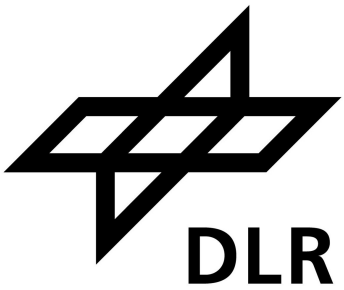
The overlap gap property and its implication for the quantum approximate optimisation algorithm

INAUGURAL-DISSERTATION ZUR
ERLANGUNG DES AKADEMISCHEN GRADES
doctor rerum naturalium (Dr. rer. nat.)
der Mathematisch-Naturwissenschaftlichen Fakultät der Universität zu Köln



vorgelegt von
Mark **Goh** Xin Hong
aus Singapur

Gutachter:	Univ.-Prof. Dr. Matthias Sperl
	Univ.-Prof. Dr. David Gross
Tag der Disputation:	09.06.2026

Teilpublikationen

- M. Goh, "The overlap gap property limits limit swapping in the QAOA," *Quantum Information & Computation*, vol. 25, no. 4, pp. 329–343, 2025. DOI: 10.2478/qic-2025-0018 [Online]. Available: https://doi.org/10.2478/qic-2025-0018
- M. Goh and T. Müller, *Mind the gap: The QAOA outperforms the MF-AOA in surpassing the OGP barrier*, forthcoming, 2026
- M. Goh and T. Müller, *Watch that gap: Parameter optimisation in the QAOA depends on the OGP*, forthcoming, 2026
- M. Goh et al., *No quantum advantage implies improved bounds and classical algorithms for the binary paint shop problem*, 2026. arXiv: 2604.00607 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2604.00607

Mark Goh Xin Hong
Köln, den 09.06.2026

Abstract

The main focus of this thesis can be framed in the following question:

> How does the presence and absence of the overlap gap property (OGP) affect quantum algorithms such as the quantum approximate optimization algorithm (QAOA)?

The OGP, a topological property that occurs when the set of near optimal solutions exhibits a form of strong clustering, is known to inhibit the performance of various classes of algorithms [5], [6]. In classical algorithms, the OGP acts a tight upper bound on many different classes of algorithms including online algorithms [7], [8] and 'stable' algorithms such as Lipschitz algorithms [9]. For quantum algorithms, the OGP is known to inhibit the algorithmic performance at shallow depth such at logarithmic depth quantum annealing [10]. This thesis focuses on how the OGP affects the QAOA and advances the understanding of the run-time required for quantum advantage. Improving upon previous results, I prove that at logarithmic depth, the OGP acts as a strict upper bound for the QAOA thus proving that there is indeed no quantum advantage at logarithmic depth [1].

Complementing the analytic result above, I similarly proved that the performance of the mean-field approximate optimization algorithm (MF-AOA), a classical algorithm inspired by the QAOA, is likewise strictly upper-bounded by the OGP [2]. Thus, for the QAOA to exhibit quantum advantage over its classical counterpart, one requires sufficiently high depth such that the QAOA is able to give an output that is greater than the OGP value. Together with my collaborator, we also provide numerical evidence that the depth required for quantum advantage in the QAOA is super-polynomial if one requires the approximation ratio to be higher than the OGP value.

While this thesis adds onto the growing body of literature of how the OGP acts as a signal of algorithmic hardness for many classes of algorithms, this work also provides the first instance of how the OGP can inform algorithmic design. This thesis provides evidence that in the presence of the OGP, the variational parameters used to optimize the QAOA's output at polynomial depth differ depending on whether one wishes to maximize the approximation ratio or to minimize the time to find the

optimal solution [3]. Understanding the right schedule is important for designing an optimal quantum control to steer a quantum state into its desired form. Our results also indicate that problems that do not exhibit the OGP are likely to be solved efficiently by quantum annealing or adiabatic quantum computing since the optimal schedule found in the QAOA follows an adiabatic schedule.

Lastly, this thesis provides a novel use of the QAOA, that of identifying use cases where efficient classical algorithms exists but are yet to be discovered. Specifically, by applying logarithmic depth QAOA to the binary paint shop problem (BPSP), we show that the performance of the QAOA surpasses that of all known polynomial time classical algorithms. However, given the failure of logarithmic depth QAOA to outperform classical algorithms on sparse optimization problems [11] such as the BPSP, this implies that there must be an efficient classical algorithm that beats all currently known heuristic methods. Applying the MF-AOA to the BPSP, we find that it significantly outperforms all known classical heuristics and applications of the QAOA (and its variants) on the BPSP [4].

Contents

Acronyms

AMP approximate message passing

AR approximation ratio

BPSP Binary Paint Shop Problem

CSP constraint satisfaction problems

LABS Low Autocorrelation Binary Sequences

MF-AOA mean-field approximate optimization algorithm

NISQ noisy intermediate-scale quantum

OGP overlap gap property

QAA quantum adiabatic algorithm

QAOA quantum approximate optimization algorithm

RSB replica symmetry breaking

TTS time-to-solution

VQA variational quantum algorithms

Symbols and Notation

$\mathbb{R}^q(n,d)$ d-regular q-uniform regular hypergraph

$\boldsymbol{\gamma}, \boldsymbol{\beta}$ Variational parameters of the QAOA

$\boldsymbol{\sigma}$ Classical spins

$\boldsymbol{x}$ Classical bitstring

$\boldsymbol{z}$ Qubits in the z basis

$\mathbb{E}_J$ Average over disorder

$\mathbb{G}^q_{ER}(n,m)$ q-uniform Erdös–Rényi hypergraph

$\mathbb{G}^q_{ER}(n,p)$ q-uniform Erdös–Rényi–Gilbert hypergraph

$\mathcal{A}$ Algorithm

$\mathcal{O}$ Order of approximation

$E_p(\boldsymbol{\gamma}, \boldsymbol{\beta})$ Expectation value of the QAOA at depth p

H Hamiltonian

J Coupling strength

n Number of qubits/size of the Hamiltonian

p Depth of QAOA circuit

$R_{\boldsymbol{\sigma\tau}}$ Overlap between $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$

X_i, Z_i Pauli Operator X/Z acting on qubit i

Chapter 1

Introduction

In this chapter, we provide an brief but concise overview on the historical development of quantum computing with a focus on quantum algorithms giving some motivation for the main research question of this thesis. We also provide an outline of what to expect in the rest of this thesis.

1.1 Quantum computing history

The start of quantum computing began in the 1980s when Benioff introduced a quantum Turing machine model [12], [13]. More famously, Feynman's speech on how 'nature isn't classical ... if you want to make a simulation of nature, you'd better make it quantum mechanical' [14] popularised the field. Focusing on quantum algorithms, the field was mainly concentrated on oracle based algorithms such as the Deutsch–Jozsa algorithm [15], the Bernstein–Vazirani algorithm [16], and Simon's algorithm [17] which were attempts to prove a clear separation between classical and quantum complexity classes. In 1994, interest in quantum computing started to gain significant traction when Shor published his algorithm that could break the widely used RSA encryption protocol in polynomial time [18], [19], something that a classical computer is unlikely to do unless $\mathbf{P} = \mathbf{NP}$. Besides Shor's algorithm, the other noteworthy algorithm is Grover's search algorithm that provides a quadratic speed up over classical algorithms for the unstructured search problem [20].

Since then, many various types of quantum algorithms have been introduced such as the HHL algorithm for linear equations [21], and the quantum singular value transformation, which generalizes many quantum algorithms such as quantum random walks and quantum signal processing [22].

1.2 Noisy Intermediate Scale Quantum (NISQ) era

While many of the quantum algorithms in section 1.1 provide some sort of quantum advantage over classical algorithms, two challenges remain. First, many of these algorithms achieve a speed up in the asymptotic run time. In practice, the constants hiding in the big $\mathcal{O}$ notation could mean that quantum advantage only occurs when the problem size is incredibly large, possibly larger than what is practically feasible. For instance, bench-marking Grover's search algorithm suggests that while there is some speed up over classical algorithms, in practice, it is often slightly smaller than the theoretical quadratic speed up and requires n to be significantly large [23]. Secondly, because quantum advantage requires the problem size to be incredibly large, we require the quantum hardware to be fault tolerant. That is to say, we require that quantum computations can be done without the fear of decoherence due to errors induced by the environment.

Current technology has not reached such a milestone yet and instead, what we have are error prone quantum computers. In other words, we are currently in the so-called noisy intermediate-scale quantum (NISQ) era, a term coined by Preskill [24]. In this era, the quantum computers that are available will have qubit numbers that are moderate in size hence the 'intermediate scale'. The number of qubits that we can build on a quantum processor are at most on the order of 10^3. For instance, Sycamore, the google gate-based quantum processor, which demonstrated quantum advantage in 2019 has 53 qubits [25] while D-wave, an annealing quantum processor, has up to 5000 qubits [26]. 'Noisy' comes about in two parts. First, the error rate that occurs whenever we implement a quantum gate operation is non-trivial with even the best performing quantum gates having an error rate of approximately 0.1% for two-qubit gate operations [27]. Without error correction, the fidelity of the quantum state after a 1000 gate operations drops to less than half. Secondly, even if the gates are error free, there is the quality of the qubits to remain in a coherent quantum state for a sufficiently long until the computation is complete. The typical decoherence time of a superconducting qubit is on the order 10^{-5}s [28], much shorter than the estimated time of a week to break RSA encription with a quantum computer [29].

Algorithms developed for current devices are known as NISQ algorithms. A widely studied class of NISQ algorithm is known as the variational quantum algorithms (VQA) [30]. This type of algorithm attempts to use a quantum computer in tandem with a classical processor to determine what quantum state to produce in order to extract the relevant information about the problem at hand. The variational in VQA comes from the variational method in quantum mechanics in the context of finding the extremal eigenstate. Typical application of VQA includes some form of

quantum machine learning, variational quantum eigensolvers, and the focus of this thesis, combinatorial optimization. Arguably, the most popular algorithm for combinatorial optimization is the quantum approximate optimization algorithm (QAOA) where the goal is to find the lowest energy eigenstate in a minimization problem. While there are now many different variants of the QAOA [31], this thesis will focus on the default version.

1.3 Outline

In chapter 2 we provide the theoretical foundations needed to understand the thesis. In particular, section 2.1 covers combinatorial optimization problems and the difference between the worst-case complexity and average-case complexity. The basics of random graph theory is covered in section 2.2 which will be useful when we consider the mean-field spin glass and diluted spin glass problems in section 2.3, the main optimization problem we consider in this thesis. Section 2.4 covers the main theoretical concept of this thesis, the overlap gap property. There we cover its relation to replica symmetry breaking and the clustering of solutions, its use to show algorithmic obstruction, and when it does fail as a heuristic for algorithmic limitations.

Chapter 3 discusses the algorithms of interest in this thesis. Namely, the QAOA and its classical counterpart, the mean-field approximate optimization algorithm (MF-AOA). In addition, we cover the locality properties of the QAOA and the proof of convergence to the optimal solution.

In chapter 4, we cover one application of the QAOA on Max-q-XORSAT, i.e. the dilute spin glass problem. We review the best known classical method, at the time of publication, to find the optimal parameters for Max-q-XORSAT up to logarithmic depth, and its relation to the optimal parameters for the mean-field spin glass problem at constant depth.

Chapter 5 begins with a recap on what is known in literature about the failure of logarithmic depth QAOA to find approximate solutions. In section 5.2, we extend this proof of logarithmic depth limitation on the random Erdös-Rényi hypergraphs to the random regular hypergraphs.

Beyond logarithmic depth optimization for the QAOA, chapter 6 covers heuristic methods for optimising the QAOA. In section 6.1, we cover current techniques of optimising the QAOA. While these methods are efficient in finding optimal parameters that optimise the two different ways of quantifying the QAOA's performance (i.e. the approximation ratio and the time-to-solution), benchmarking the QAOA's performance with respect to the time-to-solution means being able to quantify the overlap with the ground state. In practice, the ground state is not known, so these

heuristics methods only work with respect to the approximation ratio. One would hope that the variational parameters that optimises the two different metric would be approximately the same. However, we show in section 6.2 that this is only true in the absence of the OGP and that the OGP is a sufficient condition results in a significant difference in the variational parameters that optimises the approximation ratio vs. the time-to-solution.

In chapter 7, we cover some scenarios where the QAOA can offer a quantum advantage over classical algorithms. These include the QAOA's potential speed up over symmetric optimization problems [32] in section 7.1, the failure of the classical algorithms to surpass the OGP value in section 7.2, and the depth required of the QAOA to be able to give a solution that surpasses the OGP barrier in section 7.3.

Finally, in chapter 8, we apply all that we have learned in the previous chapters to an (idealised) industry problem, the binary paint shop problem (BPSP). By evaluating the performance of the QAOA on Max-2-XORSAT with constant degree 4, we improved both the known analytic and numerical upper-bound of the expected paint swap ratio for the BPSP in the thermodynamic limit. In addition, our results also suggest that there exists a classical algorithm that surpasses all currently known classical heuristics for the BPSP which we verify with the MF-AOA.

Chapter 2

Preliminaries

2.1 Constraint Satisfaction Problem (CSP)

Combinatorial optimization problems or constraint satisfaction problems (CSP) are of interest to both physicists and computer scientist. However, the language and notation varies between the two fields. Here, we primarily adopt the computer science convention where we aim to find the maximum energy of a Hamiltonian rather than the minimum, but occasionally use the physicist language when the context is clear. A table for translating between the two languages can be found in table 2.1.

Physics Notation	Computer Sci Notation
Spins $\boldsymbol{\sigma} \in \{-1, 1\}^n$	An assignment to Boolean variables
Interaction/coupling strength $J_{i_1,\ldots,i_k}$	Constraint (e.g. hyperedge)
Hamiltonian $H(\boldsymbol{\sigma})$	Value of assignment/ Objective function
Ground state energy E_{min}	Optimal Value
Mean field Model	Underlying hypergraph being complete
Dilute spin glass model	Underlying hypergraph being sparse

Table 2.1: A dictionary between the physicists spin glass models and CSPs, adapted from [33].

For now, we use a general definition of a CSP with binary inputs, i.e. $\{-1, 1\}^n$, but specify it later for specific instances.

Definition 2.1.1 (General CSP)**.** Let Λ be a distribution of functions $f : \Sigma^k \to \mathbb{R}$ with $\Sigma \in \{-1, 1\}$. A constraint is given by $J_{i_1,\ldots,i_k}$ where $i_1, \ldots, i_k$ are picked from $[n]$ with $J_{i_1,\ldots,i_k} \in \mathbb{R}$ and $[n] = [1, 2, \ldots, n]$. Draw $f \sim \Lambda$ and add the constraint e to the set of all clauses E, where e describe the clause $f_e(J_{i_1,\ldots,i_k}, \sigma_{i_1}, \ldots, \sigma_{i_k})$. The

Hamiltonian of this CSP is then given by

$$H_{CSP} = \sum_{e \in E} f_e(J_{i_1,\ldots,i_k}, \sigma_{i_1}, \ldots, \sigma_{i_k}). \tag{2.1.1}$$

Definition 2.1.2 (Satisfiability)**.** If all the constraints of a CSP can be satisfied simultaneously, we say that the CSP is *satisfiable*. Else, we say that the CSP is *unsatisfiable*.

Example 2.1.3. Let Λ contain only the predicate XOR (f = XOR function), k=2, and $J_{i_1,i_2} = J$ for some constant J. This then corresponds to a simplified Ising model without external field. The Hamiltonian is expressed as

$$H_{Is}(\boldsymbol{\sigma}) = J \sum_{(i,j) \in E} \sigma_i \sigma_j. \tag{2.1.2}$$

Given that many problems might not be satisfiable, we might still be interested to know what is the best possible satisfiable outcome (i.e. the Boolean assignment that satisfies as many constraints as possible).

Definition 2.1.4 (Optimal value of a CSP)**.** Given an instance $\mathcal{I}$ of a CSP, the optimal or maximum value is given by

$$E_{\mathrm{OPT},\mathcal{I}} = \frac{1}{n} \max_{\boldsymbol{\sigma} \in \Sigma^n} H_{\mathcal{I}}(\boldsymbol{\sigma}). \tag{2.1.3}$$

As a decision problem, if one can find a solution $\boldsymbol{\sigma}^*$ such that $H_{CSP}(\boldsymbol{\sigma}^*) > C$ for some constant C in polynomial time, we say that the problem is in the complexity class **P**. On the other hand, if we are given a possible solution $\boldsymbol{\sigma}^*$ and we can check in polynomial time if $H_{CSP}(\boldsymbol{\sigma}^*) > C$, we say that the problem is in the complexity class **NP**. It is clear that $\mathbf{P} \subset \mathbf{NP}$ but is still not known if the two complexity class are equal to each other. The widely held view is that it is the case that $\mathbf{P} \neq \mathbf{NP}$.

While computer science frequently use decision problems, real life applications often require us to phrase them as a search problem (e.g. find a solution if one such exists) and ideally as an optimisation problem (e.g. find the optimal solution for a given problem). Assuming $\mathbf{P} \neq \mathbf{NP}$, this would imply that **NP**-hard optimization problems do not have any efficient polynomial time algorithms to find the optimal solution. While an optimization problem is clearly harder than a decision problem, one would hope that relaxing the condition to require only finding approximate optimal solution, the problem would be significantly easier to solve. We state clearly now what we mean by approximate optimal solution.

Definition 2.1.5 (ϵ-optimal solution)**.** Given an instance $\mathcal{I}$ of a CSP and its optimal value $E_{\mathrm{OPT},\mathcal{I}}$, and any $\epsilon > 0$, $\boldsymbol{\sigma} \in \{-1, 1\}^n$ is said to be an *ϵ-optimal solution* if it

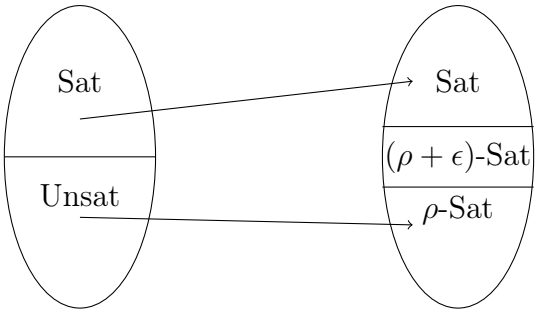


Figure 2.1: The **PCP** theorem maps instances of CSP that are satisfiable and instances of ρ-satisfied CSPs to an equivalent 3-SAT instance in polynomial time. What it leaves out are instances that satisfy $(\rho+\epsilon)$-satisfied CSPs or equivalently, arbitrary ϵ-optimal solutions. Thus, if an algorithm is able to solve a ρ-approximate solution on the right, we can invert the map to solve a ρ-approximate solution for arbitrary **NP**-Hard Problems.

satisfies the following condition

$$H_{\mathcal{I}}(\boldsymbol{\sigma}) \geq (1-\epsilon)\, E_{\mathrm{OPT},\mathcal{I}}. \tag{2.1.4}$$

Unfortunately, the **PCP** theorem shows that finding an approximate solution is just as difficult as finding the optimal solution itself. While the original interpretation of the **PCP** theorem concerns a new type of proof system as a probabilistic checkable proof [34], [35], we state below the alternative (but equivalent) formulation about the hardness of approximation.

Theorem 2.1.6 (modified from [36], PCP Theorem: Hardness of approximation)**.** *There exists a real number $\rho < 1$ such that for every $L \in$ **NP** there is a polynomial-time function f mapping boolean variables to Max-3-SAT formulas such that*

$$\begin{aligned} x \in L &\implies H(f(x)) = 1, \\ x \notin L &\implies H(f(x)) < \rho. \end{aligned}$$

What theorem 2.1.6 states is that given any problem $L \in \mathbf{NP}$, if we have a ρ-approximate algorithm for Max-3-SAT, we can convert it into an algorithm for problem L. Thus, unless $\mathbf{P} = \mathbf{NP}$, no polynomial time ϵ-approximate algorithm will exist for arbitrary $\epsilon > 0$. A graphical explanation can be found in fig. 2.1

One limitation about these computational complexity class arguments is that they study the worst case scenario. In reality, many of these **NP**-hard optimization problems can be easy to solve in practice. For instance, if one wishes to travel from Germany to New Zealand, there are a multitude of paths that one could take, but only finitely many that are reasonable and near-optimal as one can quickly figure when trying to buy flight tickets. While there are several different ways to

quantify average-case analysis [37], [38], we adopt here the approach that the underlying graph of the CSP can be generated randomly from a particular probability distribution in polynomial time. This is the topic of section 2.2.

2.2 Random Graphs

As we are primarily interested in a random instance of a CSP, this often involves generating a random hypergraph, a generalization of a graph where edges can now connect between more than 2 vertices simultaneously[1]. Given an instance of a graph $G = (V, E)$, we denote V as the set of vertices and E the set of hyperedges. For this thesis, we are primarily interested in Erdös–Rényi–(Gilbert) hypergraphs, random regular hypergraphs, and the complete hypergraph. We now give their definitions and some simple properties.

Definition 2.2.1 (degree)**.** The degree of a vertex v is defined as the total number of edges connected to v.

Definition 2.2.2 (regular graphs)**.** A graph G is d-regular for some $d \in \mathbb{Z}^+$ if every vertex in G has degree d.

Definition 2.2.3 (uniform graphs)**.** A graph G is q-uniform for some $q \in \mathbb{Z}^+$ if every hyperedge in G connects to exactly q vertices.

Definition 2.2.4 (Erdös–Rényi (ER) Hypergraphs)**.** Denote the set of Erdös–Rényi (ER) q-uniform hypergraph[2] that contains n vertices and m hyperedges by $\mathbb{G}^q_{ER}(n, m)$. An ER q-uniform hypergraph is generated by randomly selecting m hyperedges out of a uniform distribution of all possible hyperedges $E \subset V^q$.

Note that because the edges are selected randomly, the degree of a vertex follows a binomial distribution $B(m, 1/n)$. Thus it is clear that for any fixed n, it takes at most $\mathcal{O}(n^q)$ time to generate the graph. Furthermore, in the large n limit, the Poisson distribution $Po(\lambda)$ is often used to model an ER hypergraph.

Theorem 2.2.5. *In the large n limit with $np = \lambda$ held constant, $B(n, p) \approx Po(\lambda)$.*

[1] In physics language, it refers to k-body interactions where $k \geq 3$.

[2] Some authors use a similar definition but rather than fixing the number of hyperedges, they use $G = \mathbb{G}^q_{ER}(n, p)$ where each hyperedge is added with probability p.

Proof. let X be a binomial random variable with variables (n, p) and $\lambda = np$. Then

$$\begin{aligned} P(X = i) &= \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} \\ &= \frac{n!}{(n-i)!i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1)\dots(n-i+1)}{n^i} \frac{\lambda^i}{i!} \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= (1 + o(1/n)) \frac{\lambda^i}{i!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^i \\ &\xrightarrow{n\to\infty} \frac{\lambda^i}{i!} \exp\{-\lambda\}. \end{aligned}$$

which is exactly the Poisson distribution. □

Remark 2.2.6. In subsequent chapters, when the random clauses are chosen from a Poisson distribution, we are implicitly saying that the underlying graph is an ER graph in the large n limit with $np = \lambda$ held constant.

Definition 2.2.7 (Random–Regular Hypergraph)**.** Denote the set of d-regular, q-uniform hypergraphs by $\mathbb{R}^q(n, d)$ that contains n vertices each with degree d. A random regular uniform hypergraphs is selected out of the uniform distribution of all simple hypergraph that fits the regularity criteria (i.e. no repeated edges and no self-loops).

Here, we implicitly assume that $nd = qm$ for some integer m. While there are no easy unbiased methods to generate a random regular hypergraph, one can use an algorithm in the form of a while loop to generate them via the configuration model as introduced by Bollobás [39].

Definition 2.2.8 (Complete regular Hypergraph)**.** A q-uniform complete hypergraph is a graph with all possible q-hyperedges denoted by $\mathbb{K}_n^q$. The total number of hyperedges is given by the formula

$$|E| = \binom{n}{q}.$$

2.3 Spin Glass

Spin glasses are a type of magnetic alloy that are disordered. In the context of optimization, one is interested in finding the optimal energy[3] $\max_{\boldsymbol{\sigma}\in\Sigma^n} H(\boldsymbol{\sigma})$ for a given spin glass Hamiltonian $H(\boldsymbol{\sigma})$ which takes as input spin configurations $\Sigma^n =$

[3]Note that physics convention is to frame it as the ground state energy by converting it into a minimization problem rather than a maximization one.

$\{-1,1\}^n$. When the underlying graph of the Hamiltonian is a fully-connected graph, the spin glass is known as a mean-field spin glass due to its relation to the mean-field approximation. If the underlying graph is sparse, we say that we have a dilute spin glass.

2.3.1 Mean-Field spin glass

The earliest and simplest example of a mean-field spin glass was introduced by Sherrington and Kirkpatrick (SK) [40] where the Hamiltonian is defined as

$$H_{SK}(\boldsymbol{\sigma}) = \frac{1}{\sqrt{n}} \sum_{i<j}^{n} J_{ij}\sigma_i\sigma_j, \tag{2.3.1}$$

where J_{ij} for $1 \leq i < j \leq n$ are independent and identically distributed (i.i.d.) standard Gaussian random variables.

More generally, the SK model can be generalised from a two-body coupling to a q-body coupling to form the q-mean field spin glass model

$$H_q(\boldsymbol{\sigma}) = \sqrt{\frac{q!}{2n^{(q-1)}}} \sum_{j<k<\cdots<q} J_{jk\ldots q}\sigma_j\sigma_k\ldots\sigma_q. \tag{2.3.2}$$

While Sherrington and Kirkpatrick found a solution to their model via the so-called replica symmetric trick, it was quickly found to only be valid in the high-temperature regime. The trick stems from the formula

$$\mathbb{E}_J F(\beta) = \mathbb{E}_J \frac{1}{\beta n} \log Z = \lim_{N\to 0} \frac{1}{\beta n}\frac{1}{N} \log \mathbb{E}_J Z^N, \tag{2.3.3}$$

with Z the canonical partition function, $F(\beta)$ the free energy, and where

$$\mathbb{E}_J A = \int dJ\ p(J)\ A(J), \tag{2.3.4}$$

with $p(J)$ the probability distribution of the disorder J.

As part of the calculations, one defines the overlap between two spin configurations as follows

Definition 2.3.1 (Overlap)**.** Given two configuration $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \{-1,1\}^n$, define their mutual overlap as

$$R_{\boldsymbol{\sigma\tau}} = \frac{1}{n}\sum_{i=1}^{n} \sigma_i\tau_i. \tag{2.3.5}$$

One can also express the overlap in terms of the Hamming distance between two

configurations as follows

$$\frac{H_{\boldsymbol{\sigma\tau}}}{n} = \frac{1 - R_{\boldsymbol{\sigma\tau}}}{2}, \tag{2.3.6}$$

where $H_{\boldsymbol{\sigma\tau}}$ is the Hamming distance between states $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$.

Besides the obvious interpretation of the overlap between two spin configurations as a measure of how correlated they are, the overlap also quantifies correlation between Hamiltonian instances. This is a product of using the replica trick where we create N copies of the partition function Z in order to simplify the calculation of the free energy as done in eq. (2.3.3). Initially, we have i.i.d. replicas but coupled sites. Upon averaging over the disorder $\mathbb{E}_J Z^N$, the individual lattice sites become uncoupled but we now coupled the replicas [41]. Thus, the overlap between two replicas a, b comes now naturally as

$$R_{ab} = \frac{1}{n} \sum_{i=1}^{n} \sigma_i^a \sigma_i^b. \tag{2.3.7}$$

Since we have N replicas, denote the $N \times N$ matrix of overlaps by Q_{ab}. *Prima facie*, since the replicas should all be equivalent, one assumes a replica symmetry form whose matrix elements should be permutation invariant. Thus, the ansatz that Sherrington and Kirkparick made for the overlap matrix Q has the form

$$Q_{ab} = q_0 + (1 - q_0)\delta_{ab}. \tag{2.3.8}$$

The self-overlap of a replica is 1 while the overlap of any two different replicas has value q_0 determined by solving a saddle point equation.

Remark 2.3.2. Similarly to the relation between overlap and Hamming distance of two states, a corresponding view also exists for the replicas in terms of mean square displacement [42]. For the replica symmetric solution, the averaged mean square displacement probability distribution of the replicas has the form

$$P(q) = \delta(q - q_0). \tag{2.3.9}$$

We know now that the replica symmetric solution is valid only when there is no ergodicity breaking [41]. That is to say, the replica symmetric solution is valid only when one can reach any arbitrary state for all possible initial states. When ergodicity breaking happens, the phase space is broken up into separate unconnected states. The exact solution for the mean-field spin glass was conjectured by Parisi via the replica symmetry breaking (RSB) solution [43], and proven by Talagrand to give the correct solution for the even q case [44] before being generalised to all q by

Panchenko [45].

