Local Algorithms and the Failure of Log-Depth Quantum Advantage on Sparse Random CSPs

Antares Chen
University of Chicago, IL, USA

Neng Huang
University of Michigan, Ann Arbor, MI, USA

Kunal Marwaha
University of Chicago, IL, USA

Abstract

We construct and analyze a message-passing algorithm for random constraint satisfaction problems (CSPs) at large clause density, generalizing work of El Alaoui, Montanari, and Sellke for Maximum Cut [3] through a connection between random CSPs and mean-field Ising spin glasses [1, 55]. For CSPs with even predicates, the algorithm asymptotically solves a stochastic optimal control problem dual to an extended Parisi variational principle. This gives an optimal fraction of satisfied constraints among algorithms obstructed by the branching overlap gap property of Huang and Sellke [50], notably including the Quantum Approximate Optimization Algorithm and all quantum circuits on a bounded-degree architecture of up to $\epsilon \cdot \log n$ depth [24].

2012 ACM Subject Classification Theory of computation → Randomness, geometry and discrete structures; Theory of computation → Design and analysis of algorithms

Keywords and phrases Random CSP, message-passing algorithm

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2026.78

Category RANDOM

Related Version *Full Version*: `https://arxiv.org/abs/2310.01563`

1 Introduction

Today's quantum computers are starting to show signs of advantage over classical machines. Many people are seeking to use this technology in a practical application as soon as possible. One such area is combinatorial optimization: even though the problems are NP-hard, optimization has incredible value to the contemporary world. This industry is also eager to experiment with quantum computers, perhaps because rigorous worst-case analysis is not always relevant.

The most popular quantum protocol for classical optimization is the *Quantum Approximate Optimization Algorithm* (QAOA), proposed in [34]. Given an optimization problem, the protocol associates a qubit to each variable, and evolves a quantum state to a near-ground state of the problem, depending on choice of hyperparameters. The QAOA has also sparked the interest of theorists, characterizing the protocol's performance on canonical problems, and debating whether it is any more powerful than existing classical algorithms (e.g. [9, 45, 11]).

One key property of the QAOA is its *locality*: at each layer of the algorithm, a qubit interacts only with qubits associated with adjacent variables in the optimization problem. Locality is known to limit the QAOA up to logarithmic circuit depth when the optimization problem is *sparse* [32, 33]. This limitation is due to a phenomenon known as the *overlap gap property* (OGP), a type of clustering behavior of near-optimal solutions [44]. Notably, this is an *average-case* property of many random optimization problems. Recent formulations of the OGP have led to quantitative obstructions for broad families of algorithms, both quantum and classical [21, 24, 50].

In this work, we show that *with high probability* over a large family of sparse optimization problems, there is an *optimal* algorithm among those obstructed by the OGP. Furthermore, this algorithm is *classical*, so there can be no quantum advantage by the QAOA up to logarithmic depth. In fact, this no-go result extends to logarithmic-depth quantum circuits on fixed, bounded-degree architectures, such as the *brickwork* circuits [18] used in random circuit sampling [15]. This is the strongest *average-case* result known to the authors regarding quantum advantage in classical optimization.

Our algorithm is related to the theory of *message-passing* used in random inference problems [13], and inspired by rich connections to statistical physics models such as the mean-field spin glass [82]. The parameters of the algorithm are chosen by solving a stochastic control problem dual to a problem determining the quantitative barrier of the OGP, as pioneered in [1, 50]. In particular, our algorithm extends work of [3], and uses a sparse-to-dense connection observed in e.g. [26, 55]. The performance of the algorithm, like that of [68, 1, 3, 79], is dependent on a technical condition (Assumption 1) which we explain in the text.

1.1 Background

We study Constraint Satisfaction Problems (CSPs). A CSP instance consists of n variables and a set of constraints associated with a *predicate* $f : \{\pm 1\}^r \to \{0, 1\}$.

▶ **Definition 1** (Instance of a random CSP)**.** *Consider a function $f : \{\pm 1\}^r \to \{0, 1\}$. Fix a constant $\alpha > 0$. A* random CSP instance $\mathcal{I}$ *over n variables $\{x_i\}_{i\in[n]}$ and $m = \alpha \cdot n$ clauses is generated by the following procedure: For each $i = 1, \dots, m$, draw $i_1, \dots, i_r$ uniformly i.i.d from $[n]$, draw r random signs $\epsilon_{i_1}, \dots, \epsilon_{i_r}$ uniformly i.i.d from $\{\pm 1\}$, and add the constraint e to $E(\mathcal{I})$, where e describes the clause $f_e(x_e) \stackrel{\text{def}}{=} f(\epsilon_{i_1} x_{i_1}, \dots, \epsilon_{i_r} x_{i_r})$. A clause is satisfied if it evaluates to* 1.

We denote this random model as $\mathsf{CSP}_f(\alpha)$. We are interested in the best *satisfying fraction* (i.e. largest fraction of clauses satisfied by some assignment in $\{\pm 1\}^n$) an algorithm can achieve as $n \to \infty$ when α is a very large constant, i.e. so that instances of $\mathsf{CSP}_f(\alpha)$ are typically unsatisfiable. This value is related to the rich literature of *mean-field Ising spin glasses*. We summarize some relevant results, especially drawing on [1, 3, 55]. Let $\mathscr{U}$ denote the following set of functions:

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

Given a univariate polynomial $\xi : \mathbb{R} \to \mathbb{R}$ with non-negative coefficients, define the functional $\mathsf{P}_\xi : \mathscr{U} \to \mathbb{R}$ given by

$$\mathsf{P}_\xi(\mu) \stackrel{\text{def}}{=} \Phi^\mu(0, 0) - \frac{1}{2} \cdot \int_0^1 \xi''(t) \cdot t \cdot \mu(t)\, dt\,,$$

where $\Phi^\mu : [0, 1) \times \mathbb{R} \to \mathbb{R}$ is a solution of the following partial differential equation:[1]

$$\begin{aligned} &\Phi^\mu_t(t, x) = -\frac{\xi''(t)}{2} \cdot \left(\Phi^\mu_{xx}(t, x) + \mu(t) \cdot \left(\Phi^\mu_x(t, x) \right)^2 \right) && \forall t \in [0, 1), \forall x \in \mathbb{R}\,, \\ &\Phi^\mu(1, x) = |x| && \forall x \in \mathbb{R}\,. \end{aligned} \tag{1.1}$$

P_ξ is known as the Parisi functional, while Equation (1.1) are known as the Parisi equations [73]. Let the infimum of P_ξ be defined as $\mathsf{OPT}_\xi \stackrel{\text{def}}{=} \inf_{\mu\in\mathscr{U}} \mathsf{P}_\xi(\mu)$.

[1] For the multivariable function $\Phi^\mu(t, x)$, we use subscripts to denote partial derivatives.

▶ **Theorem 2** ([6, 53]). *Consider any polynomial ξ with non-negative coefficients. Then OPT_ξ is achieved by some unique $\mu \in \mathcal{U}$.*

Surprisingly, the maximum satisfying fraction of a typical instance $\mathcal{I} \sim \mathsf{CSP}_f(\alpha)$ is related to OPT_ξ:

▶ **Theorem 3** ([55, Corollary 1.2], building on [73, 85, 70, 71, 26, 80, 72]). *Consider $\mathcal{I} \sim \mathsf{CSP}_f(\alpha)$. With high probability as $n \to \infty$, the maximum satisfying fraction of $\mathcal{I}$ is*[2]

$$\mathbb{E}[f] + \frac{\mathsf{OPT}_\xi}{\sqrt{\alpha}} + o_\alpha\left(\frac{1}{\sqrt{\alpha}}\right),$$

where $\xi(s) \stackrel{\text{def}}{=} \sum_{j=1}^{r} \|f^{=j}\|^2 s^j$, and $\|f^{=j}\|^2$ is the Fourier weight of f at degree j.

We also know that some algorithms are quantitatively obstructed due to the geometry of the solution space.

▶ **Definition 4** (Overlap-concentrated algorithm, [50, Definition 2.1]). *A deterministic algorithm $\mathcal{A}$ is* overlap-concentrated *with respect to a random family of problems $\mathcal{F}$ if the following is true for every $t \in [0,1]$ and constant $\delta > 0$: Draw two instances $\mathcal{I}_1, \mathcal{I}_2 \sim \mathcal{F}$ that are t-correlated.*[3] *Let $\mathcal{A}(\mathcal{I})$ be the assignment in $\{\pm 1\}^n$ that $\mathcal{A}$ generates on instance $\mathcal{I}$. Then the event*

$$\frac{1}{n}\left|\langle \mathcal{A}(\mathcal{I}_1), \mathcal{A}(\mathcal{I}_2)\rangle - \mathop{\mathbb{E}}_{\mathcal{I}_1,\mathcal{I}_2}[\langle \mathcal{A}(\mathcal{I}_1), \mathcal{A}(\mathcal{I}_2)\rangle]\right| \le \delta \tag{1.2}$$

occurs with high probability over choice of $\mathcal{I}_1, \mathcal{I}_2$.

For example, *local* algorithms in the factors of i.i.d. model and certain quantum circuits up to logarithmic depth [55] are overlap-concentrated on $\mathsf{CSP}_f(\alpha)$. A randomized algorithm is overlap-concentrated if the w.h.p. statement and the expectation in Equation (1.2) also vary over the algorithm's internal randomness.

We briefly explain the intuition of why Definition 4 prevents algorithms from succeeding on certain optimization problems; see [40] for more detail. Suppose the optimal assignments are *clustered*; i.e. they are "needles in the haystack". Then random instances will have "needles" in random locations in the solution space. But for $t \in [0,1]$, consider instances $\mathcal{I}_2(t)$ that are t-correlated with the initial instance $\mathcal{I}_1$. If the algorithm $\mathcal{A}$ is successful, the correlation $\frac{1}{n}\mathbb{E}_{\mathcal{I}_2(t)}[\langle \mathcal{A}(\mathcal{I}_1), \mathcal{A}(\mathcal{I}_2(t))\rangle]$ measures the correlation of these "needles" (i.e. good assignments), and so it varies smoothly from 0 to 1 as t goes from 0 to 1. The *overlap-concentrated* property is a bound on the variance of this quantity, meaning that with high probability over $\{\mathcal{I}_2(t)\}_{t\in[0,1]}$, the quantity $\frac{1}{n}\langle \mathcal{A}(\mathcal{I}_1), \mathcal{A}(\mathcal{I}_2(t))\rangle$ varies smoothly with t from 0 to 1. But this is not possible on problems that exhibit an *overlap gap property* (OGP)[4] across pairs of instances; so $\mathcal{A}$ cannot be successful at locating the "needles" or good assignments.

There is an emerging literature generalizing the *overlap gap property* with increasingly strong quantitative obstructions [44, 21, 43, 50]. We use the *branching OGP* of [50], which gives the strongest quantitative bounds of algorithms known on spin glasses and random CSPs [55].

---

[2] Following [26], we use $o_x(g(x))$ to mean the value is at most some function $\kappa(n,x)$ satisfying $\frac{\kappa(n,x)}{g(x)} \to 0$ as $n \to \infty$, then $x \to \infty$.

[3] Instances $\mathcal{I}_1, \mathcal{I}_2$ are t-correlated if a t fraction of constraints are identical, and the rest are independent; see also [55, Definition 6.1].

[4] An *overlap gap property* exists when an inner product, for example $\frac{1}{n}\langle \mathcal{A}(\mathcal{I}_1), \mathcal{A}(\mathcal{I}_2)\rangle$, is forbidden to take values within a certain range $[a,b]$.

