Random Max-CSPs Inherit Algorithmic Hardness from Spin Glasses

Chris Jones* Kunal Marwaha† Juspreet Singh Sandhu‡ Jonathan Shi§

November 26, 2022

Abstract

We study random constraint satisfaction problems (CSPs) in the unsatisfiable regime. We relate the structure of near-optimal solutions for *any* Max-CSP to that for an associated spin glass on the hypercube, using the Guerra-Toninelli interpolation from statistical physics. The *noise stability* polynomial of the CSP's predicate is, up to a constant, the mixture polynomial of the associated spin glass. We prove two main consequences:

1. We relate the maximum fraction of constraints that can be satisfied in a random Max-CSP to the ground state energy density of the corresponding spin glass. Since the latter value can be computed with the Parisi formula [Par80, Tal06, AC17], we provide numerical values for some popular CSPs.
2. We prove that a Max-CSP possesses generalized versions of the *overlap gap property* if and only if the same holds for the corresponding spin glass. We transfer results from [HS21] to obstruct algorithms with *overlap concentration* on a large class of Max-CSPs. This immediately includes local classical and local quantum algorithms [CLSS22].

*Department of Computer Science, University of Chicago, Chicago, Illinois, USA. Supported by the National Science Foundation under Grant No. CCF-2008920. Email: csj@uchicago.edu.

†Department of Computer Science, University of Chicago, Chicago, Illinois, USA. Supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1746045. Email: kmarw@uchicago.edu.

‡School of Engineering & Applied Sciences, Harvard University, Cambridge, Massachusetts, USA. Supported by a Simons Investigator Fellowship, NSF grant DMS-2134157, NSF STAQ award PHY-1818914, DARPA grant W911NF2010021, DARPA ONISQ program award HR001120C0068 and DOE grant DE-SC0022199. Email: jus065@g.harvard.edu.

§Department of Computer Science, Bocconi University, Milan, Italy. Supported by European Research Council (ERC) award No. 834861 (SO-ReCoDi). Email: jonathan.shi@unibocconi.it.

Contents

1 Introduction

In this work, we formalize a general and deep connection between two intensely studied classes of optimization problems: constraint satisfaction problems (CSPs), studied in computer science, and spin glass models, studied in statistical physics. We demonstrate that for dense enough random CSPs, the geometric properties of the set of nearly-optimal solutions converge to those of a corresponding spin glass model. In these spin glass models, the very same geometric properties imply bounds on the average-case approximability achieved by broad classes of algorithms [AMS20, GJW20, GJ21]; these bounds are conjectured to be the best possible among all polynomial-time algorithms [Gam21, HS21]. The correspondence we establish here implies that the same lower bounds apply to average-case CSPs.

CSPs are paradigmatic computational tasks. Their study has led to foundational results in computational hardness, approximability, and optimization [Pas13]. In recent years, we have learned more about CSPs through methods inspired by statistical physics, especially when the clauses of the CSP are chosen randomly [DSS14, DSS16, DMS17, Sen18, DSS22]. By identifying the solution quality of a variable assignment with the *energy* of a configuration of particles, we can investigate "physical" properties of the CSP, such as phase transitions or solution clustering at different temperatures. Surprisingly, these physical properties can have computational consequences.

We study random CSP instances where the number of constraints is linear in the number of variables, and each constraint acts on a constant number of variables. If n is the number of variables, then $m = \alpha n$ is the number of constraints for some constant α. For large enough α, the CSP is unsatisfiable with probability $1 - o_n(1)$. We think of these as *Max-CSPs*.

In a Max-CSP, the goal is find a variable assignment that maximizes the number of satisfied constraints. How many constraints can be satisfied? How are the best assignments distributed around the hypercube? Can we find these assignments with efficient algorithms? Statistical physicists use questions like these to investigate the *solution geometry* of a problem. Our main result connects the solution geometry of a Max-CSP (with large enough α) to that of a *spin glass*. As a consequence, much of our mathematical and algorithmic understanding of spin glasses transfers to these CSPs.

A *spin glass* (or *mixed spin glass*) is a random system of n particles (variables) specified by a *mixture polynomial* $\xi(s) = \sum_{i \geq 1} c_i s^i$. In a spin glass, every i-tuple of particles interacts with an independent Gaussian weight of variance $c_i n^{(1-i)/2}$; this can be thought of as a random weighted CSP on the complete i-uniform hypergraph. We show that as the clause density of any random CSP increases ($\alpha \to \infty$), the solution space starts to resemble that of a spin glass. For example, at large enough clause density, Max-Cut on sparse random graphs qualitatively looks like the *Sherrington-Kirkpatrick model*, where $c_2 = 1, \{c_i\}_{i \neq 2} = 0$ [SK75, DMS17].

1.1 Main results

Formally, we relate the *free energy density* of a random Max-CSP instance to that of a particular spin glass. The associated spin glass is determined only by the Fourier weights of the CSP. In fact, the mixture polynomial of the spin glass is, up to a constant, the *noise stability polynomial* of the CSP:

Theorem 1.1 (Free energy density)**.** *Generate a random CSP instance $\mathcal{I}$ consisting of $\alpha \cdot n$ independent and*

uniform constraints of a predicate $f : \{\pm 1\}^k \to \{0,1\}$ *with randomly signed literals. Define the polynomial*

$$\xi(s) = \mathrm{Stab}_s(f) - \widehat{f}(\varnothing)^2 = \sum_{j=1}^{k} \|f^{=j}\|^2 s^j \,. \tag{1}$$

Generate a random spin glass instance with mixture polynomial ξ*, and let* $Z_{\mathcal{I}}(\beta)$ *and* $Z_{\mathsf{SG}_\xi}(\beta)$ *be the respective partition functions. Then, w.h.p. as* $n \to \infty$*,*

$$\frac{1}{\beta n} \log Z_{\mathcal{I}}(\beta) = \widehat{f}(\varnothing) + \lim_{n\to\infty} \frac{1}{\beta n} \log Z_{\mathsf{SG}_\xi}(\beta) + O\left(\frac{\beta^2}{\alpha^2}\right). \tag{2}$$

The asymptotic equivalence of the free energy density implies the equivalence of several properties of the solution geometry for large enough α. We show two specific implications of Theorem 1.1. The first is that with high probability, the optimal value of *any* random Max-CSP can be found with a spin glass calculation.

Corollary 1.2. (Optimal value equivalence). *Generate a random CSP instance* $\mathcal{I}$ *consisting of* $\alpha \cdot n$ *independent and uniform constraints of a predicate* $f : \{\pm 1\}^k \to \{0,1\}$ *with randomly signed literals. Let* $v_{\mathcal{I}}$ *be the maximum fraction of constraints of* $\mathcal{I}$ *that can be satisfied. Let* ξ *be defined as in Equation (1), and* $GSED(\mathsf{SG}_\xi)$ *as the ground state energy density of a random instance of the associated spin glass. Then, w.h.p. as* $n \to \infty$*,*

$$v_{\mathcal{I}} = \widehat{f}(\varnothing) + \lim_{n\to\infty} \frac{GSED(\mathsf{SG}_\xi)}{\sqrt{\alpha}} + O\left(\frac{1}{\alpha^{2/3}}\right). \tag{3}$$

This generalizes previous interpolations for Max-kXOR [DMS17, Sen18] and Max-kSAT [Pan18]. Computing the minimum value, or ground state energy, of a spin glass can famously be done using the *Parisi formula*. In Section 4, we use the Parisi formula to compute $GSED(\mathsf{SG}_\xi)$ for several common CSPs (our code is available online).

The second implication relates to algorithmic hardness.[1] A recent body of work uses the *overlap gap property (OGP)* to obstruct families of algorithms from $(1-\varepsilon)$-approximating average-case instances of optimization problems. We show that a generic version of the OGP, called the *branching OGP* [HS21], exists on a spin glass exactly when it exists on the associated Max-CSP for all sufficiently large α. As a result, the techniques that obstruct algorithms on certain spin glasses also work on the corresponding Max-CSPs:

Corollary 1.3 (Branching OGP equivalence, informal)**.** *Take any Max-CSP. Consider the associated spin glass* SG_ξ*, where* ξ *is defined as in Equation (1). Then* SG_ξ *has a branching OGP if and only if the Max-CSP has the same branching OGP for all sufficiently large* α*.*

Theorem 1.4 (Branching OGP obstructs algorithms with overlap concentration, informal)**.** *Consider an algorithm with* overlap concentration*. Then it cannot output arbitrarily good approximate solutions on almost all instances of Max-CSPs which have a branching OGP.*

Our proof of Theorem 1.4 follows the framework of Huang and Sellke [HS21], who prove the same result on spin glasses. Only a few changes are needed to transfer their result to Max-CSPs.

[1] For this, we need to interpolate the free energy density restricted to any given overlap and correlation structure.

The class of algorithms obstructed in Theorem 1.4 is as follows. Consider two correlated instances with correlation parameter $t \in [0,1]$. An algorithm is *overlap-concentrated* if for every t, with high probability the distribution of overlaps of output assignments on the two correlated instances falls within a narrow band. Many commonly-used algorithms have this property, including some stable algorithms [GJW20] and approximately Lipschitz algorithms [HS21] on spin glasses, and local classical and local quantum algorithms [CLSS22] on random Max-CSPs. Additionally, survey propagation with a constant number of message-passing rounds likely has this property [BMZ05, Gam21, BH22].

It is known that an OGP (and the stronger branching OGP) exists on spin glasses with even mixture polynomials without quadratic terms [HS21]. Combining this with the results that we have stated, we conclude the following:

Corollary 1.5 (informal)**.** *Consider a random Max-CSP with a predicate f such that the only nonzero Fourier coefficients of f have even degree $j \geq 4$. No algorithm with overlap concentration (e.g., local algorithms) can output arbitrarily good approximate solutions on almost all instances of the Max-CSP.*

Corollary 1.5 partially resolves [CLSS22, Conjecture 5.1]. We remark that a full characterization of spin glasses (and thereby CSPs) with an OGP is not yet known. For example, the Sherrington-Kirkpatrick model is strongly suspected to have no OGP, but this is not fully proven [ACZ20].

We cannot help but mention that these topological properties of the solution space (i.e. the solution geometry) may *precisely characterize* the algorithmic approximability of spin glasses and random CSPs [ACO08]. For spin glasses, the following is known:

Theorem 1.6. *For all spin glasses SG_ξ, there is a value ALG (given by an extended Parisi formula) such that:*

1. *[Mon19, AMS20, Sel21] Assume that the minimizer of the extended Parisi functional for SG_ξ exists. For all $\varepsilon > 0$, then there is an efficient algorithm that outputs a solution with value $ALG - \varepsilon$ on almost all instances.*
2. *[HS21] Assume the mixture polynomial ξ is* even*. For all $\varepsilon > 0$, there is a branching OGP with value $ALG + \varepsilon$, which therefore obstructs overlap-concentrated algorithms from achieving this value on almost all instances.*

It is likely the case that the same holds for CSPs for sufficiently large α. In this paper we prove that the lower bound transfers (Part 2). The upper bound (Part 1) is known for Max-Cut by [AMS21], and can likely be generalized to an arbitrary CSP (in a similar way that [AMS20] generalizes [Mon19] for spin glasses).

The paper is organized as follows. We provide formal definitions and additional motivation in Section 2. In Section 3, we give the main interpolation between every Max-CSP and a related spin glass. In Section 4, we show how this implies equivalence of optimal value (Corollary 1.2), and list numerical approximations to optimal values of several common Max-CSPs. In Section 5, we prove that the main interpolation implies the equivalence of OGPs (Corollary 1.3). In Section 6, we prove that an OGP obstructs overlap-concentrated algorithms on Max-CSPs (Theorem 1.4) and conclude Corollary 1.5. We close with a discussion in Section 7. Appendix A contains a technical proof about correlated sparse models.

1.2 Related work

Spin glasses. Spin glasses as a state of matter have been studied since the early 20th century. They were first considered as metallic alloys with many ground states, in which the magnetic spins are frustrated

(i.e. many nearby spins are mismatched). In the language of CSPs, spin glasses have exponentially many near-optimal solutions, in which many constraints are unsatisfied. In this work, we reserve the term "spin glass" for *mean-field* spin glass models, where "mean-field" means that all pairwise (or higher arity) interactions between particles are present.

The Sherrington-Kirkpatrick model is an early mathematical model of a spin glass [SK75]. It was solved (by deriving an explicit formula for the free energy density) by Parisi [Par80], but the solution relied on non-mathematically rigorous physical arguments. A later series of works [Gue03, Tal06, Pan13] proved that the formula is correct for all mixed spin glass models. Its numerical value was carefully approximated for the Sherrington-Kirkpatrick model [CR02] and more recently for other mixed spin glasses [AM20, MH22]. An introduction to the Parisi formula is given by Panchenko [Pan13].

Random CSPs through the lens of statistical physics. Statistical physicists have studied random instances of combinatorial optimization problems since at least the 1980s [FA86, MPV87]. A "dictionary" converting between the language of computer science and physics is provided in [CLSS22, Table 1].

Several modern works use the Guerra-Toninelli interpolation [GT04] in a similar technical way as our work; the interpolation method is by now a standard tool in spin glass theory. Dembo, Montanari, and Sen [DMS17] applied the interpolation to prove that the size of the Max-Cut and Max-Bisection in a random d-regular or Erdős-Rényi graph is related to the Sherrington-Kirkpatrick model in the same way as Corollary 1.2. The interpolation was later used (with different spin glass models) to determine the optimal value of random Max-kSAT [Pan18], Max-kXOR and Max q-cut [Sen18] instances in the unsatisfiable regime. Compared to these works, our Theorem 1.1 generalizes the CSP predicate to arbitrary mixtures of Boolean functions, but they must have random signs on the literals (e.g. Max-2XOR instead of Max-Cut). We also extend the Guerra-Toninelli interpolation (in Theorem 3.5) to transfer more properties of the solution geometry than just the optimal value; this is exactly what allows us to compare results on algorithmic hardness.

In this paper, we study the *unsatisfiable* regime, where the number of clauses of the CSP is αn for some large constant α. When the clause density α is smaller, for example near the satisfiability threshold of the CSP, the exact connection with spin glasses breaks down, and so the results are less unified. Nonetheless, statistical physics-inspired methods continue to give powerful insight into the solution structure of these CSPs [AM02, ART06, PT04, DSS14, DSS16, DSS22].

Overlap gaps. When near-optimal solutions are clustered, it becomes impossible for many algorithms to find them. From a geometric perspective, we can't "move" from one cluster to others without passing through a lower-value assignment. This general phenomenon was named the *overlap gap property (OGP)* and it was shown to obstruct local algorithms [GS14]. Further generalizations of overlaps [CGPR19, HS21] show stronger obstructions on wider classes of algorithms. The OGP and its generalizations have been used to obstruct algorithms from finding near-optimal solutions of various quantities in mixed spin glasses [GJW20, GJ21, HS21, Sel21], CSPs [CGPR19, BH22, CLSS22], sparse random graphs [GS14, RV17], and matrices [GL18, AWZ20]. See [Gam21] for a survey of overlaps and solution geometry.

