2}) & \xi \le s, \\ 0 & \xi \ge s. \end{cases}$$

Thus, for fixed $s \in [0, 1]$, $p(\xi, s) = \min(s, \frac{1}{2})$ with probability s, and otherwise $p(\xi, s) = 0$. We will show that for this p,

$$\limsup_{k \to \infty} \min_{s \in [0,1]} F(p(\cdot, s)) \ge \psi_1^*,$$

from which the proposition follows.

Note that if $s = \omega_k(k^{-1/2})$, then $\mathop{\mathbb{E}}_{\xi} H(p(\xi, s)) = \omega_k(\log k/k)$, and so $F(p(\cdot, s)) = \omega_k(1)$. Therefore it suffices to consider $s = O_k(k^{-1/2})$. Then,

$$\mathop{\mathbb{E}}_{\xi} H(p(\xi, s)) = (1 + o_k(1))s^2 \log \frac{1}{s}.$$

We now analyze the behavior of the denominator of $F(p(\cdot, s))$. Note that a sample $u \sim D(p(\cdot, s))$ equals $\log \frac{1}{s}$ with probability s^2, $\log \frac{1}{1-s} \le \frac{s}{1-s}$ with probability $s(1-s)$, and 0 with probability $1-s$. For $i = 1, \dots, k$, define

$$v_i = \log \frac{1}{s} \mathbb{1}\left\{u_i = \log \frac{1}{s}\right\}, \qquad \text{and} \qquad w_i = \frac{s}{1-s} \mathbb{1}\left\{u_i = \log \frac{1}{1-s}\right\}.$$

So, $u_i \le v_i + w_i$. For $u_1, \dots, u_k \sim D(p(\cdot, s))^{\otimes k}$, we have

$$\mathbb{P}\left[\sum_{i=1}^{k} u_i \ge \log k + \log\log k\right] \le \mathbb{P}\left[\sum_{i=1}^{k} v_i \ge \log k\right] + \mathbb{P}\left[\sum_{i=1}^{k} w_i \ge \log\log k\right].$$

Let $1 + \delta = \frac{\log\log k}{k \,\mathbb{E}\, w_1} = \frac{\log\log k}{ks^2}$. Because $s = O_k(k^{-1/2})$, we have $1 + \delta = \omega_k(1)$, and so $\frac{\delta^2}{2+\delta} \ge \frac{1}{2}(1+\delta)$ for sufficiently large k. By a Chernoff bound,

$$\begin{aligned}\mathbb{P}\left[\sum_{i=1}^{k} w_i \ge \log\log k\right] &\le \mathbb{P}\left[\sum_{i=1}^{k} \frac{1-s}{s} w_i \ge \frac{1-s}{s}\log\log k\right] \le \exp\left(-\frac{\delta^2}{2+\delta} \cdot ks(1-s)\right) \\ &\le \exp\left(-\frac{1}{2}(1+\delta)ks(1-s)\right) \le \exp\left(-\frac{(1-s)\log\log k}{2s}\right) \le \exp\left(-\Omega_k(k^{1/2})\right).\end{aligned}$$

To analyze the other probability, we consider cases $s > \frac{1}{k}$ and $s \le \frac{1}{k}$. We first consider $s > \frac{1}{k}$. In order to have $\sum_{i=1}^{k} v_i \ge \log k$, at least two v_i must be nonzero. This occurs with probability

$$1 - (1-s^2)^k - s^2k(1-s^2)^{k-1} \le 1 - (1+s^2k)(1-s^2)^k.$$

Thus,

$$F(p(\cdot, s)) \ge \frac{1}{\log k} \cdot \frac{\log 2 + (1+o_k(1))s^2k \log \frac{1}{s}}{1 - (1+s^2k)(1-s^2)^k + \exp(-\Omega_k(k^{1/2}))}.$$

If $s^2k = o_k(1)$, then $1 - (1+s^2k)(1-s^2)^k = O_k(s^4k^2)$, and the right-hand side is $\omega_k(1)$. So, this bound is minimized at $s = \lambda k^{-1/2}$ for constant λ, in which case

$$\frac{1}{\log k} \cdot \frac{\log 2 + (1+o_k(1))s^2k \log \frac{1}{s}}{1 - (1+s^2k)(1-s^2)^k + \exp(-\Omega_k(k^{1/2}))} \to \frac{\lambda/2}{1 - (1+\lambda)\exp(-\lambda)} = \psi_1(\lambda) \ge \psi_1^*.$$

We now consider $s \le \frac{1}{k}$. In order to have $\sum_{i=1}^{k} v_i \ge \log k$, at least one v_i must be nonzero. This occurs with probability

$$1 - (1-s^2)^k = (1+o_k(1))s^2k,$$

and so

$$F(p(\cdot, s)) \ge \frac{1}{\log k} \cdot \frac{\log 2 + (1+o_k(1))s^2k \log \frac{1}{s}}{(1+o_k(1))s^2k + \exp(-\Omega_k(k^{1/2}))}.$$

The right-hand side is $\omega_k(1)$ because $s \le \frac{1}{k}$. □

For any nonnegative integer n, we may further define

$$\psi_n(\lambda) = \frac{\lambda/(n+1)}{1 - \left(\sum_{k \le n} \frac{\lambda^k}{k!}\right)\exp(-\lambda)}$$

and $\psi_n^* = \inf_{\lambda > 0} \psi_n(\lambda)$. Over positive integers n, the largest ψ_n^* is $\psi_2^* \approx 1.716$. The following corollary gives the lower bound on κ^{**} alluded to above.

Corollary 2.9.4. *We have that $\kappa^{**} \geq \psi_2^*$.*

Proof. We will construct a suitable function family p. For any nonnegative integer n, we can define

$$p_n(\xi, s) = \begin{cases} \min(s, \frac{1}{2}) & \xi \leq s^n, \\ 0 & \xi \geq s. \end{cases} \tag{2.9.3}$$

By a similar analysis to Proposition 2.9.3, we can show for this p that

$$\limsup_{k \to \infty} \min_{s \in [0,1]} F(p_n(\cdot, s)) \geq \psi_n^*.$$

Taking $n = 2$ yields the result. □

Due to Corollary 2.9.4, a proof of Theorem 2.1.6 improving the constant κ^* below ψ_2^* will require new conceptual insights. Finally, we conjecture that Corollary 2.9.4 is in fact sharp.

Conjecture 2.9.5. *We have that $\kappa^{**} = \psi_2^*$. In particular, Theorem 2.1.6 holds for all $\kappa > \psi_2^*$.*

The following evidence supports this conjecture. In the maximin problem (2.9.1), if we restrict the maximum over p to functions such that for every s, $p(\xi, s)$ attains at most one nonzero value, then we can show by explicit computation that the maximin problem has value ψ_2^*. The idea of this proof is that for each such p, at the s minimizing $F(p(\cdot, s))$, $p(\cdot, s)$ equals (up to isomorphism of the probability space (Ξ, P_ξ)) $p_n(\cdot, s')$ for some s' and some (possibly fractional) n. We can show that fractional n do not maximize $\min_{s \in [0,1]} F(p_n(\cdot, s))$. Thus the candidate maxima are p_n for integer n, and of these p_2 is maximal, attaining value ψ_2^*. We believe that the maximum of (2.9.1) over $p \in \mathcal{P}$ is attained by p with this property.

Chapter 3

Mean-Field Spin Glasses

This chapter studies the optimization of mean-field spin glasses and is structured as follows.

- Section 3.1 states our main hardness result, Theorem 3.1.3, that overlap concentrated algorithms cannot surpass the objective ALG defined in (1.3.9) (and for spherical spin glasses, (3.1.5)) with non-negligible probability.
- In Section 3.2 we formulate Proposition 3.2.2, which establishes the main branching OGP, and prove Theorem 3.1.3 assuming this proposition.
- Sections 3.3 through 3.5 prove Proposition 3.2.2 using a many-replica version of the Guerra-Talagrand interpolation.
- Section 3.6 shows that (for spherical models without external field) the full strength of our branching OGP is necessary to show tight algorithmic hardness.
- Section 3.7 shows that approximately Lipschitz algorithms are overlap concentrated, and that natural optimization algorithms including gradient descent, AMP, and Langevin dynamics are approximately Lipschitz. This includes, in particular, the optimal AMP algorithms of [Mon19, AMS21, Sel21].

Notation and Preliminaries

We generally use ordinary lower-case letters $(x, y, \dots)$ for scalars and bold lower-case $(\boldsymbol{x}, \boldsymbol{y}, \dots)$ for vectors. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$, we denote the ordinary inner product by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^N x_i y_i$ and the normalized inner product by $R(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{N}\langle \boldsymbol{x}, \boldsymbol{y} \rangle$. We associate with these inner products the norms $\|\boldsymbol{x}\|_2^2 = \langle \boldsymbol{x}, \boldsymbol{x} \rangle$ and $\|\boldsymbol{x}\|_N^2 = R(\boldsymbol{x}, \boldsymbol{x})$. There is no confusion between the $\|\cdot\|_N$ norm and the ℓ_p norm, which will not appear in this thesis.

Ensembles of scalars over an index set $\mathbb{L}$ are denoted with an arrow $(\vec{x}, \vec{y}, \dots)$, and the entry of $\vec{x}$ indexed by $u \in \mathbb{L}$ is denoted $x(u)$. Similarly, ensembles of vectors are written in bold and with an arrow $(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}, \dots)$, and the entry of $\vec{\boldsymbol{x}}$ indexed by $u \in \mathbb{L}$ are denoted $\boldsymbol{x}(u)$. Sequences of scalars parametrizing these ensembles are also denoted with an arrow, for example $\vec{k} = (k_1, \dots, k_D)$.

We reiterate that $S_N = \{\boldsymbol{x} \in \mathbb{R}^N : \sum_{i=1}^N \boldsymbol{x}_i^2 = N\}$ and $\Sigma_N = \{-1, 1\}^N$, and that $B_N = \{\boldsymbol{x} \in \mathbb{R}^N : \sum_{i=1}^N \boldsymbol{x}_i^2 \le N\}$ and $C_N = [-1, 1]^N$ are their convex hulls. The space of Hamiltonians H_N is denoted $\mathscr{H}_N$. We identify each Hamiltonian H_N with its disorder coefficients $(\mathbf{G}^{(p)})_{p \in 2\mathbb{N}}$, which we concatenate into a vector $\boldsymbol{g} = \boldsymbol{g}(H_N)$.

For any tensor $A_p \in (\mathbb{R}^N)^{\otimes p}$, where $p \ge 1$, we define the operator norm

$$\|A_p\|_{\mathrm{op}} = \frac{1}{N} \max_{\boldsymbol{\sigma}^1, \dots, \boldsymbol{\sigma}^p \in S_N} \left| \langle A_p, \boldsymbol{\sigma}^1 \otimes \cdots \otimes \boldsymbol{\sigma}^p \rangle \right|.$$

Note that when $p = 1$, $\|A_p\|_{\mathrm{op}} = \|A_p\|_N$. The following proposition shows that with exponentially high probability, the operator norms of all constant-order gradients of H_N are bounded and $O(1)$-Lipschitz. We will prove this proposition in Section 3.8.

Proposition 3.0.1. *For fixed model (ξ, h) and $r \in [1, \sqrt{2})$, there exists a constant $c > 0$, sequence $(K_N)_{N \geq 1}$ of sets $K_N \subseteq \mathscr{H}_N$, and sequence of constants $(C_k)_{k \geq 1}$ independent of N, such that the following properties hold.*

1. $\mathbb{P}[H_N \in K_N] \geq 1 - e^{-cN}$;
2. *If $H_N \in K_N$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ satisfy $\|\boldsymbol{x}\|_N, \|\boldsymbol{y}\|_N \leq r$, then*

$$\left\|\nabla^k H_N(\boldsymbol{x})\right\|_{\mathrm{op}} \leq C_k, \tag{3.0.1}$$

$$\left\|\nabla^k H_N(\boldsymbol{x}) - \nabla^k H_N(\boldsymbol{y})\right\|_{\mathrm{op}} \leq C_{k+1} \|\boldsymbol{x} - \boldsymbol{y}\|_N . \tag{3.0.2}$$

Finally, the notations $O(\cdot), \Omega(\cdot), o(\cdot), \omega(\cdot)$ indicate asymptotic behavior in N.

3.1 Results

3.1.1 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 (1.3.2). Let

$$\widetilde{H}_N^{(1)} = \sqrt{p}\widetilde{H}_N^{[0]} + \sqrt{1-p}\widetilde{H}_N^{[1]} \qquad \text{and} \qquad \widetilde{H}_N^{(2)} = \sqrt{p}\widetilde{H}_N^{[0]} + \sqrt{1-p}\widetilde{H}_N^{[2]}$$

and define

$$H_N^{(1)}(\boldsymbol{\sigma}) = \langle \boldsymbol{h}, \boldsymbol{\sigma} \rangle + \widetilde{H}_N^{(1)}(\boldsymbol{\sigma}) \qquad \text{and} \qquad H_N^{(2)}(\boldsymbol{\sigma}) = \langle \boldsymbol{h}, \boldsymbol{\sigma} \rangle + \widetilde{H}_N^{(2)}(\boldsymbol{\sigma}).$$

We say the pair of Hamiltonians $H_N^{(1)}, H_N^{(2)}$ is p-correlated. Note that pairs of corresponding entries in $\boldsymbol{g}^{(1)} = \boldsymbol{g}(H_N^{(1)})$ and $\boldsymbol{g}^{(2)} = \boldsymbol{g}(H_N^{(2)})$ are Gaussian with covariance $\left[\begin{smallmatrix} 1 & p \\ p & 1 \end{smallmatrix}\right]$.

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

Definition 3.1.1. 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 $H_N^{(1)}, H_N^{(2)}$,

$$\mathbb{P}\left[\left| R\left(\mathcal{A}(H_N^{(1)}), \mathcal{A}(H_N^{(2)})\right) - \mathbb{E}\, R\left(\mathcal{A}(H_N^{(1)}), \mathcal{A}(H_N^{(2)})\right)\right| \geq \lambda\right] \leq \nu. \tag{3.1.1}$$

3.1.2 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 (1.3.6). Similarly to Theorem 1.3.1, Auffinger and Chen [AC17a], see also [CS17], 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 (1.3.3). 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

$$\mathsf{P}^{\mathrm{Sp}}(B, \zeta) = \frac{1}{2}\left[\frac{h^2}{B_\zeta(0)} + \int_0^1 \left(\frac{\xi''(t)}{B_\zeta(t)} + B_\zeta(t)\right)\, \mathrm{d}t\right], \tag{3.1.2}$$

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

$$B_\zeta(t) = B - \int_t^1 \xi''(q)\zeta(q)\, \mathrm{d}q. \tag{3.1.3}$$

Theorem 3.1.2 ([AC17a, Theorem 10])**.** *The following identity holds.*

$$\mathsf{OPT}^{\mathrm{Sp}} = \inf_{(B,\zeta) \in \mathscr{V}(\xi)} \mathsf{P}^{\mathrm{Sp}}(B, \zeta). \tag{3.1.4}$$

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

3.1.3 The Optimal Energy of Overlap Concentrated Algorithms

We defined $\mathsf{ALG}^{\mathrm{Is}}$ in (1.3.9) by a non-monotone extension of the variational formula in (1.3.7). We can similarly define $\mathsf{ALG}^{\mathrm{Sp}}$ by a non-monotone extension of (3.1.4). Recall the set $\mathscr{L}$ defined in (1.3.8). 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 can therefore define $\mathsf{ALG}^{\mathrm{Sp}} = \mathsf{ALG}^{\mathrm{Sp}}_{\xi,h}$ by

$$\mathsf{ALG}^{\mathrm{Sp}} = \inf_{(B,\zeta)\in\mathscr{K}(\xi)} \mathsf{P}^{\mathrm{Sp}}(B, \zeta). \tag{3.1.5}$$

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 will show that for any mixed even spherical or Ising spin glass, no overlap concentrated algorithm can attain an energy level above the algorithmic thresholds $\mathsf{ALG}^{\mathrm{Sp}}$ and $\mathsf{ALG}^{\mathrm{Is}}$ with nontrivial probability.

Theorem 3.1.3 (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 any $N \ge N_0$ and any $\nu \in [0, 1]$. For any (λ, ν) overlap concentrated $\mathcal{A} = \mathcal{A}_N : \mathscr{H}_N \to B_N$ (resp. C_N),*

$$\mathbb{P}\left[\frac{1}{N} H_N\left(\mathcal{A}(H_N)\right) \ge \mathsf{ALG} + \varepsilon\right] \le \exp(-cN) + 3(\nu/\lambda)^c.$$

Remark 3.1.4. 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, see Proposition 3.7.2. 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, see Proposition 3.7.3.

In tandem with Theorem 1.3.2 and its spherical analogue Theorem 3.1.5 below, Theorem 3.1.3 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 1.3.2). We will see in Section 3.7 that the algorithms in these two theorems are overlap concentrated.

Theorem 3.1.5 ([AMS21, Sel21])**.** *For any $\varepsilon > 0$, there exists an efficient and $O_\varepsilon(1)$-Lipschitz AMP algorithm $\mathcal{A} : \mathscr{H}_N \to B_N$ such that*

$$\mathbb{P}[H_N(\mathcal{A}(H_N))/N \ge \mathsf{ALG}^{\mathrm{Sp}} - \varepsilon] \ge 1 - o(1), \quad c = c(\varepsilon) > 0.$$

In fact, we can show the success probability of the algorithm in Theorem 3.1.5 (resp. Theorem 1.3.2) is exponentially high. By Theorem 3.7.6, for this algorithm $\mathcal{A}$ the map $H_N \mapsto \frac{1}{N} H_N(\mathcal{A}(H_N))$ is $O(N^{-1/2})$-Lipschitz with respect to the $\|\cdot\|_2$ norm when restricted to $H_N \in K_N$ (see Proposition 3.0.1). By Kirszbraun's Theorem, $H_N \mapsto \frac{1}{N} H_N(\mathcal{A}(H_N))$ agrees on K_N with an $O(N^{-1/2})$-Lipschitz $\mathcal{A}' : \mathscr{H}_N \to \mathbb{R}$. By Gaussian concentration and the bound $\mathbb{P}[H_N \in K_N] \ge 1 - e^{-cN}$, (after adjusting ε) we have $\frac{1}{N} H_N(\mathcal{A}(H_N)) \ge \mathsf{ALG}^{\mathrm{Sp}} - \varepsilon$ (resp. $\mathsf{ALG}^{\mathrm{Is}} - \varepsilon$) with probability at least $1 - e^{-c_\varepsilon N}$ for a constant $c_\varepsilon > 0$.

In the case of the spherical spin glass, the value of $\mathsf{ALG}^{\mathrm{Sp}}$ is explicit, and is given by the following proposition. We will prove this proposition in Section 3.9.

Proposition 3.1.6. *If $h^2 + \xi'(1) \ge \xi''(1)$, then*

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

and the infimum in (3.1.5) 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* (3.1.5) *is uniquely attained by* $B = \xi''(1)^{1/2}$ *and*

$$\zeta(q) = \mathbb{1}\{q \geq \widehat{q}\} \frac{\xi'''(q)}{2\xi''(q)^{3/2}} = -\mathbb{1}\{q \geq \widehat{q}\} \frac{\mathrm{d}}{\mathrm{d}q}\xi''(q)^{-1/2}. \tag{3.1.6}$$

If $h = 0$, *the infimum is not attained. It is achieved by* $B = \xi''(1)^{1/2}$ *and* ζ *given by* (3.1.6) *in the limit as* $\widehat{q} \to 0^+$.[1]

Note that $\mathsf{ALG}^{\mathrm{Sp}} = \mathsf{OPT}^{\mathrm{Sp}}$ if and only if the infimum in (3.1.5) is attained at a pair $(B, \zeta) \in \mathscr{V}(\xi)$. Thus, Proposition 3.1.6 implies that $\mathsf{ALG}^{\mathrm{Sp}} = \mathsf{OPT}^{\mathrm{Sp}}$ if and only if $h^2 + \xi'(1) \geq \xi''(1)$ or $\xi''(q)^{-1/2}$ is concave on $[\widehat{q}, 1]$. In the former case, the model is replica symmetric at zero temperature; in the latter case it is full replica symmetry breaking on $[\widehat{q}, 1]$ at zero temperature.

In the case $h^2 + \xi'(1) > \xi''(1)$, [Fyo13, BČNS21] showed that H_N has trivial complexity, in the sense that H_N has exactly two critical points on S_N with high probability, namely its global maximum and minimum.

In the pure p-spin case, $h = 0$ and $\xi(x) = x^p$ for an even $p \geq 4$. The corresponding algorithmic limit is

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

This coincides with the threshold $E_\infty(p)$ identified in [ABAČ13]. As conjectured in [ABAČ13] and proved in [Sub17], 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 3.1.7. 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 $\|\boldsymbol{h}\|_N^2$ in (3.1.2) and replace $\Phi(0, h)$ by $\frac{1}{N}\sum_{i=1}^N \Phi(0, h_i)$ in (1.3.6). This includes for instance the natural case of Gaussian external field. Here $\mathcal{A}$ can depend arbitrarily on $\boldsymbol{h}$ as long as overlap concentration holds conditionally on $\boldsymbol{h}$.

3.2 Proof of Main Impossibility Result

In this section, we prove Theorem 3.1.3 assuming Proposition 3.2.2, which establishes the main OGP. Throughout, we fix a model (ξ, h) and $\varepsilon > 0$. Let H_N be a Hamiltonian (1.3.1) with model (ξ, h). Let $\lambda > 0$ be a constant we will set later, and let $\mathcal{A} : \mathscr{H}_N \to B_N$ (resp. C_N) be (λ, ν) overlap concentrated.

3.2.1 The Correlation Function

We define the correlation function $\chi : [0,1] \to \mathbb{R}$ by

$$\chi(p) = \mathbb{E}\, R\left(\mathcal{A}(H_N^{(1)}), \mathcal{A}(H_N^{(2)})\right), \tag{3.2.1}$$

where $H_N^{(1)}, H_N^{(2)}$ are p-correlated copies of H_N. The following proposition establishes several properties of correlation functions, which we will later exploit.

Proposition 3.2.1. *The correlation function* χ *has the following properties.*

(i) For all $p \in [0,1]$, $\chi(p) \in [0,1]$.

(ii) χ *is either strictly increasing or constant on* $[0,1]$.

(iii) For all $p \in [0,1]$, $\chi(p) \leq (1-p)\chi(0) + p\chi(1)$.

We call any $\chi : [0,1] \to \mathbb{R}$ satisfying the conclusions of Proposition 3.2.1 a *correlation function*.

[1] When $h = 0$, we cannot take $\widehat{q} = 0$ in (3.1.6) because then $B = \int_0^1 \xi''(q)\zeta(q)\, \mathrm{d}q$, so $(B, \zeta) \notin \mathscr{K}(\xi)$.

Proof. In this proof, we will write $\mathcal{A}(\boldsymbol{g})$ to mean $\mathcal{A}(H_N)$ for the Hamiltonian H_N with disorder coefficients $\boldsymbol{g} = \boldsymbol{g}(H_N)$. We introduce the Fourier expansion of $\mathcal{A}$. For each nonnegative integer j, let He_j denote the j-th univariate Hermite polynomial. These are defined by $\mathrm{He}_0(x) = 1$ and for $n \geq 0$,

$$\mathrm{He}_{n+1}(x) = x\mathrm{He}_n(x) - \mathrm{He}_n'(x).$$

Recall that the renormalized Hermite polynomials $\widetilde{\mathrm{He}}_n = \frac{1}{\sqrt{n!}}\mathrm{He}_n$ form an orthonormal basis of $L^2(\mathbb{R})$ with the standard Gaussian measure, i.e. they form a complete basis and satisfy

$$\mathop{\mathbb{E}}_{g\sim\mathcal{N}(0,1)} \widetilde{\mathrm{He}}_n(g)\widetilde{\mathrm{He}}_m(g) = \mathbb{1}\{n = m\}.$$

For each multi-index $\alpha = (\alpha_1, \alpha_2, \dots,)$ of nonnegative integers that are eventually zero, define the multivariate Hermite polynomial

$$\widetilde{\mathrm{He}}_\alpha(\boldsymbol{g}) = \prod_i \widetilde{\mathrm{He}}_{\alpha_i}(\boldsymbol{g}_i),$$

These polynomials form an orthonormal basis of $L^2(\mathbb{R}^{\mathbb{N}})$ with the standard Gaussian measure, see e.g. [LMP15, Theorem 8.1.7]. Hence for each $1 \leq i \leq N$, we can write

$$\mathcal{A}_i(\boldsymbol{g}) = \sum_\alpha \widehat{\mathcal{A}}_i(\alpha)\widetilde{\mathrm{He}}_\alpha(\boldsymbol{g}) \qquad \text{where} \qquad \widehat{\mathcal{A}}_i(\alpha) = \mathbb{E}\left[\mathcal{A}_i(\boldsymbol{g})\widetilde{\mathrm{He}}_\alpha(\boldsymbol{g})\right].$$

For each multi-index α, let $|\alpha| = \sum_{i\geq 1}\alpha_i$. For each nonnegative integer j, introduce the Fourier weight

$$W_j = \frac{1}{N}\sum_{i=1}^N \sum_{|\alpha|=j} \widehat{\mathcal{A}}_i(\alpha)^2 \geq 0.$$

For $i = 1, 2$, let $\boldsymbol{g}^{(i)} = \boldsymbol{g}(H_N^{(i)})$. Let T_p denote the Ornstein-Uhlenbeck operator. We compute that

$$\begin{aligned}\chi(p) &= \frac{1}{N}\mathbb{E}\left\langle \mathcal{A}(\boldsymbol{g}^{(1)}), \mathcal{A}(\boldsymbol{g}^{(2)})\right\rangle = \frac{1}{N}\mathbb{E}\left\langle \mathcal{A}(\boldsymbol{g}), T_p\mathcal{A}(\boldsymbol{g})\right\rangle = \frac{1}{N}\mathbb{E}\left\|T_{\sqrt{p}}\mathcal{A}(\boldsymbol{g})\right\|_2^2 \\ &= \frac{1}{N}\sum_{i=1}^N \left\|T_{\sqrt{p}}\mathcal{A}_i(\boldsymbol{g})\right\|_2^2 = \frac{1}{N}\sum_{i=1}^N\sum_\alpha p^{|\alpha|}\widehat{\mathcal{A}}_i(\alpha)^2 = \sum_{j\geq 0} p^j W_j.\end{aligned}$$

It is now clear that $0 \leq \chi(p) \leq \chi(1)$. Since $\chi(1) = \mathbb{E}\left\|\mathcal{A}(H_N)\right\|_N^2 \leq 1$, this proves the first claim. The second claim follows because $\chi(p)$ is strictly increasing unless $W_j = 0$ for all $j \geq 1$, in which case $\chi(p)$ is constant. Finally, the last claim follows since χ is manifestly convex. □

3.2.2 Hierarchically Correlated Hamiltonians

Here we define the hierarchically organized ensemble of correlated Hamiltonians that will play a central role in our proofs of impossibility. Let D be a nonnegative integer and $\vec{k} = (k_1, \dots, k_D)$ for positive integers $k_1, \dots, k_D$. For each $0 \leq d \leq D$, let $V_d = [k_1] \times \cdots \times [k_d]$ denote the set of length d sequences with j-th element in $[k_j]$. The set V_0 consists of the empty tuple, which we denote $\emptyset$. Let $\mathbb{T}(\vec{k})$ denote the depth D tree rooted at $\emptyset$ with depth d vertex set V_d, where $u \in V_d$ is the parent of $v \in V_{d+1}$ if u is an initial substring of v. For nodes $u^1, u^2 \in \mathbb{T}(\vec{k})$, let

$$u^1 \wedge u^2 = \max\left\{d \in \mathbb{Z}_{\geq 0} : u^1_{d'} = u^2_{d'} \text{ for all } 1 \leq d' \leq d\right\},$$

where the set on the right-hand side always contains 0 vacuously. This is the depth of the least common ancestor of u^1 and u^2. Let $\mathbb{L}(\vec{k}) = V_D$ denote the set of leaves of $\mathbb{T}(\vec{k})$. When $\vec{k}$ is clear from context, we denote $\mathbb{T}(\vec{k})$ and $\mathbb{L}(\vec{k})$ by $\mathbb{T}$ and $\mathbb{L}$. Finally, let $K = |\mathbb{L}| = \prod_{d=1}^D k_d$.

Let sequences $\vec{p} = (p_0, p_1, \dots, p_D)$ and $\vec{q} = (q_0, q_1, \dots, q_D)$ satisfy

$$\begin{aligned}0 = p_0 \leq p_1 \leq \cdots \leq p_D = 1,\\ 0 \leq q_0 < q_1 < \cdots < q_D = 1.\end{aligned}$$

The sequence $\vec{p}$ controls the correlation structure of our ensemble of Hamiltonians, while the sequence $\vec{q}$ controls the overlap structure that we will require the inputs to these Hamiltonians to have.

We now construct an ensemble of Hamiltonians $(H_N^{(u)})_{u\in\mathbb{L}}$, such that each $H_N^{(u)}$ is marginally distributed as H_N and each pair of Hamiltonians $H_N^{(u^1)}, H_N^{(u^2)}$ is $p_{u^1\wedge u^2}$-correlated. For each $u \in \mathbb{T}$, including non-leaf nodes, let $\widetilde{H}_N^{[u]}$ be an independent copy of $\widetilde{H}_N$, generated by (1.3.2). For each $u \in \mathbb{L}$, we construct

$$\begin{aligned} H_N^{(u)}(\boldsymbol{\sigma}) &= \langle \boldsymbol{h}, \boldsymbol{\sigma}\rangle + \widetilde{H}_N^{(u)}(\boldsymbol{\sigma}), \quad \text{where} \\ \widetilde{H}_N^{(u)} &= \sum_{d=1}^{D} \sqrt{p_d - p_{d-1}} \cdot \widetilde{H}_N^{[(u_1,\ldots,u_d)]}. \end{aligned} \tag{3.2.2}$$

It is clear that this ensemble has the stated properties. Consider a state space of K-tuples

$$\vec{\boldsymbol{\sigma}} = (\boldsymbol{\sigma}(u))_{u\in\mathbb{L}} \in (\mathbb{R}^N)^K.$$

We define a grand Hamiltonian on this state space by

$$\mathcal{H}_N^{\vec{k},\vec{p}}(\vec{\boldsymbol{\sigma}}) \equiv \sum_{u\in\mathbb{L}} H_N^{(u)}(\boldsymbol{\sigma}(u)).$$

We will denote this by $\mathcal{H}_N$ when $\vec{k}, \vec{p}$ are clear from context. For states $\vec{\boldsymbol{\sigma}}^1, \vec{\boldsymbol{\sigma}}^2 \in (\mathbb{R}^N)^K$, define the overlap matrix $R = R(\vec{\boldsymbol{\sigma}}^1, \vec{\boldsymbol{\sigma}}^2) \in \mathbb{R}^{K\times K}$ by

$$R_{u^1,u^2} = R(\boldsymbol{\sigma}^1(u^1), \boldsymbol{\sigma}^2(u^2))$$

for all $u^1, u^2 \in \mathbb{L}$. We now define an overlap matrix $Q = Q^{\vec{k},\vec{q}} \in \mathbb{R}^{K\times K}$; we will control the maximum energy of $\mathcal{H}_N$ over inputs $\vec{\boldsymbol{\sigma}}$ with approximately this self-overlap. Let Q have rows and columns indexed by $u^1, u^2 \in \mathbb{L}$ and entries

$$Q_{u^1,u^2} = q_{u^1\wedge u^2}.$$

Fix a point $\boldsymbol{m} \in \mathbb{R}^N$ such that $\|\boldsymbol{m}\|_N^2 = q_0$, which we will later take to be $\boldsymbol{m} = \mathbb{E}[\mathcal{A}(H_N)]$. For a tolerance $\eta \in (0,1)$, define the band

$$B(\boldsymbol{m},\eta) = \left\{\boldsymbol{\sigma} \in \mathbb{R}^N : |R(\boldsymbol{\sigma},\boldsymbol{m}) - q_0| \le \eta\right\}.$$

Define the sets of points in S_N^K and Σ_N^K with self-overlap approximately Q and overlap with $\boldsymbol{m}$ approximately q_0 by

$$\begin{aligned} \mathcal{Q}^{\mathrm{Sp}}(Q,\boldsymbol{m},\eta) &= \left\{\vec{\boldsymbol{\sigma}} \in (S_N \cap B(\boldsymbol{m},\eta))^K : \|R(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}}) - Q\|_\infty \le \eta\right\}, \\ \mathcal{Q}^{\mathrm{Is}}(Q,\boldsymbol{m},\eta) &= \left\{\vec{\boldsymbol{\sigma}} \in (\Sigma_N \cap B(\boldsymbol{m},\eta))^K : \|R(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}}) - Q\|_\infty \le \eta\right\}. \end{aligned}$$

Let χ be a correlation function (recall Proposition 3.2.1). We say $\vec{p} = (p_0,\ldots,p_D)$ and $\vec{q} = (q_0,\ldots,q_D)$ are *χ-aligned* if the following properties hold for all $0 \le d \le D$.

- If $q_d \le \chi(1)$, then $\chi(p_d) = q_d$.
- If $q_d > \chi(1)$, then $p_d = 1$.

The following proposition controls the expected maximum energy of the grand Hamiltonian constrained on the sets $\mathcal{Q}^{\mathrm{Sp}}(Q,\boldsymbol{m},\eta)$ and $\mathcal{Q}^{\mathrm{Is}}(Q,\boldsymbol{m},\eta)$, and is the main ingredient in our proof of impossibility. We defer the proof of this proposition to Sections 3.3 through 3.5.

Proposition 3.2.2. *For any mixed even model (ξ,h) and $\varepsilon > 0$, there exists a small constant $\eta_0 \in (0,1)$ and large constants $N_0, K_0 > 0$, dependent only on ξ, h, ε, such that for all $N \ge N_0$ the following holds.*

Let $\mathsf{ALG} = \mathsf{ALG}^{\mathrm{Sp}}$ (resp. $\mathsf{ALG}^{\mathrm{Is}}$). For any correlation function χ and vector $\boldsymbol{m} \in \mathbb{R}^N$ with $\|\boldsymbol{m}\|_N^2 = \chi(0)$, there exist $D, \vec{k}, \vec{p}, \vec{q}, \eta$ as above such that $\vec{p}$ and $\vec{q}$ are χ-aligned, $\eta \ge \eta_0$, $K \le K_0$, and

$$\frac{1}{N}\,\mathbb{E}\max_{\vec{\boldsymbol{\sigma}}\in\mathcal{Q}(\eta)} \mathcal{H}_N(\vec{\boldsymbol{\sigma}}) \le K(\mathsf{ALG} + \varepsilon),$$

where $\mathcal{Q}(\eta) = \mathcal{Q}^{\mathrm{Sp}}(Q,\boldsymbol{m},\eta)$ (resp. $\mathcal{Q}^{\mathrm{Is}}(Q,\boldsymbol{m},\eta)$).

3.2.3 Extending a Branching Tree to S_N and Σ_N

To account for the possibility that $\mathcal{A}$ outputs solutions in B_N (resp. C_N) not in S_N (resp. Σ_N), we will show that a branching tree of solutions in B_N (resp. C_N) output by $\mathcal{A}$ can always be extended into a branching tree of solutions in S_N (resp. Σ_N), with only a small cost to the energies attained.

Consider χ-aligned $\vec{p}, \vec{q}$ as above. Let $\underline{D} \le D$ be the smallest integer such that $p_{\underline{D}} = 1$. Define $\vec{\underline{k}} = (k_1, \dots, k_{\underline{D}})$, $\vec{\underline{p}} = (p_0, \dots, p_{\underline{D}})$, and $\vec{\underline{q}} = (q_0, \dots, q_{\underline{D}})$. Let $\underline{\mathbb{L}} = V_{\underline{D}}$ denote the nodes of $\mathbb{T}$ at depth $\underline{D}$, and let $\underline{K} = |\underline{\mathbb{L}}| = \prod_{d=1}^{\underline{D}} k_d$.

Consider an analogous state space of $\underline{K}$-tuples

$$\vec{\underline{\boldsymbol{\sigma}}} = (\underline{\boldsymbol{\sigma}}(\underline{u}))_{\underline{u} \in \underline{\mathbb{L}}} \in (\mathbb{R}^N)^{\underline{K}}.$$

Define $\underline{Q} = \underline{Q}^{\vec{\underline{k}}, \vec{\underline{q}}} \in \mathbb{R}^{\underline{K} \times \underline{K}}$ analogously as the matrix indexed by $\underline{u}^1, \underline{u}^2 \in \underline{\mathbb{L}}$, where

$$\underline{Q}_{\underline{u}^1, \underline{u}^2} = q_{\underline{u}^1 \wedge \underline{u}^2} \wedge \chi(1).$$

Note that because $\vec{p}, \vec{q}$ are χ-aligned, $q_{\underline{D}-1} < \chi(1) \le q_{\underline{D}}$. So, the right-hand side is $\chi(1)$ if $\underline{u}^1 \wedge \underline{u}^2 = \underline{D}$ (i.e. $\underline{u}^1 = \underline{u}^2$) and $q_{\underline{u}^1 \wedge \underline{u}^2}$ otherwise. The following sets capture the overlap structure of outputs of $\mathcal{A}$.

$$\underline{\mathcal{Q}}^{\mathrm{Sp}}(\underline{Q}, \boldsymbol{m}, \eta) = \left\{ \vec{\underline{\boldsymbol{\sigma}}} \in (B_N \cap B(\boldsymbol{m}, \eta))^{\underline{K}} : \left\| R(\vec{\underline{\boldsymbol{\sigma}}}, \vec{\underline{\boldsymbol{\sigma}}}) - \underline{Q} \right\|_\infty \le \eta \right\},$$
$$\underline{\mathcal{Q}}^{\mathrm{Is}}(\underline{Q}, \boldsymbol{m}, \eta) = \left\{ \vec{\underline{\boldsymbol{\sigma}}} \in (C_N \cap B(\boldsymbol{m}, \eta))^{\underline{K}} : \left\| R(\vec{\underline{\boldsymbol{\sigma}}}, \vec{\underline{\boldsymbol{\sigma}}}) - \underline{Q} \right\|_\infty \le \eta \right\}.$$

By the construction (3.2.2), for each $\underline{u} \in \underline{\mathbb{L}}$ the Hamiltonians

$$\left\{ H_N^{(u)} : u \in \mathbb{L} \text{ is a descendant of } \underline{u} \text{ in } \mathbb{T} \right\}$$

are equal almost surely. Let $H_N^{(\underline{u})}$ denote any representative from this set.

We next define the condition S_{eigen} which guarantees existence of a suitable "extension" $\vec{\boldsymbol{\sigma}}$ of $\vec{\underline{\boldsymbol{\sigma}}} = \left(\mathcal{A}(H_N^{(\underline{u})}) \right)_{\underline{u} \in \underline{\mathbb{L}}}$. First, given a subset $S \subseteq [N]$, denote by W_S the $|S|$ dimensional subspace spanned by the elementary basis vectors $\{e_s : s \in S\}$. Below, λ_j denotes the j-th largest eigenvalue and $(\cdot)|_{W_S}$ denotes restriction to the subspace W_S as a bilinear form, or equivalently $A|_{W_S} = P_{W_S} A P_{W_S}$, where P_{W_S} is the projection onto W_S.

Definition 3.2.3. For constants δ and K, let $S_{\mathrm{eigen}}(\delta, K)$ denote the event that both of the below hold for all $\underline{u} \in \underline{\mathbb{L}}$.

1. $\lambda_{2K+1}\left(\nabla^2 H_N^{(\underline{u})}(\boldsymbol{x})|_{W_S} \right) \ge 0$ for all $S \subseteq [N]$ of size $|S| \ge \delta N$.
2. $H_N^{(\underline{u})} \in K_N$, for the K_N given by Proposition 3.0.1.

We will use the following lemma, whose proof is deferred to Subsection 3.2.6.

Lemma 3.2.4. *Fix a model ξ, h, constants $\varepsilon, \eta > 0$, and $\vec{k}, \vec{q}$ as above. Let δ be sufficiently small depending on $\xi, h, \eta, \varepsilon$, and assume that $S_{\mathrm{eigen}}(\delta, K)$ holds. For any $\vec{\underline{\boldsymbol{\sigma}}} \in \underline{\mathcal{Q}}(\eta/2)$, there exists $\vec{\boldsymbol{\sigma}} \in \mathcal{Q}(\eta)$ such that*

$$H_N^{(\underline{u})}(\boldsymbol{\sigma}(u)) \ge H_N^{(\underline{u})}(\underline{\boldsymbol{\sigma}}(\underline{u})) - N\varepsilon$$

whenever $\underline{u} \in \underline{\mathbb{L}}$ is an ancestor of $u \in \mathbb{L}$.

3.2.4 Completion of the Proof

We will now finish the proof of Theorem 3.1.3. Below we give the proof in the spherical setting; the Ising case follows verbatim up to replacing B_N by C_N and $\mathsf{ALG}^{\mathrm{Sp}}$ by $\mathsf{ALG}^{\mathrm{Is}}$ (since $C_N \subseteq B_N$).

Let $\mathsf{ALG} = \mathsf{ALG}^{\mathrm{Sp}}$. Let χ be the correlation function of $\mathcal{A}$ defined in (3.2.1) and set $\boldsymbol{m} = \mathbb{E}[\mathcal{A}(H_N)]$. Note that $\|\boldsymbol{m}\|_N^2 = \chi(0)$ by definition. For small $\varepsilon/2 > 0$ there exist N_0, K_0, η_0 and $D, \vec{k}, \vec{p}, \vec{q}, \eta, K$ as in Proposition 3.2.2 such that

$$\frac{1}{N}\,\mathbb{E}\max_{\vec{\boldsymbol{\sigma}}\in\mathcal{Q}(\eta)} \mathcal{H}_N(\vec{\boldsymbol{\sigma}}) \le K(\mathsf{ALG} + \varepsilon/2). \tag{3.2.3}$$

For $N \ge N_0$ let

$$\alpha_N = \mathbb{P}\left[\frac{1}{N} H_N(\mathcal{A}(H_N)) \ge \mathsf{ALG} + \varepsilon\right].$$

For each $\underline{u} \in \underline{\mathbb{L}}$, let $\underline{\boldsymbol{\sigma}}(\underline{u}) = \mathcal{A}(H_N^{(\underline{u})})$, and let $\vec{\underline{\boldsymbol{\sigma}}} = (\underline{\boldsymbol{\sigma}}(\underline{u}))_{\underline{u}\in\underline{\mathbb{L}}}$. We define the following events, where $\delta > 0$ is chosen so that Lemma 3.2.4 holds with parameters $\varepsilon/4, \eta, \vec{k}, \vec{q}$. In the statement of Theorem 3.1.3, we take $\lambda = \eta_0/4 \le \eta/4$.