Given a polynomial ξ with non-negative coefficients, let $\mathscr{L}_\xi$ denote the following set of functions:

$$\mathscr{L}_\xi \stackrel{\text{def}}{=} \left\{ \mu : [0,1) \to \mathbb{R}_{\geq 0} : \mu \text{ is right-continuous, } \|\xi''\mu\|_{\mathrm{TV}[0,t]} < \infty \ \forall t \in [0,1), \text{ and } \int_0^1 \xi''(t)\mu(t)\,dt < \infty \right\}.$$

Here, $\mathrm{TV}[a,b]$ over an interval $[a,b]$ is the total variance over partitions:

$$\|f\|_{\mathrm{TV}[a,b]} \stackrel{\text{def}}{=} \sup_n \sup_{a \leq t_0 < \cdots < t_n \leq b} \sum_{i \in [n]} |f(t_i) - f(t_{i-1})|\,.$$

Furthermore, the definition of P_ξ over $\mathscr{U}$ can be naturally extended to $\mathscr{L}_\xi$, with infimum

$$\mathsf{ALG}_\xi \stackrel{\text{def}}{=} \inf_{\mu \in \mathscr{L}_\xi} \mathsf{P}_\xi(\mu)\,. \tag{1.3}$$

Since $\mathscr{L}_\xi \supseteq \mathscr{U}$, we have $\mathsf{ALG}_\xi \leq \mathsf{OPT}_\xi$. It turns out that all *overlap-concentrated* algorithms over $\mathsf{CSP}_f(\alpha)$ are with high probability obstructed to some ALG_ξ:

▶ **Theorem 5** ([55, Corollary 6.10])**.** *Choose any* even *predicate f. For all $\epsilon > 0$ there is $\alpha_0 > 0$ such that the following is true: For all $\alpha \geq \alpha_0$, for every* overlap-concentrated *algorithm $\mathcal{A}$ over $\mathsf{CSP}_f(\alpha)$, and with high probability over instances $\mathcal{I} \sim \mathsf{CSP}_f(\alpha)$ as $n \to \infty$, $\mathcal{A}$ returns an assignment to $\mathcal{I}$ with satisfying fraction* at most

$$\mathbb{E}[f] + \frac{\mathsf{ALG}_\xi + \epsilon}{\sqrt{\alpha}}\,,$$

where $\xi(s) \stackrel{\text{def}}{=} \sum_{j=1}^r \|f^{=j}\|^2 s^j$, and $\|f^{=j}\|^2$ is the Fourier weight of f at degree j.

Note that if $\mathsf{ALG}_\xi \neq \mathsf{OPT}_\xi$, then these *overlap-concentrated* algorithms are obstructed from reaching the optimal solution for typical instances of $\mathsf{CSP}_f(\alpha)$.

1.2 Statement of results

The main result of our work is that there is an *optimal* algorithm among *overlap-concentrated* algorithms over $\mathsf{CSP}_f(\alpha)$, subject to a technical condition. This matches analogous results for Ising spin glasses [1], and for multi-species spherical spin glasses [84, 48].

▶ **Assumption 1** (Proposed in [79, Remark 1.7])**.** *For any polynomial ξ with non-negative coefficients and where $\xi''(0) \neq 0$,*[5] *ALG_ξ is achieved by some $\mu_* \in \mathscr{L}_\xi$. That is, the minimizer of Equation* (1.3) *exists.*

We include Assumption 1 because our algorithm uses the minimizer μ_*. As noted in [1, journal version, Remark 2.4], Assumption 1 should not have a major impact on the algorithm's performance, since by an informal continuity argument, using a near-minimizer $\widetilde{\mu}_* \in \mathscr{L}_\xi$ instead of μ_* incurs negligible error. We further discuss this in Section 5.

▶ **Theorem 6.** *Suppose Assumption 1 holds. Choose any predicate f without linear-weight terms; i.e. $\|f^{=1}\|^2 = 0$. For all $\epsilon > 0$ there is $\alpha_0 > 0$ such that the following is true: For all $\alpha \geq \alpha_0$, there exists a* randomized *algorithm $\mathcal{A}$ such that $\mathcal{A}$ is* overlap-concentrated *over $\mathsf{CSP}_f(\alpha)$, and with high probability over* both *instances $\mathcal{I} \sim \mathsf{CSP}_f(\alpha)$ and internal randomness of $\mathcal{A}$ as $n \to \infty$, $\mathcal{A}$ returns an assignment to $\mathcal{I}$ with satisfying fraction* at least

$$\mathbb{E}[f] + \frac{\mathsf{ALG}_\xi - \epsilon}{\sqrt{\alpha}}\,,$$

where $\xi(s) \stackrel{\text{def}}{=} \sum_{j=1}^r \|f^{=j}\|^2 s^j$, and $\|f^{=j}\|^2$ is the Fourier weight of f at degree j.

[5] It is shown in [79] that ALG_ξ is not achieved by any $\mu \in \mathscr{L}_\xi$ when $\xi''(0) = 0$. See also Remark 14.

This is analogous to work of [1], which proves a similar result for mean-field spin glasses. This also extends the algorithm of [3], which studies Maximum Cut on random regular graphs. In fact, all of these algorithms are related to the Approximate Message Passing (AMP) framework, popularized by the results of [13]; see [7] for a recent survey. It was suspected that such an algorithm could be written down for $\mathsf{CSP}_f(\alpha)$ [55, 41].

One immediate impact of this result is in quantum computing. In the gate-based circuit model, many quantum circuits are *overlap-concentrated* for instances of $\mathsf{CSP}_f(\alpha)$ of depth up to $c \cdot \log n$ for some c [24]. This work implies that such quantum circuits, even up to $c \cdot \log n$ depth, cannot outperform classical algorithms on typical instances of a random CSP.

▶ **Corollary 7** (Failure of quantum advantage). *Suppose Assumption 1 holds. Choose any* even *predicate* f. *For all* $\epsilon > 0$ *there exist* $\alpha_0 > 0$ *and* $c : \mathbb{R}^{>0} \to \mathbb{R}^{>0}$ *such that the following is true: For all* $\alpha \geq \alpha_0$, *and with high probability over instances* $\mathcal{I} \sim \mathsf{CSP}_f(\alpha)$ *as* $n \to \infty$, no quantum circuit *with depth up to* $c(\alpha) \cdot \log n$ *that applies 2-local gates on a fixed, bounded-degree architecture can produce an assignment with satisfying fraction that outperforms classical algorithms by more than* $\frac{\epsilon}{\sqrt{\alpha}}$.

Even until now, the promise of quantum advantage for optimization problems has been unclear. Multiple times, improved analyses of the Quantum Approximate Optimization Algorithm (QAOA) [34] suggesting a such an advantage [35, 88, 77, 11] led to better classical algorithms [9, 45, 63, 8, 3]. This work shows that for the QAOA and many other quantum circuits up to logarithmic depth, there can be *no quantum advantage* on random CSPs with even predicates of sufficiently large clause density. Note that unlike related work on the QAOA [32, 12, 4], this rules out quantum advantage *regardless* of the presence of overlap gap phenomena.

We remark on the technical condition of *even predicates*. This can likely be removed with two steps. First, the obstruction of *overlap-concentrated* algorithms only applies to *even* predicates. New techniques to prove OGPs for spherical spin glasses [47] can possibly be ported to all Ising spin glasses; these results would immediately transfer to random instances of $\mathsf{CSP}_f(\alpha)$ via [55]. Second, the algorithm we analyze applies when predicates have no linear-weight terms; adjustments taken in [79] will likely generalize this algorithm. These steps would extend Corollary 7 to all predicates f.

Previous works noting limitations of logarithmic-depth quantum algorithms on sparse random optimization problems [32, 33, 24] also rely on the concept of *overlap gap property*. As in [1, 48], this work shows the existence of a family of random optimization problems with an *optimal* algorithm up to the OGP barrier. All of the optimal algorithms are classical, which hints that quantum algorithms that cannot pass the OGP barrier offer no additional power for classical optimization in general.

1.3 Proof overview

Our proof builds on a number of other works, especially [68, 1, 3]. We outline the proof here. Due to the page limit, some proofs in this paper are deferred to the full version.

In Section 2, we define our model of CSP and describe the algorithm. Instead of $\mathsf{CSP}_f(\alpha)$, we run the algorithm on a CSP whose associated constraint hypergraph is locally a regular *hypertree*.[6] This algorithm runs in *iterations*, where on each iteration, every vertex recomputes its value using information from adjacent vertices; we parameterize this algorithm by the functions (two per *iteration number*) that recompute each vertex's value.

[6] In the hypergraph, the variables are vertices and the constraints are hyperedges.

Table 1 List of important terms used in the analysis of message-passing algorithms for mean-field spin glasses and random CSPs. Terms in the first row asymptotically approach independent Gaussians after applying state evolution, and so constitute the *Brownian motion* (up to normalization) in the stochastic picture.

Term	**[68]**	**[1]**	**[3]**	**(our work)**
Brownian motion (asymptotically independent)	u	Δ	u	u
sum of Brownian motions	–	z	–	w
nonlinearity controller	x	x	x	x
value of spin	z	$m = f(\dots)$	z	z

We can already study the performance of the parameterized algorithm, analyzing how information propagates locally through the hypertree. In Section 3, we generalize a set of σ-algebras from [3] to describe this behavior. This allows us to analyze the correlation of a variable from the algorithm with any other variable from a particular σ-algebra. With these tools in hand, we simplify the expected performance of the parameterized algorithm, deferring some detailed calculations to the full version of this paper.

We then show how the algorithm, with the right choice of parameters, solves a stochastic optimal control problem that is dual to the minimization problem in Equation (1.3).

1. First, we show that with high probability as $\alpha \to \infty$, the intermediate values in the algorithm are asymptotically *Gaussian*. We prove this in Section 4 with an argument known as *state evolution*. Notably, this argument works for sparse models (as in [3]), and for message-passing on hypergraphs, where some values are products of Gaussians. This allows us to study quantities in the algorithm output as *discrete* stochastic processes.
2. Then, we take the number of iterations to a very large constant while proportionally decreasing the step size. Under this process, each discrete stochastic process asymptotically approximates a *continuous* stochastic process. In Section 5, we show how to choose the optimal parameters. We also show that rounding the algorithm's output to a value in $\{\pm 1\}^n$ does not meaningfully affect its performance.

This work extends [3] in a spiritually similar way that [1] extends [68]; roughly, we extend from graphs to *hypergraphs*. This requires more involved calculations and in some cases new arguments. For example, we require a more complicated proof of state evolution, since the messages in our algorithm are not Gaussian at fixed degree. Furthermore, our proof of rounding is new, since we cannot rely on spectral statistics of random *graphs*. Finally, our analysis uses an additional symmetry we call *index-regularity* (defined in Section 2) which, to the best of our knowledge, does not come up in prior work. We show that enforcing this symmetry has negligible effect (proof deferred to the full version of the paper) and use this symmetry to simplify some of our analysis.

There are a number of quantities to keep track of. We annotate the important symbols in Table 1. Following notation in [3], we typically use lowercase for a quantity in the algorithm (e.g. $u^{\ell}_{i\to a}$), uppercase with step size δ for a *discrete* stochastic process (e.g. U^{δ}_{ℓ}), and uppercase with time t for a *continuous* stochastic process (e.g. U_t).

1.4 Related work

Approximate Message Passing. Approximate Message Passing (AMP) is an algorithmic framework used to design locally computable, iterative methods for a variety of high-dimensional statistical inference and recovery tasks [61, 76, 67, 64, 10]. Many of the ideas

motivating AMP originate from the study of *belief propagation*, an iterative method rediscovered multiple times in different scientific contexts; see [65, 58] for a broader history. AMP was first introduced in the context of compressed sensing and signal recovery [29, 57], partly inspired by the use of structured priors in decoding low-density parity check codes [75, 69]. In spin glass theory, AMP is synonymous with computing the fixed points of a certain Thouless–Anderson–Palmer (TAP) free energy functional [87, 59, 90, 22, 23]. In fact, our algorithm is inspired by similar connections between the TAP equations and belief propagation [66, 65].

A beneficial feature of AMP is that one can rigorously analyze the asymptotic behavior of statistics in the framework. This is typically done through the use of *state evolution* statements [14, 13]: laws of large numbers which dictate the behavior of such statistics when the dimension in the problem approaches infinity. This has been subsequently used to compute approximate ground states of both Ising and spherical mean-field spin glasses [68, 1, 47], as well as Max-Cut on random sparse graphs [3, 19]. See the surveys of [38, 7] for further discussion on AMP and its variants. See also [78, 52, 54] for connections to the sum-of-squares hierarchy.

Ising spin glasses and CSPs. Although mean-field spin glasses with Ising spins were proposed as mathematical models of materials [82], their connection to combinatorial optimization has been observed for some time [39, 66]. More recently, a correspondence has emerged between Ising spin glasses and sparse random CSPs at large enough clause density; this started with relating the supremum of the two models [26, 80, 72]. This supremum was calculated using a novel heuristic argument known as "replica symmetry breaking" [73], and proven to be correct in [85, 70] (see also [71]). Moreover, the duality preserves a phenomenon known as *overlap gap* [44], a generalization of solution clustering that obstructs certain algorithms; see [21, 43, 42, 24, 50, 55, 49], and the expositions of [40, 46]. Notably, a strong version of overlap gap phenomena [50, 47, 48] is connected to the message-passing algorithms of [68, 1, 3], and this work. Other topics at the intersection of spin glasses and CSPs include approximating partition functions (e.g. [83]), sampling from Gibbs measures (e.g. [2, 51]), and computing thresholds for random CSPs (e.g. [27, 28]).