Optimizing spin glasses and random Max-CSPs. Recently, [AMS20, Sel21] showed that a type of approximate message-passing algorithm finds the ground state energy of spin glasses without overlap gaps. Under the same assumption, a version of this algorithm was also shown to be optimal on Max-Cut for sparse random graphs with constant (but sufficiently large) degree [AMS21]. Both the Sherrington-Kirkpatrick model and Max-Cut on sparse random graphs are strongly suspected to have no overlap gaps [ACZ20].

There has been some study of a near-term quantum algorithm (the QAOA [FGG14]) optimizing spin glasses, with recently-proven rigorous performance bounds [CvD21, FGGZ22]. In fact, for large enough clause density, the performance of the QAOA is *identical* on a random instance of Max-kXOR and on its corresponding spin glass [BM21, BFM+22, BGMZ22].

2 Preliminaries

2.1 Random constraint satisfaction problems (CSPs)

A CSP instance, denoted $\mathcal{I}$, consists of n variables and a set of constraints, denoted $E(\mathcal{I})$. In a random CSP instance, the constraints $E(\mathcal{I})$ are drawn from a distribution Λ on functions $f : \Sigma^k \to \mathbb{R}$.

Definition 2.1 (Instance of a random CSP)**.** *Let Λ be a distribution on functions $f : \Sigma^k \to \mathbb{R}$ with alphabet $\Sigma = \{\pm 1\}$. Fix a constant $\alpha > 0$. Then, a random CSP over n variables $\{\sigma_i\}_{i=1}^n$ and $m = \alpha n$ clauses is generated by: for each $i = 1, \dots, m$, draw $i_1, \dots, i_k$ uniformly i.i.d from $[n]$, draw $f \sim \Lambda$, draw k random signs ε_i uniformly i.i.d from $\{\pm 1\}$, and add the constraint e to $E(\mathcal{I})$, describing the clause $f_e(\sigma_e) := f(\varepsilon_1 \sigma_{i_1}, \dots, \varepsilon_k \sigma_{i_k})$.*

Remark 2.2. *The canonical case is to take Λ which is supported on a single predicate $f : \Sigma^k \to \{0, 1\}$. For example, the OR predicate corresponds to kSAT. Our proofs apply to the more general setting of Definition 2.1, but it does not apply to $|\Sigma| > 2$ or instances without random signs.*

We denote this random model as $\mathsf{CSP}_\Lambda(\alpha)$. We study the *unsatisfiable* regime; for example, one can think of $\alpha \gg \exp(k)$.

In this work, we use the language of *Hamiltonians*. In other settings this quantity may be called the objective function, the value, or the score of the assignment.

Definition 2.3 (CSP Hamiltonian)**.** *Consider a $\mathsf{CSP}_\Lambda(\alpha)$ instance $\mathcal{I}$ on $[n]$. For any input $\sigma \in \{\pm 1\}^n$ let*

$$H^\alpha(\sigma) = \frac{1}{\alpha} \sum_{e \in E(\mathcal{I})} f_e(\sigma_e) \,. \tag{4}$$

Remark 2.4. *We divide by α so that, regardless of the value of α, the value of the Hamiltonian is in the same interval $[(\min f) \cdot n, (\max f) \cdot n]$.*

Definition 2.5 (Optimal value of a CSP instance)**.** *Define the maximum or optimal value of a $\mathsf{CSP}_\Lambda(\alpha)$ instance $\mathcal{I}$ by:*

$$v_{\mathcal{I}} = \frac{1}{n} \max_{\sigma \in \{\pm 1\}^n} H^\alpha(\sigma) \,. \tag{5}$$

When the predicates in the CSP are 0/1-valued, the maximum value of a CSP instance $\mathcal{I}$ is the maximum possible fraction of constraints that can be satisfied.

Remark 2.6. *By a union bound, with high probability all assignments to a random CSP with predicate f satisfy $\widehat{f}(\varnothing) + O(\frac{1}{\sqrt{\alpha}})$ fraction of constraints. The purpose of our work is to precisely study the behavior of the $\frac{1}{\sqrt{\alpha}}$ term.*

2.2 Mean-field spin glasses

Let n denote the number of particles or variables in the system. We introduce the mathematical objects describing a spin glass instance:

Definition 2.7 (Gaussian tensor)**.** *$J^{(p)}$ denotes an order-p Gaussian tensor, where each axis has length n. Namely, $J^{(p)}[i_1, \dots, i_p] \sim \mathcal{N}(0,1)$ i.i.d.*[2]

Definition 2.8 (Mixture polynomial)**.** *Call $\xi(s) = \sum_{p \geq 1}^{k} c_p^2 s^p$ a* mixture polynomial.

We now have enough to define a spin glass. The mixture polynomial determines the random model, and the Gaussian tensor determines the instance of the spin glass. For example, $\xi(s) = s^2$ (or $\xi(s) = s^2/2$ in some works) specifies the well-studied *Sherrington-Kirkpatrick spin glass model*.

Definition 2.9 (Finite mixed spin glass Hamiltonian)**.** *Choose $J := (J^{(p)} : p = 1, 2, \dots)$, where the $J^{(p)}$ are independent. Choose also a mixture polynomial ξ. For any $\sigma \in \{\pm 1\}^n$, define the Hamiltonian*

$$H^{\xi}(\sigma) = \sum_{p=1}^{k} \frac{c_p}{n^{(p-1)/2}} \left\langle J^{(p)}, \sigma^{\otimes p} \right\rangle = \sum_{p=1}^{k} \frac{c_p}{n^{(p-1)/2}} \sum_{[i_1,\dots,i_p] \in [n]^p} J^{(p)}[i_1, \dots, i_p] \sigma_{i_1} \dots \sigma_{i_p} \,. \tag{6}$$

We use the variable H^{ξ} to identify the instance.

Definition 2.10 (Instance of a spin glass)**.** *A* spin glass *is the collection of Hamiltonian values $(H^{\xi}(\sigma))_{\sigma \in \{\pm 1\}^n}$ for a given sample J and mixture polynomial ξ. We say that this spin glass is an* instance *of the random model SG_{ξ}.*[3]

An early motivator of spin glass theory was the conjecture (and eventual proof) of the following representation for the average-case ground state energy:

Theorem 2.11 ([Par80, Gue03, Tal06, Pan14])**.** *Fix any spin glass model SG_{ξ}. Sample a sequence of independent instances $\{H^{\xi}_{n=1}, H^{\xi}_{n=2}, \dots\}$ from SG_{ξ} for increasing values of n. Then the limit*

$$GSED(\mathsf{SG}_{\xi}) := \lim_{n \to \infty} \max_{\sigma \in \{\pm 1\}^n} \frac{1}{n} H^{\xi}_n(\sigma) \,, \tag{7}$$

almost surely exists. Furthermore,

$$GSED(\mathsf{SG}_{\xi}) \overset{a.s.}{=} \lim_{\beta \to \infty} \frac{1}{\beta} \mathcal{P}^{\xi}(\beta, 0) \,, \tag{8}$$

where $\mathcal{P}^{\xi}(\beta, h)$ is the Parisi functional[4] *with mixture polynomial ξ, inverse temperature β, and external field strength h.*

We state one other fact about spin glasses. One way to characterize a spin glass is as a Gaussian process on the state space $\{\pm 1\}^n$ with covariance structure given by the mixture polynomial ξ. The covariance of the Gaussians $H^{\xi}(\sigma_1), H^{\xi}(\sigma_2)$ is a function of the normalized inner product, or *overlap*, of σ_1 and σ_2.

[2] We follow the definition in [Pan13, Chapter 2]. Some definitions of spin glasses use a *symmetric* Gaussian tensor, which is equivalent up to scaling of the c_k and $o_n(1)$ change in the free energy density. This is because the models only differ on tensor entries $[i_1, \dots, i_p]$ such that $i_1, \dots, i_p$ are not all distinct, which only make up $o_n(1)$ fraction of all entries of the tensor.

[3] Spin glass models do not always include a $p = 1$ term, instead including a *fixed* external field. We use the $p = 1$ term to interpolate to CSPs that have nonzero first Fourier weight.

[4] A variational formula for the Parisi functional is given in Definition 4.2.

Fact 2.12 (Gaussian characterization of a spin glass). *Let $\{H^\xi(\sigma)\}_{\sigma\in\{\pm1\}^n}$ be a spin glass instance of the random model SG_ξ. Then these variables are jointly Gaussian, with mean zero and covariance*

$$\mathbb{E}\left[H^\xi(\sigma_1)H^\xi(\sigma_2)\right] = n\cdot\xi\left(\frac{\langle\sigma_1,\sigma_2\rangle}{n}\right), \tag{9}$$

where $\langle\cdot\,,\cdot\rangle$ is the standard inner product.

2.3 Solution geometry of optimization problems

Solution geometry refers to the distribution of the optimal and near-optimal solutions on the Boolean hypercube. Given a Hamiltonian $H:\{\pm1\}^n\to\mathbb{R}$, what do we mean by "near-optimal solutions"? We consider two notions. The first, used when considering computational tasks, is the set of σ such that $H(\sigma)\geq v$ for some value v. The second, a "smoother" notion from statistical physics, is that near-optimal solutions are samples from a low-temperature Gibbs distribution.

Definition 2.13 (Gibbs distribution). *Given a Hamiltonian $H:\{\pm1\}^n\to\mathbb{R}$ and an* inverse temperature *parameter $\beta\geq 0$, the* Gibbs distribution *is the distribution on $\{\pm1\}^n$ proportional to $e^{\beta H(\cdot)}$.*

Definition 2.14 (Partition function). *The* partition function *of H is the normalizing constant of the Gibbs distribution, i.e. the exponentially-weighted sum*

$$Z_H(\beta) = \sum_{\sigma\in\{\pm1\}^n} e^{\beta H(\sigma)}. \tag{10}$$

The most important geometric notion is the *overlap* of two assignments, which is exactly the normalized inner-product:

Definition 2.15 (Overlap). *The* overlap *between two assignments (configurations) $\sigma_1,\sigma_2\in\{\pm1\}^n$ is the normalized inner-product,*

$$R(\sigma_1,\sigma_2) = \frac{\langle\sigma_1,\sigma_2\rangle}{n}. \tag{11}$$

The overlap distribution is the probability density function of $R(\sigma_1,\sigma_2)$, where σ_1,σ_2 are independently sampled near-optimal solutions. This distribution is supported on $[-1,1]$ and can contain quite a lot of structure. We are interested mostly in "gaps" in the support of the overlap distribution (i.e. *overlap gaps*), which imply algorithmic barriers against some types of algorithms.

A recent body of work [CGPR19, HS21] generalizes *overlaps* to *multi-overlaps*, describing the overlaps between more than two samples from a Gibbs distribution. Some of these generalizations imply suboptimality for larger classes of algorithms and with a stronger gap in the approximation parameter. The generalization of overlap used in this paper is the *I-overlap*; given a set $\boldsymbol{\sigma}$ of $|\boldsymbol{\sigma}|=\ell$ assignments, each in $\{\pm1\}^n$, we define an overlap value for every subset of $\boldsymbol{\sigma}$.

Definition 2.16 (I-overlap). *Consider any $\boldsymbol{\sigma}:=(\sigma_1,\dots,\sigma_\ell)\in(\{\pm1\}^n)^\ell$ and any $I\subseteq[\ell]$. Define the* I-overlap *of $\boldsymbol{\sigma}$ as*

$$R_I(\boldsymbol{\sigma}) = \frac{1}{n}\sum_{j=1}^{n}\prod_{i\in I}(\sigma_i)_j. \tag{12}$$

The I-overlap recovers the definition of overlap when I has two elements.

We define an *overlap vector* to associate a set of ℓ assignments with all of its possible I-overlaps:

Definition 2.17 (Overlap vector)**.** *Consider any $\boldsymbol{\sigma} := (\sigma_1, \dots, \sigma_\ell) \in (\{\pm 1\}^n)^\ell$. Then its overlap vector $Q(\boldsymbol{\sigma}) \in [-1,1]^{2^{[\ell]}}$ lists all possible I-overlaps; that is, $Q(\boldsymbol{\sigma})_I = R_I(\boldsymbol{\sigma})$ for every $I \in 2^{[\ell]}$.*

Definition 2.18 (Overlap polytope)**.** *Consider the set of all possible overlap vectors $Q(\boldsymbol{\sigma})$ for any positive $n \in \mathbb{N}$ and sets of vectors $\boldsymbol{\sigma} \in (\{\pm 1\}^n)^\ell$. Then $\mathcal{R}^{(\ell)} \subseteq [-1,1]^{2^{[\ell]}}$ is the closure of this set. Formally,*

$$\mathcal{R}^{(\ell)} = \overline{\{Q(\boldsymbol{\sigma}) : n \in \mathbb{N}\,,\, \boldsymbol{\sigma} \in (\{\pm 1\}^n)^\ell\}}\,. \tag{13}$$

Remark 2.19. *Since $R_\varnothing(\boldsymbol{\sigma}) = 0$ for all $\boldsymbol{\sigma}$, we can ignore this coordinate, and then $\mathcal{R}^{(\ell)}$ is a non-degenerate convex polytope in $\mathbb{R}^{2^\ell - 1}$. For example, $\mathcal{R}^{(2)}$ is a regular tetrahedron in $\mathbb{R}^3$.*

Definition 2.20 (Preimage of overlap vectors)**.** *For any open subset $S \subseteq \mathcal{R}^{(\ell)}$ (in the Euclidean subset topology of $\mathcal{R}^{(\ell)}$), let $U_n^{(\ell)}(S) \subseteq (\{\pm 1\}^n)^\ell$ be the preimage of S in dimension n; i.e. the set of $\boldsymbol{\sigma} \in (\{\pm 1\}^n)^\ell$ such that $Q(\boldsymbol{\sigma}) \in S$. We drop the superscript when ℓ is in context.*

The polytope $\mathcal{R}^{(\ell)}$ captures all permutation-invariant functions of $\sigma_1, \dots, \sigma_\ell$, as we observe in the next fact.

Definition 2.21 (Permutation-invariant function)**.** *Let a permutation $\pi \in S_n$ act on $\sigma \in \{\pm 1\}^n$ as*

$$\sigma_i^\pi = \sigma_{\pi(i)}\,, \tag{14}$$

and on a function $f : \{\pm 1\}^n \to \mathbb{R}$ (extended coordinate-wise to $f : (\{\pm 1\}^n)^\ell \to \mathbb{R}$) as

$$f^\pi(\sigma) = f(\sigma^\pi)\,. \tag{15}$$

Then f is permutation-invariant if it is fixed by all π.

Fact 2.22. *Any permutation-invariant function $f : (\{\pm 1\}^n)^\ell \to \mathbb{R}$ can be expressed as a function on $\mathcal{R}^{(\ell)}$,*

$$f(\sigma_1, \dots, \sigma_\ell) = f(Q(\sigma_1, \dots, \sigma_\ell))\,. \tag{16}$$

Questions about solution geometry should likely be permutation-invariant. Thus since we are interested in permutation-invariant functions of $\sigma_1, \dots, \sigma_\ell$ when $\sigma_1, \dots, \sigma_\ell$ are independent near-optimal solutions, we should study the distribution on $\mathcal{R}^{(\ell)}$ of the overlap vector $Q(\sigma_1, \dots, \sigma_\ell)$ for this sample.[5] Our main result shows that the support of this distribution for a Max-CSP matches that of a spin glass. (We formally define gaps in the support, i.e. *overlap gaps*, in Section 5.)