Since ergodicity breaking implies that the states belong to different parts of the phase space, it is intuitive that spin configurations in the same state should have a larger overlap than comparing spin configurations that are from different states. Thus, it is clear that the ansatz eq. (2.3.8) has to take on a different form leading to the RSB solution. For instance, the 1-RSB characterization of the overlaps would have the form $Q_{aa} = 1$, $Q_{ab} = q_1$ for spin configurations in the same state, and $Q_{ab} = q_0 < q_1$ for overlaps in different states. For instance, suppose that the number of replica inside each state is 4, then the corresponding Q matrix has the following form

$$\begin{bmatrix} 1 & q_1 & q_1 & q_1 & & & & & \\ q_1 & 1 & q_1 & q_1 & & & q_0 & & \cdots \\ q_1 & q_1 & 1 & q_1 & & & & & \\ q_1 & q_1 & q_1 & 1 & & & & & \\ & & & & 1 & q_1 & q_1 & q_1 & \\ & & q_0 & & q_1 & 1 & q_1 & q_1 & \\ & & & & q_1 & q_1 & 1 & q_1 & \\ & & & & q_1 & q_1 & q_1 & 1 & \\ & \vdots & & & & & & & \ddots \end{bmatrix} \tag{2.3.10}$$

This procedure can be iterated multiple times to form a k-RSB and if done infinitely many times, we get what is known as full-RSB or ∞-RSB. In full-RSB, this implies that the overlap spectrum is now a continuous function and similarly for the mean-square displacement.

Thus far, I have given a very physically motivated picture as to what the replica symmetric and the RSB solutions are. However, it is often easier to re-express the problem in terms of a stochastic differential equation rather than to work directly with the free energy. We now re-express the Parisi solution as found by Auffinger and Chen [46].

Definition 2.3.3 (Parisi functional [46])**.** Given a function γ in

$$\mathcal{U} = \left\{ \gamma : [0,1) \to \mathbb{R}_{\geq 0} : \gamma \text{ is right-continuous and non-decreasing}, \int_0^1 \gamma(t)dt < \infty \right\}, \tag{2.3.11}$$

the Parsi functional $\mathcal{P}$ is defined as

$$\mathcal{P}(\gamma) = \Phi(0,0) - \frac{1}{2}\int_0^1 s\, \xi''(s)\gamma(s)ds, \tag{2.3.12}$$

where $\Phi(x,t) : \mathbb{R} \times [0,1] \to \mathbb{R}$ is the solution of the Parisi PDE[4]

$$\partial_t \Phi(x,t) + \frac{\xi''}{2}\left(\partial_x^2 \Phi(x,t) + \gamma(t)\left(\partial_x \Phi(x,t)\right)^2\right) = 0 \tag{2.3.13}$$

$$\Phi(x,1) = |x| \tag{2.3.14}$$

Theorem 2.3.4 (Theorem 1 of [46]). *The following identity holds almost surely*

$$\lim_{N\to\infty} \max_{\boldsymbol{\sigma}} H_q(\sigma) = \inf_{\gamma \in \mathcal{U}} \mathcal{P}(\gamma) = \Pi_q, \tag{2.3.15}$$

where Π_q *is some constant.*

The function $\gamma \in \mathcal{U}$ has a nice physical interpretation [47]. Consider a Boltzmann distribution with temperature $1/\beta$ and let σ_1, σ_2 be two independent samples from this distribution. Then, $\beta^{-1}\gamma(t)$ is the asymptotic probability of the event $\{(\sigma_1,\sigma_2) : |R_{\sigma_1\sigma_2}| \leq t\}$ (where we first take the limit $N \to \infty$ followed by $\beta \to \infty$). Thus, the non-decreasing condition of γ naturally follows (with some rescaling by β) as the limit of a sequence of cumulative distribution functions. The measure γ is further intimately linked to the replica symmetry and RSB [6]. In the replica symmetric case, we say that the γ is a delta distribution at the origin implying that overlap $R_{\sigma\tau}$ is approximately 0 with high probability, i.e. typical pairs of solutions from i.i.d. instances are nearly orthogonal to each other. If RSB occurs, then γ is distinct from the delta structure. If full RSB occurs, then γ is strictly increasing. However, for any finite k-RSB, the cumulative distribution function is not strictly increasing but has a region that is flat. It is this flatness that results in what is known at the Overlap Gap Property which we cover in section 2.4.

2.3.2 Dilute spin glass

We now turn our attention to dilute spin glass problems. Besides the q-spin glass Hamiltonian, many CSPs such as MaxCut can be thought of as a dilute spin glass as well [48]. As an example, given a hypergraph $G = (V, E)$, the Max-q-XORSAT Hamiltonian can be defined as

$$H^q_{XOR}(\boldsymbol{\sigma}) = \sum_{(i_1,\ldots,i_q)\in E} \frac{1}{2}(1 + J_{i_1 i_2 \ldots i_q}\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_q}). \tag{2.3.16}$$

The connection between dilute spin glasses and mean-field spin glasses has a rather rich history. In fact, the proof of the Parisi formula requires one to, loosely speaking, solve the dilute spin glass. Guerra first introduced his Gaussian interpola-

[4] It is also common to see it as the Hamilton–Jacobi–Bellman equation due to its relation to the Hamilton–Jacobi equation in physics.

tion scheme by interpolating the mean-field spin glass to simpler models [49]. What followed was a series of related work [50], [51], [52] that allowed Talagrand to use such an interpolation to bound and prove the Parisi formula for the even case [44] and eventually Pachenko in its full generality [45].

Given that the Parisi solution for the mean-field spin glass is Π_q, one wonders if this value appears similarly in the optimal solution for the dilute spin glass. Using techniques from an earlier work [53], Sen [54] showed that for the Erdös–Rényi graph and random regular graph, the maximum fraction of satisfied clauses in the Max-q-Cut and Max-q-XORSAT contains Π_q in the sub-leading term with high probability, as $n \to \infty$. For example, in Max-q-XORSAT, the maximum fraction of satisfied clauses has the form

$$\frac{\max_{\boldsymbol{\sigma}} H_q^{XOR}(\boldsymbol{\sigma})}{|E|} = \frac{1}{2} + \Pi_q \sqrt{\frac{q}{2d}} + \mathcal{O}(1/\sqrt{d}), \tag{2.3.17}$$

where d denotes the degree (resp. average degree) of the random regular (resp. Erdös–Rényi) hypergraph.

More generally, Jones et al. [55] found that for any instance $\mathcal{I}$ of a Max-q-CSP with random signed constraints (i.e. $J \in \{1, -1\}$) and $m = \alpha \cdot n$ edges, for sufficiently large α, in the limit $n \to \infty$, the optimal value of the CSP is given by

$$E_{\mathrm{OPT},\mathcal{I}} = \hat{f}(\varnothing) + \Pi_q \sqrt{\frac{1}{2\alpha}} + o\left(\frac{1}{\sqrt{\alpha}}\right), \tag{2.3.18}$$

where $\hat{f}$ denotes the Fourier coefficient and $\hat{f}(\varnothing) = \mathbb{E}_{\boldsymbol{\sigma} \in \{\pm 1\}^n}[f(\boldsymbol{\sigma})]$ [56].

Remark 2.3.5. For MaxCut and Max-q-XORSAT, $\hat{f}(\varnothing) = \frac{1}{2}$.

Remark 2.3.6. In [55], the result is further generalised to allow for a mixture polynomial $H(x) = \sum_{q \geq 1} c_q^2 x^q$ so that different q's are allowed in a CSP. In this case, one has a corresponding mixture of Π_q as well in the sub-leading term. Note that our conventions of the Parisi value differ by a constant of $1/\sqrt{2}$[5].

2.4 Overlap Gap Property

Within the average-case versions of **NP**-hard optimization problems, it is still an open question as to which average case problems are easy to solve and which are hard. One heuristic partition of the two cases is known as the Overlap Gap Property (OGP). While the term was first introduced in [57], the property itself was used in

[5] This difference of $\sqrt{2}$ comes about from the mathematician vs. physicist choice of placing it in the denominator of the q mean-field spin glass to ensure that the $q = 2$ case gives us the Sherrington–Kirkpatrick model.

[58], [59] by applying statistical physics ideas on constraint satisfaction problems. A detailed survey of the history and progress of the OGP can be found in [5], [6].

The OGP is a geometric property of the solution space that, roughly speaking, states that near optimal solutions are either close to each other or far from each other and no middle ground exists. What is important is that there is some measure of distance between two solutions. What constitutes as distance depends on the problem. In the spherical-spin glass model, the distance would be Euclidean distance, whereas for our purpose of the Ising-spin glass model, it is the Hamming distance.

The OGP is of particular interest for computer scientists because in some problems, the point at which the OGP happens coincides precisely where known algorithms thresholds occur. For example, in the maximum independent set problem, the largest independent set in $\mathbb{G}^2_{ER}(n,p) = \mathbb{G}^2_{ER}(n,1/2)$ is with high probability $2(1+o(1))\log_2 n$. Meanwhile, the best known efficient algorithm only finds a set of size $(1+o(1))\log_2 n$ which coincides precisely at where the OGP occurs [60], [61], [62]. A list of problems that are known to exhibit the OGP can be found in [5].

We now give a formal definition for the simplest version of the OGP.

Definition 2.4.1 (OGP in an instance $\mathcal{I}$ of a CSP.)**.** Let $\epsilon > 0$, with parameters $0 < \mu_1 < \mu_2 < 1$. We say that an instance $\mathcal{I}$ of a CSP exhibits the OGP if there exists ϵ, μ_1, μ_2 such that for any two ϵ-optimal solution σ, τ, it holds that their overlap is either less than μ_1 or greater than μ_2

$$|R_{\sigma\tau}| \in [0,\mu_1] \cup [\mu_2, 1]. \tag{2.4.1}$$

Alternatively, one notes that the Hamming distance for ϵ-optimal solution similarly has a gap since they are related by eq. (2.3.6). A graphical representation of the OGP can be found in fig. 2.2.

While it is possible to prove that an arbitrary instance $\mathcal{I}$ of a CSP exhibits the OGP with high probability using the above definition, one often uses more complicated versions of the OGP to improve the upper-bounds of algorithmic thresholds. For spin glass, a type of multi-OGP called the branching OGP has been developed that makes use of the ultrametric structure in the Parisi solution to find a tight Lipschitz hardness [9]. As preparation for section 2.4.2, we define the ensemble or coupled-OGP as an interpolation scheme between two i.i.d. instances to prove the limitations of local algorithms [33], [48].

Definition 2.4.2 (Ensemble/Coupled- OGP (e-OGP))**.** A set of problem instances Ξ satisfies e-OGP with parameters $\epsilon > 0, 0 \leq \mu_1 < \mu_2$ if for every pair of instances

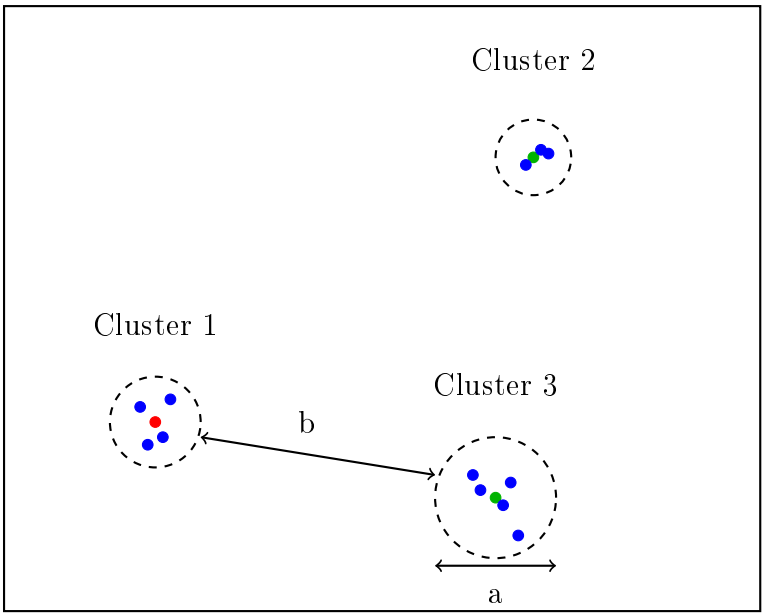


Figure 2.2: Graphical representation of the OGP. The optimal solution is found in cluster 1 and indicated by the red node. The local optimum of cluster 2/3 are given in green and every other ϵ-optimal solution is in blue. If the OGP exists, then the Hamming distance between all possible ϵ-optimal solution lie in the interval $[0, a] \cup [b, 1]$ with $0 < a < b < 1$.

$\xi, \psi \in \Xi$, for every ϵ-optimal solution $\boldsymbol{\sigma}$ of ξ and ϵ-optimal solution $\boldsymbol{\tau}$ of ψ, we have

$$R_{\sigma\tau} \leq \mu_1 \cup R_{\sigma\tau} \geq \mu_2 \tag{2.4.2}$$

In the case when ξ, ψ are sampled i.i.d., $R_{\sigma\tau} \leq \mu_1$ is not possible.

Remark 2.4.3. One example of problem instances Ξ with a total of T instances would be to create 2 i.i.d. graph ξ_1 and ξ_T generated from the same distribution (e.g. Erdös–Rényi graph). Then, for each step $\xi_t \to \xi_{t+1}$, simply add (resp. remove) an edge if it is present (resp. absent) in ξ_T.

2.4.1 OGP, symmetry breaking, and clustering

While I have discuss previously how the RSB solution gives rise to the OGP in section 2.3.1, some clarification is needed. Note that the flatness of the Parisi measure γ in the interval $[a, b]$ means that the density of the overlap in this region is 0 after taking N to infinity. However, it does not imply the absence of such overlaps [6], merely that the measure of such overlaps is 0.

The OGP is a statement about the clustering of solutions. However, we need to clarify what exactly we mean by clustering.

Definition 2.4.4 (weak clustering property [5])**.** A model exhibits weak clustering property if there exists a subset Σ of satisfying configurations which contain all but exponentially small in N fraction of satisfying configurations, and which can be partitioned into subsets/clusters separated by order $\mathcal{O}(N)$ distances, such that within each cluster one can move between any two configuration by changing at most constantly many $\mathcal{O}(1)$ bits.

Definition 2.4.5 (strong clustering property [5])**.** A model exhibits strong clustering property if all satisfying configurations can be partitioned into clusters like in the weak clustering property but with no exceptions (i.e. Σ is the set of all satisfying assignments).

An implication of the OGP is the existence of the strong clustering property since the diameter of the solution cluster is smaller than the distance between the clusters. More precisely, the OGP states that every ϵ-optimal solution is either near in Hamming distance or far when compared to another ϵ-optimal solution. This is equivalent to stating that the ϵ-optimal solution are related via either changing $\mathcal{O}(1)$ bits or $\mathcal{O}(n)$ bits and nothing in between. Furthermore, the strong clustering property implies that every solution produced by an algorithm belongs to one of these clusters. Note that the converse need not be true. That is, clustering does not imply the existence of OGP since it could be weak clustering which allows for the existence of an exponentially small number of ϵ-optimal solutions with overlaps in the interval $[\mu_1, \mu_2]$.

2.4.2 OGP as algorithmic obstruction

We now provide the general argument one uses in the literature to show how OGP leads to algorithmic failure to find arbitrary ϵ-optimal solutions. Generally, the class of algorithms that are limited in performance by the OGP are stable algorithms, stable with respect to some small perturbation. Such classes of algorithms include Lipschitz algorithms like Langevin Dynamics and Approximate Message Passing [9], [63], [64], low-degree polynomial algorithms [62], [63], and low depth quantum algorithms [65], [66], [67]. More recently, the OGP has been extended to prove limitations of online algorithms such as a greedy algorithm that is not stable [7], [8]. As the focus of this thesis is on low depth QAOA which is stable, we will provide a sketch as to how the OGP inhibits stable algorithm and refer the reader to the relevant papers for the proof of algorithmic limitation for online algorithms.

Here, we let $\mathcal{A}(\mathcal{I}_t) = \sigma_t$ denote that an algorithm $\mathcal{A}$ takes as input an instance $\mathcal{I}_t$ and outputs a solution σ_t. Now suppose we have a family of Hamiltonian instances $\mathcal{I}_t$ parameterized by $t \in [0, T]$ such that $\mathcal{I}_0$ and $\mathcal{I}_T$ are i.i.d. instances. We say that the algorithm is δ stable with respect to $\mathcal{I}_t$ if for every t, we have $R_{\sigma_t \sigma_{t+1}} \leq \delta$ (i.e.

for each step of the coupled instances, the overlap of the algorithm's output differ by at most δ.)

Now, assume that the e-OGP holds for the family of instances $\mathcal{I}_t, t \in [0,T]$ with $R_{\sigma_0\sigma_T} \geq \mu_2$, where σ_0 is an ϵ-optimal solution for $\mathcal{I}_0$ and σ_T an ϵ-optimal solution for $\mathcal{I}_T$. In addition, we assume that $\delta \leq \mu_2 - \mu_1$ (i.e. the stability of the algorithm renders it unable to jump the gap). These assumptions allow us to show, via contradiction, that any δ-stable algorithm fails to find an ϵ-optimal solution. Consider the sequence of solutions (σ_t) produced by the algorithm. We have (1) $\forall t : R_{\sigma_t\sigma_{t+1}} \leq \delta$ by stability and (2) $R_{\sigma_0\sigma_T} \geq \mu_2$ by the e-OGP. Assume, for the sake of contradiction, that every solution σ_t generated by the algorithm is an ϵ-optimal solution. By e-OGP, there exists some $t \in [0,T]$ such that $R_{\sigma_0\sigma_t} \leq \mu_1$ and $R_{\sigma_0\sigma_{t+1}} \geq \mu_2$. This implies that $\delta \geq R_{\sigma_t\sigma_{t+1}} \geq |R_{\sigma_0\sigma_{t+1}} - R_{\sigma_0\sigma_t}| \geq \mu_2 - \mu_1$ contradicting the stability requirement. Thus, stable algorithms are unable to find arbitrary ϵ-optimal solutions.

2.4.3 Failure of the OGP as indication of hardness

We have previously shown how the OGP acts as an algorithmic threshold for stable algorithms. One might then be tempted to think that once a problem exhibits the OGP, no efficient algorithm exists. However, this has recently been shown to be false and there is an easy optimization problem that can exhibit the OGP [68]. The problem in question is the shortest $s-t$ path problem. Consider a graph $G \sim \mathbb{G}_{ER}(n, \frac{C\log n}{n})$ for $C > 1$. With high probability, the shortest path has length $OPT = (1+o(1))\frac{\log n}{\log(C\log n)}$ [68]. Let $\mathcal{S}_\epsilon(G)$ be the set of all path with length $(1+\epsilon)OPT$, then with high probability, this set exhibits the OGP under the following conditions:

Theorem 2.4.6 (Theorem 2.2 of [68]). *Let $G, G' \sim \mathbb{G}_{ER}(n, \frac{C\log n}{n})$ for $C > 1$ and G' is sampled from G with each edge or non-edge with probability $1-\rho$. Then there exists some constant C' such that for every sufficiently small $\epsilon > 0$, with probability $1 - \mathcal{O}(\frac{\log\log n}{\log n})$ all pairs $(p,p') \in \mathcal{S}_\epsilon(G) \times \mathcal{S}_\epsilon(G')$ have*

$$\frac{|p \cap p'|}{\sqrt{|p| \cdot |p'|}} \in \begin{cases} [0, C'\epsilon) & \rho < \frac{1}{(1+2\epsilon)\log n} \\ [0, C'\epsilon) \cup \{1\} & \text{otherwise.} \end{cases} \tag{2.4.3}$$

In addition, with probability $1 - \mathcal{O}(\frac{\log\log n}{\log n})$, there exists p, p' with overlap $< C'\epsilon$.

As argued in section 2.4.2, this implies that stable algorithms are unable to solve the shortest path problem. However, the authors note that there exists a degree $\mathcal{O}(\log n / \log\log n)$ polynomial algorithm that can find the optimal solution and is computable in polynomial time [68].

There is an intuitive heuristic reason why the OGP fails to predict algorithmic hardness for the shortest path and why stable algorithms fail to find the optimal path. In the Ising spin glass, Max-k-XORSAT, and maximum independent set problems, the difficulty of these problems stems from the density of edges and that the addition or removal of one edge will not drastically change the problem or its solution. In other words, the problem and the solution are stable to perturbation. This is not the case for the shortest path problem where the addition or removal of a single edge can drastically alter the solution landscape. Suppose the shortest path between two points (s, t) in an unweighted graph is some sufficiently large integer $k \gg 1$. The solution can drastically change by simply adding an edge connecting s and t directly so the shortest past is only distance 1. This shows that the solution to the shortest path problem is susceptible to perturbations unlike spin glass type problems.

Remark 2.4.7. As a real life example, consider the shortest railroad path from Frankfurt to Cologne before 2000. The shortest path would have required one to pass through Koblenz before arriving at Cologne taking a total time of roughly 2.5 hours. Today, we have a direct path connecting Frankfurt to Cologne taking only slightly under an hour[6] thus changing the travel time drastically.

More rigorously, a recent work seems to point in the direction that for the OGP to act as a good heuristic, the probability of the OGP occurring has to be at least of form $1 - \mathcal{O}(1/\text{Poly}(n))$ rather than $1 - \mathcal{O}(1/\text{Polylog}(n))$ [69] but it is still an open question. Furthermore, it also seems that the framing of the optimization problem is important. While local algorithms will fail in finding the shortest $s - t$ path due to the presence of the OGP, they are likely to succeed for the shortest path trees which contains the shortest $s - t$ path [70].

[6] Assuming no (non-relativistic) time dilation due to Deutsche Bahn.

Chapter 3

Local Algorithms

We now turn to the class of algorithms that is the focus of this thesis. While we are interested in studying the Quantum Approximate Optization Algorithm and the Mean-Field Approximation Algorithm, they both fall into the general class of what is known as local algorithms. Roughly speaking, given a starting vertex on a graph v_i, at each step of the algorithm, a local algorithm explores the neighbouring vertices (i.e. all vertices which share an edge e with v_i). While local algorithms traditionally used notions such as a *factor of i.i.d.* [48], such notions fail to capture local quantum algorithms which violate Bell inequalities and thus require a more general version of local algorithms as defined in [33]. To define what exactly constitutes a local algorithm, we first give the formal definition of the p-local neighbourhood of a vertex in a (hyper)graph.

Definition 3.0.1 (p-neighbourhood and hypergraph of radius p (Definition 2.9 of [33]))**.** Let $G = (V, E)$ be a hypergraph with $v \in V$ and $p \in \mathbb{Z}^+$. The p-neighbourhood of v is defined as

$$B(v, p) = \{w \in V | w \text{ is } p \text{ hyperedges away from } v\}. \tag{3.0.1}$$

In addition, we say that (G, v) has radius p for the minimum value of p such that $B(v, p) = G$.

Definition 3.0.2 (Generic p-local algorithm (Definition 3.1 of [33]))**.** Consider a randomised algorithm $\mathcal{A}$ that takes as input a q-uniform hypergraph $G = (V, E)$ and outputs a label from some set S to each vertex in the graph (i.e. $\mathcal{A}(G) \in S^V$). We say that $\mathcal{A}$ is generic p-local if for any set $L \subseteq V$, the following conditions hold:

- (Local distribution determination): The joint marginal distribution $(\mathcal{A}(G)_v)_{v \in L}$ is identical to $(\mathcal{A}(G')_v)_{v \in L}$, where $G' = \bigcup_{v \in L} B(v, p)$
- (Local independence): The distribution $\mathcal{A}(G)_v$ is statistically independent of the joint distribution of $\mathcal{A}(G)_{v'}$ over all vertices $v' \notin B(v, 2p)$

3.1 QAOA

The Quantum Approximate Optimization Algorithm (QAOA) is a heuristic quantum algorithm developed by Farhi et al. [71] that produces approximate solutions to combinatorial optimization problems. It is a type of algorithm known as Variational Quantum Algorithms (VQAs) [30] which are of interest for the current Noisy Intermediate-Scale Quantum (NISQ) computers where fault-tolerant algorithms like Shor's are not feasible. As a type of VQA, the variational method is used to obtain the ground state energy with respect to a trial wave function. The wave function contains one or more parameters to be optimised hence the variational part.

Suppose we have a combinatorial optimization problem given by n bits and m clause C. Then we define the quantum cost function as

$$\hat{H}_C(\boldsymbol{z}) = \sum_{\alpha=1}^{m} \hat{C}_\alpha(\boldsymbol{z}), \tag{3.1.1}$$

where $\boldsymbol{z} = z_1 z_2 \dots z_n$ is an n qubit state $\boldsymbol{z} \in \{-1, 1\}^n$, and $\hat{C}$ the quantum operator of the corresponding classical clause C.

Similarly, denote sum of all single bit flip operators as

$$\hat{H}_B = \sum_{j=1}^{n} \hat{X}_j. \tag{3.1.2}$$

Define the unitary operators $\hat{U}(\gamma, \hat{H}_C), \hat{U}(\beta, \hat{H}_B)$ with $\beta, \gamma \in [0, 2\pi)$

$$\hat{U}(\gamma, \hat{H}_C) = e^{-i\gamma \hat{H}_C} = \prod_{\alpha=1}^{m} e^{-i\gamma \hat{C}_\alpha} \qquad \hat{U}(\beta, \hat{H}_B) = e^{-i\beta \hat{H}_B} = \prod_{j=1}^{n} e^{-i\beta \hat{X}_j}. \tag{3.1.3}$$

For the default QAOA, one initialize the system of qubits in the state

$$|s\rangle = |+\rangle^{\otimes n} = \frac{1}{\sqrt{2^n}} \sum_{\boldsymbol{z}} |\boldsymbol{z}\rangle, \tag{3.1.4}$$

and apply p layers of $\hat{U}(\gamma, \hat{H}_C)$ and $\hat{U}(\beta, \hat{H}_B)$. Defining $\boldsymbol{\gamma} = \gamma_1 \gamma_2 \dots \gamma_p$ and $\boldsymbol{\beta} = \beta_1 \beta_2 \dots \beta_p$. The state is prepared as

$$|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle = \hat{U}(\beta_p, \hat{H}_B)\hat{U}(\gamma_p, \hat{H}_C) \dots \hat{U}(\beta_1, \hat{H}_B)\hat{U}(\gamma_1, \hat{H}_C) |s\rangle. \tag{3.1.5}$$

For ease of notation, we now let $\hat{U}_C(\gamma_i) \equiv \hat{U}(\gamma, \hat{H}_C)$ and $\hat{U}_B(\beta_i) \equiv \hat{U}(\beta, \hat{H}_B)$. We occasionally drop γ, β when these parameters are clear from the context.

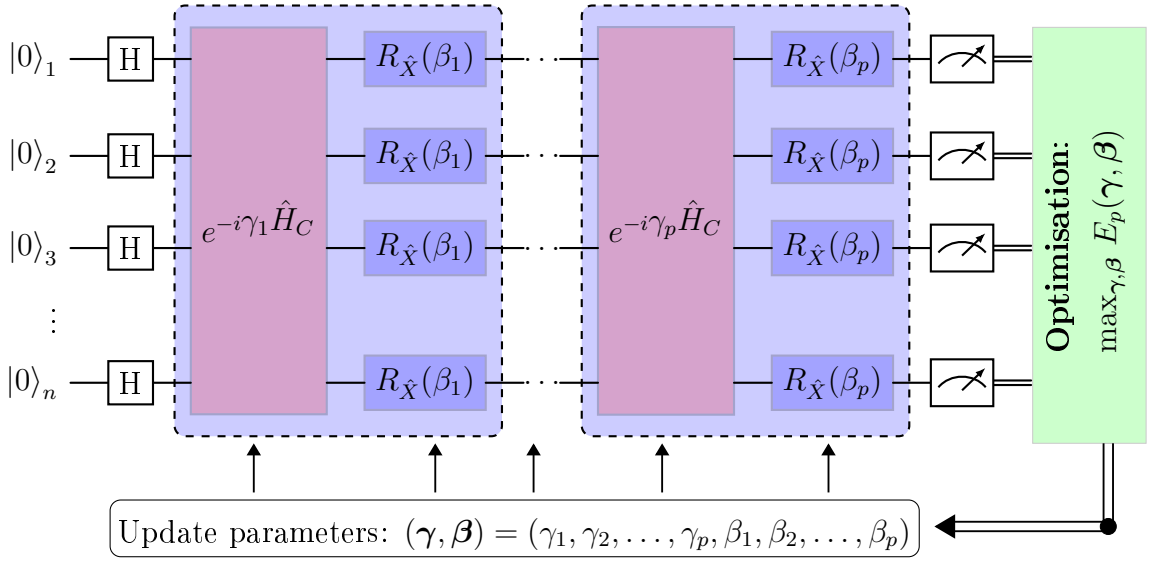


Figure 3.1: Circuit diagram of the QAOA. Note that $\exp\left\{-i\beta\hat{H}_B\right\}$ is just the single qubit $\hat{X}_i$ gate across all i qubits. Each layer which consists of $\hat{U}(\beta,\hat{H}_B)\hat{U}(\gamma,\hat{H}_C)$ is partitioned into a block for clarity.