Define the following events.

$$\begin{aligned}
S_{\mathrm{solve}} &= \left\{\frac{1}{N} H_N^{(\underline{u})}(\underline{\boldsymbol{\sigma}}(\underline{u})) \ge \mathsf{ALG} + \varepsilon \text{ for all } \underline{u} \in \underline{\mathbb{L}}(\vec{k})\right\},\\
S_{\mathrm{overlap}} &= \{\vec{\underline{\boldsymbol{\sigma}}} \in \underline{\mathcal{Q}}(\eta/2)\},\\
S_{\mathrm{eigen}} &= \{S_{\mathrm{eigen}}(\delta, K)\},\\
S_{\mathrm{ogp}} &= \left\{\frac{1}{N}\max_{\vec{\boldsymbol{\sigma}}\in\mathcal{Q}(\eta)} \mathcal{H}_N(\vec{\boldsymbol{\sigma}}) < K(\mathsf{ALG} + 3\varepsilon/4)\right\}.
\end{aligned}$$

Proposition 3.2.5. *With parameters as above,*

$$S_{\mathrm{solve}} \cap S_{\mathrm{overlap}} \cap S_{\mathrm{eigen}} \cap S_{\mathrm{ogp}} = \emptyset.$$

Proof. Suppose that the first three events hold. Then $\mathcal{A}$ outputs $\vec{\underline{\boldsymbol{\sigma}}} \in \underline{\mathcal{Q}}(\eta/2)$ such that for all $\underline{u} \in \underline{\mathbb{L}}$,

$$H_N^{(\underline{u})}(\underline{\boldsymbol{\sigma}}(\underline{u})) \ge \mathsf{ALG} + \varepsilon.$$

Lemma 3.2.4 now implies the existence of $\vec{\boldsymbol{\sigma}} \in \mathcal{Q}(\eta)$ such that for all $\underline{u} \in \underline{\mathbb{L}}$,

$$H_N^{(u)}(\boldsymbol{\sigma}(u)) \ge \mathsf{ALG} + 3\varepsilon/4.$$

This contradicts S_{ogp}. □

Proposition 3.2.6. *The following inequalities hold.*

(a) $\mathbb{P}(S_{\mathrm{solve}}) \ge \alpha_N^K$.

(b) $\mathbb{P}(S_{\mathrm{overlap}}) \ge 1 - K^2\nu - \frac{2K\nu}{\lambda}$.

(c) $\mathbb{P}(S_{\mathrm{eigen}}) \ge 1 - \exp(-cN)$ *for* $c > 0$ *depending only on* ξ, h, ε.

(d) $\mathbb{P}(S_{\mathrm{ogp}}) \ge 1 - 2\exp\left(-\frac{\varepsilon^2}{32\xi(1)}N\right)$.

We defer the proof of this proposition to after the proof of Theorem 3.1.3.

Proof of Theorem 3.1.3. Lemma 3.2.5 implies that $\mathbb{P}(S_{\mathrm{solve}}) + \mathbb{P}(S_{\mathrm{overlap}}) + \mathbb{P}(S_{\mathrm{eigen}}) + \mathbb{P}(S_{\mathrm{ogp}}) \le 3$. Because $(K^2 + 2K)^{1/K} \le 3$ for any positive integer K and $\lambda < 1$,

$$\begin{aligned}
\alpha_N &\le \left(K^2\nu + \frac{2K\nu}{\lambda}\right)^{1/K} + 2\exp\left(-\frac{\varepsilon^2}{32K\xi(1)}N\right) + e^{-cN/K}\\
&\le 3\left(\frac{\nu}{\lambda}\right)^{1/K} + 2\exp\left(-\frac{\varepsilon^2}{32K\xi(1)}N\right) + e^{-cN/K}.
\end{aligned}$$

Recall that $K \le K_0$ and K_0 is a constant depending only on ξ, h, ε. The proof is complete up to choosing an appropriate c in Theorem 3.1.3. □

3.2.5 Proofs of Probability Lower Bounds

In this section, we will prove Proposition 3.2.6. As preparation we first give two useful concentration lemmas. The first shows that $R(\mathcal{A}(H_N), \boldsymbol{m})$ concentrates around $\|\boldsymbol{m}\|_N^2$ for overlap concentrated algorithms with $\mathbb{E}[\mathcal{A}(H_N)] = \boldsymbol{m}$.

Lemma 3.2.7. *If $\mathcal{A} = \mathcal{A}_N$ is (λ, ν) overlap concentrated and $\mathbb{E}[\mathcal{A}(H_N)] = \boldsymbol{m}$, then*

$$\mathbb{P}\left[\left|R(\mathcal{A}(H_N), \boldsymbol{m}) - \|\boldsymbol{m}\|_N^2\right| > 2\lambda\right] \le \frac{2\nu}{\lambda}. \tag{3.2.4}$$

Proof. Define the convex function $\psi(t) = (|t - \|\boldsymbol{m}\|_N^2| - \lambda)_+$. Then by Jensen's inequality, for independent Hamiltonians H_N and H_N',

$$\mathbb{E}\left[\psi\left(R(\mathcal{A}(H_N), \boldsymbol{m})\right)\right] \le \mathbb{E}\left[\psi\left(R(\mathcal{A}(H_N), \mathcal{A}(H_N'))\right)\right].$$

Because $\mathcal{A}$ is (λ, ν) overlap concentrated, $\psi\left(R(\mathcal{A}(H_N), \mathcal{A}(H_N'))\right) = 0$ with probability at least $1 - \nu$. Moreover, $\psi\left(R(\mathcal{A}(H_N), \mathcal{A}(H_N'))\right) \le 2$ pointwise. So,

$$\mathbb{E}\left[\psi\left(R(\mathcal{A}(H_N), \boldsymbol{m})\right)\right] \le 2\nu.$$

By Markov's inequality,

$$\mathbb{P}\left[\left|R(\mathcal{A}(H_N), \boldsymbol{m}) - \|\boldsymbol{m}\|_N^2\right| > 2\lambda\right] = \mathbb{P}\left[\psi\left(R(\mathcal{A}(H_N), \boldsymbol{m})\right) > \lambda\right] \le \frac{2\nu}{\lambda}.$$

□

The next lemma shows subgaussian concentration for $\frac{1}{N}\max_{\vec{\boldsymbol{\sigma}} \in \mathcal{Q}(\eta)} \mathcal{H}_N(\vec{\boldsymbol{\sigma}})$.

Proposition 3.2.8. *The random variable*

$$Y = \frac{1}{N}\max_{\vec{\boldsymbol{\sigma}} \in \mathcal{Q}(\eta)} \mathcal{H}_N(\vec{\boldsymbol{\sigma}})$$

satisfies for all $t \ge 0$

$$\mathbb{P}[|Y - \mathbb{E}\, Y| \ge t] \le 2\exp\left(-\frac{Nt^2}{2K^2\xi(1)}\right).$$

Proof. For any $\vec{\boldsymbol{\sigma}} \in S_N^K$, by Cauchy-Schwarz the variance of $\mathcal{H}_N(\vec{\boldsymbol{\sigma}})$ is at most

$$\begin{aligned}
\mathbb{E}\left[\left(\mathcal{H}_N(\vec{\boldsymbol{\sigma}}) - \mathbb{E}\,\mathcal{H}_N(\vec{\boldsymbol{\sigma}})\right)^2\right] &= \sum_{u^1, u^2 \in \mathbb{L}} \mathbb{E}\,\widetilde{H}_N^{(u^1)}(\boldsymbol{\sigma}(u^1))\,\mathbb{E}\,\widetilde{H}_N^{(u^2)}(\boldsymbol{\sigma}(u^2)) \\
&\le K\sum_{u \in \mathbb{L}} \mathbb{E}\,\widetilde{H}_N^{(u)}(\boldsymbol{\sigma}(u))^2 \\
&= NK^2\xi(1).
\end{aligned}$$

The result now follows from the Borell-TIS inequality ([Bor75, CIS76], or see [Zei15, Theorem 2]). Note that both the statement and proof of Borell-TIS hold for noncentered Gaussian processes with no modification. □

We now prove each part of Proposition 3.2.6 in turn.

Proof of Proposition 3.2.6(a). For $0 \le d \le \underline{D}$, let $1^d \in \mathbb{T}$ denote the node $(1, \dots, 1)$ with d entries (so $1^0 = \emptyset$ is the root of $\mathbb{T}$), and let S_d be the event that $H_N^{(u)}(\boldsymbol{\sigma}(\underline{u})) \ge \mathsf{ALG} + \varepsilon$ for all $\underline{u} \in \mathbb{L}$ descended from the node 1^d. Let $P_d = \mathbb{P}[S_d]$. Note that $P_{\underline{D}} = \alpha_N$. We will show $P_0 \ge \alpha_N^K \ge \alpha_N^K$ by showing that for all $1 \le d \le \underline{D}$,

$$P_{d-1} \ge P_d^{k_d}.$$

The result will then follow by induction.

Recall the construction (3.2.2) of the Hamiltonians $H_N^{(u)}$ in terms of i.i.d. Hamiltonians $(\widetilde{H}_N^{[u]})_{u\in\mathbb{T}}$. Conditioned on the Hamiltonians $\Omega_{d-1} = (\widetilde{H}_N^{[1^{d'}]})_{0\le d'\le d-1}$, let $f_d(\Omega_{d-1})$ denote the conditional probability of S_d. Note that

$$P_d = \mathbb{E}\, f_d(\Omega_{d-1}).$$

By symmetry of the k_d descendant subtrees of the node 1^{d-1},

$$P_{d-1} = \mathbb{E}\, f_d(\Omega_{d-1})^{k_d}.$$

Thus $P_{d-1} \ge P_d^{k_d}$ by Jensen's inequality. □

Proof of Proposition 3.2.6(b). By definition of χ, $\mathbb{E}\, R(\underline{\boldsymbol{\sigma}}(\underline{u}^1), \underline{\boldsymbol{\sigma}}(\underline{u}^2)) = \chi(p_{\underline{u}^1\wedge\underline{u}^2})$. If $\underline{u}^1 \wedge \underline{u}^2 < \underline{D}$, then $p_{\underline{u}^1\wedge\underline{u}^2} < 1$. Because $\vec{p}, \vec{q}$ are χ-aligned, we have $\chi(p_{\underline{u}^1\wedge\underline{u}^2}) = q_{\underline{u}^1\wedge\underline{u}^2}$. If $\underline{u}^1 \wedge \underline{u}^2 = \underline{D}$, then $p_{\underline{u}^1\wedge\underline{u}^2} = 1$, so clearly $\chi(p_{\underline{u}^1\wedge\underline{u}^2}) = \chi(1)$. So, in all cases, $\mathbb{E}\, R(\underline{\boldsymbol{\sigma}}(\underline{u}^1), \underline{\boldsymbol{\sigma}}(\underline{u}^2)) = \underline{Q}_{\underline{u}^1,\underline{u}^2}$.

Using (3.1.1) and a union bound over $\underline{u}^1, \underline{u}^2 \in \underline{\mathbb{L}}$, we have

$$\left\|R(\vec{\underline{\boldsymbol{\sigma}}}, \vec{\underline{\boldsymbol{\sigma}}}) - \underline{Q}\right\|_\infty \le \lambda$$

with probability at least $1 - K^2\nu$. By Lemma 3.2.7 and a union bound, we have

$$\left|R(\mathcal{A}(H_N^{(\underline{u})}), \boldsymbol{m}) - \|\boldsymbol{m}\|_N^2\right| \le 2\lambda$$

for all $\underline{u} \in \underline{\mathbb{L}}$ with probability at least $1 - \frac{2K\nu}{\lambda}$. Recall that $\lambda = \eta_0/4 \le \eta/4$. By a final union bound,

$$\mathbb{P}[\vec{\underline{\boldsymbol{\sigma}}} \in \underline{\mathcal{Q}}(\eta/2)] \ge 1 - K^2\nu - \frac{2K\nu}{\lambda}.$$

□

Proof of Proposition 3.2.6(c). We focus on a fixed $\underline{u} \in \underline{\mathbb{L}}$. The requirements $H_N^{(u)} \in K_N$ follow from Proposition 3.0.1. The uniform eigenvalue lower bound follows by union bounding over subspaces S and a net of points $\boldsymbol{x}$. In fact it follows from exactly the same proof as [Sel20, Lemma 2.6] up to replacing each appearance of an eigenvalue λ_i to λ_{K+i}.

□

Proof of Proposition 3.2.6(d). By (3.2.3) and Proposition 3.2.8 with $t = K\varepsilon/4$,

$$\begin{aligned}\mathbb{P}\left[\frac{1}{N}\max_{\vec{\boldsymbol{\sigma}}\in\mathcal{Q}(\eta)} \mathcal{H}_N(\vec{\boldsymbol{\sigma}}) \ge K(\mathsf{ALG} + 3\varepsilon/4)\right] &\le \mathbb{P}\left[\frac{1}{N}\max_{\vec{\boldsymbol{\sigma}}\in\mathcal{Q}(\eta)} \mathcal{H}_N(\vec{\boldsymbol{\sigma}}) - \frac{1}{N}\mathbb{E}\max_{\vec{\boldsymbol{\sigma}}\in\mathcal{Q}(\eta)} \mathcal{H}_N(\vec{\boldsymbol{\sigma}}) \ge \frac{K\varepsilon}{4}\right] \\ &\le 2\exp\left(-\frac{\varepsilon^2}{32\xi(1)}N\right).\end{aligned}$$

□

3.2.6 Proof of Lemma 3.2.4

The spherical case of Lemma 3.2.4 follows from [Sub21, Remark 6] and does not require any of the axis-aligned subspace conditions. We therefore focus on the Ising case, which is a slight extension of the main result of [Sel20].

Lemma 3.2.9. *Suppose $S_{\text{eigen}}(\delta, K)$ holds. Then for any $\boldsymbol{x} \in [-1,1]^N$ with $||\boldsymbol{x}||_N^2 \le 1-\delta$, any $u \in \mathbb{L}$ and any subspace $W \subseteq \mathbb{R}^N$ of dimension $\dim(W) \ge N - K - 1$, there are mutually orthogonal vectors $\boldsymbol{y}^1, \dots, \boldsymbol{y}^K \in W \cap \boldsymbol{x}^\perp$ such that for each $i \in [K]$ the following hold where C_3 is as in Proposition 3.0.1.*

1. $\boldsymbol{x} + \boldsymbol{y}^i \in [-1,1]^N$.
2. *If $x_j \in \{-1,1\}$ then $y_j^i = 0$.*

3. $H_N^{(u)}(\boldsymbol{x}+\boldsymbol{y}^i) - H_N^{(u)}(\boldsymbol{x}) \geq -\delta \left\|\boldsymbol{y}^i\right\|_2^2$.

4. $\left\|\boldsymbol{y}^i\right\|_N \leq \frac{\delta}{10C_3}$.

5. *If* $\|\boldsymbol{x}\|_N^2 < q_d$ *for some* $1 \leq d \leq D$, *then* $\left\|\boldsymbol{x}+\boldsymbol{y}^i\right\|_N^2 \leq q_d$.

6. *At least one of the following three events holds.*

   (a) $\left\|\boldsymbol{y}^i\right\|_N = \frac{\delta}{10C_3}$.

   (b) $\boldsymbol{x}+\boldsymbol{y}^i$ *has strictly more* ± 1*-valued coordinates than* $\boldsymbol{x}$.

   (c) $||\boldsymbol{x}||_N^2 < q_d$ *and* $||\boldsymbol{x}+\boldsymbol{y}^i||_N^2 = q_d$ *for some* $1 \leq d \leq D$.

Proof. By the Markov inequality, $\boldsymbol{x}$ has a set S of at least $(1 - ||\boldsymbol{x}||_N^2)N$ coordinates not in $\{-1, 1\}$. $S_{\text{eigen}}(\delta, K)$ and the Cauchy interlacing inequality imply

$$\lambda_K \left(\nabla^2 H_N^{(u)}(\boldsymbol{x})|_{W_S \cap W}\right) \geq \lambda_{2K+1} \left(\nabla^2 H_N^{(u)}(\boldsymbol{x})|_{W_S}\right) \geq 0.$$

Let $\boldsymbol{y}^1, \dots, \boldsymbol{y}^K \in W_S(\boldsymbol{x}) \cap W$ be a corresponding choice of orthogonal eigenvectors, each satisfying

$$\left\langle \boldsymbol{y}^i, \nabla^2 H_N^{(u)}(\boldsymbol{x}) \boldsymbol{y}^i \right\rangle \geq 0.$$

Since $\boldsymbol{y}^i$ and $-\boldsymbol{y}^i$ play symmetric roles we may assume without loss of generality that $\langle \nabla H_N^{(u)}(\boldsymbol{y}), \boldsymbol{y}^i \rangle \geq 0$. Replacing $\boldsymbol{y}^i$ by $t\boldsymbol{y}^i$ for suitable $t \in [0,1]$ if needed, we may ensure that Items 1, 2, 4, 5, and 6 above hold.

Since $S_{\text{eigen}}(\delta, K)$ implies that $\left\|\nabla^3 H_N^{(u)}\right\|_{\text{op}}$ is uniformly bounded by C_3, it follows that along the line segment $\boldsymbol{x} + [0,1]\boldsymbol{y}^i$ the Hessian of $H_N^{(u)}$ varies in operator norm by at most $\frac{\delta}{5}$. This combined with $\langle \nabla H_N^{(u)}(\boldsymbol{x}), \boldsymbol{y}^i \rangle \geq 0$ implies

$$H_N^{(u)}(\boldsymbol{x}+\boldsymbol{y}^i) \geq H_N^{(u)}(\boldsymbol{x}) - \delta \left\|\boldsymbol{y}^i\right\|_2^2.$$

This completes the proof. □

Proof of Lemma 3.2.4. Take

$$\delta < \frac{\min(\varepsilon, \eta, 1 - q_{D-1})^2}{16(C_1 + C_3 + 1)}$$

sufficiently small, where C_1, C_3 are given by Proposition 3.0.1. Enumerate $\underline{u}^1, \dots, \underline{u}^K \in \mathbb{L}$. Assume the points $\boldsymbol{\sigma}(u)$ for descendants $u \in \mathbb{L}$ of $\underline{u}^1, \dots, \underline{u}^{j-1}$ have already been chosen and satisfy the conclusions of Lemma 3.2.4. We show how to define the points $\boldsymbol{\sigma}(u)$ for u a descendant of $\underline{u}^j$.

From the starting point $\boldsymbol{x}^{0,\underline{u}^j} = \boldsymbol{\sigma}(\underline{u}^j)$, we produce iterates $\boldsymbol{x}^{i,v}$ for $i \in \mathbb{N}$ and $v \in \mathbb{T}$ a descendant of $\underline{u}^j$, similarly to [Sub21] and [Sel20, Proof of Theorem 1]. First let $d_0 = d_0(\underline{u}^j) \in [D]$ be such that $||\boldsymbol{x}^{0,\underline{u}^j}||_N^2 \in [q_{d_0-1}, q_{d_0})$, and set $\boldsymbol{x}^{0,v} = \boldsymbol{x}^{0,\underline{u}^j}$ for all depth d_0 descendants v of $\underline{u}^j$ if $d_0 > \underline{D}$.

Given a point $\boldsymbol{x}^{m,v}$ with v a descendant of $\underline{u}^j$, suppose that $||\boldsymbol{x}^{m,v}||_N^2 \in (q_{|v|-1}, q_{|v|} \wedge (1-\delta))$. Then take the subspace $W^\perp$ (which changes from iteration to iteration) to be the span of $\boldsymbol{m}$ as well as all currently defined leaves of the exploration tree (including $\boldsymbol{x}^{m,v}$ itself). Hence $\dim(W^\perp) \leq K+1$ and so $\dim(W) \geq N-K-1$. (The resulting exploration tree can be constructed in arbitrary order; at any time it will have at most K leaves.)

Then there exists $\boldsymbol{y}^{m,v}$ satisfying the properties of Lemma 3.2.9 with subspace W and Hamiltonian $H_N^{(\underline{u}^j)}$. We update

$$\boldsymbol{x}^{m+1,v} = \boldsymbol{x}^{m,v} + \boldsymbol{y}^{m,v}, \quad v \in \mathbb{T}.$$

However if $||\boldsymbol{x}^{m,v}||_N^2 = q_{|v|}$, then we let $v^1, \dots, v^{k_{d+1}}$ be the children of v in $\mathbb{T}$ and generate $\boldsymbol{y}^{m,v^1}, \dots \boldsymbol{y}^{m,v^{k_{d+1}}}$ again using Lemma 3.2.9. We then define

$$\boldsymbol{x}^{m+1,v^j} = \boldsymbol{x}^{m,v} + \boldsymbol{y}^{m,v^j}, \quad j \in [k_{d+1}].$$

Continuing in this way, we eventually reach points $\boldsymbol{x}^{m+1,u}$ with $||\boldsymbol{x}^{m+1,u}||_N^2 \geq (1-\delta)$ for each $u \in \mathbb{L}$; indeed the last condition of Lemma 3.2.9 ensures that this eventually occurs for each $u \in \mathbb{L}$. We set $\boldsymbol{x}^u = \boldsymbol{x}^{m+1,u}$. Observe that by orthogonality of $\boldsymbol{x}^{m,v}$ and $\boldsymbol{y}^{m,v}$,

$$\begin{aligned} H_N^{(u)}(\boldsymbol{x}^{m+1,v}) &\geq H_N^{(u)}(\boldsymbol{x}^{m,v}) - N\delta \left\|\boldsymbol{y}^{m,v}\right\|_N^2 \\ &\geq H_N^{(u)}(\boldsymbol{x}^{m,v}) - N\delta \cdot \left(\left\|\boldsymbol{x}^{m+1,v}\right\|_N^2 - \left\|\boldsymbol{x}^{m,v}\right\|_N^2 \right). \end{aligned}$$

It follows by telescoping that (recall $\underline{u}^j \in \underline{\mathbb{L}}$ is an ancestor of $u \in \mathbb{L}$),

$$H_N^{(u)}(\boldsymbol{x}^u) \geq H_N^{(u)}(\boldsymbol{x}^{\underline{u}^j}) - N\delta \geq H_N^{(u)}(\boldsymbol{x}^{\underline{u}^j}) - N\varepsilon/2.$$

Since every update above is made orthogonally to all contemporaneous iterates, it is not difficult to see that the final iterates $(\boldsymbol{x}^u)_{u \in \mathbb{L}}$ satisfy the following.

- $R(\boldsymbol{x}^u, \boldsymbol{x}^u) \geq 1 - \delta \geq q_{u \wedge u} - \frac{\eta}{2}$.
- If $u^1 \neq u^2$ are both descendants of $\underline{u}^j \in \underline{\mathbb{L}}$ and $u^1 \wedge u^2 < d_0(\underline{u}^j)$, then

  $$R(\boldsymbol{x}^{u^1}, \boldsymbol{x}^{u^2}) = R(\boldsymbol{x}^{\underline{u}^j}, \boldsymbol{x}^{\underline{u}^j}) \leq q_{\underline{D}} + \frac{\eta}{2} \leq q_{u^1 \wedge u^2} + \frac{\eta}{2}.$$

  and

  $$R(\boldsymbol{x}^{u^1}, \boldsymbol{x}^{u^2}) \geq q_{d_0 - 1} \geq q_{u^1 \wedge u^2},$$

  hence $\left| R(\boldsymbol{x}^{u^1}, \boldsymbol{x}^{u^2}) - q_{u^1 \wedge u^2} \right| \leq \eta/2$.
- Otherwise, $R(\boldsymbol{x}^{u^1}, \boldsymbol{x}^{u^2}) = q_{u^1 \wedge u^2}$.

Moreover all updates were also orthogonal to $\boldsymbol{m}$, so $|R(\boldsymbol{m}, \boldsymbol{x}^u)| \leq \eta/2$ for all $u \in \mathbb{L}$.

Finally, to produce outputs in Σ_N, for each $u \in \mathbb{L}$ and $i \in [N]$ we independently round the coordinate x_i^u at random to $\sigma(u)_i \in \{-1, 1\}$ so that $\mathbb{E}[\boldsymbol{\sigma}(u)] = \boldsymbol{x}^u$. It is not difficult to see that

$$\mathbb{P}[|R(\boldsymbol{x}^{u^1}, \boldsymbol{x}^{u^2}) - R(\boldsymbol{\sigma}(u^1), \boldsymbol{\sigma}(u^2))| \geq \delta] \leq e^{-c(\delta)N}$$

for each $u^1, u^2 \in \mathbb{L}$, and similarly for inner products with $\boldsymbol{m}$. We conclude that $\vec{\boldsymbol{\sigma}} \in \mathcal{Q}(\eta)$ holds with probability $1 - e^{-c(\delta)N}$ (since $\delta \leq \eta/2$). Similarly $\|\boldsymbol{\sigma}(u) - \boldsymbol{x}^u\|_2^2$ is an independent sum of N terms each at most 1 and has expectation at most δ. It follows that

$$\mathbb{P}[\|\boldsymbol{\sigma}(u) - \boldsymbol{x}^u\|_N \geq 2\delta^{1/2}] \leq e^{-c(\delta)N}.$$

Now using S_{eigen}, for every $(\underline{u}, u) \in \underline{\mathbb{L}} \times \mathbb{L}$ with $\underline{u}$ an ancestor of u,

$$\begin{aligned} H_N^{(u)}(\boldsymbol{\sigma}(u)) &\geq H_N^{(u)}(\boldsymbol{x}^u) - 2C_1\delta^{1/2}N \\ &\geq H_N^{(u)}(\boldsymbol{x}^u) - N\varepsilon/2 \\ &\geq H_N^{(u)}(\boldsymbol{x}^{\underline{u}}) - N\varepsilon \end{aligned}$$

holds with probability $1 - e^{-c(\delta)N}$. In particular, the above events hold simultaneously over all $(\underline{u}, u)$ with probability at least $\frac{1}{2}$ over the random rounding step. Hence there exists some $\vec{\boldsymbol{\sigma}}$ satisfying all desired conditions. This concludes the proof. □

3.2.7 A Different Class of Algorithms Capturing The Approach of Subag

The optimization algorithm of [Sub21] in the spherical setting can be summarized as follows. Starting from any $\boldsymbol{x}^1 \in B_N$ with $\left\|\boldsymbol{x}^1\right\|_N^2 = \delta$, repeatedly compute the maximum-eigenvalue unit eigenvector $\boldsymbol{v}^i \in \mathbb{R}^N$ of $P_{(\boldsymbol{x}^i)^\perp} \nabla^2 H_N(\boldsymbol{x}^i) P_{(\boldsymbol{x}^i)^\perp}$ (the Hessian of H_N at $\boldsymbol{x}^i$ restricted to the orthogonal complement of $\boldsymbol{x}^i$). Then, set

$$\boldsymbol{x}^{i+1} = \boldsymbol{x}^i + \boldsymbol{v}^i \sqrt{\delta N} \tag{3.2.5}$$

where the sign of $\boldsymbol{v}^i$ is chosen depending on the gradient $\nabla H_N(\boldsymbol{x}^i)$. By construction, $\left\|\boldsymbol{x}^i\right\|_N^2 = i\delta$, so if $\delta^{-1} = m \in \mathbb{N}$ then $\boldsymbol{x}^m \in S_N$. By uniformly lower bounding the maximum eigenvalue of the Hessians, [Sub21] showed that this algorithm obtains energy at least $(\mathsf{ALG}^{\mathrm{Sp}} + o_\delta(1))N$ as $\delta \to 0$. Because the maximum eigenvalue is a discontinuous operation, our results do not apply to Subag's algorithm.

We consider the following variant. At each $\boldsymbol{x}^i$, let the subspace $W(\boldsymbol{x}^i)$ be the span of the top $\lfloor \delta N \rfloor$ eigenvectors of $P_{(\boldsymbol{x}^i)^\perp} \nabla^2 H_N(\boldsymbol{x}^i) P_{(\boldsymbol{x}^i)^\perp}$. Next, choose $\boldsymbol{v}^i$ *uniformly at random* from the unit sphere of $W(\boldsymbol{x}^i)$ and update using (3.2.5). This modified algorithm obeys the same guarantees as that of [Sub21] by exactly the same proof.

More generally, we define the class of δ-*subspace random walk* algorithms for $\delta > 0$ with $\delta^{-1} = m \in \mathbb{N}$, only in the spherical setting for convenience, as follows. Given H_N, let $W(\boldsymbol{x}^i) \subseteq \mathbb{R}^N$ be an arbitrary (measurable in $(H_N, \boldsymbol{x})$) subspace of dimension $\lfloor \delta N \rfloor$. Starting from arbitrary $\boldsymbol{x}^1 \in B_N$ with $\|\boldsymbol{x}_1\|_N^2 = \delta$, repeatedly choose a uniformly random unit vector $\boldsymbol{v}^i \in W(\boldsymbol{x}^i)$ and define $\boldsymbol{x}^{i+1}$ via (3.2.5), leading to the output $\boldsymbol{\sigma} = \boldsymbol{x}^m$. Note that in contrast to elsewhere in the paper, here the output $\boldsymbol{x}^{i+1}$ is random even given H_N, i.e. $\boldsymbol{x}^{i+1} = \mathcal{A}(H_N, \omega)$ for some independent random variable ω. As we now outline, for $\delta \le \delta_0(\varepsilon)$ sufficiently small depending on ε, no δ-subspace random walk algorithm can achieve energy than $\mathsf{ALG}^{\mathrm{Sp}} + \varepsilon$ with non-negligible probability.

Fixing H_N and $\boldsymbol{x}^1$, for any $j \le m$ we may generate coupled outputs $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2$ as follows. First use shared iterates $\boldsymbol{x}^{i,1} = \boldsymbol{x}^{i,2} = \boldsymbol{x}^i$ for $i \le j$ and then proceed via

$$\boldsymbol{x}^{i+1,\ell} = \boldsymbol{x}^{i,\ell} + \boldsymbol{v}^{i,\ell}\sqrt{\delta N}, \quad \ell \in \{1,2\}$$

for independent update sequences $(\boldsymbol{v}^{j,1}, \dots, \boldsymbol{v}^{m-1,1})$ and $(\boldsymbol{v}^{j,2}, \dots, \boldsymbol{v}^{m-1,2})$. Finally output $\boldsymbol{\sigma}^\ell = \boldsymbol{x}^{m,\ell}$. It is not difficult to see that for N sufficiently large,

$$\mathbb{P}\left[\left|R(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) - j\delta\right| > \eta/2\right] \le e^{-cN}$$

for some $c = c(\delta, \eta)$ thanks to the random directions of the updates $\boldsymbol{v}^{i,\ell}$. With $\mathbb{L}$ as in the earlier part of this section, we can now construct a branching tree of outputs $\boldsymbol{\sigma}(u)$ for $u \in \mathbb{L}$. As $\delta \to 0$, for appropriate $j_d = \lfloor q_d \delta^{-1} \rfloor$, the solution configuration $\vec{\boldsymbol{\sigma}}$ hence constructed satisfies

$$\mathbb{P}[\vec{\boldsymbol{\sigma}} \in \mathcal{Q}(\eta)] \le e^{-cN}$$

with $\boldsymbol{m}$ the zero vector. Because we consider a single Hamiltonian H_N, we use Proposition 3.2.2 with $\chi(p) \to 0$ for all $p < 1$. Since the statement is uniform in χ, this does not present any difficulties (we are essentially "defining" $\vec{p} = 1^D$ to be χ-aligned with arbitrary $\vec{q}$). Mimicking the proofs earlier in this section (including the argument in the proof of Proposition 3.2.6(a) which now uses Jensen's inequality on the randomness of $\mathcal{A}$), we obtain the following result.

Theorem 3.2.10. *Consider a mixed even Hamiltonian H_N with model (ξ, h). For any $\varepsilon > 0$ there are $\delta_0, c, N_0 > 0$ depending only on ξ, h, ε such that the following holds for any $N \ge N_0$ and $\delta < \delta_0$. For any δ-subspace random walk algorithm $\mathcal{A}$,*

$$\mathbb{P}\left[\frac{1}{N} H_N\left(\mathcal{A}(H_N, \omega)\right) \ge \mathsf{ALG}^{\mathrm{Sp}} + \varepsilon\right] \le \exp(-cN).$$

3.3 Guerra's Interpolation

In this section, we begin the proof of Proposition 3.2.2. We take either $\mathcal{Q}(\eta) = \mathcal{Q}^{\mathrm{Sp}}(Q, \boldsymbol{m}, \eta)$ or $\mathcal{Q}(\eta) = \mathcal{Q}^{\mathrm{Is}}(Q, \boldsymbol{m}, \eta)$ (recall $Q = Q^{\vec{k},\vec{q}}$); the proofs in this section apply uniformly to both cases. The goal of this section is to use Guerra's interpolation to upper bound the constrained free energy

$$F_N(\mathcal{Q}(\eta)) = \frac{1}{N} \log \mathbb{E} \int_{\mathcal{Q}(\eta)} \exp \mathcal{H}_N(\vec{\boldsymbol{\sigma}}) \, \mathrm{d}\mu^K(\vec{\boldsymbol{\sigma}}),$$

where μ is a (for now) arbitrary measure on S_N. In the sequel, we will take μ to be the uniform measure on S_N for spherical spin glasses, and the counting measure on Σ_N for Ising spin glasses. We develop a bound on $F_N(\mathcal{Q}(\eta))$ that holds for all $D, \vec{k}, \vec{p}, \vec{q}, \eta$, and will set these variables in the sequel to prove Proposition 3.2.2.

We will control this free energy by controlling the following related free energy. Let $\lambda \in \mathbb{R}$ be a constant we will set later. For all $\boldsymbol{\sigma} \in \mathbb{R}^N$, let $\pi(\boldsymbol{\sigma}) = \boldsymbol{\sigma} - \boldsymbol{m}$. We define the following modified grand Hamiltonian, where we add an external field $\lambda \boldsymbol{m}$ centered at $\boldsymbol{m}$:

$$\begin{aligned}\mathcal{H}_{N,\lambda}(\vec{\boldsymbol{\sigma}}) &= \mathcal{H}_N(\vec{\boldsymbol{\sigma}}) + \sum_{u \in \mathbb{L}} \langle \lambda \boldsymbol{m}, \pi(\boldsymbol{\sigma}(u)) \rangle \\ &= K\langle \boldsymbol{h}, \boldsymbol{m} \rangle + \sum_{u \in \mathbb{L}} \left[\langle \boldsymbol{h} + \lambda \boldsymbol{m}, \pi(\boldsymbol{\sigma}(u)) \rangle + \widetilde{H}_N^{(u)}(\boldsymbol{\sigma}(u)) \right].\end{aligned}$$

We define the free energy

$$F_{N,\lambda}(\mathcal{Q}(\eta)) = \frac{1}{N} \log \mathbb{E} \int_{\mathcal{Q}(\eta)} \exp \mathcal{H}_{N,\lambda}(\vec{\boldsymbol{\sigma}}) \, \mathrm{d}\mu^K(\vec{\boldsymbol{\sigma}}).$$

Since $\mathcal{Q}(\eta) \subseteq B(\boldsymbol{m}, \eta)^K$, we have $|\mathcal{H}_N(\vec{\boldsymbol{\sigma}}) - \mathcal{H}_{N,\lambda}(\vec{\boldsymbol{\sigma}})| \le NK|\lambda|\eta$ for all $\vec{\boldsymbol{\sigma}} \in \mathcal{Q}(\eta)$, and so

$$|F_N(\mathcal{Q}(\eta)) - F_{N,\lambda}(\mathcal{Q}(\eta))| \le K|\lambda|\eta. \tag{3.3.1}$$

Define the matrices $M^{\vec{k},\vec{p},1}, \dots, M^{\vec{k},\vec{p},D} \in \mathbb{R}^{K \times K}$, whose rows and columns are indexed by $\mathbb{L}$, by

$$M^{\vec{k},\vec{p},d}_{u^1,u^2} = \mathbb{1}\left\{u^1 \wedge u^2 \ge d\right\} p_{u^1 \wedge u^2}.$$

Further, define $M^{\vec{k},\vec{p},\vec{q}} : [q_0, 1) \to \mathbb{R}^{K \times K}$ as the piecewise constant matrix-valued function such that for $q \in [q_{d-1}, q_d)$, $M^{\vec{k},\vec{p},\vec{q}}(q) = M^{\vec{k},\vec{p},d}$. Define $\kappa^{\vec{k},\vec{p},\vec{q}} : [q_0, 1) \to \mathbb{R}$ by

$$\kappa^{\vec{k},\vec{p},\vec{q}}(q) = \frac{1}{K} \mathrm{Sum}(M^{\vec{k},\vec{p},\vec{q}}(q)),$$

where Sum denotes the sum of entries of a matrix. Explicitly, for $q \in [q_{d-1}, q_d)$,

$$\kappa^{\vec{k},\vec{p},\vec{q}}(q) = \sum_{j=d}^{D-1} \left[(k_{j+1} - 1) \prod_{\ell=j+2}^{D} k_\ell \right] p_j + p_D. \tag{3.3.2}$$

When $\vec{k}, \vec{p}, \vec{q}$ are clear, we will write $M^d = M^{\vec{k},\vec{p},d}$, $M(q) = M^{\vec{k},\vec{p},\vec{q}}(q)$ and $\kappa(q) = \kappa^{\vec{k},\vec{p},\vec{q}}(q)$. Consider a sequence

$$0 = \zeta_{-1} < \zeta_0 < \cdots < \zeta_D = 1,$$

which we identify with the piecewise constant CDF $\zeta : [q_0, 1) \to [0, 1]$, where for $x \in [q_d, q_{d+1})$,

$$\zeta(x) = \zeta_d, \tag{3.3.3}$$

corresponding to the discrete distribution $\zeta(\{q_d\}) = \zeta_d - \zeta_{d-1}$. We denote by $\mathcal{M}_{\vec{q}}$ the set of such CDFs ζ for a given $\vec{q}$.

Let $\mathbb{T}_D = \mathbb{N}^0 \cup \mathbb{N}^1 \cup \cdots \cup \mathbb{N}^D$ and for $\omega \in \mathbb{T}_D$, let $|\omega|$ denote the length of ω. Let $\emptyset$ denote the empty tuple. We think of $\mathbb{T}_D$ as a tree rooted at $\emptyset$, where the parent of any $\omega \ne \emptyset$ is the initial substring of ω with length $|\omega| - 1$. For $\alpha \in \mathbb{N}^D$, let $p(\alpha) = ((\alpha_1), (\alpha_1, \alpha_2), \dots, (\alpha_1, \dots, \alpha_D))$ denote the path of vertices from the root to α, not including the root. For $\alpha^1, \alpha^2 \in \mathbb{N}^D$, let $\alpha^1 \wedge \alpha^2$ denote the depth of the least common ancestor of α^1 and α^2. Recall the Ruelle cascades $(\nu_\alpha)_{\alpha \in \mathbb{N}^D}$ corresponding to $(\zeta_0, \zeta_1, \dots, \zeta_{D-1})$ which were introduced in [Rue87], see also [Pan13b, Section 2.3].

For each increasing $\psi : [q_0, 1] \to \mathbb{R}_{\ge 0}$, we define a Gaussian process $g_\psi^{(u)}(\alpha)$ indexed by $(u, \alpha) \in \mathbb{L} \times \mathbb{N}^D$ as follows. Generate $e\vec{t}a_\emptyset \in \mathbb{R}^K$ by

$$e\vec{t}a_\emptyset = (\eta_\emptyset(u))_{u \in \mathbb{L}} \sim \mathcal{N}(0, M^1).$$

Furthermore, for each non-root $\omega \in \mathbb{T}_D$, independently generate $e\vec{t}a_\omega \in \mathbb{R}^K$ by

$$e\vec{t}a_\omega = (\eta_\omega(u))_{u \in \mathbb{L}} \sim \mathcal{N}(0, M^{|\omega|}).$$

Then, for each $u \in \mathbb{L}$, set

$$g_\psi^{(u)}(\alpha) = \eta_\emptyset(u)\psi(q_0)^{1/2} + \sum_{\omega \in p(\alpha)} \eta_\omega(u)(\psi(q_{|\omega|}) - \psi(q_{|\omega|-1}))^{1/2}.$$

This is the centered Gaussian process with covariance

$$\mathbb{E}\, g_\psi^{(u^1)}(\alpha^1) g_\psi^{(u^2)}(\alpha^2) = p_{u^1 \wedge u^2} \psi(q_{\alpha^1 \wedge \alpha^2} \wedge q_{u^1 \wedge u^2}),$$

where for $x, y \in \mathbb{R}$, $x \wedge y = \min(x, y)$. Generate N i.i.d. copies of the process $g_{\xi'}^{(u)}(\alpha)$, which we denote $g_{\xi',i}^{(u)}(\alpha)$ for $i = 1, \dots, N$. Similarly, for the function

$$\theta(q) = (q - q_0)\xi'(q) - \xi(q) + \xi(q_0),$$

we generate N i.i.d. processes $g_{\theta,i}^{(u)}(\alpha)$ for $i = 1, \dots, N$. Note that for $q \in [q_0, 1)$,

$$\theta(q) = \int_{q_0}^q (\xi'(q) - \xi'(q'))\, \mathrm{d}q' \geq 0 \qquad \text{and} \qquad \theta'(q) = (q - q_0)\xi''(q) \geq 0,$$

so θ is nonnegative and increasing, as required. For $t \in [0, 1]$, define the interpolating Hamiltonian

$$\begin{aligned} \mathcal{H}_{N,\lambda,t}(\vec{\boldsymbol{\sigma}}, \alpha) = \sum_{u \in \mathbb{L}} \left[\sqrt{t} \widetilde{H}_N^{(u)}(\boldsymbol{\sigma}(u)) + \sqrt{1-t} \sum_{i=1}^N g_{\xi',i}^{(u)}(\alpha) \pi(\boldsymbol{\sigma}(u))_i + \sqrt{t} \sum_{i=1}^N g_{\theta,i}^{(u)}(\alpha) \right] \\ + K\langle \boldsymbol{h}, \boldsymbol{m} \rangle + \langle \boldsymbol{h} + \lambda \boldsymbol{m}, \pi(\boldsymbol{\sigma}(u)) \rangle \end{aligned} \tag{3.3.4}$$

and the interpolating free energy

$$\varphi(t) = \frac{1}{N} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^D} \nu_\alpha \int_{\mathcal{Q}(\eta)} \exp \mathcal{H}_{N,\lambda,t}(\vec{\boldsymbol{\sigma}}, \alpha)\, \mathrm{d}\mu^K(\vec{\boldsymbol{\sigma}}).$$

The following bound on $F_N(\mathcal{Q}(\eta))$ is the main result of this section.

Proposition 3.3.1. *The free energy $F_N(\mathcal{Q}(\eta))$ is upper bounded by*

$$F_N(\mathcal{Q}(\eta)) \leq \varphi(0) - \frac{K}{2} \int_{q_0}^1 (q - q_0)\xi''(q)\kappa(q)\zeta(q)\, \mathrm{d}q + 3K^2\xi''(1)\eta + K|\lambda|\eta.$$

where $\zeta : [q_0, 1) \to [0, 1]$ is defined in (3.3.3).