The proof heavily uses the concept of *free energy*, defined as the logarithm of the partition function.

Definition 2.23 (Free energy)**.** *The* free energy *of H is* $\log Z_H(\beta)$.

The free energy at low temperature ($\beta \to \infty$) matches the optimal value of the Hamiltonian. This is a commonly-used fact (for example, [Pan13, Equation 1.7]); we include the short proof.

[5] It is likely that the distribution converges in some sense with high probability as $n \to \infty$, and then converges again as $\beta \to \infty$, but our proof does not show this nor use this.

Fact 2.24. *The maximum value of a Hamiltonian H on a finite domain Ω is related to its free energy by*

$$\max_{\sigma\in\Omega} H(\sigma) \le \frac{1}{\beta}\log Z_H(\beta) \le \max_{\sigma\in\Omega} H(\sigma) + \frac{\log|\Omega|}{\beta}\,. \tag{17}$$

Proof. We have

$$\frac{1}{\beta}\log\sum_{\sigma\in\Omega}\exp\left(\beta H(\sigma)\right) \ge \max_{\sigma\in\Omega}\frac{1}{\beta}\log\exp\left(\beta H(\sigma)\right) = \max_{\sigma\in\Omega} H(\sigma)\,, \tag{18}$$

and

$$\frac{1}{\beta}\log\sum_{\sigma\in\Omega}\exp\left(\beta H(\sigma)\right) \le \max_{\sigma\in\Omega}\frac{1}{\beta}\log|\Omega|\exp\left(\beta H(\sigma)\right) = \max_{\sigma\in\Omega} H(\sigma) + \frac{\log|\Omega|}{\beta}\,. \tag{19}$$

□

2.4 Fourier analysis on the hypercube

Fourier analysis is commonly used to study Boolean functions [O'D14]. For example, the Fourier basis provides a convenient way to understand the action of linear operators on a Boolean function.

We consider the space of Boolean functions $f : \{\pm1\}^k \to \mathbb{R}$ with the expectation inner product, over the uniform distribution on $\{\pm1\}^n$. These functions have a canonical decomposition.

Definition 2.25 (Fourier spectrum of a Boolean function)**.** *Every function $f : \{\pm1\}^k \to \mathbb{R}$ permits a unique decomposition as a linear combination of parity functions. Specifically,*

$$f(\sigma) = \sum_{S\subseteq[k]} \widehat{f}(S)\prod_{i\in S}\sigma_i\,, \tag{20}$$

where $\widehat{f}(S)$ are called the Fourier coefficients.

One can verify that the monomials $\{\prod_{i\in S}\sigma_i\}_{S\subseteq[k]}$ form an orthonormal basis, often called the *Fourier basis*.

Recall that the average value of f is exactly the Fourier coefficient of the empty set:

Fact 2.26. $\mathbb{E}_{\sigma\in\{\pm1\}^n}[f(\sigma)] = \widehat{f}(\varnothing)$.

We also consider the *noise stability* of a Boolean function, which describes how resistant the function is to independent noise on its input bits.

Definition 2.27 (ρ-correlated)**.** *Fix $\sigma \in \{\pm1\}^k$ and $\rho \in [-1,+1]$. Then a random sample τ of the distribution $N_\rho(\sigma)$ chooses each spin τ_i independently as*

$$\tau_i = \begin{cases} \sigma_i & \text{with probability } \frac{1+\rho}{2}\,, \\ -\sigma_i & \text{with probability } \frac{1-\rho}{2}\,. \end{cases} \tag{21}$$

Since $\mathbb{E}_{\tau\sim N_\rho(\sigma)}[\tau_i\sigma_i] = \rho$, we say that τ is ρ-correlated with σ.

Definition 2.28 (Noise stability of a Boolean function around a point). *For $\rho \in [-1, +1]$, define the noise stability of the Boolean function $f : \{\pm 1\}^k \to \mathbb{R}$ around point $\sigma \in \{\pm 1\}^n$ as*

$$\mathrm{Stab}_\rho[f](\sigma) = \mathop{\mathbb{E}}_{\tau \sim N_\rho(\sigma)}\left[f(\sigma) f(\tau)\right]. \tag{22}$$

Definition 2.29 (Noise stability of a Boolean function, [O'D14, Theorem 2.49]). *The noise stability of a Boolean function $f : \{\pm 1\}^k \to \mathbb{R}$ is defined as*

$$\mathrm{Stab}_\rho[f] = \mathop{\mathbb{E}}_{\sigma \sim \{\pm 1\}^k}\left[\mathop{\mathbb{E}}_{\tau \sim N_\rho(\sigma)}\left[f(\sigma) f(\tau)\right]\right] = \mathop{\mathbb{E}}_{\sigma \sim \{\pm 1\}^k}\left[\mathrm{Stab}_\rho[f](\sigma)\right] = \sum_{S \subseteq [k]} \rho^{|S|} \widehat{f}(S)^2 . \tag{23}$$

An equivalent way of writing $\mathrm{Stab}_\rho[f]$ in the Fourier basis partitions the levels of the hypercube; that is,

$$\mathrm{Stab}_\rho[f] = \sum_{S \subseteq [k]} \rho^{|S|} \widehat{f}(S)^2 = \sum_{j=0}^{k} \rho^j \|f^{=j}\|^2 , \tag{24}$$

where,

$$\|f^{=j}\|^2 = \sum_{T \subseteq [k], |T| = j} \widehat{f}(T)^2 , \tag{25}$$

is the Fourier weight of the jth Fourier level of f.

2.5 Concentration inequalities

In the probabilistic combinatorics literature, "with high probability" means with probability $1 - o_n(1)$. All of the quantities that we consider are exponentially concentrated. Therefore, we will work with their expectations and conclude convergence in probability. We define the convergence $\lim_{n\to\infty} X_n \overset{p}{=} \lim_{n\to\infty} Y_n$ or $\lim_{n\to\infty} X_n \overset{a.s.}{=} \lim_{n\to\infty} Y_n$ as $X_n - Y_n$ converging to 0.

Definition 2.30 (Convergence in probability). *A sequence of random variables X_n converges in probability to $L \in \mathbb{R}$, written $X_n \xrightarrow{p} L$ or $\lim_{n\to\infty} X_n \overset{p}{=} L$, if for every fixed $\varepsilon > 0$,*

$$\Pr\left[|X_n - L| > \varepsilon\right] = o_n(1) . \tag{26}$$

Fact 2.31. *Suppose X_n is a sequence of random variables such that:*

1. *the limit of the expectation exists: $L = \lim_{n\to\infty} \mathbb{E}\, X_n$, and*
2. *X_n is concentrated: for every $\varepsilon > 0$ fixed, $\Pr[|X_n - \mathbb{E}\, X_n| > \varepsilon] = o_n(1)$.*

Then $X_n \xrightarrow{p} L$.

The Parisi formula (Theorem 2.11) gives a stronger form of convergence.

Definition 2.32 (Almost-sure convergence). *A sequence of random variables X_n almost surely converges to $L \in \mathbb{R}$, written $\lim_{n\to\infty} X_n \overset{a.s.}{=} L$, if*

$$\Pr\left[\lim_{n\to\infty} X_n \text{ exists and equals } L\right] = 1 . \tag{27}$$

In the remainder of this section, we state concentration inequalities for the Gibbs measures we consider.

Lemma 2.33 (Concentration under spin glass Gibbs distribution, [Pan13, Theorem 1.2]). *Let $\beta > 0$ and $f : \{-1,+1\}^n \to \mathbb{R}$ be arbitrary. Let $\{H(\sigma)\}_{\sigma\in\{-1,+1\}^n}$ be a Gaussian process such that $\mathbb{E}\, H(\sigma)^2 \le a$ for all $\sigma \in \{-1,+1\}^n$. Let $X = \log \sum_{\sigma\in\{-1,+1\}^n} \exp(\beta H(\sigma)) f(\sigma)$. Then for all $x \ge 0$,*

$$\Pr\Big[\big|X - \mathbb{E}\, X\big| \ge x\Big] \le 2\exp\left(\frac{-x^2}{4a\beta^2}\right). \tag{28}$$

For a spin glass and a function f with magnitude exponential in n, X is of order n. Since $\mathbb{E}\, H^{\xi}(\sigma)^2 = n \cdot \xi(1) = O(n)$ for a spin glass by Fact 2.12, X is concentrated up to $O(\sqrt{n})$ fluctuations around its mean.

We use a similar concentration inequality for the CSP setting.

Lemma 2.34 (Concentration under CSP Gibbs distribution, [CLSS22, Lemma 7.6]). *Let $\beta > 0$ and $f : \{-1,+1\}^n \to \mathbb{R}$ be arbitrary. Let $\{H(\sigma)\}_{\sigma\in\{-1,+1\}^n}$ be a sample from $\mathsf{CSP}_\Lambda(\alpha)$. Let*

$$X = \log \sum_{\sigma\in\{-1,+1\}^n} \exp(\beta H(\sigma)) f(\sigma).$$

Then, the following holds for every $\delta > 0$:

$$\Pr\Big[\big|X - \mathbb{E}\, X\big| \ge \delta n\Big] \le 2\exp\left(-\Omega(n)\right). \tag{29}$$

The above statement can be proved via [CLSS22, Lemma 7.6 (arXiv version)] with the observation that the free energy is also Lipschitz in the number of hyperedges.

3 Sparse and dense models have the same free energy

In this section we prove Theorem 1.1. We first define a *coupled model* for both $\mathsf{CSP}_\Lambda(\alpha)$ and SG_ξ. We connect the free energy of the coupled models via the Guerra-Toninelli interpolation as in several prior works [Pan05, CGPR19, JS20, CLSS22]; our proof is especially inspired by [Pan18].

Definition 3.1 ((A,b)-coupled models). *Let $A \in \{0,1\}^{\ell\times\ell'}$ be a 0-1 matrix and $b \in \mathbb{R}^{\ell'}_{\ge 0}$ a nonnegative vector satisfying $(Ab)_i = 1$ for all $i \in [\ell]$.*

An (A,b)-coupled model of SG_ξ is a collection of Hamiltonians $\mathcal{G}_1, \dots, \mathcal{G}_\ell$ related by

$$\begin{pmatrix} \mathcal{G}_1 \\ \vdots \\ \mathcal{G}_\ell \end{pmatrix} = A \begin{pmatrix} \mathcal{G}'_1 \\ \vdots \\ \mathcal{G}'_{\ell'} \end{pmatrix}, \tag{30}$$

where each $\mathcal{G}'_{i'}$ is the Hamiltonian for an independent instance of $\mathsf{SG}_{s\to\sqrt{b_{i'}}\xi(s)}$. The grand Hamiltonian *is $\mathcal{G} : (\{\pm1\}^n)^\ell \to \mathbb{R}$ defined by $\mathcal{G}(\mathbf{x}) = \sum_{i=1}^{\ell} \mathcal{G}_i(\mathbf{x}_i)$.*

An (A,b)-coupled model of $\mathsf{CSP}_\Lambda(\alpha)$ is a collection of Hamiltonians $\mathcal{H}_1, \dots, \mathcal{H}_\ell$ related by

$$\begin{pmatrix} \mathcal{H}_1 \\ \vdots \\ \mathcal{H}_\ell \end{pmatrix} = A \begin{pmatrix} \mathcal{H}'_1 \\ \vdots \\ \mathcal{H}'_{\ell'} \end{pmatrix}, \tag{31}$$

where each $\mathcal{H}'_{i'}$ is the Hamiltonian for an independent instance of $\mathsf{CSP}_\Lambda(b_{i'}\alpha)$. *The* grand Hamiltonian *is* $\mathcal{H} : (\{\pm 1\}^n)^\ell \to \mathbb{R}$ *defined by* $\mathcal{H}(\mathbf{x}) = \sum_{i=1}^{\ell} \mathcal{H}_i(\mathbf{x}_i)$.

In this section we use $\mathcal{G}$ for a grand Hamiltonian of a spin glass and $\mathcal{H}$ for that of a CSP. Expectations are over the randomness of the Hamiltonians.

Notice that A is $\{0,1\}$-valued. One can think of this matrix as choosing which of the ℓ' independent "hidden" instances are connected to each of the ℓ "observed" instances. The vector b gives the variances of the hidden instances, and the scaling is set so that the variance of each observed instance is the same as a single instance.

We study the free energy of our models, restricted to sets of assignments with particular overlap vectors.

Definition 3.2 (Free energy of states with given overlap)**.** *Let $S \subseteq \mathcal{R}^{(\ell)}$ be an open set. Let $\mathcal{H} : (\{\pm 1\}^n)^\ell \to \mathbb{R}$ be a grand Hamiltonian. Define $Z_{\mathcal{H},S}(\beta)$ as the partition function of $\mathcal{H}$ when the configuration tuples are restricted to those having overlap vector in S; that is,*

$$Z_{\mathcal{H},S}(\beta) = \sum_{\mathbf{x} \in U_n^{(\ell)}(S)} e^{\beta \mathcal{H}(\mathbf{x})} . \tag{32}$$

Define the free energy density of $\mathcal{H}$ as

$$\phi_{\mathcal{H},S}(\beta) = \frac{1}{\beta \ell n} \log Z_{\mathcal{H},S}(\beta) . \tag{33}$$

Remark 3.3. *We ignore the edge case $U_n^{(\ell)}(S) = \emptyset$.*

In order to use the Guerra-Toninelli interpolation, we modify the CSP model so that the number of constraints is $m \sim \mathrm{Pois}(\alpha n)$ rather than $m = \alpha n$ fixed. Because $\mathrm{Pois}(\alpha n)$ concentrates around its mean as $n \to \infty$, the free energy density is asymptotically the same:

Lemma 3.4. *Let $\phi^{(Pois)}_{\mathcal{H},S}$ and $\phi^{(exact)}_{\mathcal{H},S}$ denote the free energy density of a CSP instance with $m \sim \mathrm{Pois}(\alpha n)$ and $m = \alpha n$ clauses, respectively. Then*

$$\left| \mathbb{E}\, \phi^{(Pois)}_{\mathcal{H},S} - \mathbb{E}\, \phi^{(exact)}_{\mathcal{H},S} \right| = o_n(1) . \tag{34}$$

Proof. Let m be the number of edges in the Poisson model and αn in the exact model. Couple the two models so that the first αn edges of the Poisson model equal the exact model. Comparing the Hamiltonians of the two instances, we find

$$\frac{1}{\alpha} \sum_{e \in E\left(I^{(exact)}\right)} f_e(\sigma_e) - \frac{|f|}{\alpha} \cdot |m - \alpha n| \le \frac{1}{\alpha} \sum_{e \in E\left(I^{(Pois)}\right)} f_e(\sigma_e) \le \frac{1}{\alpha} \sum_{e \in E\left(I^{(exact)}\right)} f_e(\sigma_e) + \frac{|f|}{\alpha} |m - \alpha n| , \tag{35}$$

where $|f| = \max_{x \in \{\pm 1\}^k} f(x)$. This gives a corresponding bound on the difference in free energy density:

$$\left| \phi^{(\mathrm{Pois})}_{\mathcal{H},S} - \phi^{(\mathrm{exact})}_{\mathcal{H},S} \right| \le \frac{|f| \cdot |m - \alpha n|}{\alpha n} . \tag{36}$$

Taking the expectation,

$$\left|\mathbb{E}\,\phi^{(\mathrm{Pois})}_{\mathcal{H},S} - \mathbb{E}\,\phi^{(\mathrm{exact})}_{\mathcal{H},S}\right| \le \mathbb{E}\left|\phi^{(\mathrm{Pois})}_{\mathcal{H},S} - \phi^{(\mathrm{exact})}_{\mathcal{H},S}\right| \qquad \text{(Jensen's inequality)} \tag{37}$$

$$\le \frac{|f|\cdot \mathbb{E}_{m\sim\mathrm{Pois}(\alpha n)}\,|m-\alpha n|}{\alpha n} \qquad \text{(Equation (36))} \tag{38}$$

$$\le \frac{|f|\cdot \left(\mathbb{E}_{m\sim\mathrm{Pois}(\alpha n)}(m-\alpha n)^2\right)^{1/2}}{\alpha n} \qquad \text{(Jensen's inequality)} \tag{39}$$

$$= \frac{|f|\cdot\sqrt{\alpha n}}{\alpha n} = o_n(1)\,. \tag{40}$$

□

Now we can state our main interpolation.