Let the expectation of $\hat{H}_C$ in this state be E_p

$$E_p(\boldsymbol{\gamma},\boldsymbol{\beta}) = \langle\boldsymbol{\gamma},\boldsymbol{\beta}|\hat{H}_C|\boldsymbol{\gamma},\boldsymbol{\beta}\rangle. \tag{3.1.6}$$

Traditionally, the goal of the QAOA is to find the optimal parameters $(\boldsymbol{\gamma}^*,\boldsymbol{\beta}^*)$ that maximises the expectation value. An alternative way to quantify the performance of the QAOA is to measure the QAOA's overlap with the ground state $|\boldsymbol{z}^*\rangle$. These two ways of quantifying the QAOA's performance are equivalent to optimising the approximation ratio (AR) or the time-to-solution (TTS) defined as

$$\text{AR} = \frac{E_p(\boldsymbol{\gamma},\boldsymbol{\beta})}{E_{max}}, \quad \text{TTS} = \frac{p}{|\langle\boldsymbol{z}^*|\boldsymbol{\gamma},\boldsymbol{\beta}\rangle|^2}. \tag{3.1.7}$$

Remark 3.1.1. There are several variations of AR or TTS. One alternative method of defining AR is to normalised it to the lowest energy value

$$\frac{E_p(\boldsymbol{\gamma},\boldsymbol{\beta}) - E_{min}}{E_{max} - E_{min}}. \tag{3.1.8}$$

Similarly for TTS, it is sometimes defined without the numerator p since one typically assumes constant depth p or that $p \ll 1/|\langle\boldsymbol{z}^*|\boldsymbol{\gamma},\boldsymbol{\beta}\rangle|^2$. It is also useful to define it via

$$\frac{1}{|\langle\boldsymbol{z}^*|\boldsymbol{\gamma},\boldsymbol{\beta}\rangle|^2} = 2^{c(p)N}, \tag{3.1.9}$$

Algorithm 1 Quantum Approximate Optimization Algorithm **QAOA**$(\hat{H}_C, \hat{H}_B, p, \boldsymbol{\gamma}, \boldsymbol{\beta})$

Require: Number of layers p and $2p$ variational parameters.
Ensure: Solution arbitrarily close to ground state of $\hat{H}_C$.

1. Prepare ground state $|s\rangle$
2. Apply p layers of alternating unitary $\hat{U}_B(\beta_i)\hat{U}_C(\gamma_i)$ to get $|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle$
3. Measure repeatedly to obtain $E_p(\boldsymbol{\gamma}, \boldsymbol{\beta})$
4. Use a classical optimizer to update $\boldsymbol{\gamma}, \boldsymbol{\beta}$ and repeat the process again until within acceptable errors.

where $c(p)$ is a function of the QAOA circuit depth p in order to study the scaling exponent.

In the original QAOA, one takes repeated measurements of $E_p(\boldsymbol{\gamma}, \boldsymbol{\beta})$ before using a classical optimizer to update the parameters in order to find the optimal value. The circuit diagram of a p-layer QAOA can be found in fig. 3.1 while a summary of the algorithm is given in algorithm 1. Heuristic methods to choose $\boldsymbol{\gamma}, \boldsymbol{\beta}$ can be found for example in [72] using interpolation methods and have been generalised in [73].

The QAOA has been applied to several optimization problems such as MaxCut [71], Maximum Independent Set [65], Binary Paint Shop Problem [74], and spin glasses [66], [75]. However, whether the QAOA or its variants offer any advantage over classical algorithms is still an open problem. For instance, it has been shown that logarithmic depth local quantum algorithms like the QAOA have no quantum advantage in sparse optimization problems [11].

3.1.1 Proof of convergence

One might see the similarity to the quantum adiabatic algorithm (QAA). Here, the starting state $|s\rangle$ is the same and is the lowest energy state of H_B and we are similarly looking for the lowest energy state of H_C. This similarity is not coincidental as the QAOA was developed as a heuristic algorithm based on the QAA [71]. In fact, one requires the reduction of the QAOA to the QAA to show prove of convergence. We follow the work of Binkowski et al. that made the proof of convergence rigorous [76].

Proof. We begin the proof first with the following lemma.

Lemma 3.1.2 (Lemma 11 of [76]). *For any $\epsilon > 0$, let $\hat{V}_i, \hat{W}_i$ be families of unitary operators such that*

$$\|\hat{V}_i - \hat{W}_i\| < \epsilon. \tag{3.1.10}$$

Then the following estimate is valid:

$$\left\| \prod_{i=1}^{p} \hat{V}_i - \prod_{i=1}^{p} \hat{W}_i \right\| < p\epsilon \tag{3.1.11}$$

Proof. Assuming eq. (3.1.10), we can thus find a linear operator $\hat{R}_i$ with $\|\hat{R}_i\| \leq 1$ and $\forall i : \hat{V}_i = \hat{W}_i + \epsilon \hat{R}_i$. We now prove the identity via induction. The base case $p = 1$ is trivial. Now we assume it hold for p and prove for $p + 1$:

$$\begin{aligned}
\left\| \prod_{i=1}^{p+1} \hat{V}_i - \prod_{i=1}^{p+1} \hat{W}_i \right\| &= \left\| \left(\prod_{i=1}^{p} \hat{V}_i \right) \hat{V}_{p+1} - \left(\prod_{i=1}^{p} \hat{W}_i \right) \hat{W}_{p+1} \right\| \\
&\leq \left\| \hat{V}_{p+1} \left(\prod_{i=1}^{p} \hat{V}_i - \prod_{i=1}^{p} \hat{W}_i \right) \right\| + \left\| (\hat{V}_{p+1} - \hat{W}_{p+1}) \prod_{i=1}^{p} \hat{W}_i \right\| \\
&\leq \left\| \prod_{i=1}^{p} \hat{V}_i - \prod_{i=1}^{p} \hat{W}_i \right\| + \left\| (\hat{V}_{p+1} - \hat{W}_{p+1}) \right\| \\
&\leq p\epsilon + \epsilon \\
&= (p+1)\epsilon.
\end{aligned}$$

By induction, this is true for all $p \in \mathbb{N}$. □

The time dependent Hamiltonian with $t \in [0, T]$ in the QAA is given by

$$\hat{H}_{QAA}(t) = (1 - t)\hat{H}_C + t\hat{H}_B. \tag{3.1.12}$$

First, we note that $\hat{H}_B$ has only positive off-diagonals. By Perron–Frobenius theorem, there exists a unique largest real eigenvalue. It follows that the difference between the largest and second largest eigenvalue is strictly greater than 0 (i.e. there exists a gap in the energy spectrum).

Let $\hat{U}_{QAA}(t)$ denote the unitary operator associated with the adiabatic evolution. We then express the Hamiltonian of the QAA as a discretised version with sufficiently high depth q so that, for any $\epsilon > 0$, we get $\epsilon/2$ close to the QAA

$$\left| \prod_{j=1}^{q} \exp\left\{ -i\hat{H}_{QAA}\left(\frac{jT}{q}\right)\frac{jT}{q} \right\} - \hat{U}_{QAA}(T) \right| < \frac{\epsilon}{2} \tag{3.1.13}$$

Set $\hat{V}_j = \exp\left\{ -i\hat{H}_{QAA}\left(\frac{jT}{q}\right)\frac{jT}{q} \right\}$. We then express each time-step $s \in [1, q]$ of

the discretized quantum adiabatic algorithm as a Trotterize approximation

$$\begin{aligned}&\exp\left\{-i\left((1-s)\hat{H}_B + s\hat{H}_C\right)s\right\}\\&= \lim_{n\to\infty}\left(\exp\left\{-i\left(\frac{1-s}{n}s\right)\hat{H}_B\right\}\exp\left\{-i\frac{s}{n}s\ \hat{H}_C\right\}\right)^n.\end{aligned} \tag{3.1.14}$$

Similarly, the Trotter product formula implies that for any $s \in [1, q]$, there exists a $n_s \in \mathbb{Z}^+$ such that for any $n_p > n_s$, we have

$$\left|\left(\exp\left\{-i\left(\frac{1-\frac{jT}{q}}{n_p}\frac{jT}{q}\right)\hat{H}_B\right\}\exp\left\{-i\frac{\frac{jT}{q}}{n_p}\frac{jT}{q}\ \hat{H}_C\right\}\right)^{n_p} - \hat{V}_j\right| < \frac{\epsilon}{2q}. \tag{3.1.15}$$

Choose the maximum n_s among each s so that we have a total of $p = n_s^q$ angles with $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)$ and

$$(\gamma_i)_j = \frac{1-\frac{ijT}{q}}{n_s}\frac{ijT}{q} \tag{3.1.16}$$

$$(\beta_i)_j = \frac{\frac{ijT}{q}}{n_s}\frac{ijT}{q}, \tag{3.1.17}$$

for $i \in [1, q]$ and $j \in [1, n_s]$.

Denote the QAOA trotterized evolution as $V(\gamma_i, \beta_i)$ for $i \in [1, q]$ so that we have

$$\|\hat{V}(\gamma_i, \beta_i) - \hat{V}_i\| < \frac{\epsilon}{2q}. \tag{3.1.18}$$

Finally, we combine our results with theorem 3.1.2 to get

$$\left\|\prod_{i=1}^{q}\hat{V}(\gamma_i,\beta_i) - \hat{U}_{QAA}(T)\right\| \leq \left\|\prod_{i=1}^{q}\hat{V}(\gamma_i,\beta_i) - \prod_{i=1}^{q}\hat{V}_i\right\| + \left\|\prod_{i=1}^{q}\hat{V}_i - \hat{U}_{QAA}(T)\right\| \tag{3.1.19}$$

$$\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \tag{3.1.20}$$

This proves that when $p \to \infty$, we get the exact solution. □

3.1.2 Locality Properties

We now verify that the QAOA fulfil the locality conditions of theorem 3.0.2.

Recall that for the initial state $|+\rangle^{\otimes N}$, we associate each qubit as the vertex of a graph. For every hyperedge in the graph, the operators are of the form $e(z) = Z_i \ldots Z_q$. Denote the unitary operator in the QAOA as $U = U(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \prod_{i=1}^{p} e^{-i\beta_i H_B} e^{-i\gamma_i H_C}$. Consider the Heisenberg picture of the QAOA circuit so that the state $|+\rangle^{\otimes N}$ is fixed while the measurement evolves over time as $U^\dagger e(z) U$. Con-

sider $p = 1$ and $e(z) = Z_iZ_j$ for simplicity. Notice that we can commute all the operators X_k that are not i or j so the expression is simplified to

$$U^\dagger(\gamma_1)e^{i\beta_1(X_i+X_j)}Z_iZ_je^{-i\beta_1(X_i+X_j)}U(\gamma_1). \tag{3.1.21}$$

Similarly, all the terms in $U(\gamma_1)$ that do not involve Z_i, Z_j commute through. In a graphical picture, the only operators we need to keep are the edges adjacent to Z_iZ_j and Z_iZ_j itself.

Denote the measurement of a qubit v in the computational basis as $Z(v)$. Because the measurement statistics is independent of the ordering of measurement (i.e. $Z(v)$ commutes with $Z(w)$ for any qubit w), the Heisenberg picture $U^\dagger Z(v)U$ must similarly commute. Further, for $L \subseteq V$, let $S(L) = \{U^\dagger Z(v)U | v \in L\}$.

We first show that the joint marginal distribution of any set $L \subseteq V$ is identical to the $p-$local neighbourhood of L. Consider the measurement of the qubits in $S(L)$. Since the action of U on L on depends on vertices that are at most p-distance away, the distribution of the output is entirely determined by the p-neighbourhood of L.

Now we show that the distribution is statistically independent of vertices outside of the $2p$-local neighbourhood. Consider the set $L' \subseteq V\backslash B(v, 2p)$. As argued above, $S(L)$ acts only on the p-neighbourhood of L so L' is outside of the action of $S(L)$. Similarly, $S(L')$ does not act on any vertices in the p-radius of L. Thus, the measurement of $S(L)$ and $S(L')$ are independent of each other.

3.2 MF-AOA

The Mean-Field Approximate Optimization Algorithm (MF-AOA) [77] is a quantum-inspired classical limit of the QAOA where the time evolution of the quantum spins is now replaced by classical spin dynamics via the mean-field approximation.

For a concrete example of how the MF-AOA works, consider the mean-field q spin glass with spins $\boldsymbol{Z} \in \{-1, 1\}^N$ with a local magnetic field $h_i \in \mathbb{R}$:

$$H_{h,q}(\boldsymbol{Z}) = \sum_{i=1}^{N} h_iZ_i + \sum_{j<k<\cdots<q} J_{jk\ldots q}Z_k \ldots Z_q, \tag{3.2.1}$$

where we excluded the normalising constant for ease of notation.

As a quantum-inspired algorithm of the QAOA, one similarly uses a problem Hamiltonian and a driving Hamiltonian of the following form:

$$H_P = -H_{h,q}(\boldsymbol{Z}) \tag{3.2.2a}$$

$$H_D = -\sum_{i=1}^{N} \Delta_i X_i, \quad \Delta_i > 0. \tag{3.2.2b}$$

where the positivity of Δ_i ensures that the ground state of H_D is given by $|+\rangle^{\otimes N}$.

Define the classical spin vector as

$$\begin{aligned} \boldsymbol{n}_i(t) &= (n_i^x(t), n_i^y(t), n_i^z(t)) \\ &= (\mathrm{Tr}[\rho X_i(t)], \mathrm{Tr}[\rho Y_i(t)], \mathrm{Tr}[\rho Z_i(t)]), \end{aligned} \tag{3.2.3}$$

with ρ being the density matrix in the mean-field approximation $\rho = \otimes_{i=1}^{N} \rho_i$.

In the mean-field approximation, the total Hamiltonian has the form

$$H(t) = -\gamma(t) \sum_{i=1}^{N} \left[h_i + \sum_{j<\cdots<q} J_{ij\ldots q} n_j^z \ldots n_q^z \right] n_i^z - \beta(t) \sum_{i=1}^{N} \Delta_i n_i^x, \tag{3.2.4}$$

where $\beta(t)$ and $\gamma(t)$ are piecewise-constant functions. Define also the effective magnetization

$$m_i(t) = h_i + \sum_{1 \le i_1, \ldots, i_{p-1} \le N} J_{i,i_1,\ldots,i_{q-1}} n_{i_1}^z(t) \ldots n_{i_{q-1}}^z(t). \tag{3.2.5}$$

The dynamics of this system boils down solving the $3N$ differential equations for each spin in their effective magnetic field

$$\partial_t \boldsymbol{n}(t) = \boldsymbol{n}(t) \times \boldsymbol{B}(t), \tag{3.2.6}$$

with $B_i(t) = \langle 2(1 - s(t)), 0, 2s(t) m_i(t) \rangle$.

Remark 3.2.1. For the pure q-spin glass and Max-q-XORSAT without an external field (i.e. $\forall i \; h_i = 0$) the effective magnetization remains 0. In order for the algorithm to work, one could fix the last spin of the state e.g. $n_N^z = 1$. Alternatively, one uses a so-called catalyst Hamiltonian [78] to temporarily induce a magnetization. While the external field can be arbitrary subjected to the boundary conditions $\lambda_i(0) = \lambda_i(T_f) = 0$, it suffices to introduce an external term

$$H_{ex} = \sum_i \lambda_i s^2(t)(1 - s(t)), \tag{3.2.7}$$

where λ_i are chosen from a normal distribution $\mathcal{N}(0, \sigma^2)$.

The effective magnetization for i-th spin thus has the form

$$m_i(s) = \lambda_i s^2(t)(1 - s(t)) + \sum_{1 \le i_1, \ldots, i_{p-1} \le N} J_{i,i_1,\ldots,i_{q-1}} n_{i_1}^z(t) \ldots n_{i_{q-1}}^z(t), \tag{3.2.8}$$

The piecewise solution to the mean-field Hamiltonian is thus given by

$$n_i(p) = \prod_{t=1}^{p} V_i^D(t) V_i^P(t) n_i(0), \tag{3.2.9}$$

with

$$V_i^D(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(2\Delta_i\beta_t) & \sin(2\Delta_i\beta_t) \\ 0 & -\sin(2\Delta_i\beta_t) & \cos(2\Delta_i\beta_t) \end{pmatrix}, \tag{3.2.10a}$$

$$V_i^P(t) = \begin{pmatrix} \cos(2m_{i,t-1}\gamma_t) & \sin(2m_{i,t-1}\gamma_t) & 0 \\ -\sin(2m_{i,t-1}\gamma_t) & \cos(2m_{i,t-1}\gamma_t) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.2.10b}$$

Since the QAOA is guaranteed to find the exact solution by reduction to the adiabatic algorithm, it suffices to take linear functions of γ and β for a sufficiently long run-time T to ensure the algorithm achieves the best possible outcome. We now explain the algorithm as described in algorithm 2. Starting with N classical spins in the x-axis, for each time step, apply the mean-field evolution and calculate the new effective magnetization. Once finished, round off the z component of each spin vector to obtain the bitstring $\sigma^* = (\text{sign}(n_1^z), \ldots, \text{sign}(n_N^z))$.

Algorithm 2 Mean-Field Approximate Optimization Algorithm (H, T_f)

Require: Hamiltonian H, Runtime T_f

1. Initialize N classical spin vectors in state $n_i(0) = (1, 0, 0) \quad \forall i$.
2. **for** $t \in [0, T_f]$ **do** :
   (a) Apply mean-field evolution operators using eq. (3.2.9).
   (b) Calculate the new effective magnetization m_i using eq. (3.2.5).
3. **end for**
4. Round the z-component of the spin vectors to obtain $\sigma^* = (\text{sign}(n_1^z), \ldots, \text{sign}(n_N^z))$.
5. **return** σ^*

We note that by construction, the spins are independent from each other at the start of the algorithm. However, given a subset of vertices $L \subset V$, at each step of the algorithm, when we calculate the new effective magnetization, this is akin to exploring the neighbourhood of L by exploring the vertices connected to L via an edge. Thus, following similar arguments to section 3.1.2, one can show that the MF-AOA also satisfy the requirements of a local algorithm.

Chapter 4

Shallow depth QAOA

In this chapter, we review the recent application of the QAOA to the Max-q-XORSAT problem on random regular graphs and sparse graphs in general, and on the q mean-field spin glasses. The main emphasis is on showing the equivalence of the QAOA performance at constant depth p between the q mean-field spin glasses and the Max-q-XORSAT problem.

4.1 Max-q-XORSAT

Given a q-uniform hypergraph $G = (V, E)$ where $E \subset V^q$, and a given signed weight $J_{i_1,\ldots,i_q} \in \{-1, +1\}$, the Max-$q$-XORSAT problem can be expressed as

$$H_q^{XOR}(\boldsymbol{\sigma}) = \sum_{(i_1,\ldots,i_q)\in E} \frac{1}{2}(1 + J_{i_1 i_2 \ldots i_q} z_{i_1} z_{i_2} \ldots z_{i_q}). \tag{4.1.1}$$

where a constraint is satisfied if and only if $z_{i_1} z_{i_2} \ldots z_{i_q} = J_{i_1 i_2 \ldots i_q}$.

We say that an instance of the problem is satisfied if there is an assignment of values to the bit-string $\boldsymbol{\sigma}$ which satisfies all the clauses (i.e. $H_{XOR}^q(\boldsymbol{\sigma}) = |E|$), otherwise, we say it is unsatisfiable. It is known that for an instance $\mathcal{I}$ of a random q-XORSAT problem, given $|E|$ hyperedges and problem size n, the following theorem holds

Theorem 4.1.1 (Theorem 1 of [79]). *Let $q > 2$ be fixed. Let $|E|/n = \delta$. In the limit $n \to \infty$,*

1. *if $\delta > 1$, then a random formula from $\mathcal{I}$ is unsatisfiable with high probability.*
2. *if $\delta < 1$, then a random formula from $\mathcal{I}$ is satisfiable with high probability.*

As we will be interested in random regular hypergraphs, this condition implies that if the average degree d is greater than q, then a random formula is unsatisfiable

with high probability. Suppose we fix $d > q$ sufficiently large so that we are in the unsatisfiability regime. The maximum number of satisfiable equations in an instance of random XORSAT for both $\mathbb{G}^q_{ER}(n,p)$, and $\mathbb{R}^q(n,d)$ has been found [54] to be in the form of eq. (2.3.17) which we repeat below

$$\frac{1}{|E|}\max_{\boldsymbol{\sigma}} H^q_{XOR}(\boldsymbol{z}) = \frac{1}{2} + \Pi_q\sqrt{\frac{q}{2d}} + \mathcal{O}(1/\sqrt{d}). \tag{4.1.2}$$

4.2 QAOA on large girth graphs

Recall from section 3.1 that the QAOA is a local algorithm. Consider a $(D+1)$-regular graph with girth greater than $2p+1$ (i.e. the smallest cycle in the graph is greater than $2p+1$). In this setting, the subgraph explored by the QAOA with p layers will appear as D regular trees.

In evaluating the performance of the QAOA on large-girth $(D+1)$ regular hypergraphs, it turns out that the optimal parameter for $\boldsymbol{\gamma}$ is of order $1/\sqrt{D}$. Thus, in order to find the optimal γ_i's, it will be useful to prepare the QAOA state using a scaled unitary operator $U(\gamma, H^S_{XOR})$ with

$$H^S_{XOR} = -\frac{1}{\sqrt{D}}\sum_{(i_1,\ldots,i_q)\in E} J_{i_1 i_2 \ldots i_q} z_{i_1} z_{i_2} \ldots z_{i_q}, \tag{4.2.1}$$

where we drop the factor of $1/2$ since it just adds a phase factor.

Knowing that the asymptotic solution for the Max-q-XORSAT is eq. (2.3.17), we can thus express the output of the QAOA in a similar form. We state the theorem with said expression found by Basso et al. in [80].

Theorem 4.2.1 (Theorem 2 of [80])**.** *For H^q_{XOR} on any $(D+1)$-regular q-uniform hypergraph with girth $> 2p+1$. Let $|\boldsymbol{\gamma}_S, \boldsymbol{\beta}\rangle$ be the QAOA state generated using H^S_{XOR}. Then, for any choice of coupling J,*

$$\frac{\langle \boldsymbol{\gamma}_S, \boldsymbol{\beta}| H^q_{XOR} |\boldsymbol{\gamma}_S, \boldsymbol{\beta}\rangle}{|E|} = \frac{1}{2} + \nu^q_p(D, \boldsymbol{\gamma}_S, \boldsymbol{\beta})\sqrt{\frac{q}{2D}}, \tag{4.2.2}$$

where $\nu^q_p(D, \boldsymbol{\gamma}_S, \boldsymbol{\beta})$ is independent of J and can be numerically evaluated with an iteration using $\mathcal{O}(p4^{pq})$ time and $\mathcal{O}(4^p)$ memory. In the infinite D limit, $\lim_{D\to\infty} \nu^q_p(D, \boldsymbol{\gamma}_S, \boldsymbol{\beta}) = \nu^q_p(\boldsymbol{\gamma}_S, \boldsymbol{\beta})$ can be evaluated with an iteration using $\mathcal{O}(p^2 4^p)$ time and $\mathcal{O}(p^2)$ memory.

Remark 4.2.2. At the time of writing, I was made aware of that the algorithm to evaluate $\nu^2_p(D, \boldsymbol{\gamma}, \boldsymbol{\beta})$ is sub-optimal. By applying tensor network techniques in evaluating $\nu^2_p(D, \boldsymbol{\gamma}, \boldsymbol{\beta})$, one can improve the time complexity to $\mathcal{O}(4^p)$ [81], a quadratic speed up that could potentially apply to $\nu^q_p(D, \boldsymbol{\gamma}, \boldsymbol{\beta})$ as well!

We will now explain how one evaluates the performance of the QAOA (i.e. $\nu_p^q(D, \boldsymbol{\gamma}_S, \boldsymbol{\beta})$ before giving their proofs. First, we index vectors in the following order:

$$\boldsymbol{a} = (a_1, a_2, \ldots, a_p, a_0, a_{-p}, \ldots, a_{-2}, a_{-1}). \tag{4.2.3}$$

Define for $1 \leq r \leq p$ the $(2p+1)$ component vector

$$\Gamma_r = \gamma_r, \quad \Gamma_0 = 0, \quad \Gamma_{-r} = -\gamma_r. \tag{4.2.4}$$

Furthermore, let

$$f(\boldsymbol{a}) = \frac{1}{2} \langle a_1 | e^{i\beta_1 X} | a_2 \rangle \ldots \langle a_p | e^{i\beta_p X} | a_0 \rangle \ \langle a_0 | e^{-i\beta_p X} | a_{-p} \rangle \ldots \langle a_{-2} | e^{-i\beta_1 X} | a_{-1} \rangle \tag{4.2.5}$$

where $a_i \in \{-1, 1\}$ enumerates the computation basis states and

$$\langle a_1 | e^{i\beta X} | a_2 \rangle = \begin{cases} \cos\beta & a_1 = a_2 \\ i \sin\beta & a_1 \neq a_2. \end{cases} \tag{4.2.6}$$

Let $H_D^{(m)} : \{-1, 1\}^{2p+1} \to \mathbb{C}$ for $0 \leq m \leq p$. Let $H_D^{(0)}(\boldsymbol{a}) = 1$ and $H_D^{(m)}$ be defined recursively as follows:

$$H_D^{(m)}(\boldsymbol{a}) = \left(\sum_{\boldsymbol{b^1}, \ldots \boldsymbol{b^{q-1}}} \cos\left[\frac{1}{\sqrt{D}} \boldsymbol{\Gamma} \cdot (\boldsymbol{a} \boldsymbol{b}^1 \ldots \boldsymbol{b}^{q-1}) \right] \prod_{i=1}^{q-1} f(\boldsymbol{b}^i) H_D^{(m-1)}(\boldsymbol{b}^i) \right)^D \tag{4.2.7}$$

where we denote $\boldsymbol{abc}$ as the entry-wise product $(\boldsymbol{abc})_i = a_i b_i c_i$.

Iterating till $m = p$ gives us $H_D^{(p)}(\boldsymbol{a})$ which we use to compute

$$v_p^q(D, \gamma, \beta) = i \sqrt{\frac{D}{2q}} \sum_{\boldsymbol{a^1} \ldots \boldsymbol{a^q}} \sin\left[\frac{1}{\sqrt{D}} \boldsymbol{\Gamma} \cdot \boldsymbol{a}^1 \ldots \boldsymbol{a}^q \right] \prod_{i=1}^{q} \left[a_0^i f(\boldsymbol{a}^i) f(\boldsymbol{b}) H_D^{(p)}(\boldsymbol{a}^1) \right]. \tag{4.2.8}$$

In the infinite D limit, let $G^{(m)} \in \mathbb{C}^{(2p+1)\times(2p+1)}$ for $0 \leq m \leq p$ be defined as follows: for $j, k \in \{1, \ldots, p, 0, -p, \ldots, 1\}$, let the base case be $G_{j,k}^{(0)} = \sum_{\boldsymbol{a}} f(\boldsymbol{a}) a_j a_k$, then $G^{(m)}$ is defined recursively as

$$G_{j,k}^{(m)} = \sum_{\boldsymbol{a}} f(\boldsymbol{a}) a_j a_k \exp\left(-\frac{1}{2} \sum_{j', k' = -p}^{p} \left(G_{j',k'}^{(m-1)} \right)^{q-1} \Gamma_{j'} \Gamma_{k'} a_{j'} a_{k'} \right). \tag{4.2.9}$$

After p iterations, $G^{(p)}$ is used to compute

$$\nu_p^q(\boldsymbol{\gamma}_S, \boldsymbol{\beta}) = \frac{i}{\sqrt{2q}} \sum_{j=-p}^{p} \Gamma_j (G_{0,p}^{(p)})^q. \tag{4.2.10}$$

4.2.1 Proof of iteration

Since the subgraph explored looks like a D-regular hypertree with a total of $n = q[(q-1)^p D^p + \cdots + (q-1)D + 1]$ vertices. Focus on a single hyperedge so we want to evaluate the expression

$$\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | Z_1 \dots Z_q | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle .$$

For the QAOA of depth p, we insert $2p+1$ complete sets of the computational Z-basis with labels $\boldsymbol{z}^{[i]}, i \in [-p, p]$. For example, with $p = 1$, we have the following expression

$$\begin{aligned}
\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | Z_1 \dots Z_q | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle &= \frac{1}{2^n} \sum_{\{\boldsymbol{z}^{[i]}\}} e^{i\gamma_1 H_{XOR}^S(\boldsymbol{z}^{[\mathbf{1}]})} \left\langle \boldsymbol{z}^{[1]} \middle| e^{i\beta B} \middle| \boldsymbol{z}^{[0]} \right\rangle z_1^{[0]} \dots z_q^{[0]} \\
&\qquad \times \left\langle \boldsymbol{z}^{[0]} \middle| e^{-i\beta B} \middle| \boldsymbol{z}^{[-1]} \right\rangle e^{-i\gamma_1 H_{XOR}^S(\boldsymbol{z}^{[-\mathbf{1}]})} \\
&= \frac{1}{2^n} \sum_{\{\boldsymbol{z}^{[i]}\}} e^{i\gamma_1 H_{XOR}^S(\boldsymbol{z}^{[\mathbf{1}]}) - i\gamma_1 H_{XOR}^S(\boldsymbol{z}^{[-\mathbf{1}]})} z_1^{[0]} \dots z_q^{[0]} \\
&\qquad \prod_{v=1}^{n} \left\langle z_v^{[1]} \middle| e^{i\beta B} \middle| z_v^{[0]} \right\rangle \left\langle z_v^{[0]} \middle| e^{-i\beta B} \middle| z_v^{[-1]} \right\rangle .
\end{aligned} \tag{4.2.11}$$

Using our definition of $f(\boldsymbol{z}_v)$, $\boldsymbol{\Gamma}$, and H_{XOR}^S in eqs. (4.2.1), (4.2.4) and (4.2.5), we can re-write the expression as

$$\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | Z_1 \dots Z_q | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle = \sum_{\{\boldsymbol{z}^{[i]}\}} z_1^{[0]} \dots z_q^{[0]} \exp\left(-\frac{i}{\sqrt{D}} \sum_{j_1 \dots j_q \in E} \boldsymbol{\Gamma} \cdot (\boldsymbol{z}_{j_1}) \dots (\boldsymbol{z}_{j_q}) \right) \prod_{v=1}^{n} f(\boldsymbol{z}_v), \tag{4.2.12}$$

where in the last step, $\boldsymbol{z}_v = (z_v^{[1]}, z_v^{[0]}, z_v^{[-1]})$, thus replacing the sum over $2p+1$ complete sets to an equivalent sum over the bit configuration of each vertex. It's clear that this formula holds for any p.