Lemma 3.3.2 (Guerra's interpolation bound)**.** *For all $t \in [0, 1]$ and $\eta \in (0, 1)$,*

$$\varphi'(t) \leq 3K^2\xi''(1)\eta.$$

Proof. Let $\langle \cdot \rangle_t$ denote the average with respect to the Gibbs measure on $\mathcal{Q}(\eta) \times \mathbb{N}^D$ given by

$$G(\vec{\boldsymbol{\sigma}}, \alpha) \propto \nu_\alpha \exp \mathcal{H}_{N,\lambda,t}(\vec{\boldsymbol{\sigma}}, \alpha).$$

By Gaussian integration by parts [Pan13b, Lemma 1.4],

$$\begin{aligned} \varphi'(t) &= \frac{1}{N} \mathbb{E} \left\langle \frac{\partial \mathcal{H}_{N,\lambda,t}}{\partial t}(\vec{\boldsymbol{\sigma}}, \alpha) \right\rangle_t \\ &= \frac{1}{N} \mathbb{E} \left\langle \mathbb{E} \frac{\partial \mathcal{H}_{N,\lambda,t}}{\partial t}(\vec{\boldsymbol{\sigma}}^1, \alpha^1) \mathcal{H}_{N,\lambda,t}(\vec{\boldsymbol{\sigma}}^1, \alpha^1) - \mathbb{E} \frac{\partial \mathcal{H}_{N,\lambda,t}}{\partial t}(\vec{\boldsymbol{\sigma}}^1, \alpha^1) \mathcal{H}_{N,\lambda,t}(\vec{\boldsymbol{\sigma}}^2, \alpha^2) \right\rangle_t, \end{aligned} \tag{3.3.5}$$

where $(\vec{\boldsymbol{\sigma}}^1, \alpha^1)$ and $(\vec{\boldsymbol{\sigma}}^2, \alpha^2)$ are independent samples from the Gibbs measure. Recall (3.3.4). For any realizations $(\vec{\boldsymbol{\sigma}}^1, \alpha^1)$ and $(\vec{\boldsymbol{\sigma}}^2, \alpha^2)$,

$$\begin{aligned}
&\frac{2}{N}\,\mathbb{E}\,\frac{\partial \mathcal{H}_{N,\lambda,t}}{\partial t}(\vec{\boldsymbol{\sigma}}^1, \alpha^1)\mathcal{H}_{N,t}(\vec{\boldsymbol{\sigma}}^2, \alpha^2) \\
&= \sum_{u^1,u^2\in\mathbb{L}} p_{u^1\wedge u^2}\left[\xi(R(\boldsymbol{\sigma}^1(u^1),\boldsymbol{\sigma}^2(u^2))) - R(\pi(\boldsymbol{\sigma}^1(u^1)),\pi(\boldsymbol{\sigma}^2(u^2)))\xi'(q_{\alpha^1\wedge\alpha^2}\wedge q_{u^1\wedge u^2}) + \theta(q_{\alpha^1\wedge\alpha^2}\wedge q_{u^1\wedge u^2})\right] \\
&= \sum_{u^1,u^2\in\mathbb{L}} p_{u^1\wedge u^2}\big[\xi(R(\boldsymbol{\sigma}^1(u^1),\boldsymbol{\sigma}^2(u^2))) \\
&\quad - \big(R(\boldsymbol{\sigma}^1(u^1),\boldsymbol{\sigma}^2(u^2)) - R(\boldsymbol{\sigma}^1(u^1),\boldsymbol{m}) - R(\boldsymbol{\sigma}^2(u^2),\boldsymbol{m}) + R(\boldsymbol{m},\boldsymbol{m})\big)\,\xi'(q_{\alpha^1\wedge\alpha^2}\wedge q_{u^1\wedge u^2}) \\
&\quad + \theta(q_{\alpha^1\wedge\alpha^2}\wedge q_{u^1\wedge u^2})\big] \\
&= \sum_{u^1,u^2\in\mathbb{L}} p_{u^1\wedge u^2}\big[C\left(R(\boldsymbol{\sigma}^1(u^1),\boldsymbol{\sigma}^2(u^2)), q_{\alpha^1\wedge\alpha^2}\wedge q_{u^1\wedge u^2}\right) \\
&\quad + \big(R(\boldsymbol{\sigma}^1(u^1),\boldsymbol{m}) + R(\boldsymbol{\sigma}^2(u^2),\boldsymbol{m}) - 2q_0\big)\,\xi'(q_{\alpha^1\wedge\alpha^2}\wedge q_{u^1\wedge u^2}) + \xi(q_0)\big],
\end{aligned}$$

where

$$C(x,y) = \xi(x) - \xi(y) - (x-y)\xi'(y) = \int_y^x \int_y^z \xi''(w)\,\mathrm{d}w\,\mathrm{d}z. \tag{3.3.6}$$

Because $\boldsymbol{\sigma}^1(u^1), \boldsymbol{\sigma}^2(u^2) \in B(\boldsymbol{m}, \eta)$,

$$\left|\big(R(\boldsymbol{\sigma}^1(u^1),\boldsymbol{m}) + R(\boldsymbol{\sigma}^2(u^2),\boldsymbol{m}) - 2q_0\big)\,\xi'(q_{\alpha^1\wedge\alpha^2}\wedge q_{u^1\wedge u^2})\right| \le 2\xi'(1)\eta.$$

Hence using (3.3.5) and noting that $q_{\alpha^1\wedge\alpha^1} = 1$, we obtain

$$\begin{aligned}
\varphi'(t) \le \frac{1}{2} \sup_{\substack{\vec{\boldsymbol{\sigma}}^1,\vec{\boldsymbol{\sigma}}^2\in\mathcal{Q}(\eta)\\ \alpha^1,\alpha^2\in\mathbb{N}^D}} \sum_{u^1,u^2\in\mathbb{L}} &\left[C\left(R(\boldsymbol{\sigma}^1(u^1),\boldsymbol{\sigma}^1(u^2)), q_{u^1\wedge u^2}\right) - C\left(R(\boldsymbol{\sigma}^1(u^1),\boldsymbol{\sigma}^2(u^2)), q_{\alpha^1\wedge\alpha^2}\wedge q_{u^1\wedge u^2}\right)\right] \\
&+ 2K^2\xi'(1)\eta.
\end{aligned}$$

By (3.3.6), $0 \le C(x,y) \le |x-y|^2\xi''(1)$. Since $|R(\boldsymbol{\sigma}^1(u^1),\boldsymbol{\sigma}^1(u^2)) - q_{u^1\wedge u^2}| \le \eta$ for $\vec{\boldsymbol{\sigma}}^1 \in \mathcal{Q}(\eta)$,

$$C\left(R(\boldsymbol{\sigma}^1(u^1),\boldsymbol{\sigma}^1(u^2)), q_{u^1\wedge u^2}\right) \le \xi''(1)\eta^2.$$

Moreover,

$$C\left(R(\boldsymbol{\sigma}^1(u^1),\boldsymbol{\sigma}^2(u^2)), q_{\alpha^1\wedge\alpha^2}\wedge q_{u^1\wedge u^2}\right) \ge 0.$$

So,

$$\varphi'(t) \le \frac{1}{2}K^2\xi''(1)\eta^2 + 2K^2\xi'(1)\eta \le 3K^2\xi''(1)\eta.$$

□

We will now evaluate $\varphi(1)$ to complete the proof of Proposition 3.3.1.

Lemma 3.3.3. *The following identity holds.*

$$\varphi(1) = F_{N,\lambda}(\mathcal{Q}(\eta)) + \frac{K}{2}\sum_{d=0}^{D-1}\kappa(q_d)\zeta_d(\theta(q_{d+1}) - \theta(q_d)).$$

Proof. It is clear that

$$\varphi(1) = F_{N,\lambda}(\mathcal{Q}(\eta)) + \frac{1}{N}\log\mathbb{E}\sum_{\alpha\in\mathbb{N}^D}\nu_\alpha \exp\sum_{u\in\mathbb{L}}\sum_{i=1}^{N} g_{\theta,i}^{(u)}(\alpha).$$

We will evaluate the last term by the recursive evaluation of Ruelle cascades. For $1 \le d \le D$, independently generate $\vec{\boldsymbol{\eta}}_d = (\boldsymbol{\eta}_d(u))_{u\in\mathbb{L}} \in (\mathbb{R}^N)^K$ by generating, independently for each $1 \le i \le N$,

$$(\vec{\boldsymbol{\eta}}_d)_i = (\boldsymbol{\eta}_d(u)_i)_{u\in\mathbb{L}} \sim \mathcal{N}(0, M^d).$$

(Because $\theta(q_0) = 0$, we will not need $\vec{\boldsymbol{\eta}}_0$, corresponding to the root $\emptyset$ of $\mathbb{T}_D$.) Let

$$X_D = \sum_{u\in\mathbb{L}} \sum_{i=1}^{N} \sum_{d=1}^{D} \boldsymbol{\eta}_d(u)_i \left(\theta(q_d) - \theta(q_{d-1})\right)^{1/2},$$

and for $0 \le d \le D-1$ let

$$X_d = \frac{1}{\zeta_d} \log \mathbb{E}_d \exp \zeta_d X_{d+1}, \tag{3.3.7}$$

where $\mathbb{E}_d$ denotes expectation with respect to $\vec{\boldsymbol{\eta}}_{d+1}$. By properties of Ruelle cascades [Pan13b, Theorem 2.9],

$$\frac{1}{N} \log \mathbb{E} \sum_{\alpha\in\mathbb{N}^D} \nu_\alpha \exp \sum_{u\in\mathbb{L}} \sum_{i=1}^{N} g_{\theta,i}^{(u)}(\alpha) = \frac{1}{N} X_0.$$

Here we use that the depth-zero term $\eta_\emptyset(u)\theta(q_0)^{1/2}$ of $g_\theta^{(u)}(\alpha)$ is zero because $\theta(q_0) = 0$. We now evaluate X_0 by (3.3.7). For each $1 \le d \le D$, $\sum_{u\in\mathbb{L}} \sum_{i=1}^{N} \boldsymbol{\eta}_d(u)_i$ has variance

$$\mathbb{E}\left(\sum_{u\in\mathbb{L}} \sum_{i=1}^{N} \boldsymbol{\eta}_d(u)_i\right)^2 = N\mathrm{Sum}(M^d) = NK\kappa(q_{d-1}).$$

So,

$$\begin{aligned}\frac{1}{\zeta_d} \log \mathbb{E}_d \exp \zeta_d \left(\sum_{u\in\mathbb{L}} \sum_{i=1}^{N} \boldsymbol{\eta}_{d+1}(u)_i\right) \left(\theta(q_{d+1}) - \theta(q_d)\right)^{1/2} &= \frac{1}{\zeta_d} \log \exp\left(\frac{NK}{2} \kappa(q_d)\zeta_d^2 (\theta(q_{d+1}) - \theta(q_d))\right) \\ &= \frac{NK}{2} \kappa(q_d)\zeta_d(\theta(q_{d+1}) - \theta(q_d)).\end{aligned}$$

A straightforward induction argument using this computation gives

$$\frac{1}{N} X_0 = \frac{K}{2} \sum_{d=0}^{D-1} \kappa(q_d)\zeta_d(\theta(q_{d+1}) - \theta(q_d)),$$

completing the proof. □

Corollary 3.3.4. *For the distribution function $\zeta : [q_0, 1) \to [0, 1]$ defined in (3.3.3),*

$$\varphi(1) = F_{N,\lambda}(\mathcal{Q}(\eta)) + \frac{K}{2} \int_{q_0}^{1} (q - q_0)\xi''(q)\kappa(q)\zeta(q)\, \mathrm{d}q$$

Proof. On each interval $[q_d, q_{d+1})$, the functions $\kappa(q)$ and $\zeta(q)$ are constant. Moreover, recall that $\theta'(q) = (q - q_0)\xi''(q)$. The result follows from Lemma 3.3.3. □

Proof of Proposition 3.3.1. By Lemma 3.3.2 and Corollary 3.3.4,

$$F_{N,\lambda}(\mathcal{Q}(\eta)) \le \varphi(0) - \frac{K}{2} \int_{q_0}^{1} (q - q_0)\xi''(q)\kappa(q)\zeta(q)\, \mathrm{d}q + 3K^2\xi''(1)\eta.$$

The result follows from (3.3.1). □

In the following two sections, we will use Proposition 3.3.1 to upper bound $F_N(\mathcal{Q}(\eta))$ in the spherical and Ising settings by estimating

$$\varphi(0) = KR(\boldsymbol{h},\boldsymbol{m}) + \frac{1}{N}\log \mathbb{E}\sum_{\alpha\in\mathbb{N}^D}\nu_\alpha \int_{\mathcal{Q}(\eta)} \exp\sum_{u\in\mathbb{L}}\left[\langle \boldsymbol{h}+\lambda\boldsymbol{m}, \pi(\boldsymbol{\sigma}(u))\rangle + \sum_{i=1}^N g_{\xi,i}^{(u)}(\alpha)\pi(\boldsymbol{\sigma}(u))_i\right] \mathrm{d}\mu^K(\vec{\boldsymbol{\sigma}}). \tag{3.3.8}$$

In the spherical and Ising settings, μ is respectively the uniform measure on S_N and the counting measure on Σ_N. We denote $\varphi(0)$ in these settings by $\varphi^{\mathrm{Sp}}(0)$ and $\varphi^{\mathrm{Is}}(0)$. We will also denote F_N in these settings by F_N^{Sp} and F_N^{Is}.

3.4 Overlap-Constrained Upper Bound on the Spherical Grand Hamiltonian

In this section, we complete the proof of Proposition 3.2.2 in the spherical setting. Denote the expected overlap-constrained maximum energy of the grand Hamiltonian by

$$\mathrm{GS}_N^{\mathrm{Sp}}(\mathcal{Q}(\eta)) = \frac{1}{N}\mathbb{E}\max_{\vec{\boldsymbol{\sigma}}\in\mathcal{Q}(\eta)} \mathcal{H}_N(\vec{\boldsymbol{\sigma}}).$$

Let $\underline{\mathscr{L}}$ and $\overline{\mathscr{L}}$ denote the subsets of $\mathscr{L}$ supported on $[0,q_0)$ and $[q_0,1)$, respectively. The function κ defined in (3.3.2) is an element of $\overline{\mathscr{L}}$. Moreover (recall (3.3.3)) $\mathcal{M}_{\vec{q}} \subseteq \overline{\mathscr{L}}$. For $\beta > 0$ and $\zeta \in \mathcal{M}_{\vec{q}}$, let $\beta\kappa\zeta \in \overline{\mathscr{L}}$ denote the pointwise product $\beta\kappa\zeta(q) = \beta\kappa(q)\zeta(q)$. For any $\underline{\zeta} \in \underline{\mathscr{L}}$, let $\underline{\zeta}+\beta\kappa\zeta \in \mathscr{L}$ be the function

$$(\underline{\zeta}+\beta\kappa\zeta)(q) = \begin{cases} \underline{\zeta}(q) & q < q_0, \\ \beta\kappa\zeta(q) & q \geq q_0. \end{cases}$$

We will develop the following bound on $\mathrm{GS}_N^{\mathrm{Sp}}(\mathcal{Q}(\eta))$ for all $D,\vec{k},\vec{p},\vec{q},\eta,\beta$.

Proposition 3.4.1. *Let $\zeta \in \mathcal{M}_{\vec{q}}$ and $\underline{\zeta} \in \underline{\mathscr{L}}$ be arbitrary. Let $\beta > 0$ and suppose that $(B, \underline{\zeta}+\beta\kappa\zeta) \in \mathscr{K}(\xi)$, $B \geq \beta^{-1}$. There exists a constant C, depending only on ξ, h, such that for $N \geq C\log\max(K,2)$,*

$$\mathrm{GS}_N^{\mathrm{Sp}}(\mathcal{Q}(\eta)) \leq K\mathsf{P}^{\mathrm{Sp}}(B,\underline{\zeta}+\beta\kappa\zeta) + CK^2\left(\beta\eta + B\eta + \frac{\log\frac{1}{\eta}}{\beta} + \frac{1}{\sqrt{N}}\right).$$

Crucially, in the input of the Parisi functional, the increasing function ζ is pointwise multiplied by κ, which (by selecting appropriate parameters $\vec{k},\vec{p},\vec{q}$) can be arranged to decrease as rapidly as desired. This multiplication by κ allows us to pass from increasing functions $\zeta \in \mathcal{M}_{\vec{q}}$ to arbitrary bounded variation functions, in the sense that $\beta\kappa\zeta$ can approximate any element of $\overline{\mathscr{L}}$. Consequently, $\underline{\zeta}+\beta\kappa\zeta$ can approximate any element of $\mathscr{L}$, and $\mathsf{P}^{\mathrm{Sp}}(B,\underline{\zeta}+\beta\kappa\zeta)$ can be made arbitrarily close to $\mathsf{ALG}^{\mathrm{Sp}}$. We will prove Proposition 3.2.2 by setting the parameters in Proposition 3.4.1 such that $(B,\underline{\zeta}+\beta\kappa\zeta)$ approximates the minimizer of P^{Sp} and the error term is small.

Our proof of Proposition 3.2.2 proceeds in three steps. In Subsection 3.4.1 we use the machinery of the previous section to prove Proposition 3.4.2, an upper bound on the free energy $F_N^{\mathrm{Sp}}(\mathcal{Q}(\eta))$. In Subsection 3.4.2, we take this bound to low temperature to prove Proposition 3.4.1. In Subsection 3.4.3, we complete the proof of Proposition 3.2.2 by setting appropriate parameters in Proposition 3.4.1.

3.4.1 The Free Energy Upper Bound

In this subsection, we will use Proposition 3.3.1 to upper bound $F_N^{\mathrm{Sp}}(\mathcal{Q}(\eta))$. We take μ to be the uniform measure on S_N. The main result of this subsection is the following upper bound on $F_N^{\mathrm{Sp}}(\mathcal{Q}(\eta))$, which holds for all $D,\vec{k},\vec{p},\vec{q},\eta$.

Proposition 3.4.2. *Let $\zeta \in \mathcal{M}_{\vec{q}}$ and $\underline{\zeta} \in \underline{\mathscr{L}}$ be arbitrary. Suppose $(B, \underline{\zeta} + \kappa\zeta) \in \mathscr{K}(\xi)$, $B \geq 1$, and $N \geq 2$. Then,*

$$F_N^{\mathrm{Sp}}(\mathcal{Q}(\eta)) \leq K\mathsf{P}^{\mathrm{Sp}}(B, \underline{\zeta} + \kappa\zeta) + 3K^2\xi''(1)\eta + KB\eta.$$

The crux of this argument is to upper bound $\varphi^{\mathrm{Sp}}(0)$ so that we may apply Proposition 3.3.1. We equip the state space $(\mathbb{R}^N)^K$ with the natural inner product

$$\langle \vec{\boldsymbol{y}}^1, \vec{\boldsymbol{y}}^2 \rangle = \sum_{u \in \mathbb{L}} \langle \boldsymbol{y}^1(u), \boldsymbol{y}^2(u) \rangle$$

and norm $\|\vec{\boldsymbol{y}}\|^2 = \langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{y}} \rangle$. Generate $\vec{\boldsymbol{\eta}}_0 = (\boldsymbol{\eta}_0(u)) \in (\mathbb{R}^N)^K$ by generating, independently for each $1 \leq i \leq N$,

$$(\vec{\boldsymbol{\eta}}_0)_i = (\boldsymbol{\eta}_0(u)_i)_{u \in \mathbb{L}} \sim \mathcal{N}(0, M^1). \tag{3.4.1}$$

Similarly, for $1 \leq d \leq D$, independently generate $\vec{\boldsymbol{\eta}}_d = (\boldsymbol{\eta}_d(u))_{u \in \mathbb{L}} \in (\mathbb{R}^N)^K$ by generating, independently for each $1 \leq i \leq N$,

$$(\vec{\boldsymbol{\eta}}_d)_i = (\boldsymbol{\eta}_d(u)_i)_{u \in \mathbb{L}} \sim \mathcal{N}(0, M^d). \tag{3.4.2}$$

Let $\vec{\boldsymbol{m}} = (\boldsymbol{m}(u))_{u \in \mathbb{L}} \in (\mathbb{R}^N)^K$ and $\vec{\boldsymbol{h}} = (\boldsymbol{h}(u))_{u \in \mathbb{L}} \in (\mathbb{R}^N)^K$ satisfy $\boldsymbol{m}(u) = \boldsymbol{m}$ and $\boldsymbol{h}(u) = \boldsymbol{h}$ for all $u \in \mathbb{L}$. For $\vec{\boldsymbol{\sigma}} \in (\mathbb{R}^N)^K$, define $\pi(\vec{\boldsymbol{\sigma}}) = \vec{\boldsymbol{\sigma}} - \vec{\boldsymbol{m}}$. We define the following functions on $(\mathbb{R}^N)^K$. Let

$$\begin{aligned} G_D(\vec{\boldsymbol{y}}) &= \log \int_{\mathcal{Q}(\eta)} \exp\langle \vec{\boldsymbol{y}}, \pi(\vec{\boldsymbol{\sigma}}) \rangle \, \mathrm{d}\mu^K(\vec{\boldsymbol{\sigma}}) \\ &= -\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{m}} \rangle + \log \int_{\mathcal{Q}(\eta)} \exp\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{\sigma}} \rangle \, \mathrm{d}\mu^K(\vec{\boldsymbol{\sigma}}). \end{aligned}$$

and for $0 \leq d \leq D-1$, let

$$G_d(\vec{\boldsymbol{y}}) = \frac{1}{\zeta_d} \log \mathbb{E} \exp \zeta_d G_{d+1}\left(\vec{\boldsymbol{y}} + \vec{\boldsymbol{\eta}}_{d+1}(\xi'(q_{d+1}) - \xi'(q_d))^{1/2}\right).$$

By properties of Ruelle cascades,

$$\varphi^{\mathrm{Sp}}(0) = \frac{1}{N} \mathbb{E}\, G_0((\vec{\boldsymbol{h}} + \lambda\vec{\boldsymbol{m}}) + \vec{\boldsymbol{\eta}}_0\xi'(q_0)^{1/2}) + KR(\boldsymbol{h}, \boldsymbol{m}).$$

We will estimate the spherical integral G_D, and through it the functions G_d for $0 \leq d \leq D-1$, by comparison with a Gaussian integral. This step relies on the following lemma, which is a straightforward extension of [Tal06a, Lemma 3.1]; we defer the proof to the end of this section. For $B \geq 1$, let ν_B denote the measure of $\mathcal{N}(0, \frac{1}{B})$. Let $\chi^2(d)$ denote a χ^2 random variable with d degrees of freedom.

Lemma 3.4.3. *For all $\vec{\boldsymbol{y}} \in (\mathbb{R}^N)^K$,*

$$\exp G_D(\vec{\boldsymbol{y}}) \leq \mathbb{P}\left(\chi^2(N) \geq BN\right)^{-K} \exp\left(-\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{m}} \rangle\right) \int \exp\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{\rho}} \rangle \, \mathrm{d}\nu_B^{N \times K}(\vec{\boldsymbol{\rho}}).$$

The probability term in this lemma can be controlled by the following standard bound, whose proof we also defer.

Lemma 3.4.4. *If $B \geq 1$ and $N \geq 2$, then*

$$\mathbb{P}(\chi^2(N) \geq BN) \geq \exp(-BN/2).$$

It remains to analyze the terms in Lemma 3.4.3 involving $\vec{\boldsymbol{y}}$. Define further

$$\begin{aligned} G_D'(\vec{\boldsymbol{y}}) &= -\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{m}} \rangle + \log \int \exp\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{\rho}} \rangle \, \mathrm{d}\nu_b^K(\vec{\boldsymbol{\rho}}) = \frac{\|\vec{\boldsymbol{y}}\|_2^2}{2B} - \langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{m}} \rangle, \\ G_d'(\vec{\boldsymbol{y}}) &= \frac{1}{\zeta_d} \log \mathbb{E} \exp \zeta_d G_{d+1}\left(\vec{\boldsymbol{y}} + \vec{\boldsymbol{\eta}}_{d+1}(\xi'(q_{d+1}) - \xi'(q_d))^{1/2}\right) \qquad \text{for } 0 \leq d \leq D-1, \end{aligned}$$

Henceforth, suppose $N \geq 2$. Lemmas 3.4.3 and 3.4.4 imply that

$$\varphi^{\mathrm{Sp}}(0) \leq \frac{1}{N}\,\mathbb{E}\,G_0'((\vec{\boldsymbol{h}} + \lambda\vec{\boldsymbol{m}}) + \vec{\boldsymbol{\eta}}_0\xi'(q_0)^{1/2}) + KR(\boldsymbol{h},\boldsymbol{m}) + \frac{1}{2}KB. \tag{3.4.3}$$

Consider a new state space $\mathbb{R}^K$ with elements $\vec{y} = (y(u))_{u\in\mathbb{L}}$ where $y(u) \in \mathbb{R}$, equipped with the natural inner product

$$\langle \vec{y}^1, \vec{y}^2 \rangle = \sum_{u\in\mathbb{L}} \vec{y}^1(u)\vec{y}^2(u)$$

and norm $\|\vec{y}\|_2^2 = \langle \vec{y}, \vec{y}\rangle$. Generate the $\mathbb{R}^K$-valued Gaussians $e\vec{t}a_0 \sim \mathcal{N}(0, M^1)$ and, for $1 \leq d \leq D$, $e\vec{t}a_d \sim \mathcal{N}(0, M^d)$. Recall that $\boldsymbol{h} = (h, \ldots, h)$. Let $\boldsymbol{m} = (m_1, \ldots, m_N)$, and let $\vec{1} \in \mathbb{R}^K$ denote the all-1 vector. For $1 \leq i \leq N$, define the following functions on $\mathbb{R}^K$.

$$\begin{aligned}
\Gamma_D^i(\vec{y}) &= \frac{\|\vec{y}\|^2}{2B} - m_i\langle \vec{1}, \vec{y}\rangle, \\
\Gamma_d^i(\vec{y}) &= \frac{1}{\zeta_d}\log \mathbb{E}\exp\zeta_d\Gamma_{d+1}^i\left(\vec{y} + e\vec{t}a_{d+1}(\xi'(q_{d+1}) - \xi'(q_d))^{1/2}\right) \qquad \text{for } 0 \leq d \leq D-1.
\end{aligned}$$

By independence of the $1 \leq i \leq N$ coordinates in the G_d', (3.4.3) implies

$$\varphi^{\mathrm{Sp}}(0) \leq \frac{1}{N}\sum_{i=1}^N \mathbb{E}\,\Gamma_0^i((h+\lambda m_i)\vec{1} + e\vec{t}a_0\xi'(q_0)^{1/2}) + KR(\boldsymbol{h},\boldsymbol{m}) + \frac{1}{2}KB. \tag{3.4.4}$$

It remains to compute the Gaussian integrals Γ_d^i. For this, we rely on the following lemma. We defer the proof, which is a standard computation with Gaussian integrals. Let $\mathbb{S}_K$ denote the set of $K \times K$ positive definite matrices, and let $|\cdot|$ denote the matrix determinant.

Lemma 3.4.5. *Suppose $\zeta > 0$ and $\Lambda, \Sigma \in \mathbb{S}_K$ satisfy $\Lambda - \zeta\Sigma \in \mathbb{S}_K$. If $\vec{v} \in \mathbb{R}^K$ and $e\vec{t}a \sim \mathcal{N}(0,\Sigma)$, then*

$$\begin{aligned}
&\frac{1}{\zeta}\log\mathbb{E}\exp\frac{1}{2}\zeta\left[(\vec{y}+e\vec{t}a)^\top\Lambda^{-1}(\vec{y}+e\vec{t}a) - 2\vec{v}^\top(\vec{y}+e\vec{t}a)\right] \\
&= \frac{1}{2}\left[\vec{y}^\top(\Lambda-\zeta\Sigma)^{-1}\vec{y} - 2\vec{v}^\top\Lambda(\Lambda-\zeta\Sigma)^{-1}\vec{y}\right] + \frac{1}{2\zeta}\log\frac{|\Lambda|}{|\Lambda-\zeta\Sigma|} + \frac{1}{2}\vec{v}^\top(\zeta\Sigma)(\Lambda-\zeta\Sigma)^{-1}\Lambda\vec{v}.
\end{aligned}$$

We can compute the expectations in (3.4.4) by applying this lemma recursively. Define

$$\overline{\mathscr{K}}(\xi) = \left\{(B,\zeta) \in \mathbb{R}^+ \times \overline{\mathscr{L}} : B > \int_{q_0}^1 \xi''(q')\zeta(q')\,\mathrm{d}q'\right\}.$$

Proposition 3.4.6. *Let $\zeta \in \mathcal{M}_{\vec{q}}$, and suppose $(B,\kappa\zeta) \in \overline{\mathscr{K}}(\xi)$. Then, for $B_{\kappa\zeta}$ defined as in (3.1.3),*

$$\mathbb{E}\,\Gamma_0^i((h+\lambda m_i)\vec{1} + e\vec{t}a_0\xi'(q_0)^{1/2}) \leq \frac{K}{2}\left[\frac{(h+(\lambda-B)m_i)^2 + \xi'(q_0)}{B_{\kappa\zeta}(q_0)} + \int_{q_0}^1 \frac{\xi''(q)}{B_{\kappa\zeta}(q)}\,\mathrm{d}q - Bm_i^2\right].$$

Proof. Let $\Lambda_D = BI_K$, and for $0 \leq d \leq D-1$, let

$$\Lambda_d = \Lambda_{d+1} - \zeta_d(\xi'(q_{d+1}) - \xi'(q_d))M^{d+1}.$$

We will first show that $\Lambda_0, \ldots, \Lambda_D \in \mathbb{S}_K$, so that we can apply Lemma 3.4.5. For $q \in [q_0, 1]$, we define

$$\Lambda(q) = BI_K - \int_q^1 \xi''(q')M(q')\zeta(q')\,\mathrm{d}q'.$$

Note that $\Lambda_d = \Lambda(q_d)$ for all $0 \leq d \leq D$. Since $M(q) \preceq \kappa(q)I_K$ in the Loewner order,

$$\Lambda(q) \succeq \left(B - \int_q^1 \xi''(q')\kappa(q')\zeta(q')\,\mathrm{d}q'\right)I_K = B_{\kappa\zeta}(q)I_K. \tag{3.4.5}$$

So, the hypothesis $(B, \kappa\zeta) \in \overline{\mathscr{K}}(\xi)$ implies $\Lambda(q) \in \mathbb{S}_K$ for all $q \in [q_0, 1]$. In particular $\Lambda_0, \ldots, \Lambda_D \in \mathbb{S}_K$.

Further, define $\vec{v}_D = m_i\vec{1}$, and for $0 \le d \le D-1$, define $\vec{v}_d = \Lambda_d^{-1}\Lambda_{d+1}\vec{v}_{d+1}$. This implies that $\vec{v}_d = Bm_i\Lambda_d^{-1}\vec{1}$. We can write Γ_D^i as

$$\Gamma_D^i(\vec{y}) = \frac{1}{2}\left(\vec{y}^\top\Lambda_D^{-1}\vec{y} - 2\vec{v}_D^\top\vec{y}\right).$$

By a recursive computation with Lemma 3.4.5 (which applies because $\Lambda_0, \ldots, \Lambda_D \in \mathbb{S}_K$), we have for all $0 \le d \le D$ that

$$\begin{aligned}\Gamma_d^i(\vec{y}) &= \frac{1}{2}\left[\vec{y}^\top\Lambda_d^{-1}\vec{y} - 2\vec{v}_d^\top\vec{y} + \sum_{d'=d}^{D-1}\frac{1}{\zeta_{d'}}\log\frac{|\Lambda_{d'+1}|}{|\Lambda_{d'}|} + \sum_{d'=d}^{D-1}\vec{v}_{d'+1}(\Lambda_{d'+1}-\Lambda_{d'})\Lambda_{d'}^{-1}\Lambda_{d'+1}\vec{v}_{d'+1}\right] \\ &= \frac{1}{2}\left[\vec{y}^\top\Lambda_d^{-1}\vec{y} - 2Bm_i\vec{1}^\top\Lambda_d^{-1}\vec{y} + \sum_{d'=d}^{D-1}\frac{1}{\zeta_{d'}}\log\frac{|\Lambda_{d'+1}|}{|\Lambda_{d'}|} + B^2m_i^2\sum_{d'=d}^{D-1}\vec{1}^\top\Lambda_{d'+1}^{-1}(\Lambda_{d'+1}-\Lambda_{d'})\Lambda_{d'}^{-1}\vec{1}\right].\end{aligned}$$

Note that

$$\sum_{d'=d}^{D-1}\vec{1}^\top\Lambda_{d'+1}^{-1}(\Lambda_{d'+1}-\Lambda_{d'})\Lambda_{d'}^{-1}\vec{1} = \sum_{d'=d}^{D-1}\vec{1}^\top(\Lambda_{d'}^{-1}-\Lambda_{d'+1}^{-1})\vec{1} = \vec{1}^\top(\Lambda_d^{-1}-\Lambda_D^{-1})\vec{1} = \vec{1}^\top\Lambda_d^{-1}\vec{1} - \frac{K}{B}.$$

So,

$$\begin{aligned}\Gamma_0^i(\vec{y}) &= \frac{1}{2}\left[\vec{y}^\top\Lambda_0^{-1}\vec{y} - 2Bm_i\vec{1}^\top\Lambda_0^{-1}\vec{y} + B^2m_i^2\vec{1}^\top\Lambda_0^{-1}\vec{1} + \sum_{d=0}^{D-1}\frac{1}{\zeta_d}\log\frac{|\Lambda_{d+1}|}{|\Lambda_d|} - KBm_i^2\right] \\ &= \frac{1}{2}\left[(\vec{y}-Bm_i\vec{1})^\top\Lambda_0^{-1}(\vec{y}-Bm_i\vec{1}) + \sum_{d=0}^{D-1}\frac{1}{\zeta_d}\log\frac{|\Lambda_{d+1}|}{|\Lambda_d|} - KBm_i^2\right]\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathbb{E}\,\Gamma_0^i((h+\lambda m_i)\vec{1} + e\vec{t}a_0\xi'(q_0)^{1/2}) \\ &= \frac{1}{2}\left[(h+(\lambda-B)m_i)^2\mathrm{Tr}(\Lambda_0^{-1}\vec{1}\vec{1}^\top) + \xi'(q_0)\mathrm{Tr}(\Lambda_0^{-1}M^1) + \sum_{d=0}^{D-1}\frac{1}{\zeta_d}\log\frac{|\Lambda_{d+1}|}{|\Lambda_d|} - KBm_i^2\right].\end{aligned}$$

By Jacobi's formula,

$$\frac{\mathrm{d}}{\mathrm{d}q}\log|\Lambda(q)| = \xi''(q)\zeta(q)\mathrm{Tr}(\Lambda(q)^{-1}M(q)),$$

so

$$\frac{1}{\zeta_d}\log\frac{|\Lambda_{d+1}|}{|\Lambda_d|} = \int_{q_d}^{q_{d+1}}\xi''(q)\mathrm{Tr}(\Lambda(q)^{-1}M(q))\,\mathrm{d}q.$$

Therefore,

$$\begin{aligned}&\mathbb{E}\,\Gamma_0^i((h+\lambda m_i)\vec{1} + e\vec{t}a_0\xi'(q_0)^{1/2}) \\ &= \frac{1}{2}\left[(h+(\lambda-B)m_i)^2\mathrm{Tr}(\Lambda(q_0)^{-1}\vec{1}\vec{1}^\top) + \xi'(q_0)\mathrm{Tr}(\Lambda(q_0)^{-1}M(q_0)) + \int_{q_0}^{1}\mathrm{Tr}(\Lambda(q)^{-1}M(q))\,\mathrm{d}q - KBm_i^2\right].\end{aligned}$$

Finally, for each $q \in [q_0, 1)$, (3.4.5) implies $\Lambda(q)^{-1} \preceq \frac{I_K}{B_{\kappa\zeta}(q)}$, so

$$\mathrm{Tr}(\Lambda(q)^{-1}M(q)) \le \mathrm{Tr}\left(\frac{M(q)}{B_{\kappa\zeta}(q)}\right) = \frac{K}{B_{\kappa\zeta}(q)},$$

and similarly $\mathrm{Tr}(\Lambda(q_0)^{-1}\vec{1}\vec{1}^\top) \le \frac{K}{B_{\kappa\zeta}(q_0)}$. This implies the result. □

Proposition 3.4.6 and (3.4.4) readily imply the following bound on $F_N^{\mathrm{Sp}}(\mathcal{Q}(\eta))$.

Proposition 3.4.7. *Let $B \geq 1$, $N \geq 2$, and $\lambda \in \mathbb{R}$. Let $\zeta \in \mathcal{M}_{\vec{q}}$, and suppose $(B, \kappa\zeta) \in \overline{\mathscr{K}}(\xi)$. Then,*

$$F_N^{\mathrm{Sp}}(\mathcal{Q}(\eta)) \leq \frac{K}{2}\left[\frac{\|\boldsymbol{h} + (\lambda - B)\boldsymbol{m}\|_N^2 + \xi'(q_0)}{B_{\kappa\zeta}(q_0)} + 2R(\boldsymbol{h}, \boldsymbol{m}) + \int_{q_0}^1 \left(\frac{\xi''(q)}{B_{\kappa\zeta}(q)} + B_{\kappa\zeta}(q)\right) \mathrm{d}q\right] + 3K^2\xi''(1)\eta + K|\lambda|\eta.$$

Proof. By averaging Proposition 3.4.6 over $1 \leq i \leq N$, we get

$$\frac{1}{N}\sum_{i=1}^N \mathbb{E}\,\Gamma_0^i((h + \lambda m_i)\vec{1} + e\vec{t}a_0\xi'(q_0)^{1/2}) \leq \frac{K}{2}\left[\frac{\|\boldsymbol{h} + (\lambda - B)\boldsymbol{m}\|_N^2 + \xi'(q_0)}{B_{\kappa\zeta}(q_0)} + \int_{q_0}^1 \frac{\xi''(q)}{B_{\kappa\zeta}(q)}\,\mathrm{d}q - Bq_0\right]$$

where we used that $\|\boldsymbol{m}\|_N^2 = q_0$. Equation (3.4.4) implies that

$$\varphi^{\mathrm{Sp}}(0) \leq \frac{K}{2}\left[\frac{\|\boldsymbol{h} + (\lambda - B)\boldsymbol{m}\|_N^2 + \xi'(q_0)}{B_{\kappa\zeta}(q_0)} + 2R(\boldsymbol{h}, \boldsymbol{m}) + \int_{q_0}^1 \frac{\xi''(q)}{B_{\kappa\zeta}(q)}\,\mathrm{d}q + (1 - q_0)B\right].$$

By Proposition 3.3.1, this implies

$$F_N^{\mathrm{Sp}}(\mathcal{Q}(\eta)) \leq \frac{K}{2}\left[\frac{\|\boldsymbol{h} + (\lambda - B)\boldsymbol{m}\|_N^2 + \xi'(q_0)}{B_{\kappa\zeta}(q_0)} + 2R(\boldsymbol{h}, \boldsymbol{m}) + \int_{q_0}^1 \frac{\xi''(q)}{B_{\kappa\zeta}(q)}\,\mathrm{d}q + (1 - q_0)B - \int_{q_0}^1 (q - q_0)\xi''(q)\kappa(q)\zeta(q)\,\mathrm{d}q\right] + 3K^2\xi''(1)\eta + K|\lambda|\eta$$

By integration by parts,

$$\begin{aligned}-\int_{q_0}^1 (q - q_0)\xi''(q)\kappa(q)\zeta(q)\,\mathrm{d}q &= (q - q_0)\int_q^1 \xi''(q')\kappa(q')\zeta(q')\,\mathrm{d}q'\Big|_{q=q_0}^1 - \int_{q_0}^1\int_q^1 \xi''(q')\kappa(q')\zeta(q')\,\mathrm{d}q'\,\mathrm{d}q \\ &= \int_{q_0}^1 B_{\kappa\zeta}(q)\,\mathrm{d}q - (1 - q_0)B,\end{aligned}$$

which yields the result. □

The next lemma upper bounds our estimates for $F_N^{\mathrm{Sp}}(\mathcal{Q}(\eta))$ in terms of the Parisi functional uniformly in $\boldsymbol{m}$.

Lemma 3.4.8. *Let $q_0 \in [0, 1]$. For $(B, \zeta) \in \mathscr{K}(\xi)$, $\boldsymbol{h} = (h, \dots, h)$, $\|\boldsymbol{m}\|_N^2 = q_0$, there exists $\lambda \in [0, B]$ such that*

$$\frac{1}{2}\left[\frac{\|\boldsymbol{h} + (\lambda - B)\boldsymbol{m}\|_N^2 + \xi'(q_0)}{B_\zeta(q_0)} + 2R(\boldsymbol{h}, \boldsymbol{m}) + \int_{q_0}^1 \left(\frac{\xi''(q)}{B_\zeta(q)} + B_\zeta(q)\right) \mathrm{d}q\right] \leq \mathsf{P}^{\mathrm{Sp}}(\zeta).$$

Proof. We take $\lambda = \int_0^1 \xi''(q)\zeta(q)\,\mathrm{d}q$. The condition $(B, \zeta) \in \mathscr{K}(\xi)$ implies that $\lambda \in [0, B]$. Note that $\lambda - B = -B_\zeta(0)$. It suffices to prove that

$$\frac{\|\boldsymbol{h} - B_\zeta(0)\boldsymbol{m}\|_N^2 + \xi'(q_0)}{B_\zeta(q_0)} + 2R(\boldsymbol{h}, \boldsymbol{m}) \leq \frac{\|\boldsymbol{h}\|_N^2}{B_\zeta(0)} + \int_0^{q_0} \left(\frac{\xi''(q)}{B_\zeta(q)} + B_\zeta(q)\right) \mathrm{d}q.$$

Note that

$$\frac{\xi'(q_0)}{B_\zeta(q_0)} = \int_0^{q_0} \frac{\xi''(q)}{B_\zeta(q_0)}\,\mathrm{d}q \leq \int_0^{q_0} \frac{\xi''(q)}{B_\zeta(q)}\,\mathrm{d}q \qquad \text{and} \qquad q_0 B_\zeta(0) \leq \int_0^{q_0} B_\zeta(q)\,\mathrm{d}q.$$

So, it suffices to prove that

$$\frac{\|\boldsymbol{h} - B_\zeta(0)\boldsymbol{m}\|_N^2}{B_\zeta(q_0)} + 2R(\boldsymbol{h}, \boldsymbol{m}) \leq \frac{\|\boldsymbol{h}\|_N^2}{B_\zeta(0)} + q_0 B_\zeta(0).$$

This rearranges to (using that $\|\boldsymbol{m}\|_N^2 = q_0$)

$$0 \le \left(\frac{1}{B_\zeta(0)} - \frac{1}{B_\zeta(q_0)}\right)\left(\|\boldsymbol{h}\|_N^2 - 2B_\zeta(0)R(\boldsymbol{h},\boldsymbol{m}) + B_\zeta(0)^2\|\boldsymbol{m}\|_N^2\right),$$

which follows from Cauchy-Schwarz. □

We are now ready to prove Proposition 3.4.2.

Proof of Proposition 3.4.2. Recall that the restriction of $\underline{\zeta} + \kappa\zeta \in \mathscr{L}$ on $[q_0, 1)$ is $\kappa\zeta$. Because $(B, \underline{\zeta} + \kappa\zeta) \in \mathscr{K}(\xi)$, we have $(B, \kappa\zeta) \in \overline{\mathscr{K}}(\xi)$, and so Proposition 3.4.7 applies. Combining this with Lemma 3.4.8 applied on $(B, \underline{\zeta} + \kappa\zeta)$ gives the result. □

3.4.2 From Free Energy to Ground State Energy

Next, we will prove Proposition 3.4.1 by taking Proposition 3.4.2 to low temperature. We introduce the following temperature-scaled free energy. For $\beta > 0$ and $\eta \in (0,1)$, let

$$F_N^{\mathrm{Sp}}(\beta, \mathcal{Q}(\eta)) = \frac{1}{N}\log \mathbb{E}\int_{\mathcal{Q}(\eta)} \exp \beta\mathcal{H}_N(\vec{\boldsymbol{\sigma}})\, \mathrm{d}\mu^K(\vec{\boldsymbol{\sigma}}).$$

This free energy can be upper bounded by the following application of Proposition 3.4.2.