Theorem 3.5 (Interpolation of random Max-CSPs and spin glasses; generalized version of Theorem 1.1)**.** *Choose two positive integers ℓ and ℓ'. Consider any set $S \subseteq \mathcal{R}^{(\ell)}$. Let $\mathcal{G}$ and $\mathcal{H}$ be grand Hamiltonians for (A,b)-coupled models of SG_ξ and $\mathsf{CSP}_\Lambda(\alpha)$, respectively, where ξ is related to Λ as in Equation (1).*

Then the following holds, where the asymptotic is as $\frac{\beta}{\alpha}\to 0$ but is independent of n:

$$\mathbb{E}\,\phi_{\mathcal{H},S}(\beta) = \mathop{\mathbb{E}}_{f\sim\Lambda}[f] + \frac{1}{\sqrt{\alpha}}\,\mathbb{E}\,\phi_{\mathcal{G},S}(\beta) + O\left(\frac{\beta^2}{\alpha^2}\right). \tag{41}$$

As a result, for all $\beta = o(\alpha^{3/4})$ we have:

$$\lim_{\alpha\to\infty}\sqrt{\alpha}\left(\mathbb{E}\,\phi_{\mathcal{H},S}(\beta) - \mathop{\mathbb{E}}_{f\sim\Lambda}[f]\right) - \mathbb{E}\,\phi_{\mathcal{G},S}(\beta) = 0\,. \tag{42}$$

Furthermore, by the concentration arguments in Section 2.5, we may replace the expectations by convergence in probability as $n\to\infty$ (after taking the limit as $\alpha\to\infty$), and for the case $S = \mathcal{R}^{(\ell)}$, almost-sure convergence.

3.1 Proof of Theorem 3.5

For notational convenience, we assume in this proof that Λ is supported on a single predicate f. The general case is handled by converting $\widehat{f}(\emptyset)$ back to $\mathbb{E}_{f\sim\Lambda}[f]$.

Fix positive integers ℓ, ℓ'. We define an interpolated Hamiltonian

$$\mathcal{K}(t,\mathbf{x}) = \mathcal{H}^{\alpha(1-t)n}(\mathbf{x}) - (1-t)\ell n\widehat{f}(\varnothing) + \sqrt{\frac{t}{\alpha}}\mathcal{G}(\mathbf{x})\,, \tag{43}$$

where $\mathcal{H}^{\alpha(1-t)n}$ is the grand Hamiltonian of an (A,b)-coupled model of $\mathsf{CSP}_\Lambda(\alpha(1-t))$. The parameter t controls the interpolation from the Max-CSP (when $t=0$) to the spin glass (when $t=1$).

We let $\beta > 0$ be a parameter independent of n. We will later choose β such that $1 \ll \beta \ll \alpha$ as $\alpha\to\infty$.

Fix an open subset $S \subseteq \mathcal{R}^{(\ell)}$. We write the average free energy density of $\mathcal{K}$, at inverse temperature β, among states that produce overlap vectors in S:

$$\phi_\beta(t) := \mathop{\mathbb{E}}_{\mathcal{H},\mathcal{G}}\phi_{\mathcal{K}(t,\cdot),S}(\beta) \tag{44}$$

$$= \frac{1}{\ell n}\frac{1}{\beta}\mathop{\mathbb{E}}_{\mathcal{H},\mathcal{G}}\log\sum_{\mathbf{x}\in U_n(S)}\exp\left(\beta\mathcal{K}(t,\mathbf{x})\right). \tag{45}$$

In this proof, we upper-bound the derivative $\frac{\mathrm{d}}{\mathrm{d}t}\phi_\beta(t)$, and thereby the difference between $\phi_\beta(0)$ and $\phi_\beta(1)$, showing that the free energy densities of $\mathcal{G}$ and $\mathcal{H}$ are close.

3.1.1 Taking the derivative of $\phi_\beta(t)$

Let's calculate $\frac{\mathrm{d}}{\mathrm{d}t}\phi_\beta(t)$. First, we generalize $\mathcal{K}$ and ϕ_β so the t-dependence of $\mathcal{G}$ and $\mathcal{H}$ are controlled by independent parameters:

$$\widetilde{\mathcal{K}}(\rho,\gamma,\mathbf{x}) := \mathcal{H}^\rho(\mathbf{x}) - \frac{\rho\ell}{\alpha}\widehat{f}(\varnothing) + \frac{\gamma}{\sqrt{\alpha}}\mathcal{G}(\mathbf{x})\,, \tag{46}$$

$$\widetilde{\phi}_\beta(\rho,\gamma) := \frac{1}{\beta\ell n}\mathop{\mathbb{E}}_{\mathcal{H}^\rho,\mathcal{G}}\log\sum_{\mathbf{x}\in U_n(S)}\exp\left(\beta\widetilde{\mathcal{K}}(\rho,\gamma,\mathbf{x})\right). \tag{47}$$

When $\rho = (1-t)\alpha n$ and $\gamma = \sqrt{t}$, we recover the original expressions:

$$\mathcal{K}(t,\mathbf{x}) = \widetilde{\mathcal{K}}((1-t)\alpha n,\ \sqrt{t},\mathbf{x}) \tag{48}$$

$$\phi_\beta(t) = \widetilde{\phi}_\beta((1-t)\alpha n,\ \sqrt{t}) \tag{49}$$

We then take a derivative using the chain rule:

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi_\beta(t) = \left(\frac{\partial}{\partial\rho}\widetilde{\phi}_\beta(\rho,\gamma)\right)\frac{\mathrm{d}\rho}{\mathrm{d}t} + \left(\frac{\partial}{\partial\gamma}\widetilde{\phi}_\beta(\rho,\gamma)\right)\frac{\mathrm{d}\gamma}{\mathrm{d}t} \tag{50}$$

$$= -\alpha n\left(\frac{\partial}{\partial\rho}\widetilde{\phi}_\beta(\rho,\gamma)\right) + \frac{1}{2\sqrt{t}}\left(\frac{\partial}{\partial\gamma}\widetilde{\phi}_\beta(\rho,\gamma)\right) \tag{51}$$

We also introduce a Gibbs expectation operator to use when computing the partial derivatives. Given a function $p(\mathbf{x})$, the average of p with respect to the Gibbs distribution of $\widetilde{\mathcal{K}}(\rho,\gamma,x)$ is

$$\langle p\rangle_{\mathbf{x}} := \frac{\sum_{\mathbf{x}\in U_n(S)} p(\mathbf{x})\cdot\exp\left(\beta\widetilde{\mathcal{K}}(\rho,\gamma,\mathbf{x})\right)}{\sum_{\mathbf{x}\in U_n(S)}\exp\left(\beta\widetilde{\mathcal{K}}(\rho,\gamma,\mathbf{x})\right)}\,. \tag{52}$$

3.1.2 The Poisson derivative

To calculate $\frac{\partial}{\partial\rho}\widetilde{\phi}_\beta(\rho,\gamma)$, we introduce ℓ' new variables $\rho_1,\dots,\rho_{\ell'}$ to parameterize the independent Poisson instances used to construct $\mathcal{H}^\rho$. We introduce intermediate functions

$$\widetilde{\mathcal{K}}(\rho_1,\dots,\rho_{\ell'},\gamma,\mathbf{x}) = \sum_{i\in[\ell]}\left[\left(A\begin{pmatrix}\mathcal{H}_1'^{\rho_1}\\ \vdots\\ \mathcal{H}_{\ell'}'^{\rho_{\ell'}}\end{pmatrix}\right)_i(\mathbf{x}_i) - \left(A\begin{pmatrix}\rho_1\\ \vdots\\ \rho_{\ell'}\end{pmatrix}\right)_i\frac{\widehat{f}(\varnothing)}{\alpha} + \frac{\gamma}{\sqrt{\alpha}}\mathcal{G}_i(\mathbf{x}_i)\right], \tag{53}$$

$$\widetilde{\phi}_\beta(\rho_1,\dots,\rho_{\ell'},\gamma) = \frac{1}{\beta\ell n}\mathop{\mathbb{E}}_{\mathcal{H}_1'^{\rho_1},\dots,\mathcal{H}_{\ell'}'^{\rho_{\ell'}}}\mathop{\mathbb{E}}_{\mathcal{G}}\left[\log\sum_{\mathbf{x}\in U_n^{(\ell)}(S)}\exp\left(\beta\widetilde{\mathcal{K}}(\rho_1,\dots,\rho_{\ell'},\gamma,\mathbf{x})\right)\right]. \tag{54}$$

When $\rho_{i'} = b_{i'}\rho$ for all $i'\in[\ell']$, we recover the original functions $\widetilde{\mathcal{K}}$ and $\widetilde{\phi}_\beta$. In this case, through another application of the chain rule,

$$\frac{\partial}{\partial\rho}\widetilde{\phi}_\beta(\rho,\gamma) = \sum_{i'\in[\ell']}\left(\frac{\partial}{\partial\rho_{i'}}\widetilde{\phi}_\beta(\rho_1,\dots,\rho_{\ell'},\gamma)\right)\frac{\partial\rho_{i'}}{\partial\rho} = \sum_{i\in[\ell']} b_{i'}\frac{\partial}{\partial\rho_{i'}}\widetilde{\phi}_\beta(\rho_1,\dots,\rho_{\ell'},\gamma)\,. \tag{55}$$

We use the explicit derivative of a Poisson variable:

Fact 3.6 (Derivative of a Poisson variable)**.**

$$\frac{\partial}{\partial\lambda}\mathop{\mathbb{E}}_{X\sim\mathrm{Pois}(\lambda)}\Big[f(X)\Big]=\mathop{\mathbb{E}}_{X\sim\mathrm{Pois}(\lambda)}\Big[f(X+1)-f(X)\Big] \tag{56}$$

Because of Fact 3.6,

$$\frac{\partial}{\partial\rho_{i'}}\widetilde{\phi}_\beta(\rho_1,\dots,\rho_{\ell'},\gamma)=\widetilde{\phi}_\beta(\rho_1,\dots,\rho_{i'}+1,\dots,\rho_{\ell'},\gamma)-\widetilde{\phi}_\beta(\rho_1,\dots,\rho_{i'},\dots,\rho_{\ell'},\gamma) \tag{57}$$

$$=\frac{1}{\beta\ell n}\mathop{\mathbb{E}}_{\mathcal{H}_1'^{\rho_1},\dots,\mathcal{H}_{\ell'}'^{\rho_{\ell'}}}\mathop{\mathbb{E}}_{\mathcal{G}}\log\left\langle\exp\left(\beta\widetilde{\mathcal{K}}(\rho_1,\dots,\rho_{i'}+1,\dots,\rho_{\ell'},\gamma,\mathbf{x})-\beta\widetilde{\mathcal{K}}(\rho_1,\dots,\rho_{i'},\dots,\rho_{\ell'},\gamma,\mathbf{x})\right)\right\rangle_{\mathbf{x}} \tag{58}$$

$$=\frac{1}{\beta\ell n}\mathop{\mathbb{E}}_{\mathcal{H}_1'^{\rho_1},\dots,\mathcal{H}_{\ell'}'^{\rho_{\ell'}}}\log\left\langle\exp\left(\beta\sum_{j\in[\ell]}A_{ji'}\big((\mathcal{H}_{i'}'^{\rho_{i'}+1}-\mathcal{H}_{i'}'^{\rho_{i'}})(\mathbf{x}_j)-\widehat{f}(\varnothing)\big)\right)\right\rangle_{\mathbf{x}}. \tag{59}$$

The difference of $\mathcal{H}_{i'}'^{\rho_{i'}+1}$ and $\mathcal{H}_{i'}'^{\rho_{i'}}$ is a single extra clause (normalized by α), applied to k random indices of the input with random signs. Let f^* be the extra clause. Then

$$=\frac{1}{\beta\ell n}\mathop{\mathbb{E}}_{f^*}\log\left\langle\exp\left(\frac{\beta}{\alpha}\sum_{j\in[\ell]}A_{ji'}\big(f^*(\mathbf{x}_j)-\widehat{f}(\varnothing)\big)\right)\right\rangle_{\mathbf{x}}. \tag{60}$$

When $\frac{\beta}{\alpha}$ is small, this term can be Taylor-expanded as $\log z=\log\left(1-(1-z)\right)=-\sum_{p\geq1}\frac{(1-z)^p}{p}$:

$$=\frac{-1}{\beta\ell n}\mathop{\mathbb{E}}_{f^*}\sum_{p\geq1}\frac{1}{p}\left(1-\left\langle\exp\left(\frac{\beta}{\alpha}\sum_{j\in[\ell]}A_{ji'}\big(f^*(\mathbf{x}_j)-\widehat{f}(\varnothing)\big)\right)\right\rangle_{\mathbf{x}}\right)^p. \tag{61}$$

We introduce additional "replicas" $\mathbf{x}_{(s)}$ of $\mathbf{x}$. Precisely, each $\mathbf{x}_{(s)}$ is an i.i.d. copy of $\mathbf{x}$. For any function w and any set of replicas $y_0,y_1,\dots$, we have the identity $\langle w(y_0)\rangle^p=\left\langle\prod_{s\in[p]}w(y_s)\right\rangle$. As a result,

$$=\frac{-1}{\beta\ell n}\mathop{\mathbb{E}}_{f^*}\left\langle\sum_{p\geq1}\frac{1}{p}\prod_{s=1}^{p}\left(1-\exp\left(\frac{\beta}{\alpha}\sum_{j\in[\ell]}A_{ji'}\big(f^*(\mathbf{x}_{(s)_j})-\widehat{f}(\varnothing)\big)\right)\right)\right\rangle_{\mathbf{x}_{(1)},\mathbf{x}_{(2)},\cdots}. \tag{62}$$