In order to evaluate the expression, consider the leaf vertices $w_1, \dots, w_{q-1}$ of a hyperedge and denote their parent vertex as $p(w_1)$. Summing over the bit configuration for $z_{w_1}, \dots, z_{w_{q-1}}$ gives us

$$\sum_{z_{w_1}, \dots, z_{w_{q-1}}} \exp\left\{ -\frac{i}{\sqrt{D}} \boldsymbol{\Gamma} \cdot (\boldsymbol{z}_{p(w_1)} \boldsymbol{z}_{w_1}, \dots, \boldsymbol{z}_{w_{q-1}}) \right\} \prod_{j=1}^{q-1} f(\boldsymbol{z}_{\boldsymbol{w_j}}). \tag{4.2.13}$$

Note that this term involves the parent vertex bit configuration $\boldsymbol{z}_{p(w_1)}$. Since this is a $D+1$ regular hypergraph, there are D branching hyperedges each with equal contribution so we get

$$H_D^{(1)}(\boldsymbol{z}_{p(w_1)}) = \left(\sum_{z_{w_1},\ldots,z_{w_{q-1}}} \exp\left\{ -\frac{i}{\sqrt{D}} \boldsymbol{\Gamma} \cdot (\boldsymbol{z}_{p(w_1)} \boldsymbol{z}_{w_1}, \ldots, \boldsymbol{z}_{w_{q-1}}) \right\} \prod_{j=1}^{q-1} f(\boldsymbol{z}_{\boldsymbol{w_j}}) \right)^D . \tag{4.2.14}$$

Notice that $f(\boldsymbol{a}) = f(-\boldsymbol{a})$. With this, we can show by induction that $H_D^{(m)}(-\boldsymbol{a}) = H_D^{(m)}(\boldsymbol{a})$ too thus implying that the function is even. This allows us to simplify the expression above as

$$H_D^{(1)}(\boldsymbol{z}_{p(w_1)}) = \left(\sum_{z_{w_1},\ldots,z_{w_{q-1}}} \cos\left(\frac{1}{\sqrt{D}} \boldsymbol{\Gamma} \cdot (\boldsymbol{z}_{p(w_1)} \boldsymbol{z}_{w_1}, \ldots, \boldsymbol{z}_{w_{q-1}}) \right) \prod_{j=1}^{q-1} f(\boldsymbol{z}_{\boldsymbol{w_j}}) \right)^D . \tag{4.2.15}$$

Let $p(w_1) = u_1$. We can repeat this process again looking at the parent of u_1 and consider all the other vertices that share the same parent $p(u_1)$. Summing them up and taking into account $H_D^{(1)}(\boldsymbol{z}_{p(u_j)})$, we end up with

$$H_D^{(2)}(\boldsymbol{z}_{p(u_1)}) = \left(\sum_{z_{w_1},\ldots,z_{w_{q-1}}} \cos\left(\frac{1}{\sqrt{D}} \boldsymbol{\Gamma} \cdot (\boldsymbol{z}_{p(u_1)} \boldsymbol{z}_{u_1}, \ldots, \boldsymbol{z}_{u_{q-1}}) \right) \prod_{j=1}^{q-1} f(\boldsymbol{z}_{\boldsymbol{u_j}}) H_D^{(1)}(\boldsymbol{z}_{p(u_j)}) \right)^D . \tag{4.2.16}$$

Repeating this process until we reach the root of the tree, we can re-express eq. (4.2.12) as

$$\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | Z_1 \ldots Z_q | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle = \sum_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_q} z_1^{[0]} \ldots z_q^{[0]} \exp\left(-\frac{i}{\sqrt{D}} \boldsymbol{\Gamma} \cdot \boldsymbol{z}_1 \ldots \boldsymbol{z}_q \right) \prod_{j=1}^{q} f(\boldsymbol{z}_j) H_D^{(p)}(\boldsymbol{z}_j). \tag{4.2.17}$$

Since the expression must be the same when $\boldsymbol{z}_1 \to -\boldsymbol{z}_1$, this implies that the solution must be odd so we are left with

$$\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | Z_1 \ldots Z_q | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle = -i \sum_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_q} \sin\left(\frac{1}{\sqrt{D}} \boldsymbol{\Gamma} \cdot \boldsymbol{z}_1 \ldots \boldsymbol{z}_q \right) \prod_{j=1}^{q} z_j^{[0]} f(\boldsymbol{z}_j) H_D^{(p)}(\boldsymbol{z}_j). \tag{4.2.18}$$

As for the infinite D limit, we simply perform a Taylor expansion of the cosine

function in $H_D^{(m)}(\boldsymbol{a})$ and we use the fact that $\sum_b f(\boldsymbol{b}) H_D^{(m-1)}(\boldsymbol{b}) = 1$ [80] to get

$$\lim_{D\to\infty} \left(\sum_{\boldsymbol{b^1},\ldots\boldsymbol{b^{q-1}}} [1 - \frac{1}{2D} \left(\boldsymbol{\Gamma}\cdot(\boldsymbol{a}\boldsymbol{b}^1\ldots\boldsymbol{b}^{q-1})\right)^2 + \mathcal{O}(1/D^2)] \prod_{i=1}^{q-1} f(\boldsymbol{b}^i) H_D^{(m-1)}(\boldsymbol{b}^i) \right)^D$$
$$= \exp\left\{ -\frac{1}{2} \left(\boldsymbol{\Gamma}\cdot(\boldsymbol{a}\boldsymbol{b}^1\ldots\boldsymbol{b}^{q-1})\right)^2 \prod_{i=1}^{q-1} f(\boldsymbol{b}^i) H_D^{(m-1)}(\boldsymbol{b}^i) \right\}. \tag{4.2.19}$$

Expanding the dot product and defining $G_{j,k}^{(m)} = \sum_{\boldsymbol{a}} f(\boldsymbol{a}) H^{(m)} a_j a_k$, we can re-express the iteration as

$$G_{j,k}^{(m)} = \sum_{\boldsymbol{a}} f(\boldsymbol{a}) a_j a_k \exp\left\{ -\frac{1}{2} \sum_{j',k'=-p}^{p} \left(G_{j',m'}^{(m-1)}\right)^{q-1} \right\} \Gamma_{j'}\Gamma_{k'} a_{j'} a_{k'}. \tag{4.2.20}$$

Finally, using a Taylor expansion again on the sine function, we can give the expression of $\nu_p^q(\boldsymbol{\gamma}_S, \boldsymbol{\beta})$ in the infinite D limit as

$$\lim_{D\to\infty} \nu_p^q(\boldsymbol{\gamma}_S, \boldsymbol{\beta}) = \frac{i}{\sqrt{2q}} \sum_{k=-p}^{p} \Gamma_k \prod_{j=1}^{q} z_j^{[0]} z_j^{[k]} f(\boldsymbol{z}_j) H^{(p)}(\boldsymbol{z}_j).$$
$$= \frac{i}{\sqrt{2q}} \sum_{k=-p}^{p} \Gamma_k (G_{0,k}^{(p)})^q. \tag{4.2.21}$$

4.3 Equivalence of performance

A follow up work by some of the co-authors analysed the performance of the QAOA at constant depth on spin glass type Hamiltonians [66]. Specifically, their results require the following assumption:

Assumption 4.3.1. *Assume that the constraints $J_{i_1,\dots,i_q}$ are i.i.d. with mean zero symmetric distribution and finite second moment. Further assume that the normalised expectation value $\mathbb{E}[e^{i\lambda J_{i_1,\dots,i_q}}]$ is real and positive for large n. Denote $g_{q,n} = n^{q-1}\log \mathbb{E}[e^{i\lambda J_{i_1,\dots,i_q}}]$. For any fixed λ, the following equations exist and all functions are differentiable:*

$$\lim_{n\to\infty} \frac{g''_{q,n}(\lambda)}{n} = 0, \tag{4.3.1a}$$

$$\lim_{n\to\infty} g_{q,n}(\lambda) = g_q(\lambda), \tag{4.3.1b}$$

$$\lim_{n\to\infty} g'_{q,n}(\lambda) = g'_q(\lambda). \tag{4.3.1c}$$

With theorem 4.3.1, the authors managed to derive an expression for any CSP at constant depth.

Theorem 4.3.2 (Theorem 1 of [66])**.** *Let H_J be a random spin-glass type CSP drawn from a graph distribution $\mathbb{G}$ that satisfy theorem 4.3.1. Then, for any constant p and parameters $(\boldsymbol{\gamma}, \boldsymbol{\beta}) \in \mathbb{R}^{2p}$, we have*

$$\lim_{n\to\infty} \mathbb{E}_{J\sim\mathbb{G}(n)} \left[\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | H_J/n | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle \right] = V_p(\mathbb{G}, \boldsymbol{\gamma}, \boldsymbol{\beta}), \tag{4.3.2}$$

and

$$\lim_{n\to\infty} \mathbb{E}_{J\sim\mathbb{G}(n)} \left[\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | (H_J/n)^2 | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle \right] = [V_p(\mathbb{G}, \boldsymbol{\gamma}, \boldsymbol{\beta})]^2. \tag{4.3.3}$$

Their formula for $V_p(\mathbb{G}, \boldsymbol{\gamma}, \boldsymbol{\beta})$ requires a generalised multinomial theorem with a rather technical proof. Previously, it was shown that the ensemble average performance of the QAOA for the Sherrington–Kirkpatrick model is equal to the one on a D regular hypertree at constant depth in the limit $D \to \infty$ [80]. Their novel contribution was to show that at constant depth, the performance of the QAOA on sparse or dense graphs are equivalent which we state below:

Theorem 4.3.3 (Theorem 2 of [66])**.** *Let $\mathbb{G}_{d,q}$ be a generic ensemble of CSP with only q-body couplings and (average) degree d that satisfy theorem 4.3.1. Suppose that*

$$\lim_{d\to\infty} \lim_{n\to\infty} (g^d_{q,n}(\lambda), {g^d_{q,n}}'(\lambda)) = \left(-\frac{\lambda^2}{2}, -\lambda \right), \quad \forall \lambda \in \mathbb{R}. \tag{4.3.4}$$

Then the asymptotic performance of constant p QAOA on a sparse $\mathbb{G}_{d,q}$ is the same as that on $\mathbb{K}^q$ (the mean-field q-spin model) in the large degree limit:

$$V_p(\mathbb{K}^q, \boldsymbol{\gamma}, \boldsymbol{\beta}) = \lim_{d\to\infty} V_p(\mathbb{G}_{d,q}, \boldsymbol{\gamma}, \boldsymbol{\beta}). \tag{4.3.5}$$

Because of the equivalence to sparse $\mathbb{G}_{d,q}$ graph, it turns out that this expression is related to the algebraic expression we got by analysing the QAOA's performance on large-girth regular graphs. In fact, it is equivalent to the performance for the corresponding q mean-field spin glass models.

Theorem 4.3.4 (Theorem 3 of [66])**.** *For any constant p and parameters $(\boldsymbol{\gamma}, \boldsymbol{\beta})$, we have*

$$V_p(\mathbb{K}^q, \boldsymbol{\gamma}, \boldsymbol{\beta}) = \sqrt{2}\nu_p^q(\sqrt{q}\boldsymbol{\gamma}, \boldsymbol{\beta}). \tag{4.3.6}$$

Remark 4.3.5. Note that the factor of $\sqrt{2}$ is due to a convention difference as mentioned in section 2.3.2. Thus, for the case of $q = 2$, if we standardised the convection to follow that of this thesis, we have in fact

$$V_p(\mathbb{K}^2, \boldsymbol{\gamma}, \boldsymbol{\beta}) = \nu_p^2(\boldsymbol{\gamma}, \boldsymbol{\beta}). \tag{4.3.7}$$

while for the general case, it will be of the form

$$V_p(\mathbb{K}^q, \boldsymbol{\gamma}, \boldsymbol{\beta}) = \nu_p^q(\sqrt{q/2}\boldsymbol{\gamma}, \boldsymbol{\beta}). \tag{4.3.8}$$

Chapter 5

Logarithmic depth failure

Given that the QAOA is a local algorithm, it is clear that if the runtime of the QAOA is shallow, (i.e. the circuit depth of the algorithm is shallow), then if the underlying graph is sparse, the QAOA does not 'see' the whole graph and thus present a barrier for its performance. Indeed a hard upper-bound was proven in [65], [82] where in one case, the limitation was due to the presence of the OGP and in the other worst case scenario, the QAOA only observes a tree as in theorem 4.2.1. In section 5.1, I first give a recap on how the OGP acts as a limitation for the QAOA at constant depth using the algorithm described in section 4.2. In section 5.2 we give the main result of the publication [1], that the algorithm is actually valid up to logarithmic depth and provide a sharp equality that shows that the QAOA at logarithmic depth gives an output strictly less than the OGP threshold.

5.1 Limitations of QAOA

In order to prove that the QAOA is limited in performance at shallow depth, we first recall the result of theorem 4.3.2 that states that for the constant depth QAOA, the output of the QAOA is given by

$$\lim_{n\to\infty} \mathbb{E}_{J\sim\mathbb{G}(n)} \left[\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | H_J/n | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle \right] = V_p(\mathbb{G}, \boldsymbol{\gamma}, \boldsymbol{\beta}), \tag{5.1.1}$$

$$\lim_{n\to\infty} \mathbb{E}_{J\sim\mathbb{G}(n)} \left[\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | (H_J/n)^2 | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle \right] = [V_p(\mathbb{G}, \boldsymbol{\gamma}, \boldsymbol{\beta})]^2. \tag{5.1.2}$$

where $V_p(\mathbb{G}, \boldsymbol{\gamma}, \boldsymbol{\beta})$ has an explicit formula defined in [66].

As a corollary, the output of the QAOA concentrates about the expectation value at large n.

Corollary 5.1.1 (corollary of [66]). *At constant depth, the output of the QAOA for spin glass type problems is concentrated over both instances and measurements.*

Proof. To show concentration properties, we first note that the variance over instances and over measurements are given respectively by

$$\text{Var(instances)} : \mathbb{E}_J \left[\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | H_J/n | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle^2 \right] - \mathbb{E}_J^2 \left[\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | H_J/n | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle \right], \tag{5.1.3}$$

$$\text{Var(measurements)} : \mathbb{E}_J \left[\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | (H_J/n)^2 | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle - \langle \boldsymbol{\gamma}, \boldsymbol{\beta} | H_J/n | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle^2 \right]. \tag{5.1.4}$$

Adding them up and taking the large n limit, we have

$$\lim_{n\to\infty} \left(\mathbb{E}_J \left[\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | (H_J/n)^2 | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle \right] - \mathbb{E}_J^2 \left[\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | H_J/n | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle \right] \right) = 0. \tag{5.1.5}$$

Since the variance is strictly non-negative, then theorem 4.3.2 implies that in the limit $n \to \infty$, both variances vanish. Thus, by Chebyshev's inequality, the QAOA is concentrated over both instances and measurements. □

We cite the result of [33] that states at logarithmic depth, the probability that the output of the QAOA has an energy value greater than the OGP value is exponentially small.

Theorem 5.1.2 (Modified version of [33] Corollary 4.4)**.** *Let the underlying graph for the Max-q-XORSAT be an Erdös–Rényi graph. For every even $q \geq 4$ and $E_{OGP} < E_{OPT}$, the following property holds. There exists a constant c such that at $p \leq c \log n$, for arbitrary parameters $(\boldsymbol{\gamma}, \boldsymbol{\beta})$, the probability that the QAOA's output has an energy greater than E_{OGP} is exponentially small:*

$$\Pr[\, \langle \boldsymbol{\gamma}, \boldsymbol{\beta} | H_{XOR}^q | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle \geq E_{OGP}] \leq \exp\{-\mathcal{O}(n^a)\}, \tag{5.1.6}$$

for some constant a.

In the large n limit, with probability converging to unity, the performance of the QAOA at $p = c \log n$ satisfies $\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | H_{XOR}^q | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle \leq E_{OGP}$.

Remark 5.1.3. Note that a recent paper by Chen et al. [11] further generalize this result that argues that the is no quantum advantage offered by logarithmic depth QAOA for sparse optimization problems.

It is thus easy to see that for any constant p, the QAOA is unable to surpass the OGP barrier even if it sees the whole graph using theorem 5.1.2 and theorem 4.3.3 since we have

$$V_p(\mathbb{K}^q, \boldsymbol{\gamma}, \boldsymbol{\beta}) = \lim_{d\to\infty} V_p(\mathbb{G}_{ER}^q, \boldsymbol{\gamma}, \boldsymbol{\beta}) = \sqrt{2} \nu_p^q(\sqrt{q}\boldsymbol{\gamma}, \beta) \leq E_{OGP}. \tag{5.1.7}$$

The novelty of eq. (5.1.7) is that at any constant p, the mean-field model is non-local and the QAOA sees the entire graph. Thus, providing the first proof of

limitation of the QAOA on a dense graph. An improvement upon this result by [67] shows limitation on dense graphs at super constant depth $p \sim \mathcal{O}(\log \log n)$.

5.2 Extension of limitation

While previous results of the limitation of QAOA were shown to hold on Erdös–Rényi hypergraph [33], their results do not apply directly to a regular hypergraph. At best, Basso et al. showed that the QAOA is limited in performance at constant p for a regular hypergraph but not when p is of order $\mathcal{O}(\epsilon \log n)$ [66]. A similar result showed by Fahri et al. argued that the output of the QAOA on MaxCut, and easily extended to Max-q-XORSAT, is at most $\frac{1}{2} + \mathcal{O}(1/\sqrt{d})$ at some logarithmic depth but does not give an exact equation as to what the upper bound is [82].

The main result that was proven in [1] is given in the following theorem:

Theorem 5.2.1. *Given a local algorithm $\mathcal{A}$ that is limited in performance up to depth $p = \epsilon \log n$ on an Erdös–Rényi hypergraph with sufficiently large average degree D, $\mathcal{A}$ is also limited in performance up to depth p on a random D-regular hypergraph for sufficiently large D.*

We delay the proof to section 5.2.1 and note the implications of the theorem first. Theorem 5.2.1 extends the validity of the optimal parameters found in [80] to logarithmic depth as well as the QAOA's limitation in performance. We do this by showing that the following limits are equal

$$\lim_{n\to\infty} \lim_{p\to\epsilon\log n} \frac{1}{|E|} \langle \boldsymbol{\gamma}, \boldsymbol{\beta} | H^q_{XOR} | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle = \lim_{p\to\infty} \lim_{n\to\infty} \frac{1}{|E|} \langle \boldsymbol{\gamma}, \boldsymbol{\beta} | H^q_{XOR} | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle . \tag{5.2.1}$$

This equality allows us to improve the upper bound into a strict equality by some value that is less than the OGP value. Namely

$$\lim_{n\to\infty} \lim_{p\to\epsilon\log n} \frac{1}{|E|} \langle \boldsymbol{\gamma}, \boldsymbol{\beta} | H^q_{XOR} | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle = \frac{1}{2} + \nu^{[q]}_\infty(D, \boldsymbol{\gamma}, \boldsymbol{\beta}) \sqrt{\frac{q}{2D}}, \tag{5.2.2}$$

with $\nu^{[q]}_\infty(D, \boldsymbol{\gamma}, \boldsymbol{\beta}) < \nu^{[q]}_\infty(\boldsymbol{\gamma}, \boldsymbol{\beta}) \leq E_{OGP} < E_{OPT} = \Pi_q$.

While the proof does not apply to dense graphs at logarithmic depth, there is a simple reason why this iteration fails to find the Parisi value Π_q for the mean-field q spin glass. The main idea is that the QAOA attempts to optimise a different problem structure that is not exhibited by the mean-field q spin glass. Analysing the QAOA using asymptotic analysis at constant depth, the underlying hypergraph that the QAOA explores is a hypertree. However, it can be easily shown that Max-q-XORSAT on a tree-like structure does not exhibit the OGP. Since the problem structure that the QAOA solves via asymptotic analysis is forbidden by the OGP,

it is clear that these two problems are in-equivalent and thus, the QAOA is unable to find the Parisi value. We now prove that Max-q-XORSAT does not exhibit the OGP on a bipartite graph.

Theorem 5.2.2. *The OGP does not exists for Max-q-XORSAT when the underlying graph is a bipartite graph.*

Proof. For simplicity, we prove it on Max-2-XORSAT where we do not need to take into account the non-degeneracy of the optimal solution since for a general $q \geq 3$, the vertices in the leaf have a permutation symmetry which is not the case for a graph. Thus, if Max-2-XORSAT does not have the OGP, then so does the general Max-q-XORSAT since the permutation symmetry makes the presence of the OGP less likely.

First, we note that because it is a bipartite graph, the problem is exactly solvable up to $\mathbb{Z}_2$ symmetry. We denote the two possible solutions as $\boldsymbol{\sigma}^+$ and $\boldsymbol{\sigma}^-$ with their overlap $|R_{\boldsymbol{\sigma}^+,\boldsymbol{\sigma}^-}| = 1$. We now convert to the Hamming distance representation and note that due to $\mathbb{Z}_2$ symmetry, we only need to consider Hamming distances in the interval $[0, n/2]$. Thus, we consider the set of Hamming distances away from the $\boldsymbol{\sigma}^+$ solution. If the OGP is present, then there exists some $\epsilon > 0$ such that taking the set of solutions that are ϵ-optimal results in a gap in their Hamming distance.

Consider the first sub-optimal solution in which a single constraint is not satisfied. This can be done by taking any arbitrary edge e_i and setting all vertices going away from the root to be the $\boldsymbol{\sigma}^-$ solution while all other vertices take on the $\boldsymbol{\sigma}^+$ solution. It's clear that the set of Hamming distance S has multiple gaps. For instance, in a regular $D+1$ tree, we have

$$S = \{1, 2, D, D+1, 2D+2, \ldots, D^2+D, \ldots\}$$

Taking the next sub-optimal solution, we similarly take another edge $e_j \neq e_i$ and repeat the same step to partition the graph into $\boldsymbol{\sigma}^-$ and $\boldsymbol{\sigma}^+$ solution which does not create a single gap in the set of Hamming distance. This procedure continues until the set of Hamming distance is dense. □

5.2.1 Proof of theorem 5.2.1

Proof. The proof is as follows: first, we need to show that for some $D \in \mathbb{Z}^+$, a D-regular q-uniform hypertree can be embedded into an Erdös–Rényi hypergraph of sufficiently high average connectivity. Then, we show that a D-regular q-uniform hypergraph can be generated from said hypertree. Finally, we show that algorithm $\mathcal{A}$ must also fail to find solutions arbitrarily close to the optimal solution in a D-regular q-uniform hypergraph as doing so would result in a contradiction.

We note that a COP instance with m chosen edges can be converted into a regular instance changing only $o_D(1/\sqrt{D})$[1] edges. This reduction has been used several times before as in [53], [54] and recently in [11]. We follow the proof of [11] but simplify one of the lemmas using $G^q_{ER}(n,p)$ later to avoid the combinatoric arguments needed when considering $G^q_{ER}(n,m)$, and extend the argument from constant depth l to $l \leq \epsilon \log n$.

Here, we show that an Erdös–Rényi hypergraph can be converted into a hypertree by changing only $o(1/\sqrt{D})$ edges. Given $\mathbb{G}^q_{ER}(n,m)$ with average degree D, define $D' = \lceil D + \sqrt{D}\log D \rceil$. Let d_i be the degree of vertex v_i. Modify the graph as follows:

1. Remove edges until $d_i \leq D'$ for all vertices.

2. Add edges to all vertices until $d_i = D'$ where each vertex is chosen with probability proportional to $D' - d_i$.

In order to prove that an Erdös–Rényi hypergraph can be converted into a hypertree by changing only $o(1/\sqrt{D})$ edges, we will need the following lemmas.

Lemma 5.2.3. *In the limit $n \to \infty$, the number of edges removed from $\mathbb{G}^q_{ER}(n,m)$ is at most $n \cdot \mathcal{O}_D(1/D^{c\log D})$ for some constant $c > 0$.*

Proof. Note that the distribution of edges in an Erdös–Rényi graph follows a binomial distribution $B(m, 1/n)$. For each vertex v_i, the number of edges removed is either 0 or $d_i - D'$ so $\Delta_i := \max(D',0)$. The first moment of Δ_i is bounded by

$$\mathbb{E}[\Delta_i] = \sum_{d>D'} P(\Delta_i = k)(k - D') = \sum_{d>D'} P(\Delta_i \geq k) \leq \int_{D'}^{\infty} P(d_i \geq x)dx. \tag{5.2.3}$$

Where we used the fact that the expectation value of a random variable equals the cumulative function in the second equality.

The second moment can also be bounded as

$$\mathbb{E}[\Delta_i^2] = \sum_{d>D'} P(\Delta_i = k)(k - D')^2 \leq \int_{D'}^{\infty} P(d_i \geq x)2(x - D')dx. \tag{5.2.4}$$

Using Chernoff's bound for the binomial distribution, we have

$$P(d_i \geq x) \leq 2\exp\left\{-\frac{(x-D')^2}{3D'}\right\}. \tag{5.2.5}$$

[1] By $o_D(g(D))$, we mean that there is a function $f(n,D)$ such that $\frac{f(n,D)}{g(D)} \to 0$ as $n \to \infty$ for fixed D, then setting $D \to \infty$.

Applying Eq. (5.2.5) to the second moment bound, we have

$$\int_{D'}^{\infty} 2(x-D')\exp\left\{-\frac{(x-D')^2}{3D'}\right\}dx = 3D'\exp\left\{-\frac{(D'-D)^2}{3D}\right\} = \mathcal{O}_D\left(D\exp\left\{-c(\log D)^2\right\}\right). \tag{5.2.6}$$

for some $c > 0$. Thus, both the first and second moment are bounded by $\mathcal{O}_D(d^{-c'\log d})$ for some constant $c' > 0$.