Corollary 3.4.9. *Let $\zeta \in \mathcal{M}_{\vec{q}}$ and $\underline{\zeta} \in \underline{\mathscr{L}}$ be arbitrary. Let $\beta > 0$ and suppose $(B, \underline{\zeta} + \beta\kappa\zeta) \in \mathscr{K}(\xi)$, $B \ge \beta^{-1}$, and $N \ge 2$. Then,*

$$\frac{1}{\beta}F_N^{\mathrm{Sp}}(\beta, \mathcal{Q}(\eta)) \le K\mathsf{P}^{\mathrm{Sp}}(B, \underline{\zeta} + \beta\kappa\zeta) + 3K^2\xi''(1)\beta\eta + KB\eta.$$

Proof. The hypothesis $(B, \underline{\zeta} + \beta\kappa\zeta) \in \mathscr{K}(\xi)$ implies $(\beta B, \beta^{-1}\underline{\zeta} + \kappa\zeta) \in \mathscr{K}(\beta^2\xi)$. The hypothesis $B \ge \beta^{-1}$ implies $\beta B \ge 1$. By Proposition 3.4.2 with parameters $(\beta^2\xi, \beta h)$ (corresponding to the Hamiltonian βH_N), ζ, βB, and $\beta^{-1}\underline{\zeta}$,

$$F_N^{\mathrm{Sp}}(\beta, \mathcal{Q}(\eta)) \le K\mathsf{P}^{\mathrm{Sp}}_{\beta^2\xi,\beta h}(\beta B, \beta^{-1}\underline{\zeta} + \kappa\zeta) + 3K^2\xi''(1)\beta^2\eta + KB\beta\eta. \tag{3.4.6}$$

We can verify that

$$\mathsf{P}^{\mathrm{Sp}}_{\beta^2\xi,\beta h}(\beta B, \beta^{-1}\underline{\zeta} + \kappa\zeta) = \beta\mathsf{P}^{\mathrm{Sp}}_{\xi,h}(B, \underline{\zeta} + \beta\kappa\zeta).$$

So, dividing (3.4.6) by β gives the result. □

The following lemma relates the ground state energy $\mathrm{GS}_N^{\mathrm{Sp}}(\mathcal{Q}(\eta))$ to this free energy at large inverse temperature β. We defer the proof, which is a relatively standard approximation argument.

Lemma 3.4.10. *There exists a constant C depending only on ξ, h such that for all $\beta > 0$, $\eta \in (0, \frac{1}{2})$, and $N \ge C\log\max(K, 2)$,*

$$\mathrm{GS}_N^{\mathrm{Sp}}(\mathcal{Q}(\eta)) \le \frac{1}{\beta}F_N^{\mathrm{Sp}}(\beta, \mathcal{Q}(2\eta)) + CK\left(\eta + \frac{\log\frac{1}{\eta}}{\beta} + \frac{1}{\sqrt{N}}\right).$$

Proof of Proposition 3.4.1. Let C be large enough that Lemma 3.4.10 is satisfied and $C\log 2 \ge 2$. For all $N \ge C\log\max(K,2)$, Corollary 3.4.9 (with 2η in place of η) and Lemma 3.4.10 imply that

$$\mathrm{GS}_N^{\mathrm{Sp}}(\eta) \le K\mathsf{P}^{\mathrm{Sp}}(B, \underline{\zeta} + \beta\kappa\zeta) + 6K^2\xi''(1)\beta\eta + 2KB\eta + CK\left(\eta + \frac{\log\frac{1}{\eta}}{\beta} + \frac{1}{\sqrt{N}}\right).$$

By applying the estimate $K \le K^2$ and absorbing constants depending on only ξ, h into C, we deduce

$$\mathrm{GS}_N^{\mathrm{Sp}}(\eta) \le K\mathsf{P}^{\mathrm{Sp}}(B, \underline{\zeta} + \beta\kappa\zeta) + CK^2\left(\beta\eta + B\eta + \eta + \frac{\log\frac{1}{\eta}}{\beta} + \frac{1}{\sqrt{N}}\right).$$

Finally, because $B \ge \beta^{-1}$, we have $\beta + B \ge \beta + \beta^{-1} \ge 2$, so by increasing the constant C we may drop the term η from the sum. □

3.4.3 Proof of the Main Upper Bound

We now complete the proof of Proposition 3.2.2. We will set the parameters of Proposition 3.4.1 such that $(B, \zeta + \beta\kappa\zeta)$ approximates the minimizer of P^{Sp} in $\mathscr{L}$ and the error term is small.

For $\zeta \in \mathscr{L}$ and $\delta, x \in [0,1)$, we define a perturbation $\zeta_{\delta,x} \in \mathscr{L}$ of ζ by

$$\zeta_{\delta,x}(q) = \begin{cases} \zeta(x+\delta) & q \in [x, x+\delta), \\ \zeta(q) & \text{otherwise.} \end{cases}$$

Note that $\zeta_{0,x} = \zeta$.

We now set several constants depending only on ξ, h, ε. Let C be the constant given by Proposition 3.4.1. By continuity of the Parisi functional P^{Sp} on $\mathscr{K}(\xi)$, we may pick $(B^*, \zeta^*) \in \mathscr{K}(\xi)$ and a small constant $\Delta \in (0,1)$ such that the following properties hold.

(a) ζ^* is positive-valued, right-continuous, and piecewise constant with finitely many jump discontinuities $0 < x_1 < \cdots < x_r < 1$.

(b) For all $\delta \in [0, \Delta]$ and $x \in [0,1)$, $(B^*, \zeta^*_{\delta,x}) \in \mathscr{K}(\xi)$ and

$$\mathsf{P}^{\mathrm{Sp}}(B^*, \zeta^*_{\delta,x}) \le \mathsf{ALG} + \frac{\varepsilon}{2}. \tag{3.4.7}$$

The perturbations $\zeta^*_{\delta,x}$ will be used in the following way. Given $q_0 \in [0,1]$, we will apply Proposition 3.4.1 with $\underline{\zeta} + \beta\kappa\zeta = \zeta^*_{(1-q_0)\Delta,q_0}$. In particular, we will construct β, $\kappa = \kappa^{\vec{k},\vec{p},\vec{q}}$ and $\zeta \in \mathcal{M}_{\vec{q}}$ such that $\beta\kappa\zeta = \zeta^*_{(1-q_0)\Delta,q_0}$ on $[q_0, 1)$. Because ζ is increasing, we must construct a κ that decreases rapidly enough to make this equality hold. In the below proof, the fact that $\zeta^*_{(1-q_0)\Delta,q_0}$ does not have any discontinuities in $[q_0, q_0 + (1-q_0)\Delta]$ implies that $q_1 > q_0 + (1-q_0)\Delta$, which implies that $p_1 > \Delta$ for any χ-aligned $\vec{p}, \vec{q}$. This allows us to construct a suitable κ while keeping $K = \prod_{d=1}^{D} k_d$ bounded by a constant.

Proof of Proposition 3.2.2, spherical case. We first set the constants K_0, η_0, N_0. For $x \in (0,1]$, let $\zeta^*(x^-) = \lim_{y \to x^-} \zeta^*(y)$. Let

$$K_0 = \prod_{j=1}^{r} \left(\left\lfloor \frac{\zeta^*(x_j)}{\Delta \zeta^*(x_j^-)} \right\rfloor + 1 \right).$$

This is well-defined because ζ^* is positive-valued. Let $\eta_0 \in (0, \frac{1}{2})$ satisfy the inequalities

$$CK_0 \left(B^* \eta_0 + \eta_0^{1/2} + \eta_0^{1/2} \log \frac{1}{\eta_0} \right) \le \frac{\varepsilon}{4}, \tag{3.4.8}$$

$$\eta_0 \le (B^*)^2, \tag{3.4.9}$$

$$\eta_0 < \zeta^*(1^-)^{-2}. \tag{3.4.10}$$

Finally, let N_0 satisfy $N_0 \ge C \log \max(K_0, 2)$ and

$$\frac{CK_0}{\sqrt{N_0}} \le \frac{\varepsilon}{4}. \tag{3.4.11}$$

We emphasize that K_0, η_0, N_0 depend only on ξ, h, ε.

In the below analysis, we always set $\eta = \eta_0$ (this clearly satisfies $\eta \ge \eta_0$) and $\beta = \eta_0^{-1/2}$.

We are given a correlation function $\chi : [0,1] \to [0,1]$ and a point $\boldsymbol{m} \in \mathbb{R}^N$ with $\|\boldsymbol{m}\|_N^2 = \chi(0)$. We set $q_0 = \chi(0)$; we will set the rest of $\vec{q}$ below. We will construct $D, \vec{k}, \vec{p}, \vec{q}, \zeta$ such that on $[q_0, 1)$,

$$\beta \kappa^{\vec{k},\vec{p},\vec{q}} \zeta = \zeta^*_{(1-q_0)\Delta,q_0}. \tag{3.4.12}$$

Let

$$S = \{x_1, \ldots, x_r\} \cap (q_0 + (1-q_0)\Delta, 1).$$

Set $D-1=|S|$. Set $\vec{q}$ such that $(q_1,\dots,q_{D-1})$ is the set S in increasing order and $q_D=1$.

By Proposition 3.2.1(ii), χ is either strictly increasing or constant. If χ is strictly increasing, set $\vec{p}=(p_0,\dots,p_D)$ by $p_d=\chi^{-1}(q_d)$ for all $q_d\le\chi(1)$ and $p_d=1$ for all $q_d>\chi(1)$. If χ is constant, its unique value is $q_0=\chi(0)$; set $p_0=0$ and $p_d=1$ for all $1\le d\le D$. In either case, $\vec{p},\vec{q}$ are clearly χ-aligned. Moreover, we always have $p_1>\Delta$: if χ is increasing, this follows from $q_1>q_0+(1-q_0)\Delta$ and Proposition 3.2.1(iii), while if χ is constant this is obvious.

Set $k_1=1$, and for $1\le d\le D-1$, set

$$k_{d+1}=\left\lfloor\frac{\zeta^*(q_d^-)}{\Delta\zeta^*(q_d)}\right\rfloor+1.$$

Because $q_1,\dots,q_{D-1}$ are a subset of $x_1,\dots,x_r$, we indeed have $K=\prod_{d=1}^D k_d\le K_0$.

This constructs $D,\vec{k},\vec{p},\vec{q},\eta$, which defines $\mathcal{H}_N^{\vec{k},\vec{p}}$, $\mathcal{Q}(\eta)=\mathcal{Q}^{\mathrm{Sp}}(Q^{\vec{k},\vec{q}},\boldsymbol{m},\eta)$, and $\kappa^{\vec{k},\vec{p},\vec{q}}$. Finally, we construct the sequence $(\zeta_{-1},\zeta_0,\dots,\zeta_D)$ satisfying

$$0=\zeta_{-1}<\zeta_0<\cdots<\zeta_D=1 \tag{3.4.13}$$

such that the $\zeta\in\mathcal{M}_{\vec{q}}$ defined by (3.3.3) satisfies (3.4.12) on $[q_0,1)$. In particular, we define ζ_d for $0\le d\le D-1$ by

$$\zeta_d=\frac{\zeta^*_{(1-q_0)\Delta,q_0}(q_d)}{\beta\kappa^{\vec{k},\vec{p},\vec{q}}(q_d)}.$$

For this choice of ζ_d, (3.4.12) holds at $q_0,q_1,\dots,q_{d-1}$ by inspection. Because ζ, $\kappa^{\vec{k},\vec{p},\vec{q}}$ and $\zeta^*_{(1-q_0)\Delta,q_0}$ are all piecewise constant and right-continuous on $[q_0,1)$ with jump discontinuities only at $q_1,\dots,q_{D-1}$, (3.4.12) holds on $[q_0,1)$. It remains to verify that this choice of ζ_d satisfies the increasing condition (3.4.13). Because $\zeta^*_{(1-q_0)\Delta,q_0}$ is positive-valued, $\zeta_0>\zeta_{-1}=0$. At each $1\le d\le D-1$, we have

$$\frac{\zeta_d}{\zeta_{d-1}}=\frac{\zeta^*_{(1-q_0)\Delta,q_0}(q_d)}{\zeta^*_{(1-q_0)\Delta,q_0}(q_{d-1})}\cdot\frac{\kappa^{\vec{k},\vec{p},\vec{q}}(q_{d-1})}{\kappa^{\vec{k},\vec{p},\vec{q}}(q_d)}$$

By (3.3.2),

$$\kappa^{\vec{k},\vec{p},\vec{q}}(q_d)\le\sum_{j=d+1}^{D-1}\left[(k_{j+1}-1)\prod_{\ell=j+2}^{D}k_\ell\right]+1=\prod_{\ell=d+2}^{D}k_\ell,$$

where we upper bounded all the p_d by 1. So,

$$\frac{\kappa^{\vec{k},\vec{p},\vec{q}}(q_{d-1})}{\kappa^{\vec{k},\vec{p},\vec{q}}(q_d)}=1+\frac{(k_{d+1}-1)\prod_{\ell=d+2}^{D}k_\ell}{\kappa^{\vec{k},\vec{p},\vec{q}}(q_d)}p_d\ge 1+(k_{d+1}-1)p_d\ge k_{d+1}p_d\ge k_{d+1}\Delta.$$

Here we used that $p_d\ge p_1\ge\Delta$. Further noting that $\zeta^*_{(1-q_0)\Delta,q_0}(q_{d-1})=\zeta^*(q_d^-)$, we have

$$\frac{\zeta_d}{\zeta_{d-1}}\ge\frac{\Delta\zeta^*_{(1-q_0)\Delta,q_0}(q_d)}{\zeta^*_{(1-q_0)\Delta,q_0}(q_{d-1})}\cdot k_{d+1}=\frac{\Delta\zeta^*(q_d)}{\zeta^*(q_d^-)}\cdot k_{d+1}>1$$

by definition of k_{d+1}. Thus $\zeta_d>\zeta_{d-1}$ for $1\le d\le D-1$. Finally, because $\kappa^{\vec{k},\vec{p},\vec{q}}(q_{D-1})=1$,

$$\zeta_{D-1}=\frac{\zeta^*_{(1-q_0)\Delta,q_0}(q_{D-1})}{\beta}=\eta_0^{1/2}\zeta^*(1^-)<1=\zeta_D,$$

using (3.4.10). Thus the ζ we constructed satisfies (3.4.12) and (3.4.13).

Define $\underline{\zeta}\in\underline{\mathscr{L}}$ by $\underline{\zeta}=\zeta^*$ on $[0,q_0)$. Thus, as elements of $\mathscr{L}$,

$$\underline{\zeta}+\beta\kappa^{\vec{k},\vec{p},\vec{q}}\zeta=\zeta^*_{(1-q_0)\Delta,q_0}.$$

By construction, $(B^*, \zeta^*_{(1-q_0)\Delta,q_0}) \in \mathscr{K}(\xi)$, and (3.4.9) implies $B^* \geq \beta^{-1}$. By Proposition 3.4.1,

$$\frac{1}{N}\,\mathbb{E}\max_{\vec{\boldsymbol{\sigma}}\in\mathcal{Q}(\eta)} \mathcal{H}_N(\vec{\boldsymbol{\sigma}}) \leq K\mathsf{P}^{\mathrm{Sp}}(B^*, \zeta^*_{(1-q_0)\Delta,q_0}) + CK^2\left(B^*\eta + \eta^{1/2} + \eta^{1/2}\log\frac{1}{\eta} + \frac{1}{\sqrt{N}}\right).$$

By (3.4.7),

$$K\mathsf{P}^{\mathrm{Sp}}(B^*, \zeta^*_{(1-q_0)\delta,q_0}) \leq K\left(\mathsf{ALG} + \frac{\varepsilon}{2}\right).$$

By (3.4.8),

$$CK^2\left(B^*\eta + \eta^{1/2} + \eta^{1/2}\log\frac{1}{\eta}\right) \leq \frac{K\varepsilon}{4}.$$

Finally, by (3.4.11),

$$\frac{CK^2}{\sqrt{N}} \leq \frac{K\varepsilon}{4}.$$

Combining the last four inequalities gives the result. □

3.4.4 Deferred Proofs

Here we give the proofs of Lemmas 3.4.3, 3.4.4, 3.4.5, and 3.4.10, which are all relatively standard. We recall the following lemma, due to Talagrand, from which Lemma 3.4.3 readily follows.

Lemma 3.4.11 ([Tal06a, Lemma 3.1])**.** *For all $\boldsymbol{y} \in \mathbb{R}^N$, the following inequality holds.*

$$\int_{S_N} \exp\langle \boldsymbol{y}, \boldsymbol{\sigma}\rangle \,\mathrm{d}\mu(\boldsymbol{\sigma}) \leq \mathbb{P}\left(\chi^2(N) \geq BN\right)^{-1} \int \exp\langle \boldsymbol{y}, \boldsymbol{\rho}\rangle \,\mathrm{d}\nu_B^N(\boldsymbol{\rho}).$$

Proof of Lemma 3.4.3. Using $\mathcal{Q}(\eta) \subseteq S_N^K$ and Lemma 3.4.11, we get

$$\begin{aligned}
\exp G_D(\vec{\boldsymbol{y}}) &= \exp(-\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{m}}\rangle) \int_{\mathcal{Q}(\eta)} \exp\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{\sigma}}\rangle \,\mathrm{d}\mu_N^K(\vec{\boldsymbol{\sigma}}) \\
&\leq \exp(-\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{m}}\rangle) \int_{S_N^K} \exp\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{\sigma}}\rangle \,\mathrm{d}\mu_N^K(\vec{\boldsymbol{\sigma}}) \\
&= \exp(-\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{m}}\rangle) \prod_{u\in\mathbb{L}} \int_{S_N} \exp\langle \boldsymbol{y}(u), \boldsymbol{\sigma}(u)\rangle \,\mathrm{d}\mu_N(\vec{\boldsymbol{\sigma}}(u)) \\
&\leq \mathbb{P}\left(\chi^2(N) \geq BN\right)^{-K} \exp(-\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{m}}\rangle) \prod_{u\in\mathbb{L}} \int \exp\langle \boldsymbol{y}(u), \boldsymbol{\rho}(u)\rangle \,\mathrm{d}\nu_B^N(\vec{\boldsymbol{\rho}}(u)) \\
&= \mathbb{P}\left(\chi^2(N) \geq BN\right)^{-K} \exp(-\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{m}}\rangle) \int \exp\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{\rho}}\rangle \,\mathrm{d}\nu_B^{N\times K}(\vec{\boldsymbol{\rho}}).
\end{aligned}$$

□

Proof of Lemma 3.4.4. Using the probability density of $\chi^2(N)$, we compute:

$$\begin{aligned}
\mathbb{P}(\chi^2(N) \geq BN) &= \int_{BN}^{\infty} \frac{x^{N/2-1}e^{-x/2}}{2^{N/2}\Gamma\left(\frac{N}{2}\right)} \,\mathrm{d}x \\
&= \frac{(N/2)^{N/2}}{\Gamma\left(\frac{N}{2}\right)} \int_B^{\infty} y^{N/2-1}e^{-Ny/2} \,\mathrm{d}y \\
&\geq \frac{(N/2)^{N/2}}{\Gamma\left(\frac{N}{2}\right)} \int_B^{\infty} e^{-Ny/2} \,\mathrm{d}y \\
&= \frac{(N/2)^{N/2-1}}{\Gamma\left(\frac{N}{2}\right)} e^{-BN/2} \\
&\geq e^{-BN/2},
\end{aligned}$$

where the last step uses that $(N/2)^{N/2-1} \geq \Gamma\left(\frac{N}{2}\right)$ for $N \geq 2$. □

Proof of Lemma 3.4.5. By a straightforward computation,

$$\begin{aligned}
&\mathbb{E}\exp\frac{1}{2}\zeta\left[(\vec{y}+e\vec{t}a)^\top\Lambda^{-1}(\vec{y}+e\vec{t}a)-2\vec{v}^\top(\vec{y}+e\vec{t}a)\right]\\
&=|\Sigma|^{-1/2}(2\pi)^{-K/2}\int\exp\left[-\frac{1}{2}\left(\vec{x}^\top\Sigma^{-1}\vec{x}-\zeta(\vec{y}+\vec{x})^\top\Lambda^{-1}(\vec{y}+\vec{x})+2\zeta\vec{v}^\top(\vec{y}+\vec{x})\right)\right]\,\mathrm{d}\vec{x}\\
&=|\Sigma|^{-1/2}(2\pi)^{-K/2}\int\exp\left[-\frac{1}{2}\left(\vec{x}^\top\left(\Sigma^{-1}-\zeta\Lambda^{-1}\right)\vec{x}-2\zeta(\Lambda^{-1}\vec{y}-\vec{v})^\top\vec{x}-\zeta\vec{y}^\top\Lambda^{-1}\vec{y}+2\zeta\vec{v}^\top\vec{y}\right)\right]\,\mathrm{d}\vec{x}\\
&=|\Sigma|^{-1/2}|\Sigma^{-1}-\zeta\Lambda^{-1}|^{-1/2}\exp\frac{1}{2}\left(\zeta^2(\Lambda^{-1}\vec{y}-\vec{v})^\top\left(\Sigma^{-1}-\zeta\Lambda^{-1}\right)^{-1}(\Lambda^{-1}\vec{y}-\vec{v})+\zeta\vec{y}^\top\Lambda^{-1}\vec{y}-2\zeta\vec{v}^\top\vec{y}\right)\\
&=\frac{|\Lambda|^{1/2}}{|\Lambda-\zeta\Sigma|^{1/2}}\exp\frac{\zeta}{2}\left(\vec{y}^\top(\Lambda-\zeta\Sigma)^{-1}\vec{y}-2\vec{v}^\top\Lambda(\Lambda-\zeta\Sigma)^{-1}\vec{y}+\vec{v}^\top(\zeta\Sigma)(\Lambda-\zeta\Sigma)^{-1}\Lambda\vec{v}\right).
\end{aligned}$$

Taking logarithms and dividing by ζ yields the result. □

Proof of Lemma 3.4.10. Define the random variable

$$\vec{\boldsymbol{\sigma}}^* = \arg\max_{\vec{\boldsymbol{\sigma}}\in\mathcal{Q}(\eta)}\mathcal{H}_N(\vec{\boldsymbol{\sigma}}),$$

where we break ties arbitrarily. For $\delta>0$, define

$$\mathcal{B}(\vec{\boldsymbol{\sigma}}^*,\delta)=\left\{\vec{\boldsymbol{\sigma}}\in S_N^K:\|\boldsymbol{\sigma}(u)-\boldsymbol{\sigma}^*(u)\|_N\le\delta\text{ for all }u\in\mathbb{L}\right\}.$$

If $\vec{\boldsymbol{\sigma}}\in\mathcal{B}(\vec{\boldsymbol{\sigma}}^*,\eta/3)$, then for each $u\in\mathbb{L}$ we can write $\boldsymbol{\sigma}(u)=\boldsymbol{\sigma}^*(u)+\delta(u)\boldsymbol{\rho}(u)$, where $\boldsymbol{\rho}(u)\in S_N$ and $0\le\delta(u)\le\eta/3$. Then, for all $u\in\mathbb{L}$,

$$|R(\boldsymbol{\sigma}(u),\boldsymbol{m})-q_0|\le|R(\boldsymbol{\sigma}^*(u),\boldsymbol{m})-q_0|+\delta(u)|R(\boldsymbol{\rho}(u),\boldsymbol{m})|\le\eta+\eta/3\le 2\eta,$$

and for all $u,v\in\mathbb{L}$,

$$\begin{aligned}
&|R(\boldsymbol{\sigma}(u),\boldsymbol{\sigma}(v))-q_{u\wedge v}|\\
&\le|R(\boldsymbol{\sigma}^*(u),\boldsymbol{\sigma}^*(v))-q_{u\wedge v}|+\delta(u)|R(\boldsymbol{\sigma}^*(u),\boldsymbol{\rho}(v))|+\delta(v)|R(\boldsymbol{\sigma}^*(v),\boldsymbol{\rho}(u))|+\delta(u)\delta(v)|R(\boldsymbol{\rho}(u),\boldsymbol{\rho}(u))|\\
&\le\eta+\eta/3+\eta/3+\eta/3=2\eta.
\end{aligned}$$

So, $\mathcal{B}(\vec{\boldsymbol{\sigma}}^*,\eta/3)\subseteq\mathcal{Q}(2\eta)$.

Let constants c,C_1 be given by Proposition 3.0.1. By this proposition, the event

$$S=\left\{\sup_{u\in\mathbb{L}}\sup_{\boldsymbol{\sigma}\in S_N}\left\|\nabla H_N^{(u)}(\boldsymbol{\sigma})\right\|_N\le C_1\right\}$$

has probability $\mathbb{P}(S)\ge 1-Ke^{-cN}$. Here we use the fact that for $\boldsymbol{v}\in\mathbb{R}^N$, $\|\boldsymbol{v}\|_N=\|\boldsymbol{v}\|_{\mathrm{op}}$. On S,

$$\mathcal{H}_N(\vec{\boldsymbol{\sigma}})\ge\mathcal{H}_N(\vec{\boldsymbol{\sigma}}^*)-\frac{C_1NK\eta}{3}$$

for all $\vec{\boldsymbol{\sigma}}\in\mathcal{B}(\vec{\boldsymbol{\sigma}}^*,\eta/3)$. So,

$$\begin{aligned}
F_N^{\mathrm{Sp}}(\beta,\mathcal{Q}(2\eta))&=\frac{1}{N}\log\mathbb{E}\int_{\mathcal{Q}(2\eta)}\exp\beta\mathcal{H}_N(\vec{\boldsymbol{\sigma}})\,\mathrm{d}\mu^K(\vec{\boldsymbol{\sigma}})\\
&\ge\frac{1}{N}\log\mathbb{E}\,\mathbb{1}\{S\}\int_{\mathcal{B}(\vec{\boldsymbol{\sigma}}^*,\eta/3)}\exp\beta\mathcal{H}_N(\vec{\boldsymbol{\sigma}})\,\mathrm{d}\mu^K(\vec{\boldsymbol{\sigma}})\\
&\ge\frac{1}{N}\log\mathbb{E}\,\mathbb{1}\{S\}\int_{\mathcal{B}(\vec{\boldsymbol{\sigma}}^*,\eta/3)}\exp\beta\left(\mathcal{H}_N(\vec{\boldsymbol{\sigma}}^*)-\frac{C_1NK\eta}{3}\right)\,\mathrm{d}\mu^K(\vec{\boldsymbol{\sigma}})\\
&\ge\frac{1}{N}\log\mathbb{E}\,\mathbb{1}\{S\}\exp\beta\mathcal{H}_N(\vec{\boldsymbol{\sigma}}^*)-\frac{\beta C_1K\eta}{3}+\frac{1}{N}\log\mu^K(\mathcal{B}(\vec{\boldsymbol{\sigma}}^*,\eta/3))\\
&=\beta\mathrm{GS}_N^{\mathrm{Sp}}(\mathcal{Q}(\eta))-\frac{\beta C_1K\eta}{3}+\frac{1}{N}\log\mu^K(\mathcal{B}(\vec{\boldsymbol{\sigma}}^*,\eta/3))\\
&\quad+\frac{1}{N}\log\mathbb{E}\,\mathbb{1}\{S\}\exp\beta\left(\mathcal{H}_N(\vec{\boldsymbol{\sigma}}^*)-\mathbb{E}\,\mathcal{H}_N(\vec{\boldsymbol{\sigma}}^*)\right).
\end{aligned}$$

The set $\mathcal{B}(\vec{\boldsymbol{\sigma}}^*, \eta/3)$ is the product of K spherical caps in S_N. By elementary properties of the spherical measure, there exists a large C such that $\mu^K(\mathcal{B}(\vec{\boldsymbol{\sigma}}^*, \eta/3)) \le \eta^{CNK}$, and so

$$\frac{1}{N}\log \mu^K(\mathcal{B}(\vec{\boldsymbol{\sigma}}^*, \eta/3)) \ge -CK\log\frac{1}{\eta}.$$

By Proposition 3.2.8,

$$\mathbb{P}\left(\mathcal{H}_N(\vec{\boldsymbol{\sigma}}^*) - \mathbb{E}\,\mathcal{H}_N(\vec{\boldsymbol{\sigma}}^*) \le -K\sqrt{4\log 2\cdot\xi(1)N}\right) \le \frac{1}{2}.$$

By a union bound, the complement of this event and S simultaneously hold with probability at least $\frac{1}{2} - Ke^{-cN}$. Thus,

$$\frac{1}{N}\log \mathbb{E}\,\mathbb{1}\{S\}\exp\beta\left(\mathcal{H}_N(\vec{\boldsymbol{\sigma}}^*) - \mathbb{E}\,\mathcal{H}_N(\vec{\boldsymbol{\sigma}}^*)\right) \ge -\beta K\sqrt{\frac{4\log 2\cdot\xi(1)}{N}} + \frac{1}{N}\log\left(\frac{1}{2} - Ke^{-cN}\right).$$

Putting this all together, we can choose a large C dependent only on ξ, h such that

$$F_N^{\mathrm{Sp}}(\beta, \mathcal{Q}(2\eta)) \ge \beta \mathrm{GS}_N^{\mathrm{Sp}}(\mathcal{Q}(\eta)) - CK\beta\eta - CK\log\frac{1}{\eta} - \frac{CK\beta}{\sqrt{N}} - \frac{1}{N}\log\frac{1}{\frac{1}{2} - Ke^{-cN}}.$$

By choosing C large enough, we can ensure that if $N \ge C\log\max(K,2)$, then $Ke^{-cN} \le \frac{1}{4}$. Then, we may absorb the last term into the term $CK\log\frac{1}{\eta}$. Rearranging yields the result. □

3.5 Overlap-Constrained Upper Bound on the Ising Grand Hamiltonian

In this section we upper-bound $\varphi^{\mathrm{Is}}(0)$. We take the reference measure μ to be the counting measure so that integrals over $\mathcal{Q}^{\mathrm{Is}}(\eta)$ become sums.

We define $(Z_0, \dots, Z_D)$ similarly to G_d of the previous section, but as a sum over all of $(\Sigma_N)^K$ directly. As before, define $\vec{\boldsymbol{\eta}}_0, \dots \vec{\boldsymbol{\eta}}_D$ to be independent Gaussians as in (3.4.1) and (3.4.2). For $\vec{\boldsymbol{y}} \in (\mathbb{R}^K)^N$, define

$$\begin{aligned}
Z_D(\vec{\boldsymbol{y}}) &= \log \sum_{\vec{\boldsymbol{\sigma}}\in(\Sigma_N)^K} \exp\sum_{u\in\mathbb{L}}\langle \boldsymbol{h} + \lambda\boldsymbol{m} + \boldsymbol{y}(u), \pi(\boldsymbol{\sigma}(u))\rangle \\
&= \log\prod_{i=1}^N\prod_{u\in\mathbb{L}}\left(2\cosh\left(h + \lambda m_i + \boldsymbol{y}(u)_i\right)\exp\left(-m_i(h+\lambda m_i + \boldsymbol{y}(u)_i)\right)\right) \\
&= \sum_{i=1}^N\sum_{u\in\mathbb{L}}\Big(\log\left(2\cosh\left(h+\lambda m_i + \boldsymbol{y}(u)_i\right)\right) - m_i(h+\lambda m_i + \boldsymbol{y}(u)_i)\Big).
\end{aligned}$$

Given the sequence $0 = \zeta_{-1} < \zeta_0 < \zeta_1 < \cdots < \zeta_L = 1$, recursively set

$$Z_d(\vec{\boldsymbol{y}}) = \frac{1}{\zeta_d}\mathbb{E}\,\zeta_d Z_{d+1}\left(\vec{\boldsymbol{y}} + \vec{\boldsymbol{\eta}}_{d+1}(\xi'(q_{d+1}) - \xi'(q_d))^{1/2}\right).$$

Then $Z_0 \equiv Z_0(0)$ is a deterministic function of $\boldsymbol{m}$ and h.

Proposition 3.5.1. *For any $\boldsymbol{m} \in [-1,1]^N$,*

$$\varphi^{\mathrm{Is}}(0) \le \frac{1}{N}Z_0 + KR(\boldsymbol{h}, \boldsymbol{m}).$$

Proof. Recall from (3.3.8) that

$$\varphi^{\mathrm{Is}}(0) = KR(\boldsymbol{h},\boldsymbol{m}) + \frac{1}{N}\log\mathbb{E}\sum_{\alpha\in\mathbb{N}^D}\nu_\alpha\sum_{\vec{\boldsymbol{\sigma}}\in\mathcal{Q}^{\mathrm{Is}}(\eta)}\exp\left(\sum_{u\in\mathbb{L}}\langle\boldsymbol{h}+\lambda\boldsymbol{m}, \pi(\boldsymbol{\sigma}(u))\rangle + \sum_{u\in\mathbb{L}}\sum_{i=1}^N g^{(u)}_{\xi',i}(\alpha)\pi(\boldsymbol{\sigma}(u))_i\right).$$

Summing over all of $(\Sigma_N)^K$ gives the upper bound

$$\varphi^{\mathrm{Is}}(0) \le KR(\boldsymbol{h},\boldsymbol{m}) + \frac{1}{N}\log \mathbb{E} \sum_{\alpha\in\mathbb{N}^D} \nu_\alpha \sum_{\vec{\boldsymbol{\sigma}}\in(\Sigma_N)^K} \exp\left(\sum_{u\in\mathbb{L}} \langle \boldsymbol{h}+\lambda \boldsymbol{m}, \pi(\boldsymbol{\sigma}(u))\rangle + \sum_{u\in\mathbb{L}}\sum_{i=1}^{N} g^{(u)}_{\xi',i}(\alpha)\pi(\boldsymbol{\sigma}(u))_i \right).$$

Similarly to previous sections or as in [Pan13b, Theorem 2.9], properties of Ruelle cascades imply that the right hand side above equals

$$KR(\boldsymbol{h},\boldsymbol{m}) + \frac{1}{N} Z_0$$

because the coordinates $i \in [N]$ now decouple. □

3.5.1 Properties of Parisi PDEs

Here we review properties of Parisi PDEs. We begin with the 1-dimensional case for general $\zeta \in \mathscr{L}$ and consider the PDE

$$\partial_t \Phi_\zeta(t,x) + \frac{1}{2}\xi''(t)\left(\partial_{xx}\Phi_\zeta(t,x) + \zeta(t)(\partial_x \Phi_\zeta(t,x))^2\right) = 0 \tag{3.5.1}$$
$$\Phi_\zeta(1,x) = f_0(x).$$

For $\beta > 0$ we will consider the initial conditions $f_0(x) = \log(\cosh(\beta x)/\beta) - ax$ for $a = m_i \in [-1,1]$ which leads to solution $\Phi^\beta_{a,\zeta}$. When not specified, we take $\beta = 1$ and $a = 0$, so for instance $\Phi_{a,\zeta} = \Phi^1_{a,\zeta}$ and $\Phi^\beta_\zeta = \Phi^\beta_{0,\zeta}$. We also allow the $\beta = \infty$ case $\Phi^\infty_{a,\zeta}$ corresponding to $f_0(x) = |x| - ax$. Note that (1.3.4) corresponds to the case $(a,\beta) = (0,\infty)$. Regularity properties for solutions to (3.5.1) were derived in several works such as [JT16, Che17] for $\zeta \in \mathscr{U}$. We draw on the results[2] of [AMS21] for $\zeta \in \mathscr{L}$.

Proposition 3.5.2. *[AMS21, Proposition 6.1(b) and Lemma 6.4] For $\zeta \in \mathscr{L}$ and $(a,\beta) \in [-1,1]\times(0,\infty]$, the function $\Phi^\beta_{a,\zeta}$ is continuous on $[0,1]\times\mathbb{R}$ and 2-Lipschitz in x. Moreover both*

$$\partial_{xx}\Phi^\beta_{a,\zeta}(t,x) \quad \text{and} \quad \partial_t \Phi^\beta_{a,\zeta}(t,x)$$

are uniformly bounded on $(t,x) \in [0,1-\varepsilon]\times\mathbb{R}$ for any $\varepsilon > 0$. Finally $\Phi^\beta_{a,\zeta}(t,x)$ is convex in x.

Proposition 3.5.3. *[AMS21, Lemma 6.5] For $\zeta \in \mathscr{L}$ the SDE*

$$\mathrm{d}X_t = \xi''(t)\zeta(t)\partial_x \Phi^\beta_{a,\zeta}(t,X_t)\,\mathrm{d}t + \sqrt{\xi''(t)}\,\mathrm{d}B_t, \quad X_0 = X_0 \tag{3.5.2}$$

has strong and pathwise unique solution.

Proposition 3.5.4. *[AMS21, Proposition 6.1(c)] For $\zeta_1,\zeta_2 \in \mathscr{L}$, and $\beta \in (0,\infty]$,*

$$|\Phi^\beta_{\zeta_1} - \Phi^\beta_{\zeta_2}| \le \int_0^1 \xi''(t)|\zeta_1(t) - \zeta_2(t)|dt.$$

The Multi-Dimensional Parisi PDE

Here we define the Parisi PDE on $\mathbb{R}^K$. For simplicity we restrict attention to finitely supported $\zeta \in \mathcal{M}_{\vec{q}}$. We construct $\Phi^{\mathbb{L}}$ via the Hopf-Cole transformation and verify that it solves a version of (3.5.1).

Recall the definition of $M^d = M^{\vec{k},\vec{p},d} \in \mathbb{R}^{K\times K}$ given by

$$M^{\vec{k},\vec{p},d}_{u^1,u^2} = \mathbb{1}\left\{u^1 \wedge u^2 \ge d\right\} p_{u^1\wedge u^2}.$$

As before, $M(t) = M^d$ for $t \in [q_{d-1}, q_d)$.

[2] Technically the cited results from [AMS21] assume f_0 is 1-Lipschitz and even. The evenness is not used in the proofs of the statement below. In our case f_0 is 2-Lipschitz when $|a| \le 1$, which is equivalent up to a rescaling as in (3.5.7).

For an atomic measure $\zeta \in \mathcal{M}_{\vec{q}}$ consider the function $\Phi^{\mathbb{L}}_{\zeta}(t,\vec{x}) : [0,1] \times \mathbb{R}^K \to \mathbb{R}$ defined as as follows. The $t = 1$ boundary condition is

$$\Phi^{\mathbb{L}}_{a,\zeta}(1,\vec{x}) = \sum_{u \in \mathbb{L}} \log\left(2\cosh x(u)\right) - ax(u).$$

For $t \in [q_0, 1)$, $\Phi^{\mathbb{L}}_{a,\zeta}$ is defined recursively by

$$\Phi^{\mathbb{L}}_{a,\zeta}(t,x) = \frac{1}{\zeta(t)} \log \mathbb{E} \exp\left(\zeta(t)\Phi^{\mathbb{L}}_{a,\zeta}(q_{d+1}, \vec{x} + e\vec{t}a_{d+1} \cdot (\xi'(q_{d+1}) - \xi'(t))^{1/2})\right), \quad t \in [q_d, q_{d+1})$$

where $e\vec{t}a_0 \sim \mathcal{N}(0, M^1)$ and $e\vec{t}a_d \sim \mathcal{N}(0, M^d)$ for $1 \le d \le D$ are independent Gaussian vectors in $\mathbb{R}^K$. For $t \in [0, q_0)$, we extend the definition of ζ so that $\zeta(t) = 0$ and define

$$\Phi^{\mathbb{L}}_{a,\zeta}(t,x) = \mathbb{E}\, \Phi^{\mathbb{L}}_{a,\zeta}(q_0, \vec{x} + e\vec{t}a_0 \cdot (\xi'(q_0) - \xi'(t))^{1/2}).$$

Proposition 3.5.5. *For any $\zeta \in \mathcal{M}_{\vec{q}}$,*

$$Z_0 = \frac{1}{N} \sum_{i=1}^{N} \Phi^{\mathbb{L}}_{m_i,\zeta}(0, (h + \lambda m_i)\vec{1}).$$

Proof. This follows from Lemma 3.5.6 since the recursive definition of $\Phi^{\mathbb{L}}_{a,\zeta}(t,x)$ restricted to times $t \in \{q_d\}_{d \in [D]}$ is exactly that of Z_0 up to an spatial shift of $(h + \lambda m_i)\vec{1}$. □

We defer the proof of the next lemma, which is a standard computation.

Lemma 3.5.6. *The function $\Phi^{\mathbb{L}}_{a,\zeta}$ is smooth on each time interval $[q_d, q_{d+1}] \times \mathbb{R}^K$. Moreover it is continuous and solves the K-dimensional Parisi PDE*

$$\partial_t \Phi^{\mathbb{L}}_{a,\zeta}(t,\vec{x}) = -\frac{\xi''(t)}{2}\left(\langle M(t), \nabla^2 \Phi^{\mathbb{L}}_{a,\zeta}\rangle + \zeta(t)\langle M(t), (\nabla \Phi^{\mathbb{L}}_{a,\zeta})^{\otimes 2}\rangle\right). \tag{3.5.3}$$

Finally $|\partial_{x(u)} \Phi^{\mathbb{L}}_{a,\zeta}(t,\vec{x})| \le 1 + |a|$ holds for all $(t,\vec{x},u) \in [0,1] \times \mathbb{R}^K \times \mathbb{L}$.

Auffinger-Chen Representation

As shown by [AC15] the Parisi PDE admits a stochastic control formulation. We now recall such representations in the cases of interest starting with the 1-dimensional case. For $0 \le t_1 \le t_2 \le 1$ let $\mathcal{D}[t_1, t_2]$ be the space of processes $v \in C([t_1, t_2]; \mathbb{R})$ with $\sup_{t_1 \le r \le t_2} |v_r| \le 2$ which are progressively measurable with respect to filtration supporting a standard Brownian motion B_t. Define the functional

$$\mathcal{X}^{t_1,t_2}_{a,\zeta}(x,v) = \mathbb{E}\left[\mathcal{Y}^{t_1,t_2}_{a,\zeta}(x,v) - \mathcal{Z}^{t_1,t_2}_{a,\zeta}(v)\right]$$

where

$$\mathcal{Y}^{t_1,t_2}_{a,\zeta}(x,v) \equiv \Phi^{\beta}_{a,\zeta}\left(t_2, x + \int_{t_1}^{t_2} \zeta(r)\xi''(r)v_r \, \mathrm{d}r + \int_{t_1}^{t_2} \sqrt{\xi''(r)} \, \mathrm{d}B_r\right),$$

$$\mathcal{Z}^{t_1,t_2}_{a,\zeta}(v) \equiv \frac{1}{2}\int_{t_1}^{t_2} \zeta(r)\xi''(r)v_r^2 \, \mathrm{d}r.$$

Note that since $|v_r| \le 2$ is uniformly bounded and $||\xi'' \cdot \zeta||_1 < \infty$ there are no continuity issues near $t = 1$. The next proposition, whose standard proof we defer, relates $\Phi^{\beta}_{a,\zeta}$ to stochastic control.

Proposition 3.5.7. *For any $\zeta \in \mathscr{L}$, $[t_1, t_2] \subseteq [0,1]$, $a \in [-1,1]$ and $\beta \in (0,\infty]$, the function $\Phi^{\beta}_{a,\zeta}$ satisfies*

$$\Phi^{\beta}_{a,\zeta}(t_1,x) = \sup_{v \in \mathcal{D}[t_1,t_2]} \mathcal{X}^{t_1,t_2}_{a,\zeta}(x,v). \tag{3.5.4}$$

Moreover the maximum in (3.5.4) is achieved by

$$v_r = \partial_x \Phi^{\beta}_{a,\zeta}(r, X_r)$$

where X_r solves the SDE (3.5.2) with initial condition $X_{t_1} = x$.