We Taylor-expand this expression in $\frac{\beta}{\alpha}$. The first line comes from the $p=1$ term and the second line comes from the $p=2$ term:

$$=\frac{-1}{\beta\ell n}\mathop{\mathbb{E}}_{f^*}\Bigg\langle-\frac{\beta}{\alpha}\sum_{j\in[\ell]}A_{ji'}\big(f^*(\mathbf{x}_j)-\widehat{f}(\varnothing)\big)-\frac{\beta^2}{2\alpha^2}\left(\sum_{j\in[\ell]}A_{ji'}\big(f^*(\mathbf{x}_j)-\widehat{f}(\varnothing)\big)\right)^2$$
$$+\frac{\beta^2}{2\alpha^2}\left(\sum_{j\in[\ell]}A_{ji'}\big(f^*(\mathbf{x}_j)-\widehat{f}(\varnothing)\big)\right)\left(\sum_{j\in[\ell]}A_{ji'}\big(f^*(\mathbf{y}_j)-\widehat{f}(\varnothing)\big)\right)+O(\frac{\beta^3}{\alpha^3})\Bigg\rangle_{\mathbf{x},\mathbf{y}}. \tag{63}$$

Notice that because the clause applies random signs to its input,

$$\mathop{\mathbb{E}}_{f^*}\Big[\big\langle f^*(\mathbf{x}_j)\big\rangle_{\mathbf{x}_j}\Big]=\widehat{f}(\varnothing)\,. \tag{64}$$

As a result, the only terms that remain are

$$= \frac{-\beta}{2\alpha^2 \ell n} \mathop{\mathbb{E}}_{f^*} \left\langle \sum_{i,j \in [\ell]} A_{ii'} A_{ji'} \Big(- f^*(\mathbf{x}_i) f^*(\mathbf{x}_j) + f^*(\mathbf{x}_i) f^*(\mathbf{y}_j) \Big) \right\rangle_{\mathbf{x},\mathbf{y}} + \frac{1}{\ell n} O(\frac{\beta^2}{\alpha^3}) \,. \tag{65}$$

From here, we rewrite the correlation of f^* as a function of the noise stability of f. Specifically, let $u_a = \varepsilon_a(\mathbf{x}_i)_{d_a}$ and $v_a = \varepsilon_a(\mathbf{y}_j)_{d_a}$ for some uniformly chosen $\varepsilon \in_{\mathrm{R}} \{\pm 1\}^k$ and $d \in_{\mathrm{R}} [\ell]^k$. Then u, v are marginally uniform points in the hypercube that are $R(\mathbf{x}_i, \mathbf{y}_j)$-correlated.

Write $f^*(\sigma) = f(\varepsilon_1 \sigma_{d_1}, \dots, \varepsilon_k \sigma_{d_k})$. Then

$$\left\langle f^*(\mathbf{x}_i) f^*(\mathbf{y}_j) \right\rangle_{\mathbf{x}_i, \mathbf{y}_j} = \langle f(u) f(v) \rangle_{\mathbf{x}_i, \mathbf{y}_j} = \left\langle \mathrm{Stab}_{R(\mathbf{x}_i, \mathbf{y}_j)}[f] \right\rangle_{\mathbf{x}_i, \mathbf{y}_j} \tag{66}$$

where $\mathrm{Stab}_{R(\mathbf{x}_i, \mathbf{y}_j)}[f]$ is defined as in Definition 2.29. Using this, we get

$$= \frac{-\beta}{2\alpha^2 \ell n} \sum_{i,j \in [\ell]} A_{ii'} A_{ji'} \left\langle - \mathrm{Stab}_{R(\mathbf{x}_i, \mathbf{x}_j)}[f] + \mathrm{Stab}_{R(\mathbf{x}_i, \mathbf{y}_j)}[f] \right\rangle_{\mathbf{x},\mathbf{y}} + \frac{1}{\ell n} O(\frac{\beta^2}{\alpha^3}) \tag{67}$$

$$= \frac{\beta}{2\alpha^2 \ell n} \sum_{i,j \in [\ell]} A_{ii'} A_{ji'} \left\langle \xi(R(\mathbf{x}_i, \mathbf{x}_j)) - \xi(R(\mathbf{x}_i, \mathbf{y}_j)) \right\rangle_{\mathbf{x},\mathbf{y}} + \frac{1}{\ell n} O(\frac{\beta^2}{\alpha^3}) \,. \tag{68}$$

We can now write the ρ-derivative of $\widetilde{\phi}_\beta$:

$$\frac{\partial}{\partial \rho} \widetilde{\phi}_\beta = \frac{\beta}{2\alpha^2 \ell n} \sum_{i' \in [\ell']} \sum_{i,j \in [\ell]} b_{i'} A_{ii'} A_{ji'} \left\langle \xi(R(\mathbf{x}_i, \mathbf{x}_j)) - \xi(R(\mathbf{x}_i, \mathbf{y}_j)) \right\rangle_{\mathbf{x},\mathbf{y}} + \frac{1}{\ell n} O(\frac{\beta^2}{\alpha^3}) \tag{69}$$

3.1.3 The Gaussian derivative

We compute the derivative of $\widetilde{\phi}_\beta(\rho, \gamma)$ with respect to γ. Since the variance of each $\mathcal{G}_i$ is independent of γ, the derivative pulls into the expectation operator:

$$\frac{\partial}{\partial \gamma} \widetilde{\phi}_\beta(\rho, \gamma) = \frac{1}{\beta \ell n} \mathop{\mathbb{E}}_{\mathcal{H}^\rho, \mathcal{G}} \frac{\partial}{\partial \gamma} \log \left[\sum_{\mathbf{x} \in U_n(S)} \exp\left(\beta \widetilde{\mathcal{K}}(\rho, \gamma, \mathbf{x}) \right) \right] \tag{70}$$

By the chain rule, the γ-derivative of the partition function is proportional to the Gibbs average of the γ-derivative of the Hamiltonian $\widetilde{\mathcal{K}}$:

$$= \frac{1}{\ell n} \mathop{\mathbb{E}}_{\mathcal{H}^\rho, \mathcal{G}} \left[\left\langle \frac{\partial}{\partial \gamma} \left(\widetilde{\mathcal{K}}(\rho, \gamma, \mathbf{x}) \right) \right\rangle_{\mathbf{x}} \right] \tag{71}$$

Notice that the expectation is a correlation between a Gaussian process and a Gibbs measure, so we can use the following formula:

Lemma 3.7 (Stein's lemma with Gibbs average [Pan13, Lemma 1.1]). *Consider two jointly Gaussian processes $\{Y(\sigma)\}_\sigma$ and $\{Z(\sigma)\}_\sigma$. For any w, let $\langle \cdot \rangle_{\sigma_1, \dots, \sigma_w}$ be the w-product Gibbs measure with respect to the process $\{Z(\sigma)\}_\sigma$. Then*

$$\mathop{\mathbb{E}}_{Y,Z} [\langle Y(\sigma) \rangle_\sigma] = \mathop{\mathbb{E}}_{Z} \left[\left\langle \mathop{\mathbb{E}}_{Y} \left[Y(\sigma) Z(\sigma) \right] \right\rangle_\sigma - \left\langle \mathop{\mathbb{E}}_{Y} \left[Y(\sigma) Z(\sigma') \right] \right\rangle_{\sigma, \sigma'} \right] . \tag{72}$$

Consider the following Gaussian processes:

$$S_{\mathcal{H}^\rho,\mathcal{G}} : (\mathbf{x}) \to \frac{\partial}{\partial\gamma}\left(\widetilde{\mathcal{K}}(\rho,\gamma,\mathbf{x})\right) \tag{73}$$

$$T_{\mathcal{H}^\rho,\mathcal{G}} : (\mathbf{x}) \to \beta\widetilde{\mathcal{K}}(\rho,\gamma,\mathbf{x}) \tag{74}$$

Applying Lemma 3.7 then yields

$$= \frac{1}{\ell n}\mathop{\mathbb{E}}_{\mathcal{H}'^\rho,\mathcal{G}}\left[\left\langle \mathop{\mathbb{E}}_{\mathcal{H}'^\rho,\mathcal{G}}\left[S_{\mathcal{H}^\rho,\mathcal{G}}(\mathbf{x})T_{\mathcal{H}^\rho,\mathcal{G}}(\mathbf{x})\right]\right\rangle_{\mathbf{x}} - \left\langle \mathop{\mathbb{E}}_{\mathcal{H}'^\rho,\mathcal{G}}\left[S_{\mathcal{H}^\rho,\mathcal{G}}(\mathbf{x})T_{\mathcal{H}^\rho,\mathcal{G}}(\mathbf{y})\right]\right\rangle_{\mathbf{x},\mathbf{y}}\right]. \tag{75}$$

The derivative only acts on the Gaussian components of the Hamiltonian $\widetilde{\mathcal{K}}$:

$$= \frac{1}{\sqrt{\alpha}\ell n}\mathop{\mathbb{E}}_{\mathcal{H}^\rho,\mathcal{G}}\left[\left\langle \mathop{\mathbb{E}}_{\mathcal{G}}\left[\mathcal{G}(\mathbf{x})T_{\mathcal{H}^\rho,\mathcal{G}}(\mathbf{x})\right]\right\rangle_{\mathbf{x}} - \left\langle \mathop{\mathbb{E}}_{\mathcal{G}}\left[\mathcal{G}(\mathbf{x})T_{\mathcal{H}^\rho,\mathcal{G}}(\mathbf{y})\right]\right\rangle_{\mathbf{x},\mathbf{y}}\right] \tag{76}$$

$$= \frac{\beta}{\sqrt{\alpha}\ell n}\mathop{\mathbb{E}}_{\mathcal{H}^\rho,\mathcal{G}}\left[\left\langle \mathop{\mathbb{E}}_{\mathcal{G}}\left[\mathcal{G}(\mathbf{x})\Big(\mathcal{H}^\rho(\mathbf{x}) - \frac{\rho\ell}{\alpha}\widehat{f}(\varnothing) + \frac{\gamma}{\sqrt{\alpha}}\mathcal{G}(\mathbf{x})\Big)\right]\right\rangle_{\mathbf{x}} - \left\langle \mathop{\mathbb{E}}_{\mathcal{G}}\left[\mathcal{G}(\mathbf{x})\Big(\mathcal{H}^\rho(\mathbf{y}) - \frac{\rho\ell}{\alpha}\widehat{f}(\varnothing) + \frac{\gamma}{\sqrt{\alpha}}\mathcal{G}(\mathbf{y})\Big)\right]\right\rangle_{\mathbf{x},\mathbf{y}}\right] \tag{77}$$

The constant terms (proportional to $\widehat{f}(\varnothing)$) cancel. Furthermore, $\mathcal{G}$ is centered and is independent of $\mathcal{H}$: For each independent instance $\mathcal{G}'_i$, $\mathbb{E}[\mathcal{G}'_i\mathcal{H}^\rho] = 0$, and by linearity of expectation, $\mathbb{E}[\mathcal{G}\mathcal{H}^\rho] = 0$. All that remains is

$$= \frac{\beta\gamma}{\alpha\ell n}\mathop{\mathbb{E}}_{\mathcal{G}}\left[\left\langle \mathop{\mathbb{E}}_{\mathcal{G}}\left[\mathcal{G}(\mathbf{x})\mathcal{G}(\mathbf{x})\right]\right\rangle_{\mathbf{x}} - \left\langle \mathop{\mathbb{E}}_{\mathcal{G}}\left[\mathcal{G}(\mathbf{x})\mathcal{G}(\mathbf{y})\right]\right\rangle_{\mathbf{x},\mathbf{y}}\right]. \tag{78}$$

Using the definition of an (A,b)-coupled model, this is

$$= \frac{\beta\gamma}{\alpha\ell n}\mathop{\mathbb{E}}_{\mathcal{G}}\left[\left\langle \sum_{i,j\in[\ell]}\sum_{i',j'\in[\ell']} A_{ii'}A_{ji'}\sqrt{b_{i'}b_{j'}}\mathop{\mathbb{E}}_{\mathcal{G}}[\mathcal{G}'_{i'}(\mathbf{x}_i)\mathcal{G}'_{j'}(\mathbf{x}_j)]\right\rangle_{\mathbf{x}} - \left\langle \sum_{i,j\in[\ell]}\sum_{i',j'\in[\ell']} A_{ii'}A_{ji'}\sqrt{b_{i'}b_{j'}}\mathop{\mathbb{E}}_{\mathcal{G}}[\mathcal{G}'_{i'}(\mathbf{x}_i)\mathcal{G}'_{j'}(\mathbf{y}_j)]\right\rangle_{\mathbf{x},\mathbf{y}}\right]. \tag{79}$$

By Fact 2.12, we can write the covariance as a function of ξ:

$$\mathop{\mathbb{E}}_{\mathcal{G}}[\mathcal{G}'_{i'}(\mathbf{x}_i)\mathcal{G}'_{j'}(\mathbf{y}_j)] = \begin{cases} n\xi(R(\mathbf{x}_i,\mathbf{y}_j)) & i' = j' \\ 0 & \text{otherwise} \end{cases} \tag{80}$$

Thus, we have

$$= \frac{\beta\gamma}{\alpha\ell}\mathop{\mathbb{E}}_{\mathcal{G}}\left[\left\langle \sum_{i,j\in[\ell]}\sum_{i'\in[\ell']} A_{ii'}A_{ji'}b_{i'}\Big(\xi(R(\mathbf{x}_i,\mathbf{x}_j)) - \xi(R(\mathbf{x}_i,\mathbf{y}_j))\Big)\right\rangle_{\mathbf{x},\mathbf{y}}\right]. \tag{81}$$

3.1.4 Putting it all together

Now we can calculate the total derivative of $\phi_\beta(t)$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi_\beta(t) = -\alpha n\left(\frac{\partial}{\partial\rho}\widetilde{\phi}_\beta(\rho,\gamma)\right) + \frac{1}{2\sqrt{t}}\left(\frac{\partial}{\partial\gamma}\widetilde{\phi}_\beta(\rho,\gamma)\right) \tag{82}$$

$$= \left(\frac{-\beta}{2\alpha\ell} + \frac{\beta\gamma}{2\alpha\ell\sqrt{t}}\right)\sum_{i'\in[\ell']}\sum_{i,j\in[\ell]} b_{i'}A_{ii'}A_{ji'}\left\langle \xi(R(\mathbf{x}_i,\mathbf{x}_j)) - \xi(R(\mathbf{x}_i,\mathbf{y}_j))\right\rangle_{\mathbf{x},\mathbf{y}} - \frac{1}{\ell}O(\frac{\beta^2}{\alpha^2}) \tag{83}$$

$$= O(\frac{\beta^2}{\alpha^2})\,, \tag{84}$$

since $\gamma = \sqrt{t}$ and ℓ is a constant. So

$$\left|\phi_\beta(1) - \phi_\beta(0)\right| \le \max_{t\in[0,1]}\frac{\mathrm{d}}{\mathrm{d}t}\phi_\beta(t) = O(\frac{\beta^2}{\alpha^2})\,. \tag{85}$$

This proves Theorem 3.5.