Quantum algorithms for classical optimization. Recent demonstrations of quantum advantage (e.g. [5, 91]) have led to increased excitement around using contemporary (i.e. NISQ [74]) quantum computers for a practically relevant problem. Although classical optimization is such a problem, quantum algorithms here are not proven to have any benefit over classical algorithms. The most popular proposal is the Quantum Approximate Optimization Algorithm (QAOA) [34]. Earlier suggestions of quantum advantage with the QAOA [35, 88, 77, 89] led to better classical algorithms [9, 45, 63, 8]; note that all of these comparisons were for very small circuit depth. Recently, [36, 11, 12] find an expression for the performance of the QAOA on spin glasses and sparse CSPs at any constant depth. For some problems, this is competitive with known classical algorithms; see [37, 16].

The QAOA is a *local* algorithm. It was noted that this feature can limit its performance [32, 33]. The authors in [24] unify definitions of quantum and classical local algorithms to show that the QAOA is obstructed by overlap gap phenomena on sparse problems. Because of this, [3] rules out quantum advantage through $\epsilon \cdot \log n$-depth QAOA on random regular Max-Cut instances of large degree (up to Assumption 1); our work generalizes this fact to CSPs with even predicates. Regardless of theory, there are many recent demonstrations of the QAOA on state-of-the-art quantum computing systems; see for example [31, 81, 62]. It is also known that the QAOA can solve satisfiability faster than brute force [17]; see also [25] for a related

protocol with a more-than-quadratic improvement. Some quantum algorithms for machine learning are known to provide no quantum advantage [86]. Very recently, [56] proposed a quantum algorithm for classical optimization, but we do not yet know if it gives a quantum advantage.

2 Local algorithms for sparse CSPs

A CSP instance has an implicit *directed* hypergraph, with variables as vertices and constraints as hyperedges.

▶ **Definition 8** (Directed hypergraph of a CSP)**.** *Consider a r-uniform CSP instance $\mathcal{I}$ with n variables and constraints $E(\mathcal{I})$, where each constraint $e \in E(\mathcal{I})$ involves r (ordered) variables in $[n]$. Then $\mathcal{I}$ is associated with the directed hypergraph with vertices $[n]$ and hyperedges $E(\mathcal{I})$; i.e. $G([n], E(\mathcal{I}))$.*

We consider a hyperedge *directed* when the involved vertices are ordered. For example, if constraint e involves vertices $\{v_1, v_2, v_3\}$ in order $[2, 3, 1]$, then the associated directed hyperedge is $[v_2, v_3, v_1]$. We denote this ordered list as ∂e. We often refer to the CSP instance by its hypergraph. The instance is fully specified by its hypergraph, the CSP function f, and the (ordered) randomness $[\epsilon_{e_1}, \ldots, \epsilon_{e_r}]$ on each constraint $e \in E(\mathcal{I})$. For each variable i, we use ∂i to denote the set of constraints i participates in.

The *factor graph* of a hypergraph $G(V, E)$ is a bipartite graph of order $|V| + |E|$, with edges between each hyperedge $e \in E$ and each vertex in ∂e. For more information on the factor graph formalism, see [60]. Let $\mathsf{B}_i(\ell)$ be the induced subgraph of G given by the ℓ-local neighborhood around vertex $i \in V$.[7]

▶ **Definition 9.** *A finite hypergraph $G(V, E)$ is (ℓ, ϵ)-*locally treelike *if the factor graph of $\mathsf{B}_i(\ell + 1)$ is isomorphic to a tree for at least $(1 - \epsilon) \cdot n$ vertices $i \in V$.*

This definition is a generalization of locally treelike graphs used in [3]; see Figure 1 for an illustration.

Every *message-passing* algorithm is contained in the model of local algorithms presented in [24]. In this model, the value associated with each variable is uniquely determined by its local neighborhood.

▶ **Definition 10** (Generic p-local algorithms, [24, Definition 3.1])**.** *Consider a randomized algorithm $\mathcal{A}$ that inputs a r-uniform hypergraph $G(V, E)$ and outputs a label in an alphabet Σ to each vertex in the graph, i.e. $\mathcal{A}(G) \in \Sigma^V$. Then $\mathcal{A}$ is* generic p-local *if for any set $S \subseteq V$, the following conditions hold:*

1. (Local distribution determination) *The joint marginal distribution of $(\mathcal{A}(G)_v)_{v \in S}$ is identical to that of $(\mathcal{A}(G')_v)_{v \in S}$, where $G' = \bigcup_{v \in S} \mathsf{B}_v(p)$.*
2. (Local independence) *$\mathcal{A}(G)_v$ is statistically independent of $\mathcal{A}(G)_{v'}$ over all vertices $v' \notin \mathsf{B}_v(2p)$.*

This model also includes the Quantum Approximate Optimization Algorithm [34] on typical instances of $\mathsf{CSP}_f(\alpha)$. In fact, any depth-p quantum circuit with a fixed architecture of bounded degree can be seen as generic p-local by including the architecture in the input hypergraph. [24] proves that for all $p \leq c(\alpha) \cdot \log n$, generic p-local algorithms are overlap-concentrated over instances of $\mathsf{CSP}_f(\alpha)$ with high probability as $n \to \infty$.

[7] Note that we define distance as the number of *hypergraph vertices* in the shortest path; vertices that share a hyperedge are distance 1 apart.

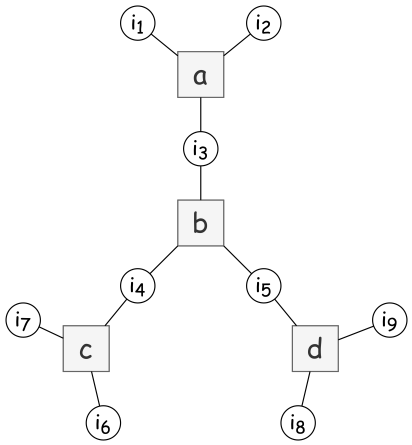

Figure 1 Factor graph of the hypergraph with vertices $\{i_1, \ldots, i_9\}$ and hyperedges $\{a, b, c, d\}$. Note that this hypergraph is locally treelike.

In our analysis, we make some convenient modifications to the input. First, since our algorithm is local and the instances are locally treelike on all but a negligible fraction of vertices[8], we analyze the algorithm on an instance which is locally a *hypertree*, i.e. a hypergraph whose factor graph is a tree. We also make the hypergraph *regular*, by embedding our instance of $\mathsf{CSP}_f(\alpha)$ in a locally treelike d-regular instance with $d = r \cdot \alpha + o_\alpha(1)$.

On hypergraphs, we need one additional property, which we call *index-regular*. Here, we mean that for each $\iota \in [r]$, each variable is used in *index* ι of a constraint the same number of times.

▶ **Definition 11** (d-index-regular CSP). *Consider a (r-uniform) CSP instance $\mathcal{I}$ with n variables and constraints $E(\mathcal{I})$, such that each constraint $e \in E(\mathcal{I})$ has associated variables $\partial e = [v_{e,1}, \ldots, v_{e,r}]$. Then $\mathcal{I}$ is* d-index-regular *if every variable $v \in [n]$ is involved in d constraints, and moreover, for each $\iota \in [r]$, v is the ι^{th}-indexed variable in exactly $\frac{d}{r}$ constraints.*

We describe our choice of algorithm in Figure 2. This involves some notation we describe here. A function that depends on zero inputs is a constant; i.e. $A_i^0 = F_0$ and $A_{i\to a}^0 = F_{\to,0}$. For any scalar quantity that depends on vertex, a bolded version represents a vector quantity. For example, given $\partial a = [i_1, \ldots, i_r]$, $\boldsymbol{z}_{\partial a \to a}^\ell$ is the *ordered* list $[z_{i_1\to a}^\ell, \ldots, z_{i_r\to a}^\ell]$, and $\boldsymbol{z}_{\partial a}^\ell$ is the *ordered* list $[z_{i_1}^\ell, \ldots, z_{i_r}^\ell]$. Similarly, $\boldsymbol{z}^L$ is the vector in $\mathbb{R}^V$ such that $\left(\boldsymbol{z}^L\right)_i = z_i^L$.

[8] In particular, for any finite ℓ, with high probability our instances are (ℓ, ϵ)-locally treelike for some ϵ, where $\epsilon \to 0$ as $n \to \infty$.

Input: The timestep $\delta > 0$, and a d-index-regular hypergraph with girth at least $2L+2$, where $L = \lfloor \frac{1}{\delta} \rfloor$.
Parameters: The CSP function $f : \{\pm 1\}^r \to \{0,1\}$, constant $K > 0$, and nonlinearities

$$\big\{F_{\to,\ell} : \mathbb{R}^\ell \to [-K,K]\big\}_{\ell=0,\dots,L-1} \qquad \big\{F_\ell : \mathbb{R}^\ell \to [-K,K]\big\}_{\ell=0,\dots,L-1}$$

Algorithm:

1. Initialize the following:
   a. $z_i^0 \sim \mathcal{N}(0,\delta)$ for all $i \in V$ and $z_{i\to a}^0 = z_i^0$ for all $i \in \partial a$ and $a \in E$.
   b. $w_i^0 = 0$ for all $i \in V$ and $w_{i\to a}^0 = 0$ for all $i \in \partial a$ and $a \in E$.
2. For each iteration $\ell = 0, \dots, L-1$:
   a. Let $A_{i\to a}^\ell = F_{\to,\ell}\big(u_{i\to a}^1, \dots, u_{i\to a}^\ell\big)$, and let $A_i^\ell = F_\ell\big(u_i^1, \dots, u_i^\ell\big)$.
   b. For each factor $a \in E$ and involved variable $i \in \partial a$, compute:

$$w_{i\to a}^{\ell+1} = \frac{1}{\sqrt{d-1}} \cdot \sum_{b \in \partial i \setminus a} \mathsf{D}_{i;b} f(\boldsymbol{z}_{\partial b \to b}^\ell)$$
$$u_{i\to a}^{\ell+1} = w_{i\to a}^{\ell+1} - w_{i\to a}^\ell$$
$$z_{i\to a}^{\ell+1} = z_{i\to a}^0 + \sum_{s=1}^{\ell+1} A_{i\to a}^{s-1} \cdot u_{i\to a}^s$$

   c. For each variable $i \in V$, compute:

$$w_i^{\ell+1} = \frac{1}{\sqrt{d}} \cdot \sum_{b \in \partial i} \mathsf{D}_{i;b} f(\boldsymbol{z}_{\partial b \to b}^\ell)$$
$$u_i^{\ell+1} = w_i^{\ell+1} - w_i^\ell$$
$$z_i^{\ell+1} = z_i^0 + \sum_{s=1}^{\ell+1} A_i^{s-1} \cdot u_i^s$$

Output: For every $i \in V$, output 1 with probability p_i and -1 with probability $1-p_i$, where $p_i = \begin{cases} 1 & \text{if } z_i^L \geq 1, \\ \frac{1+z_i^L}{2} & \text{if } |z_i^L| < 1, \\ 0 & \text{if } z_i^L \leq -1. \end{cases}$

Figure 2 Description of the message-passing algorithm.

Our analysis makes extensive use of the partial derivative operator. For any factor a and variable $j \in \partial a$, let $\mathsf{D}_{j;a} f$ be the partial derivative of f at the *coordinate* ι such that j is the ιth coordinate of ∂a. For example, if $\partial a = [k,i,j]$ and $f(x_1,x_2,x_3) = x_1x_2 + x_2x_3$, then $\mathsf{D}_{k;a} f(x_1,x_2,x_3) = x_2$ and $\mathsf{D}_{i;a} f(x_1,x_2,x_3) = x_1 + x_3$. Note the following:

▶ Remark 12. Since the domain of f is $\{\pm 1\}^r$, f has a representation as a multilinear polynomial. As a result, for all factors b and variables $i \in \partial b$, $\mathsf{D}_{i;b} f$ is independent of the entry with the same coordinate as i in ∂b.

We additionally require the nonlinearities to satisfy the following two properties:

▶ **Property 1** (Similar node and node-to-factor evolution). *For every $\ell \geq 0$, variable i, factor $a \in \partial i$, and $m \in \mathbb{N}$, we have $\mathbb{E}[|A_i^\ell - A_{i\to a}^\ell|^m] = O_d(\frac{1}{d^{m/2}})$.*

▶ **Property 2** (Normalization of second moment). *For every $\ell \geq 0$, variable i, and factor $a \in \partial i$, we have $\mathbb{E}[(A_i^\ell)^2] = \mathbb{E}[(A_{i\to a}^\ell)^2] = \frac{\delta}{\nu_{\ell+1}}$, where $\nu_\ell \stackrel{\text{def}}{=} \frac{\xi'(\ell\delta) - \xi'((\ell-1)\delta)}{r}$, and ξ is defined as in Theorem 6.*

In Section 5, we show how to choose the nonlinearities based on Equation (1.1), and show that our choice indeed satisfies Property 1 and Property 2.