For the total number of edges removed Δ, we note that by a union bound, $\Delta \leq \tilde{\Delta}$, where $\tilde{\Delta} := \sum_i \Delta_i$. Furthermore, we have $\mathbb{E}[\tilde{\Delta}] = n\mathbb{E}[\Delta_i]$. Unless Δ_i and Δ_j share the same edge, they are independent, so

$$\mathrm{Var}(\tilde{\Delta}) \leq n\mathbb{E}[\Delta_i^2] + 2\sum_{i,j}\mathrm{Cov}(\Delta_i, \Delta_j). \tag{5.2.7}$$

Since the degree of each vertex is not independent as they follow a multinomial distribution (i.e. increasing the degree of one vertex results in a decrease in the degree of another), the covariance term is negative. Therefore, as $n \to \infty$, Δ is at most $n\mathcal{O}_D(d^{-c\log d})$ for some $c > 0$ with high probability. □

The process of removing edges does not create cycles (i.e. destroy tree-like property). However, we need to ensure that the graph was initially tree-like and remains tree-like after adding edges. Rather than working with $G = \mathbb{G}^q_{ER}(n, m)$, we will work with $G = \mathbb{G}^q_{ER}(n, p)$ for the next lemma and use the fact that a graph containing a k-cycle is a monotone increasing property and that for any monotone increasing graph property $\mathcal{P}$, $P(\mathcal{P} \in \mathbb{G}^q_{ER}(n, m)) \leq C \cdot P(\mathcal{P} \in \mathbb{G}^q_{ER}(n, p))$ for some constant C [83].

Lemma 5.2.4. *Fix any constant D and $l \leq \epsilon \log n$. With high probability as $n \to \infty$, 1-$o(1)$ fraction of the l-local neighbourhood are treelike.*

Proof. Let $p = \frac{c+\log n + D\log\log n}{\binom{n}{q}} \sim \mathcal{O}(\log n/n^{q-1})$ for some constant $c > 1$, and $k \in \mathbb{N}$. Let X be the number of k-cycles. Leaving k in the big-$\mathcal{O}$ notation to account for k as a function of n later, we have

$$\mathbb{E}X = \binom{n}{k}p^k \sim \mathcal{O}\left((q!)^k\frac{n^{2k}\log^k n}{n^{kq}k!}\right). \tag{5.2.8}$$

By Markov's inequality, we thus have

$$P(X > 0) \leq \mathcal{O}\left(\frac{(q!)^k\log^k n}{n^{k(q-2)}k!}\right), \tag{5.2.9}$$

which vanishes in the limit $n \to \infty$ implying that the number of constant k-cycles is $o(1)$. The same argument implies that cycles of size $\log n$ and $(\log n)^{\log n}$ also vanish.

Now we show that the l-local neighbourhood of an arbitrary vertex has at most $o(1)$ k-cycles. In the limit $n \to \infty$, the degree of each vertex follows a Poisson distribution with mean D. Let Y denote the degree of an arbitrary vertex. The probability that any vertex v has degree at most $\log n$ is given by

$$P(Y \leq \log n) = 1 - P(Y > \log n) \tag{5.2.10}$$

From the Chernoff bound for the Poisson distribution, we have that

$$P(Y > x) \leq e^{-D} \frac{(eD)^x}{x^x} \sim \mathcal{O}(c^x/x^x) = o(1), \tag{5.2.11}$$

for some constant c so all vertices have degree at most $\log n$ with high probability.

Thus, the l-local neighbourhood has at most $(\log n)^{\epsilon \log n}$ vertices. Repeating the same argument for the number of k-cycles shows that only $o(1)$ of the l-local neighbourhood will contain a cycle. □

Lemma 5.2.5. *Fix any D, there exists some $\epsilon > 0$ such that for $l \leq \epsilon \log n$, with high probability as $n \to \infty$, adding edges preserves trees in 1-$o(1)$ fraction of the l-local neighbourhood.*

Proof. Right after removing edges, every vertex has at most degree D' so given some constant D and $l \leq \epsilon \log n$, the l-local neighbourhood $B_G(v, l)$ is upper-bound by $D'^{\epsilon \log n}$ and is a hypertree. Then, we have to add on average $n(D \log D) \sim \Theta(n)$ edges but since $B_G(v, l)$ is of order $\mathcal{O}(D'^{\epsilon \log n})$, the probability that an added hyperedge contains at least two vertices in $B_G(v, l)$ is $\mathcal{O}(D'^{\epsilon \log n}/n)$.

Choose $\epsilon < 1/\log D'$. As a result, adding clauses results in $o(1)$ fraction of l-local neighbourhood forming a cycle. □

Now, we show that a D-regular, q-uniform hypergraph is locally also a hypertree.

Lemma 5.2.6. *Fix any $D > 1$ and $p \leq \epsilon \log n$ for some $\epsilon > 0$, with high probability as $n \to \infty$, $1 - o_D(1)$ fraction of vertices in the p-local neighbourhood are treelike.*

Proof. As we are interested in the large n limit, we first show that for fixed p, the dominant term in the probability that a cycle is formed in the large n limit is given by $\frac{(q-1)^p D^p}{n-1-\cdots-D^{p-1}(q-1)^{p-1}}$. Consider $p = 1$ and choose any hyperedge. Then, the first $(q-1)$ vertices form no cycle with probability 1. The next hyperedge added will form a cycle with probability $\frac{q-1}{n-1}$. This process repeats until we reach the last hyperedge for the root vertex (i.e. the D hyperedge) where the probability of forming a cycle is given by $\frac{(D-1)(q-1)}{n-1}$ so the dominant term is of the form $\frac{D(q-1)}{n-1}$. In other words,

the term that contributes the highest probability of forming a cycle at depth p is when we are filling up the last hyperedge. For $p = 2$ and higher, choosing the first hyperedge already has a non-trivial probability of forming a cycle as we might add a vertex at the $p-1$ level. Focusing on $p = 2$, this means that adding the first $(q-1)$ vertices has a probability of $\frac{D(q-1)-1}{n-1}$ to form a vertex. If we are in the middle of filling up the second layer (i.e. some of the $p = 1$ vertices already have degree D), then the adding of the next hyperedge and vertex would form a cycle with a $p = 1$ vertex with $\frac{D(q-1)-c}{n-1-c}$ for some constant c while the probability that it forms a cycle with a $p = 2$ vertex is given by $\frac{c*D(q-1)}{n-1-D(q-1)}$. For the very last hyperedge added in $p = 2$, the probability of forming a cycle is given by $\frac{(q-1)^2(D^2-1)}{n-1-D(q-1)}$ which is the dominant term. This process can be iterated to show that the dominant term is of the form $\mathcal{O}(\frac{q^p D^p}{n-\cdots-D^{p-1}(q-1)^{p-1}})$ at depth p.

For any fixed D and $p \leq \epsilon \log n$ for some constant $\epsilon > 0$, the probability that at p distance away from any vertex v_i remains tree-like is given by

$$\max\left(0, 1 - \frac{(q-1)D}{n-1} - \cdots - \frac{(q-1)^p D^p}{n-1-\cdots-D^{p-1}(q-1)^{p-1}}\right), \tag{5.2.12}$$

since

$$\mathcal{O}\left(\frac{c^{\epsilon \log n}}{n}\right) = \mathcal{O}\left(\frac{n^{\epsilon \log c}}{n}\right), \tag{5.2.13}$$

choose $\epsilon < \frac{1}{\log c} = \frac{1}{\log(Dq)}$ so that $\lim_{n\to\infty}(n^{\epsilon \log c - 1})$ goes to 0 . Then the probability that at $\epsilon \log n$ distance away from any vertex is tree like converges to unity for $n \to \infty$,

$$\lim_{n\to\infty} 1 - \cdots - \frac{(q-1)^p D^p}{n-1-\cdots-D^{p-1}(q-1)^{p-1}} = 1. \tag{5.2.14}$$

□

Now we can show that the OGP is also an obstruction in random regular hypergraphs via contradiction. Assume that an algorithm $\mathcal{A}$ at logarithmic depth is able to find solutions arbitrarily close to the optimal solution for the Max-q-XORSAT on a regular hypergraph. Then this would imply that $\mathcal{A}$ is also able to find such solutions when performed on an Erdös–Rényi hypergraph since both graphs are p-locally the same. However, this contradicts Theorem 5.1.2 and thus, the OGP must also restrict the performance of logarithmic depth local algorithms when applied to a regular hypergraph. □

We provide a simpler proof in Appendix A that relies on a conjecture about the monotonicity of the OGP.

Chapter 6

Heuristic methods for optimization

While the methods in chapter 4 were done by analysing the QAOA at constant depth, it turns out that the iteration of theorem 4.2.1 is valid up to logarithmic depth [1] as argued in section 5.2. It is thus a reasonable question as to whether one can analyse or find optimal parameters for the QAOA beyond logarithmic depth. Unfortunately, there are no known analytic techniques to find the parameters $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ beyond logarithmic depth for the QAOA. Likewise, there are no known hard bounds for the QAOA beyond logarithmic depth.

With respect to the difficulty of finding optimal parameters, the difficulty is often attributed to the phenomenon known as barren plateaus [84], [85]. Loosely speaking, it states that the landscape of the variational parameter space is flat as the probability of a non-zero gradient is exponentially small [86]. As for limitations, one often uses locality properties of the QAOA so that, at shallow depth, the QAOA has concentration of measure [65], [66] which might not hold at larger depth [87].

Rather than analytic results, one often uses heuristics methods to find optimal parameters. We give a brief review in section 6.1 of the methods currently used to find good initial starting points to find optimal parameters. In section 6.2, we show that the heuristic methods to find parameters that optimises the expectation value fail to find the optimal solution when the problem exhibits the OGP.

The results of section 6.2 is based on a joint work with my collaborator Thorge Müller (TM) [3]. The project was initialized by me and I crafted the code to generate all the Hamiltonian instances verifying whether they exhibit the OGP. The code used for simulating the QAOA was crafted by TM but with inputs by myself on what to change on a high level overview to suit the purpose of the project. Multiple discussions occurred between the two of us regarding the analysis of the numerical results so there is no clear demarcation on individual contribution. However, I contributed the insight of there being a difference between optimising the QAOA with respect to the approximation ratio vs. time-to-solution when the problem exhibits the OGP as well as the possibility that the phase transition of $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ for the ap-

proximation ratio occurs when variational parameters for the approximation ratio and time-to-solution are the same at high p depth. The paper was largely written by me with corrections and additional write up on the numerical analysis written by TM.

6.1 Optimising expectation values

As per the standard goal of the QAOA, we want to maximize the expectation value

$$\max_{\boldsymbol{\gamma},\boldsymbol{\beta}} \langle \boldsymbol{\gamma}, \boldsymbol{\beta} | H | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle \,.$$

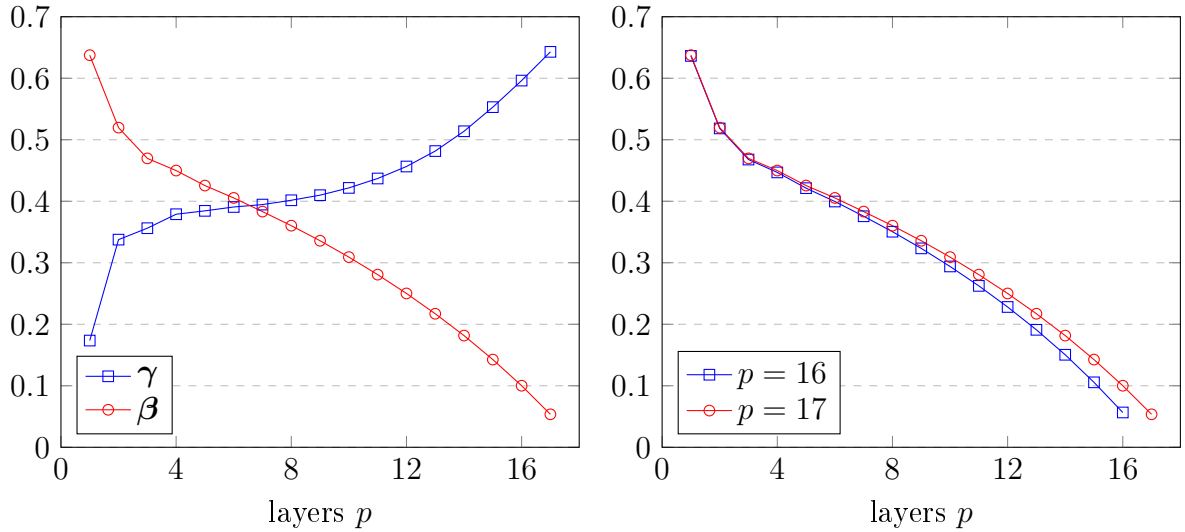
In other words, we wish to find the variational parameters $\boldsymbol{\gamma}, \boldsymbol{\beta}$ that would maximize the average output of the QAOA.

For a shallow depth QAOA, it turns out that optimising $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ with respect to the expectation value results in what appears as an adiabatic curve. Recalling that the QAOA can be thought of as a Trotterization of the quantum adiabatic algorithm [71], [88], we know that we will recover the adiabatic parameters $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ in the limit $p \to \infty$. Interestingly, even at finite depth p, the values of $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ are reminiscent of the adiabatic algorithm. For instance, the optimal parameters of $\boldsymbol{\gamma}$ (resp. $\boldsymbol{\beta}$) found in [80] for the QAOA at constant depth for the Max-q-XORSAT and mean-field q-spin glass is monotonically increasing (resp. decreasing) akin to the adiabatic evolution. A plot of the parameters can be found in fig. 6.1a.

Furthermore, it has been noticed that given the optimal parameters at a given p, the optimal parameters at the next layer $(\boldsymbol{\gamma}_{p+1}, \boldsymbol{\beta}_{p+1})$ closely resemble that of $(\boldsymbol{\gamma}_p, \boldsymbol{\beta}_p)$. An example of this can be seen in fig. 6.1b. Based on this observation, Zhou et al. proposed two heuristic methods of optimisation [72] using either linear interpolation or a change of basis. We review both of these methods as future heuristic methods such as [73] similarly makes uses of an interpolation scheme and/or re-express the parameters $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ in terms of an orthonormal basis like the Fourier basis.

6.1.1 Linear interpolation

The first strategy uses a linear interpolation to find initial starting points for the optimal parameters. The procedure is as follows: Starting at $p = 1$, find the global, or minimally local, optimal parameters. Now suppose we have found a local optimum at depth p, with parameters $(\boldsymbol{\gamma}_p^l, \boldsymbol{\beta}_p^l)$ where the superscript denotes that it is a local optimum and subscript denoting the depth. In order to find a good initial starting point for layer $p+1$, we reuse the parameters at layer p as initial points $(\boldsymbol{\gamma}_{p+1}^{\text{int}}, \boldsymbol{\beta}_{p+1}^{\text{int}})$



(a) Optimal parameters of $\boldsymbol{\gamma}, \boldsymbol{\beta}$ for $p = 17$ (b) optimal $\boldsymbol{\beta}$ for $p = 16$ and $p = 17$

Figure 6.1: Parameters for the optimal $\boldsymbol{\gamma}, \boldsymbol{\beta}$ of the QAOA applied to Sherrington–Kirkpatrick model and Max-2-XORSAT with large girth using data obtained from [80] for fixed $p = 16, 17$. (a) Note that the parameters follow that of an adiabatic schedule with $\boldsymbol{\gamma}$ (resp. $\boldsymbol{\beta}$) being monotonically increasing (resp. decreasing). (b) Notice that the curve of $\boldsymbol{\beta}$ for $p = 16$ and $p = 17$ do not change drastically and that the initial and final values are roughly the same. This similarly provides the motivation for various interpolation strategies.

via a rescalling:

$$\left[\boldsymbol{\gamma}_{p+1}^{\text{int}}\right]_i = \frac{i-1}{p}\left[\boldsymbol{\gamma}_p^l\right]_{i-1} + \frac{p-i+1}{p}\left[\boldsymbol{\gamma}_p^l\right]_i, \tag{6.1.1}$$

with $i \in [1, p+1]$. Note that we denote $[\boldsymbol{\gamma}_p]_i = \gamma_i$ as the i-th parameter of a p layer QAOA with the boundary condition $[\boldsymbol{\gamma}_p^l]_0 = [\boldsymbol{\gamma}_p^l]_{p+1} = 0$. Similarly, the rescalling for $\boldsymbol{\beta}_{p+1}^{\text{int}}$ is the same with $\gamma \mapsto \beta$. Once we have the initial starting points $(\boldsymbol{\gamma}_{p+1}^{\text{int}}, \boldsymbol{\beta}_{p+1}^{\text{int}})$, we can use them to find a local optimum $(\boldsymbol{\gamma}_{p+1}^l, \boldsymbol{\beta}_{p+1}^l)$ using an optimizer such as gradient descent or the BFGS routine [89], [90], [91], [92].

6.1.2 Fourier decomposition

In the second heuristic method, we re-parameterized the QAOA from the $2p$ parameters $(\boldsymbol{\gamma}, \boldsymbol{\beta}) \in \mathbb{R}^{2p}$ to $2q$ parameters $(\boldsymbol{u}, \boldsymbol{v}) \in \mathbb{R}^{2q}$. For $i \in [1, p]$, each parameter γ_i, β_i is parameterized by $(\boldsymbol{u}, \boldsymbol{v})$ via the equations

$$\gamma_i = \sum_{k=1}^{q} u_k \sin\left[\left(k - \frac{1}{2}\right)\left(i - \frac{1}{2}\right)\frac{\pi}{p}\right], \tag{6.1.2a}$$

$$\beta_i = \sum_{k=1}^{q} v_k \cos\left[\left(k - \frac{1}{2}\right)\left(i - \frac{1}{2}\right)\frac{\pi}{p}\right], \tag{6.1.2b}$$

where u_i, v_i are the amplitudes of the k-frequency for $\boldsymbol{\gamma}, \boldsymbol{\beta}$.

The restriction of using sine for $\boldsymbol{\gamma}$ and cosine for $\boldsymbol{\beta}$ stems from the fact that $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ typically follow an adiabatic schedule so $\boldsymbol{\gamma}$ starts from 0 and increases in value while $\boldsymbol{\beta}$ starts at some constant and decreases to 0. Thus, performing a discrete Fourier transform on $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ should result in their decomposition into the respective Fourier basis.

Several variants of the Fourier interpolation method were proposed: Denoting the general interpolation as FOURIER$[q, R]$, it depends on the two parameters $q, R \in \mathbb{Z}^+$ [72]. The parameter q denotes the maximum frequency allowed in the Fourier decomposition $(\boldsymbol{u}, \boldsymbol{v})$. As we are working with the discrete Fourier transform, the range of q is $[1, p]$. Given the smoothness and adiabatic character of $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ at low depth, one might restrict q to be a small constant so that the number of parameters remains bounded even when p goes to infinity. The second parameter R denotes the number of additional initial starting points after adding a perturbation so that the variational parameters do not get stuck in a local optimum and potentially find a better one.

The basic version of this interpolation scheme FOURIER$[p, 0]$ works similarly to the linear interpolation method. Given the optimal parameters $(\boldsymbol{\gamma}_p^l, \boldsymbol{\beta}_p^l)$ at depth p, we can decompose them into the $\boldsymbol{u}_p^l, \boldsymbol{v}_p^l$ basis and reuse them as initial starting points at depth $p+1$. Specifically, given $(\boldsymbol{u}_p^l, \boldsymbol{v}_p^l)$, we generate the initial starting point

$$\boldsymbol{u}_{p+1}^{\text{int}} = (\boldsymbol{u}_p^l, 0), \quad \boldsymbol{v}_{p+1}^{\text{int}} = (\boldsymbol{v}_p^l, 0). \tag{6.1.3}$$

With $(\boldsymbol{u}_{p+1}^{\text{int}}, \boldsymbol{v}_{p+1}^{\text{int}})$ as the initial starting point, we can again run an optimizer to find a local optimum $(\boldsymbol{u}_{p+1}^l, \boldsymbol{v}_{p+1}^l)$.

As for the general version FOURIER$[p, R > 0]$, in addition to the local optimum found by the basic FOURIER$[p, 0]$ method, we now include $R+1$ additional starting points with R of them having some random perturbation added for a total of $R+2$ local optimum. Specifically, given the best local optimum $(\boldsymbol{u}_p^B, \boldsymbol{v}_p^B)$ among the $R+2$ local optimum found at layer p, create the initial starting points

$$\boldsymbol{u}_{p+1}^{\text{int},r} = \begin{cases} (\boldsymbol{u}_p^B, 0), & r = 0 \\ (\boldsymbol{u}_p^B + \delta \boldsymbol{u}_p^{P,r}, 0), & 1 \le r \le R \end{cases} \tag{6.1.4a}$$

$$\boldsymbol{v}_{p+1}^{\text{int},r} = \begin{cases} (\boldsymbol{v}_p^B, 0), & r = 0 \\ (\boldsymbol{v}_p^B + \delta \boldsymbol{v}_p^{P,r}, 0), & 1 \le r \le R \end{cases} \tag{6.1.4b}$$

with $\delta \in \mathbb{R}$, and $\boldsymbol{u}_p^{P,r}, \boldsymbol{v}_p^{P,r}$ a set of random numbers drawn from the normal distri-

bution

$$\left[\boldsymbol{u}_p^{P,r}\right]_k \sim N\left(0, \left[\boldsymbol{u}_p^{B}\right]_k^2\right), \tag{6.1.5a}$$

$$\left[\boldsymbol{v}_p^{P,r}\right]_k \sim N\left(0, \left[\boldsymbol{v}_p^{B}\right]_k^2\right). \tag{6.1.5b}$$

Remark 6.1.1. In the subsequent sections and chapters, though we state we use the Fourier decomposition, we occasional use the insight of [73] that rather than decomposing $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ into sines and cosines, one can decompose them into any orthonormal functions such as Chebyshev and Legendre polynomials in addition to trigonometric functions. Besides the usage of different polynomials, the only other difference of [73] as compared to the Fourier decomposition of [72] is that given depth p, the former allows for optimization at depths $p + k$ with $k \geq 1$ while the latter method only optimises for $p + 1$. We will stick to the standard method of optimising layer $p + 1$ given p to ensure we find a value close to the global optimum at each $p \in \mathbb{N}$.

6.1.3 Comparison between Linear and Fourier

As noted in [72], the Fourier method seems to perform better than the linear interpolation technique at larger depths. We replicate this result on Max-4-XORSAT problems that exhibit the OGP as seen in fig. 6.2. Here we only use the FOURIER$[p, 0]$ method but find that it already surpasses the Linear interpolation method at depth $p \geq 15$ for $n = 20$. Interestingly, the equivalence of the methods at low depth is a reflection of the universality of the QAOA's parameter as seen in theorem 5.1.1.

From now onward, we will simply use the Fourier decomposition method for finding near-optimal parameters at larger depth.

6.2 Optimising ground states

In section 3.1, we stated that besides quantifying the QAOA's performance via its approximation ratio (AR), one can similarly quantify it via the time-to-solution (TTS). Thus, rather than maximizing the expectation value, we wish to maximize the overlap with the ground state solution

$$\max_{\boldsymbol{\gamma},\boldsymbol{\beta}} \langle \boldsymbol{z}^* | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle .$$

A priori, the ground state solution is not known and the only measurable quantity is the expectation value. Thus, one hopes that the parameters that maximizes the AR are approximately the same to those that maximises the TTS. However, we show in this section that when the problem has the OGP, the variational parameters differ

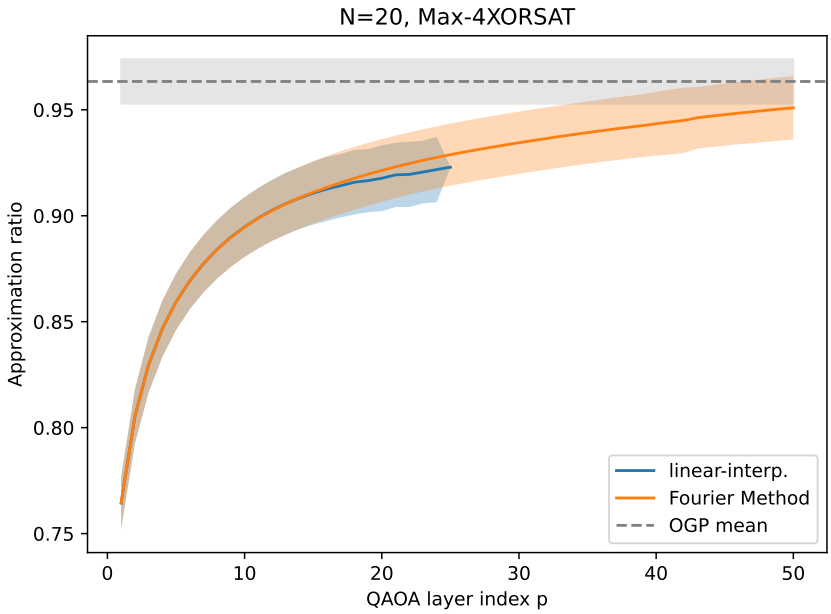


Figure 6.2: The solid line corresponds to the average QAOA output over 50 instances of Max-4-XORSAT with the OGP. The shaded interval corresponds to the standard deviation of the observed values across the various instances.

between each other at polynomial depth.

Denote the parameters that maximises the overlap with the ground state as $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ and denote the parameters that maximises the expectation value as $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$. The two different methods of quantifying the QAOA raises several questions such as the following,

1. Are there scenarios where the parameters to maximise the expectation value are different when optimising to find the exact solution?

2. If yes, in which scenario should one optimise the QAOA to optimise the expectation value and when to optimise the probability to find exact solutions?

A recent paper provides an affirmative answer to question 1 where they showed that for the Low Autocorrelation Binary Sequences (LABS) problem, these two parameters differ significantly up to depth $p = 100$ for problem sizes $n \geq 24$ [87]. When the depth p of the QAOA is sufficiently large, these two parameters should be approximately the same since the QAOA's output roughly equals to that of the optimal solution. What they showed is that at low depth, there is a difference between $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ and $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$. More generally, as we will argue in section 6.2.1 and section 6.2.2, these parameters differ when the problem exhibits the OGP.

6.2.1 LABS has the OGP

We first define the LABS problem. For a given sequence of bitstring $\boldsymbol{\sigma} \in \{-1, 1\}^n$, the autocorrelation is defined as

$$C_k(\boldsymbol{\sigma}) = \sum_{i=1}^{n-k} \sigma_i \sigma_{i+k}. \tag{6.2.1}$$

The LABS problem is to find a bit-string that minimises the energy

$$E(\boldsymbol{\sigma}) = \sum_{k=1}^{n-1} C_k^2(\boldsymbol{\sigma}), \tag{6.2.2}$$

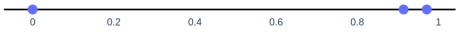
or equivalently, maximise the merit factor

$$F(\boldsymbol{\sigma}) = \frac{n^2}{2E(\boldsymbol{\sigma})}. \tag{6.2.3}$$

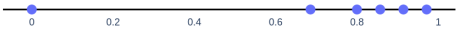
The authors of [87] mentioned implicitly that one possible reason why the parameters that optimises the AR and the TTS differ in the LABS problem is due to the lack of correlation between the Hamming distance of a bitstring from the optimal solution and the objective value of the bitstring unlike MaxCut which exhibits some correlation. We propose a much more explicit condition, that this difference occurs because LABS exhibits the OGP. Here we provide a heuristic argument why the OGP should exist in the LABS problem as well with numerical evidence to support it.

Note that the LABS problem is the same as solving the MaxCut problem on a hypergraph with hyperedges of cardinality 2 and 4 when one expands the term in eq. (6.2.2). Further note that the total number of edges with cardinality 2 is of order $O(n^2)$ while the number of hyperedges of cardinality 4 is of order $O(n^3)$. Thus, for sufficiently large n, in order to minimise the energy, it is clear that the energy contribution of the degree 4 interaction becomes dominant.

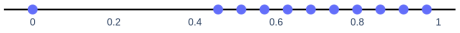
Under the widely believed assumption that the Sherrington–Kirkpatrick model does not exhibit the OGP [46], then MaxCut on a random graph does not exhibit the OGP. However, it is known that for even $q \geq 4$, MaxCut on random q-uniform Erdös–Rényi hypergraphs exhibits the OGP [48] for sufficiently large size n and sufficiently large average degree. While the hypergraph for the LABS problem is an instance of an Erdös–Rényi hypergraphs with sufficiently high degree of order $O(n^2)$, the OGP only holds with certainty in the limit $n \to \infty$ and that for any finite n, there is a non-zero probability that an instant of MaxCut on such a graph does not exhibit it. We are unable to rule out that all LABS problem are such instances. However, for sufficiently large n, the probability that it does not exhibit the OGP



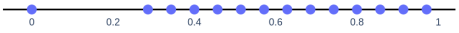
(a) Overlap of optimal solutions. Instances typically has 2 or more optimal solutions.



(b) Overlap including first sub-optimal solution.



(c) Overlap including first 3 sub-optimal solution.



(d) Overlap including first 6 sub-optimal solution.

Figure 6.3: Typical evolution of the overlap spectrum of the LABS problem. There are usually at least 2 optimal solution that results in a gap in the overlap spectrum. Taking into account additional ϵ-optimal solutions slowly fills the gap until it eventually closes. The figures above correspond to the $n = 35$ instance.