The corresponding stochastic control formulation in $\mathbb{R}^K$ is as follows. For $0 \le t_1 \le t_2 \le 1$ let $\mathcal{D}^{\mathbb{L}}[t_1, t_2]$ be the space of processes $\vec{v} \in C([t_1, t_2]; \mathbb{R}^K)$ with $\sup_{t_1 \le r \le t_2} |\vec{v}_r|_\infty \le 2$ which are progressively measurable with respect to a filtration supporting an $\mathbb{R}^K$ valued Brownian motion $\vec{B}_r = (B_r^u)_{u \in \mathbb{L}}$. Define the functional

$$\mathcal{X}_{a,\zeta}^{\mathbb{L},t_1,t_2}(\vec{x}, \vec{v}) \equiv \mathbb{E}\left[\mathcal{Y}_{a,\zeta}^{\mathbb{L},t_1,t_2}(\vec{x}, \vec{v}) - \mathcal{Z}_{a,\zeta}^{\mathbb{L},t_1,t_2}(\vec{v})\right]$$

where

$$\mathcal{Y}_{a,\zeta}^{\mathbb{L},t_1,t_2}(\vec{x}, \vec{v}) \equiv \Phi_{a,\zeta}^{\mathbb{L}}\left(t_2, \vec{x} + \int_{t_1}^{t_2} \zeta(r)\xi''(r)M(r)\vec{v}_r \, \mathrm{d}r + \int_{t_1}^{t_2} \sqrt{\xi''(r)M(r)} \, \mathrm{d}\vec{B}_r\right),$$
$$\mathcal{Z}_{a,\zeta}^{\mathbb{L},t_1,t_2}(\vec{v}) \equiv \frac{1}{2}\int_{t_1}^{t_2} \zeta(r)\xi''(r)\langle M(r), \vec{v}_r^{\otimes 2}\rangle \, \mathrm{d}r.$$

In the multi-dimensional case we restrict attention to finitely supported $\zeta \in \mathcal{M}_{\vec{q}}$ to avoid the by-now routine process of extending regularity properties of $\Phi_\zeta^{\mathbb{L}}$ to general ζ. The proof is again deferred.

Proposition 3.5.8. *For any $\zeta \in \mathcal{M}_{\vec{q}}$, $[t_1, t_2] \subseteq [0, 1]$ and $a \in [-1, 1]$, the function $\Phi_{a,\zeta}^{\mathbb{L}}$ satisfies*

$$\Phi_{a,\zeta}^{\mathbb{L}}(t_1, \vec{x}) = \sup_{\vec{v} \in \mathcal{D}^{\mathbb{L}}[t_1,t_2]} \mathcal{X}_{a,\zeta}^{\mathbb{L},t_1,t_2}(\vec{x}, \vec{v}). \tag{3.5.5}$$

Moreover (3.5.5) is maximized by $\vec{v}_s = \nabla\Phi_{a,\zeta}^{\mathbb{L}}(s, \vec{X}_s)$ where the $\mathbb{R}^K$-valued process $\vec{X}_s$ solves

$$\vec{X}_s = \vec{x} + \int_{t_1}^{s} \zeta(r)\xi''(r)M(r)\nabla\Phi_{a,\zeta}^{\mathbb{L}}(r, \vec{X}_r) \, \mathrm{d}r + \int_{t_1}^{s} \sqrt{\xi''(r)M(r)} \, \mathrm{d}\vec{B}_r, \qquad s \in [t_1, t_2].$$

3.5.2 Relations Among Parisi PDEs

Following [CPS18, Section 8] we relate $\Phi_{a,\zeta}$ to Φ_ζ. Note that we always consider times $t \in [0, 1]$ with endpoint conditions at $t = 1$, while [CPS18] defines the boundary condition for $\Phi_{a,\zeta}$ at time $t = 1 - q_0$, see e.g. Equation (3.25) therein.

Proposition 3.5.9. *For any $a \in [-1, 1]$ and $\zeta \in \mathscr{L}$, with $y = x - a\int_0^1 \xi''(t)\zeta(t)dt$,*

$$\Phi_\zeta(0, y) - ay = \Phi_{a,\zeta}(0, x) + \frac{a^2}{2}\int_0^1 \xi''(t)\zeta(t)dt.$$

Proof. By setting $y = x - a\int_0^1 \xi''(s)\zeta(s) \, \mathrm{d}s$, it suffices to show that for all $t \in [0, 1]$,

$$\Phi_{a,\zeta}(t, x) = \Phi_\zeta\left(t, x - a\int_t^1 \xi''(s)\zeta(s) \, \mathrm{d}s\right) - ax + \frac{a^2}{2}\int_t^1 \xi''(s)\zeta(s) \, \mathrm{d}s.$$

(In particular the desired result is obtained by setting $t = 0$.) It suffices to show this for ζ continuous. Set

$$f(t, x) \equiv \Phi_\zeta\left(t, x - a\int_t^1 \xi''(s)\zeta(s) \, \mathrm{d}s\right) - ax + \frac{a^2}{2}\int_t^1 \xi''(s)\zeta(s) \, \mathrm{d}s$$

and define

$$b(t, x) \equiv x - a\int_t^1 \xi''(s)\zeta(s) \, \mathrm{d}s.$$

Then we compute

$$\partial_t f(t, x) = \partial_t\Phi_\zeta(t, b(t, x)) + a\xi''(t)\zeta(t)\partial_x\Phi_\zeta(t, b(t, x)) - \frac{a^2}{2}\xi''(t)\zeta(t)$$

and

$$\partial_x f(t,x) = \partial_x \Phi_\zeta(t, b(t,x)) - a,$$
$$\partial_{xx} f(t,x) = \partial_{xx} \Phi_\zeta(t, b(t,x)).$$

It follows that

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

Note that at time 1, $f(1,x) = \log(2\cosh(x)) - ax = \Phi_{a,\zeta}(1,x)$. Uniqueness of solutions to the Parisi PDE as in [JT16, Lemma 13] completes the proof.

□

Lemma 3.5.10. *For any $\zeta, \gamma \in \mathscr{L}$ and any $(t,x,\beta) \in [0,1] \times \mathbb{R} \times (0,\infty]$,*

$$\Phi^\beta_{a,\zeta}(t,x) \le \Phi^\beta_{a,\zeta+\gamma}(t,x).$$

Proof. We use the Auffinger-Chen representation (3.5.4) for $\Phi^\beta_{a,\zeta}$ and $\Phi^\beta_{a,\zeta+\gamma}$. For any control v, consider the modified control

$$w_t \equiv \frac{\zeta(t) v_t}{\zeta(t) + \gamma(t)}.$$

It is not difficult to see that

$$\mathcal{Y}^{t,1}_{a,\zeta}(x,v) = \mathcal{Y}^{t,1}_{a,\zeta+\gamma}(x,w)$$

since the resulting SDE is the same, while

$$\mathcal{Z}^{t,1}_{a,\zeta}(v) \ge \mathcal{Z}^{t,1}_{a,\zeta+\gamma}(w).$$

Therefore

$$\mathcal{X}^{t,1}_{a,\zeta}(x,v) \le \mathcal{X}^{t,1}_{a,\zeta+\gamma}(x,w)$$

Since v was arbitrary, we are done by Proposition 3.5.7. □

Define $\overline{\zeta} = \zeta|_{[q_0,1]}$ and $\underline{\zeta} = \zeta|_{[0,q_0]}$ when $\zeta \in \mathscr{L}$ and $q_0 \in [0,1]$ are given. The next lemma is analogous to Lemma 3.4.8 and will be used to connect our estimates for $\varphi(0)$ to the Parisi functional uniformly in $\boldsymbol{m}$.

Lemma 3.5.11. *For $\zeta \in \mathscr{L}$, with $\lambda = \int_0^1 \xi''(t)\zeta(t)dt$,*

$$\frac{1}{N}\sum_{i=1}^N \Phi^\infty_{m_i,\overline{\zeta}}(0, h + \lambda m_i) - \frac{1}{2}\int_{q_0}^1 (t - q_0)\overline{\zeta}(t)\xi''(t)dt + R(\boldsymbol{h},\boldsymbol{m}) \le \mathsf{P}^{\mathrm{Is}}_{\xi,h}(\zeta).$$

Proof. Define the constants

$$I = \int_0^1 t\xi''(t)\zeta(t)dt, \quad J = \lambda = \int_0^1 \xi''(t)\zeta(t)dt,$$
$$\overline{I} = \int_{q_0}^1 t\xi''(t)\overline{\zeta}(t)dt, \quad \overline{J} = \int_{q_0}^1 \xi''(t)\overline{\zeta}(t)dt,$$
$$\underline{I} = \int_0^{q_0} t\xi''(t)\underline{\zeta}(t)dt, \quad \underline{J} = \int_0^{q_0} \xi''(t)\underline{\zeta}(t)dt.$$

Then $I = \overline{I} + \underline{I}$ and $J = \overline{J} + \underline{J}$ and $q_0\underline{J} \ge \underline{I}$. Recalling that $\mathsf{P}^{\mathrm{Is}}_{\xi,h}(\zeta) = \Phi^\infty_\zeta(0,h) - \frac{I}{2}$, we estimate

$$\begin{aligned}
\mathsf{P}^{\mathrm{Is}}_{\xi,h}(\zeta) &= \Phi^\infty_\zeta(0,h) - \tfrac{I}{2} && \text{Prop 3.5.2}\\
&= \tfrac{1}{N}\textstyle\sum_{i=1}^N \left(\Phi^\infty_{m_i,\zeta}(0, h+\lambda m_i) + \frac{m_i^2 J}{2}\right) - \frac{I}{2} + R(\boldsymbol{h},\boldsymbol{m}) && \|\boldsymbol{m}\|_N^2 = q_0\\
&= \tfrac{1}{N}\textstyle\sum_{i=1}^N \Phi^\infty_{m_i,\zeta}(0, h+\lambda m_i) + \frac{q_0 J}{2} - \frac{I}{2} + R(\boldsymbol{h},\boldsymbol{m}) && \zeta = \overline{\zeta} + \underline{\zeta}\\
&= \tfrac{1}{N}\textstyle\sum_{i=1}^N \Phi^\infty_{m_i,\zeta}(0, h+\lambda m_i) + \frac{q_0\overline{J}}{2} - \frac{\overline{I}}{2} + \frac{q_0\underline{J} - \underline{I}}{2} + R(\boldsymbol{h},\boldsymbol{m}) && q_0\underline{J} \ge \underline{I}\\
&\ge \tfrac{1}{N}\textstyle\sum_{i=1}^N \Phi^\infty_{m_i,\zeta}(0, h+\lambda m_i) + \frac{q_0\overline{J}}{2} - \frac{\overline{I}}{2} + R(\boldsymbol{h},\boldsymbol{m}) && \text{Lem 3.5.10}\\
&\ge \tfrac{1}{N}\textstyle\sum_{i=1}^N \Phi^\infty_{m_i,\overline{\zeta}}(0, h+\lambda m_i) + \frac{q_0\overline{J}}{2} - \frac{\overline{I}}{2} + R(\boldsymbol{h},\boldsymbol{m}).
\end{aligned}$$

This is exactly what we wanted to show.

□

The next crucial lemma upper-bounds $\Phi^{\mathbb{L}}_{a,\zeta}$ using the 1-dimensional function $\Phi_{a,\kappa\zeta}$. As in the spherical case, multiplying by κ will allow us to pass from increasing $\zeta \in \mathcal{M}_{\vec{q}}$ to arbitrary functions in $\mathscr{L}$.

Lemma 3.5.12. *For any $\zeta \in \mathcal{M}_{\vec{q}}$, $\vec{x} \in \mathbb{R}^K$, $a \in [-1,1]$ and $t \in [0,1]$,*

$$\Phi^{\mathbb{L}}_{a,\zeta}(t,\vec{x}) \leq \sum_{u\in\mathbb{L}} \Phi_{a,\kappa\zeta}(t,x(u)). \tag{3.5.6}$$

Proof. Define

$$\widetilde{\mathcal{Z}}^{\mathbb{L},t_1,t_2}_{a,\zeta}(\vec{v}) \equiv \frac{1}{2}\int_{t_1}^{t_2} \zeta(r)\xi''(r)\kappa(r)^{-1}\langle M(r)^2, \vec{v}_r^{\otimes 2}\rangle \,\mathrm{d}r.$$

Since $M(r) \preceq \kappa(r) I_K$, in the Loewner order, it follows that

$$\kappa(r)^{-1} M(r)^2 \preceq M(r).$$

Hence $\widetilde{\mathcal{Z}}^{\mathbb{L},t_1,t_2}_{a,\zeta}(\vec{v}) \leq \mathcal{Z}^{\mathbb{L},t_1,t_2}_{a,\zeta}(\vec{v})$ for any $\vec{x}$ and $\vec{v}$. Setting

$$\widetilde{\mathcal{X}}^{\mathbb{L},t_1,t_2}_{a,\zeta}(\vec{x},\vec{v}) \equiv \mathcal{Y}^{\mathbb{L},t_1,t_2}_{a,\zeta}(\vec{x},\vec{v}) - \widetilde{\mathcal{Z}}^{\mathbb{L},t_1,t_2}_{a,\zeta}(\vec{v}),$$

it follows that

$$\mathcal{X}^{t,1}_{a,\zeta}(\vec{x},\vec{v}) \leq \widetilde{\mathcal{X}}^{t,1}_{a,\zeta}(\vec{x},\vec{v})$$

always holds. Next for any $\vec{v} \in \mathcal{D}^{\mathbb{L}}[t,1]$ and $r \in [t,1]$, define $\vec{V}_r = \frac{M(r)\vec{v}_r}{\kappa(r)} \in \mathbb{R}^K$. Then

$$\langle M(r)^2, \vec{v}_r^{\otimes 2}\rangle = \|M(r)\vec{v}_r\|_2^2 = \kappa(r)^2 \left\|\vec{V}_r\right\|^2$$

and so (including the relevant Brownian motions as arguments in a slight abuse of notation),

$$\widetilde{\mathcal{Z}}^{\mathbb{L},t,1}_{a,\zeta}(\vec{v},\vec{B}) = \sum_{u\in\mathbb{L}} \mathcal{Z}^{t,1}_{a,\kappa\zeta}(V(u),B(u)).$$

Moreover since $\kappa(r)\vec{V}_r(u) = M(r)\vec{v}_r(u)$,

$$\widetilde{\mathcal{Y}}^{\mathbb{L},t,1}_{a,\zeta}(\vec{x},\vec{v},\vec{B}) = \sum_{u\in\mathbb{L}} \mathcal{Y}^{t,1}_{a,\kappa\zeta}(x(u),V(u),B(u)).$$

Since each coordinate $B_r(u)$ of $\vec{B}_r$ has the marginal law of a 1-dimensional Brownian motion,

$$\widetilde{\mathcal{X}}^{\mathbb{L},t,1}_{a,\zeta}(\vec{x},\vec{v}) = \sum_{u\in\mathbb{L}} \mathcal{X}^{t,1}_{a,\kappa\zeta}(x(u),V(u)).$$

Therefore we obtain

$$\begin{aligned}\mathcal{X}^{t,1}_{a,\zeta}(\vec{x},\vec{v}) &\leq \widetilde{\mathcal{X}}^{t,1}_{a,\zeta}(\vec{x},\vec{v}) \\ &= \sum_{u\in\mathbb{L}} \mathcal{X}^{t,1}_{a,\kappa\zeta}(x(u),V(u)) \\ &\leq \sum_{u\in\mathbb{L}} \Phi_{a,\kappa\zeta}(t,x(u)).\end{aligned}$$

Since $\vec{v} \in \mathcal{D}^{\mathbb{L},t,1}$ was arbitrary this concludes the proof. □

3.5.3 Zero Temperature Limit

We now apply the above results with $(\beta^2\xi, \beta\boldsymbol{h}, \beta\lambda)$ in place of $(\xi, \boldsymbol{h}, \lambda)$, which corresponds to scaling $\mathcal{H}_N$ to $\beta\mathcal{H}_N$. We accordingly define $\Phi_{\beta^2\xi,\zeta}$ and $\varphi^{\mathrm{Is}}_{\beta^2\xi}(0)$ by making this substitution in their definitions. It is not hard to derive the scaling relation

$$\Phi_{\beta^2\xi,\zeta}(t,\beta x) = \beta \cdot \Phi^{\beta}_{\beta\zeta}(t,x), \quad (t,x) \in [0,1]\times\mathbb{R} \tag{3.5.7}$$

for any $\beta \in (0,\infty)$ and $\zeta \in \mathscr{L}$.

We will also use the following simple estimate to pass to the zero temperature limit.

Proposition 3.5.13. $\sup_\zeta \left|\Phi^{\beta}_{\zeta}(t,x) - \Phi^{\infty}_{\zeta}(t,x)\right| \le \frac{\log 2}{\beta}$.

Proof. Recall that $\Phi^{\beta}_{\zeta}(1,x)$ is convex and 1-Lipschitz while $\Phi^{\infty}_{\zeta}(1,x) = |x| \le \Phi^{\beta}_{\zeta}(1,x)$. It follows that

$$\begin{aligned}\sup_{x\in\mathbb{R}}\left|\Phi^{\beta}_{\zeta}(1,x) - \Phi^{\infty}_{\zeta}(1,x)\right| &= \left|\Phi^{\beta}_{\zeta}(1,0) - \Phi^{\infty}_{\zeta}(1,0)\right| \\ &= \frac{\log 2}{\beta}.\end{aligned}$$

Hence

$$\left|\mathcal{X}^{0,1}_{\zeta,\beta}(v,x) - \mathcal{X}^{0,1}_{\zeta,\infty}(v,x)\right| \le \frac{\log 2}{\beta}$$

holds for any control v, since the only difference is from the boundary value at time $t=1$ in $\mathcal{Y}$. Proposition 3.5.7 now implies the desired result. □

Below, recall the definition $\overline{\zeta} = \zeta|_{[q_0,1]}$.

Lemma 3.5.14. *Let $(\vec{p},\vec{q},\vec{k})$ be as in Section 3.3, and fix $\beta > 0$ and $\zeta \in \mathscr{L}$ such that $\overline{\zeta} \in \mathcal{M}_{\vec{q}}$. With*

$$\lambda = \int_0^1 \xi''(t)\kappa(t)\zeta(t)\,\mathrm{d}t$$

we have

$$F^{\mathrm{Is}}_N(\beta, \mathcal{Q}(\eta)) \le \beta K \mathsf{P}^{\mathrm{Is}}(\beta\kappa\zeta) + 3\beta^2K^2\xi''(1)\eta + K\beta\lambda\eta + K\log 2.$$

Proof. Applying Proposition 3.3.1 with $\overline{\zeta}$ and $(\beta^2\xi, \beta\boldsymbol{h}, \beta\lambda)$ in place of $(\xi,\boldsymbol{h},\lambda)$ in the first line,

$$\begin{aligned}
F^{\mathrm{Is}}_N(\beta,\mathcal{Q}(\eta)) - 3\beta^2K^2\xi''(1)\eta - K\beta\lambda\eta &\le \varphi^{\mathrm{Is}}_{\beta^2\xi}(0) - \tfrac{\beta^2K}{2}\textstyle\int_{q_0}^1 (q-q_0)\xi''(q)\kappa(q)\zeta(q)\,\mathrm{d}q \\
&\le \tfrac{1}{N}\textstyle\sum_{i=1}^N \Phi^{\mathbb{L}}_{\beta^2\xi,m_i,\overline{\zeta}}(q_0,\beta h+\beta\lambda m_i) - \tfrac{\beta^2K}{2}\int_{q_0}^1 (q-q_0)\xi''(q)\kappa(q)\zeta(q)\,\mathrm{d}q + \beta K R(\boldsymbol{h},\boldsymbol{m}) && \text{Props 3.5.1, 3.5.5} \\
&\le \tfrac{K}{N}\textstyle\sum_{i=1}^N \Phi_{\beta^2\xi,m_i,\kappa\overline{\zeta}}(q_0,\beta h+\beta\lambda m_i) - \tfrac{\beta^2K}{2}\int_{q_0}^1 (q-q_0)\xi''(q)\kappa(q)\zeta(q)\,\mathrm{d}q + \beta K R(\boldsymbol{h},\boldsymbol{m}) && \text{Lem 3.5.12} \\
&= \tfrac{\beta K}{N}\textstyle\sum_{i=1}^N \Phi^{\beta}_{m_i,\beta\kappa\overline{\zeta}}(q_0,h+\lambda m_i) - \tfrac{\beta^2K}{2}\int_{q_0}^1 (q-q_0)\xi''(q)\kappa(q)\zeta(q)\,\mathrm{d}q + \beta K R(\boldsymbol{h},\boldsymbol{m}) && \text{(3.5.7)} \\
&= \tfrac{\beta K}{N}\textstyle\sum_{i=1}^N \Phi^{\infty}_{m_i,\beta\kappa\overline{\zeta}}(q_0,h+\lambda m_i) - \tfrac{\beta K}{2}\int_{q_0}^1 (q-q_0)\xi''(q)\beta\kappa(q)\zeta(q)\,\mathrm{d}q \\
&\quad + \beta K R(\boldsymbol{h},\boldsymbol{m}) + K\log 2 && \text{Prop 3.5.13} \\
&\le \beta K \mathsf{P}^{\mathrm{Is}}(\beta\kappa\zeta) + K\log 2. && \text{Lem 3.5.11}
\end{aligned}$$

Here terms modified from the previous line are in red text.

□

All that remains is to approximate an arbitrary $\zeta_* \in \mathscr{L}$ by $\beta\kappa\zeta$ on $[q_0,1)$ for $\zeta \in \mathcal{M}_{\vec{q}}$ and choose parameters appropriately. We do this now.

Proof of Proposition 3.2.2, Ising case. First choose $\zeta_* = \zeta_*(\xi, h, \varepsilon) \in \mathscr{L}$ such that

$$\mathsf{P}^{\mathrm{Is}}(\zeta_*) \le \inf_{\zeta \in \mathscr{L}} \mathsf{P}^{\mathrm{Is}}(\zeta) + \frac{\varepsilon}{10} = \mathsf{ALG}^{\mathrm{Is}} + \frac{\varepsilon}{10}. \tag{3.5.8}$$

Since $\int_0^1 \xi''(t)\zeta_*(t)\,\mathrm{d}t < \infty$, the monotone convergence theorem guarantees

$$\lim_{\beta \to \infty} \int_0^1 \xi''(t) \cdot |\min(\zeta(t), \beta) - \zeta(t)|\,\mathrm{d}t = 0.$$

Define $\zeta_\beta(t) = \min(\zeta(t), \beta)$. Therefore there exists

$$\beta = \beta(\zeta_*, \xi, \varepsilon) = \beta(\xi, h, \varepsilon) \ge \frac{20 \log 2}{\varepsilon} \tag{3.5.9}$$

sufficiently large so that (recall Proposition 3.5.4)

$$\mathsf{P}^{\mathrm{Is}}(\zeta_\beta) - \mathsf{P}^{\mathrm{Is}}(\zeta_*) \le 2 \int_0^1 \xi''(t) \cdot |\zeta_\beta(t) - \zeta(t)|\,\mathrm{d}t \le \frac{\varepsilon}{10}. \tag{3.5.10}$$

For $\delta > 0$, let $q_0^\delta = q_0$ and $q_\delta^{d+1} = \min(q_d^\delta + \delta, 1)$. This determines D which satisfies $q_{D-1} < q_D = 1$. Since $\zeta_\beta \in \mathscr{L}$ is bounded and has bounded variation, there exists $\delta = \delta(\xi, \zeta_\beta, \varepsilon) = \delta(\xi, \boldsymbol{h}, \varepsilon) > 0$ such that the function

$$\zeta_{\beta,\delta}(t) = \begin{cases} \zeta_\beta(t), & t \in [0, q_0) \\ \max\left(\delta, \zeta_\beta(q_j^\delta)\right), & t \in [q_j^\delta, q_{j+1}^\delta), j \ge 0 \end{cases}$$

satisfies

$$\mathsf{P}^{\mathrm{Is}}(\zeta_\beta) - \mathsf{P}^{\mathrm{Is}}(\zeta_{\beta,\delta}) \le 2 \int_0^1 \xi''(t)\,|\zeta_\beta(t) - \zeta_{\beta,\delta}(t)|\,\mathrm{d}t \le \frac{\varepsilon}{10}. \tag{3.5.11}$$

(Note in particular that δ does not depend on q_0.) Observe that $\zeta_{\beta,\delta}(t) \in [\delta, \beta]$ holds for all $t \in [0, 1]$. Next define

$$k_1 = k_2 = \cdots = k_D = k_* \equiv \left\lceil \frac{\beta}{\delta^2} \right\rceil.$$

This leads to $p_d^\delta = \chi^{-1}(q_d^\delta)$ with $\delta \le p_1^\delta \le p_D^\delta = 1$ and hence $\kappa(t) = \kappa_d$ for $t \in [q_d^\delta, q_{d+1}^\delta)$, where

$$\delta k_*^{D-d} \le \kappa_d \le k_*^{D-d}.$$

Next define

$$\widehat{\zeta}_{\beta,\delta}(t) \equiv \frac{\zeta_{\beta,\delta}(t)}{\beta\kappa(t)}, \quad t \in [q_0, 1]$$

so that $\beta\kappa\widehat{\zeta}_{\beta,\delta} = \zeta_{\beta,\delta}$. Note that

$$\sup_{t \in [0,1]} \widehat{\zeta}_{\beta,\delta}(t) \le \frac{\sup_{s \in [0,1]} \zeta_{\beta,\delta}(s)}{\beta} \le 1.$$

Additionally $\widehat{\zeta}_{\beta,\delta}$ is nondecreasing since if $q_d \le t_d < q_{d+1}^\delta \le t_{d+1} \le q_{d+2}^\delta$, then

$$\begin{aligned} \frac{\widehat{\zeta}_{\beta,\delta}(t_d)}{\widehat{\zeta}_{\beta,\delta}(t_{d+1})} &= \frac{\zeta_{\beta,\delta}(t_d)}{\zeta_{\beta,\delta}(t_d)} \cdot \frac{\kappa_{d+1}}{\kappa_d} \\ &\le \frac{\beta}{\delta^2 k_*} \le 1 \end{aligned}$$

by definition of k_*. Set

$$\lambda = \int_0^1 \xi''(t)\kappa(t)\widehat{\zeta}_{\beta,\delta}(t)\,\mathrm{d}t$$

and

$$\eta = \frac{\varepsilon}{30\beta K\xi''(1) + 10\lambda}. \tag{3.5.12}$$

We now show that using $\widehat{\zeta}_{\beta,\delta}$ in the interpolation implies Proposition 3.2.2. Take $\vec{p}, \vec{q}, \vec{k}, D, \beta, \eta$ as above.

$$\begin{aligned}
\tfrac{1}{N}\,\mathbb{E}\left[\max_{\vec{\sigma}\in\mathcal{Q}^{\mathrm{Is}}(\eta)} \mathcal{H}_N^{\vec{k},\vec{p}}(\vec{\sigma})\right] &\le F_N^{\mathrm{Is}}(\beta, \mathcal{Q}(\eta))/\beta && \\
&\le K\mathsf{P}^{\mathrm{Is}}(\beta\kappa\widehat{\zeta}_{\beta,\delta}) + 3\beta K^2\xi''(1)\eta + K\lambda\eta + \tfrac{K\log 2}{\beta} && \text{Lem 3.5.14} \\
&\le K\cdot\mathsf{P}^{\mathrm{Is}}(\zeta_{\beta,\delta}) + \tfrac{2K\varepsilon}{10} && \text{(3.5.9), (3.5.12)} \\
&\le K\cdot\mathsf{P}^{\mathrm{Is}}(\zeta_{\beta}) + \tfrac{3K\varepsilon}{10} && \text{(3.5.11)} \\
&\le K\cdot\mathsf{P}^{\mathrm{Is}}(\zeta_{*}) + \tfrac{4K\varepsilon}{10} && \text{(3.5.10)} \\
&\le K\cdot\mathsf{ALG}^{\mathrm{Is}} + \tfrac{5K\varepsilon}{10}. && \text{(3.5.8)}
\end{aligned}$$

Moreover the values D, η and K above are bounded depending only on ξ, h and ε. Indeed $D \le \delta^{-1}+1$, η is bounded as in (3.5.12), and $K = \prod_{d=1}^{D} k_i = k_*^D = \left\lceil \frac{\beta}{\delta^2} \right\rceil^D$. Meanwhile β as defined in (3.5.9) also depends only on ξ, h, ε. This concludes the proof.

□

3.5.4 Deferred Proofs

Here we give the missing proofs for this section, which are all relatively standard.

Proof of Lemma 3.5.6. We assume $\zeta(t) > 0$ as the $\zeta(t) = 0$ case is clear. We consider only the case $t \in [q_{D-1}, 1)$ as the remaining cases are identical by induction. Let $\vec{y} = \vec{y}(t) \in \mathbb{R}^K$ be the Gaussian random vector

$$\vec{y} = e\vec{t}a_D(\xi''(1) - \xi''(t))^{1/2}.$$

Below A always denotes

$$A(\vec{x}+\vec{y}) = \Phi^{\mathbb{L}}_{a,\zeta}(1, \vec{x}+\vec{y})$$

and for convenience we set $m = \zeta_D = \zeta(t)$ for $t \in [q_{D-1}, 1)$. First note that since $|\partial_{x(u)}\Phi^{\mathbb{L}}_{a,\zeta}(1,\vec{x})| \le 1 + |a|$ holds, there are no issues of convergence in any of the expectations even though $\vec{y}$ has unbounded support.

By differentiating in the endpoint value $\vec{x}+\vec{y}$ before taking expectation in $\vec{y}$ it follows that

$$\nabla\Phi^{\mathbb{L}}_{a,\zeta} = \frac{\mathbb{E}[\nabla A e^{mA}]}{\mathbb{E}[e^{mA}]}.$$

This immediately implies that $|\partial_{x(u)}\Phi^{\mathbb{L}}_{a,\zeta}(t,\vec{x})| \le 1 + |a|$. Similarly one has

$$\partial_{x_ix_j}\Phi^{\mathbb{L}}_{a,\zeta} = \frac{\mathbb{E}\left[\partial_{x_ix_j}A + m(\partial_{x_i}A)(\partial_{x_j}A)e^{mA}\right]}{\mathbb{E}[e^{mA}]} - m\left(\frac{\mathbb{E}[\partial_{x_i}e^{mA}]}{\mathbb{E}[e^{mA}]}\right)\left(\frac{\mathbb{E}[\partial_{x_j}e^{mA}]}{\mathbb{E}[e^{mA}]}\right).$$

Combining, we compute

$$\langle T, \nabla^2\Phi^{\mathbb{L}}_{a,\zeta}\rangle + m\langle T, (\nabla L)^{\otimes 2}\rangle = \frac{1}{\mathbb{E}[e^{mA}]}\,\mathbb{E}[\left(\langle T, \nabla^2 A\rangle + m\langle T, (\nabla A)^{\otimes 2}\rangle\right)e^{mA}]$$

Next, note that the time-derivative of the covariance of $\vec{y}(t)$ is $M(t)$. Since $M(t)$ is positive semidefinite we can couple together $(\vec{y}(t))_{t\in[q_{L-1},1]}$ via

$$\vec{y}(t) = \int_t^1 \sqrt{\xi''(r)M(r)}\,\mathrm{d}\vec{B}_r$$

where $\vec{B}_r$ is a standard Brownian motion in $\mathbb{R}^K$. Applying Ito's formula backward in time now implies

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}\,e^{mA(\vec{x},\vec{y}(t))} = -\frac{1}{2}m\,\mathbb{E}\left[\left(\langle T, \nabla^2 A\rangle + m\langle M(t), (\nabla A)^{\otimes 2}\rangle\right)e^{mA}\right].$$

Therefore we conclude

$$\begin{aligned}\partial_t \Phi^{\mathbb{L}}_{a,\zeta} &= -\frac{\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}e^{mA(\vec{x},\vec{y}(t))}}{m\,\mathbb{E}\,e^{mA(\vec{x},\vec{y}(t))}} \\ &= -\frac{1}{2}\langle T, \nabla^2 \Phi^{\mathbb{L}}_{a,\zeta}\rangle + m\langle T, (\nabla \Phi^{\mathbb{L}}_{a,\zeta})^{\otimes 2}\rangle.\end{aligned}$$

□

Proof of Proposition 3.5.7. Set

$$W_s = x + \int_{t_1}^{s} \zeta(r)\xi''(r)v_r \,\mathrm{d}r + \int_{t_1}^{s} \sqrt{\xi''(r)}\,\mathrm{d}B_r$$

and

$$V_s \equiv \Phi^{\beta}_{a,\zeta}(s, W_s) - \frac{1}{2}\int_{t_1}^{s} \zeta(r)\xi''(r)v_r^2 \,\mathrm{d}r.$$

Ito's formula gives

$$\mathrm{d}V_t = \left(\partial_t \Phi^{\beta}_{a,\zeta}(t,W_t) + \zeta(t)\xi''(t)v_t\partial_x\Phi^{\beta}_{a,\zeta}(t,W_t) + \frac{1}{2}\xi''(t)\partial_{xx}\Phi^{\beta}_{a,\zeta}(t,W_t) - \frac{1}{2}\zeta(t)\xi''(t)v_t^2\right)\,\mathrm{d}t + Y_t\,\mathrm{d}B_t.$$

Here Y_t is irrelevant and (3.5.1) lets us rewrite the finite variation part of $\mathrm{d}V_t$ as

$$\begin{aligned}&\partial_t \Phi^{\beta}_{a,\zeta}(t,X_t) + \zeta(t)\xi''(t)v_t\partial_x\Phi^{\beta}_{a,\zeta}(t,W_t) + \frac{1}{2}\xi''(t)\partial_{xx}\Phi^{\beta}_{a,\zeta}(t,W_t) - \frac{1}{2}\zeta(t)\xi''(t)v_t^2 \\ &\qquad = -\frac{1}{2}\zeta(t)\xi''(t)\left(v_t - \partial_x\Phi^{\beta}_{a,\zeta}(t,W_t)\right)^2 \\ &\qquad \le 0.\end{aligned}$$

We conclude that

$$\Phi^{\beta}_{\zeta}(t_1,x) \ge \mathcal{X}^{t_1,t_2}_{\zeta}(x,v)$$

with equality when $v_r = \partial_x\Phi^{\beta}_{\zeta}(r,W_r)$ holds for all $r \in [t_1,t_2]$. By uniqueness of solutions for SDEs with Lipschitz coefficients, this implies $W_r = X_r$.

□

Proof of Proposition 3.5.8. The proof is similar to the 1-dimensional case. First, the SDE defining $\vec{X}_t$ has strong and pathwise unique solutions since $\nabla\Phi^{\mathbb{L}}_{a,\zeta}(t,\vec{x})$ is uniformly bounded and Lipschitz in $\vec{x}$. Set

$$\vec{W}_s = \vec{x} + \int_{t_1}^{s} \zeta(r)\xi''(r)M(r)\vec{v}_r \,\mathrm{d}r + \int_{t_1}^{s} \sqrt{\xi''(r)M(r)}\,\mathrm{d}\vec{B}(r)$$

and

$$V_s^{\mathbb{L}} \equiv \Phi^{\mathbb{L}}_{a,\zeta}\left(s,\vec{X}_s\right) - \frac{1}{2}\int_{t_1}^{s} \zeta(r)\xi''(r)\langle M(r), \vec{v}_r^{\otimes 2}\rangle \,\mathrm{d}r.$$

By Ito's formula,

$$\mathrm{d}V_t^{\mathbb{L}} = \left(\partial_t\Phi^{\mathbb{L}}_{\zeta}(t,\vec{W}_t) + \zeta(t)\xi''(t)\vec{v}_t\partial_x\Phi^{\mathbb{L}}_{\zeta}(t,\vec{W}_t) + \frac{1}{2}\xi''(t)\partial_{xx}\Phi^{\mathbb{L}}_{\zeta}(t,\vec{X}_t) - \frac{\xi''(t)}{2}\langle M(t),\vec{v}_t^{\otimes 2}\rangle\right)\,\mathrm{d}t + Y_t^{\mathbb{L}}\,\mathrm{d}B_t.$$

Here $Y_t^{\mathbb{L}}$ is again irrelevant. By (3.5.3) the finite variation part of $\mathrm{d}V_t^{\mathbb{L}}$ is

$$\begin{aligned}&\partial_t\Phi^{\mathbb{L}}_{a,\zeta}(t,\vec{W}_t) + \left\langle M(t), \vec{v}_t \otimes \nabla\Phi^{\mathbb{L}}_{a,\zeta}(t,\vec{W}_t)\right\rangle + \frac{1}{2}\xi''(t)\partial_{xx}\Phi^{\mathbb{L}}_{a,\zeta}(t,\vec{W}_t) - \frac{\xi''(t)}{2}\langle M(t),\vec{v}_t^{\otimes 2}\rangle \\ &\qquad = -\frac{1}{2}\left\langle M(t), \left(\vec{v}_t - \nabla\Phi^{\mathbb{L}}_{a,\zeta}(t,\vec{W}_t)\right)^{\otimes 2}\right\rangle \\ &\qquad \le 0.\end{aligned}$$

We conclude that

$$\Phi^{\mathbb{L}}_{a,\zeta}(t_1,\vec{x}) \geq \mathcal{X}^{\mathbb{L},t_1,t_2}_{a,\zeta}(\vec{x},\vec{v})$$

with equality when

$$\vec{v}_r = \nabla \Phi^{\mathbb{L}}_{a,\zeta}(r,\vec{W}_r)$$

holds for all $r \in [t_1,t_2]$. Again, uniqueness of solutions to SDEs with Lipschitz coefficients implies $\vec{W}_r = \vec{X}_r$. □

3.6 Necessity of Full Branching Trees

In this section we show, roughly speaking, that it is necessary to use a full branching tree to obtain our results within the overlap gap framework. We restrict for convenience to the setting of spherical models with null external field $h = 0$ and set $\mathsf{ALG}^{\mathrm{Sp}}_{\xi} = \mathsf{ALG}^{\mathrm{Sp}}_{\xi,0} = \int_0^1 \xi''(t)^{1/2}\,\mathrm{d}t$ (recall Proposition 3.1.6) and $\mathsf{OPT}^{\mathrm{Sp}}_{\xi} = \mathsf{OPT}^{\mathrm{Sp}}_{\xi,0}$.

A consequence of Theorem 3.6.13, proved near the end of this section, can be expressed informally as follows for any ξ with $\mathsf{ALG}^{\mathrm{Sp}}_{\xi} < \mathsf{OPT}^{\mathrm{Sp}}_{\xi}$. Recall the canonical bijection between finite ultrametric spaces and edge-weighted rooted trees (or see Subsection 3.6.2 for a reminder). For all finite ultrametric spaces X of diameter at most $\sqrt{2}$ whose corresponding rooted tree does not contain a subdivision of a full binary subtree of depth D, with probability at least $1 - e^{-\Omega(N)}$ the following holds. There exists an isometric (up to the scaling factor $\sqrt{N}$) embedding $\iota : X \to S_N$ such that

$$H_N(\iota(x)) \geq (\mathsf{ALG}^{\mathrm{Sp}}_{\xi} + \varepsilon_{\xi,D})N, \quad \forall x \in X.$$

Here $\varepsilon_{\xi,D} > 0$ is a constant depending only on ξ and D, and in particular is independent of the size of the ultrametric X. In other words, to rule out algorithms achieving better than $\mathsf{ALG}^{\mathrm{Sp}}_{\xi} + \varepsilon$ using forbidden ultrametrics, as $\varepsilon \to 0$ it is necessary to take $D \to \infty$, in effect using the full power of Proposition 3.2.2.

The full statement of Theorem 3.6.13 shows that in fact a super-constant amount of branching must occur at all "depths" in $[0,1]$ where $\xi''(t)^{-1/2}$ is strictly convex. We also show in Theorem 3.6.14 that there exists an embedding ι as above with large average energy

$$\frac{1}{|X|}\sum_{x \in X} H_N(\iota(x)) \geq (\mathsf{ALG}^{\mathrm{Sp}}_{\xi} + \varepsilon_{\xi,D})N$$

unless "almost all of" X branches a super-constant amount at "almost all such depths". Note that this average energy is what the Guerra-Talagrand interpolation actually allows one to upper bound. Throughout this section we always consider just a single Hamiltonian H_N. This corresponds to the case $\vec{p} \approx (1,1,\ldots,1)$, i.e. a correlation function $\chi(p)$ which sharply increases near $p = 1$ such as $\chi(p) = p^{100}$.

Our plan to prove Theorem 3.6.13 is as follows. If $\mathsf{ALG}^{\mathrm{Sp}}_{\xi} < \mathsf{OPT}^{\mathrm{Sp}}_{\xi}$, there exists an interval $[a,b] \subseteq [0,1]$ on which $(\xi'')^{-1/2}$ is strictly convex. Let $\mathbb{T}$ be the finite rooted tree with leaf set corresponding to the ultrametric space X. Let $\varepsilon > 0$ be a small constant depending only on ξ and D. We use the algorithm of [Sub21] to find embeddings of ancestor points $\iota(x_a)$ for each $x \in X$ of norm $\|\iota(x_a)\|_2 = \sqrt{aN}$ which satisfy

$$H_N(\iota(x_a)) \geq \left(\int_0^a \xi''(t)^{1/2}\,\mathrm{d}t - \varepsilon\right) N.$$

Next we embed the depth $[a,b]$ parts of $\mathbb{T}$ so that the resulting depth b ancestor embeddings $\iota(x_b)$ satisfy

$$H_N(\iota(x_b)) \geq \left(\int_0^b \xi''(t)^{1/2}\,\mathrm{d}t + 2\varepsilon\right) N.$$

In other words, from radius $\sqrt{aN}$ to $\sqrt{bN}$, the embedded points' energy grows by $\int_a^b \xi''(t)^{1/2}\,\mathrm{d}t + 3\varepsilon$, which exceeds the maximum possible growth of an overlap concentrated algorithm by a small constant 3ε. This is the main step of our procedure, and it succeeds whenever the portion of $\mathbb{T}$ at depths in $[a,b]$ does not contain a full binary tree of depth D. The proof uses induction on D, and the $D = 1$ case is described in

Figure 3.1. We remark that our proof is essentially constructive assuming access to an oracle to find many orthogonal near-maximizers of H_N on arbitrary bands as guaranteed by Lemma 3.6.4.

Finally we again use the algorithm of [Sub21] to define embeddings of the leaves $\iota(x) \in S_N$ for $x \in X$ with

$$H_N(\iota(x)) \geq \left(\int_0^1 \xi''(t)^{1/2} \, \mathrm{d}t + \varepsilon \right) N.$$

We remark that in previous multi-OGP arguments, ultrametricity of the forbidden configuration does not explicitly enter. However in these arguments, it is always *possible* that the structure of replicas identified is an ultrametric. Specifically, in a "star" multi-OGP [RV17, GS17, GK21a] all the replicas are pairwise equidistant. For the "ladder" OGP implementations of [Wei20, BH21], the forbidden structure is defined by applying some stopping rule to choose a finite number of solutions from a "stably evolving" sequence of algorithmic outputs. In both settings it is possible that the resulting configuration is a star ultrametric with all pairwise nonzero distances equal. However, the rooted tree corresponding to such an ultrametric does not contain even a full binary tree of depth $D = 2$. Therefore Theorem 3.6.13 strongly suggests that existing OGP arguments are incapable of ruling out Lipschitz $\mathcal{A}$ from achieving energies down to the algorithmic threshold $\mathsf{ALG}_\xi^{\mathrm{Sp}}$.