▶ Remark 13. Our analysis of the algorithm above holds even if the CSP predicate depends on factor a, so long as the absolute value of each Fourier component is fixed (this implies that all f_a are associated with the same ξ). For example, this implies that the algorithm achieves the same value for *every* choice of signs $\{\epsilon_{i,\iota}\}_{i\in[n],\iota\in[r]}$ in Definition 1.

▶ Remark 14. If $\|f^{=2}\| = 0$, the associated ξ does not satisfy $\xi''(0) \neq 0$ as in Assumption 1. We handle this case via the following argument. First, it can be shown that ALG_ξ is continuous in each coefficient of ξ using an explicit form of the Parisi functional and the fact that $\mathsf{ALG}_\xi \geq 0$ [1, Proposition 4.1]. Then, we can add a small quadratic term to f, and run the algorithm on the perturbed instance. The error produced by this perturbation may be absorbed in the ϵ of Theorem 6.

3 Independence claims

Given random variables $\{X_1, \ldots, X_k\}$, let $\sigma(X_1, \ldots, X_k) = \sigma(\{X_1, \ldots, X_k\})$ be the *σ-algebra* generated by $\{X_1, \ldots, X_k\}$. For an introduction to σ-algebra and related concepts, see for example [30].

Consider the algorithm in Figure 2. For every $0 \leq \ell \leq L$,[9] variable v, and factor $a \in \partial v$, we'll reuse the notation $\mathsf{B}_v(\ell)$ as the set of *nodes* of distance[10] at most ℓ from v in the corresponding hypergraph. We define the set $\mathsf{B}_{v\to a}(\ell)$ as the set of nodes of distance at most ℓ from v, *excluding* nodes that reach v through the factor a. We define

$$\mathcal{G}_v^\ell \stackrel{\text{def}}{=} \sigma\left(\{z_i^0 \mid i \in \mathsf{B}_v(\ell)\}\right), \qquad \mathcal{G}_{v\to a}^\ell \stackrel{\text{def}}{=} \sigma\left(\{z_i^0 \mid i \in \mathsf{B}_{v\to a}(\ell)\}\right).$$

Intuitively, these $\mathcal{G}_v^\ell$ algebras are the random variables determined by the initial Gaussian variables in the radius-ℓ neighborhood around v; here, ℓ is a *distance*. We formalize a series of information-theoretic claims using these σ-algebras, similar to [3, Lemma 2.2]. See also an illustration in Figure 3.

The first fact about independence informally says that node-to-factor messages never use information on "the other side" of the factor; once information goes across a, it never comes back.

▶ **Lemma 15.** *For every $0 \leq \ell_1, \ell_2 \leq L$, distinct variables i, j, and factors $a \in \partial i, b \in \partial j$, if every path of length at most $2L$ from i to j goes through a and b, then $\mathcal{G}_{i\to a}^{\ell_1}$ and $\mathcal{G}_{j\to b}^{\ell_2}$ are independent.*

[9] Recall that because the instance has girth at least $2L+2$, all local neighborhoods of radius up to L are treelike.

[10] Once again, we consider *distance* of i and j as the number of *nodes* required to pass from i to j; the factors are not counted.

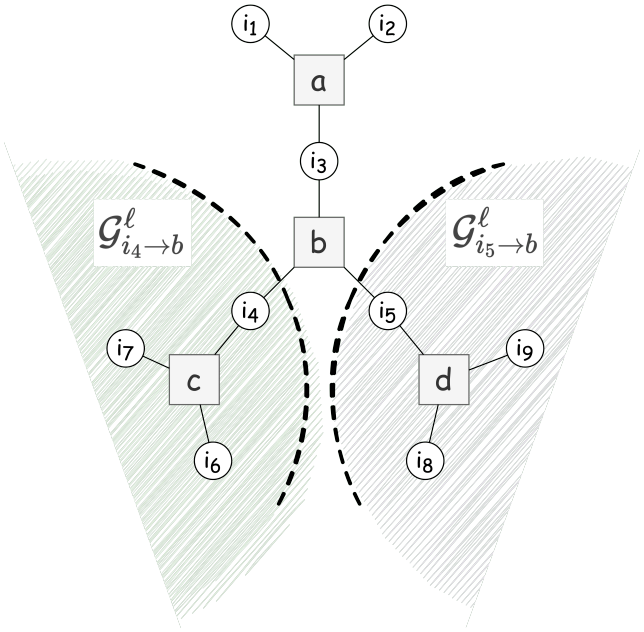

Figure 3 An illustration of σ-algebras on the hypertree. Since the path from i_4 to i_5 goes through factor b, $\mathcal{G}^{\ell}_{i_4\to b}$ and $\mathcal{G}^{\ell}_{i_5\to b}$ are independent by Lemma 15. In addition, $\mathcal{G}^{\ell+1}_{i_3\to a} = \sigma\left(\mathcal{G}^{0}_{i_3} \bigcup \mathcal{G}^{\ell}_{i_4\to b} \bigcup \mathcal{G}^{\ell}_{i_5\to b}\right)$ by Lemma 16.

Lemma 15 holds even if the factors are the same, i.e. $a = b$. We refer to this case as Lemma 15 *around* factor a.

The next fact concerns the geometry of local neighborhoods. It allows us to decompose each σ-algebra into σ-algebras with distance reduced by one, which is handy for induction.

▶ **Lemma 16.** *For every $1 \le \ell \le L$, variable i, and factor $a \in \partial i$, we have*

$$\mathcal{G}^{\ell}_{i} = \sigma\left(\mathcal{G}^{\ell}_{i\to a} \cup \left(\bigcup_{j\in\partial a\setminus i} \mathcal{G}^{\ell-1}_{j\to a}\right)\right), \qquad \mathcal{G}^{\ell}_{i\to a} = \sigma\left(\mathcal{G}^{0}_{i} \cup \left(\bigcup_{b\in\partial i\setminus a}\ \bigcup_{v\in\partial b\setminus i} \mathcal{G}^{\ell-1}_{v\to b}\right)\right).$$

The next statement sorts quantities appearing in the algorithm (Figure 2) into appropriate σ-algebras. It proves the intuition that messages at iteration number ℓ are determined by the local neighborhood of distance at most ℓ:

▶ **Lemma 17.** *For every $1 \le \ell \le L$, variable i, and factor a, we have $u^{\ell}_{i\to a}, w^{\ell}_{i\to a}, z^{\ell}_{i\to a} \in \mathcal{G}^{\ell}_{i\to a}$ and $u^{\ell}_{i}, w^{\ell}_{i}, z^{\ell}_{i} \in \mathcal{G}^{\ell}_{i}$.*

Lemma 17 implies that the ℓ^{th} node message at variable i is determined by the initial Gaussian variables in the radius-ℓ neighborhood around i. Note that the node-to-factor message $u^{\ell}_{i\to a}$ is *not* influenced by Gaussians at nodes that pass through a to reach i.

Next, we prove that messages at node i and iteration number $\ell+1$ are uncorrelated with the initial Gaussians in the radius-ℓ neighborhood around i. Put another way, the "extra randomness" at distance $\ell+1$ removes any correlation with $\mathcal{G}^{\ell}_{i}$. To do this, we show that the messages $\{u^{\ell}_{j\to b}\}_{j\in\partial b\setminus i}$ are uncorrelated with each other, conditioned on $\mathcal{G}^{\ell}_{i}$.[11]

[11] This is subtle. Although $\{u^{\ell}_{j\to b}\}_{j\in\partial b\setminus i}$ are independent by Lemma 15 around factor b, independence does not imply *conditional* independence.

▶ **Lemma 18.** *Choose any variable i, factor $b \in \partial i$, and $1 \le \ell \le L$. Choose also $s_j \ge 0$ for every $j \in \partial b \backslash i$. Then*

$$\mathbb{E}\left[\prod_{j\in\partial b\backslash i}(u^{\ell}_{j\to b})^{s_j}\;\middle|\;\mathcal{G}^{\ell}_i\right] = \prod_{j\in\partial b\backslash i}\mathbb{E}\left[(u^{\ell}_{j\to b})^{s_j}\;\middle|\;\mathcal{G}^{\ell}_i\right].$$

▶ **Lemma 19.** *Choose any variable i, factor $a \in \partial i$, and $0 \le \ell \le L-1$. Then $\mathbb{E}\left[u^{\ell+1}_{i\to a}\mid \mathcal{G}^{\ell}_i\right] = 0$ and $\mathbb{E}\left[u^{\ell+1}_{i}\mid \mathcal{G}^{\ell}_i\right] = 0$.*

▶ **Corollary 20.** *Fix any variable i, factor $b \in \partial i$, and $0 \le \ell \le L$. Then $\mathbb{E}\left[\mathsf{D}_{i;b}f(\boldsymbol{z}^{\ell}_{\partial b\to b}) \mid \mathcal{G}^{\ell}_i\right] = \mathsf{D}_{i;b}f(\boldsymbol{z}^{\ell-1}_{\partial b\to b})$ if $\ell \ge 1$ and 0 otherwise.*

3.1 Simplifying the main sum

What is the average value achieved by the algorithm? By linearity, we can look at the value of any factor a; i.e. $\mathbb{E}[f(\boldsymbol{z}^{L}_{\partial a})]$. Note that $\mathbb{E}[f(\boldsymbol{z}^{0}_{\partial a})] = f(0,\dots,0) = \mathbb{E}[f]$, and

$$\begin{aligned}\mathbb{E}[f(\boldsymbol{z}^{L}_{\partial a})] &= \mathbb{E}[f(\boldsymbol{z}^{0}_{\partial a})] + \sum_{\ell=0}^{L-1}\left(\mathbb{E}[f(\boldsymbol{z}^{\ell+1}_{\partial a})] - \mathbb{E}[f(\boldsymbol{z}^{\ell}_{\partial a})]\right)\\ &= \mathbb{E}[f] + \sum_{\ell=0}^{L-1}\left(\mathbb{E}[f(\boldsymbol{z}^{\ell+1}_{\partial a})] - \mathbb{E}[f(\boldsymbol{z}^{\ell}_{\partial a})]\right).\end{aligned}$$

It suffices to compute the difference $\mathbb{E}[f(\boldsymbol{z}^{\ell+1}_{\partial a})] - \mathbb{E}[f(\boldsymbol{z}^{\ell}_{\partial a})]$ for every $0 \le \ell \le L-1$. We use the Fourier expansion of f to express this as a sum of differences of monomials.

First, we simplify terms of the form $u_i^{s_i}Y_i^{s_i-1}\prod_{v\in\partial a\backslash i}Y_v^{s_v}$, which we encounter along the way. This relies on the following consequence of the monotone class theorem:

▶ **Lemma 21** (e.g. [30, Theorem 5.2.2], [3, Lemma 2.3])**.** *Let $\{\mathcal{F}_1,\dots,\mathcal{F}_m,\mathcal{A}\}$ be a set of σ-algebras so that $\mathcal{F} \stackrel{\text{def}}{=} \sigma(\mathcal{F}_1,\dots,\mathcal{F}_m) \subseteq \mathcal{A}$. Choose a random variable $X \in \mathcal{A}$,[12] and suppose for any $Y_1 \in \mathcal{F}_1,\dots,Y_m \in \mathcal{F}_m$, $\mathbb{E}[XY_1\dots Y_m] = 0$ when the expectation exists. Then $\mathbb{E}[XY] = 0$ (when the expectation exists) for every $Y \in \mathcal{F}$.*

As a result, to show an expectation is zero, we may replace a variable in σ-algebra $\mathcal{F}$ with a product of variables from σ-algebras that cover $\mathcal{F}$.

▶ **Lemma 22** (Generalization of [3, Lemma 2.5])**.** *Choose factor a, integers $s_v \ge 0$ for $v \in \partial a$, and let $Y_v^{s_v} \in \mathcal{G}_v^{s_v}$. Let $i \in \partial a$ be such that $s_i + 1 \ge s_v$ for all $v \in \partial a$. Then*

$$\mathbb{E}\left[u_i^{s_i+1}\prod_{v\in\partial a}Y_v^{s_v}\right] = \frac{1}{\sqrt{d}}\,\mathbb{E}\left[\left(\mathsf{D}_{i;a}f(\boldsymbol{z}^{s_i}_{\partial a\to a}) - \mathbb{1}_{[s_i\ge 1]}\mathsf{D}_{i;a}f(\boldsymbol{z}^{s_i-1}_{\partial a\to a})\right)\prod_{v\in\partial a}Y_v^{s_v}\right],$$

where $\mathbb{1}_{[x\ge y]} = 1$ if $x \ge y$ and 0 otherwise.