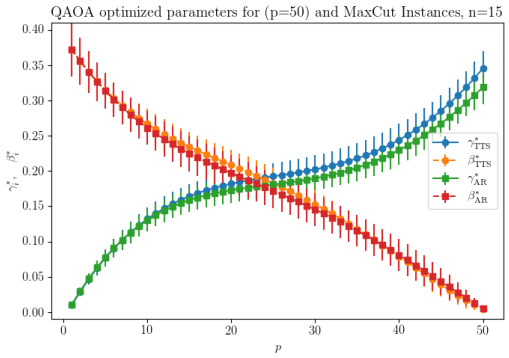
is exponentially small (i.e. $\exp\{-cn\}$ for some $c > 0$). Furthermore, even if some instances of LABS do not exhibit the OGP, as an instance of a Max-Cut problem on a hypergraph with hyperedges of cardinality 2 and 4, one might be able to create an interpolation from the LABS instance to an instance that exhibits the OGP so as to form some version of the multi-OGP to prove algorithmic limitations.

Rather than proving it rigorously, we compute numerically the overlap spectrum of LABS problem for sizes n up to $n \leq 35$ via an exhaustive search. We observe that instances of the OGP occurs when $n \geq 11$ though with the exceptions of $n \in \{14, 22, 23\}$. A typical evolution of the overlap spectrum in the LABS problem can be seen in fig. 6.3.

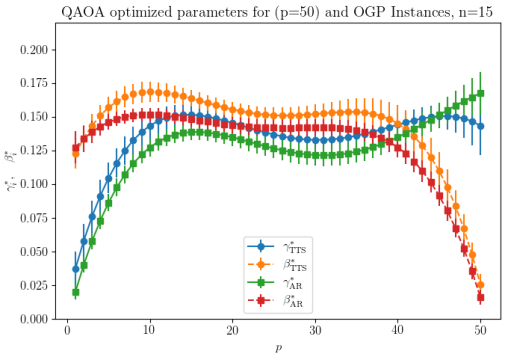
6.2.2 Parameter difference in the presence of the OGP

As mentioned above, for a general CSP, the ground state is not known so when the problem size n is large enough for an NP-hard problem such as max-cut, it takes incredibly long to find it. Thus, for bench-marking the QAOA in determining if the parameters that optimises the AR are approximately the same as those that optimise the TTS, we generate instances of Max-2-XORSAT and Max-4-XORSAT of small problem sizes where the optimal solution can be found easily, and where we can identify early on if $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ and $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ differ or converge. More precisely, we generate instances of size $n \in [15, 20]$, identify the optimal solution, and whether they exhibit the OGP. Then, we use the Fourier decomposition to find the near optimal parameters for both parameter schedules.

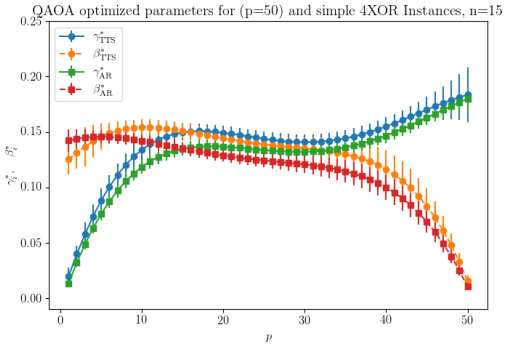
Our numerical analysis reveal that in all instances of Max-4-XORSAT that exhibited the OGP, $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ and $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ are different from each other at intermediate



(a) Typical parameter landscape for Max-2-XORSAT

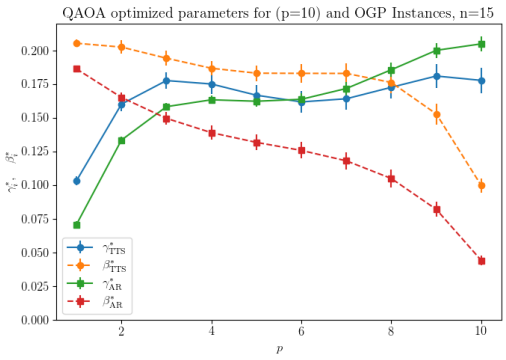


(b) Typical parameter landscape for Max-4-XORSAT with the OGP

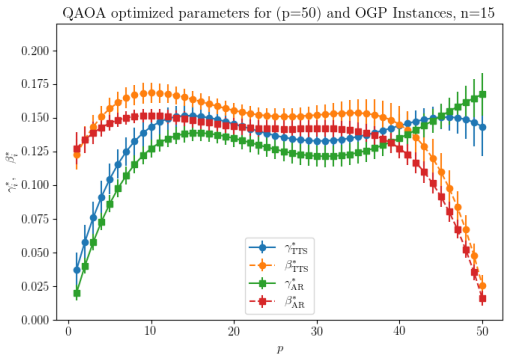


(c) Typical parameter landscape for Max-4-XORSAT without the OGP

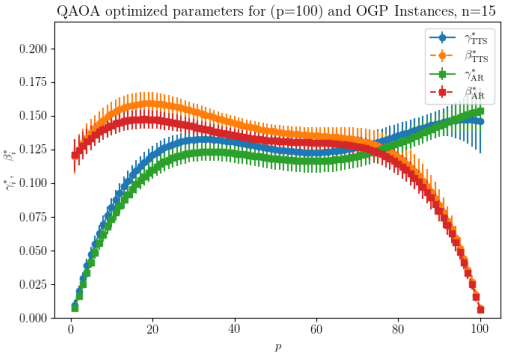
Figure 6.4: Typical difference in parameter optimization between $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ and $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ for Max-2-XORSAT and Max-4-XORSAT with and without the OGP. Here we used the landscape for $n = 15$ and $p = 50$.



(a) $p = 10$



(b) $p = 50$



(c) $p = 100$

Figure 6.5: Evolution of how $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ converges to $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ as p increases. Note that $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ is consistently larger than $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ even as p increases.

depth. On the other hand, all instances of Max-2-XORSAT generated showed no significant difference between $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ and $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$. An example can be found in fig. 6.4.

In addition, we observe consistently that $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ typically has values larger than $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$. As the circuit depth p increases, the gap between $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ and $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ decreases. A graphical representation can be found in fig. 6.5. A more fine-grained snapshot of images and analysis of the variational parameters can be found in appendix B. Interestingly, for most instances of Max-4-XORSAT that do not exhibit the OGP, we found that the optimal schedule exhibited a sort of mixture between the optimal schedule for MaxCut and Max-4-XORSAT as seen in fig. 6.4. More precisely, $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ is still typically larger than $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ at shallow p but qualitatively, they converge at a faster rate than instances with the OGP. Moreover, the overall schedule retains the 'adiabatic schedule'.

However, significant differences between $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ and $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ were observed in a small number of non-OGP instances. In these instances that do not exhibit the OGP, we find that they exhibit a weak clustering property as defined in definition 2.4.4.

Because of the difference between $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ and $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ at low depth, this

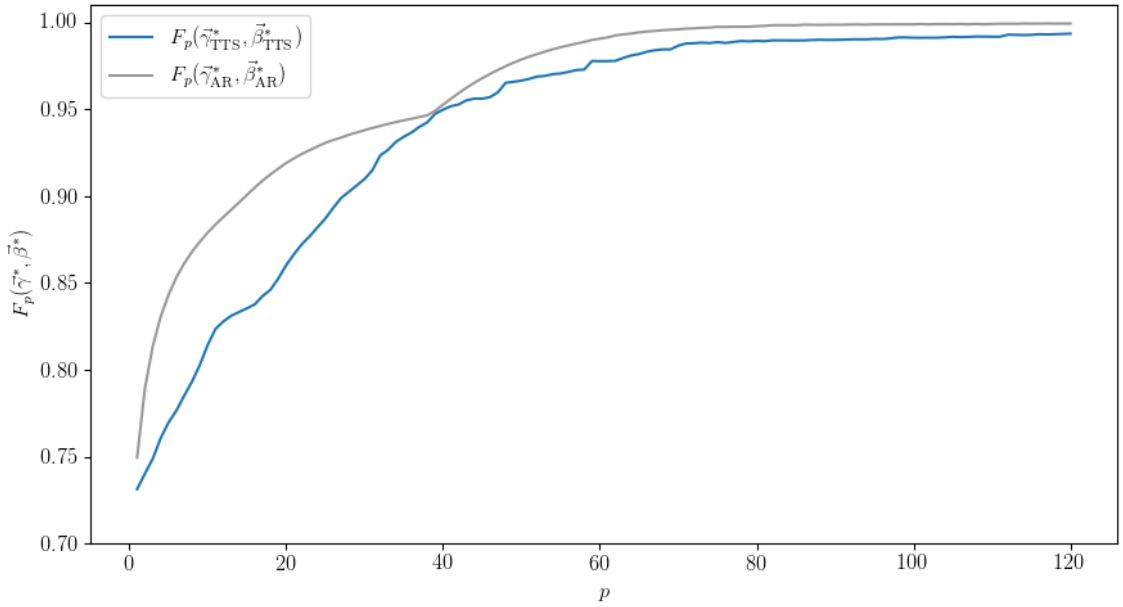


Figure 6.6: A typical example of the 'phase transition' in the scaling behaviour for the expectation value $\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | H | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle = F_p(\boldsymbol{\gamma}, \boldsymbol{\beta})$ of the QAOA. Here we have used both $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ and $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ in calculating the expectation value. Note that while the behaviour of $\langle H \rangle$ follows $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ initially, at a critical point p^*, the scaling behaviour switches to that of $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$.

results in a 'phase transition' in the scaling behaviour of the expectation value $\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | H | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle$. More precisely, because we are guaranteed to find the optimal solution as p goes to infinity, there exists some critical depth p^* when $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS} \approx (\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$. As previously found in [87], there exists a trade off between using $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ and $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ for the LABS problem. This behaviour is similarly found for Max-4-XORSAT with the OGP up to the point when they are approximately equal. Thus, it is clear that if the scaling behaviour of $\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | H | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle$ is different when using $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ as compared to $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ for shallow depth, then before the critical point p^*, the scaling behaviour follows those that optimises AR. Past the critical point p^*, the scaling behaviour changes and follows the TTS behaviour. An example can be found in fig. 6.6 and in appendix B.

These numerical findings motivate us to propose that if the problem is likely to have the OGP with high probability, then one should optimise the QAOA such that it minimises the TTS rather than AR for a quantum advantage. More precisely, the numerical findings gives us reason to believe the following conjecture:

Conjecture 6.2.1. *When a class of CSP exhibits the OGP, the QAOA should be optimised to find exact solutions to minimise the time to solution.*

To reiterate, this conjecture only makes sense if the depth of the QAOA is sub-exponential since for a large enough circuit depth, the QAOA finds the exact solution. In the current NISQ-era quantum computing where we do not have access

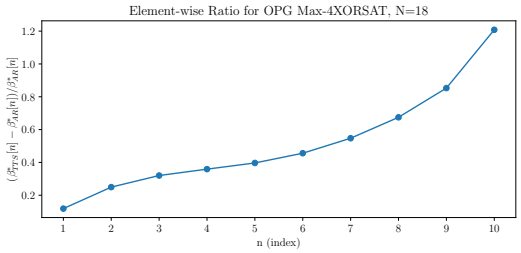
to exponential circuit depths, deciding what to optimise becomes important and our conjecture suggests that for problems exhibiting the OGP, it is much better to optimize the TTS rather than the expectation value.

6.2.3 From AR to TTS

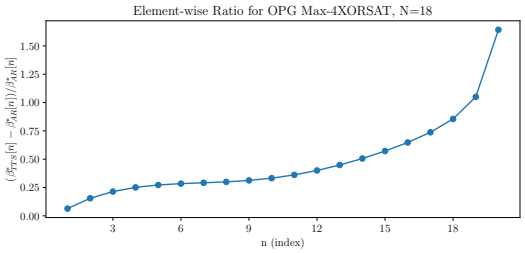
While the previous subsection has provided us with reason to believe that $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ is different from $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ for CSPs exhibiting the OGP, there is an operational problem. Namely that for sufficiently large problem instances, we do not know the ground state and hence, no method of figuring out what $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ is. The only metric that can be measured with repeated runs of a quantum computer is $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$. Thus, the problem we face now is given $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$, how can we guess what $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ is? We propose two different method

1. Extrapolation (Ext)
    (a) Generate instance for small n where its easy to find the optimal solution.
    (b) Optimize $\boldsymbol{\gamma}, \boldsymbol{\beta}$ at layer p using a classical computer to maximise likelihood of finding optimal solution.
    (c) Extrapolate $\boldsymbol{\gamma}, \boldsymbol{\beta}$ to layer $p+1$ using heuristics methods such as linear interpolation.
    (d) Reuse $\boldsymbol{\gamma}, \boldsymbol{\beta}$ for larger n where the optimal solution is unknown.
2. Rescaling (ReScal)
    (a) For fixed n, optimize $\boldsymbol{\gamma}, \boldsymbol{\beta}_{AR}$ at layer p using a classical computer to maximise the expectation value.
    (b) Systematically shift $\boldsymbol{\gamma}, \boldsymbol{\beta}_{AR}$ up by a constant factor to estimate $\boldsymbol{\gamma}, \boldsymbol{\beta}_{TTS}$.

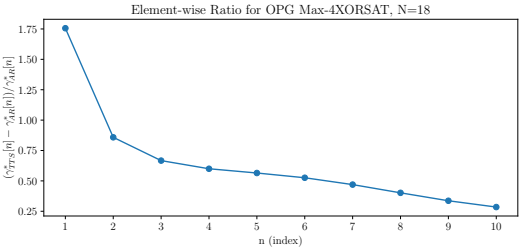
We do not implement the Ext strategy in this thesis as this would imply knowing the ground state of a large n problem in order to benchmark it and leave it as a further work to see if this strategy, or any sort of parameter transfer strategy, would work. Furthermore, there is some indirect evidence that Ext works in [87] for the LABS problem. Up to a rescalling, by optimising $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ for both AR and TTS for small n values and reusing them for larger n, they found that the performance of the QAOA using fixed parameters (i.e. the Ext strategy) performed considerably well on both metric. A key point to note is that for the LABS problem, there is only 1 unique instance for fixed n and the underlying graph of a problem with size n can be embedded into one with size $n+1$. For the Max-4-XORSAT problems considered in this thesis, the problem structure is not fixed and there is no clear



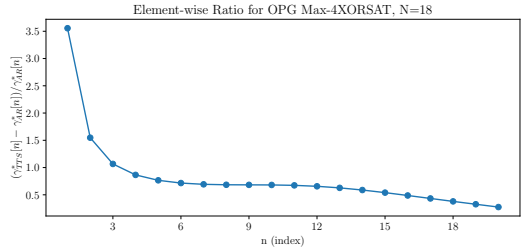
(a) Average ratio for $p = 10$ for $n = 18$ 4-XORSAT.



(b) Average ratio for $p = 20$ for $n = 18$ 4-XORSAT.



(c) Average ratio for $p = 10$ for $n = 18$ 4-XORSAT.



(d) Average ratio for $p = 20$ for $n = 18$ 4-XORSAT.

Figure 6.7: The average ratio obtained for random Max-4-XORSAT instances exhibiting the OGP using eqs. (6.2.4) and (6.2.5)

relation between two independent problem structure except for the degree of the underlying hypergraph.

The strategy ReScal proceeds as follows: Assuming that we have found $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ and $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ for (a set of) instances with circuit depth p, we would calculate the rescaling ratio r_i via the equation

$$r_{b,i} = \frac{\beta_{i,TTS} - \beta_{i,AR}}{\beta_{i,AR}}, \tag{6.2.4}$$

$$r_{g,i} = \frac{\gamma_{i,TTS} - \gamma_{i,AR}}{\gamma_{i,AR}}. \tag{6.2.5}$$

We normalised $r_{b,i}$ with respect to β_i since near the end of the schedule, β_p approaches 0 and without this normalisation, the parameter $r_{b,p}$ would blow up (similarly for the start of the schedule for γ_1). A sample of how the ratio appears for varying p can be found in fig. 6.7.

Given the rescaling ratio for each variation parameter, we now apply $\boldsymbol{r}$ to a new problem instance where $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ is known but not $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ before hand. Comparing the overlap of $|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle$ with the ground state as compared to $|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle$ obtained via ReScal, we find that this heuristic method provides us with an improvement from an average probability of approximately 0.05 to 0.15 for $p = 15$ on $n = 18$ instances as observed in fig. 6.8. Of course for sufficiently large p, this method is no longer expected to work and becomes redundant since we expect $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR} \approx (\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ and

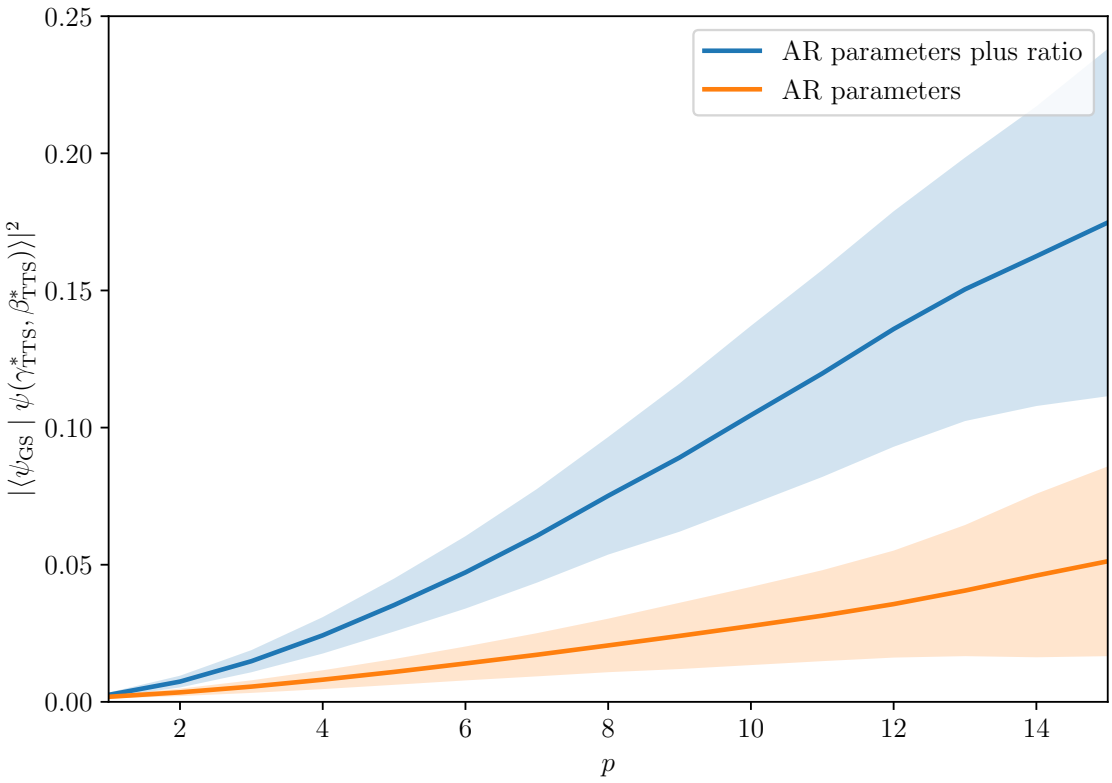


Figure 6.8: The probability of overlap with the ground state using $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ (AR parameters) as compared to using one obtained after ReScal (AR parameters plus ratio) on random $n = 18$ Max-4-XORSAT instances.

we're only interested in validating that this method works at shallow depth. Moreover, we do not claim that ReScale provides the optimal parameters to optimise the TTS. The goal of this heuristic is to merely provide an improvement of the TTS as compared to blindly using the variational parameters that optimises the AR which we have numerical evidence for it in fig. 6.8.

Chapter 7

Quantum Advantage

In this chapter, we identify various scenarios where the QAOA can offer a form of quantum advantage. In summary, the following are required for the QAOA to obtain a good solution with polynomial — possibly constant — depth.

1. The QAOA needs to see the whole graph.

2. The CSP does not exhibit the OGP.

Item 1 has already been known since the QAOA is a local algorithm, it requires at least logarithmic depth to explore the whole graph on a sparse optimization problem [65], [82]. For item 2, we provide an example of it in section 7.1 where the QAOA can find the optimal solution with $\Omega(1)$ probability with just a single layer when applied to symmetric CSPs, a non-local problem where the QAOA sees the whole graph at $p = 1$.

When the CSP has the OGP, the QAOA can still offer some advantage but it requires larger depths. We show in section 7.2 that the QAOA has an advantage over its classical counterpart, the MF-AOA, since the latter is unable to surpass the OGP threshold. In section 7.3, we numerically estimate the scaling behaviour of the QAOA required to surpass such a barrier.

The results of section 7.2 and section 7.3 is based on a joint work with my collaborator Thorge Müller (TM) [2]. The project was initialized by me and I individually related the MF-AOA to the approximate message passing (AMP) algorithm in section 7.2. I crafted the code to generate all the Hamiltonian instances verifying whether they exhibit the OGP. The code used for simulating the QAOA was crafted by TM but with inputs by myself on what to change on a high level overview to suit the purpose of the project. Multiple discussions occurred between the two of us regarding the analysis of the numerical results so there is no clear demarcation on individual contribution. Though I contributed by suggesting to analyse the scaling exponent as a power-law and as a 'logarithmic'-power law. The paper was largely

written by me with corrections and additional write up on the numerical analysis written by TM.

7.1 Symmetric CSP

The aim of this optimization problem is to find the bit string that minimises (eqiv. maximises) the Min-CSPs (eqiv. Max-CSPs) with a planted solution. The cost function $C : \{0,1\}^n \to \mathbb{Z}$ counts the number of constraints violated and is minimise by some $\boldsymbol{x}^* \in \{0,1\}^n$. By construction, for all $\boldsymbol{x} \neq \boldsymbol{x}^*$, we have $C(\boldsymbol{x}) \geq C(\boldsymbol{x}^*)$.

Definition 7.1.1 (Symmetric CSPs def 2.1 of [32])**.** Given a permutation group $G \subseteq S_n$ acting on $[n]$ bits and a bit string $\boldsymbol{x}^* \in \{0,1\}^n$, a Min-CSP with cost function $C(\boldsymbol{x}) : \{0,1\}^n \to \mathbb{Z}$ is G-symmetric with respect to $\boldsymbol{x}^*$ if $C(\pi(\boldsymbol{x} \oplus \boldsymbol{x}^*) \oplus \boldsymbol{x}^*) = C(\boldsymbol{x})$ for all $x \in \{0,1\}^*$ and $\pi \in G$. Where $\oplus$ refers to bitwise modulo-2 addition and $[\pi(\boldsymbol{x})]_i = \boldsymbol{x}_{\pi(i)}$

For S_n symmetric problems, we have $C(\boldsymbol{x}) = c(|\boldsymbol{x} \oplus \boldsymbol{x}^*|)$ for some function $c : \{0, \ldots, n\} \to \mathbb{Z}$, where $|\boldsymbol{x} \oplus \boldsymbol{x}^*|$ is the Hamming distance between $\boldsymbol{x}$ and $\boldsymbol{x}^*$. Note that C uses bit string inputs and c for Hamming distance inputs with respect to $\boldsymbol{x}^*$.

In order to prove the potential speedup of the QAOA, Montanaro and Zhou assumed that the symmetric CSP have the following property:

Assumption 7.1.2 (3.1 and 3.2 of [32])**.** *Either one of the two conditions are true*

1. $c(k) \neq c(l)$ *for all* $k \neq l$.

2. $Pr_{\boldsymbol{x},\boldsymbol{y}}[c(|\boldsymbol{x}|) = c(|\boldsymbol{y}|) \cap |\boldsymbol{x}| \cap |\boldsymbol{y}|] = o(1/\sqrt{n})$ *for bit strings* $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^n$ *drawn uniformly at random.*

Under assumption 7.1.2, the authors proved that the 1-step QAOA finds the solution with probability $O(1/\sqrt{n})$ [32]. They also proved the following result in the same paper.

Theorem 7.1.3 (Theorem 3.4 in [32])**.** *For all* $l, a \in \mathbb{Z}^+$ *with* $0 \leq a \leq l$*, if the symmetric CSP cost function has the following form in leading order in* n*:*

$$c(k) = n^l \left(\frac{k}{n}\right)^a \left(1 - \frac{k}{n}\right)^{l-a} + O(n^{l-1}), \tag{7.1.1}$$

the 1-step QAOA finds the exact solution with $\Omega(1)$.

Example 7.1.4 (Example 1 of [32])**.** Consider the SAT formula with the planted solution $\boldsymbol{x}^* = 0^n$ for simplicity:

$$\Phi = \Phi_1 \wedge \Phi_2, \qquad \text{where} \quad \Phi_1 = \neg x_1 \wedge \neg x_2 \wedge \cdots \wedge \neg x_n, \quad \Phi_2 = \bigwedge_{i \neq j \neq k} (\neg x_i \vee x_j \vee x_k). \tag{7.1.2}$$

While Φ_1 ensures that $\boldsymbol{x}^* = 0^n$ is the only satisfying solution, Φ_2 encourages a local search algorithm to increase the Hamming weight of the current attempt if the current bit-string is sufficiently far from the solution. The cost function is given by $C(\boldsymbol{x}) = \sum_i x_i + \sum_{i \neq j \neq k} x_i(1 - x_j)(1 - x_k)$ which, converted into the Hamming distance cost function, can be expressed simply as

$$c(k) = k + k(n - k)(n - k - 1). \tag{7.1.3}$$

Note that this cost function fulfills the leading order requirement for the cost function in theorem 7.1.3. Further note that the cost function has a global minimum at $k = 0$ and a local minimum at $k = n-1$. The global maximum is located at approximately $k = n/3$ for sufficiently large n and thus, if a local algorithm starts with a bit-string with Hamming distance $\geq n/3$, an algorithm based on gradient descent would end up in the local minimum.

For $S_{n_1} \times S_{n_2}$-symmetric CSP, the authors similarly proved that the 1-step QAOA succeeds with probability $\Omega(1)$ assuming the following:

Assumption 7.1.5 (Modified assumption 3.6 of [32])**.** *Consider a family of $S_{n_1} \times S_{n_2}$-symmetric CSP in the limit $n \to \infty$ with $\alpha_i = n_i/n > 0$ fixed. Assume that there exists $l \in \mathbb{Z}^+$, a finite sequence of non-negative integer tuple $(a_\mu, b_\mu, c_\mu, d_\mu)$ satisfying $a_\mu + b_\mu + c_\mu + d_\mu = l$ for all μ, and a sequence of $k_\mu \in \mathbb{R}$, such that up to leading order in n, the cost function can be written down as*

$$c(k_1, k_2) = n^l f(k_1/n_1, k_2/n_2) + O(n^{l-1}), \tag{7.1.4}$$

$$\text{where} \quad f(x_1, x_2) = \sum_\mu k_\mu x_1^{a_\mu}(1 - x_1)^{b_\mu} x_2^{c_\mu}(1 - x_2)^{d_\mu}. \tag{7.1.5}$$

Remark 7.1.6. Note that both in [32] as well as the corresponding talk at QIP, it is said that this paper provides evidence for exponential advantage of the QAOA with polynomial circuit depth. Upon careful reading, what the paper actually shows is that against un-optimised general purpose classical algorithms, there is some evidence of exponential advantage. Thus, the evidence of exponential advantage is on slim ice.

7.1.1 Absence of the OGP in symmetric CSP

We now prove the absence of the OGP in symmetric S_n problems.

S_n- symmetric CSP

We prove that S_n-symmetric CSP do not exhibit the OGP if the cost function are of the form in eq. (7.1.1). By construction, the minima of the S_n-symmetric CSP are located at $k = 0$ and $k = n$ up to leading order.

We now show that if one uses a $\{-1, 1\}^n$ notation by converting the binary variables to an Ising spin representation to calculate the overlaps, it does not exhibit the OGP if one calculates the overlaps $|R_{1,2}|$ of near optimal solutions.

Theorem 7.1.7. *Assuming that a S_n-symmetric CSP has a cost function of the form in eq. (7.1.1), than the OGP is not present.*

Proof. Without lost of generality, assume that $c(n-1) < c(1)$. We find the set of all possible overlaps of the optimal solution and those whose hamming distance is n and $n-1$ away have the normalised overlap

$$\mathcal{S} = \left\{1, 1 - \frac{2}{n}, 1 - \frac{4}{n}\right\}.$$

It is then intuitive that adding the overlap of the current set of solutions with those with Hamming distance $n-1-x$ for $x \geq 1$ monotonically increases $\mathcal{S}$. More precisely, if the cost function is symmetric, then we alternate between including solutions with Hamming distance $n-x$ and $1+x$ for $x \geq 1$ so no gap is created in the process. Without lost of generality, assume that the cost function is right skewed so that $c(1) > c(n-1-x)$ for some $x > 0$. In this case, we include solutions with Hamming distance $k = n-1-x$ when calculating the set of overlaps. This monotonically increases the set of overlaps until we have $c(1) \leq c(n-1-x)$ which is dependent on parameters a, l. At which point, we include solutions with Hamming distance 1 from the optimal solution. However, the overlap of solutions with Hamming distance 1 with the other bitstrings considered so far differs by $2/n$ or $4/n$ when comparing the overlap between them and $\boldsymbol{x}^*$. Thus, no gap in the set of overlap is produced after including the set of bit-strings with Hamming distance 1. This procedure continues until the set $\mathcal{S}$ is dense. □

Since the symmetric CSP considered does not exhibit the OGP, it is clear that previous results [33], [65], [66], [67] using concentration of measure to limit the performance of the QAOA fails to apply to the symmetric CSP setting. This does not contradict the claim that previous hardness result do not apply to the symmetric CSP due to "highly nonlocal geometry" of the problem structure [32]. Rather, the

speed up is best understood as both the absence of a complicated geometry in the solution space via the OGP, and because the underlying graph of a symmetric CSP is dense and thus a highly non-local geometry, that enables the speed up.