3.6.1 Preparation

For given ξ and $t \in [0,1]$, define

$$\mathsf{ALG}_\xi^{\mathrm{Sp}}(t) = \int_0^t \xi''(s)^{1/2} \, \mathrm{d}s$$

so that $\mathsf{ALG}_\xi^{\mathrm{Sp}}(1) = \mathsf{ALG}_\xi^{\mathrm{Sp}}$. Define also

$$\mathsf{ALG}_\xi^{\mathrm{Sp}}([a,b]) = \mathsf{ALG}_\xi^{\mathrm{Sp}}(b) - \mathsf{ALG}_\xi^{\mathrm{Sp}}(a).$$

Define

$$\xi_a(t) = \xi(t) - \xi(a) - (t-a)\xi'(a).$$

Note that $\xi_a(a) = \xi_a'(a) = 0$, and $\xi_a''(t) = \xi''(t)$ for all t. Define the rescaled mixture function

$$\xi_{[a,b]}(t) = \xi_a\left(a + (b-a)t\right).$$

We derive

$$\mathsf{ALG}_{\xi_{[a,b]}}^{\mathrm{Sp}} = \int_0^1 \sqrt{\xi''_{[a,b]}(t)} \, \mathrm{d}t = \int_a^b \sqrt{\xi_a''(s)} \, \mathrm{d}s = \int_a^b \sqrt{\xi''(s)} \, \mathrm{d}s = \mathsf{ALG}_\xi^{\mathrm{Sp}}([a,b]).$$

Correspondingly, define

$$\mathsf{OPT}_\xi^{\mathrm{Sp}}([a,b]) = \mathsf{OPT}_{\xi_{[a,b]}}^{\mathrm{Sp}}.$$

Proposition 3.6.1. *Suppose* $\frac{\mathrm{d}^2}{\mathrm{d}t^2}(\xi''(t)^{-1/2}) > 0$ *for* $t \in [a,b] \subseteq [0,1]$. *Then*

$$\mathsf{OPT}_\xi^{\mathrm{Sp}}([a,b]) > \mathsf{ALG}_\xi([a,b]). \tag{3.6.1}$$

Proof. The result follows from Proposition 3.1.6 applied to $\xi_{[a,b]}$. □

The next proposition follows from the work [Sub18] and ensures the existence of many approximately orthogonal replicas which each approximately achieve the ground state energy in spherical spin glasses without external field. In Lemma 3.6.4 we make several simple modifications to this result, for instance requiring that the replicas be exactly orthogonal.

Proposition 3.6.2. *Suppose* $\frac{\mathrm{d}^2}{\mathrm{d}t^2}(\xi''(t)^{-1/2}) > 0$ *for* $t \in [0,1]$. *Then for any* $C, \varepsilon > 0$ *and* $k \in \mathbb{N}$, *for* $N \geq N_0 = N_0(\xi, C, \varepsilon, k)$, *with probability at least* $1 - e^{-CN}$ *either* $H_N \notin K_N$ *(recall Proposition 3.0.1) or the following holds. There exist* k *points* $\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_k \in S_N$ *with*

$$|R(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j)| \leq \varepsilon, \quad 1 \leq i < j \leq k$$

and

$$H_N(\boldsymbol{\sigma}_i) \geq N(\mathsf{OPT}_\xi^{\mathrm{Sp}} - \varepsilon), \quad i \in [k].$$

Proof. With the absence of external field, it follows from [Sub18, Lemma 42] that 0 is multi-samplable. Let $\mathcal{Q}_k(\varepsilon) \subseteq B(\boldsymbol{m},\varepsilon)^k \cap S_N^k$ denote the set of $\vec{\boldsymbol{\sigma}}$ with $|R(\boldsymbol{\sigma}_i,\boldsymbol{\sigma}_j)| \le \varepsilon$ for $i \neq j$. Let μ be the uniform measure on S_N. Define

$$F_{N,\beta} = \frac{1}{\beta N} \log \int_{S_N} \exp \beta H_N(\boldsymbol{\sigma}) \, \mathrm{d}\mu(\boldsymbol{\sigma})$$

to be the quenched free energy of H_N on S_N at inverse temperature β and

$$\widetilde{F}_{N,\beta}(\boldsymbol{m}) = \widetilde{F}_{N,\beta}(\boldsymbol{m}, k_N, \varepsilon) \equiv \frac{1}{\beta N k_N} \log \int_{\mathcal{Q}_{k_N}(\varepsilon)} \exp \beta \sum_{i=1}^{k_N} H_N(\boldsymbol{\sigma}_i) \, \mathrm{d}\mu^k(\vec{\boldsymbol{\sigma}}).$$

Here k_N grows to ∞ with N at a suitably slow rate. By [Sub18, Proposition 1 and Theorem 3][3] it follows that for $N \ge N_0$ sufficiently large,

$$\mathbb{P}\left[\mathbb{E}\,\widetilde{F}_{N,\beta}(\boldsymbol{m}) - \mathbb{E}\,F_{N,\beta} \ge -\varepsilon\right] \ge 1 - e^{-CN}.$$

Therefore there exists some $\vec{\boldsymbol{\sigma}} \in \mathcal{Q}_{k_N}(\varepsilon)$ satisfying

$$\sum_{i=1}^{k_N} H_N(\boldsymbol{\sigma}_i) \ge N k_N (\mathsf{OPT}_\xi^{\mathrm{Sp}} - \varepsilon - o_\beta(1)).$$

Here $o_\beta(1)$ is a value tending to 0 as $\beta \to \infty$, uniformly in everything else. Assuming $H_N \in K_N$, the values $\frac{1}{N}|H_N(\boldsymbol{\sigma}_i)|$ are uniformly bounded by a constant C_1 (because $H_N(0) = 0$). It follows by Markov's inequality that at least $k_N\left(\frac{\varepsilon}{10C_1} - \varepsilon - o_\beta(1)\right)$ of the $\boldsymbol{\sigma}_i$ satisfy $H_N(\boldsymbol{\sigma}_i) \ge N(\mathsf{OPT}_\xi^{\mathrm{Sp}} - \frac{\varepsilon}{2} - o_\beta(1))$. Since $k_N \to \infty$, eventually

$$k \le \left\lfloor k_N \left(\frac{\varepsilon}{10C_1} - \varepsilon - o_\beta(1)\right)\right\rfloor$$

for suitably large β, which completes the proof. □

For fixed $\boldsymbol{m}$, define the first-order Taylor expansion

$$\overline{H}_N^{\boldsymbol{m}}(\boldsymbol{\sigma}) = H_N(\boldsymbol{m}) + \langle \nabla H_N(\boldsymbol{m}), \boldsymbol{\sigma} - \boldsymbol{m}\rangle.$$

of H_N and write

$$H_N = \overline{H}_N^{\boldsymbol{m}} + \widehat{H}_N^{\boldsymbol{m}}.$$

For $0 \le a \le b \le 1$ with $\boldsymbol{m} \in \sqrt{a}\cdot S_N$, define $B(\boldsymbol{m},0,b) = B(\boldsymbol{m},0) \cap \sqrt{b}\cdot S_N$ and its convex hull $B(\boldsymbol{m},0,[a,b])$.

Lemma 3.6.3. *For any fixed $\boldsymbol{m}$, the law of $\widehat{H}_N^{\boldsymbol{m}}$ restricted to $B(\boldsymbol{m},0)$ is a Gaussian process with covariance*

$$\mathbb{E}[\widehat{H}_N^{\boldsymbol{m}}(\boldsymbol{\sigma}^1)\widehat{H}_N^{\boldsymbol{m}}(\boldsymbol{\sigma}^2)] = N\xi_a(R(\boldsymbol{\sigma}^1,\boldsymbol{\sigma}^2)). \tag{3.6.2}$$

Moreover the restrictions of $\widehat{H}_N^{\boldsymbol{m}}$ and $\overline{H}_N^{\boldsymbol{m}}$ to $B(\boldsymbol{m},0)$ are independent.

Proof. Note that for all $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2 \in B(\boldsymbol{m},0)$,

$$R(\boldsymbol{\sigma}^1 - \boldsymbol{m}, \boldsymbol{\sigma}^2 - \boldsymbol{m}) = R(\boldsymbol{\sigma}^1,\boldsymbol{\sigma}^2) - a.$$

Since $\xi_a(t)$ has all derivatives non-negative for $t \ge a$, we may sample a centered Gaussian process $\widetilde{H}_N$ on $B(\boldsymbol{m},0,[a,1])$ with covariance given by

$$\begin{aligned}\mathbb{E}[\widetilde{H}_N(\boldsymbol{\sigma}^1)\widetilde{H}_N(\boldsymbol{\sigma}^2)] &= N\xi_a(R(\boldsymbol{\sigma}^1 - \boldsymbol{m}, \boldsymbol{\sigma}^2 - \boldsymbol{m}) + a)\\ &= N\xi_a(R(\boldsymbol{\sigma}^1,\boldsymbol{\sigma}^2)).\end{aligned}$$

[3] In the statement of [Sub18, Theorem 3], there are values δ_N, ρ_N which also shrink with N. We are taking $\varepsilon = \rho_N$ a small constant and ignoring the constraint from δ, so our value of $\widetilde{F}_{N,\beta}(\boldsymbol{m})$ is larger than that of [Sub18]. Therefore the lower bound on $\widetilde{F}_{N,\beta}(\boldsymbol{m})$ we use is somewhat weaker than in the results cited.

Next, generate the independent centered Gaussian process $\underline{H}_N$ by

$$\mathbb{E}[\underline{H}_N(\boldsymbol{\sigma}^1)\underline{H}_N(\boldsymbol{\sigma}^2)] = N\left(\xi(a) + \xi'(a)\left(R(\boldsymbol{\sigma}^1,\boldsymbol{\sigma}^2) - a\right)\right).$$

It follows by adding covariances (with $x = R(\boldsymbol{\sigma}^1,\boldsymbol{\sigma}^2)$ in the definition of ξ_a) that

$$\widetilde{H}_N + \underline{H}_N \overset{d}{=} H_N$$

when restricted to $B(\boldsymbol{m},0)$. Since $\xi_a(a) = \xi_a'(a) = 0$, it follows that $\widetilde{H}_N(\boldsymbol{m}) = 0$ and $\nabla\widetilde{H}_N(\boldsymbol{m}) = 0$ hold almost surely. Therefore $\underline{H}_N = \overline{H}_N^{\boldsymbol{m}}$ is the first-order Taylor expansion of H_N around $\boldsymbol{m}$, and then also $\widetilde{H}_N = \widehat{H}_N^{\boldsymbol{m}}$. Moreover $\widetilde{H}_N$ and $\underline{H}_N$ are independent by construction. This concludes the proof. □

In the following Lemma 3.6.4, we refine Proposition 3.6.2 in several simple but convenient ways. In particular, Lemma 3.6.3 implies the same result uniformly over all bands $B(\boldsymbol{m},0,b)$; it also guarantees exact orthogonality. Lemma 3.6.4 will serve as a useful tool for embedding more complicated ultrametric trees. Roughly speaking, it gives a way to gain on the embedding algorithm of [Sub21] (stated later as Proposition 3.6.10).

Lemma 3.6.4. *Suppose* $\frac{\mathrm{d}^2}{\mathrm{d}t^2}(\xi''(t)^{-1/2}) > 0$ *for* $t \in [a,b] \subseteq [0,1]$. *Then there exists* $\varepsilon > 0$ *depending only on* ξ, a, b *such that for any* k, *for* $N \geq N_0(\xi,a,b,k)$ *sufficiently large and some* $c = c(\xi,a,b,k)$, *with probability* $1 - e^{-cN}$ *the following holds.*

For any $\boldsymbol{m}$ *with* $||\boldsymbol{m}||_N^2 = a \leq 1$ *and any linear subspace* $W \subseteq \mathbb{R}^N$ *with* $\dim(W) \geq N - k$, *there exist* k *points* $\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_k \in W + \boldsymbol{m}$ *such that*

$$R(\boldsymbol{\sigma}_i - \boldsymbol{m}, \boldsymbol{\sigma}_j - \boldsymbol{m}) = (b-a)\cdot \mathbb{1}\{i=j\} \quad \forall i,j \in [k] \tag{3.6.3}$$

and

$$H_N(\boldsymbol{\sigma}_i) \geq H_N(\boldsymbol{m}) + N(\mathsf{ALG}_\xi^{\mathrm{Sp}}([a,b]) + \varepsilon) \quad \forall i \in [k]. \tag{3.6.4}$$

Proof. Consider a (non-random) $\eta\sqrt{N}$-net $\mathcal{N}_\eta$ on $\sqrt{a}\cdot S_N$ of size at most $(10/\eta)^N$. For any $\boldsymbol{m} \in \sqrt{a}\cdot S_N$, the Hamiltonian $\widehat{H}_N^{\boldsymbol{m}}(\boldsymbol{\sigma})$ restricted to $B(\boldsymbol{m},0,b)$ has covariance

$$\begin{aligned}\mathbb{E}\,\widehat{H}_N^{\boldsymbol{m}}(\boldsymbol{\sigma}_1)\widehat{H}_N^{\boldsymbol{m}}(\boldsymbol{\sigma}_2) &= N\xi_a(R(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2)) \\ &= N\xi_{[a,b]}\left(R\left(\frac{\boldsymbol{\sigma}_1-\boldsymbol{m}}{\sqrt{b-a}}, \frac{\boldsymbol{\sigma}_2-\boldsymbol{m}}{\sqrt{b-a}}\right)\right).\end{aligned}$$

Since

$$||\boldsymbol{\sigma} - \boldsymbol{m}||_2 = \sqrt{N(b-a)}$$

for $\boldsymbol{\sigma} \in B(\boldsymbol{m},0,b)$, we conclude that $\widehat{H}_N^{\boldsymbol{m}}$ is exactly an $N-1$ dimensional spin glass with mixture $\xi_{[a,b]}$ on $B(\boldsymbol{m},0,b)$ up to rescaling the input.

Fix a large constant M, and choose ε sufficiently small depending on M. We apply Proposition 3.6.2 to $\widehat{H}_N^{\boldsymbol{m}}$ with mixture $\xi_{[a,b]}(t)$ based on the observation just above. Recall that the constant C in Proposition 3.6.2 can be arbitrarily large. It follows by a union bound that with probability $1 - e^{-C_1 N}$, for all $\boldsymbol{n} \in \mathcal{N}_\eta$ there exist $\widetilde{\boldsymbol{\sigma}}_1(\boldsymbol{n}),\dots,\widetilde{\boldsymbol{\sigma}}_M(\boldsymbol{n})$ satisfying

$$|R(\widetilde{\boldsymbol{\sigma}}_i(\boldsymbol{n}) - \boldsymbol{n}, \widetilde{\boldsymbol{\sigma}}_j(\boldsymbol{n}) - \boldsymbol{n}) - (b-a)\cdot\mathbb{1}\{i=j\}| \leq \varepsilon \quad \forall 1 \leq i < j \leq M$$

and

$$\widehat{H}_N(\widetilde{\boldsymbol{\sigma}}_i(\boldsymbol{n})) \geq N(\mathsf{OPT}_\xi^{\mathrm{Sp}}([a,b]) - \varepsilon) \quad \forall i \in [M]. \tag{3.6.5}$$

For any $\boldsymbol{m} \in \sqrt{a}\cdot S_N$, there exists by definition $\boldsymbol{n} \in \mathcal{N}_\eta$ with $||\boldsymbol{m} - \boldsymbol{n}|| \leq \eta\sqrt{N}$. Then with $\widetilde{\boldsymbol{\sigma}}_i = \widetilde{\boldsymbol{\sigma}}_i(\boldsymbol{n})$ as above,

$$|R(\widetilde{\boldsymbol{\sigma}}_i - \boldsymbol{m}, \widetilde{\boldsymbol{\sigma}}_j - \boldsymbol{m}) - (b-a)\cdot\mathbb{1}\{i=j\}| \leq \varepsilon_1 \quad \forall 1 \leq i < j \leq M$$

for some $\varepsilon_1 = o_{\varepsilon,\eta}(1)$ tending to 0 as $\varepsilon,\eta \to 0$. Define the linear subspace $\widetilde{W} \subseteq W$ by

$$\widetilde{W} = W \cap \boldsymbol{m}^\perp \cap (\nabla H_N)^\perp$$

where $(\cdot)^\perp$ denotes orthogonal complement. Let $P_{\widetilde{W}^\perp}$ be the orthogonal projection matrix onto $\widetilde{W}^\perp$. It is easy to see that

$$\left|\left|\sum_{i=1}^{M}(\widetilde{\boldsymbol{\sigma}}_i - \boldsymbol{m})^{\otimes 2}\right|\right|_2^2 \le (1+M\varepsilon)N \le 2N$$

for ε sufficiently small. Then

$$\begin{aligned}\sum_{i=1}^{M}||P_{\widetilde{W}^\perp}(\widetilde{\boldsymbol{\sigma}}_i)||_2^2 &= \left\langle P_{\widetilde{W}^\perp}, \sum_{i=1}^{M}(\widetilde{\boldsymbol{\sigma}}_i - \boldsymbol{m})^{\otimes 2}\right\rangle \\ &\le ||P_{\widetilde{W}^\perp}||_2^2 \cdot \left|\left|\sum_{i=1}^{M}(\boldsymbol{\sigma}_i - \boldsymbol{m})^{\otimes 2}\right|\right|_2^2 \\ &\le 2(k+2)N.\end{aligned}$$

By the pigeonhole principle, at most $M-k$ values $i \in [M]$ can satisfy

$$||P_{\widetilde{W}^\perp}(\boldsymbol{\sigma}_i - \boldsymbol{m})||_2^2 \ge \frac{2(k+2)N}{M-k}.$$

It follows that there exist a subset $\widetilde{\boldsymbol{\sigma}}_{i_1}, \dots, \widetilde{\boldsymbol{\sigma}}_{i_k}$ with

$$||P_{\widetilde{W}^\perp}(\widetilde{\boldsymbol{\sigma}}_{i_j} - \boldsymbol{m})||_2^2 \le \eta N, \quad j \in [k]$$

where $\eta \le \frac{2(k+2)}{M-k}$ is arbitrarily small (by choosing M large depending on k). Defining $\boldsymbol{\sigma}'_{i_1}, \dots, \boldsymbol{\sigma}'_{i_k}$ by

$$\boldsymbol{\sigma}'_{i_j} - \boldsymbol{m} = P_{\widetilde{W}^\perp}(\widetilde{\boldsymbol{\sigma}}_{i_j} - \boldsymbol{m}),$$

we have

$$\boldsymbol{\sigma}'_{i_1}, \dots, \boldsymbol{\sigma}'_{i_k} \in \boldsymbol{m} + \widetilde{W}$$

satisfying

$$|R(\boldsymbol{\sigma}'_{i_j} - \boldsymbol{m}, \boldsymbol{\sigma}_{i_\ell} - \boldsymbol{m}) - (b-a)\cdot \mathbb{1}\{j=\ell\}| \le \varepsilon_2, \quad j,\ell \in [k]$$

and

$$||\boldsymbol{\sigma}'_{i_j} - \widetilde{\boldsymbol{\sigma}}_{i_j}||_2^2 \le \eta N, \quad j \in [k].$$

Here $\varepsilon_2 = o_{\varepsilon_1,\eta}(1)$ tends to 0 as ε_1 and η tend to 0. Using Gram-Schmidt orthonormalization inside the affine subspace $B(\boldsymbol{m}, 0)$, for $\varepsilon_3 = o_{\varepsilon_2}(1)$ we may find $\widehat{\boldsymbol{\sigma}}_1, \dots, \widehat{\boldsymbol{\sigma}}_k \in B(\boldsymbol{m}, [a,b]) \cap W$ satisfying

$$R(\widehat{\boldsymbol{\sigma}}_i - \boldsymbol{m}, \widehat{\boldsymbol{\sigma}}_j - \boldsymbol{m}) = (b-a)\cdot \mathbb{1}\{i=j\} \quad \forall 1 \le i < j \le k$$

and

$$||\widehat{\boldsymbol{\sigma}}_j - \boldsymbol{\sigma}'_{i_j}||_2^2 \le \varepsilon_3 N \quad \forall j \in [k].$$

Assuming H_N is $C_1\sqrt{N}$-Lipschitz with respect to the $\|\cdot\|_2$ norm (recall Proposition 3.0.1), this implies based on (3.6.5) that for some $\varepsilon_4 = o_{\varepsilon_3}(C_1+1)$,

$$\widehat{H}_N(\widehat{\boldsymbol{\sigma}}_j) \ge N(\mathsf{OPT}_\xi^{\mathrm{Sp}} - \varepsilon_4) \quad \forall i \in [k]$$

and

$$||\widehat{\boldsymbol{\sigma}}_j - \widetilde{\boldsymbol{\sigma}}_{i_j}||_2^2 \le 2(\varepsilon_3 + \eta)N \quad \forall j \in [k].$$

Recalling Proposition 3.6.1, this completes the proof. □

3.6.2 Trees and Ultrametrics

We recall the well known connection between trees and ultrametric spaces. Here and throughout given a rooted tree $\mathbb{T}$ with root $r(\mathbb{T})$ we denote by $\mathrm{pa}(v)$ the parent of $v \in V(\mathbb{T})\backslash\{r(\mathbb{T})\}$

Definition 3.6.5. [BD98] A dated, rooted tree $\mathbb{T}$ with range $[a,b] \subseteq [0,1]$ is a finite tree rooted at $r(\mathbb{T}) \in V(\mathbb{T})$ together with a height function

$$|\cdot| : V(\mathbb{T}) \to [a,b]$$

satisfying the following properties.

- $|r(\mathbb{T})| = a$.
- $|v| = b$ for all leaves $v \in L(\mathbb{T})$.
- $|\mathrm{pa}(v)| < |v|$ for all $v \in V(\mathbb{T})\backslash\{r(\mathbb{T})\}$.

We say $\mathbb{T}$ is reduced if no $v \in V(\mathbb{T})$ except possibly $r(\mathbb{T})$ has exactly 1 child.

In a rooted tree, let $u \wedge v \in V(\mathbb{T})$ denote the least common ancestor of vertices u and v. To any dated rooted tree $\mathbb{T}$, we associate a metric $d_T : V(\mathbb{T}) \times V(\mathbb{T}) \to [0, \sqrt{2}]$ characterized by

$$|u \wedge v| = \frac{|u| - d_T(u,v)^2 + |v|}{2}, \qquad u, v \in V(\mathbb{T}). \tag{3.6.6}$$

When $u, v \in L(\mathbb{T})$ are leaves and $\mathbb{T}$ has range $[a,b]$, this becomes

$$|u \wedge v| = b - \frac{d_T(u,v)^2}{2}, \qquad u, v \in L(\mathbb{T}). \tag{3.6.7}$$

Crucially, observe that for $u, v \in L(\mathbb{T})$, the value $d_T(u,v)$ is a strictly decreasing function of $|u \wedge v|$. Therefore d_T defines an ultrametric on $L(\mathbb{T})$, or in fact the set of vertices at any fixed height. The specific decreasing bijection between $|u \wedge v| \in [0,1]$ and $d_T(u,v) \in [0, \sqrt{2}]$ for $u, v \in L(\mathbb{T})$ can in general be arbitrary; the one above is suited for embeddings into Euclidean space since

$$R(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) = \frac{R(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^1) - ||\boldsymbol{\sigma}^1 - \boldsymbol{\sigma}^2||_N^2 + R(\boldsymbol{\sigma}^2, \boldsymbol{\sigma}^2)}{2}, \qquad \boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2 \in \mathbb{R}^N. \tag{3.6.8}$$

The following type of result is well known and seems to be folklore.

Proposition 3.6.6. *[RTV86, Section 6],[BD98] For any finite set X, (3.6.6) defines a bijection between the following two isomorphism classes.*

1. *Dated, rooted reduced trees with range $[0,1]$ and leaf set X.*
2. *Ultrametric structures on X with diameter at most $\sqrt{2}$.*

Any dated, rooted tree can be naturally reduced by removing vertices with a single child and connecting their parent and child. Hence we will consider general dated, rooted trees to give ourselves more flexibility. We are interested in embeddings of the leaves $L(\mathbb{T})$ into level sets $\{\boldsymbol{\sigma} \in \mathbb{R}^N : H_N(\boldsymbol{\sigma}) \geq (\mathsf{ALG} + \varepsilon)N\}$ which are isometries up to the scaling factor $\sqrt{N}$. It will be convenient to embed the entire vertex set $V(\mathbb{T})$.

Definition 3.6.7. A *Euclidean embedding* of a dated, rooted tree $\mathbb{T}$ to is a function $\iota : V(\mathbb{T}) \to \mathbb{R}^N$ satisfying

$$R(\iota(u), \iota(v)) = |u \wedge v| \qquad \forall u, v \in V(\mathbb{T}).$$

or equivalently (by (3.6.8)),

$$||\iota(u) - \iota(v)||_N = d_T(u,v) \qquad \forall u, v \in V(\mathbb{T}).$$

The next lemma gives an alternate characterization of Euclidean embeddings. .

Lemma 3.6.8. *$\iota : V(\mathbb{T}) \to \mathbb{R}^N$ is a Euclidean embedding if and only if the following properties hold. Below we implicitly define $|\mathrm{pa}(r(\mathbb{T}))| = 0$ and $\iota(\mathrm{pa}(r(\mathbb{T})) = 0$.*

1. *$\iota(r(\mathbb{T})) = a$.*
2. *For all $v \in V(\mathbb{T})$,*
$$||\iota(v) - \iota(\mathrm{pa}(v))||_N = |v| - |\mathrm{pa}(v)|.$$
3. *For all distinct $u, v \in V(\mathbb{T})$,*
$$R(\iota(u) - \iota(\mathrm{pa}(u)), \iota(v) - \iota(\mathrm{pa}(v))) = 0.$$

Proof. First if ι is a Euclidean embedding, then clearly the first property holds. The second holds because

$$\begin{aligned} ||\iota(v) - \iota(\mathrm{pa}(v))||_N^2 &= R(\iota(v) - \iota(\mathrm{pa}(v)), \iota(v) - \iota(\mathrm{pa}(v))) \\ &= |v \wedge v| - |v \wedge \mathrm{pa}(v)| - |\mathrm{pa}(v) \wedge v| + |\mathrm{pa}(v) \wedge \mathrm{pa}(v)| \\ &= |v| - |\mathrm{pa}(v)|. \end{aligned}$$

For the third property, we compute

$$R(\iota(u) - \iota(\mathrm{pa}(u)), \iota(v) - \iota(\mathrm{pa}(v))) = |u \wedge v| - |v \wedge \mathrm{pa}(v)| - |\mathrm{pa}(u) \wedge v| + |\mathrm{pa}(u) \wedge \mathrm{pa}(v)|.$$

Since $u \neq v$, without loss of generality suppose $v \neq (u \wedge v)$. Then v is an ancestor of neither u nor $\mathrm{pa}(u)$. The third property then follows because

$$u \wedge v = u \wedge \mathrm{pa}(v) \qquad \text{and} \qquad \mathrm{pa}(u) \wedge v = \mathrm{pa}(u) \wedge \mathrm{pa}(v).$$

In the other direction, let us assume the three properties hold and show ι is a Euclidean embedding. Consider vertices $u = u_n$ and $v = v_m$ with ancestor paths

$$(r(\mathbb{T}) = u_0, u_1, \ldots, u_{n-1}), \qquad (r(\mathbb{T}) = v_0, v_1, \ldots, v_{m-1}).$$

Suppose that $u \wedge v = u_d = v_d$, so that $u_j = v_j$ if and only if $j \leq d$. Then we expand

$$\begin{aligned} R(\iota(u), \iota(v)) &= a + \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}} R\big(\iota(u_i) - \iota(u_{i-1}), \iota(v_j) - \iota(v_{j-1})\big) \\ &= a + \sum_{1 \leq i \leq d} R\big(\iota(u_i) - \iota(u_{i-1}), \iota(u_i) - \iota(u_{i-1})\big) \\ &= a + \sum_{1 \leq i \leq d} |u_i| - |u_{i-1}| \\ &= a + |u_d| - |r(\mathbb{T})| \\ &= |u \wedge v|. \end{aligned}$$

□

Next, define the depth D rooted binary tree $\mathbb{T}_D^2$ on vertex set

$$V(\mathbb{T}_D^2) = \emptyset \cup [2] \cup [2]^2 \cup \cdots \cup [2]^D.$$

As usual, a vertex $v \in [2]^j$ is the parent of $u \in [2]^{j+1}$ if and only if v is an initial substring of u. We say the rooted tree $\mathbb{T}$ *contains* $\mathbb{T}_D^2$ if there exists an ancestry-preserving injection

$$\phi : V(\mathbb{T}_D^2) \to V(\mathbb{T})$$

(which need not preserve the root). Define the branching depth $D(\mathbb{T})$ to be the largest D such that $\mathbb{T}$ contains $\mathbb{T}_D^2$. For $v \in V(\mathbb{T})$, define $D(v) = D(\mathbb{T}_v)$ where $\mathbb{T}_v \subseteq \mathbb{T}$ is the subtree rooted at v.

Proposition 3.6.9. *For any rooted tree $\mathbb{T}$, the set*

$$V_D = \{v \in V(\mathbb{T}) : D(v) = D(\mathbb{T})\} \tag{3.6.9}$$

is a simple path graph with one endpoint at $r(\mathbb{T})$.

Proof. Let $D = D(\mathbb{T})$. Clearly V_D is closed under ancestry and contains $r(\mathbb{T})$. Suppose for sake of contradiction that V_D is not a path with $r(\mathbb{T})$ as an endpoint. Then V_D contains vertices v and w neither of which is an ancestor of the other. But if the disjoint subtrees rooted at v and w each contain $\mathbb{T}^2_D$, then $\mathbb{T}$ contains $\mathbb{T}^2_{D+1}$, a contradiction. □

We will use the following slight generalization of the main result of [Sub21]. It can be seen as the "default" embedding procedure which ensures energy $\mathsf{ALG}^{\mathrm{Sp}}_\xi$ at the leaves, while Lemma 3.6.4 gives a tool to improve over this embedding on intervals $[a, b]$ where $\xi''(t)^{-1/2}$ is convex.

Proposition 3.6.10. *For any $\varepsilon > 0$ and $k \in \mathbb{Z}^+$, there exist c and N_0 depending on ξ, ε, k such that with probability $1 - e^{-cN}$ the following holds for all $N \geq N_0$.*

For any $\boldsymbol{m}$ with $||\boldsymbol{m}||^2_N = q \leq 1$, any dated, rooted tree $\mathbb{T}$ of order $|V(\mathbb{T})| \leq k$ with range $[q, 1]$, and any linear subspace $W \subseteq \mathbb{R}^N$ with $\dim(W) \geq N - k$, there is an embedding ι of X into $W + \boldsymbol{m}$ such that

$$||\iota(u) - \iota(v)||_N = d(u, v) \quad \forall u, v \in V(\mathbb{T}) \tag{3.6.10}$$

and

$$H_N(\iota(x)) \geq H_N(\boldsymbol{m}) + N \cdot (\mathsf{ALG}^{\mathrm{Sp}}_\xi(\|\iota(u)\|^2_N - \mathsf{ALG}^{\mathrm{Sp}}_\xi(\|\boldsymbol{m}\|^2_N) - \varepsilon) \quad \forall v \in V(\mathbb{T}). \tag{3.6.11}$$

Proof. The proof is essentially contained in [Sub21, Theorem 4 and Remark 6]. The restriction to $W + \boldsymbol{m}$ has no effect on the proof whenever $k \leq o(N)$. Indeed, a GOE matrix has $\Omega_\varepsilon(N)$ eigenvalues at least $2 - \varepsilon$ with probability $1 - e^{-\Omega_\varepsilon(N^2)}$. This property implies existence of an eigenvalue at least $2 - \varepsilon$ when a GOE matrix is restricted to any subspace of dimension at least $N - \Omega_\varepsilon(N)$ by the Courant-Fisher theorem. Repeating the proof of [Sub21, Theorem 4] with this minor modification establishes the result. □

The following simple lemma is a slightly more general statement of Proposition 3.6.10. It will be used to extend partial embeddings of ancestor-closed subsets of $V(\mathbb{T})$ to all of $V(\mathbb{T})$.

Lemma 3.6.11. *For any $\varepsilon > 0$ and finite dated rooted tree $\mathbb{T}$, there exist c and N_0 depending on ξ, ε, T such that with probability $1 - e^{-cN}$ the following holds for all $N \geq N_0$.*

For any ancestor-closed subset $U \subseteq V(\mathbb{T})$, let $\iota_U : U \to \mathbb{R}^N$ be a Euclidean embedding. Then there is a Euclidean embedding $\iota : V(\mathbb{T}) \to \mathbb{R}^N$ extending ι_U such that for any $v \in V(\mathbb{T})$ with lowest U-ancestor $u \in U$,

$$H_N(\iota(v)) \geq H_N(\iota(u)) + N \cdot (\mathsf{ALG}^{\mathrm{Sp}}_\xi(|v|) - \mathsf{ALG}^{\mathrm{Sp}}_\xi(|u|) - \varepsilon). \tag{3.6.12}$$

Proof. The result follows by repeated application of Proposition 3.6.10. Indeed, $V(\mathbb{T})\backslash U$ consists of a disjoint union of subtrees $\mathbb{T}_i$ rooted at vertices $u_1, \ldots, u_k$ in U. For each $j \in [k]$, given a Euclidean embedding ι^{j-1}_U of

$$U_{j-1} = U \cup \left(\bigcup_{1 \leq i \leq j-1} \mathbb{T}_i \right),$$

we extend it to $\mathbb{T}_j$ as follows. Let

$$W_j = \mathrm{span}(\iota(u)_{u \in U_{j-1}})^\perp$$

be the orthogonal complement of the span of the already-embedded vertices. Then applying Proposition 3.6.10 with $W = W_j$ and $\boldsymbol{m} = \iota(u_j)$, we obtain a Euclidean embedding of $\mathbb{T}_j$ into $W_j + \iota(u_j)$, which joins with ι^{j-1}_U to form an embedding ι^j_U on U_j. It follows from Lemma 3.6.8 that ι^j_U is again a Euclidean embedding of U_j. Moreover (3.6.11) ensures that (3.6.12) is satisfied for each $v \in \mathbb{T}_j$. Repeating this inductively for each $j \in [k]$ completes the proof. □

3.6.3 Proof of Theorems 3.6.13 and 3.6.14

We now show that full binary trees are necessary for our results, in the sense that trees $\mathbb{T}$ not containing $\mathbb{T}_D^2$ fail as obstructions to energy $(\mathsf{ALG}_\xi^{\mathrm{Sp}} + \varepsilon_{\xi,D})N$ for some $\varepsilon_{\xi,D} > 0$ independent of $|V(\mathbb{T})|$. The main arguments are devoted to proving Lemma 3.6.12, which implies the two main theorems. Lemma 3.6.12 is proved by induction on D, and a representative case for $D = 1$ is depicted in Figure 3.1.

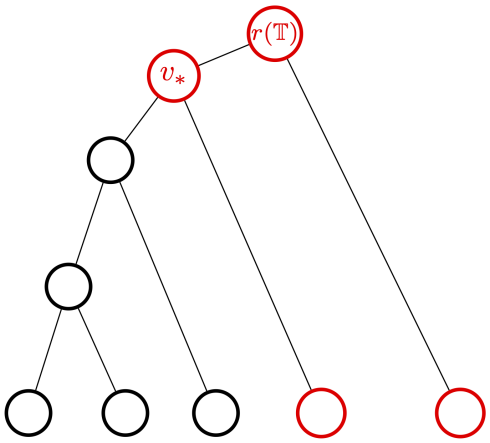


Figure 3.1: A stylized instance of Lemma 3.6.12 in the case $D = 1$ and $[a, b] = [0, 1]$ is displayed. By definition of branching depth, when $D = 1$ the non-leaves of $\mathbb{T}$ consist of a single path. We choose a vertex v_* along this path with small depth $|v_*| = a_2$, and embed v_* to have energy at least $(\mathsf{ALG}(a_2) + 2\varepsilon)N$ using Lemma 3.6.4. The leaves with parent on the segment connecting v_* to $r(\mathbb{T})$ (shown in red) can be embedded one at a time using Lemma 3.6.4. The remaining subtree under v_* is embedded all at once using Proposition 3.6.10. This results in a Euclidean embedding $\iota : V(\mathbb{T}) \to \mathbb{R}^N$ satisfying $H_N(\iota(v)) \geq (\mathsf{ALG} + \varepsilon)N$ for all $v \in L(\mathbb{T})$. For $D > 1$, we repeat this idea recursively.

In the proofs below, we will repeatedly use the principle that $\mathbb{T}$ can be subdivided by placing additional vertices on the edges of $\mathbb{T}$. This only makes constructing an Euclidean embedding more difficult. In particular, we may assume that all leaves have an ancestor of any given height. We sometimes make this explicit by considering the subgraph $\mathbb{T}_{[a,a']}$ of a tree $\mathbb{T}$ with range $[a, b]$, for $a < a' < b$. Precisely, $\mathbb{T}_{[a,a']}$ is the subgraph of vertices with heights in $[a, a']$, where we implicitly assume via subdivision that each leaf in $L(\mathbb{T})$ has ancestors at heights exactly a and a'. Note that unless $a = 0$, $\mathbb{T}_{[a,a']}$ is not in general a tree but is a disjoint union of dated rooted trees each with range $[a, a']$. We similarly define $\mathbb{T}_{[a]}$ to consist of the subset of $V(\mathbb{T})$ at heights exactly a, which without loss of generality contains exactly one ancestor of each leaf of $\mathbb{T}$.

Lemma 3.6.12. *Fix a mixture ξ, and suppose $\frac{\mathrm{d}^2}{\mathrm{d}t^2}(\xi''(t)^{-1/2}) > 0$ for $t \in [a, b] \subseteq [0, 1]$. Fix $D \in \mathbb{N}$ and sufficiently small constants $c, \varepsilon > 0$ depending only on ξ, a, b and D. Then for any $a_1 \in \left[a, \frac{a+b}{2}\right]$, any $k \in \mathbb{N}$, and any dated rooted tree $\mathbb{T}$ with range $[a, b]$, with probability $1 - O(e^{-cN})$ over the random choice of H_N, the following holds.*

For any $\boldsymbol{m} \in \sqrt{a_1} \cdot S_N$ and any linear subspace $W \subseteq \mathbb{R}^N$ with $\dim(W) \geq N - k$, there exists a Euclidean embedding

$$\iota : V(\mathbb{T}) \to W + \boldsymbol{m}$$

with $\iota(r(\mathbb{T})) = \boldsymbol{m}$ such that for all $v \in L(\mathbb{T})$,

$$H_N(\iota(v)) \geq H_N(\boldsymbol{m}) + (\mathsf{ALG}_\xi^{\mathrm{Sp}}(b) - \mathsf{ALG}_\xi^{\mathrm{Sp}}(a_1)) + \varepsilon)N. \tag{3.6.13}$$

Proof. We proceed by induction on D.

Base Case In the base case $D = 0$, the tree $\mathbb{T}$ contains only a single leaf v. It then suffices to find a single point $\boldsymbol{\sigma} \in W + \boldsymbol{m}$ with $\|\boldsymbol{\sigma}\|_N^2 = b$ such that

$$H_N(\boldsymbol{\sigma}) \geq H_N(\boldsymbol{m}) + (\mathsf{ALG}_\xi^{\mathrm{Sp}}(b) - \mathsf{ALG}_\xi^{\mathrm{Sp}}(a') + \varepsilon)N.$$

Indeed such a $\boldsymbol{\sigma}$ exists by Lemma 3.6.4.

Inductive Step For $D \geq 1$, assume the result holds for branching depths up to $D-1$. Our strategy is to first embed the path V_D (recall (3.6.9)), and then apply the inductive hypothesis on the remainder of $\mathbb{T}$ to complete the embedding. We will assume in the remainder of the proof that H_N is $C_1\sqrt{N}$-Lipschitz with respect to the $\|\cdot\|_2$ norm for some constant C_1 as in Proposition 3.0.1.

Define $a_2, a_3 \in \left[a_1, \frac{3a+b}{4}\right]$ such that

$$\sqrt{a_3^2 - a_2^2} = \sqrt{a_2^2 - a_1^2} \leq \frac{\varepsilon}{4C_1}.$$

Let $t = \max_{v \in V_D} |v|$ denote the maximum height of any $v \in V_D$. (Note that t is not affected by adding extraneous vertices to $\mathbb{T}$.)

Case 1: $t \leq a_2$ Let $v_t \in V_D$ satisfy $|v_t| = t$. Take

$$\iota : V_D :\to W + \boldsymbol{m}$$

to be an arbitrary Euclidean embedding with codomain $W + \boldsymbol{m}$. Without loss of generality, we may assume that the children of each $v \in V_D$ have height at most a_3. Next, extend ι to a still arbitrary Euclidean embedding on Q_D, which consists of V_D together with all children of vertices in V_D.

For each vertex $v \in Q_D$, the Lipschitz property implies

$$\begin{aligned} H_N(\iota(v)) &\geq H_N(\boldsymbol{m}) - C_1\sqrt{a_3^2 - a_1^2}N \\ &\geq H_N(\boldsymbol{m}) - \varepsilon_1 N. \end{aligned}$$

Observe that any $v \in Q_D \backslash V_D$ satisfies $D(v) \leq D - 1$. Because of this, we can now apply the inductive hypothesis to each subtree $\mathbb{T}_v$ rooted at some $v \in Q_D \backslash V_D$ in an arbitrary order over the v's. More precisely, suppose a Euclidean embedding mapping a subset $U \subseteq V(\mathbb{T})$ to $W + \boldsymbol{m}$ is given, and that U contains no strict descendants of $v \in Q_D \backslash V_D$. Then we know that $|v| \leq a_3 \leq \frac{3a+b}{4}$. Define the affine subspace

$$W_v = \mathrm{span}(\iota(u)_{u \in U})^\perp.$$

Then by the inductive hypothesis (using the same values a, b), there exists ε_2 depending only on $\xi, a, b, D-1$ such that ι extends to a Euclidean embedding

$$\iota : V \cup \mathbb{T}_v \to W + \boldsymbol{m}$$

such that $\iota(u) \in W_v$ for all $u \in \mathbb{T}_v$, and which satisfies

$$H_N(\iota(u)) \geq H_N(\iota(v)) + \left(\mathsf{ALG}_\xi^{\mathrm{Sp}}(b) - \mathsf{ALG}_\xi^{\mathrm{Sp}}(|v| + \varepsilon_2)\right) N, \qquad \forall u \in L(\mathbb{T}_v).$$

In particular, the above procedure can be repeated for each v, resulting in an embedding ι defined on all of $V(\mathbb{T})$. Finally for any $u \in L(\mathbb{T})$, we must have $u \in L(\mathbb{T}_v)$ for some v as above, and so

$$\begin{aligned} H_N(\iota(u)) &\geq H_N(\iota(v)) + (\mathsf{ALG}_\xi^{\mathrm{Sp}}(b) - \mathsf{ALG}_\xi^{\mathrm{Sp}}(|v| + \varepsilon_2))N \\ &\geq H_N(\boldsymbol{m}) + (\mathsf{ALG}_\xi^{\mathrm{Sp}}(b) - \mathsf{ALG}_\xi^{\mathrm{Sp}}(|v|) + \varepsilon_2 - \varepsilon_1)N \\ &\geq H_N(\boldsymbol{m}) + (\mathsf{ALG}_\xi^{\mathrm{Sp}}(b) - \mathsf{ALG}_\xi^{\mathrm{Sp}}(a_3) + \varepsilon_1)N + \left(\varepsilon_2 - 2\varepsilon_1 - \xi'(1)\sqrt{a_3^2 - a_1^2}\right) N. \end{aligned}$$

Note that

$$2\varepsilon_1 + \xi'(1)\sqrt{a_3^2 - a_1^2} \le \varepsilon_1 \cdot \left(2 + \frac{\xi'(1)}{C_1}\right).$$

Since ε_2 depended only on ξ, a, b, D and ε_1 was chosen sufficiently small depending on the same parameters, we may assume that

$$\varepsilon_2 - 2\varepsilon_1 - \xi'(1)\sqrt{a_3^2 - a_1^2} \ge 0.$$

Choosing $\varepsilon = \varepsilon_1$ finishes Case 1 of the inductive step.