▶ **Proposition 23.** *Fix $0 \le \ell \le L-1$, factor a, and $S \subseteq \partial a$. Then*

$$\mathbb{E}\left[\prod_{j\in S}z_j^{\ell+1}\right] - \mathbb{E}\left[\prod_{j\in S}z_j^{\ell}\right] = \sum_{i\in S}\mathbb{E}\left[A_i^{\ell}u_i^{\ell+1}\cdot\prod_{j\in S\backslash i}z_j^{\ell}\right] + O_d\left(\frac{1}{d}\right). \tag{3.1}$$

[12] $X \in \mathcal{A}$ means that $\sigma(X) \subseteq \mathcal{A}$.

▶ **Corollary 24.** *Choose $0 \le \ell \le L-1$ and factor a. Then*

$$\mathbb{E}\left[f(\boldsymbol{z}_{\partial a}^{\ell+1})\right] - \mathbb{E}\left[f(\boldsymbol{z}_{\partial a}^{\ell})\right] = \sum_{i \in \partial a} \mathbb{E}\left[A_i^\ell u_i^{\ell+1} \cdot \mathsf{D}_{i;a} f(\boldsymbol{z}_{\partial a}^{\ell})\right] + O_d\left(\frac{1}{d}\right).$$

We simplify this expression further, assuming the algorithm's node and node-to-factor nonlinearities are similar (Property 1). We choose nonlinearities exhibiting this property in Section 5.

▶ **Proposition 25.** *Assume Property 1. Let $0 \le \ell \le L-1$, a be a factor, and $i \in \partial a$. Then*

$$\frac{1}{\alpha n}\sum_{a \in E}\left(\mathbb{E}\left[f(\boldsymbol{z}_{\partial a}^{\ell+1})\right] - \mathbb{E}\left[f(\boldsymbol{z}_{\partial a}^{\ell})\right]\right) = \frac{r}{\sqrt{d}} \cdot \frac{1}{n}\sum_{i \in V} \mathbb{E}\left[A_i^\ell (u_i^{\ell+1})^2\right] + O_d\left(\frac{1}{d}\right).$$

By these propositions, we get the following:

▶ **Theorem 26.** *Assume Property 1. Fix a constant $L \ge 1$. Then*

$$\frac{1}{\alpha n}\sum_{a \in E} \mathbb{E}\left[f(\boldsymbol{z}_{\partial a}^{L})\right] = \mathbb{E}[f] + \frac{r}{\sqrt{d}} \cdot \sum_{\ell=0}^{L-1}\left(\frac{1}{n}\sum_{i \in V} \mathbb{E}\left[A_i^\ell (u_i^{\ell+1})^2\right]\right) + O_d\left(\frac{1}{d}\right).$$

4 State evolution for message-passing on hypergraphs

So far, we've shown that the value of the algorithm takes a simple form (i.e. Theorem 26). From here, we prove a *state evolution* statement: with some weak conditions on the nonlinearities (Property 2), we analyze the asymptotic behavior of statistics of the algorithm. State evolution as applied to AMP first appeared in [13]; this was inspired by a proof in a paper computing the fixed points of the TAP equations for the Sherrington–Kirkpatrick model [14]. Since then, state evolution has been used in AMP algorithms applied to a variety of tasks; see for example [38].

Here, we consider functions of the messages $\{u_{i\to a}^\ell\}_{\ell \in [L]}$ (or $\{u_i^\ell\}_{\ell \in [L]}$) that are *pseudo-Lipschitz*:

▶ **Definition 27.** *Fix $s > 0$. We say that a given function $\psi : \mathbb{R}^s \to \mathbb{R}$ is* pseudo-Lipschitz *if there exists a constant $C > 0$ such that*

$$|\psi(\boldsymbol{x}) - \psi(\boldsymbol{y})| \le C \cdot \left(1 + \|\boldsymbol{x}\| + \|\boldsymbol{y}\|\right) \cdot \|\boldsymbol{x} - \boldsymbol{y}\|,$$

for every $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^s$.

As $d \to \infty$, the expected value of a pseudo-Lipschitz function ψ with messages $\{u_{i\to a}^\ell\}_{\ell \in [L]}$ (or $\{u_i^\ell\}_{\ell \in [L]}$) as input approaches the expected value of ψ with independent Gaussians as input.

▶ **Definition 28.** *For any predicate $f : \{\pm 1\}^r \to \{0,1\}$, let ξ be defined as in Theorem 6. For any $\ell \ge 1$, define the quantity*[13]

$$\nu_\ell \stackrel{\text{def}}{=} \frac{\xi'(\ell\delta) - \xi'((\ell-1)\delta)}{r}.$$

[13] Note this is the same parameter ν_ℓ in Property 2.

▶ **Theorem 29.** *Assume Property 2. Then for any pseudo-Lipschitz function $\psi : \mathbb{R}^\ell \to \mathbb{R}$, variable i, factor $a \in \partial i$, $1 \le \ell \le L$, we have*

$$\mathbb{E}\left[\psi(u^1_{i\to a}, \dots, u^\ell_{i\to a})\right] = \mathbb{E}\left[\psi(U_1, \dots, U_\ell)\right] + o_d(1)\,, \tag{4.1}$$

$$\mathbb{E}\left[\psi(u^1_{i}, \dots, u^\ell_{i})\right] = \mathbb{E}\left[\psi(U_1, \dots, U_\ell)\right] + o_d(1)\,, \tag{4.2}$$

where $U_s \sim \mathcal{N}(0, \nu_s)$ independently.

Theorem 29 is inspired by the state evolution statement in [3]. These statements are for *sparse* problems, and so do not require a correction term to ensure variables remain Gaussian (often named the *Onsager correction*). In [3], the messages $\{u^\ell_{i\to a}\}_{\ell\in[L]}$ are Gaussian even at fixed degree! In our algorithm, this is not the case for arbitrary predicates f; we must take additional care showing that these quantities are close to Gaussian.

We defer the proof of Theorem 29 to the full version of the paper. As a quick overview of the proof, we first establish that given Property 2, $u^\ell_{i\to a}$ is *close* to the Gaussian $\mathcal{N}(0, \nu_\ell)$ when $d \to \infty$, independently of $\mathcal{G}_i^{\ell-1}$. Then, we show that pseudo-Lipschitz functions are smooth enough so that an input $u^\ell_{i\to a}$ may be exchanged with an independent Gaussian input, and use this argument inductively to finish the proof.

Given Theorem 29, we may convert each spin $\{z_i^\ell\}_{0\le\ell\le L}$ to a certain *Gaussian process*. In Section 5, we do exactly this, and choose the nonlinearities $\{A_i^\ell\}_{0\le\ell\le L}$ to maximize the expression in Theorem 26 (subject to $|z_i^L| \le 1$).

5 Choosing the nonlinearities

We now design the nonlinearities to maximize Theorem 26. In light of Section 4, we consider the stochastic optimal control problem of [1]. Define $\xi(s) \stackrel{\text{def}}{=} \sum_{j=1}^r \|f^{=j}\|^2 s^j$. Consider any $\mu \in \mathscr{L}_\xi$, and Φ^μ chosen as in Equation (1.1). Define the stochastic differential equation[14]

$$\mathrm{d}X_t = \xi''(t)\mu(t)\Phi^\mu_x(t, X_t)\mathrm{d}t + \sqrt{\xi''(t)}\mathrm{d}B_t\,, \quad X_0 = 0\,, \tag{5.1}$$

along with associated martingale

$$Z_t = \int_0^t \sqrt{\xi''(s)}\Phi^\mu_{xx}(s, X_s)\mathrm{d}B_s\,. \tag{5.2}$$

We now crucially use Assumption 1. Let $\mu_* \in \mathscr{L}$ be the minimizer of Equation (1.3).

▶ **Lemma 30** ([1, Theorem 3])**.** *Suppose Assumption 1 holds. Then for every $\epsilon > 0$, there exists $\eta > 0$ such that*

$$\int_0^{1-\eta} \xi''(t)\,\mathbb{E}[\Phi^{\mu_*}_{xx}(t, X_t)]\mathrm{d}t \ge \mathsf{ALG}_\xi - \epsilon\,.$$

For mean-field spin glasses, Parisi predicted that the minimizer of $\inf_{\mu\in\mathscr{U}} \mathsf{P}_\xi(\mu)$ led to the correct description of the ground state. While this was true, uniqueness (Theorem 2) required a stochastic representation of the minimization problem. Later, [68, 1] formulated an algorithm whose optimal value can be written as a stochastic optimal control problem dual to Equation (1.3). Surprisingly, the study of overlap gap phenomena is also linked to Equation (1.3) [50, 55], so this algorithm is optimal among a broad family.[15]

[14] Here, $(B_t)_{t\ge0}$ is standard Brownian motion.
[15] This is the family of *overlap-concentrated* algorithms; see Definition 4.

In a sense, the optimal message-passing algorithm updates its local information in a way concordant with the Parisi minimizer μ_*, *assuming* that μ_* exists. Nonetheless, even if Assumption 1 is false, we informally argue that this algorithm succeeds. Since Φ^μ_x and Φ^μ_{xx} are continuous in μ (e.g., [20, Theorem 4]), using a near-minimizer $\widetilde{\mu}_*$ of Equation (1.3) changes the value in Lemma 30 by a negligible amount, which can be absorbed into the choice of ϵ. Furthermore, $\widetilde{\mu}_*$ can be precomputed, since finding $\widetilde{\mu}_*$ is a task required only once per *CSP predicate* f, not once per run of the algorithm. Mathematizing this argument, however, is a valuable open problem; see [1, Section 6].

In this section, we prove that the algorithm achieves a value affine to Lemma 30 after taking two sets of limits.

5.1 A finite difference equation

Consider discrete versions of Equations (5.1) and (5.2). Choose $\delta > 0$, and consider independent *standard* Gaussians $\{B_1, B_2, \dots\}$. Then consider the finite difference

$$X^\delta_{\ell+1} - X^\delta_\ell = \xi''(\delta\ell)\mu_*(\delta\ell)\Phi^{\mu_*}_x(\delta\ell, X^\delta_\ell)\cdot\delta + \sqrt{\xi'(\delta(\ell+1)) - \xi'(\delta\ell)}\cdot B_{\ell+1} \qquad X^\delta_0 = 0\,, \tag{5.3}$$

$$Z^\delta_{\ell+1} - Z^\delta_\ell = \sqrt{\xi'(\delta(\ell+1)) - \xi'(\delta\ell)}\cdot\Phi^{\mu_*}_{xx}(\delta\ell, X^\delta_\ell)\cdot B_{\ell+1} \qquad Z^\delta_0 = 0\,. \tag{5.4}$$

The variables are named suggestively. By Section 4, the messages $u^\ell_{i\to a}, u^\ell_i$ behave as independent Gaussians as $d\to\infty$, which we use to imitate B_ℓ above. To this end, we define the sequences

$$x^{\ell+1}_i = x^\ell_i + \xi''(\delta\ell)\mu_*(\delta\ell)\Phi^{\mu_*}_x(\delta\ell, x^\ell_i)\cdot\delta + \sqrt{r}\cdot u^{\ell+1}_i \qquad x^0_i = 0\,,$$
$$x^{\ell+1}_{i\to a} = x^\ell_{i\to a} + \xi''(\delta\ell)\mu_*(\delta\ell)\Phi^{\mu_*}_x(\delta\ell, x^\ell_{i\to a})\cdot\delta + \sqrt{r}\cdot u^{\ell+1}_{i\to a} \qquad x^0_{i\to a} = 0\,.$$

We show that in the limit $d\to\infty$, the messages $\{u^\ell_i, u^\ell_{i\to a}\}$ behave as stochastic input to Equations (5.3) and (5.4). Note that $\Phi^\mu_x(t,x)$ and $\Phi^\mu_{xx}(t,x)$ are Lipschitz in x, since all x-derivatives are of $\Phi^\mu_x(t,x)$ are bounded:

▶ **Lemma 31** ([1, Lemma 6.4])**.** *For every $\mu\in\mathscr{L}$, $\epsilon>0$, and $j\ge 2$, there exists $K = K(\mu,j,\epsilon) > 0$ such that the j^{th} derivative in x of Φ^μ is within $[-K,K]$ for all $t\in[0,1-\epsilon]$ and $x\in\mathbb{R}$. In other words, $|\frac{\partial^j}{\partial x^j}\Phi^\mu(t,x)|\le K$ for all $t\in[0,1-\epsilon]$ and $x\in\mathbb{R}$.*

By induction and using Lemma 31, x^ℓ_i and $\Phi^{\mu_*}_{xx}(t,x^\ell_i)$, and $x^\ell_{i\to a}$ and $\Phi^{\mu_*}_{xx}(t,x^\ell_{i\to a})$ are Lipschitz in the messages $\{u^s_i\}_{s\le\ell}$ and $\{u^s_{i\to a}\}_{s\le\ell}$, respectively. So we can apply Theorem 29:

▶ **Proposition 32.** *Assume Property 2. Choose $1\le\ell\le L$ and any variable i. Then for every integer $k\ge 1$, we have*

$$\mathbb{E}[\Phi^{\mu_*}_{xx}(\delta\ell, x^\ell_i)^k] = \mathbb{E}[\Phi^{\mu_*}_{xx}(\delta\ell, X^\delta_\ell)^k] + o_d(1)\,.$$

We now choose the nonlinearities:

$$A^\ell_i = \Phi^{\mu_*}_{xx}(\delta\ell, x^\ell_i)\cdot\sqrt{\frac{\delta}{\nu_{\ell+1}\cdot\mathbb{E}[\Phi^{\mu_*}_{xx}(t,x^\ell_i)^2]}}\,, \tag{5.5}$$

$$A^\ell_{i\to a} = \Phi^{\mu_*}_{xx}(\delta\ell, x^\ell_{i\to a})\cdot\sqrt{\frac{\delta}{\nu_{\ell+1}\cdot\mathbb{E}[\Phi^{\mu_*}_{xx}(t,x^\ell_{i\to a})^2]}}\,. \tag{5.6}$$

These nonlinearities behave as we promised earlier. First of all, by Lemma 31, the nonlinearities are uniformly bounded for all d (presumed in Figure 2). Furthermore, by construction, the second moment of the nonlinearities satisfy Property 2. Finally, the nonlinearities do satisfy Property 1:

▶ **Proposition 33.** *Equations* (5.5) *and* (5.6) *satisfy Property 1. That is, for every* $m \in \mathbb{N}$, *we have* $\mathbb{E}[|A_i^\ell - A_{i\to a}^\ell|^m] = O_d(\frac{1}{d^{m/2}})$.