4 Optimal value of a random Max-CSP

As a corollary of Theorem 3.5, we prove that in the large α limit, the optimal value of a coupled Max-CSP among solutions with given overlap structure is determined by that of a spin glass.

Corollary 4.1. *For all ℓ, ℓ' and (A,b)-coupled models $\mathcal{G}$ of SG_ξ and $\mathcal{H}$ of $\mathsf{CSP}_\Lambda(\alpha)$, and all open sets $S \subseteq \mathcal{R}^{(\ell)}$,*

$$\mathbb{E}\max_{\mathbf{x}\in U_n^{(\ell)}(S)}\frac{1}{\ell n}\mathcal{H}(\mathbf{x}) = \mathop{\mathbb{E}}_{f\sim\Lambda}[f] + \mathbb{E}\max_{\mathbf{x}\in U_n^{(\ell)}(S)}\frac{1}{\sqrt{\alpha}}\frac{1}{\ell n}\mathcal{G}(\mathbf{x}) + O(\alpha^{-2/3})\,. \tag{86}$$

By the concentration arguments in Section 2.5, we may replace the expectations by convergence in probability in the limit as $n\to\infty$, and for the case $S = \mathcal{R}^{(\ell)}$, almost-sure convergence.

Proof. Using Theorem 3.5, we have

$$\mathbb{E}\,\phi_{\mathcal{H},S}(\beta) = \mathop{\mathbb{E}}_{f\sim\Lambda}[f] + \frac{1}{\sqrt{\alpha}}\,\mathbb{E}\,\phi_{\mathcal{G},S}(\beta) + O(\frac{\beta^2}{\alpha^2})\,. \tag{87}$$

Recall that there are at most $2^{\ell n}$ choices of $\mathbf{x}$. Applying Fact 2.24 for $\mathcal{G}$ and for $\mathcal{H}$, we get

$$\left|\phi_{\mathcal{G},S}(\beta) - \max_{\mathbf{x}\in U_n(S)}\frac{1}{\ell n}\mathcal{G}(\mathbf{x})\right| \le \frac{\log|U_n(S)|}{\beta\ell n} = O(\frac{1}{\beta})\,, \tag{88}$$

$$\left|\phi_{\mathcal{H},S}(\beta) - \max_{\mathbf{x}\in U_n(S)}\frac{1}{\ell n}\mathcal{H}(\mathbf{x})\right| = O(\frac{1}{\beta})\,. \tag{89}$$

By Jensen's inequality, the same holds when using expectations. Plugging in these bounds and choosing $\beta = \alpha^{2/3}$ proves the claim. □

In the special case $\ell = \ell' = 1$, $(A,b) = (\mathbb{I},(1))$, and $S = \mathcal{R}^{(\ell)} = [-1,1]$, we conclude Corollary 1.2.

4.1 Numerical calculations

We compute the value of the Parisi formula for the spin glasses associated with popular CSPs in Table 1. This code can be run for any choice of spin glass and is available online[6]. Our code uses the zero-temperature representation of the Parisi functional from [AC17], which we restate below:

Definition 4.2 (Parisi functional at zero temperature [Par80, AC17])**.** *The Parisi functional at zero temperature and zero external magnetic field $\mathcal{P}^{\xi}(\infty, 0)$ gives the a.s. limit of the ground state energy density of a spin glass SG_{ξ} in the thermodynamic limit ($n \to \infty$) as*

$$\lim_{n\to\infty} \frac{1}{n}\left[\max_{\sigma\in\{\pm 1\}^n} H^{\xi}(\sigma)\right] \overset{a.s.}{=} \mathcal{P}^{\xi}(\infty, 0)\,, \tag{90}$$

where the Parisi functional is the following variational formula,

$$\mathcal{P}^{\xi}(\infty, 0) = \inf_{\zeta\in\mathcal{U}}\left(\Phi_{\zeta}(0,0) - \frac{1}{2}\int_0^1 s\xi''(s)\zeta(s)ds\right), \tag{91}$$

and the function $\Phi_{\zeta}(x;t) : \mathbb{R}\times[0,1] \to \mathbb{R}$ is the solution of the following Hamilton-Jacobi-Bellman equation (known as the Parisi PDE),

$$\partial_t\Phi_{\zeta}(x;t) + \frac{\xi''(t)}{2}\left(\partial_{xx}\Phi_{\zeta}(x;t) + \zeta(t)\left(\partial_x\Phi_{\zeta}(x;t)\right)^2\right) = 0\,, \tag{92}$$

with the initial condition,

$$\Phi_{\zeta}(x;1) = |x|\,, \tag{93}$$

and

$$\mathcal{U} = \left\{\zeta : [0,1) \to \mathbb{R}_{\geq 0} : \zeta \text{ is right-continuous, non-decreasing, and } \int_0^1 \zeta(t)dt < \infty\right\}. \tag{94}$$

k	**Max 1-in-k SAT**	**Max k NAESAT**	**Max k SAT**	**Max k XOR**
2	$\frac{1}{2} + \frac{0.54}{\sqrt{\alpha}}$	$\frac{1}{2} + \frac{0.54}{\sqrt{\alpha}}$	$\frac{3}{4} + \frac{0.40}{\sqrt{\alpha}}$	$\frac{1}{2} + \frac{0.54}{\sqrt{\alpha}}$
3	$\frac{3}{8} + \frac{0.54}{\sqrt{\alpha}}$	$\frac{3}{4} + \frac{0.47}{\sqrt{\alpha}}$	$\frac{7}{8} + \frac{0.33}{\sqrt{\alpha}}$	$\frac{1}{2} + \frac{0.58}{\sqrt{\alpha}}$
4	$\frac{1}{4} + \frac{0.48}{\sqrt{\alpha}}$	$\frac{7}{8} + \frac{0.37}{\sqrt{\alpha}}$	$\frac{15}{16} + \frac{0.26}{\sqrt{\alpha}}$	$\frac{1}{2} + \frac{0.58}{\sqrt{\alpha}}$
5	$\frac{5}{32} + \frac{0.41}{\sqrt{\alpha}}$	$\frac{15}{16} + \frac{0.28}{\sqrt{\alpha}}$	$\frac{31}{32} + \frac{0.20}{\sqrt{\alpha}}$	$\frac{1}{2} + \frac{0.59}{\sqrt{\alpha}}$

Table 1: Optimal value of a random k-CSP with n variables and αn clauses, as $n \to \infty$. The calculation uses the form of the Parisi formula in Definition 4.2. All values are rounded to two decimal places. The values for kXOR match those in [MH22]. We expect the values to be accurate to two significant figures, based on consistency with independently calculated values for 2XOR and 3XOR [AMS20].

[6] https://github.com/marwahaha/csp-parisi/

5 Overlap gaps in a random Max-CSP

Theorem 3.5 implies that many quantities are equivalent for $\mathsf{CSP}_\Lambda(\alpha)$ at large α and its associated spin glass model. For example:

1. Free energy of a single instance (proving Theorem 1.1).
   $(\ell = 1, \ell' = 1, A = [[1]], b = [1], S = \mathcal{R})$
2. Free energy of a single instance, restricting the Hamming weight to $W \subseteq [-1, 1]$ (related to [JS20]).
   $(\ell = 1, \ell' = 1, A = [[1]], b = [1], S = \{r \in \mathcal{R} : r_{\{1\}} \in W\})$
3. Existence (or non-existence) of an overlap gap property in the overlap range (s, t).
   $(\ell = 2, \ell' = 1, A = [[1], [1]], b = [1], S = \{r \in \mathcal{R} : s < r_{\{1,2\}} < t\})$
4. Existence (or non-existence) of an η-coupled overlap gap property in the overlap range (s, t) [CGPR19].
   $(\ell = 2, \ell' = 3, A = [[1,\ 1,\ 0], [1,\ 0,\ 1]], b = [\eta,\ 1-\eta,\ 1-\eta], S = \{r \in \mathcal{R} : s < r_{\{1,2\}} < t\})$
5. Existence (or non-existence) of the branching overlap gap property in [HS21].
   (Choice of parameters in Section 5.1)

We use Theorem 3.5 to show how $\mathsf{CSP}_\Lambda(\alpha)$ inherits an overlap gap property from a spin glass. This proof is a generalization of [CGPR19, Proof of Theorem 5] and [CLSS22, Lemma 8.14] to hold for arbitrary (A, b)-coupled instances. We use the same argument to transfer the branching OGP of [HS21], a hierarchical style of OGP described by a rooted tree.

Definition 5.1 (OGP [Gam21]). *Given a Hamiltonian $H : \{\pm 1\}^n \to \mathbb{R}$, an overlap gap property (OGP) with value v exists if there are $-1 \le s < t \le 1$ such that for all σ_1, σ_2 with $H(\sigma_1) \ge v, H(\sigma_2) \ge v$, we have*

$$R(\sigma_1, \sigma_2) \notin (s, t)\,. \tag{95}$$

Definition 5.2 (Average-OGP). *A Hamiltonian H has an average-OGP with value v if the same holds whenever $\frac{1}{2}(H(\sigma_1) + H(\sigma_2)) \ge v$.*

Remark 5.3. *An average-OGP implies an OGP. On the other hand, an OGP implies an average-OGP with a weakened (larger) value. The interpolation in this work transfers the average-OGP.*

Proposition 5.4. *If an average-OGP with value v holds for SG_ξ with high probability, then for all $\varepsilon > 0$, for all sufficiently large α, an average-OGP with value $\mathbb{E}_{f\sim\Lambda}[f] + \frac{v+\varepsilon}{\sqrt{\alpha}}$ holds for $\mathsf{CSP}_\Lambda(\alpha)$ with high probability, when ξ is related to Λ as in Equation (1).*

Proof. Let $\ell = 2$, and let $S = (a, b)$ be the overlap gap for the spin glass model SG_ξ. Using Corollary 4.1, for sufficiently large α, we have

$$\mathbb{E} \max_{\mathbf{x} \in U_n^{(\ell)}(S)} \frac{1}{\ell n}\mathcal{H}(\mathbf{x}) \le \mathop{\mathbb{E}}_{f\sim\Lambda}[f] + \frac{\mathbb{E} \max_{\mathbf{x} \in U_n^{(\ell)}(S)} \frac{1}{\ell n}\mathcal{G}(\mathbf{x}) + \varepsilon}{\sqrt{\alpha}}\,. \tag{96}$$

The overlap gap property (plus concentration inequality Lemma 2.33) implies that

$$\max_{\mathbf{x} \in U_n^{(\ell)}(S)} \frac{1}{\ell n}\mathcal{G}(\mathbf{x}) \le v + o_n(1)\,. \tag{97}$$

Hence (using concentration inequality Lemma 2.34), with high probability we also have

$$\max_{\mathbf{x}\in U_n^{(\ell)}(S)} \frac{1}{\ell n}\mathcal{H}(\mathbf{x}) \le \mathop{\mathbb{E}}_{f\sim\Lambda}[f] + \frac{v+2\varepsilon}{\sqrt{\alpha}}\,. \tag{98}$$

□

Analogously, we can transfer a generic version of the OGP on (A,b)-coupled models.

Definition 5.5 (Generic OGP). *For a relatively open subset $S \subseteq \mathcal{R}^{(\ell)}$, random Hamiltonians $H_1,\dots,H_\ell : \{\pm1\}^n \to \mathbb{R}$ exhibit an S-*OGP *with value v if*

$$\max_{(\sigma_1,\dots,\sigma_\ell)\in U_n^{(\ell)}(S)} \frac{1}{\ell n}\left(H_1(\sigma_1)+\cdots+H_\ell(\sigma_\ell)\right) \le v\,. \tag{99}$$

The size *of the OGP is ℓ.*

Proposition 5.6. *If an S-OGP with value v holds for an (A,b)-coupled model of SG_ξ with high probability, then for all $\varepsilon>0$, for all sufficiently large α, an S-OGP with value $\mathbb{E}_{f\sim\Lambda}[f]+\frac{v+\varepsilon}{\sqrt{\alpha}}$ holds for the (A,b)-coupled model of $\mathsf{CSP}_\Lambda(\alpha)$ with high probability, when ξ is related to Λ as in Equation (1).*

5.1 The branching OGP

The branching OGP gives *tight* algorithmic bounds for a class of spin glass models, matching the performance of certain approximate message passing algorithms [HS21]. We present the somewhat involved definition and show that it is captured by our framework.

Definition 5.7 (Tree-coupled ensemble). *A* tree-coupled ensemble of Hamiltonians *is defined by:*

1. *A rooted tree of height D defined by the vector $\vec{k}\in\mathbb{Z}_+^D$, so that every node at depth d has k_{d+1} children. We describe the nodes of the tree as $\mathbb{T}(\vec{k})$ and the leaves of the tree as $\mathbb{L}(\vec{k})$.*
2. *A coupling vector $\vec{p}\in\mathbb{Z}_+^{D+1}$, so that*

$$0 = p_0 \le p_1 \le \cdots \le p_D = 1\,. \tag{100}$$

Given these parameters, we generate a family of Hamiltonians $(H^{(u)})_{u\in\mathbb{L}(\vec{k})}$ as follows. Generate independent instances of the Hamiltonian $\widetilde{H}^{(v)}$ for every non-root $v\in\mathbb{T}(\vec{k})$, and scale[7] each by a factor of p_d-p_{d-1} where d is the depth of v in the tree. Use these to construct the Hamiltonians $H^{(u)}$ for each $u\in\mathbb{L}(\vec{k})$ defined by

$$H^{(u)} = \sum_{d=0}^{D-1} \widetilde{H}^{(a(u,d))}\,, \tag{101}$$

where $a(u,d)$ is the dth ancestor of u. The size *of the ensemble is $\left|\mathbb{L}(\vec{k})\right| = \prod_{i=1}^D \vec{k}_i$.*

Definition 5.8 (Branching OGP). *A tree-coupled ensemble of Hamiltonians exhibits a* branching OGP *with value v and gap η if:*

- *there is $\vec{q}\in\mathbb{Z}_+^{D+1}$ with $0\le q_0<q_1<\cdots<q_D=1$, and $\mathbf{m}\in\mathbb{R}^n$ with $R(\mathbf{m},\mathbf{m})=q_0$,*

[7] For the spin glass, scale the variance of each Gaussian by p_d-p_{d-1}; for the CSP, scale the number of clauses by p_d-p_{d-1}.

- *define* $Q \in \mathbb{R}^{\mathbb{L}(\vec{k})\times\mathbb{L}(\vec{k})}$ *by* $Q_{u,v} = q_{lca(u,v)}$, *where* $lca(u,v)$ *is the depth of the least common ancestor of* u *and* v,
- *define*

$$\mathcal{Q}(\eta) = \Big\{\vec{\sigma} \in (\{\pm 1\}^n)^{\mathbb{L}(\vec{k})} : \forall u, v \in \mathbb{L}(\vec{k}).\ \left|R(\sigma^{(u)}, \sigma^{(v)}) - Q_{u,v}\right| \le \eta \\ \text{and } \forall u \in \mathbb{L}(\vec{k}).\ \left|R(\sigma^{(u)}, \mathbf{m}) - q_0\right| \le \eta\Big\}$$

- *it holds that for all* $(\sigma^{(u)})_{u\in\mathbb{L}(\vec{k})} \in \mathcal{Q}(\eta)$, *the average value over* $u \in \mathbb{L}(\vec{k})$ *of* $H^{(u)}(\sigma^{(u)})$ *is at most* v.