Case 2: $t \ge a_2$ Let $v_* \in V(\mathbb{T})$ denote the unique vertex on V_D at height a_2 – such a v_* exists without loss of generality. Then applying Lemma 3.6.4 on $\mathbb{T}_{[a_1,a_2]}$, it follows that there exists $\boldsymbol{\sigma} \in W + \boldsymbol{m}$ with $||\boldsymbol{\sigma}||_N^2 = a_2$ such that

$$H_N(\boldsymbol{\sigma}) \ge H_N(\boldsymbol{m}) + (\mathsf{ALG}_\xi^{\mathrm{Sp}}(a_2) - \mathsf{ALG}_\xi^{\mathrm{Sp}}(a_1) + \varepsilon_2)N$$

for some ε_2 depending only on ξ, a, b. Set $\iota(v_*) = \boldsymbol{\sigma}$. Next we apply Proposition 3.6.10 to the subtree $\mathbb{T}_{v_*}$ rooted at v_*, obtaining a Euclidean embedding

$$\iota : V_D \cup \mathbb{T}_{v_*} \to W + \boldsymbol{m}$$

such that

$$H_N(\iota(x)) \ge H_N(\boldsymbol{m}) + (\mathsf{ALG}_\xi^{\mathrm{Sp}}(a_2) - \mathsf{ALG}_\xi^{\mathrm{Sp}}(a_1) + \varepsilon_3)N$$

for $\varepsilon_3 = \varepsilon_2/2$. Extending to ι to the remainder of $V(\mathbb{T})$ proceeds exactly as in Case 1. □

Below we use Lemma 3.6.12 to show that to rule out energies greater than $\mathsf{ALG}^{\mathrm{Sp}}$, $\mathbb{T}$ must have large branching depth when restricted to any height interval on which $\xi''(t)^{-1/2}$ is convex.

Theorem 3.6.13. *Fix ξ and suppose $\frac{\mathrm{d}^2}{\mathrm{d}t^2}(\xi''(t)^{-1/2}) > 0$ for all $t \in [a,b] \subseteq [0,1]$. Fix $D \in \mathbb{N}$ and sufficiently small constants $c, \varepsilon > 0$ depending only on ξ, a, b and D. Then for any dated rooted tree $\mathbb{T}$ with range $[0,1]$ such that every connected component of $\mathbb{T}_{[a,b]}$ has branching depth at most D, with probability $1 - O(e^{-cN})$ over the random choice of H_N, there exists a Euclidean embedding*

$$\iota : V(\mathbb{T}) \to \mathbb{R}^N$$

such that for all $v \in L(\mathbb{T})$,

$$H_N(\iota(v)) \ge (\mathsf{ALG}_\xi^{\mathrm{Sp}}(|v|) + \varepsilon)N. \tag{3.6.14}$$

Proof of Theorem 3.6.13. We let $\varepsilon > 0$ be sufficiently small throughout the argument. By Proposition 3.6.10, there exists a Euclidean embedding $\iota : \mathbb{T}_{[0,a]} \to \mathbb{R}^N$ such that for all $v \in \mathbb{T}_{[a]}$,

$$H_N(\iota_a(v)) \ge (\mathsf{ALG}_\xi^{\mathrm{Sp}}(a) - \varepsilon)N. \tag{3.6.15}$$

Here as usual we assume without loss of generality that all leaves in $\mathbb{T}$ have an ancestor at height a. Next we extend ι to a Euclidean embedding

$$\iota : \mathbb{T}_{[0,b]} \to \mathbb{R}^N$$

such that for all $v \in V(\mathbb{T}_{[b]})$ with ancestor u at height a,

$$H_N(\iota(v)) \ge H_N(\iota(u)) + (\mathsf{ALG}_\xi^{\mathrm{Sp}}(b) - \mathsf{ALG}_\xi^{\mathrm{Sp}}(a) + 3\varepsilon)N. \tag{3.6.16}$$

In fact the existence of such an extension follows directly from Lemma 3.6.12 for ε sufficiently small. Here as before we repeatedly apply Lemma 3.6.12 to individual subtrees in $\mathbb{T}_{[a,b]}$, using the subspace W in the statement to ensure the orthogonality constraints are satisfied.

Finally extend ι to $\mathbb{T}_{[b,1]}$ using Lemma 3.6.11 on each component. The result is that for any $x \in L(\mathbb{T})$ with ancestor v at height b,

$$H_N(\iota(x)) \ge H_N(\iota(v)) + (\mathsf{ALG}_\xi^{\mathrm{Sp}}(1) - \mathsf{ALG}_\xi^{\mathrm{Sp}}(b) - \varepsilon)N. \tag{3.6.17}$$

Combining (3.6.15), (3.6.16), and (3.6.17) completes the proof. □

In the Guerra-Talagrand interpolation used for our main argument, it was only possible to directly estimate the *average* energy of the replicas instead of the minimum. In the following final result, we show that to force the average of $H_N(v)$ over the leaves $v \in L(\mathbb{T})$ to be small, it is necessary to use a tree which branches a superconstant amount in any height interval $[a,b]$ as above, *on a set of components of* $\mathbb{T}_{[a,b]}$ *ancestral to almost all leaves.*

Let us illustrate the difference between Theorems 3.6.13 and 3.6.14 by an example. Form $\mathbb{T}$ by starting with a full symmetric tree as in Proposition 3.2.2 and adding many children of the root as additional leaves. Then by construction (recall also Proposition 3.2.8), with probability $1 - e^{-\Omega(N)}$ any Euclidean embedding $\iota : \mathbb{T} \to \mathbb{R}^N$ satisfies

$$\min_{v \in L(\mathbb{T})} H_N(\iota(v)) \le (\mathsf{ALG} + \varepsilon)N$$

for $\varepsilon > 0$ as in Proposition 3.2.2 arbitrarily small given ξ. However the same is not true for the average energy. Indeed, Theorem 3.6.13 with $D = 1$ implies that the additional leaves in $\mathbb{T}$ can be embedded to each have energy at least $(\mathsf{ALG} + 2\varepsilon')N$ where $\varepsilon' > 0$ depends only on ξ. If the additional leaves form a sufficiently large fraction of $L(\mathbb{T})$, then any Euclidean extension ι to all of $\mathbb{T}$ satisfies

$$\frac{1}{|L(\mathbb{T})|} \sum_{v \in L(\mathbb{T})} H_N(\iota(v)) \ge (\mathsf{ALG} + \varepsilon')N$$

assuming $H_N \in K_N$.

Theorem 3.6.14. *Fix a mixture ξ and $\delta > 0$, and suppose $\frac{\mathrm{d}^2}{\mathrm{d}t^2}(\xi''(t)^{-1/2}) > 0$ for $t \in [a,b] \subseteq [0,1]$. Fix $D \in \mathbb{N}$ and sufficiently small constants $c, \varepsilon > 0$ depending only on ξ, a, b, D and δ. Consider any tree $\mathbb{T}$ with range $[0,1]$ and $|L(\mathbb{T})| = n$ leaves such that for at least δn of the leaves $v \in |L(\mathbb{T})|$, the subtree of $\mathbb{T}_{[a,b]}$ consisting of ancestors of v has branching depth at most D. With probability $1 - O(e^{-cN})$ over the random choice of H_N, there exists a Euclidean embedding*

$$\iota : V(\mathbb{T}) \to \mathbb{R}^N$$

such that

$$\frac{1}{L(\mathbb{T})} \sum_{v \in L(\mathbb{T})} H_N(\iota(v)) \ge (\mathsf{ALG}_\xi^{\mathrm{Sp}} + \varepsilon)N. \tag{3.6.18}$$

Proof. Take $\varepsilon_0 > 0$ sufficiently small. For $v \in L(\mathbb{T})$ and $t \in [0,1]$, let v^t denote the ancestor of v at height v. As before, Proposition 3.6.10 shows that there exists a Euclidean embedding $\iota : \mathbb{T}_{[0,a]} \to \mathbb{R}^N$ such that for all $v^a \in \mathbb{T}_{[a]}$,

$$H_N(\iota_a(v^a)) \ge (\mathsf{ALG}_\xi^{\mathrm{Sp}}(a) - \delta\varepsilon_0)N. \tag{3.6.19}$$

Let $\widetilde{\mathbb{T}}_{[a,b]} \subseteq \mathbb{T}_{[a,b]}$ consist of all connected components in $\mathbb{T}_{[a,b]}$ of branching depth at most D. Next we extend ι to a Euclidean embedding

$$\iota : \mathbb{T}_{[0,a]} \cup \widetilde{\mathbb{T}}_{[a,b]} \to \mathbb{R}^N$$

such that for all $v^b \in L(\widetilde{\mathbb{T}}_{[a,b]})$ with ancestor v^a at height a,

$$H_N(\iota(v^b)) \ge H_N(\iota(v^a)) + (\mathsf{ALG}_\xi^{\mathrm{Sp}}(b) - \mathsf{ALG}_\xi^{\mathrm{Sp}}(a) + 4\varepsilon_0)N. \tag{3.6.20}$$

Lemma 3.6.12 allows ι to extend to the remainder of $V(\mathbb{T}_{[a,b]})$ such that

$$H_N(\iota(v^b)) \ge H_N(\iota(v^a)) + (\mathsf{ALG}_\xi^{\mathrm{Sp}}(b) - \mathsf{ALG}_\xi^{\mathrm{Sp}}(a) - \delta\varepsilon_0)N. \tag{3.6.21}$$

holds for all $v \in V(\mathbb{T}_{[a,b]})$. Since at least $\delta|L(\mathbb{T})|$ leaves v satisfy $v^a \in \widetilde{\mathbb{T}}_{[a,b]}$, (3.6.20) and (3.6.21) imply

$$\frac{1}{|L(\mathbb{T})|} \sum_{v \in L(\mathbb{T})} H_N(\iota(v^a)) - H_N(\iota(v^b)) \ge (\mathsf{ALG}_\xi^{\mathrm{Sp}}(b) - \mathsf{ALG}_\xi^{\mathrm{Sp}}(a) + 3\delta\varepsilon_0)N \tag{3.6.22}$$

As before we finish by extending ι to $\mathbb{T}_{[b,1]}$, using Lemma 3.6.11 one component at a time. Then for any $v \in L(\mathbb{T})$,

$$H_N(\iota(v)) \geq H_N(\iota(v^b)) + (\mathsf{ALG}_\xi^{\mathrm{Sp}}(1) - \mathsf{ALG}_\xi^{\mathrm{Sp}}(b) - \delta\varepsilon_0)N. \tag{3.6.23}$$

By combining (3.6.19), (3.6.22) and (3.6.23), it follows that the average of $H_N(\iota(v))$ over $v \in L(\mathbb{T})$ is

$$\begin{aligned}\frac{1}{|L(\mathbb{T})|}\sum_{v\in L(\mathbb{T})}[H_N(\iota(v))] &\geq \frac{1}{|L(\mathbb{T})|}\sum_{v\in L(\mathbb{T})}\Big(H_N(\iota(v)) - H_N(\iota(v^b)) + H_N(\iota(v^b)) - H_N(\iota(v^a)) + H_N(\iota(v^a))\Big)\\ &\geq \mathsf{ALG}_\xi^{\mathrm{Sp}}(1) + \delta\varepsilon_0.\end{aligned}$$

Taking $\varepsilon = \delta\varepsilon_0$ completes the proof. □

3.7 Overlap Concentration of Standard Optimization Algorithms

In this section we prove using Gaussian concentration of measure and Kirszbraun's theorem that approximately τ-Lipschitz functions $\mathcal{A} : \mathscr{H}_N \to B_N$ are $(\lambda, e^{-c_{\lambda,\tau}N})$ overlap concentrated. We also show that common optimization algorithms such as gradient descent, AMP, and Langevin dynamics are approximately Lipschitz.

3.7.1 Overlap Concentration of Approximately Lipschitz Algorithms

Recall that we identify each Hamiltonian H_N with its disorder coefficients $(\mathbf{G}^{(p)})_{p\in 2\mathbb{N}}$, which we concatenate into an infinite vector $\boldsymbol{g} = \boldsymbol{g}(H_N)$. We can define a (possibly infinite) distance on these Hamiltonians by

$$\|H_N - H'_N\|_N = \frac{1}{\sqrt{N}}\|\boldsymbol{g}(H_N) - \boldsymbol{g}(H'_N)\|_2. \tag{3.7.1}$$

We consider algorithms $\mathcal{A} : \mathscr{H}_N \to B_N$ that are τ-Lipschitz with respect to the $\|\cdot\|_N$ norms, i.e. $\mathcal{A}$ satisfying

$$\|\mathcal{A}(H_N) - \mathcal{A}(H'_N)\|_N \leq \tau\|H_N - H'_N\|_N. \tag{3.7.2}$$

for all $H_N, H'_N \in \mathscr{H}_N$. This is the same notion of Lipschitz as in Theorem 1.3.3, though the current scaling with $\|\cdot\|_N$ norms will be more convenient for proofs.

We will show overlap concentration for the following class of algorithms that relax the Lipschitz condition to a high probability set of inputs.

Definition 3.7.1. Let $\tau, \nu > 0$. An algorithm $\mathcal{A} : \mathscr{H}_N \to B_N$ is (τ,ν)-approximately Lipschitz if there exists a τ-Lipschitz $\mathcal{A}' : \mathscr{H}_N \to B_N$ with

$$\mathbb{P}[\mathcal{A}(H_N) = \mathcal{A}'(H_N)] \geq 1 - \nu. \tag{3.7.3}$$

Proposition 3.7.2. *If $\mathcal{A} : \mathscr{H}_N \to B_N$ is τ-Lipschitz, then for all $\lambda > 0$ it is $\left(\lambda, \exp\left(-\frac{\lambda^2}{8\tau^2}N\right)\right)$ overlap concentrated.*

Proof. We write $\mathcal{A}(\boldsymbol{g})$ to mean $\mathcal{A}(H_N)$ for the Hamiltonian H_N with disorder coefficients $\boldsymbol{g} = \boldsymbol{g}(H_N)$. Let $\mathcal{A}_i(\boldsymbol{g})$ denote the i-th coordinate of $\mathcal{A}(\boldsymbol{g})$, so $\mathcal{A}(\boldsymbol{g}) = (\mathcal{A}_1(\boldsymbol{g}), \dots, \mathcal{A}_N(\boldsymbol{g}))$. Define the gradient matrix $\nabla\mathcal{A}(\boldsymbol{g}) \in \mathbb{R}^{\mathbb{N}\times N}$ by

$$\nabla\mathcal{A}(\boldsymbol{g}) = \begin{bmatrix}\nabla\mathcal{A}_1(\boldsymbol{g}) & \nabla\mathcal{A}_2(\boldsymbol{g}) & \cdots & \nabla\mathcal{A}_N(\boldsymbol{g})\end{bmatrix}.$$

Because $\mathcal{A}$ is τ-Lipschitz, we have for all $\boldsymbol{g}, \boldsymbol{g}' \in \mathbb{R}^{\mathbb{N}}$ that

$$\lambda \geq \frac{\|\mathcal{A}(\boldsymbol{g}) - \mathcal{A}(\boldsymbol{g}')\|_N}{\|\boldsymbol{g} - \boldsymbol{g}'\|_N}.$$

By taking the limit $\boldsymbol{g}' \to \boldsymbol{g}$ from the best direction, we conclude that for all $\boldsymbol{g} \in \mathbb{R}^{\mathbb{N}}$,

$$\lambda \geq s_{\max}(\nabla\mathcal{A}(\boldsymbol{g})), \tag{3.7.4}$$

where $s_{\max}$ denotes the largest singular value.

Consider any $p \in [0,1]$. We can generate p-correlated $\mathbf{g}^{(1)}, \mathbf{g}^{(2)} \in \mathbb{R}^{\mathbb{N}}$ by generating i.i.d. $\mathbf{g}^{[0]}, \mathbf{g}^{[1]}, \mathbf{g}^{[2]} \in \mathbb{R}^{\mathbb{N}}$, each with i.i.d. standard Gaussian entries, and setting, for $i = 1, 2$,

$$\mathbf{g}^{(i)} = \sqrt{p}\mathbf{g}^{[0]} + \sqrt{1-p}\mathbf{g}^{[i]}.$$

We will apply Gaussian concentration to the function

$$F(\mathbf{g}^{[0]}, \mathbf{g}^{[1]}, \mathbf{g}^{[2]}) = R\left(\mathcal{A}(\mathbf{g}^{(0)}), \mathcal{A}(\mathbf{g}^{(1)})\right),$$

which is a function of i.i.d. standard Gaussians. For each $i \in \mathbb{N}$, let $\nabla \mathcal{A}_{\cdot,i}(\boldsymbol{g})$ denote the i-th row of $\nabla \mathcal{A}(\boldsymbol{g})$, i.e.

$$\nabla \mathcal{A}_{\cdot,i}(\boldsymbol{g}) = \left[\frac{\partial \mathcal{A}_1}{\partial \boldsymbol{g}_i}(\boldsymbol{g}) \quad \frac{\partial \mathcal{A}_2}{\partial \boldsymbol{g}_i}(\boldsymbol{g}) \quad \cdots \quad \frac{\partial \mathcal{A}_N}{\partial \boldsymbol{g}_i}(\boldsymbol{g})\right].$$

We can compute that

$$\frac{\partial F}{\partial \mathbf{g}_i^{[0]}}(\mathbf{g}^{[0]}, \mathbf{g}^{[1]}, \mathbf{g}^{[2]}) = \frac{\sqrt{p}}{N}\left[\nabla \mathcal{A}_{\cdot,i}(\mathbf{g}^{(1)})\mathcal{A}(\mathbf{g}^{(2)}) + \nabla \mathcal{A}_{\cdot,i}(\mathbf{g}^{(2)})\mathcal{A}(\mathbf{g}^{(1)})\right], \tag{3.7.5}$$

$$\frac{\partial F}{\partial \mathbf{g}_i^{[1]}}(\mathbf{g}^{[0]}, \mathbf{g}^{[1]}, \mathbf{g}^{[2]}) = \frac{\sqrt{1-p}}{N}\nabla \mathcal{A}_{\cdot,i}(\mathbf{g}^{(1)})\mathcal{A}(\mathbf{g}^{(2)}), \tag{3.7.6}$$

$$\frac{\partial F}{\partial \mathbf{g}_i^{[2]}}(\mathbf{g}^{[0]}, \mathbf{g}^{[1]}, \mathbf{g}^{[2]}) = \frac{\sqrt{1-p}}{N}\nabla \mathcal{A}_{\cdot,i}(\mathbf{g}^{(2)})\mathcal{A}(\mathbf{g}^{(1)}). \tag{3.7.7}$$

By the inequality $(x+y) \le 2x^2 + 2y^2$, (3.7.5) implies

$$\frac{\partial F}{\partial \mathbf{g}_i^{[0]}}(\mathbf{g}^{[0]}, \mathbf{g}^{[1]}, \mathbf{g}^{[2]})^2 \le \frac{2p}{N^2}\left[\left(\nabla \mathcal{A}_{\cdot,i}(\mathbf{g}^{(1)})\mathcal{A}(\mathbf{g}^{(2)})\right)^2 + \left(\nabla \mathcal{A}_{\cdot,i}(\mathbf{g}^{(2)})\mathcal{A}(\mathbf{g}^{(1)})\right)^2\right].$$

Similarly, (3.7.6) and (3.7.7) imply

$$\frac{\partial F}{\partial \mathbf{g}_i^{[1]}}(\mathbf{g}^{[0]}, \mathbf{g}^{[1]}, \mathbf{g}^{[2]})^2 \le \frac{2(1-p)}{N^2}\left(\nabla \mathcal{A}_{\cdot,i}(\mathbf{g}^{(1)})\mathcal{A}(\mathbf{g}^{(2)})\right)^2,$$

$$\frac{\partial F}{\partial \mathbf{g}_i^{[2]}}(\mathbf{g}^{[0]}, \mathbf{g}^{[1]}, \mathbf{g}^{[2]})^2 \le \frac{2(1-p)}{N^2}\left(\nabla \mathcal{A}_{\cdot,i}(\mathbf{g}^{(2)})\mathcal{A}(\mathbf{g}^{(1)})\right)^2.$$

Summing over the last three inequalities and over $i \in \mathbb{N}$ gives

$$\begin{aligned}\left\|\nabla F(\mathbf{g}^{[0]}, \mathbf{g}^{[1]}, \mathbf{g}^{[2]})\right\|_2^2 &\le \frac{2}{N^2}\sum_{i\in\mathbb{N}}\left(\nabla \mathcal{A}_{\cdot,i}(\mathbf{g}^{(1)})\mathcal{A}(\mathbf{g}^{(2)})\right)^2 + \frac{2}{N^2}\sum_{i\in\mathbb{N}}\left(\nabla \mathcal{A}_{\cdot,i}(\mathbf{g}^{(2)})\mathcal{A}(\mathbf{g}^{(1)})\right)^2 \\ &= \frac{2}{N^2}\left\|\nabla \mathcal{A}(\mathbf{g}^{(1)})\mathcal{A}(\mathbf{g}^{(2)})\right\|_2^2 + \frac{2}{N^2}\left\|\nabla \mathcal{A}(\mathbf{g}^{(2)})\mathcal{A}(\mathbf{g}^{(1)})\right\|_2^2.\end{aligned}$$

Since $\mathcal{A}(\mathbf{g}^{(1)}), \mathcal{A}(\mathbf{g}^{(2)}) \in B_N$, this implies

$$\left\|\nabla F(\mathbf{g}^{[0]}, \mathbf{g}^{[1]}, \mathbf{g}^{[2]})\right\|_2^2 \le \frac{2}{N}s_{\max}\left(\nabla \mathcal{A}(\mathbf{g}^{(1)})\right)^2 + \frac{2}{N}s_{\max}\left(\nabla \mathcal{A}(\mathbf{g}^{(2)})\right)^2 \le \frac{4\tau^2}{N}$$

for all $\mathbf{g}^{[0]}, \mathbf{g}^{[1]}, \mathbf{g}^{[2]} \in \mathbb{R}^{\mathbb{N}}$. The last inequality uses (3.7.4). By Gaussian concentration,

$$\mathbb{P}\left[\left|F(\mathbf{g}^{[0]}, \mathbf{g}^{[1]}, \mathbf{g}^{[2]}) - \mathbb{E}\,F(\mathbf{g}^{[0]}, \mathbf{g}^{[1]}, \mathbf{g}^{[2]})\right| \ge \lambda\right] \le \exp\left(-\frac{\lambda^2}{8\tau^2}N\right).$$

Note that Gaussian concentration of measure applies in infinite-dimensional abstract Weiner spaces as explained just before [Led01, Theorem 2.7] regarding Equation (2.10) therein. Alternatively if one wishes to avoid infinite-dimensional Gaussian measures, it suffices to prove the present proposition for the (still τ-Lipschitz) conditional expectations

$$\mathcal{A}_p(H_N) = \mathbb{E}[\mathcal{A}(H_N)|\mathbf{G}^{(2)}, \ldots, \mathbf{G}^{(p)}]$$

and observe that $\lim_{p\to\infty}\mathcal{A}_p(H_N) = \mathcal{A}(H_N)$ holds almost surely and in L^1. □

Proposition 3.7.3. *Suppose $\mathcal{A} : \mathscr{H}_N \to B_N$ is (τ, ν)-approximately Lipschitz. Then, for any $\lambda > 0$, it is $\left(\lambda, \exp\left(-\frac{(\lambda-4\nu)_+^2}{8\tau^2}N\right) + 2\nu\right)$ overlap concentrated.*

Proof. If $\lambda \le 4\nu$ the result is trivial, so suppose $\lambda > 4\nu$. Let $\mathcal{A}'$ be such that (3.7.3) holds.

Let $p \in [0,1]$, and let $H_N^{(1)}, H_N^{(2)}$ be p-correlated. We have

$$\left|R\left(\mathcal{A}(H_N^{(1)}), \mathcal{A}(H_N^{(2)})\right) - R\left(\mathcal{A}'(H_N^{(1)}), \mathcal{A}'(H_N^{(2)})\right)\right| \le 2$$

pointwise. Furthermore, (3.7.3) implies that

$$\mathcal{A}(H_N^{(1)}) = \mathcal{A}'(H_N^{(1)}) \qquad \text{and} \qquad \mathcal{A}(H_N^{(2)}) = \mathcal{A}'(H_N^{(2)}) \tag{3.7.8}$$

with probability at least $1 - 2\nu$. So,

$$\left|\mathbb{E}\, R\left(\mathcal{A}(H_N^{(1)}), \mathcal{A}(H_N^{(2)})\right) - \mathbb{E}\, R\left(\mathcal{A}'(H_N^{(1)}), \mathcal{A}'(H_N^{(2)})\right)\right| \le 4\nu. \tag{3.7.9}$$

By Proposition 3.7.3, we have

$$\left|R\left(\mathcal{A}'(H_N^{(1)}), \mathcal{A}'(H_N^{(2)})\right) - \mathbb{E}\, R\left(\mathcal{A}'(H_N^{(1)}), \mathcal{A}'(H_N^{(2)})\right)\right| \le \lambda - 4\nu \tag{3.7.10}$$

with probability at least $1 - \exp\left(-\frac{(\lambda-4\nu)^2}{8\tau^2}N\right)$. The events (3.7.8) and (3.7.10) occur simultaneously with probability at least $1 - \exp\left(-\frac{(\lambda-4\nu)^2}{8\tau^2}N\right) - 2\nu$. On this event, (3.7.9) and (3.7.10) imply

$$\left|R\left(\mathcal{A}(H_N^{(1)}), \mathcal{A}(H_N^{(2)})\right) - \mathbb{E}\, R\left(\mathcal{A}(H_N^{(1)}), \mathcal{A}(H_N^{(2)})\right)\right| \le \lambda,$$

as desired. □

3.7.2 Standard Deterministic Optimization Algorithms are Approximately Lipschitz

Fix constants $T_0, T, k_0 \in \mathbb{N}$ and $r \in [1, \sqrt{2})$. We take as initialization a sequence $(\boldsymbol{x}^{-T_0}, \dots, \boldsymbol{x}^{-1})$ of vectors in B_N, which is independent of the Hamiltonian H_N. We consider rather general k_0-th order optimization algorithms which compute

$$\boldsymbol{x}^{t+1} = f_t\left((\boldsymbol{x}^s)_{-T_0 \le s \le t}, \left(\nabla^k H_N(\boldsymbol{x}^s)\right)_{1 \le k \le k_0, -T_0 \le s \le t}\right), \quad 0 \le t \le T-1 \tag{3.7.11}$$

and output $\boldsymbol{x}^T$. Here, $(f_0, f_1, \dots, f_{T-1})$ is a deterministic sequence of functions such that $f_0, \dots, f_{T-2}$ have codomain rB_N, f_{T-1} has codomain B_N, and these functions are all Lipschitz in the sense that there exist constants $c_0, \dots, c_{T-1} > 0$ such that

$$\begin{aligned}
&\left\| f_t\left((\boldsymbol{x}^s)_{-T_0 \le s \le t}, (A_k^s)_{1 \le k \le k_0, -T_0 \le s \le t}\right) - f_t\left((\boldsymbol{y}^s)_{-T_0 \le s \le t}, (B_k^s)_{1 \le k \le k_0, -T_0 \le s \le t}\right)\right\|_N \\
&\le c_t\left[\sum_{s=-T_0}^{t} \|\boldsymbol{x}^s - \boldsymbol{y}^s\|_N + \sum_{k=1}^{k_0}\sum_{s=-T_0}^{t} \|A_k^s - B_k^s\|_{\mathrm{op}}\right].
\end{aligned} \tag{3.7.12}$$

As we review below, the majority of standard convex optimization algorithms fall into this class. However we remark that some optimization algorithms for highly smooth and convex functions, such as Newton's method and the recent advances [GDG+19, Nes21], do not fall into this class. This is because they require inverting a Hessian matrix or solving another inverse problem each iteration.

Example 3.7.4. Projected gradient descent is of the form in (3.7.11) via

$$f_k = \rho\left(\boldsymbol{x}^k - \eta_k \nabla H_N(\boldsymbol{x}^k)\right).$$

Here ρ is the projection map onto either B_N or C_N and the learning rate parameters $(\eta_1, \dots,)$ are arbitrary constants. Other variants such as accelerated gradient descent, ISTA, and FISTA (see e.g. [Bub15]) can similarly be expressed in the form (3.7.11).

Example 3.7.5. Approximate message passing (AMP) with arbitrary Lipschitz non-linearities can be expressed in the form of (3.7.11). Given a deterministic sequence of Lipschitz functions $f_t : \mathbb{R}^{t+1} \to \mathbb{R}$ for each $t \geq 0$, the AMP iterates are given by

$$\boldsymbol{x}^{t+1} = \nabla \widetilde{H}_N(f_t(\boldsymbol{x}^0, \dots, \boldsymbol{x}^t)) - \sum_{s=1}^{t} d_{t,s} f_{s-1}(\boldsymbol{x}^0, \dots, \boldsymbol{x}^{s-1}), \tag{3.7.13}$$

$$d_{t,s} = \xi''\left(R\left(f_t(\boldsymbol{x}^0, \dots, \boldsymbol{x}^t), f_{s-1}(\boldsymbol{x}^0, \dots, \boldsymbol{x}^{s-1})\right)\right) \cdot \mathbb{E}\left[\frac{\partial f_t}{\partial X^s}(X^0, \dots, X^t)\right]. \tag{3.7.14}$$

Here $X^0 \sim p_0$ is a uniformly bounded random variable, and $\boldsymbol{x}^0$ has i.i.d. coordinates generated from the same law. The non-linearities f_t are applied entry-wise as functions $f_t : \mathbb{R}^{N \times (t+1)} \to \mathbb{R}^N$. The sequence $(X^t)_{t \geq 1}$ is an independent centered Gaussian process with covariance $Q_{t,s} = \mathbb{E}[X^t X^s]$ defined recursively by

$$Q_{t+1,s+1} = \xi'\left(\mathbb{E}\left[f_t\left(X^0, \dots, X^t\right) f_s\left(X^0, \dots, X^s\right)\right]\right), \quad t, s \geq 0. \tag{3.7.15}$$

It is not difficult to see that the iteration (3.7.13) is captured by (3.7.11), by defining the non-linearities $f_t(\boldsymbol{x}^0, \dots, \boldsymbol{x}^t)$ as additional iterates $\boldsymbol{x}^\ell$ so that their gradients can be evaluated.

Theorem 3.7.6. *For any functions $f_0, \dots, f_{T-1}$ as above and any initialization $(\boldsymbol{x}^{-T_0}, \dots, \boldsymbol{x}^{-1})$ of vectors in B_N, there exist constants τ, c such that the map $H_N \to \boldsymbol{x}^T$ defined by the iteration (3.7.11) is (τ, ν)-approximately Lipschitz with $\nu = e^{-cN}$.*

Proof. We will first show the existence of τ such that the map $H_N \to \boldsymbol{x}^T$, with domain restricted to K_N (recall Proposition 3.8.1), is τ-Lipschitz with respect to the $\|\cdot\|_N$ norms. Consider running the iteration (3.7.11) on two Hamiltonians $H_N, H_N' \in K_N$ with the same initializaton $(\boldsymbol{x}^{-T_0}, \dots, \boldsymbol{x}^{-1})$; call the respective iterates $\boldsymbol{x}^0, \dots, \boldsymbol{x}^T$ and $\boldsymbol{y}^0, \dots, \boldsymbol{y}^T$. A straightforward induction using Proposition 3.8.2 and (3.7.12) gives constants $C_0, \dots, C_T$ such that for $0 \leq t \leq T$,

$$\left\|\boldsymbol{x}^t - \boldsymbol{y}^t\right\|_N \leq C_t \left\|H_N - H_N'\right\|_N.$$

In particular, we may take $\tau = C_T$.

By Kirszbraun's theorem, there exists a τ-Lipschitz $\mathcal{A}'$ such that $\mathcal{A}(H_N) = \mathcal{A}'(H_N)$ for $H_N \in K_N$. By Proposition 3.8.1, there exists c such that $\mathbb{P}(H_N \in K_N) \geq 1 - e^{-cN}$. Therefore $\mathcal{A}$ is (τ, ν)-approximately Lipschitz for $\nu = e^{-cN}$. □

The following corollary follows immediately from Theorem 3.7.6 and Proposition 3.7.3.

Corollary 3.7.7. *For any functions $f_0, \dots, f_{T-1}$ as above and any initialization $(\boldsymbol{x}^{-T_0}, \dots, \boldsymbol{x}^{-1})$ of vectors in B_N, for every $\lambda > 0$ there exists a constant c_λ such that for sufficiently large N, the map $H_N \to \boldsymbol{x}^T$ defined by the iteration (3.7.11) is $(\lambda, e^{-c_\lambda N})$ overlap concentrated.*

3.7.3 Reflected Langevin Dynamics are Approximately Lipschitz

Here we show that a natural version of Langevin dynamics, run for bounded time, is approximately Lipschitz for almost any realization of the driving Brownian motion and hence falls into the scope of our main results. The Langevin dynamics for a Hamiltonian H_N are given by the diffusion

$$\mathrm{d}X_t = \frac{\beta}{2} \nabla H_N \, \mathrm{d}t + \mathrm{d}B_t.$$

When X_t can range over all of space, the SDE above may explode to infinity in finite time. We therefore modify the naïve dynamics above by enforcing an inward-normal reflecting boundary for the convex body $\mathcal{K} = rB_N$ or $\mathcal{K} = rC_N$. We refer the reader to [Pil14] for the relevant definitions. In short, the result is a stochastic differential equation of the form

$$\mathrm{d}X_t = \frac{\beta}{2} \nabla H_N(X_t) \, \mathrm{d}t + \mathrm{d}B_t - v_t \, \mathrm{d}\ell_t. \tag{3.7.16}$$

Here ℓ_t is non-decreasing and only increases at times when $X_t \in \partial\mathcal{K}$. Meanwhile $v_t \in \mathbb{R}^N$ is contained in the outward normal cone of $X_t \in \partial\mathcal{K}$ for all t. Note that there may be several inequivalent choices for such a reflected process; our results apply to any of these choices. The Langevin dynamics we consider consists of solving (3.7.16) for a constant time T starting from X_0 which is independent of H_N, and then projecting X_T onto B_N or C_N.

The corresponding Skorokhod problem was shown to have a Lipschitz solution for convex polyhedra such as rC_N in [DI91, Proposition 2.2]. In this case, solving (3.7.16) reduces to solving an SDE with Lipschitz coefficients as explained in [Pil14, Section 2.2]. As a result, the solutions to (3.7.16) from different starting points X_0 (but with a shared Brownian motion) can be coupled together to give a continuous stochastic flow (see [RW94, Chapter 5, Section 13]). In the case of a smooth boundary such as B_N, although the Skorokhod problem does not have a Lipschitz solution, the results of [LS84] imply the existence of a stochastic flow as explained in [Bur09].

Lemma 3.7.8. *Let X_t, Y_t solve (3.7.16) inside a convex body $\mathcal{K}$ with the same Brownian motion. Then*

$$\int_0^t \langle X_t - Y_t, v_t^X \, \mathrm{d}\ell_t^X - v_t^Y \, \mathrm{d}\ell_t^Y \rangle \geq 0.$$

Here (v_t^X, ℓ_t^X) denote the reflecting boundary terms for X_t and similarly for Y_t.

Proof. Recall that ℓ_t^X, ℓ_t^Y are increasing. Moreover $\langle X_t - Y_t, v_t^X \rangle \geq 0$ whenever $X_t \in \partial\mathcal{K}$ by the definition of the normal cone, and similarly $\langle Y_t - X_t, v_t^Y \rangle \geq 0$ whenever $Y_t \in \partial\mathcal{K}$. The result follows. □

Theorem 3.7.9. *Both variants of Langevin dynamics above define, for any initialization $X_0 \in B_N$ and for almost every path $(B_t)_{t\in[0,T]}$, a (τ, ν) approximately Lipschitz map $\mathcal{A} : \mathscr{H}_N \to B_N$ with $\tau = O_{\xi,h,T}(1)$ and $\nu \leq e^{-\Omega(N)}$.*

Proof. Fix Hamiltonians

$$H_N^X, H_N^Y \in K_N \subseteq \mathscr{H}_N$$

satisfying $\left\|H_N^X - H_N^Y\right\|_N = \Delta$. Let X_t, Y_t be the solutions to (3.7.16) driven by a shared Brownian motion with H_N^X and H_N^Y for H_N respectively, and with shared initial condition $X_0 = Y_0$. We will show that

$$\|X_T - Y_T\|_N \leq C\Delta$$

holds almost surely for some constant $C = C(\xi, h, T)$. This suffices to imply the result. (Note that $\mathcal{A}$ might not be defined on all of $\mathscr{H}_N$, but it suffices for it to be well-defined and Lipschitz on K_N.)

First observe that $X_t - Y_t$ is a finite variation process, i.e. it has no Brownian component. With ℓ^X and ℓ^Y the corresponding finite variation processes in (3.7.16), Ito's formula gives

$$\begin{aligned}\frac{1}{2}\,\mathrm{d}\,\|X_t - Y_t\|_2^2 &= \langle X_t - Y_t, \, \mathrm{d}X_t - \, \mathrm{d}Y_t \rangle \, \mathrm{d}t \\ &= \langle X_t - Y_t, -v_t^X \, \mathrm{d}\ell_t^X + v_t^Y \, \mathrm{d}\ell_t^Y \rangle \, \mathrm{d}t + \beta \langle X_t - Y_t, \nabla H_N^X(X_t) - \nabla H_N^Y(Y_t) \rangle \, \mathrm{d}t.\end{aligned}$$

Integrating and using Lemma 3.7.8, we find

$$\begin{aligned}\|X_t - Y_t\|_2^2 &\leq \int_0^t \langle X_s - Y_s, -v_s^X \, \mathrm{d}\ell_s^X + v_s^Y \, \mathrm{d}\ell_s^Y \rangle \, \mathrm{d}s + \beta \int_0^t \langle X_s - Y_s, \nabla H_N^X(X_s) - \nabla H_N^Y(Y_s) \rangle \, \mathrm{d}s \\ &\leq \beta \int_0^t \langle X_s - Y_s, \nabla H_N^X(X_s) - \nabla H_N^Y(Y_s) \rangle \, \mathrm{d}s.\end{aligned}$$

By Proposition 3.8.2 with $C = C_1'$,

$$\left\|\nabla H_N^X(X_t) - \nabla H_N^Y(Y_t)\right\|_N \leq C(\Delta + \|X_t - Y_t\|_N).$$

Using AM-GM and rescaling, we obtain for each $t \in [0, T]$ the self-bounding inequality

$$\begin{aligned}\|X_t - Y_t\|_N^2 &\le C \int_0^t \Delta \|X_s - Y_s\|_N + \|X_s - Y_s\|_N^2 \ \mathrm{d}t \\ &\le 2C \int_0^t \Delta^2 + \|X_s - Y_s\|_N^2 \ \mathrm{d}t \\ &\le 2C\Delta^2 T + 2C \int_0^t \|X_s - Y_s\|_N^2 \ \mathrm{d}t.\end{aligned}$$

Grönwall's inequality now implies $\|X_T - Y_T\|_N^2 \le 2C\Delta^2 T e^{2CT}$. This concludes the proof. □

3.8 Appendix 1: Bounds on Hamiltonian Derivatives

In this section we will prove high-probability bounds on the derivatives of H_N, including Proposition 3.0.1. We write $H_N(\boldsymbol{\sigma}) = \langle \boldsymbol{h}, \boldsymbol{\sigma} \rangle + \widetilde{H}_N(\boldsymbol{\sigma})$ for

$$\widetilde{H}_N(\boldsymbol{\sigma}) = \sum_{p \in 2\mathbb{N}} \gamma_p H_{N,p}(\boldsymbol{\sigma}),$$

where the p-tensor component is

$$H_{N,p}(\boldsymbol{\sigma}) = \frac{1}{N^{(p-1)/2}} \langle \mathbf{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p} \rangle.$$

By slight abuse of notation, we also denote $\frac{1}{N^{(p-1)/2}} \mathbf{G}^{(p)} = H_{N,p}$.

Proposition 3.8.1. *There exists universal constants $c, C > 0$ such that for all sufficiently large N,*

$$\|H_{N,p}\|_{\mathrm{op}} \le C\sqrt{p}$$

for all $p \in 2\mathbb{N}$ with probability at least $1 - e^{-cN}$

Proof. By [BASZ20, Equation B.6] with $k = p$, we have for some universal constant K and all $p \in 2\mathbb{N}$,

$$\mathbb{P}\left[\|H_{N,p}\|_{\mathrm{op}} \ge 2K\sqrt{p}\right] \le e^{-K^2 pN/2}.$$

Take $C = 2K$. The result follows by a union bound over $p \in 2\mathbb{N}$. □

Proof of Proposition 3.0.1. Let $K_N \subseteq \mathscr{H}_N$ be the set of Hamiltonians H_N satisfying the conclusion of Proposition 3.8.1. We will take

$$C_k = C \sum_{p \in 2\mathbb{N}, p \ge k} \gamma_p r^{p-k} p^{\underline{k}} \sqrt{p} + h \mathbb{1}\{k = 1\},$$

where C is given by Proposition 3.8.1 and $p^{\underline{k}} = p(p-1)\cdots(p-k+1)$ denotes the k-th falling power of p. This is finite because $r < \sqrt{2}$ and $\sum_{p \in 2\mathbb{N}} \gamma_p^2 2^p < \infty$ implies $\limsup_{p \to \infty} \frac{\gamma_{p+2}}{\gamma_p} \le \frac{1}{2}$.