We are unable to prove if the OGP is present in symmetric CSPs that only makes use of theorem 7.1.2 without knowing the exact form of $c(k)$. However, suppose that the local minima exists at Hamming distance k^* and $c(k)$ is sufficiently smooth, then we can prove that if we calculate the Hamming distance between the optimal solution and the local minima, the OGP is not present at this level. For simplicity, we assume that the symmetric CSP exhibits $\mathbb{Z}_2$ symmetry so we can restrict the Hamming distance to $k \in [0, n/2]$.

Theorem 7.1.8. *Suppose that the cost function $c(k)$ is sufficiently smooth and assuming that a local minima exists at Hamming distance k^*. For any $k^* \in [0, n/2]$, and a planted bit-string $\boldsymbol{x}^*$, the OGP is not present when taking the set of Hamming distance between $\boldsymbol{x}^*$ and all bit-strings with hamming distance k^* from $\boldsymbol{x}^*$.*

Proof. Without lost of generality, set $x^* = 0^n$ as the all 0 bit-string. Then, denote $[i, j, l, \dots]$ as the bit-string that alternates between 0 and 1 with the first i bits as 1, the next j bits as 0, then l 1 bits and so on. Now, for all $i \in [1, k^*]$, consider the Hamming distance between $[k^*, k^*, n-2k^*]$ and $[k^*-i, i, n-k^*-i]$. This contains all possible even values between 0 and $2k^*$ (i.e. $H_{1,2} = 2m$ for $m \in [0, k]$). If k^* is even, then there are no more Hamming distance possible. If k^* is odd, then we include k^* in the set of Hamming distance and no other Hamming distance is possible. In either case, the set of Hamming distances does not fall into the definition of the OGP so the OGP is not present at this level.

If $c(k)$ is 'smooth' such that the next sub-optimal solution is at $k^* \pm 1$ or $k = 1$, then the set of Hamming distance would include all the odd integers between 0 and $2k^*$ as well and no gap would be produced. □

Remark 7.1.9. Note that if one can formulate a version of a symmetric CSP such that the local minima is at Hamming distance $(3/4 \pm \epsilon)n$ for some arbitrarily small $\epsilon > 0$ so that we do not have approximate $\mathbb{Z}_2$ symmetry. Then assuming $c(k)$ is sufficiently 'smooth' as mentioned above, such a symmetric CSP could exhibit the OGP since we would have a gap in the set of Hamming distances. We do not know if such a symmetric CSP can be formulated but we believe that it is unlikely. If it were possible, then in the worst case scenario with the local minimum located at $k = 3n/4$, there are $\binom{n}{3n/4}$ solutions for the local minimum.

$S_{n_1} \times S_{n_2}$-symmetric CSP

We now similarly prove that the OGP does not exists in $S_{n_1} \times S_{n_2}$-symmetric CSP.

Theorem 7.1.10. *Assuming that a $S_{n_1} \times S_{n_2}$-symmetric CSP has a cost function of the form in eq. (7.1.4), than the OGP is not present.*

Proof. We note that within each subset of bits n_i, by theorem 7.1.7, the OGP is not present. Thus, we only have to show that after concatenating the near optimal solution of each subset $(\boldsymbol{x}^1, \boldsymbol{x}^2) \in \{0,1\}^{n_1 \times n_2}$, the absence of the OGP is preserved. Denote $n = n_1 + n_2$, the optimal solution within each subset as $\boldsymbol{x}^i_{opt}$, and the first sub-optimal solution as $\boldsymbol{x}^i_{sub}$. Then, taking the overlap of the optimal and first sub-optimal solution, we have 4 distinct possibility to consider: (1) $\boldsymbol{x}^1_{opt}$ and $\boldsymbol{x}^2_{opt}$, (2) $\boldsymbol{x}^1_{opt}$ and $\boldsymbol{x}^2_{sub}$, (3) $\boldsymbol{x}^2_{opt}$ and $\boldsymbol{x}^1_{sub}$, (4) $\boldsymbol{x}^1_{sub}$ and $\boldsymbol{x}^2_{sub}$.

It is clear that the minima points for $\boldsymbol{x}^1$ are at $k_1 = 0$ and $k_1 = n_1$. Similarly for $\boldsymbol{x}^2$ with $k_2 = 0$ and $k_2 = n_2$. The rest of the argument follows that in the proof of theorem 7.1.7. If the cost function is symmetric, then the OGP is absent since we can similarly keep adding solutions into the overlap that differ by Hamming distance 1 from the current set. Without loss of generality, assume that the cost function is right skewed so the local minimum is $c(k_1, 1) > c(k_1, n_2 - 1 - x)$ where $x \geq 0$. Following the same argument as in theorem 7.1.7, we find that set of possible overlaps is

$$S = \left\{1, 1 - \frac{2}{n}, 1 - \frac{4}{n}\right\}.$$

For instance, concatenating $\boldsymbol{x}^1_{opt}$ and $\boldsymbol{x}^2_{sub}$ and calculating their overlap with the optimal solution gives us

$$\frac{n_1}{n} + \frac{n_2 - 2}{n} = 1 - \frac{2}{n}. \tag{7.1.6}$$

Thus, it clear that the set S increases monotonically following the same argument as in theorem 7.1.7.

□

7.2 Limitations of the MF-AOA

In this section, we prove that the algorithmic threshold of the MF-AOA is bounded by the OGP barrier. We do this by relating it to AMP algorithms which is similarly known to be unable to surpass the OGP barrier.

Denote the binary list $\{-1, 1\}^n = \mathcal{B}^n$ and the Hilbert cube $[-1, 1]^n = \mathcal{H}^n$.

Given an $n-$tensor array $J = (J_{i_1,\ldots,i_q}, 1 \leq i_1, \ldots, i_q \leq n) \in (\mathbb{R}^n)^{\otimes q}$ of order q and an n-vector $u \in \mathbb{R}^n$, define the inner tensor product as

$$J(u) = \sum_{1 \leq i_1,\ldots,i_q \leq n} J_{i_1,\ldots,i_q} u_{i_1} \ldots u_{i_q}. \tag{7.2.1}$$

Let $\|J\|_2$ denote the Frobenius norm

$$\|J\|_2 = \sqrt{\sum_{1\le i_1,\ldots,i_q\le n} J^2_{i_1,\ldots,i_q}}. \tag{7.2.2}$$

Now for any $u \in \mathbb{R}^n$, denote

$$y = J(\cdot, u) \in \mathbb{R}^n \tag{7.2.3}$$

with

$$y_i = \sum_{1\le i_1,\ldots,i_{q-1}\le N} J_{i,i_1,\ldots,i_{q-1}} u_{i_1} \ldots u_{i_{q-1}}. \tag{7.2.4}$$

AMP algorithms are a class of low-complexity iterative thresholding algorithms originally proposed in [93]. In spin glass theory, the AMP algorithm is closely related to the Thouless–Anderson–Palmer (TAP) free energy functional [94], [95]. Assuming the absence of the OGP, there exists an AMP algorithm that gets arbitrarily close to finding the optimal solution for the Ising mean-field k spin glass [96]. While specific AMP type algorithm exists for the k-spin glass and not for general Max-CSP problems like Max-q-XORSAT, it is suspected that there might be an AMP type algorithm for all such problems [55]. A survey of the AMP algorithm and its variants can be found in [97].

We now introduce the relevant assumption used to define a class of AMP algorithms as in [64]. For fixed $M, T \in \mathbb{Z}^+$, consider 2 functions $f_t : [-M, M]^t \to \mathbb{R}$ and $F_t : \mathbb{R} \times [-M, M]^t \to \mathbb{R}$, for $1 \le t \le T$.

Assumption 7.2.1. *$f_t(0) = 0$. In addition, the functions f_t, F_t are Lipschitz continuous on their respective domains. Specifically, $\exists K \in \mathbb{R}^+$ such that for all $1 \le t \le T$,*

$$\sup_{u,v\in[-M,M]^t} |f_t(u) - f_t(v)| \le K\|u - v\|_2 \tag{7.2.5}$$

$$\sup_{u,v\in\mathbb{R}\times[-M,M]^t} |F_t(u) - F_t(v)| \le K\|u - v\|_2 \tag{7.2.6}$$

Furthermore, fix a constant $M > 1$ and let $x_M = \max(-M, \min(x, M))$ to denote an M-truncation for $x \in \mathbb{R}$. When x is a vector, x_M is applied coordinate wise. We now define the iterations used in the AMP algorithm. Fix $U^0 \in [-M, M]^n$ and define the sequence $U^t \in \mathbb{R}^n, 1 \le t \le T$ as follows

$$U^t = \left[F_t(J(\cdot, f_t(U^0, \ldots, U^{t-1})), U^0, \ldots, U^{t-1})\right]_M \in [-M, M]^n, \tag{7.2.7}$$

where F_t, f_t and M are applied component-wise.

In words, at step t of the algorithm, the vector $f_t(U^0, \dots, U^{t-1}) \in \mathbb{R}^n$ is formed by applying f_t coordinate-wise. Then, compute the vector $J(\cdot, f_t(U^0, \dots, U^{t-1}))$ using eq. (7.2.4). This vector is concatenated with the vectors $U^0, \dots, U^{t-1}$ to form an $n \times (t+1)$ matrix before applying the function F_t coordiante wise. Finally, perform an M-truncation to each coordinate to get U^t. Once the iterations are complete, the algorithm generate a binary string in $\mathcal{B}^n$ from $\mathcal{H}^n$ via some projection $\Pi_n : \mathcal{H}^n \to \mathcal{B}^n$. A summary of the algorithm can be found in algorithm 3.

Algorithm 3 AMP Algorithm

Require: $J, U^0, M, T, (f_t, 1 \le t \le T), (F_t, 1 \le t \le T), \Pi_n$

1. **for** $t \in [0, T_f]$ **do** :
   (a) Compute U^t using eq. (7.2.7).
2. **end for**
3. Project U^T to $\mathcal{H}^n$ by applying $x \to [x]_1 = \max(-1, \min(x, 1))$ coordinate-wise with resulting vector denoted by $V \in \mathcal{H}^n$.
4. Compute $\sigma^* = \Pi_n(V)$,
5. **return** σ^*.

Using assumption 7.2.1 and the conjecture that the OGP similarly exists in the domain $\mathcal{H}^n$, one can then prove that AMP algorithms are unable to find arbitrary near optimal solution in problems that exhibit the OGP.

Theorem 7.2.2 (informal version of theorem 3.3 of [64])**.** *For every even $p \ge 4$ in the domain $\mathcal{H}^n$, assume that the OGP exists with probability $1 - \exp\{-cN\}$ for some $c > 0$ for large $n > 0$. Further assume that f_t and F_t satisfy assumption 7.2.1. Then, for a CSP with tensor structure J, there exists an $\epsilon > 0$ and $c > 0$ such that the AMP algorithm A satisfies*

$$\mathbb{P}\left(\frac{A(J)}{n} \le \frac{\max_{\boldsymbol{\sigma} \in \mathcal{B}^n} J}{n} - \epsilon\right) \ge 1 - \exp\{-cn\} \tag{7.2.8}$$

We now prove that the MF-AOA is a form of AMP algorithm as described in algorithm 3 modulo the truncation step.

Theorem 7.2.3. *The MF-AOA is a special case of the AMP algorithm as defined in algorithm 3 modulo the truncation step.*

Proof. Let $V_i(t, m(t-1)) = V_i^D(t) V_i^P(t, m_i(t-1))$. At the start of time step t, assuming we have $m_i(t-1)$, the unitary operator V_i is generated.

At the end of each time-step of the MF-AOA, one updates the effective magne-

tization of each spin using eq. (3.2.8). Denote

$$J_i(u) = \sum_{1\le i_1,\dots,i_{p-1}\le N} J_{i,i_1,\dots,i_{p-1}} u_{i_1}\dots u_{i_{p-1}}. \tag{7.2.9}$$

Since the magnetization depends on $n_i(t)$, using eq. (3.2.8), the magnetization as a function has the following form at time step t

$$m_i(s(t)) = \lambda_i s^2(t)(1-s(t)) + \sum_{i\in e\in E} J_{i,i_1.....i_{q-1}} \prod_{i_j=i_1}^{i_{q-1}} \left[\prod_{t=0}^{t} V_{i_j}(t, m(t-1)) n_{i_j}^x(0)\right]^z, \tag{7.2.10}$$

with the condition that $m_i(0) = 0$.

In order to align the MF-AOA with the AMP algorithm, we note that at the very first step $t = 1$ for the MF-AOA, because $m_i(0) = 0$, we have $V_i(1, m(0))$ as the identity operator for all i. Thus the start of the computation really begins in computing $m_i(1)$ and one can denote calculating $m_i(1)$ as the beginning of the algorithm as per algorithm 3 (i.e. define U^0 of the AMP as V^1 of MF-AOA).

Let $U^0 = \mathbb{1} = V(1, m(0))$. For $t \in [1, T]$:

$$f_t(V^1,\dots,V^t) = \sum_{i\in e\in E} J_{i,i_1\dots i_{q-1}} \prod_{i_j=i_1}^{i_{q-1}} \left[\prod_{k=1}^{t} V_{i_j}(k, m(k-1)) n_{i_j}^x(0)\right]^z, \tag{7.2.11}$$

$$m(\cdot, f_t(V^1,\dots,V^t)) = \lambda s^2(t)(1-s(t)) + f_t(V^1,\dots,V^t), \tag{7.2.12}$$

$$F_t(m, V^1,\dots,V^t) = \prod_{x=1}^{t} V(x, m(x-1)) = \prod_{x=1}^{t} V^D(x) V^P(x, m(x-1)). \tag{7.2.13}$$

In other words, f_t is a function that calculates the effective magnetization, at time t, and F_t is the matrix product of all the MF-AOA evolution operator.

We now verify that the functions f_t and F_t fulfil theorem 7.2.1. We have already shown that $f_t(0) = 0$ and move on to verify that the functions are Lipschitz continuous. For f_t, consider m and note that $\lambda_i = \mathcal{N}(0, \sigma^2)$. For sufficiently large M and sufficiently small σ, the probability that λ_i is greater than N vanishes. Denote the maximum degree of the graph problem as D. From this we see that the maximum possible magnetization is

$$|m_i| = \frac{4}{9}|\lambda_i| + |D|, \tag{7.2.14}$$

which is bounded so for any input t, so we see that f_t is indeed Liptschitz continuous.

As for F_t, we note that the matrix elements is either 1, sine, or cosine. Thus, it is clear that F_t is Lipschitz continuous. □

Corollary 7.2.4. *The MF-AOA is unable to surpass the OGP threshold*

Proof. The proof follows from theorem 7.2.2 and theorem 7.2.3. □

7.3 Scaling Behaviour of the QAOA

Since the MF-AOA — and AMP algorithms in general— is unable to surpass the OGP barrier, this provides an avenue for quantum advantage where the QAOA can outperform the MF-AOA. However, no known theoretical or numerical results exists about the depth required to surpass it though it has been suspected to be exponential depth. The closest paper that addresses this problem is the application of the QAOA to k-SAT instances at the satisfiability threshold [98], a problem known to exhibit the OGP [58]. In the study, the authors found that the QAOA exhibit an exponential scaling behaviour in the median run-time while providing a slight polynomial speedup over state of the art general purpose classical SAT solvers.

In this section, we perform the first numerical analysis as to the depth required to surpass the OGP threshold. Our methodology is as follows. First, we generate instances of Max-4-XORSAT of various problem sizes with $n \in [15, 20]$. Since the OGP occurs only in the highly unsatisfiable regime, we fixed the average degree of the graph to be sufficiently large (e.g. $d = 16$) such that the problem is unsatisfiable and instances have a non-zero probability of exhibiting the OGP. Since Max-4-XORSAT has the OGP with high probability in the limit $n \to \infty$, we only include instances that do have the OGP to identify the scaling behaviour of the QAOA. We also kept instances of Max-4-XORSAT that do not exhibit the OGP in order to test if the QAOA exhibit any significant difference in scaling behaviour or parameter optimization compared to those that do have the OGP. As a final comparison between OGP and non-OGP problem instances, we also ran the QAOA on instances of the Max-2-XORSAT problem since it is believed that it does not exhibit the OGP.

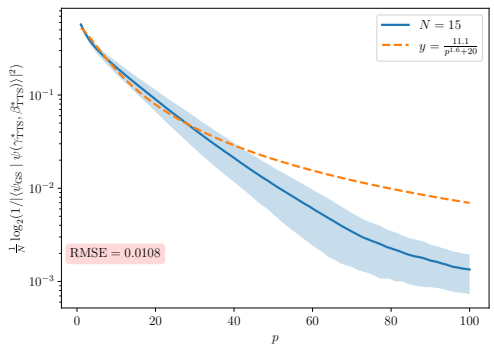
In order to estimate the time complexity of the QAOA, we perform a numerical fix using the equation

$$\frac{1}{|\langle \boldsymbol{z}^*|\boldsymbol{\gamma},\boldsymbol{\beta}\rangle|^2} = 2^{c(p)N} \tag{7.3.1}$$

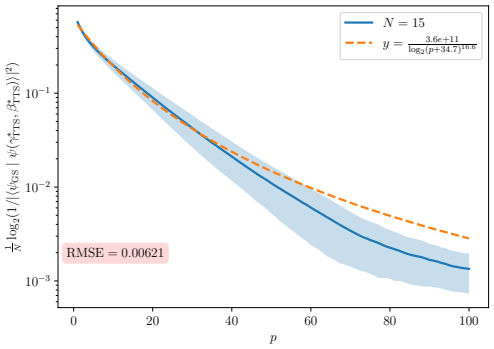
$$\frac{-\log_2 |\langle \boldsymbol{z}^*|\boldsymbol{\gamma},\boldsymbol{\beta}\rangle|^2}{N} = c(p), \tag{7.3.2}$$

where $\boldsymbol{z}^*$ is the bit-string corresponding to the optimal solution, and $c(p)$ some function of the circuit depth p.

Whether the QAOA is able to find the optimal solution with polynomial circuit depth depends on the scaling exponent $c(p)$. Here, we perform 3 different numerical



(a) The scaling coefficient for Max-4-XORSAT using a power-law fit.



(b) The scaling coefficient for Max-4-XORSAT using a logarithmic power-law fit.

Figure 7.1: Various fits of the scaling coefficient c in $1/p^{opt} = 2^{c(p)N}$ for the Max-4-XORSAT instances with the OGP.

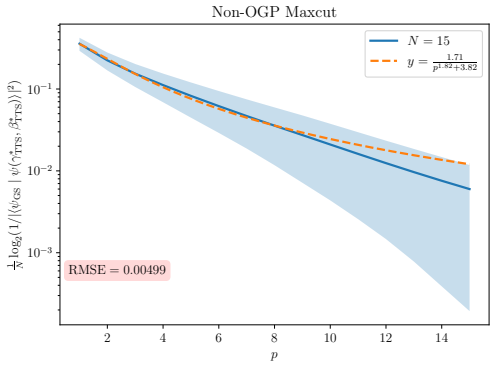
fits which corresponds to the different time-complexity often encountered in the analysis of algorithms

$$c(p) = \begin{cases} \dfrac{c_1}{p^{c_2} + c_3} & \text{Polynomial,} \\ \dfrac{c_1}{\log(p+c_3)^{c_2}} & \text{Quasi-Polynomial,} \\ \dfrac{c_1}{\log(p+c_3)} & \text{Exponential.} \end{cases} \tag{7.3.3}$$

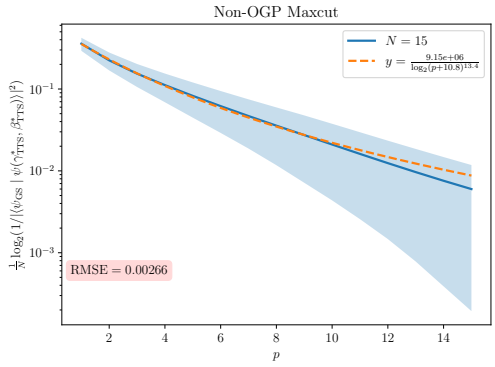
The first case corresponds to a power-law behaviour and allows the QAOA to solve a problem efficiently (i.e. polynomial depth) since when $p \sim \mathrm{Poly}(n)$, the number of measurements required is of $\mathcal{O}(1)$. Similarly, in the worst case scenario where we expect a time complexity of the order $2^{\Theta(n)}$, this correspond to the inverse logarithm of p since it is only when $p \sim 2^n$ that we expect $\mathcal{O}(1)$ measurements. The quasi-polynomial case is the scenario when we expect a super-polynomial time complexity that is sub-exponential (i.e. faster than $2^{\Theta(n)}$).

Our results suggests that on average, problems that are unlikely to exhibit the OGP such as Max-2-XORSAT are likely to require only polynomial depth, consistent with various literature [99], [100] that found a power-law like behaviour when it comes to the scaling behaviour of the QAOA. On the other hand, there is clear qualitative evidence that the QAOA does not exhibit a power law behaviour for instances of Max-4-XORSAT that do exhibit the OGP as seen in fig. 7.1. While the best fit we found was the quasi-polynomial fit, it still did not accurately capture the correct scaling. Nevertheless, the improvement of fit indicates that the QAOA likely require a super-polynomial — but sub-exponential — circuit depth p to efficiently solve problems that exhibit the OGP. Further details can be found in fig. 7.1.

While both the power-law and quasi-polynomial fit work on Max-2-XORSAT as



(a) The scaling coefficient for Max-2-XORSAT using a power-law fit.



(b) The scaling coefficient for Max-2-XORSAT using a logarithmic power-law fit.

Figure 7.2: Various fits of the scaling coefficient c in $1/p^{opt} = 2^{c(p)N}$ for the Max-2-XORSAT.

observed in fig. 7.2, we believe that this is simply a result of over-fitting. Given that i) a similar power-law behaviour was found in other examples of the QAOA on non-OGP problems [99], [100], ii) evidence that one requires $\mathcal{O}(N)$ gates to efficiently solve the Sherrington–Kirkpatrick model [101], iii) that our numerical simulation found that the QAOA can get a significant overlap with the ground state when $p \approx n$, these all suggest that the QAOA can solve non-OGP problems with polynomial circuit depth.

Remark 7.3.1. We note that [98] found a power law behaviour despite the existence of the OGP in random k-SAT. One reason for this is that they randomly generated up to 10000 instances for problem size $n \in \{12, \ldots, 20\}$ with the only requirement that the problem is satisfiable and did not check for the OGP condition. This could skew the result since it is only guaranteed that a random instance will exhibit the OGP with high probability when n goes to infinity. In our generation of Max-4-XORSAT, we found that for a random sample of 1000, the number of instances that exhibit the OGP is likely to be less than half. More precisely, we found that the probability of generating an instance exhibiting the OGP being somewhere between 30-50% per run.

Chapter 8

Application: Binary Paint Shop Problem

Thus far we have been focused on identifying when does the QAOA exhibit a quantum advantage over classical algorithms and when it does not. However, what if we have been going about this the wrong way? It is entirely possible that we would never achieve fault tolerant quantum computers in the future. If this is true, then the study of quantum algorithms, while they might provide us with a theoretical algorithm to solve problems efficiently, would be in practice work done in vain.

This chapter aims to demonstrate that the study of such quantum algorithms is far from futile, even if implementing them at the large circuit depths required for quantum advantage remains out of reach. Instead, the QAOA can be used as a tool to identify problems where one could, in principle, develop better classical algorithms that surpass the performance of all existing classical methods. Such a use of a quantum/classical algorithm is not new. The authors of the Mean-Field Approximate Optimization Algorithm (MF-AOA) developed the algorithm in part to identify problems where the QAOA would exhibit true quantum advantage over classical algorithms [77]. In a similar spirit, one research direction from Phasecraft is to use low-depth QAOA as a pre-processeing step to speed up classical algorithms. [102], [103].

In section 8.1, we will review the Binary Paint Shop Problem (BPSP) and the performance of the best known classical and quantum algorithms. In section 8.2, we will provide an algorithm that converts the BPSP into an Ising problem to implement on the QAOA. Finally in section 8.3, we show how the QAOA is able to provide better a better upper-bound for the BPSP and why it also implies that there must exist an efficient classical algorithm that outperform all known current methods.

This chapter is based on an ongoing joint work with Lara Caroline Pereira dos Santos (LCPdS) [4]. I initiated the project and derived all the theoretical results, crafted the initial python codes to evaluate the performance of the QAOA for the

BPSP in the large n limit as well as generating all instances of the BPSP. LCPdS will include additional analysis using the D-wave quantum annealer. All results in this thesis is based solely on my contribution.

8.1 Binary Paint Shop Problem

The paint shop problem is a set of various optimization problem in the automobile industry where given a sequence of cars with a certain constraints on how to paint them, minimize the number of times we need to swap the paint colour. The problem was first introduced in [104] where it was proven to be an NP-complete problem if either the number of colours or the number of car body types are unbounded. Loosely speaking, in decreasing order of difficulty, there are three main type of paint shop problem i) the multi-car multi-colour paint shop problem where we have multiple occurrence of each car body type with multiple paint colours ii) the multi-car paint shop problem where we restrict to only two colouring, iii) the binary paint shop problem (BPSP) where each car type only occurs twice. The focus of this thesis will be on the BPSP which can be defined as the following:

Definition 8.1.1. (Binary Paint Shop Problem (BPSP))
Let $A = \{a_1, a_2, \dots a_n\}$ be the set of car body types. Let $S = (s_1, \dots, s_{2n})$ be a sequence of car instances such that each $a_i \in A$ appears exactly twice in S. Denote the two paint colours as the set $C = \{0, 1\}$. The BPSP is to find a colouring sequence $f = (f_1, \dots, f_{2n}) \in \{0, 1\}^{2n}$ that minimises the colouring changes Δ_C between consecutive cars with the constraint that for each car body type, the two occurrence must be coloured differently (i.e. for any car body type a_i, if $s_j = s_k = a_i$ for $j \neq k$, we have $f_j \neq f_k$). This can be expressed mathematically as

$$\begin{aligned} &\text{minimize } \Delta_C = \sum_{i=1}^{2n-1} [f_i \neq f_{i+1}], \\ &\text{subject to } \quad f_i \neq f_j \quad \forall i, j : s_i = s_j, i \neq j, \end{aligned} \tag{8.1.1}$$

where $[\cdots]$ denotes the Iverson bracket which equals to 1 if the condition is satisfied and 0 otherwise.

Example 8.1.2. Consider the following instance of an $n = 3$ BPSP problem with the sequence

$$S = (a_1, a_2, a_1, a_3, a_3, a_2).$$

A valid solution corresponds to the colouring

$$f' = (0, 0, 1, 1, 0, 1), \quad \Delta'_C = 3.$$

However, this is not optimal since we have

$$f' = (0, 1, 1, 1, 0, 0), \quad \Delta'_C = 2.$$

The BPSP is an NP-complete problem in its decision form [104] and APX-Hard as an optimization problem [105]. Thus, unless P=NP, no polynomial time algorithm can find an arbitrary approximate solution on all instances, let alone solving them exactly. In order to evaluate the performance of various algorithms, consider the following expression: for fixed n, let the normalised expectation of Δ_C over all possible sequences of length $2n$ be given by $\mathbb{E}\,\Delta_C/n$. The performance of a classical greedy algorithm and recursive greedy algorithm have been found [106] to produce an expected colour change of

$$\begin{aligned} \text{greedy} \ &: \mathbb{E}\frac{\Delta_C}{n} = 0.5 + \mathcal{O}(1/n) \\ \text{recursive greedy} \ &: \mathbb{E}\frac{\Delta_C}{n} = 0.4 + \mathcal{O}(1/n). \end{aligned}$$

This raises the following interesting questions. Namely, what is the value of $\mathbb{E}\Delta_C/n$ in the limit $n \to \infty$ and does there exists a classical algorithm that achieves it? It was conjectured in [107] that $\mathbb{E}\Delta_C/n = o(n)$ but this has been shown to be false in [108] where they proved the following theorem.