The size *of the OGP is* $\left|\mathbb{L}(\vec{k})\right| = \prod_{i=1}^{D} \vec{k}_i$.

Remark 5.9. *Unlike [HS21], we assume that* $\mathbf{m} = \overline{m}\mathbb{1}^n$ *is a scalar multiple of the all-1s vector. Notice that* $R(\sigma, \overline{m}\,\mathbb{1}^n) = \overline{m}\sum \sigma_i/n$. *As a result, the requirement* $|R(\sigma^{(u)}, \overline{m}\,\mathbb{1}^n) - q_0| < \eta$ *constrains the magnetization (a.k.a the Hamming weight) of the* $\sigma^{(u)}$. *This loses no generality when the cost function is permutation-invariant.*

A tree-coupled ensemble is an instance of an (A, b)-coupled model, Definition 3.1, using the following choice of parameters:

- $\ell = |\mathbb{L}(\vec{k})|$.
- $\ell' = |\mathbb{T}(\vec{k})|$.
- $A_{u,v} = 1$ if $u = v$, or $v \in \mathbb{T}(\vec{k})$ is a non-root ancestor of $u \in \mathbb{L}(\vec{k})$; otherwise $A_{u,v} = 0$.
- $b_v = p_d - p_{d-1}$ for every $v \in \mathbb{T}(\vec{k})$ at the dth level of the tree.

The presence of the branching OGP is determined by the set

$$S = \{r \in \mathcal{R} : \forall u, v \in \mathbb{L}(\vec{k})\ .\ \left|r_{\{u,v\}} - Q_{u,v}\right| < \eta \text{ and } \forall u \in \mathbb{L}(\vec{k})\ .\ \left|r_{\{u\}}\overline{m} - q_0\right| < \eta\}\,. \tag{102}$$

Therefore, by Proposition 5.6, the branching OGP transfers from the spin glass to the CSP.

6 Implications for algorithmic hardness

We prove that the OGP has consequences for the performance of many Max-CSP algorithms, specifically those which are *overlap-concentrated.*

Definition 6.1 (t-correlated instances of a CSP)**.** *A* t-correlated *pair of instances* $\mathcal{I}_t^1$ *and* $\mathcal{I}_t^2$ *of* $\mathsf{CSP}_\Lambda(\alpha)$ *are created by:*

1. *Generate an instance of* $\mathsf{CSP}_\Lambda(\alpha)$ *with* $\mathrm{Pois}(\alpha t n)$ *shared clauses. Let* $\mathcal{I}_t^1$ *and* $\mathcal{I}_t^2$ *both be initially equal to this instance.*
2. *Independently generate two additional sets of* $\mathrm{Pois}((1-t)\alpha n)$ *clauses. Add the first set of clauses to* $\mathcal{I}_t^1$ *and the second set to* $\mathcal{I}_t^2$.

Definition 6.2 (Overlap-concentrated algorithm)**.** *A (possibly randomized) algorithm* $\mathcal{A}$ *is* (λ, μ)-overlap-concentrated *if for every* $t \in [0, 1]$,

$$\Pr\left[\left|R(\mathcal{A}(H_t^1), \mathcal{A}(H_t^1)) - \mathbb{E}\, R(\mathcal{A}(H_t^1), \mathcal{A}(H_t^2))\right| \ge \lambda\right] \le \mu \tag{103}$$

where H_t^1, H_t^2 are t-correlated instances.

Definition 6.2 is a modification of [HS21, Definition 2.1] for Max-CSPs.

When the cost function is permutation-invariant, no algorithm is better than a *permutation-invariant* algorithm. We thus focus on obstructions to permutation-invariant algorithms.[8]

Observation 6.3. *Let H be a random Hamiltonian sampled from a permutation-invariant distribution, and $\mathcal{A}$ be an algorithm. Define a new algorithm $\mathcal{A}'$ that randomly permutes the input, then runs $\mathcal{A}$. Then $\mathcal{A}'$ is permutation-invariant and achieves the same performance as $\mathcal{A}$.*

Fix an algorithm $\mathcal{A}$. For $t \in [0,1]$, let H_t^1 and H_t^2 be a t-correlated pair of Hamiltonians for $\mathsf{CSP}_\Lambda(\alpha)$. We define the *correlation function* $\chi : [0,1] \to \mathbb{R}$ for $\mathcal{A}$ by

$$\chi(t) = \mathbb{E}\left[R(\mathcal{A}(H_t^1), \mathcal{A}(H_t^2))\right]. \tag{104}$$

The correlation function χ satisfies the exact same properties as in [HS21, Proposition 3.1].

Proposition 6.4 (Poisson equivalent of [HS21, Proposition 3.1])**.** *For any algorithm $\mathcal{A}$ run on an instance of $\mathsf{CSP}_\Lambda(\alpha)$, $\chi : [0,1] \to [-1,1]$ satisfies*

1. *χ is continuous,*
2. *χ is either constant, or strictly increasing,*
3. *for all $t \in [0,1]$, $\chi(t) \le (1-t)\cdot\chi(0) + t\cdot\chi(1)$.*

Proof sketch. Unlike the Gaussian case, we need to use the general Efron-Stein decomposition (instead of the Hermite polynomials). Notice that conditioning on the outcomes of a Poisson random variable results in an appropriately parameterized binomial distribution. See Appendix A for a full proof. □

Definition 6.5 (Correlation function)**.** *A function $\chi : [0,1] \to [-1,+1]$ that satisfies the properties of Proposition 6.4 is called a* correlation function.

Definition 6.6 (η-forbid)**.** *We use three versions of this statement:*

- *An OGP with gap (s,t) η-forbids an algorithm $\mathcal{A}$ if $\mathbb{E}\, R(\mathcal{A}(H_1), \mathcal{A}(H_2))$ is in the η-interior of the interval (s,t) for H_1 and H_2 a pair of r-correlated instances for some $r \in [0,1]$. In other words, the open radius-η ball around the point is contained in (s,t).*
- *An S-OGP η-forbids an algorithm $\mathcal{A}$ if $\mathbb{E}\, Q(\mathcal{A}(H_1), \dots, \mathcal{A}(H_\ell))$ is in the η-interior of S.*
- *An S-OGP η-forbids a correlation function χ if $S \subseteq [-1,1]^{\binom{\ell}{2}}$ only depends on pairwise overlaps, and $(\chi(\rho_{i,j}))_{i,j}$ is in the η-interior of S where $\rho_{i,j} = \mathbb{E}\, H_i(\sigma)H_j(\sigma)$ is the correlation of Hamiltonians H_i and H_j.*

[8] This is also why we lose no generality with Remark 5.9. When proving algorithmic hardness, we set the parameter $\mathbf{m}$ in the method of [HS21, Hua22] to the expected value of the output of a candidate algorithm. For permutation-invariant algorithms, this sets $n\overline{m}$ to the Hamming weight of the algorithm's expected output.

Theorem 6.7 ([HS21, Proposition 3.2]). *For all even functions ξ and $v > ALG$, there is $\eta > 0$ and an integer ℓ such that for all correlation functions χ there is a coupled model for SG_ξ on at most ℓ Hamiltonians and a set S η-forbidding χ such that with high probability the coupled model exhibits an S-OGP with value v. Specifically, the coupled model is a tree-coupled ensemble and the OGP is a branching OGP.*

We now show that the presence of a branching OGP on a random Max-CSP upper-bounds the performance of any overlap-concentrated algorithm.

Theorem 6.8 (Formal version of Theorem 1.4). *Suppose that $\mathsf{CSP}_\Lambda(\alpha)$ has a size-ℓ branching OGP of value $\mathbb{E}_{f\sim\Lambda}[f] + \frac{v^*}{\sqrt{\alpha}}$ with high probability. Suppose that an (λ,μ)-overlap-concentrated algorithm $\mathcal{A}$ is η-forbidden by the OGP, and $\lambda \le \eta$. Fix $\varepsilon > 0$. Then if α is sufficiently large, the probability that $\mathcal{A}$ satisfies at least $\mathbb{E}_{f\sim\Lambda}[f] + \frac{v^*+\varepsilon}{\sqrt{\alpha}}$ proportion of clauses is at most*

$$\exp(-cn) + 3(\mu/\lambda)^c, \tag{105}$$

where $c > 0$ depends only on α, Λ, ε, and ℓ.

This proof is essentially contained in [HS21, Section 3.4]. We extract the proof here. Note that [HS21] obstruct a broader class of algorithms that can output solutions in the convex hull of the solution space; since we focus on algorithms that output valid values, we do not use the technical considerations of [HS21, Section 3.3].

Proof. Consider running algorithm $\mathcal{A}$ on each of the correlated instances $H^{(u)}$ in the OGP. Let $\mathcal{E}$ be the event that the average value of the output solutions is at least $v + \varepsilon$. On the one hand,

$$\Pr[\mathcal{E}] \ge p^\ell \tag{106}$$

where p is the probability that $\mathcal{A}$ outputs a solution of value at least $v + \varepsilon$ when run on a single instance [HS21, Proof of Proposition 3.6(a)].

On the other hand, $\mathcal{E}$ is unlikely, because the expected output of $\mathcal{A}$ lies in the forbidden region of the OGP and therefore $\mathcal{A}$ is likely to output solutions in the forbidden region since it's overlap-concentrated. Specifically, by a union bound,

$$\Pr[\mathcal{E}] \le (1 - \Pr[\text{OGP occurs}]) + \left(1 - \Pr\left[\mathcal{A}(H^{(u)}) \text{ in forbidden region}\right]\right). \tag{107}$$

Because of concentration, the OGP occurs with all but exponentially small probability: $\Pr[\text{OGP occurs}] \ge 1 - \exp(-c'n)$ for c' depending only on ℓ and Λ. This follows [HS21, Proof of Proposition 3.6(d)]), but with Lemma 2.34 replacing [HS21, Proposition 3.8].

The probability that $\mathcal{A}(H^{(u)})$ lies outside the forbidden region of the branching OGP can be union-bounded across the $\binom{\ell}{2}$ pairwise overlaps and the ℓ magnetization parameters. In order to lie outside the region, the deviation from the expectation must be at least η in one of the coordinates. By overlap concentration (and the setting $\eta \le \lambda$), the deviation probability is at most $\ell \cdot \frac{2\mu}{\lambda}$ for the magnetizations [HS21, Lemma 3.7] and $\binom{\ell}{2}\mu$ for the pairwise overlaps. We conclude

$$\Pr[\mathcal{E}] \le \exp(-c'n) + \binom{\ell}{2}\mu + \ell \cdot \frac{2\mu}{\lambda}. \tag{108}$$

Combining Equation (106) and Equation (108) and taking the ℓ-th root yields the claim. □

A branching OGP is known to exist at a suboptimal value for a large class of mixed spin glasses [HS21]. Using Proposition 5.6 and Theorem 6.8, we conclude the following:

Corollary 6.9 (Formal version of Corollary 1.5)**.** *Let $\mathcal{I}$ be an instance of $\mathsf{CSP}_\Lambda(\alpha)$ such that for every predicate f in the support of Λ, the only nonzero Fourier coefficients of f have even degree $j \geq 4$. There are constants $\alpha_0, \varepsilon, c, \lambda_0 > 0$ such that if $\alpha \geq \alpha_0$ and $\mathcal{A}$ is an (λ, μ)-overlap-concentrated algorithm with $\lambda \leq \lambda_0$, then*

$$\Pr\left[\frac{1}{n}H^\alpha(\mathcal{A}(\mathcal{I})) \geq v_{\mathcal{I}} - \frac{\varepsilon}{\sqrt{\alpha}}\right] \leq \exp(-cn) + 3(\mu/\lambda)^c\,. \tag{109}$$

7 Discussion

We establish a formal average-case link between spin glasses and Max-CSPs of certain minimum clause densities. This link shows an equivalence of optimal value, overlap gaps, and hardness for a large class of algorithms. Curiously, every Max-CSP with the same noise stability polynomial is linked to the same spin glass.

As part of this work, we extend the list of Max-CSPs known to have an OGP. It is an open question to completely classify which spin glasses (and thus which Max-CSPs) have an OGP. In the spherical spin glass setting [Sub18, Proposition 1], the weight of the quadratic terms determine the presence of an OGP; the same may be true on the hypercube. There is a technical hurdle to proving the existence of OGPs on spin glasses when the mixture polynomial is not even. For example, the associated spin glass to Max-3XOR is not proven to have an OGP, although it is expected to have one [AMS20].

It is also possible that the onset of the branching OGP of [HS21] marks the hardness threshold for *all* efficient algorithms for spin glasses and Max-CSPs, and not just overlap-concentrated algorithms. Furthermore, it is possible that a single algorithm (namely, suitably-applied message-passing) is the optimal algorithm. This is remarkably similar to the situation for worst-case analysis vis-à-vis the *Unique Games Conjecture (UGC)* [Kho02]. The UGC implies that the standard SDP is the optimal approximation algorithm for any CSP [Rag08]. It is interesting to investigate whether other algorithms, such as Sum-of-Squares algorithms, can be designed in a way that has equivalent performance to approximate message-passing, although there are existing certification lower bounds for a variety of models [BGL17, GJJ+20]. Our curiosity is heightened by the observation that the Parisi formula, and algorithms for spin glasses (and hence average-case CSPs), only use the "degree-2" part of the overlap distribution on $\mathcal{R}^{(\ell)}$, analogous to how the basic SDP is also a "degree-2" algorithm.

How general is overlap concentration? [HS21] show that Langevin dynamics and certain families of approximate message-passing algorithms are overlap-concentrated on spin glasses, and therefore obstructed in the presence of a branching OGP to a constant given by the extended Parisi formula. Algorithms representable as low-degree polynomials are known to be *stable* on spin glasses, and obstructed for some spin glasses [GJW20] and CSPs [BH22]. However, it remains a technical challenge to show that low-degree polynomials are overlap-concentrated:

Conjecture 7.1 (Low-degree polynomials are overlap-concentrated on CSPs and spin glasses)**.** *Any low-degree polynomial algorithm that solves a typical instance of $\mathsf{CSP}_\Lambda(\alpha)$ or SG_ξ as defined in [GJW20, Definition 2.3] is overlap-concentrated.*

There may be other forms of equivalence between spin glasses and Max-CSPs. We conjecture that the two models have the same *distribution* of overlap vectors:

Conjecture 7.2 (Equivalence of distribution of overlap vector, informal)**.** *Take any Max-CSP and consider the associated spin glass* SG_ξ*, where* ξ *is defined as in Equation (1). For any coupled ensemble of the models, the distribution of* $Q(\sigma_1, \ldots, \sigma_\ell)$ *for random near-optimal solutions converges in some sense as* $n \to \infty$*. Furthermore, for* $\alpha \to \infty$ *the distribution for the CSP converges to the distribution for the spin glass.*

It is also likely that algorithms beyond the QAOA have identical average-case performance on a random instance of Max-CSP and that of the corresponding spin glass. In fact, we suspect that every Max-CSP has an optimal algorithm related to message-passing that obeys this equivalence.