Note that with this choice of nonlinearities, the value of the spin $\boldsymbol{z}^\ell$ differs slightly from Equation (5.4) in the large degree limit. By Figure 2, the spin value evolves as

$$\widetilde{Z}_{\ell+1}^\delta - \widetilde{Z}_\ell^\delta = B_{\ell+1} \cdot \frac{\sqrt{\delta} \cdot \Phi_{xx}^{\mu_*}(\delta\ell, X_\ell^\delta)}{\sqrt{\mathbb{E}[\Phi_{xx}^{\mu_*}(\delta\ell, X_\ell^\delta)^2]}}, \quad \widetilde{Z}_0^\delta \sim N(0, \delta)\,.$$

However, the difference between Z_ℓ^δ and $\widetilde{Z}_\ell^\delta$ is inconsequential for small enough δ:

▶ **Proposition 34** ([1, Equation 5.7])**.** *Suppose Assumption 1 holds. Then there exist* $\delta_0, C > 0$ *such that for every* $\delta < \delta_0$ *and* $\ell < \frac{1}{\delta}$, *we have*

$$\left| \frac{\xi'((\ell+1)\delta) - \xi'(\ell\delta)}{\delta} \cdot \mathbb{E}[\Phi_{xx}^{\mu_*}(\delta\ell, X_\ell^\delta)^2] - 1 \right| \le C\sqrt{\delta}\,.$$

5.2 Achieving the extended Parisi value

Here we show that Theorem 26 achieves a value affine to ALG_ξ, the infimum of a Parisi minimization problem. First, note that the expectation $\mathbb{E}[A_i^\ell(u_i^{\ell+1})^2]$ can be simplified at large degree:

▶ **Proposition 35.** *For every variable* i *and* $0 \le \ell \le L-1$, *we have*

$$\left|\mathbb{E}\left[A_i^\ell(u_i^{\ell+1})^2\right] - \nu_{\ell+1} \cdot \mathbb{E}\left[A_i^\ell\right]\right| = O_d\left(\frac{1}{\sqrt{d}}\right).$$

After the large degree limit, we take a second limit sending the timestep to 0; i.e. $\delta \to 0$ or $\frac{1}{\delta} \to \infty$. As a result, the finite difference describing $x_i^\ell, x_{i\to a}^\ell$ converges to Equation (5.1), and the finite difference describing $z_i^\ell, z_{i\to a}^\ell$ converges to Equation (5.2). We bound the error when making this approximation, using a statement from [1]:[16]

▶ **Proposition 36** ([1, Proposition 5.3], see also [3, Proposition 4.8])**.** *Suppose Assumption 1 holds. Then there exists a coupling between* $(X_t, Z_t)_{t\ge 0}$ *and* $(X_\ell^\delta, \widetilde{Z}_\ell^\delta)_{\ell\in\mathbb{N}}$ *and a constant* $C > 0$ *such that for every small enough* δ *and* $\ell \le \frac{1}{\delta}$, *we have* $\mathbb{E}[|X_{\delta\ell} - X_\ell^\delta|^2] \le C\delta$ *and* $\mathbb{E}[|Z_{\delta\ell} - \widetilde{Z}_\ell^\delta|^2] \le C\delta$.

We are now in a position to simplify Theorem 26:

▶ **Theorem 37.** *Suppose Assumption 1 holds. Then for all* $\epsilon > 0$, *there exist* $d_0 \in \mathbb{N}$ *and* $\delta_0 > 0$ *such that for all* $d \ge d_0$ *and* $0 < \delta \le \delta_0$,

$$\frac{1}{\alpha n}\sum_{a\in E}\mathbb{E}\left[f(\boldsymbol{z}_{\partial a}^L)\right] \ge \mathbb{E}[f] + \frac{\mathsf{ALG}_\xi - \epsilon}{\sqrt{\alpha}}\,.$$

[16] The curious reader may refer to Table 1 to convert between the notation here and in [1].

Proof. Recall that $\alpha = \frac{d}{r}$ is the clause density. By Theorem 26 and Proposition 35, we have

$$\frac{1}{\alpha n}\sum_{a\in E}\mathbb{E}\left[f(\boldsymbol{z}^L_{\partial a})\right] = \mathbb{E}[f] + \frac{r}{\sqrt{d}}\cdot\sum_{\ell=0}^{L-1}\left(\frac{1}{n}\sum_{i\in V}\nu_{\ell+1}\cdot\mathbb{E}\left[A_i^\ell\right]\right) + O_d\left(\frac{1}{d}\right).$$

Let $\Delta_{\ell,i} \stackrel{\text{def}}{=} \nu_{\ell+1}\cdot\mathbb{E}[A_i^\ell]$. At large enough d, Proposition 32 and Proposition 34 imply the following for all i:[17]

$$\Delta_{\ell,i} = \nu_{\ell+1}\cdot\mathbb{E}[\Phi^{\mu_*}_{xx}(\delta\ell, X^\delta_\ell)]\cdot\sqrt{r}(1+O_{\frac{1}{\delta}}(\sqrt{\delta})) + o_d(1).$$

By definition of $\nu_{\ell+1}$,

$$\Delta_{\ell,i} = \frac{\xi'(\delta(\ell+1))-\xi'(\delta\ell)}{\sqrt{r}}\cdot\mathbb{E}[\Phi^{\mu_*}_{xx}(\delta\ell, X^\delta_\ell)]\cdot(1+O_{\frac{1}{\delta}}(\sqrt{\delta})) + o_d(1).$$

By Proposition 36 and Lipschitzness of $\Phi^{\mu_*}_{xx}(t,x)$ in x, using the continuous $X_{\delta\ell}$ introduces multiplicative error:

$$\Delta_{\ell,i} = \frac{\xi'(\delta(\ell+1))-\xi'(\delta\ell)}{\sqrt{r}}\cdot\mathbb{E}[\Phi^{\mu_*}_{xx}(\delta\ell, X_{\delta\ell})]\cdot(1+O_{\frac{1}{\delta}}(\sqrt{\delta}))^2 + o_d(1).$$

Putting it all together,

$$\begin{aligned}
&\frac{1}{\alpha n}\sum_{a\in E}\mathbb{E}\left[f(\boldsymbol{z}^L_{\partial a})\right]\\
&= \mathbb{E}[f] + \frac{r}{\sqrt{d}}\cdot\sum_{\ell=0}^{L-1}\left(\frac{1}{n}\sum_{i\in V}\Delta_{\ell,i}\right) + O_d\left(\frac{1}{d}\right)\\
&= \mathbb{E}[f] + \sqrt{\frac{r}{d}}\cdot\sum_{\ell=0}^{L-1}(\xi'(\delta(\ell+1))-\xi'(\delta\ell))\cdot\mathbb{E}[\Phi^{\mu_*}_{xx}(\delta\ell, X_{\delta\ell})]\cdot(1+O_{\frac{1}{\delta}}(\sqrt{\delta}))^2 + o_d\left(\frac{1}{\sqrt{d}}\right)\\
&= \mathbb{E}[f] + \frac{1}{\sqrt{\alpha}}\cdot\sum_{\ell=0}^{L-1}(\xi'(\delta(\ell+1))-\xi'(\delta\ell))\cdot\mathbb{E}[\Phi^{\mu_*}_{xx}(\delta\ell, X_{\delta\ell})]\cdot(1+o_{\frac{1}{\delta}}(1)) + o_d\left(\frac{1}{\sqrt{d}}\right).
\end{aligned}$$

Applying Lemma 30 with large enough d and small enough δ proves the result. ◀

5.3 Rounding

We have nearly achieved our objective. However, the entries of $\boldsymbol{z}^L$ are *continuous*, not in $\{\pm 1\}$. We account for this by *rounding* each entry of $\boldsymbol{z}^L$ without significant impact to the value of the algorithm.

Let $\mathrm{trnc}:\mathbb{R}\to[-1,1]$ be the truncation function, defined by

$$\mathrm{trnc}(x) = \begin{cases} -1 & \text{if } x < -1,\\ x & \text{if } -1 \le x \le 1,\\ 1 & \text{if } x > 1.\end{cases}$$

We use a randomized rounding procedure R that acts independently on each vertex i: it outputs $+1$ at vertex i with probability $\frac{1+\mathrm{trnc}(z_i^L)}{2}$, and -1 with probability $\frac{1-\mathrm{trnc}(z_i^L)}{2}$. The

[17] Note that an error bound in Proposition 32 could improve the scaling beyond $o_d(1)$.

average value of the algorithm is unchanged when replacing $\mathrm{trnc}(\boldsymbol{z}^L)$ with $R(\boldsymbol{z}^L)$.[18] This is because averaging over the randomness in R, we have for every set $S \subseteq V$,

$$\mathbb{E}\left[\prod_{i\in S} R(z_i^L)\right] = \prod_{i\in S} \mathrm{trnc}(z_i^L)\,.$$

We also compare the average value of the algorithm when replacing $\boldsymbol{z}^L$ with $\mathrm{trnc}(\boldsymbol{z}^L)$; i.e. the average value of $\mathbb{E}[f(\mathrm{trnc}(\boldsymbol{z}_{\partial a}^L))] - \mathbb{E}[f(\boldsymbol{z}_{\partial a}^L)]$ across factors a. Intuitively, trnc has no effect because each entry of $\boldsymbol{z}^L$ is almost surely bounded by 1 after taking the limits $d \to \infty$ and $\delta \to 0$:

▶ **Lemma 38** ([1, Lemma 6.2 and 6.5]). *Suppose Assumption 1 holds with minimizer $\mu_* \in \mathscr{L}$, and consider the martingale Z_t in Equation (5.2) corresponding to μ_*. Then for all $t \in [0,1]$, $|Z_t| \le 1$ almost surely.*

For each factor a, the difference $\mathbb{E}[f(\mathrm{trnc}(\boldsymbol{z}_{\partial a}^L))] - \mathbb{E}[f(\boldsymbol{z}_{\partial a}^L)]$ can be split into at most 2^r terms of the form

$$\mathbb{E}\left[\left(\prod_{i\in S}\mathrm{trnc}(z_i^L)\right) - \left(\prod_{i\in S} z_i^L\right)\right],$$

for some set $S \subseteq \partial a$ where $|S| \ge 2$ (since f has no linear terms). We show that each of these terms is at most $\delta \cdot O_d(\frac{1}{\sqrt{d}})$, where δ is the timestep of the algorithm.

Arbitrarily *order* the elements of set S; then the above term can be written as

$$\begin{aligned}&\mathbb{E}\left[\left(\prod_{i\in S}\mathrm{trnc}(z_i^L)\right) - \left(\prod_{i\in S} z_i^L\right)\right]\\ &= \sum_{\kappa=1}^{|S|}\mathbb{E}\left[\left(\prod_{\iota=1}^{\kappa-1}\mathrm{trnc}(z_{S[\iota]}^L)\right)\left(\mathrm{trnc}(z_{S[\kappa]}) - z_{S[\kappa]}\right)\left(\prod_{\iota=\kappa+1}^{|S|} z_{S[\iota]}^L\right)\right].\end{aligned}$$

We show that the average value of such terms across factors a is negligible with d as $n \to \infty$.