If $H_N \in K_N$, for each $\boldsymbol{\sigma}^1, \dots, \boldsymbol{\sigma}^k \in S_N$ we have

$$\begin{aligned}\frac{1}{N}\left\langle \nabla^k \widetilde{H}_N(\boldsymbol{x}), \boldsymbol{\sigma}^1 \otimes \cdots \otimes \boldsymbol{\sigma}^k \right\rangle &= \sum_{p \in 2\mathbb{N}, p \ge k} \frac{\gamma_p}{N} \left\langle \nabla^k H_{N,p}(\boldsymbol{x}), \boldsymbol{\sigma}^1 \otimes \cdots \otimes \boldsymbol{\sigma}^k \right\rangle \\ &= \sum_{p \in 2\mathbb{N}, p \ge k} \frac{\gamma_p p^{\underline{k}}}{N} \left\langle H_{N,p}, \boldsymbol{x}^{\otimes(p-k)} \otimes \boldsymbol{\sigma}^1 \otimes \cdots \otimes \boldsymbol{\sigma}^k \right\rangle \\ &\le \sum_{p \in 2\mathbb{N}, p \ge k} \gamma_p r^{p-k} p^{\underline{k}} \|H_{N,p}\|_{\mathrm{op}} \\ &\le C \sum_{p \in 2\mathbb{N}, p \ge k} \gamma_p r^{p-k} p^{\underline{k}} \sqrt{p},\end{aligned}$$

by Proposition 3.8.1. Thus

$$\left\|\nabla^k \widetilde{H}_N(\boldsymbol{x})\right\|_{\mathrm{op}} \le C \sum_{p\in 2\mathbb{N}, p\ge k} \gamma_p r^{p-k} p^{\underline{k}} \sqrt{p}.$$

For $k \ge 2$, $\nabla^k H_N(\boldsymbol{x}) = \nabla^k \widetilde{H}_N(\boldsymbol{x})$, and for $k = 1$, $\|\nabla H_N(\boldsymbol{x})\|_{\mathrm{op}} \le \left\|\nabla \widetilde{H}_N(\boldsymbol{x})\right\|_{\mathrm{op}} + h$. This proves the first claim. Similarly,

$$\begin{aligned}
&\frac{1}{N}\left\langle \nabla^k H_N(\boldsymbol{x}) - \nabla^k H_N(\boldsymbol{y}), \boldsymbol{\sigma}^1 \otimes \cdots \otimes \boldsymbol{\sigma}^k \right\rangle \\
&= \sum_{p\in 2\mathbb{N}, p\ge k} \frac{\gamma_p}{N}\left\langle \nabla^k H_{N,p}(\boldsymbol{x}) - \nabla^k H_{N,p}(\boldsymbol{y}), \boldsymbol{\sigma}^1 \otimes \cdots \otimes \boldsymbol{\sigma}^k \right\rangle \\
&= \sum_{p\in 2\mathbb{N}, p\ge k} \frac{\gamma_p p^{\underline{k}}}{N}\left\langle H_{N,p}, \left(\boldsymbol{x}^{\otimes(p-k)} - \boldsymbol{y}^{\otimes(p-k)}\right) \otimes \boldsymbol{\sigma}^1 \otimes \cdots \otimes \boldsymbol{\sigma}^k \right\rangle \\
&= \sum_{p\in 2\mathbb{N}, p\ge k} \frac{\gamma_p p^{\underline{k}}}{N} \sum_{j=0}^{p-k-1} \left\langle H_{N,p}, (\boldsymbol{x}-\boldsymbol{y}) \otimes \boldsymbol{x}^{\otimes(p-k-1-j)} \otimes \boldsymbol{y}^{\otimes j} \otimes \boldsymbol{\sigma}^1 \otimes \cdots \otimes \boldsymbol{\sigma}^k \right\rangle \\
&\le \sum_{p\in 2\mathbb{N}, p\ge k} \gamma_p r^{p-k-1} p^{\underline{k}} (p-k) \|\boldsymbol{x}-\boldsymbol{y}\|_N \|H_{N,p}\|_{\mathrm{op}} \\
&\le C_{k+1} \|\boldsymbol{x}-\boldsymbol{y}\|_N ,
\end{aligned}$$

so $\left\|\nabla^k H_N(\boldsymbol{x}) - \nabla^k H_N(\boldsymbol{y})\right\|_{\mathrm{op}} \le C_{k+1}\|\boldsymbol{x}-\boldsymbol{y}\|_N$, proving the second claim. □

Proposition 3.8.2. *Fix a model (ξ, h) and a constant $r \in [1, \sqrt{2})$. Let K_N be given by Proposition 3.0.1. There exists a sequence of constants $(C'_k)_{k\ge 1}$ independent of N such that for all $H_N, H'_N \in K_N$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ with $\|\boldsymbol{x}\|_N, \|\boldsymbol{y}\|_N \le r$,*

$$\left\|\nabla^k H_N(\boldsymbol{x}) - \nabla^k H'_N(\boldsymbol{y})\right\|_{\mathrm{op}} \le C'_k\left(\|\boldsymbol{x}-\boldsymbol{y}\|_N + \|H_N - H'_N\|_N\right),$$

where $\|H_N - H'_N\|_N$ is defined by (3.7.1).

Note that when $\|H_N - H'_N\|_N$ is infinite, this proposition is vacuously true.

Proof. We have that

$$\left\|\nabla^k H_N(\boldsymbol{x}) - \nabla^k H'_N(\boldsymbol{y})\right\|_{\mathrm{op}} \le \left\|\nabla^k H_N(\boldsymbol{x}) - \nabla^k H'_N(\boldsymbol{x})\right\|_{\mathrm{op}} + \left\|\nabla^k H'_N(\boldsymbol{x}) - \nabla^k H'_N(\boldsymbol{y})\right\|_{\mathrm{op}},$$

and by (3.0.2),

$$\left\|\nabla^k H'_N(\boldsymbol{x}) - \nabla^k H'_N(\boldsymbol{y})\right\|_{\mathrm{op}} \le C_{k+1}\|\boldsymbol{x}-\boldsymbol{y}\|_N .$$

For all $\boldsymbol{\sigma}^1, \ldots, \boldsymbol{\sigma}^k \in S_N$, we have

$$\begin{aligned}
\frac{1}{N}\left\langle \nabla^k H_N(\boldsymbol{x}) - \nabla^k H'_N(\boldsymbol{x}), \boldsymbol{\sigma}^1 \otimes \cdots \otimes \boldsymbol{\sigma}^k \right\rangle &= \sum_{p\in 2\mathbb{N}, p\ge k} \frac{\gamma_p}{N}\left\langle \nabla^k H_{N,p}(\boldsymbol{x}) - \nabla^k H'_{N,p}(\boldsymbol{x}), \boldsymbol{\sigma}^1 \otimes \cdots \otimes \boldsymbol{\sigma}^k \right\rangle \\
&= \sum_{p\in 2\mathbb{N}, p\ge k} \frac{\gamma_p p^{\underline{k}}}{N}\left\langle H_{N,p} - H'_{N,p}, \boldsymbol{x}^{\otimes(p-k)} \otimes \boldsymbol{\sigma}^1 \otimes \cdots \otimes \boldsymbol{\sigma}^k \right\rangle \\
&\le \sum_{p\in 2\mathbb{N}, p\ge k} \gamma_p r^{p-k} p^{\underline{k}} \left\|H_{N,p} - H'_{N,p}\right\|_{\mathrm{op}} .
\end{aligned}$$

Moreover,

$$\begin{aligned}
\left\|H_{N,p} - H'_{N,p}\right\|_{\mathrm{op}} &= \frac{1}{N^{(p+1)/2}} \max_{\boldsymbol{\sigma}^1,\ldots,\boldsymbol{\sigma}^p \in S_N} \langle \mathbf{G}^{(p)} - \mathbf{G}'^{(p)}, \boldsymbol{\sigma}^1 \otimes \cdots \otimes \boldsymbol{\sigma}^p \rangle \\
&\le \frac{1}{\sqrt{N}} \left\|\mathbf{G}^{(p)} - \mathbf{G}'^{(p)}\right\|_2 \\
&\le \|H_N - H'_N\|_N .
\end{aligned}$$

Thus we have

$$\frac{1}{N}\left\langle \nabla^k H_N(\boldsymbol{x}) - \nabla^k H'_N(\boldsymbol{x}), \boldsymbol{\sigma}^1 \otimes \cdots \otimes \boldsymbol{\sigma}^k \right\rangle \le \sum_{p\in 2\mathbb{N}, p\ge k} \gamma_p r^{p-k} p^{\underline{k}} \cdot \|H_N - H'_N\|_N .$$

Because this holds for all $\boldsymbol{\sigma}^1, \dots, \boldsymbol{\sigma}^k \in S_N$, we have

$$\left\|\nabla^k H_N(\boldsymbol{x}) - \nabla^k H'_N(\boldsymbol{x})\right\|_{\mathrm{op}} \le \sum_{p\in 2\mathbb{N}, p\ge k} \gamma_p r^{p-k} p^{\underline{k}} \cdot \|H_N - H'_N\|_N .$$

The result follows by taking C'_k to be the larger of C_{k+1} and $\sum_{p\in 2\mathbb{N}, p\ge k} \gamma_p r^{p-k} p^{\underline{k}}$. □

3.9 Appendix 2: Explicit Formula for the Spherical Algorithmic Threshold

In this section, we will prove Proposition 3.1.6, which gives an explicit formula for $\mathsf{ALG}^{\mathrm{Sp}}_{\xi,h}$.

We first remark that the $\widehat{q}$ defined in the second case of Proposition 3.1.6 exists and is unique. Define $f(q) = q\xi''(q) - \xi'(q) = \sum_{p\in 2\mathbb{N}} p(p-2)\gamma_p^2 q^{p-1}$. If we are in the second case of the proposition, then $h^2 + \xi'(1) < \xi''(1)$, so $f(1) > h^2$. Since $f(0) = 0 \le h^2$, existence of $\widehat{q}$ follows from the Intermediate Value Theorem. Moreover, $f(1) > h^2 \ge 0$ implies $\gamma_p > 0$ for some $p > 2$, so $f(q)$ is strictly increasing for $q \in [0,1]$. This implies uniqueness.

Recall that the spherical Parisi functional P^{Sp} (3.1.2) is defined in terms of a function $B_\zeta(t) = B - \int_t^1 \xi''(q)\zeta(q)\,\mathrm{d}q$. As (B,ζ) ranges over $\mathscr{K}(\xi)$, $B_\zeta(t)$ ranges over all continuous, nondecreasing functions from $[0,1]$ to $\mathbb{R}_{>0}$. We can thus reparametrize the minimizaton (3.1.5) as one over continuous and nondecreasing $B : [0,1] \to \mathbb{R}_{>0}$. By slight abuse of notation, for continuous and nondecreasing $B : [0,1] \to \mathbb{R}_{>0}$ define

$$\mathsf{P}^{\mathrm{Sp}}(B) = \mathsf{P}^{\mathrm{Sp}}_{\xi,h}(B) = \frac{1}{2}\left[\frac{h^2}{B(0)} + \int_0^1 \left(\frac{\xi''(q)}{B(q)} + B(q)\right)\,\mathrm{d}q.\right]$$

Proof of Proposition 3.1.6. We first handle the case $h = 0$. By AM-GM,

$$\mathsf{P}^{\mathrm{Sp}}(B) = \frac{1}{2}\int_0^1 \left(\frac{\xi''(q)}{B(q)} + B(q)\right)\,\mathrm{d}q \ge \int_0^1 \xi''(q)^{1/2}\,\mathrm{d}q.$$

Equality holds when $B(q) = \xi''(q)^{1/2}$ for all $q \in [0,1]$. However, this requires $B(0) = 0$, so this objective is not attained, though approximations to this B get arbitrarily close. Thus $\mathsf{ALG}^{\mathrm{Sp}} = \int_0^1 \xi''(q)^{1/2}\,\mathrm{d}q$. We will show this $\mathsf{ALG}^{\mathrm{Sp}}$ equals the value claimed. If $\gamma_p > 0$ for some $p > 2$, then $\xi'(1) < \xi''(1)$, so we are in the second case of the proposition. Since $\widehat{q} = 0$, we are done. Otherwise, $\gamma_p = 0$ for all $p > 2$, and $\xi'(1) = \xi''(1)$. Then $\xi''(q)$ is constant, so $\mathsf{ALG}^{\mathrm{Sp}} = \xi''(1)^{1/2} = \xi'(1)^{1/2}$ as claimed.

Otherwise, $h > 0$. We extend the definition of $\widehat{q}$ to

$$\widehat{q} = \sup\left\{q \in [0,1] : h^2 + \xi'(q) \ge q\xi''(q)\right\}.$$

This gives $\widehat{q} = 1$ in the first case of the proposition, and matches the definition of $\widehat{q}$ in the second case. Note that $\widehat{q} > 0$. Define

$$\widehat{B} = \left(\frac{h^2 + \xi'(\widehat{q})}{\widehat{q}}\right)^{1/2}.$$

We will prove both cases simultaneously by showing that for any continuous and nondecreasing $B : [0,1] \to \mathbb{R}_{>0}$, we have

$$\mathsf{P}^{\mathrm{Sp}}(B) \ge \widehat{q}^{\,1/2}\left(h^2 + \xi'(\widehat{q})\right)^{1/2} + \int_{\widehat{q}}^1 \xi''(q)^{1/2}\,\mathrm{d}q,$$

with equality if and only if

$$B(q) = \begin{cases} \widehat{B} & q \le \widehat{q}, \\ \xi''(q)^{1/2} & q > \widehat{q}. \end{cases}$$

It is easy to check that this B is continuous and nondecreasing (i.e. if $\widehat{q} < 1$, then $\widehat{B} = \xi''(\widehat{q})^{1/2}$) and that it corresponds to the equality cases claimed in the proposition. By AM-GM,

$$\frac{1}{2}\int_{\widehat{q}}^{1}\left(\frac{\xi''(q)}{B(q)} + B(q)\right)\,\mathrm{d}q \geq \int_{\widehat{q}}^{1}\xi''(q)^{1/2}\,\mathrm{d}q, \tag{3.9.1}$$

with equality if and only if $B(q) = \xi''(q)^{1/2}$ on $(\widehat{q}, 1]$. Define the truncated Parisi operator

$$\mathsf{P}^{\mathrm{Sp},\widehat{q}}(B) = \frac{1}{2}\left[\frac{h^2}{B(0)} + \int_0^{\widehat{q}}\left(\frac{\xi''(q)}{B(q)} + B(q)\right)\,\mathrm{d}q\right].$$

Let $\widetilde{B} : [0, \widehat{q}] \to \mathbb{R}_{>0}$ be given by $\widetilde{B}(q) = \widehat{B}$, and note that $\mathsf{P}^{\mathrm{Sp},\widehat{q}}(\widetilde{B}) = \widehat{q}^{1/2}\left(h^2 + \xi'(\widehat{q})\right)^{1/2}$. We will show that for continuous and nondecreasing $B : [0, \widehat{q}] \to \mathbb{R}_{>0}$, we have $\mathsf{P}^{\mathrm{Sp},\widehat{q}}(B) \geq \mathsf{P}^{\mathrm{Sp},\widehat{q}}(\widetilde{B})$, with equality if and only if $B \equiv \widetilde{B}$ on $[0, \widehat{q}]$. Along with (3.9.1), this implies the conclusion. We consider two cases.

Case 1: $B(0) < \widehat{B}$**.** Define

$$\widetilde{q} = \sup\left\{q \in [0, \widehat{q}] : B(q) \leq \widehat{B}\right\}.$$

It is possible that $\widetilde{q} = \widehat{q}$. For $q \in [\widetilde{q}, \widehat{q}]$, we have $B(q) \geq \widehat{B}$, so

$$\int_{\widetilde{q}}^{\widehat{q}}\left(\frac{\xi''(q)}{B(q)} + B(q)\right) - \int_{\widetilde{q}}^{\widehat{q}}\left(\frac{\xi''(q)}{\widehat{B}} + \widehat{B}\right) = \int_{\widetilde{q}}^{\widehat{q}}\left(\frac{1}{\widehat{B}} - \frac{1}{B(q)}\right)\left(B(q)\widehat{B} - \xi''(q)\right)\,\mathrm{d}q.$$

Because

$$B(q)\widehat{B} \geq \widehat{B}^2 \geq \frac{h^2 + \xi'(\widehat{q})}{\widehat{q}} \geq \xi''(\widehat{q}) \geq \xi''(q),$$

we have

$$\int_{\widetilde{q}}^{\widehat{q}}\left(\frac{\xi''(q)}{B(q)} + B(q)\right) \geq \int_{\widetilde{q}}^{\widehat{q}}\left(\frac{\xi''(q)}{\widehat{B}} + \widehat{B}\right). \tag{3.9.2}$$

Moreover, for $q \in [0, \widetilde{q}]$, we have $B(q) \leq \widehat{B}$, so

$$\begin{aligned}
2\left(\mathsf{P}^{\mathrm{Sp},\widetilde{q}}(B) - \mathsf{P}^{\mathrm{Sp},\widetilde{q}}(\widetilde{B})\right) &= h^2\left(\frac{1}{B(0)} - \frac{1}{\widehat{B}}\right) - \int_0^{\widetilde{q}}\left(B(q)\widehat{B} - \xi''(q)\right)\left(\frac{1}{B(q)} - \frac{1}{\widehat{B}}\right)\,\mathrm{d}q \\
&\geq h^2\left(\frac{1}{B(0)} - \frac{1}{\widehat{B}}\right) - \int_0^{\widetilde{q}}\left(\widehat{B}^2 - \xi''(q)\right)\left(\frac{1}{B(q)} - \frac{1}{\widehat{B}}\right)\,\mathrm{d}q \\
&= h^2\left(\frac{1}{B(0)} - \frac{1}{\widehat{B}}\right) - \int_0^{\widetilde{q}}\left(\frac{h^2 + \xi'(\widehat{q})}{\widehat{q}} - \xi''(q)\right)\left(\frac{1}{B(q)} - \frac{1}{\widehat{B}}\right)\,\mathrm{d}q \\
&\geq h^2\left(\frac{1}{B(0)} - \frac{1}{\widehat{B}}\right) - \int_0^{\widetilde{q}}\left(\frac{h^2 + \xi'(\widehat{q})}{\widehat{q}} - \xi''(q)\right)\left(\frac{1}{B(0)} - \frac{1}{\widehat{B}}\right)\,\mathrm{d}q \\
&= \left(\frac{1}{B(0)} - \frac{1}{\widehat{B}}\right)\left[h^2 - \int_0^{\widetilde{q}}\left(\frac{h^2 + \xi'(\widehat{q})}{\widehat{q}} - \xi''(q)\right)\,\mathrm{d}q\right] \\
&\geq \left(\frac{1}{B(0)} - \frac{1}{\widehat{B}}\right)\left[h^2 - \int_0^{\widehat{q}}\left(\frac{h^2 + \xi'(\widehat{q})}{\widehat{q}} - \xi''(q)\right)\,\mathrm{d}q\right] \\
&= 0.
\end{aligned}$$

Thus $\mathsf{P}^{\mathrm{Sp},\widetilde{q}}(B) \geq \mathsf{P}^{\mathrm{Sp},\widetilde{q}}(\widetilde{B})$, with equality only if $\widetilde{q} = \widehat{q}$ and $B(q) = \widehat{B}$ for all $q \in [0, \widetilde{q}]$. Combining this with (3.9.2) gives that $\mathsf{P}^{\mathrm{Sp},\widehat{q}}(B) \geq \mathsf{P}^{\mathrm{Sp},\widehat{q}}(\widetilde{B})$, with equality only if $B \equiv \widetilde{B}$ on $[0, \widehat{q}]$.

Case 2: $B(0) \geq \widehat{B}$**.** In this case, $B(q) \geq \widehat{B}$ for all $q \in [0, \widehat{q}]$. So,

$$\begin{aligned}
2\left(\mathsf{P}^{\mathrm{Sp},\widehat{q}}(B) - \mathsf{P}^{\mathrm{Sp},\widehat{q}}(\widetilde{B})\right) &= -h^2\left(\frac{1}{\widehat{B}} - \frac{1}{B(0)}\right) + \int_0^{\widehat{q}} \left(B(q)\widehat{B} - \xi''(q)\right)\left(\frac{1}{\widehat{B}} - \frac{1}{B(q)}\right)\,\mathrm{d}q \\
&\geq -h^2\left(\frac{1}{\widehat{B}} - \frac{1}{B(0)}\right) + \int_0^{\widehat{q}} \left(\widehat{B}^2 - \xi''(q)\right)\left(\frac{1}{\widehat{B}} - \frac{1}{B(q)}\right)\,\mathrm{d}q \\
&= -h^2\left(\frac{1}{\widehat{B}} - \frac{1}{B(0)}\right) + \int_0^{\widehat{q}} \left(\frac{h^2 + \xi'(\widehat{q})}{\widehat{q}} - \xi''(q)\right)\left(\frac{1}{\widehat{B}} - \frac{1}{B(q)}\right)\,\mathrm{d}q \\
&\geq -h^2\left(\frac{1}{\widehat{B}} - \frac{1}{B(0)}\right) + \int_0^{\widehat{q}} \left(\frac{h^2 + \xi'(\widehat{q})}{\widehat{q}} - \xi''(q)\right)\left(\frac{1}{\widehat{B}} - \frac{1}{B(0)}\right)\,\mathrm{d}q \\
&= \left(\frac{1}{\widehat{B}} - \frac{1}{B(0)}\right)\left[-h^2 + \int_0^{\widehat{q}} \left(\frac{h^2 + \xi'(\widehat{q})}{\widehat{q}} - \xi''(q)\right)\,\mathrm{d}q\right] \\
&= 0.
\end{aligned}$$

For equality to hold, we must have $B(q) = \widetilde{B}$ for all $q \in [0, \widehat{q}]$, so $B \equiv \widetilde{B}$ on $[0, \widehat{q}]$. □

Chapter 4

Open Problems and Conclusion

We close with a discussion of related OGP work, future directions, and some open problems that remain.

Branching OGP, the constant factor gap, and other CSPs. An important open problem is to close the remaining constant factor gap between algorithms for random k-SAT and our hardness result. As discussed in Section 2.9, the ladder OGP stalls at a clause density lower bounded by $1.716 \cdot 2^k \log k/k$. One natural approach is to apply the branching OGP to random k-SAT. However, the analog of Guerra's interpolation for an overlap constrained system of k-SAT assignments fails, and new ideas are needed to establish the branching OGP in this setting.

More speculatively, we expect the limiting clause density for efficient algorithms to coincide with the clustering threshold, even in lower order terms. [KMRT+07, Equation 6] gives the more precise expression $\frac{2^k}{k}\left(\log k + \log\log k + 1 + O(\frac{\log\log k}{\log k})\right)$ for the clustering threshold. On the algorithmic side, we expect this clause density to be attained (in the large-radius limit) by suitable refinements of `Fix`, where the radius of the neighborhood used to make each decision grows from 3 to a large constant.

It would also be interesting to show hardness results for other random CSPs. [ACO08], the seminal paper linking clustering to algorithmic hardness, predicted that the connection between computational hardness and solution geometry holds in substantial generality for random CSPs. We believe that recent developments in multi-OGP methodology make possible similar hardness results in other CSPs. Showing a general hardness result for random CSPs would be a significant advancement of the field.

When clustering does not imply hardness. Recent work on the symmetric Ising perceptron [ALS21b, PX21] showed that clustering (in the sense that is linked to hardness in random CSPs) does not always imply hardness. In particular, at any positive constraint density in the symmetric Ising perceptron, all but an exponentially small fraction of solutions are isolated, forming clusters *of size one*, even though efficient algorithms that find a solution exist at some positive constraint densities [BS19]. The popular wisdom regarding this apparent paradox is that in the Ising perceptron, although most solutions are isolated, there is a distinguished well-connected cluster of solutions that is accessible to algorithms. This picture was confirmed at some constraint densities in [ALS21a].

Forthcoming work [GK21c] shows that at a constraint density only slightly above where efficient algorithms exist, a multi-OGP rules out stable algorithms. It is plausible that a refinement of this multi-OGP will match the best algorithms. Thus, while the rigorous connection between multi-OGP and the failure of stable algorithms still holds, the heuristic that clustering implies hardness breaks down. It would be enlightening to clarify this heuristic and identify a refined notion of clustering that does match the limits of algorithms.

Limitations of OGP for low degree hardness. Current OGP techniques to show low degree hardness only rule out quite large success probabilities. This limitation arises because these arguments use that low degree polynomials are stable, which occurs with small but nontrivial probability. In contrast, OGP techniques to show hardness for (for example) local algorithms leverage these algorithms' concentration

properties, which occur with high probability; this allows us to show these algorithms cannot succeed with even small probability. To see this difference, compare Proposition 2.3.7(a) with Proposition 2.6.4(b). It would be nice to lower the success probability that low degree hardness results rule out, perhaps by leveraging a property stronger than stability. Concentration style OGPs also allow the construction of more complex forbidden structures such as the branching OGP, which appears difficult to replicate by stability style OGPs. Allowing the use of these structures is another potential benefit of leveraging a property stronger than stability.

It would also be interesting to prove a low degree hardness result that does not exclude the interval $(-1, 1)$ in the rounding scheme (and thus, does not reference the normalization parameter γ). Such a hardness result would be based on the inherent stability of polynomial threshold functions, rather than the stability imposed by a variance condition in γ. Note that a generalization of the Gotsman-Linial conjecture [GL92] to non-binary product spaces, plugged in modularly in place of Proposition 2.5.3, would yield a version of Theorem 2.1.6 in this setting at $\delta = \exp(-CD\sqrt{N}\log N)$. One could hope to devise a different OGP argument that improves this probability.

Planted problems. Another variant of OGP has been used to study the computational hardness of estimation problems and problems with planted structure [GZ17, GZ19, GJS19, BAWZ20], see also [CM19]. This notion of OGP tracks the overlap between a single solution and the planted truth, instead of between two or more solutions. OGP occurs if the best loss attained by a solution at some overlap with the planted truth, as a function of the overlap, is nonmonotone with one minimum at high overlap and another at low overlap. If this OGP occurs, gradient descent or any local Markov chain with worst-case initialization will be unable to efficiently find the planted truth. An open problem is to extend these hardness results for planted problems from local Markov chains to arbitrary stable algorithms.

Bibliography

[ABA13] 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.

[ABAČ13] 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.

[ABE+05] 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.

[ABM04] Dimitris Achlioptas, Paul Beame, and Michael Molloy. Exponential bounds for DPLL below the satisfiability threshold. In *Proceedings of 15th SODA*, pages 139–140, 2004.

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

[AC17a] Antonio Auffinger and Wei-Kuo Chen. On the energy landscape of spherical spin glasses. *Advances in Mathematics*, 330, 02 2017.

[AC17b] 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.

[Ach09] Dimitris Achlioptas. *Random satisfiability*, volume 185, pages 245–270. IOS Press, 2009.

[ACO08] Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *Proceedings of 49th FOCS*, pages 793–802, 2008.

[AJK+21] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence in high-dimensional expanders: Modified log-Sobolev inequalities for fractionally log-concave polynomials and the Ising model. *arXiv preprint arXiv:2106.04105*, 2021.

[ALS21a] Emmanuel Abbe, Shuangping Li, and Allan Sly. Binary perceptron: efficient algorithms can find solutions in a rare well-connected cluster. *arXiv preprint 2111.03084*, 2021.

[ALS21b] Emmanuel Abbe, Shuangping Li, and Allan Sly. Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron. *arXiv preprint 2102.13069*, 2021.

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

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

[AS00] Dimitris Achlioptas and Gregory B. Sorkin. Optimal myopic algorithms for random 3-SAT. In *Proceedings of 41st FOCS*, pages 590–600, 2000.

[BADG06] 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.

[BAGJ20] Gérard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Bounding flows for spherical spin glass dynamics. *Communications in Mathematical Physics*, 373(3):1011–1048, 2020.

[BASZ20] 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.

[BAWZ20] Gérard Ben Arous, Alexander S. Wein, and Ilias Zadik. Free energy wells and overlap gap property in sparse PCA. In *Proceedings of 33rd COLT*, pages 479–482, 2020.

[BB20] Matthew Brennan and Guy Bresler. Reducibility and statistical-computational gaps from secret leakage. In *Proceedings of 33rd COLT*, pages 648–847, 2020.

[BBH+12] 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.

[BBH+21] Matthew Brennan, Guy Bresler, Samuel B. Hopkins, Jerry Li, and Tselil Schramm. Statistical query algorithms and low-degree tests are almost equivalent. *Proceedings of 34th COLT*, page 774, 2021.

[BBK+21] Afonso S. Bandeira, Jess Banks, Dmitriy Kunisky, Cristopher Moore, and Alexander S. Wein. Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs. *Proceedings of 34th COLT*, pages 410–473, 2021.

[BCKM98] 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.

[BCN20] Charles Bordenave, Simon Coste, and Raj Rao Nadakuditi. Detection thresholds in very sparse matrix completion. *arXiv preprint arXiv:2005.06062*, 2020.

[BČNS21] 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.

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

[BGT10] Mohsen Bayati, David Gamarnik, and Prasad Tetali. Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. In *Proceedings of 42nd STOC*, pages 105–114, 2010.

[BH21] Guy Bresler and Brice Huang. The algorithmic phase transition of random k-sat for low degree polynomials. *arXiv preprint arXiv:2106.02129*, 2021.

[BKW20] Afonso S. Bandeira, Dmitriy Kunisky, and Alexander S. Wein. Computational hardness of certifying bounds on constrained PCA problems. In *Proceedings of 11th ITCS*, 2020.

[BM11] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011.

[BMZ05] Alfredo Braunstein, Marc Mézard, and Riccardo Zecchina. Survey propagation: an algorithm for satisfiability. *Random Structures & Algorithms*, 27(2):201–226, 2005.

[Bor75] Christer Borell. The Brunn-Minkowski inequality in Gauss space. *Inventiones mathematicae*, 30(2):207–216, 1975.

[BS19] Nikhil Bansal and Joel H. Spencer. On-line balancing of random inputs. *arXiv preprint arXiv:1903.06898*, 2019.

[Bub15] Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends in Machine Learning*, 8(3-4):231–357, 2015.

[Bur09] Krzysztof Burdzy. Differentiability of stochastic flow of reflected brownian motions. *Electronic Journal of Probability*, 14:2182–2240, 2009.

[CGPR19] Wei-Kuo Chen, David Gamarnik, Dmitry Panchenko, and Mustazee Rahman. Suboptimality of local algorithms for a class of max-cut problems. *Annals of Probability*, 47(3):1587–1618, 2019.

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

[Che17] Wei-Kuo Chen. Variational representations for the Parisi functional and the two-dimensional Guerra–Talagrand bound. *The Annals of Probability*, 45(6A):3929–3966, 2017.

[CHK+20] Yeshwanth Cherapanamjeri, Samuel B. Hopkins, Tarun Kathuria, Prasad Raghavendra, and Nilesh Tripuraneni. Algorithms for heavy-tailed statistics: Regression, covariance estimation, and beyond. In *Proceedings of 52nd STOC*, pages 601–609, 2020.

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

[CIS76] Boris S. Cirel'son, Ildar A. Ibragimov, and V.N. Sudakov. Norms of Gaussian sample functions. In *Proceedings of the Third Japan—USSR Symposium on Probability Theory*, pages 20–41. Springer, 1976.

[CK94] 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.

[CLR03] 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.

[CLR05] Andrea Crisanti, Luca Leuzzi, and Tommaso Rizzo. Complexity in mean-field spin-glass models: Ising p-spin. *Physical Review B*, 71(9):094202, 2005.

[CM19] Michael Celentano and Andrea Montanari. Fundamental barriers to high-dimensional regression with convex penalties. *arXiv preprint arXiv:1903.10603*, 2019.

[CO10] Amin Coja-Oghlan. A better algorithm for random k-SAT. *SIAM Journal on Computing*, 39:2823–2864, 2010.

[COE15] Amin Coja-Oghlan and Charilaos Efthymiou. On independent sets in random graphs. *Random Structures & Algorithms*, 47(3):436–486, 2015.

[COFF+09] Amin Coja-Oghlan, Uriel Feige, Alan Frieze, Michael Krivelevich, and Dan Vilenchik. On smoothed k-CNF formulas and the walksat algorithm. In *Proceedings of 20th SODA*, pages 451–460, 2009.

[COHH17] Amin Coja-Oghlan, Amir Haqshenas, and Samuel Hetterich. Walksat stalls well below the satisfiability threshold. *SIAM Journal on Discrete Mathematics*, 31:1160–1173, 2017.

[Coo71] Stephen Cook. The complexity of theorem proving procedures. In *Proceedings of 3rd STOC*, pages 151–158, 1971.

[COP16] Amin Coja-Oghlan and Konstantinos Panagiotou. The asymptotic k-SAT threshold. *Advances in Mathematics*, 288:985–1068, 2016.

[CPS18] Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. The generalized TAP free energy. arXiv:1812.05066, 2018.

[CR92] Václav Chvátal and Bruce Reed. Mick gets some (the odds are on his side). In *Proceedings of 33th FOCS*, pages 620–627, 1992.

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

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

[DEZ15] 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.

[DI91] Paul Dupuis and Hitoshi Ishii. On lipschitz continuity of the solution mapping to the skorokhod problem, with applications. *Stochastics: An International Journal of Probability and Stochastic Processes*, 35(1):31–62, 1991.

[DKWB20] Yunzi Ding, Dmitriy Kunisky, Alexander S. Wein, and Afonso S. Bandeira. Subexponential-time algorithms for sparse PCA. *arXiv preprint arXiv:1907.11635*, 2020.

[DLL61] Martin Davis, George Logemann, and Donald Loveland. A machine program for theorem proving. *Communications of the ACM*, 5(7):394–397, 1961.

[DMM09] David L. Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

[DMMZ08] Hervé Daudé, Marc Mézard, Thierry Mora, and Riccardo Zecchina. Pairs of SAT assignment in random boolean formulae. *Theoretical Computer Science*, 393:260–279, 2008.

[DP60] Martin Davis and Hilary Putnam. A computing procedure for quantification theory. *Journal of the ACM*, 7(3):201–205, 1960.

[DSS15] Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. In *Proceedings of 47th STOC*, pages 59–68, 2015.

[EKZ21] 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.

[FP83] John Franco and Marvin Paull. Probabilistic analysis of the Davis-Putnam procedure for solving the satisfiability problem. *Discrete Applied Mathematics*, 5(1):77–87, 1983.

[Fri90] Alan Frieze. On the independence number of random graphs. *Discrete Mathematics*, 81(2):171–175, 1990.

[FS96] Alan Frieze and Stephen Suen. Analysis of two simple heuristics on a random instance of k-SAT. *Journal of Algorithms*, 20:312–355, 1996.

[FS18] Charles Fefferman and Pavel Shvartsman. Sharp finiteness principles for Lipschitz selections. *Geometric and Functional Analysis*, 28(6):1641–1705, 2018.

[Fyo13] Yan V. Fyodorov. High-dimensional random fields and random matrix theory. *arXiv preprint arXiv:1307.2379*, 2013.

[Gam21] David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41), 2021.

[GDG+19] Alexander Gasnikov, Pavel Dvurechensky, Eduard Gorbunov, Evgeniya Vorontsova, Daniil Selikhanovych, César A Uribe, Bo Jiang, Haoyue Wang, Shuzhong Zhang, Sébastien Bubeck, et al. Near optimal methods for minimizing convex functions with lipschitz p-th derivatives. In *Conference on Learning Theory*, pages 1392–1393. PMLR, 2019.

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

[GJS19] David Gamarnik, Aukosh Jagannath, and Subhabrata Sen. The overlap gap property in principal submatrix recovery. *arXiv preprint arXiv:1908.09959*, 2019.

[GJW20] David Gamarnik, Aukosh Jagannath, and Alexander S. Wein. Low-degree hardness of random optimization problems. In *Proceedings of 61st FOCS*, pages 131–140, 2020.

[GJW21] David Gamarnik, Aukosh Jagannath, and Alexander S. Wein. Circuit lower bounds for the p-spin optimization problem. *arXiv preprint arXiv:2109.01342*, 2021.

[GK21a] David Gamarnik and Eren C. Kızıldağ. Algorithmic obstructions in the random number partitioning problem. *arXiv preprint arXiv:2103.01369*, 2021.

[GK21b] 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.

[GK21c] David Gamarnik and Eren C. Kızıldağ. A curious case of symmetric binary perceptron model: algorithms and barriers. Simons Institute presentation https://youtu.be/io2OXE1Xw04, October 2021.

[GL92] Craig Gotsman and Nathan Linial. The equivalence of two problems on the cube. *Journal of Combinatorial Theory, Series A*, 61(1):142–146, 1992.

[GPB82] Allen T. Goldberg, Paul W. Purdom, and Cynthia Brown. Average time analysis of simplified Davis-Putnam procedures. *Information Processing Letters*, 15:72–75, 1982.

[GS14] David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *Proceedings of 5th ITCS*, pages 369–376, 2014.

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

[GZ17] David Gamarnik and Ilias Zadik. High-dimensional regression with binary coefficients. estimating squared error and a phase transition. In *Proceedings of 30th COLT*, pages 948–953, 2017.

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

[Het16] Samuel Hetterich. Analysing Survey Propagation guided decimation on random formulas. In *Proceedings of 43rd ICALP*, 2016.

[HKP+17] Samuel B. Hopkins, Pravesh K. Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. The power of sum-of-squares for detecting hidden structures. In *Proceedings of 58th FOCS*, pages 720–731, 2017.

[Hop18] Samuel B. Hopkins. *Statistical Inference and the Sum of Squares Method*. PhD thesis, Cornell University, 2018.

[HS17] Samuel B. Hopkins and David Steurer. Efficient Bayesian estimation from few samples: community detection and related problems. In *Proceedings of 58th FOCS*, pages 379–390, 2017.

[HS21] Brice Huang and Mark Sellke. Tight Lipschitz hardness for optimizing mean field spin glasses. *arXiv preprint arXiv:2110.07847*, 2021.

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

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

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

[Kar76] Richard M. Karp. *The probabilistic analysis of some combinatorial search algorithms*, pages 1–19. Academic Press, 1976.

[Kiv21] Pax Kivimae. The ground state energy and concentration of complexity in spherical bipartite models. *arXiv preprint arXiv:2107.13138*, 2021.

[KKKS98] Lefteris M. Kirousis, Evangelos Kranakis, Danny Krizanc, and Yannis C. Stamatiou. Approximating the unsatisfiability threshold of random formulas. *Random Structures & Algorithms*, 12(3):253–269, 1998.

[KMRT+07] 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*, 104:10318–10323, 2007.

[Kut02] Samuel Kutin. Extensions to McDiarmid's inequality when differences are bounded with high probability. Technical report, University of Chicago, Department of Computer Science, 2002.

[KWB19] Dmitriy Kunisky, Alexander S. Wein, and Afonso S. Bandeira. Notes on computational hardness of hypothesis testing: predictions using the low-degree likelihood ratio. *arXiv preprint arXiv:1907.11636*, 2019.

[Led01] M. Ledoux. The concentration of measure phenomenon. In *Mathematical Surveys and Monographs*, volume 89. American Mathematical Society, Providence, RI, 2001.

[LMP15] Alessandra Lunardi, Michele Miranda, and Diego Pallara. Infinite dimensional analysis. In *19th Internet Seminar*, volume 2016, 2015.

[LMS98] Michael G. Luby, Michael Mitzenmacher, and M. Amin Shokrollahi. Analysis of random processes via and-or tree evaluation. In *Proceedings of 9th SODA*, pages 364–373, 1998.

[LS84] Pierre-Louis Lions and Alain-Sol Sznitman. Stochastic differential equations with reflecting boundary conditions. *Communications on Pure and Applied Mathematics*, 37(4):511–537, 1984.

[LZ20] Yuetian Luo and Anru R. Zhang. Tensor clustering with planted structures: statistical optimality and computational limits. *arXiv preprint arXiv:2005.10743*, 2020.

[McK21] Benjamin McKenna. Complexity of bipartite spherical spin glasses. *arXiv preprint arXiv:2105.05043*, 2021.

[Mon19] Andrea Montanari. Optimization of the Sherrington-Kirkpatrick hamiltonian. In *Proceedings of 60th FOCS*, pages 1417–1433, 2019.

[MPZ02] Marc Mézard, Giorgio Parisi, and Riccardo Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297:812–815, 2002.

[MRTS07] Andrea Montanari, Federico Ricci-Tersenghi, and Guilhem Semerjian. Solving constraint satisfaction problems through Belief Propagation-guided decimation. In *Proceedings of 45th Allerton*, pages 352–359, 2007.

[MTF90] Chao Ming-Te and John Franco. Probabilistic analysis of a generalization of the unit-clause literal selection heuristic for the k-satisfiability problem. *Information Sciences*, 51:289–314, 1990.

[Nes21] Yurii Nesterov. Superfast second-order methods for unconstrained convex optimization. *Journal of Optimization Theory and Applications*, pages 1–30, 2021.

[NSS20] Danny Nam, Allan Sly, and Youngtak Sohn. One-step replica symmetry breaking of random regular NAE-k-SAT. *arXiv preprint arXiv:2011.14270*, 2020.

[O'D14] Ryan O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, 2014.

[Pan13a] Dmitry Panchenko. The Parisi ultrametricity conjecture. *Annals of Mathematics*, pages 383–393, 2013.

[Pan13b] Dmitry Panchenko. *The Sherrington-Kirkpatrick model*. Springer Science & Business Media, 2013.

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

[Pap91] Christos H. Papadimitriou. On selecting a satisfying truth assignment. In *Proceedings of 32nd FOCS*, pages 163–169, 1991.

[Par79] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754, 1979.

[Par06] Giorgio Parisi. Computing the number of metastable states in infinite-range models. *arXiv preprint arXiv:cond-mat/0602349*, 2006.

[Pil14] Andrey Pilipenko. *An introduction to stochastic differential equations with reflection*, volume 1. Universitätsverlag Potsdam, 2014.

[PX21] Will Perkins and Changji Xu. Frozen 1-RSB structure of the symmetric Ising perceptron. *arXiv preprint arXiv:2102.05163*, 2021.

[PY95] Krzysztof Przesławski and David Yost. Lipschitz retracts, selectors, and extensions. *Michigan Mathematical Journal*, 42(3):555–571, 1995.

[RTV86] Rammal Rammal, Gérard Toulouse, and Miguel Angel Virasoro. Ultrametricity for physicists. *Reviews of Modern Physics*, 58(3):765, 1986.

[Rue87] David Ruelle. A mathematical reformulation of Derrida's REM and GREM. *Communications in Mathematical Physics*, 108(2):225–239, 1987.

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

[RW94] L. Chris G. Rogers and David Williams. *Diffusions, Markov processes and martingales: Volume 2, Itô calculus*, volume 2. Cambridge university press, 1994.

[Sel20] Mark Sellke. Approximate ground states of hypercube spin glasses are near corners. *arXiv preprint arXiv:2009.09316*, 2020.

[Sel21] Mark Sellke. Optimizing mean field spin glasses with external field. *arXiv preprint arXiv:2105.03506*, 2021.

[Shv84] Pavel Shvartsman. Lipshitz selections of multivalued mappings and traces of the Zygmund class of functions to an arbitrary compact, dokl. acad. nauk sssr 276 (1984), 559-562. In *English transl. in Soviet Math. Dokl*, volume 29, pages 565–568, 1984.

[Shv02] Pavel Shvartsman. Lipschitz selections of set-valued mappings and Helly's theorem. *The Journal of Geometric Analysis*, 12(2):289–324, 2002.

[SK75] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35(6), 1975.

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

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

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

[SW20] Tselil Schramm and Alexander S. Wein. Computational barriers to estimation from low-degree polynomials. *arXiv preprint arXiv:2008.02269*, 2020.

[SZ21] Eliran Subag and Ofer Zeitouni. Concentration of the complexity of spherical pure p-spin models at arbitrary energies. *arXiv preprint arXiv:2109.03163*, 2021.

[Tal06a] Michel Talagrand. Free energy of the spherical mean field model. *Probability Theory and Related Fields*, 134:339–382, 03 2006.

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

[Wei20] Alexander S. Wein. Optimal low-degree hardness of maximum independent set. *arXiv preprint arXiv:2010.06563*, 2020.

[Zei15] Ofer Zeitouni. Gaussian Fields Notes for Lectures. `https://www.wisdom.weizmann.ac.il/~zeitouni/notesGa` 2015.