Spin glasses may also be related to more classes of CSPs, such as those with non-Boolean inputs and those without random literal signs. Also of interest is the problem of refuting the CSP when it is unsatisfiable, which typically requires α superconstant. It is not known how to use statistical physics methods to study refutation [HMX21]. It may also be possible to determine the optimal average-case value to higher-order terms in α.

Acknowledgements

We thank Antares Chen for collaborating during the early stages of this project.

We also thank Peter J. Love for helpful feedback on a previous version of this manuscript.

KM thanks Ryan Robinett for a tip on improving the numerical integration scheme. JSS did some of this work as a visiting student at Bocconi University.

References

[AC17] Antonio Auffinger and Wei-Kuo Chen. Parisi formula for the ground state energy in the mixed p-spin model. *The Annals of Probability*, 45(6B):4617–4631, 2017. i, 19

[ACO08] Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 793–802. IEEE, 2008. 3

[ACZ20] Antonio Auffinger, Wei-Kuo Chen, and Qiang Zeng. The SK model is infinite step replica symmetry breaking at zero temperature. *Communications on Pure and Applied Mathematics*, 73(5), 2020. 3, 4

[AM02] Dimitris Achlioptas and Cristopher Moore. The asymptotic order of the random k-SAT threshold. In *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.*, pages 779–788. IEEE, 2002. 4

[AM20] Ahmed El Alaoui and Andrea Montanari. Algorithmic Thresholds in Mean Field Spin Glasses. *arXiv preprint arXiv:2009.11481*, 2020. 4

[AMS20] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *arXiv preprint arXiv:2001.00904*, 2020. 1, 3, 4, 19, 25

[AMS21] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Local algorithms for Maximum Cut and Minimum Bisection on locally treelike regular graphs of large degree. *arXiv preprint arXiv:2111.06813*, 2021. 3, 4

[ART06] Dimitris Achlioptas and Federico Ricci-Tersenghi. On the solution-space geometry of random constraint satisfaction problems. In *Proceedings of the thirty-eighth annual ACM Symposium on Theory of Computing*, pages 130–139, 2006. 4

[AWZ20] Gérard Ben Arous, Alexander S Wein, and Ilias Zadik. Free Energy Wells and Overlap Gap Property in Sparse PCA. In *Conference on Learning Theory*, pages 479–482. PMLR, 2020. 4

[BFM+22] Joao Basso, Edward Farhi, Kunal Marwaha, Benjamin Villalonga, and Leo Zhou. The Quantum Approximate Optimization Algorithm at High Depth for MaxCut on Large-Girth Regular Graphs and the Sherrington-Kirkpatrick Model. In *17th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2022)*, pages 7:1–7:21, 2022. 5

[BGL17] Vijay Bhattiprolu, Venkatesan Guruswami, and Euiwoong Lee. Sum-of-Squares Certificates for Maxima of Random Tensors on the Sphere. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2017)*, pages 31:1–31:20, 2017. 25

[BGMZ22] Joao Basso, David Gamarnik, Song Mei, and Leo Zhou. Performance and limitations of the QAOA at constant levels on large sparse hypergraphs and spin glass models. *arXiv preprint arXiv:2204.10306*, 2022. 5

[BH22] Guy Bresler and Brice Huang. The Algorithmic Phase Transition of Random k-SAT for Low Degree Polynomials. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 298–309. IEEE, 2022. 3, 4, 25

[BM21] Sami Boulebnane and Ashley Montanaro. Predicting parameters for the Quantum Approximate Optimization Algorithm for MAX-CUT from the infinite-size limit. *arXiv preprint arXiv:2110.10685*, 2021. 5

[BMZ05] Alfredo Braunstein, Marc Mézard, and Riccardo Zecchina. Survey propagation: An algorithm for satisfiability. *Random Structures & Algorithms*, 27(2):201–226, 2005. 3

[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. 4, 7, 11, 20

[CLSS22] Chi-Ning Chou, Peter J. Love, Juspreet Singh Sandhu, and Jonathan Shi. Limitations of Local Quantum Algorithms on Random MAX-k-XOR and Beyond. In *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*, pages 41:1–41:20, 2022. i, 3, 4, 11, 20

[CR02] A. Crisanti and T. Rizzo. Analysis of the ∞-replica symmetry breaking solution of the Sherrington-Kirkpatrick model. *Physical Review E*, 65(4), Apr 2002. 4

[CvD21] Jahan Claes and Wim van Dam. Instance Independence of Single Layer Quantum Approximate Optimization Algorithm on Mixed-Spin Models at Infinite Size. *Quantum*, 5:542, Sep 2021. 5

[DMS17] Amir Dembo, Andrea Montanari, and Subhabrata Sen. Extremal cuts of sparse random graphs. *The Annals of Probability*, 45(2):1190–1217, 2017. 1, 2, 4

[DSS14] Jian Ding, Allan Sly, and Nike Sun. Satisfiability threshold for random regular NAE-SAT. In *Proceedings of the forty-sixth annual ACM Symposium on Theory of Computing*, pages 814–822, 2014. 1, 4

[DSS16] Jian Ding, Allan Sly, and Nike Sun. Maximum independent sets on random regular graphs. *Acta Mathematica*, 217(2):263–340, 2016. 1, 4

[DSS22] Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. *Annals of Mathematics*, 196(1):1–388, 2022. 1, 4

[FA86] Y Fu and P W Anderson. Application of statistical mechanics to NP-complete problems in combinatorial optimisation. *Journal of Physics A: Mathematical and General*, 19(9):1605–1620, Jun 1986. 4

[FGG14] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014. 5

[FGGZ22] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Leo Zhou. The Quantum Approximate Optimization Algorithm and the Sherrington-Kirkpatrick Model at Infinite Size. *Quantum*, 6:759, July 2022. 5

[Gam21] David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41):e2108492118, 2021. 1, 3, 4, 20

[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. 1, 4

[GJJ+20] Mrinalkanti Ghosh, Fernando Granha Jeronimo, Chris Jones, Aaron Potechin, and Goutham Rajendran. Sum-of-Squares Lower Bounds for Sherrington-Kirkpatrick via Planted Affine Planes. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 954–965. IEEE, 2020. 25

[GJW20] David Gamarnik, Aukosh Jagannath, and Alexander S Wein. Low-degree hardness of random optimization problems. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 131–140. IEEE, 2020. 1, 3, 4, 25

[GL18] David Gamarnik and Quan Li. Finding a large submatrix of a Gaussian random matrix. *The Annals of Statistics*, 46(6A):2511–2561, 2018. 4

[GS14] David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *Proceedings of the 5th conference on Innovations in Theoretical Computer Science*, pages 369–376, 2014. 4

[GT04] Francesco Guerra and Fabio Lucio Toninelli. The high temperature region of the Viana–Bray diluted spin glass model. *Journal of Statistical Physics*, 115(1):531–555, 2004. 4

[Gue03] Francesco Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Communications in Mathematical Physics*, 233(1):1–12, 2003. 4, 6

[HMX21] Jun-Ting Hsieh, Sidhanth Mohanty, and Jeff Xu. Certifying solution geometry in random CSPs: counts, clusters and balance. *arXiv preprint arXiv:2106.12710*, 2021. 26

[HS21] Brice Huang and Mark Sellke. Tight Lipschitz hardness for optimizing mean field spin glasses. *arXiv preprint arXiv:2110.07847*, 2021. i, 1, 2, 3, 4, 7, 20, 21, 22, 23, 24, 25, 31, 32

[Hua22] Brice Huang. *Computational Hardness in Random Optimization Problems from the Overlap Gap Property*. PhD thesis, Massachusetts Institute of Technology, 2022. 23

[JS20] Aukosh Jagannath and Subhabrata Sen. On the unbalanced cut problem and the generalized Sherrington–Kirkpatrick model. *Annales de l'Institut Henri Poincaré D*, 8(1):35–88, 2020. 11, 20

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the thiry-fourth annual ACM Symposium on Theory of Computing*, pages 767–775, 2002. 25

[KLLM21] Peter Keevash, Noam Lifshitz, Eoin Long, and Dor Minzer. Global hypercontractivity and its applications. *arXiv preprint arXiv:2103.04604*, 2021. 31

[MH22] Kunal Marwaha and Stuart Hadfield. Bounds on approximating Max kXOR with quantum and classical local algorithms. *Quantum*, 6:757, 2022. 4, 19

[Mon19] A. Montanari. Optimization of the Sherrington-Kirkpatrick Hamiltonian. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1417–1433, Los Alamitos, CA, USA, nov 2019. IEEE Computer Society. 3

[MPV87] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987. 4

[O'D14] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014. 9, 10, 31

[Pan05] Dmitry Panchenko. Free energy in the generalized Sherrington–Kirkpatrick mean field model. *Reviews in Mathematical Physics*, 17(07):793–857, 2005. 11

[Pan13] Dmitry Panchenko. *The Sherrington-Kirkpatrick model*. Springer Science & Business Media, 2013. 4, 6, 8, 11, 16

[Pan14] Dmitry Panchenko. The Parisi formula for mixed p-spin models. *The Annals of Probability*, 42(3):946–958, 2014. 6

[Pan18] Dmitry Panchenko. On the K-sat model with large number of clauses. *Random Structures & Algorithms*, 52(3):536–542, 2018. 2, 4, 11

[Par80] Giorgio Parisi. A sequence of approximated solutions to the SK model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115, 1980. i, 4, 6, 19

[Pas13] Vangelis Th Paschos. *Paradigms of Combinatorial Optimization: Problems and New Approaches*, volume 2. John Wiley & Sons, 2013. 1

[PT04] Dmitry Panchenko and Michel Talagrand. Bounds for diluted mean-fields spin glass models. *Probability Theory and Related Fields*, 130(3):319–336, 2004. 4

[Rag08] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the fortieth annual ACM Symposium on Theory of Computing*, pages 245–254, 2008. 25

[RV17] Mustazee Rahman and Balint Virag. Local algorithms for independent sets are half-optimal. *The Annals of Probability*, 45(3):1543–1577, 2017. 4

[Sel21] Mark Sellke. Optimizing Mean Field Spin Glasses with External Field. *arXiv preprint arXiv:2105.03506*, 2021. 3, 4

[Sen18] Subhabrata Sen. Optimization on sparse random hypergraphs and spin glasses. *Random Structures & Algorithms*, 53(3):504–536, 2018. 1, 2, 4

[SK75] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35(26):1792, 1975. 1, 4

[Sub18] Eliran Subag. Following the ground-states of full-RSB spherical spin glasses, 2018. 25

[Tal06] Michel Talagrand. The parisi formula. *Annals of Mathematics*, pages 221–263, 2006. i, 4, 6

[Zha21] Yu Zhao. *Generalizations and Applications of Hypercontractivity and Small-Set Expansion.* PhD thesis, Carnegie Mellon University, 2021. 31

Appendix

A Correlation functions of sparse Hamiltonians

In this section, we prove Proposition 6.4. The proof follows [HS21, Proof of Proposition 3.1]. We must use the more general Efron-Stein basis and a mild generalization of the noise operator defined in Definition 2.27.

Definition A.1 (Efron-Stein decomposition)**.** *For a function $f \in L^2(\Omega^m, \pi^{\otimes m})$, we define $f^{\subseteq S}(x) = \mathbb{E}_{x_{\overline{S}}}[f(x) \mid x_S]$, where the expectation means that the variables x_i for $i \in \overline{S}$ are independently resampled from π, and $f^{=S}(x) = \sum_{J \subseteq S} (-1)^{|S|-|J|} f^{\subseteq J}(x)$.*

A core property of this decomposition, as stated in [KLLM21, Section II.2.2], is that $f^{=S}$ depends only on x_i for $i \in S$, and furthermore $\left\langle f^{=S}, g \right\rangle = 0$ for any g that does not depend on all x_i for $i \in S$. Additionally, by [O'D14, Theorem 8.35 and Proposition 8.36],

$$f = \sum_{S \subseteq [m]} f^{=S} = \sum_{S \subseteq [m]} \left(\sum_{\alpha \in \mathbb{N}^m_{|\Omega|},\, supp(\alpha) = S} \hat{f}(\alpha) \phi_\alpha \right), \tag{110}$$

where $\mathbb{N}^m_{|\Omega|} = \{0, \ldots, |\Omega| - 1\}^m$ and,

$$\phi_\alpha = \prod_{j=1}^{m} \phi_{\alpha_i}, \tag{111}$$

with $\phi_0 = \mathsf{Id}$.

This decomposition implies that any $\mathcal{A} : \Omega^m \to [-1, 1]^n$ under the input measure $\pi^{\otimes m}$ can be written as

$$\mathcal{A} = (f_1, \ldots, f_n), \tag{112}$$

where each f_i can be described as in Definition A.1.

We now define the noise operator on a domain Ω^m with an arbitrary product measure $\pi^{\otimes m}$:

Definition A.2 (Noise operator, [KLLM21, Zha21])**.** *Given a probability space (Ω, π) and a noise parameter $\rho \in [0, 1]$, define the* noise operator *T_ρ, which acts on a function $f \in L_2(\Omega^m, \pi^{\otimes m})$ as*

$$T_\rho[f](\sigma_1, \ldots, \sigma_m) = \mathop{\mathbb{E}}_{\tau \sim N_\rho(\sigma)} \Big[f(\tau_1, \ldots, \tau_m) \Big], \tag{113}$$

where the distribution $N_\rho(\sigma)$ chooses each coordinate τ_i independently as

$$\begin{cases} \tau_i = \sigma_i & \text{with probability } \rho, \\ \tau_i \sim \pi & \text{with probability } 1 - \rho. \end{cases} \tag{114}$$

Proof of Proposition 6.4. Define the Fourier weight on a level j as

$$W_j = \frac{1}{n} \sum_{i=1}^{n} \sum_{S \subseteq [m], |S| = j} (f_i^{=S})^2. \tag{115}$$

Then, similarly to [HS21, Proof of Proposition 3.1], when x, y are generated as in Definition 6.1 with $t = p$,

$$\begin{aligned}\chi(p) &= \mathbb{E}[R(\mathcal{A}(x), \mathcal{A}(y))] = \frac{1}{n}\mathbb{E}[\langle \mathcal{A}(x), \mathcal{A}(y)\rangle] = \frac{1}{n}\mathbb{E}[\langle \mathcal{A}(x), T_p[\mathcal{A}](x)\rangle] \\ &= \frac{1}{n}\mathbb{E}[\langle [T_{\sqrt{p}}A](x), [T_{\sqrt{p}}A](x)\rangle] = \frac{1}{n}\sum_{i=1}^{n}\sum_{S\subseteq[m]} p^{|S|}(f_i^{=S})^2 = \sum_{j\geq 0} p^{|S|}W_j\,. \end{aligned} \tag{116}$$

This follows for two reasons. First, the noise operator acts on a $p \cdot m$-sized random subset of $[m]$, equivalent to the (average) number of shared clauses generated by Definition 6.1. Second, the semi-group action of $T_p = T_{\sqrt{p}} \circ T_{\sqrt{p}}$. The desired properties for $\chi(p)$ now follow for the same reasons as [HS21, Proof of Proposition 3.1]. □