▶ **Proposition 39.** *For every integer $L \ge 1$, clause a, and disjoint subsets $T_1, T_2 \subseteq \partial a$ such that $T_1 \cup T_2 \ne \varnothing$, we have*

$$\mathbb{E}\left[\left(\Big(\prod_{i\in T_1} z_i^L\Big)\prod_{i\in T_2}\mathrm{trnc}(z_i^L) - \Big(\prod_{i\in T_1} z_{i\to a}^L\Big)\prod_{i\in T_2}\mathrm{trnc}(z_{i\to a}^L)\right)^2\right] = O_d\left(\frac{1}{d}\right).$$

▶ **Proposition 40.** *Fix an integer $L \ge 1$ and variable i. For each $a \in \partial i$, fix disjoint subsets $T_{1,a}, T_{2,a} \subseteq \partial a \backslash i$ with $T_{1,a} \cup T_{2,a} \ne \varnothing$. Then*

$$\left|\mathbb{E}\left[(z_i^L - \mathrm{trnc}(z_i^L))\cdot\frac{1}{d}\sum_{a\in\partial i}\prod_{j\in T_{1,a}} z_j^L \prod_{j\in T_{2,a}}\mathrm{trnc}(z_j^L)\right]\right| = O_d\left(\frac{\mathbb{E}[(z_i^L - \mathrm{trnc}(z_i^L))^2]^{1/2}}{\sqrt{d}}\right).$$

Although this is enough to show that the difference

$$\frac{1}{\alpha n}\sum_{a\in E}\left(\mathbb{E}[f(\mathrm{trnc}(\boldsymbol{z}_{\partial a}^L))] - \mathbb{E}[f(\boldsymbol{z}_{\partial a}^L)]\right) = O_d(\tfrac{1}{\sqrt{d}}),$$

we want this quantity to be negligible even among terms of size $O_d(\frac{1}{\sqrt{d}})$. We show this as the timestep δ also goes to 0.

[18] For convenience, we use the notation $\mathrm{trnc}(\boldsymbol{z}^L)$ or $R(\boldsymbol{z}^L)$ to apply trnc or R entrywise on $\boldsymbol{z}^L$.

▶ **Proposition 41.** *Suppose Assumption 1 holds. Then for every integer $L \geq 1$, there is a C such that*

$$\frac{1}{n}\sum_{i\in V}\mathbb{E}[(z_i^L - \mathrm{trnc}(z_i^L))^2] \leq C\delta + o_d(1)\,.$$

Proof. Consider the pseudo-Lipschitz function $\psi(x) \stackrel{\text{def}}{=} (x - \mathrm{trnc}(x))^2$. By Theorem 29,

$$\mathbb{E}[\psi(z_i^L)] = \mathbb{E}[\psi(\widetilde{Z}_L^\delta)] + o_d(1)\,.$$

We consider the behavior as δ goes to zero. Since ψ is bounded by a quadratic function,

$$\mathbb{E}[\psi(\widetilde{Z}_L^\delta)] \leq \mathbb{E}[|\psi(Z_{\delta L})|] + \mathbb{E}[|\widetilde{Z}_L^\delta - Z_{\delta L}|^2]\,.$$

The first term is zero by Lemma 38. The statement follows by Proposition 36. ◀

We combine these statements with Theorem 37 to achieve our goal:

▶ **Theorem 42.** *Suppose Assumption 1 holds. Fix a factor a. Then for all $\epsilon > 0$, there exist $d_0 \in \mathbb{N}$ and $\delta_0 > 0$ such that for all $d \geq d_0$ and $0 < \delta \leq \delta_0$,*

$$\frac{1}{\alpha n}\sum_{a\in E}\mathbb{E}\left[f(R(\boldsymbol{z}_{\partial a}^L))\right] \geq \mathbb{E}[f] + \frac{\mathsf{ALG}_\xi - \epsilon}{\sqrt{\alpha}}\,.$$

We conclude this section by showing the value of the algorithm *concentrates* around its mean. This implies that the algorithm is not only good on average, but it finds a good solution with high probability as $n \to \infty$.

▶ **Proposition 43** (Generalization of [3, Lemma 4.10])**.** *Fix a integer $L \geq 1$. Then there exists a constant C such that for all $t \geq 0$,*

$$\mathbb{P}\left(\left|\frac{1}{\alpha n}\sum_{a\in E}\left(f(R(\boldsymbol{z}_{\partial a}^L)) - \mathbb{E}[f(R(\boldsymbol{z}_{\partial a}^L))]\right)\right| \geq t\right) \leq Ce^{-nt^2/(2C)}\,.$$

Proof. Partition the nodes of the graph G into $C = C(L)$ sets $\{W_1, \dots, W_C\}$ such that any two nodes $i, j \in W_q$ have distance at least $2L + 3$. Then the value $\sum_{a\in E}\left(f(R(\boldsymbol{z}_{\partial a}^L)) - \mathbb{E}[f]\right)$ can be split into C sums as in

$$\Delta \stackrel{\text{def}}{=} \sum_{a\in E}\left(f(R(\boldsymbol{z}_{\partial a}^L)) - \mathbb{E}[f]\right) = \frac{1}{r}\sum_{i\in V}R(z_i^L)\sum_{a\in\partial i}\mathsf{D}_{i;a}f(R(\boldsymbol{z}_{\partial a}^L)) = \frac{1}{r}\sum_{\iota=1}^{C}\Delta_\iota\,,$$

where $\Delta_\iota \stackrel{\text{def}}{=} \sum_{i\in W_\iota}R(z_i^L)\sum_{a\in\partial i}\mathsf{D}_{i;a}f(R(\boldsymbol{z}_{\partial a}^L))$. By a union bound, $\mathbb{P}((\Delta - \mathbb{E}[\Delta]) \geq t\alpha n) \leq \sum_{\iota=1}^{C}\mathbb{P}((\Delta_\iota - \mathbb{E}[\Delta_\iota]) \geq tdn/C)$. Since each Δ_ι is a sum of up to n independent variables each bounded by d, this probability is at most $C\exp\left(-\frac{(tdn/C)^2}{8nd^2}\right) = C\exp\left(-\frac{nt^2}{8C^2}\right)$. ◀

6 Open questions

The first challenge is to either prove Assumption 1, or mathematize the continuity argument that using a near-minimizer $\widetilde{\mu}_* \in \mathscr{L}_\xi$ in the algorithm incurs negligible error. This would address the chief concern of Corollary 7.

As mentioned before, our algorithm assumes the CSP predicate f has no linear terms; e.g. f cannot represent k-SAT. We expect that a preprocessing step fixes this issue, similar to [79]'s improvement to [1].

Current techniques obstructing overlap-concentrated algorithms [50, 55] are restricted to *even* Ising spin glasses and CSP predicates f. However, a recent technique [47] on spherical spin glasses seems to circumvent this issue. Applying this technique to Ising spin glasses could lead to obstructions beyond even models, which automatically transfer to $\mathsf{CSP}_f(\alpha)$ by [55]. We suspect that message-passing is optimal among *overlap-concentrated* algorithms for *all* Ising spin glasses and CSP predicates f.

Furthermore, this work deepens a curious connection between $\mathsf{CSP}_f(\alpha)$ and Ising spin glasses [26, 55]. For example, this algorithm and the one in [1] are very similar, and achieve the same performance up to affine shift. In fact, the same is true for the QAOA for any constant depth and choice of hyperparameters [11, 12]! (This is known for monomial predicates f but likely extends to all predicates.) Perhaps there is a way to characterize the equivalence in energy landscape that explains why some algorithms "can't see the difference" between the two models. This equivalence extends beyond overlap gap phenomena, which are defined only at near-optimal energies.

It seems that understanding optimal algorithms for $\mathsf{CSP}_f(\alpha)$ without asymptotics in α requires new techniques.

References

1 Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *The Annals of Probability*, 49(6):2922–2960, 2021. arXiv:2001.00904.

2 Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from the sherrington-kirkpatrick gibbs measure via algorithmic stochastic localization. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 323–334. IEEE, 2022. doi:10.1109/FOCS54457.2022.00038.

3 Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Local algorithms for maximum cut and minimum bisection on locally treelike regular graphs of large degree, 2023. arXiv:2111.06813.

4 Anurag Anshu and Tony Metger. Concentration bounds for quantum states and limitations on the QAOA from polynomial approximations. *Quantum*, 7:999, May 2023. doi:10.22331/q-2023-05-11-999.

5 Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandao, David A. Buell, Brian Burkett, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Andrew Dunsworth, Edward Farhi, Brooks Foxen, Austin Fowler, Craig Gidney, Marissa Giustina, Rob Graff, Keith Guerin, Steve Habegger, Matthew P. Harrigan, Michael J. Hartmann, Alan Ho, Markus Hoffmann, Trent Huang, Travis S. Humble, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Paul V. Klimov, Sergey Knysh, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Mike Lindmark, Erik Lucero, Dmitry Lyakh, Salvatore Mandrà, Jarrod R. McClean, Matthew McEwen, Anthony Megrant, Xiao Mi, Kristel Michielsen, Masoud Mohseni, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Murphy Yuezhen Niu, Eric Ostby, Andre Petukhov, John C. Platt, Chris Quintana, Eleanor G. Rieffel, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Kevin J. Sung, Matthew D. Trevithick, Amit Vainsencher, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, and John M. Martinis. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, October 2019. doi:10.1038/s41586-019-1666-5.

6 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. arXiv:1606.05335.

7 Antonio Auffinger, Andrea Montanari, and Eliran Subag. Optimization of random high-dimensional functions: Structure and algorithms. In Patrick Charbonneau, Enzo Marinari, Giorgio Parisi, Federico Ricci-Tersenghi, Gabriele Sicuro, Francesco Zamponi, and Marc Mézard, editors, *Spin Glass Theory and Far Beyond: Replica Symmetry Breaking After 40 Years*, chapter 29. World Scientific, 2023. arXiv:2206.10217.

8 Boaz Barak and Kunal Marwaha. Classical algorithms and quantum limitations for maximum cut on high-girth graphs. In *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*, 2022. doi:10.4230/LIPIcs.ITCS.2022.14.

9 Boaz Barak, Ankur Moitra, Ryan O'Donnell, Prasad Raghavendra, Oded Regev, David Steurer, Luca Trevisan, Aravindan Vijayaraghavan, David Witmer, and John Wright. Beating the random assignment on constraint satisfaction problems of bounded degree. In *APPROX/RANDOM 2015*, pages 110–123, 2015. arXiv:1505.03424.

10 Aseem Baranwal, Kimon Fountoulakis, and Aukosh Jagannath. Optimality of message-passing architectures for sparse graphs. *Advances in Neural Information Processing Systems*, 36:40320–40341, 2023. doi:10.48550/arXiv.2305.10391.

11 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*, 2022. arXiv:2110.14206.

12 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. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, October 2022. doi:10.1109/focs54457.2022.00039.

13 Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011. arXiv:1001.3448.

14 Erwin Bolthausen. An iterative construction of solutions of the TAP equations for the Sherrington-Kirkpatrick model. *Communications in Mathematical Physics*, 325(1):333–366, 2014. arXiv:1201.2891.

15 Adam Bouland, Bill Fefferman, Chinmay Nirkhe, and Umesh Vazirani. On the complexity and verification of quantum random circuit sampling. *Nature Physics*, 15(2):159–163, October 2018. doi:10.1038/s41567-018-0318-2.

16 Sami Boulebnane, Abid Khan, Minzhao Liu, Jeffrey Larson, Dylan Herman, Ruslan Shaydulin, and Marco Pistoia. Evidence that the Quantum Approximate Optimization Algorithm optimizes the Sherrington-Kirkpatrick model efficiently in the average case, 2025. arXiv:2505.07929.

17 Sami Boulebnane and Ashley Montanaro. Predicting parameters for the Quantum Approximate Optimization Algorithm for MAX-CUT from the infinite-size limit, 2021. arXiv:2110.10685.

18 Anne Broadbent, Joseph Fitzsimons, and Elham Kashefi. Universal blind quantum computation. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, October 2009. doi:10.1109/focs.2009.36.

19 Houssam El Cheairi and David Gamarnik. Algorithmic universality, low-degree polynomials, and Max-Cut in sparse random graphs, 2024. doi:10.48550/arXiv.2412.18014.

20 Wei Kuo Chen. Variational representations for the Parisi functional and the two-dimensional Guerra-Talagrand bound. *Annals of Probability*, 45(6):3929–3966, 2017. arXiv:1501.06635.

21 Wei-Kuo Chen, David Gamarnik, Dmitry Panchenko, and Mustazee Rahman. Suboptimality of local algorithms for a class of max-cut problems. *The Annals of Probability*, 47(3):1587–1618, 2019. arXiv:1707.05386.

22 Wei-Kuo Chen and Dmitry Panchenko. On the TAP free energy in the mixed p-spin models. *Communications in Mathematical Physics*, 362:219–252, 2018. arXiv:1709.03468.

23 Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. Generalized TAP free energy. *Communications on Pure and Applied Mathematics*, 76(7):1329–1415, 2023. arXiv:1812.05066.

24 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)*, 2022. doi:10.4230/LIPIcs.ICALP.2022.41.

25 Alexander M Dalzell, Nicola Pancotti, Earl T Campbell, and Fernando GSL Brandão. Mind the gap: Achieving a super-Grover quantum speedup by jumping to the end. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1131–1144, 2023. arXiv:2212.01513.

26 Amir Dembo, Andrea Montanari, and Subhabrata Sen. Extremal cuts of sparse random graphs. *The Annals of Probability*, 45(2):1190–1217, March 2017. doi:10.1214/15-AOP1084.

27 Jian Ding, Allan Sly, and Nike Sun. Maximum independent sets on random regular graphs. *Acta Mathematica*, 217:263–340, 2016. arXiv:1310.4787.

28 Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. *Annals of Mathematics*, 196(1):1–388, 2022. arXiv:1411.0650.

29 David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009. arXiv:0907.3574.

30 Richard Durrett. *Probability: Theory and Examples*. Cambridge University Press, fifth edition, 2019. doi:10.1017/9781108591034.

31 Sepehr Ebadi, Alexander Keesling, Madelyn Cain, Tout T Wang, Harry Levine, Dolev Bluvstein, Giulia Semeghini, Ahmed Omran, J-G Liu, Rhine Samajdar, et al. Quantum optimization of maximum independent set using rydberg atom arrays. *Science*, 376(6598):1209–1215, 2022. arXiv:2202.09372.

32 Edward Farhi, David Gamarnik, and Sam Gutmann. The Quantum Approximate Optimization Algorithm needs to see the whole graph: A typical case, 2020. arXiv:2004.09002.

33 Edward Farhi, David Gamarnik, and Sam Gutmann. The Quantum Approximate Optimization Algorithm needs to see the whole graph: Worst case examples, 2020. arXiv:2005.08747.

34 Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A Quantum Approximate Optimization Algorithm, 2014. arXiv:1411.4028.

35 Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A Quantum Approximate Optimization Algorithm applied to a bounded occurrence constraint problem, 2015. arXiv:1412.6062.

36 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. doi:10.22331/q-2022-07-07-759.

37 Edward Farhi, Sam Gutmann, Daniel Ranard, and Benjamin Villalonga. Lower bounding the MaxCut of high girth 3-regular graphs using the QAOA, 2025. doi:10.48550/arXiv.2503.12789.

38 Oliver Y Feng, Ramji Venkataramanan, Cynthia Rush, and Richard J Samworth. A unifying tutorial on approximate message passing. *Foundations and Trends® in Machine Learning*, 15(4):335–536, 2022. arXiv:2105.02180.

39 Yaotian Fu and Philip W. Anderson. Application of statistical mechanics to NP-complete problems in combinatorial optimisation. *Journal of Physics A: Mathematical and General*, 19(9):1605–1620, June 1986. doi:10.1088/0305-4470/19/9/033.

40 David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41), October 2021. doi:10.1073/pnas.2108492118.

41 David Gamarnik. Barriers for the performance of graph neural networks (gnn) in discrete random structures. a comment, 2023. arXiv:2306.02555.

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

43 David Gamarnik, Aukosh Jagannath, and Alexander S Wein. Hardness of random optimization problems for boolean circuits, low-degree polynomials, and Langevin dynamics. *SIAM Journal on Computing*, 53(1):1–46, 2024. arXiv:2004.12063.

44 David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *Proceedings of the 5th conference on Innovations in Theoretical Computer Science*, pages 369–376, 2014. arXiv:1304.1831.

45 M. B. Hastings. Classical and quantum bounded depth approximation algorithms, 2019. arXiv:1905.07047.

46 Brice Huang. *Computational Hardness in Random Optimization Problems from the Overlap Gap Property*. Master's thesis, Massachusetts Institute of Technology, 2022. URL: https://dspace.mit.edu/handle/1721.1/143164.

47 Brice Huang and Mark Sellke. Algorithmic threshold for multi-species spherical spin glasses, 2023. doi:10.48550/arXiv.2303.12172.

48 Brice Huang and Mark Sellke. Optimization algorithms for multi-species spherical spin glasses. *Journal of Statistical Physics*, 191(2):29, 2024. arXiv:2308.09672.

49 Brice Huang and Mark Sellke. Strong low degree hardness for stable local optima in spin glasses, 2025. doi:10.48550/arXiv.2501.06427.

50 Brice Huang and Mark Sellke. Tight Lipschitz hardness for optimizing mean field spin glasses. *Communications on Pure and Applied Mathematics*, 78(1):60–119, 2025. arXiv:2110.07847.

51 Neng Huang, Will Perkins, and Aaron Potechin. Hardness of sampling for the anti-ferromagnetic Ising model on random graphs, 2024. doi:10.48550/arXiv.2409.03974.

52 Misha Ivkov and Tselil Schramm. Semidefinite programs simulate approximate message passing robustly. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 348–357, 2024. arXiv:2311.09017.

53 Aukosh Jagannath and Ian Tobasco. A dynamic programming approach to the Parisi functional. *Proceedings of the American Mathematical Society*, 144(7):3135–3150, December 2015. doi:10.1090/proc/12968.

54 David Jekel, Juspreet Singh Sandhu, and Jonathan Shi. Potential Hessian ascent: The Sherrington-Kirkpatrick model. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 5307–5387. SIAM, 2025. arXiv:2408.02360.

55 Chris Jones, Kunal Marwaha, Juspreet Singh Sandhu, and Jonathan Shi. Random Max-CSPs inherit algorithmic hardness from spin glasses. In *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, 2023. doi:10.4230/LIPIcs.ITCS.2023.77.

56 Stephen P Jordan, Noah Shutty, Mary Wootters, Adam Zalcman, Alexander Schmidhuber, Robbie King, Sergei V Isakov, Tanuj Khattar, and Ryan Babbush. Optimization by decoded quantum interferometry, 2024. arXiv:2408.08292.

57 Yoshiyuki Kabashima. A CDMA multiuser detection algorithm on the basis of belief propagation. *Journal of Physics A: Mathematical and General*, 36(43):11111, 2003. doi:10.1088/0305-4470/36/43/030.

58 Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT press, 2009. URL: https://dl.acm.org/doi/abs/10.5555/1795555.

59 Florent Krzakala, Marc Mézard, Francois Sausset, Yifan Sun, and Lenka Zdeborová. Probabilistic reconstruction in compressed sensing: Algorithms, phase diagrams, and threshold achieving matrices. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(08):P08009, 2012. arXiv:1206.3953.

60 Frank R Kschischang, Brendan J Frey, and H-A Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on information theory*, 47(2):498–519, 2001. doi:10.1109/18.910572.

61 Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(7):073403, 2017. arXiv:1701.00858.

62 Filip B Maciejewski, Stuart Hadfield, Benjamin Hall, Mark Hodson, Maxime Dupont, Bram Evert, James Sud, M Sohaib Alam, Zhihui Wang, Stephen Jeffrey, et al. Design and execution of quantum circuits using tens of superconducting qubits and thousands of gates for dense Ising optimization problems. *Physical Review Applied*, 22(4):044074, 2024. arXiv:2308.12423.

63 Kunal Marwaha. Local classical MAX-CUT algorithm outperforms p=2 QAOA on high-girth regular graphs. *Quantum*, 5:437, April 2021. doi:10.22331/q-2021-04-20-437.

64 Sohir Maskey, Ron Levie, Yunseok Lee, and Gitta Kutyniok. Generalization analysis of message passing neural networks on large random graphs. *Advances in neural information processing systems*, 35:4805–4817, 2022. arXiv:2202.00645.

65 Marc Mézard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, 2009. doi:10.1093/acprof:oso/9780198570837.001.0001.

66 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. doi:10.1142/0271.

67 Marco Mondelli and Ramji Venkataramanan. Approximate message passing with spectral initialization for generalized linear models. In *International Conference on Artificial Intelligence and Statistics*, pages 397–405. PMLR, 2021. arXiv:2010.03460.

68 Andrea Montanari. Optimization of the Sherrington-Kirkpatrick Hamiltonian. *SIAM Journal on Computing*, 2021. arXiv:1812.10897.

69 Andrea Montanari, Yonina C. Eldar, and Gitta Kutyniok. Graphical models concepts in compressed sensing. *Compressed Sensing*, pages 394–438, 2012. arXiv:1011.4328.

70 Dmitry Panchenko. The Parisi ultrametricity conjecture. *Annals of Mathematics*, pages 383–393, 2013. arXiv:1112.1003.

71 Dmitry Panchenko. *The Sherrington-Kirkpatrick Model*. Springer Science & Business Media, 2013. doi:10.1007/978-1-4614-6289-7.

72 Dmitry Panchenko. On the k-sat model with large number of clauses. *Random Structures & Algorithms*, 52(3):536–542, 2018. arXiv:1608.06256.

73 Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Phys. Rev. Lett.*, 43:1754–1756, December 1979. doi:10.1103/PhysRevLett.43.1754.

74 John Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, August 2018. doi:10.22331/q-2018-08-06-79.

75 Tom Richardson and Ruediger Urbanke. *Modern Coding Theory*. Cambridge University Press, 2008. doi:10.1017/CBO9780511791338.

76 Cynthia Rush, Adam Greig, and Ramji Venkataramanan. Capacity-achieving sparse superposition codes via approximate message passing decoding. *IEEE Transactions on Information Theory*, 63(3):1476–1500, 2017. arXiv:1501.05892.

77 Ciarán Ryan-Anderson. *Quantum Algorithms, Architecture, and Error Correction*. PhD thesis, The University of New Mexico, 2018. arXiv:1812.04735.

78 Juspreet Singh Sandhu and Jonathan Shi. Sum-of-squares & Gaussian processes I: Certification, 2024. arXiv:2401.14383.

79 Mark Sellke. Optimizing mean field spin glasses with external field. *Electronic Journal of Probability*, 29:1–47, 2024. arXiv:2105.03506.

80 Subhabrata Sen. Optimization on sparse random hypergraphs and spin glasses. *Random Structures & Algorithms*, 53(3):504–536, 2018. doi:10.1002/rsa.20774.

81 Ruslan Shaydulin and Marco Pistoia. Qaoa with $n \cdot p \geq 200$. In *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 1, pages 1074–1077. IEEE, 2023. arXiv:2303.02064.

82 David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35(26):1792, 1975. doi:10.1103/PhysRevLett.35.1792.

83 Allan Sly and Nike Sun. The computational hardness of counting in two-spin models on d-regular graphs. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 361–369. IEEE, 2012. arXiv:1203.2602.

84 Eliran Subag. Following the ground states of full-RSB spherical spin glasses. *Communications on Pure and Applied Mathematics*, 74(5):1021–1044, 2021. arXiv:1812.04588.

85 Michel Talagrand. The Parisi formula. *Annals of Mathematics*, pages 221–263, 2006. URL: https://www.jstor.org/stable/20159953.

86 Ewin Tang. Dequantizing algorithms to understand quantum advantage in machine learning. *Nature Reviews Physics*, 4:692–693, 2022. doi:10.1038/s42254-022-00511-w.

87 David J Thouless, Philip W Anderson, and Robert G Palmer. Solution of 'solvable model of a spin glass'. *Philosophical Magazine*, 35(3):593–601, 1977. doi:10.1080/14786437708235992.

88 Zhihui Wang, Stuart Hadfield, Zhang Jiang, and Eleanor G. Rieffel. Quantum approximate optimization algorithm for MaxCut: A fermionic view. *Phys. Rev. A*, 97:022304, February 2018. doi:10.1103/PhysRevA.97.022304.

89 Jonathan Wurtz and Peter Love. MaxCut Quantum Approximate Optimization Algorithm performance guarantees for p>1. *Physical Review A*, 103(4), April 2021. doi:10.1103/physreva.103.042612.

90 Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016. arXiv:1511.02476.

91 Han-Sen Zhong, Hui Wang, Yu-Hao Deng, Ming-Cheng Chen, Li-Chao Peng, Yi-Han Luo, Jian Qin, Dian Wu, Xing Ding, Yi Hu, Peng Hu, Xiao-Yan Yang, Wei-Jun Zhang, Hao Li, Yuxuan Li, Xiao Jiang, Lin Gan, Guangwen Yang, Lixing You, Zhen Wang, Li Li, Nai-Le Liu, Chao-Yang Lu, and Jian-Wei Pan. Quantum computational advantage using photons. *Science*, 370(6523):1460–1463, 2020. doi:10.1126/science.abe8770.