Theorem 8.1.3 (Modified Theorem 1.4 of [108])**.** *The expected colour change $\mathbb{E}\Delta_C/n$ is lower bounded by*

$$2H^{-1}\left(\frac{1}{2}\right) - o(1) = 0.220.. - o(1), \tag{8.1.2}$$

where H^{-1} is the inverse of the binary entropy function. In addition, it is upper bounded by $0.4 + o(1)$.

Remark 8.1.4. The authors also remarked that the lower bound can be slightly improved to 0.227 and provided an outline of the proof.

The authors of [109] independently proved a weaker lower bound and slightly improved the upper bound to $0.4 - \epsilon$ for sufficiently small $\epsilon > 0$. In addition, they developed a new recursive star greedy algorithm that is conjectured to have an upper bound of 0.361. Even if their conjecture is true, there is still a substantial gap between the upper and lower bound.

On the quantum side, the first use of the QAOA for BPSP demonstrated that in the large n limit, the QAOA with a circuit depth of $p = 4$ outperforms the greedy algorithm while $p = 7$ outperforms the recursive greedy algorithm, the best classical algorithm at the time of publication, with an expected performance of 0.393 [74]. Subsequent works have evaluated the performance of QAOA variants for the BPSP. For example, [110] studied the performance of Recursive-QAOA (RQAOA) on the BPSP and found that reusing the optimal parameters in standard QAOA for RQAOA (i.e. parameter transfer) showed no noticeable reduction in solution quality. In addition, it was found that RQAOA outperformed the standard QAOA. A similar paper [111] studied the performance of RQAOA and eXpressive-QAOA (XQAOA) with circuit depth $p = 1$. Their numerical results showed that XQAOA not only outperforms RQAOA with an expected performance of 0.357, they also found evidence that the performance of RQAOA dimishes as the problem size increase.

8.2 BPSP as an Ising model

The algorithm to convert any instance of the BPSP into an Ising problem (more precisely a weighted MaxCut) was developed in [74] and made explicit independently in [110] and [111]. For completeness, we describe the steps in transforming the BPSP into an Ising problem. Each car body type i is associate with a single qubit z_i with the sign of the spin corresponding to the colour assigned to the first instance of car body type in the sequence. Let $\Omega : S \to A$ be a map that labels each s_i to an instance of a car body type in A. The coupling between the spins are calculated by going down the car sequence S and adding the coupling between adjacent cars s_i and s_{i+1} depending on their occurrence and car body type. If both car body type are distinct and s_i and s_{i+1} are both the first or second appearance of their car type, a coupling strength of -1 between the corresponding spins in the Ising formulation is added. If one of the car is the first occurrence of their car type and the other the second occurrence, a coupling strength of $+1$ is added between the corresponding spins. If both cars are of the same body type, a constant is added since we have $Z_iZ_i = 1$ for all i. Finally, a factor of $1/2$ is added for each term so that the energy of the Ising formulation corresponding to the number of colour changes. Thus, for n car body types, the BPSP Ising Hamiltonian is given by

$$H_{\text{BPSP}} = \frac{1}{2}\sum_{i=1}^{2n-1}\left(J'_{s_is_{i+1}}Z_{\Omega(s_i)}Z_{\Omega(s_{i+1})} + 1\right) = \frac{1}{2}\sum_{i=1}^{2n-1}J'_{s_is_{i+1}}Z_{\Omega(s_i)}Z_{\Omega(s_{i+1})} + \frac{2n-1}{2}, \tag{8.2.1}$$

where $J'_{s_i s_{i+1}}$ is the coupling strength assigned by the relation between the adjacent cars $s_i s_{i+1}$ as defined above and $\Omega(s_i)$ the car body type of the i-th car in the sequence.

The equation above can be further simplified by converting the sum over adjacent cars in the sequence into a sum over the car body types A. The details of which can be found in appendix C. Doing so results in a a simplified Hamiltonian given by

$$H_{\text{BPSP}} = \frac{1}{2} \sum_{(i,j)\in E} J_{ij} Z_i Z_j + \frac{c'}{2} \tag{8.2.2}$$

where J_{ij} is the effective coupling the Ising spins corresponding to the i and j car body type, E is the set of edges in the corresponding Ising graph, and $c' = 2n + \#\text{adjacent pairs of the same car type} - 1$. Note that J is a triangular matrix and not a symmetric matrix as with the standard Ising model formulation.

8.3 Improved bounds and classical algorithm

As we are interested in averaging over all sequences S in the limit $n \to \infty$ to find the average colour change $\mathbb{E}\Delta_C/n$, we note that $\lim_{n\to\infty} \mathbb{E}c'/n = 1$ (details in appendix C). Furthermore, it has been shown previously that in the large n limit, the coupling strength only take on the values $J_{ij} \in \{-1, +1\}$ with -1 occurring with probability 2/3 and $+1$ with probability 1/3 and that the underlying graph approaches that of a 4-regular tree [74]. Thus, we have

$$\lim_{n\to\infty} \mathbb{E}\frac{H_{\text{BPSP}}}{n} = \mathbb{E}\frac{1}{2n} \sum_{(i,j)\in E} J_{ij} Z_i Z_j + 1, \tag{8.3.1}$$

where $|E| = 2n$.

In the large n limit, this problem is equivalent to the Max-2-XORSAT problem. It is known that up to logarithmic-depth, the performance of the QAOA on sparse instances of Max-2-XORSAT is independent of the choice of J and so we can simply assume a MaxCut problem [11], [33]. Since the underlying graph is a 4-regular tree, we can use the method as outlined in chapter 4 to evaluate the performance of the QAOA on $D+1$-regular trees where the expectation of the QAOA at constant depth p is given by

$$\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | Z_i Z_j | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle = -\frac{2\nu_p(D, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\sqrt{D}}. \tag{8.3.2}$$

In order to test the validity of eq. (8.3.2), we reproduced the results of [74] up

depth p \Method	$\nu_p(3,\boldsymbol{\gamma},\boldsymbol{\beta})_{\text{SK}}$	$\nu_p(3,\boldsymbol{\gamma},\boldsymbol{\beta})_{\text{best}}$	QAOA simulation	$\nu_p(\boldsymbol{\gamma},\boldsymbol{\beta})$
1	0.679	0.675	0.675	0.650
2	0.572	0.568	0.568	0.529
3	0.508	0.503	0.503	0.454
4	0.467	0.462	0.462	0.405
5	0.438	0.432	0.432	0.368
6	0.417	0.410	0.411	0.339
7	0.402	0.393	0.393	0.317

Table 8.1: Comparison of the values obtained by using the algorithm of [80] as compared to values obtained by explicitly simulating the QAOA circuit as reported in [74] up to $p = 7$. Note that the values of $\nu_p(3,\boldsymbol{\gamma},\boldsymbol{\beta})_{\text{SK}}$ were obtained by using optimal $(\boldsymbol{\gamma},\boldsymbol{\beta})$ for the Sherrington–Kirkpatrick model while $\nu_p(3,\boldsymbol{\gamma},\boldsymbol{\beta})_{\text{best}}$ were obtained after rescaling the parameters found in [74] and rounded off to 3 decimal place.

to $p = 7$. We find good agreement between the results of the tree tensor network method of [74] and eq. (8.3.2) as can be found in table 8.1.

Beyond constant depth QAOA analysis, we found that in chapter 5, eq. (8.3.2) is in fact valid up to logarithmic depth. As a result, the performance of log-depth QAOA on eq. (8.3.1) can be simplified to

$$\mathbb{E}\lim_{n\to\infty}\lim_{p\to\epsilon\log n}\langle\boldsymbol{\gamma},\boldsymbol{\beta}|\frac{H_{\text{BPSP}}}{n}|\boldsymbol{\gamma},\boldsymbol{\beta}\rangle = 1-\frac{2\nu(3,\boldsymbol{\gamma},\boldsymbol{\beta})}{\sqrt{3}}. \tag{8.3.3}$$

Since we are unable to evaluate this term explicitly, we fit the curve obtained using $\nu_p(3,\boldsymbol{\gamma},\boldsymbol{\beta})$ of various depth p to estimate the value of $\nu(3,\boldsymbol{\gamma},\boldsymbol{\beta})$. Using the values up to $p = 7$ and assuming a power law fit to the function

$$\nu_p(3,\gamma,\beta) \approx d - \frac{a}{p^b + c}, \tag{8.3.4}$$

with parameters (a, b, c, d). The fitted parameters are

$$a = 1.0401 \pm 0.0357, \tag{8.3.5a}$$
$$b = 1.0497 \pm 0.0761, \tag{8.3.5b}$$
$$c = 1.958 \pm 0.1908, \tag{8.3.5c}$$
$$d = 0.6328 \pm 0.0122. \tag{8.3.5d}$$

With a goodness of fit $1 - R^2 = 2.0707 \times 10^{-6}$.

Surprisingly, using a power law fit for the QAOA's performance on the Sherrington–Kirkpatrick model, a similar behaviour was found in [87] where they fitted the curve

according to

$$\nu_p(\gamma,\beta) = (1+d)\Pi_2 - \frac{a}{p^b + c} \tag{8.3.6}$$

Using values from $p = 1$ to $p = 80$, this yields the fitted parameters of

$$a = 1.4912 \pm 0.10083, \tag{8.3.7a}$$
$$b = 0.9635 \pm 0.01617, \tag{8.3.7b}$$
$$c = 2.2540 \pm 0.1001, \tag{8.3.7c}$$
$$d = 0.001610 \pm 0.002252. \tag{8.3.7d}$$

With a goodness of fit $1 - R^2 = 1.805 \times 10^{-5}$.

Assuming that the power law fit holds, this suggests that at logarithmic depth, the performance of the QAOA for the binary paint shop problem is

$$\lim_{n\to\infty} \mathbb{E}\Delta_C/n = 1 - \frac{2\nu(3,\boldsymbol{\gamma},\boldsymbol{\beta})}{\sqrt{3}}, \quad \text{with} \tag{8.3.8a}$$
$$0.255 \leq 1 - \frac{2\nu(3,\boldsymbol{\gamma},\boldsymbol{\beta})}{\sqrt{3}} \leq 0.283. \tag{8.3.8b}$$

Using the fitted value gives us an expected value of 0.269 while the range of values comes about from the uncertainty of the fit.

As the underlying graph is locally a tree up to logarithmic depth in the large n limit, this implies that there is no quantum advantage offered by log-depth QAOA as argued in chapter 5. More precisely, [11] shows that a suitably applied approximate message passing algorithm will outperform logarithmic depth QAOA on sparse optimization problems such as the BPSP. Thus, one can construct an AMP algorithm to solve the weighted maxcut problem which would be equal — if not better — in performance to logarithmic depth QAOA. We end off with a conjecture regarding the performance of message passing algorithms on the BPSP.

Conjecture 8.3.1. *In the limit $n \to \infty$, AMP algorithms such as the MF-AOA will outperform all currently known classical heuristic algorithms. In particular, the MF-AOA achieves a performance of approximately*

$$\lim_{n\to\infty} \mathbb{E}\Delta_C/n = 1 - \frac{2\nu(3,\boldsymbol{\gamma},\boldsymbol{\beta})}{\sqrt{3}} \gtrsim 0.269 \tag{8.3.9}$$

Remark 8.3.2. We use $\gtrsim$ because the value is likely an overestimation of the parameter fitting which can similarly be observed if one attempts a similar fit with the Sherrinton–Kirkpatrick model with values up to $p = 7$ resulting in an overestimation of Π_2 and similarly with values of to $p = 80$ as seen in [87].

In order to test this conjecture, we perform a numerical simulation of the MF-AOA modifying some of the codes that were used in the original paper [112]. To estimate the algorithm's performance in the large n limit, we generated 1000 instances of the BPSP with problem sizes up to $n = 10000$. Note that given a problem with size n, there is a question in deciding the number of time-steps T which is proportional to the runtime of the algorithm. One could, in principle, set T to be arbitrarily large but we found that when we set $T \gg n$, the results were worse than simply setting $T = n$. For example, using the same $n = 2000$ instances, 4 different runs of the MF-AOA were done with $T \in \{1000, 2000, 5000, 10000\}$. We found that most instances had the best results using $T = 2000$ while a few others had slightly better results with $T = 5000$. Thus, simply setting T to be arbitrarily large does not seem to be the optimal solution. In order to simplify the analysis, we shall set $T = \max(1000, n)$.

We found that the MF-AOA had a expected paint swap ration of $\mathbb{E}\Delta_C/n = 0.27993$ with a standard deviation of $\mathbf{s} = 0.001469$ when $n = 10000$ outperforming all known classical heuristics, applications of the QAOA and its variants on the BPSP. Further details can be found in fig. 8.1 and table 8.2. We believe, similar to the authors of [11], that AMP algorithms are optimal for the average case sparse optimization problem akin to how it is believed that the Goemans–Williamson algorithm for MaxCut is optimal for the worst possible case assuming the unique games conjecture [113].

The difficulty with proving conjecture 8.3.1 is that, to the best of the author's knowledge, the performance guarantee of AMP algorithms is only well understood when, loosely speaking, one takes the large n limit followed by setting the average degree of the graph to infinity (i.e. $\lim_{D\to\infty}\lim_{n\to\infty}$). For the BPSP, the degree is fixed at $D = 4$. The performance of AMP algorithms on MaxCut is known [114] to be bounded by

$$\frac{1}{2} + \Pi_2\sqrt{\frac{1}{D}} + o_D\left(\frac{1}{\sqrt{D}}\right), \tag{8.3.10}$$

which converted to the BPSP, gives us a value of

$$\begin{aligned}\mathbb{E}\Delta_C/n &= 1 - \Pi_2 + o_D\left(\frac{1}{\sqrt{D}}\right)\\ &= 0.2368 + o_D\left(\frac{1}{\sqrt{D}}\right).\end{aligned} \tag{8.3.11}$$

For $D = 4$, the sub-leading term is not suppressed and cannot be ignored leading to a correction that is difficult to quantify. Similarly, assuming that the bounds for MaxCut and Max-2-XORSAT (and hence BPSP) are similar, then using the

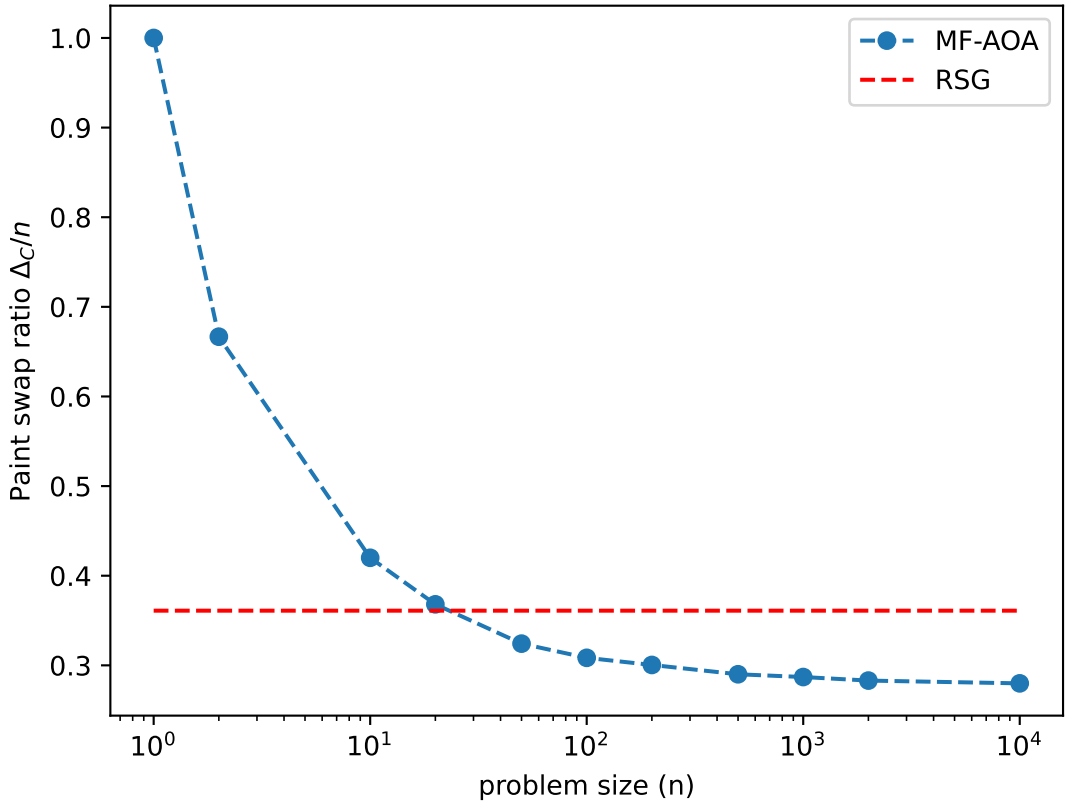


Figure 8.1: A semi-log plot of the average paint swap ratio achieved by the MF-AOA over a 1000 random generated instances for problem sizes $n \in \{10, 20, 50, 100, 200, 500, 1000, 2000, 10000\}$. In order to verify the accuracy of the Monte Carlo sampling, we similarly did it for the $n = 2$ case which is known to have an expected paint swap ratio of $2/3$. Averaging over a thousand instances, we found that the MF-AOA gets an average paint swap ratio of 0.6726 and a variance of 0.06 thus validating our method.

known upper-bound of 0.8683 and lower-bound of 0.8333 for Maxcut on a 4-regular graph [115], the expected paint swap ratio would be lower-bounded by 0.2634 and upper-bounded by 0.3334. Our numerical results of an expected paint swap ratio of approximately 0.27993 fits within these bounds which gives us good reason to believe that the algorithm is near-optimal or possibly even optimal.

n	$\mathbb{E}\Delta_C/n$	$\mathbf{s}(\Delta_C/n)$
10	0.4	0.1
20	0.37	0.07
50	0.325	0.035
100	0.31	0.02
200	0.30	0.01
500	0.289	0.008
1000	0.287	0.005
2000	0.283	0.004
10000	0.27993	0.0014

Table 8.2: Results from the numerical simulations of the MF-AOA over 1000 randomly generated BPSP instances for each n.

Chapter 9

Conclusion

In this chapter, we summarize the main result and findings of the thesis. The first half of the thesis is concerned with laying the ground work for studying the quantum approximate optimization algorithm (QAOA). In chapter 2, we defined the framework for what counts as an average case hardness for a combinatorial optimization problem with a focus on spin glass. While significant effort has been spend to identify the energy of the optimal solution in a typical spin glass instance, many efficient classical algorithms are unable to achieve them resulting in a statistical-computation gap. One method to identify when and where this gap occurs is the overlap gap property (OGP) which we reviewed in section 2.4. The occurrence of the OGP implies algorithmic limitation of a wide variety of algorithms which includes logarithmic depth quantum algorithms. In chapter 3, we investigate the properties of the QAOA and verify that it is a local algorithm that converges to the global solution as the circuit depth approaches infinity alongside its classical counterpart, the Mean-Field Approximate Optimization Algorithm. In chapter 4, we review the classical algorithm of [80] that evaluates the performance of constant depth QAOA on a regular graph.

In the second half of the thesis, three aspects of the QAOA and its relation to optimization problems were studied. The first aspect is on the limitations of logarithmic depth QAOA. In chapter 5, I proved that the algorithm of [80], to evaluate the performance of the QAOA, is in fact valid up till logarithmic depth. I moreover simplified the proof that the QAOA is unable to surpass the OGP barrier with logarithmic circuit depth. The second aspect concerns finding good variational parameters for the QAOA. In chapter 6, we reviewed the heuristic method of linear interpolation and Fourier decomposition of [72]. We then showed that the OGP is a sufficient condition that causes the variational parameters that optimises the approximation ratio (AR) to differ from those that optimises the time-to-solution (TTS) and propose a heuristic method to estimate $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ given $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$. The last aspect focuses on when the QAOA can exhibit a quantum advantage over classi-

cal algorithms. In chapter 7, we identify what are the qualities required in a problem for the QAOA to succeed over classical algorithms. Ideally, we want the QAOA to exhibit a quantum advantage at shallow depth. This is possible under several conditions. First, the QAOA requires a sufficiently large circuit depth to explore the full graph problem. This means a circuit depth greater than logarithmic depth for sparse optimization problem and as such, constant depth advantage can only work on dense problems. In addition, the problem must not exhibit the OGP. An example of which we covered in section 7.1 where $p = 1$ QAOA requires at most $\mathcal{O}(\sqrt{n})$ measurements to find the optimal solution for a symmetric optimization problem, a non-local problem. We then showed that the QAOA can outperform its classical counterpart, the mean-field approximate optimization algorithm (MF-AOA) since the latter is a type of approximate message passing algorithm which is limited in performance by the OGP. Finally we studied the circuit depth required by the QAOA in order to surpass the OGP barrier.

In the final part of the thesis, we use all of the knowledge gained in the previous chapters on a real-world industry problem, albeit a simplified one. By applying the QAOA to the Binary Paint Shop Problem in chapter 8, I improved both the analytic and numerical upper-bound on the expected number of colour change for the BPSP in the large n limit. Furthermore, given the results of chapter 5, this implies that there must exists an efficient classical algorithm that is able to outperform the QAOA and verified numerically with the MF-AOA. While it is beyond this thesis to prove the performance guarantee of such algorithms, the application of the QAOA to identify scenarios where efficient classical algorithms have yet to be discovered is novel. Thus, even if the application of polynomial circuit depth quantum algorithms remain far out of reach, their effect can already be felt by creating novel classical algorithms today.

In view of the findings of this thesis, some open question remain which we list below:

Q1. What are necessary and sufficient condition for non-adiabatic schedule in the QAOA?

Q2. What are the necessary and sufficient condition for a difference in parameter optimization between those that optimises the approximation ratio (AR) and those that optimises the time-to-solution (TTS)?

Q3. Is an adiabatic schedule the optimal schedule for problems that do not exhibit the OGP and/or weak clustering property such as MaxCut?

Q4. Are the variational parameters that optimises the AR and the TTS the same in problems that do not exhibit the OGP and/or weak clustering property

such as MaxCut?

From the results of section 6.2, the answer to question 1 and 2 seems to be that the OGP is a sufficient condition of non-adiabatic schedule in $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ as well as a difference in the optimal schedule for between the two methods of quantifying the QAOA. However, it does not seem to be a necessary condition since similar effects were observed in small number of instances that exhibit a weak clustering property.

Of greatest importance, at least to the author, would be the answer to the third and fourth question. This is because for the QAOA up to logarithmic depth, the optimal parameters that optimises the AR for spin glass type problems are universal and work well across any instance. From our numerical analysis in section 6.2, it seems that the adiabatic schedule and the universality of parameters across different instances hold beyond logarithmic depth and might even hold up to exponential depth. If true, then problems such as the barren plateau becomes less relevant as a hindrance for variational quantum algorithms when applied to problems such as MaxCut since one can simply reuse optimal parameters in one instance as ansatz for another.

Appendix A

Alternative proof of theorem 5.2.1

We provide here an alternate proof of theorem 5.2.1 that relies on a conjecture that seems likely to be true as [116] notes that in the planted clique problem, the occurrence of the OGP is related to the monotonicity of another graph property. For the Max-q-XORSAT, it is reasonable to think that the OGP is related to the density of edges in the graph which is clearly monotonically increasing.

Conjecture .1. *The OGP of Max-q-XORSAT on any graph is a monotonically increasing property in the sense that if the graph G exhibits the overlap gap property, then adding an additional edge does not destroy the graph exhibiting the OGP i.e. $G + e$ exhibits the OGP.*

Another result that we need comes from the fact that we can embed an Erdös–Rényi hypergraph into a random regular hypergraph.

Theorem .2 (Theorem 1 and Corollary 2 of [117])**.** *For each $q \geq 2$ there is a positive constant C such that if for some real $\gamma = \gamma(n)$ and positive integer $d = d(n)$,*

$$C\left(\left(d/n^{q-1} + \log(n)/d\right)^{1/3} + 1/n\right) \leq \gamma < 1, \tag{.1}$$

and $m = (1-\gamma)nd/q$ is an integer, then there is a joint distribution of $\mathbb{G}^q_{ER}(n,m)$ and $\mathbb{G}^q(n,d)$ with

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}^q_{ER}(n,m) \subset \mathbb{R}^q(n,d)) = 1. \tag{.2}$$

Furthermore, let $\mathcal{P}$ be a monotone increasing property. If $\log n \ll d \ll n^{q-1}$, for some $m \leq (1-\gamma)nd/q$ where γ satisfies Eq. (.1), then if $\mathbb{G}^q_{ER}(n,m) \in \mathcal{P}$ as $n \to \infty$, then $\mathbb{R}^q(n,d)) \in \mathcal{P}$ as $n \to \infty$.

For finite n and d, the proof trivially follows from the assumption of monotonicity. In the case of infinite n and d, we have to prove that our condition for large, potentially infinite, girth fits this condition for the embedding.

The maximum girth g of a d-regular q-uniform hypergraph is known [118] to be bounded by

$$\frac{\log n - \log 4}{\log(q-1) + \log(d-1)} - 1 < g \leq \frac{2\log n}{\log(q-1) + \log(d-1)} + 2. \tag{.3}$$

For constant q, an infinite girth requires $d \ll n$

Let $d \sim \mathcal{O}(n^\epsilon)$ for sufficiently small $\epsilon > 0$ and $n^\epsilon > q$. Such a constraint satisfies the large girth requirement. Substituting d into Eq. (.1) gives us

$$C\left(\left(\frac{n^\epsilon}{n^{q-1}} + \frac{\log n}{n^\epsilon}\right)^{1/3} + 1/n\right) \leq \gamma < 1, \tag{.4}$$

where in the large n limit, we see that the lhs. approaches 0. Thus, for $m = (1-\gamma)nd/q \sim \mathcal{O}(n^{1+\epsilon})$, an embedding can be performed.

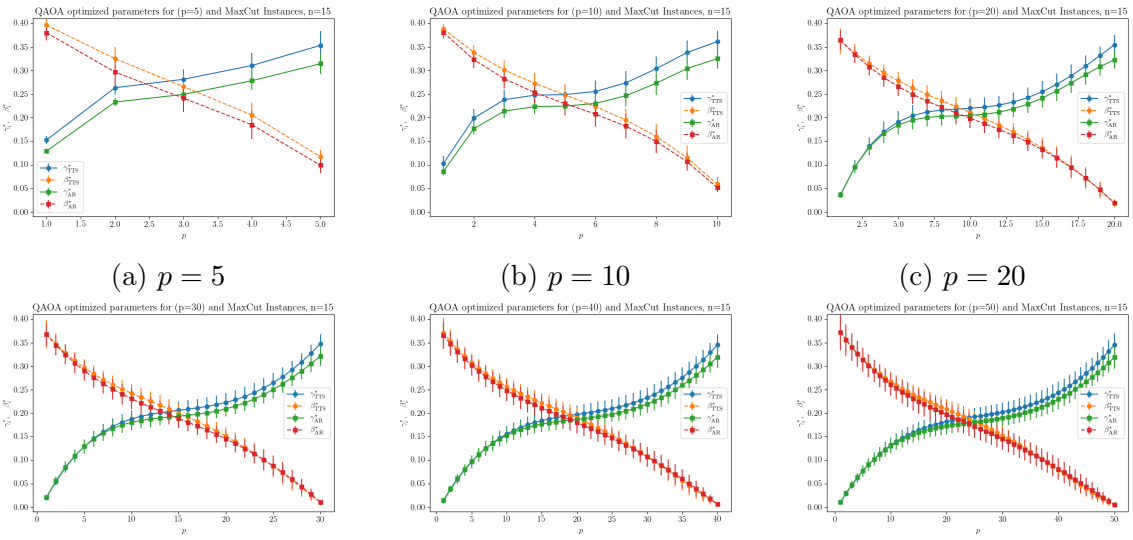
From this, it follows that for some $m^* \leq (1-\gamma)nd/q$, if $\mathbb{G}^q_{ER}(n,m)$ has a monotone increasing property $\mathcal{P}$, then $\mathbb{R}^q(n,d)$ also has it as well. For m to be drawn from a Poisson distribution where the graph has connectivity λ, we require $m = n(\log n + (\lambda-1)\log\log n + c)/d \sim \mathcal{O}(n^{1-\epsilon}\log n)$ for some finite constant c [119]. Thus, there exists a d_0 such that, for $d > d_0$, the existence of the OGP is present in the solution space of Max-q-XORSAT on Random Regular hypergraph $\mathbb{R}^q(n,d)$ with high probability meaning that the result of [33] also applies to random regular hypergraphs.

Appendix B

Numerical analysis of the QAOA

A Behaviour of variational parameters

Here, we provide the figures detailing the evolution of how the parameters $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ evolve as the depth p increases for the Max-2-XORSAT instances in fig. B.1 and the Max-4-XORSAT instances that exhibit the OGP in fig. B.2 and without the OGP in fig. B.3.

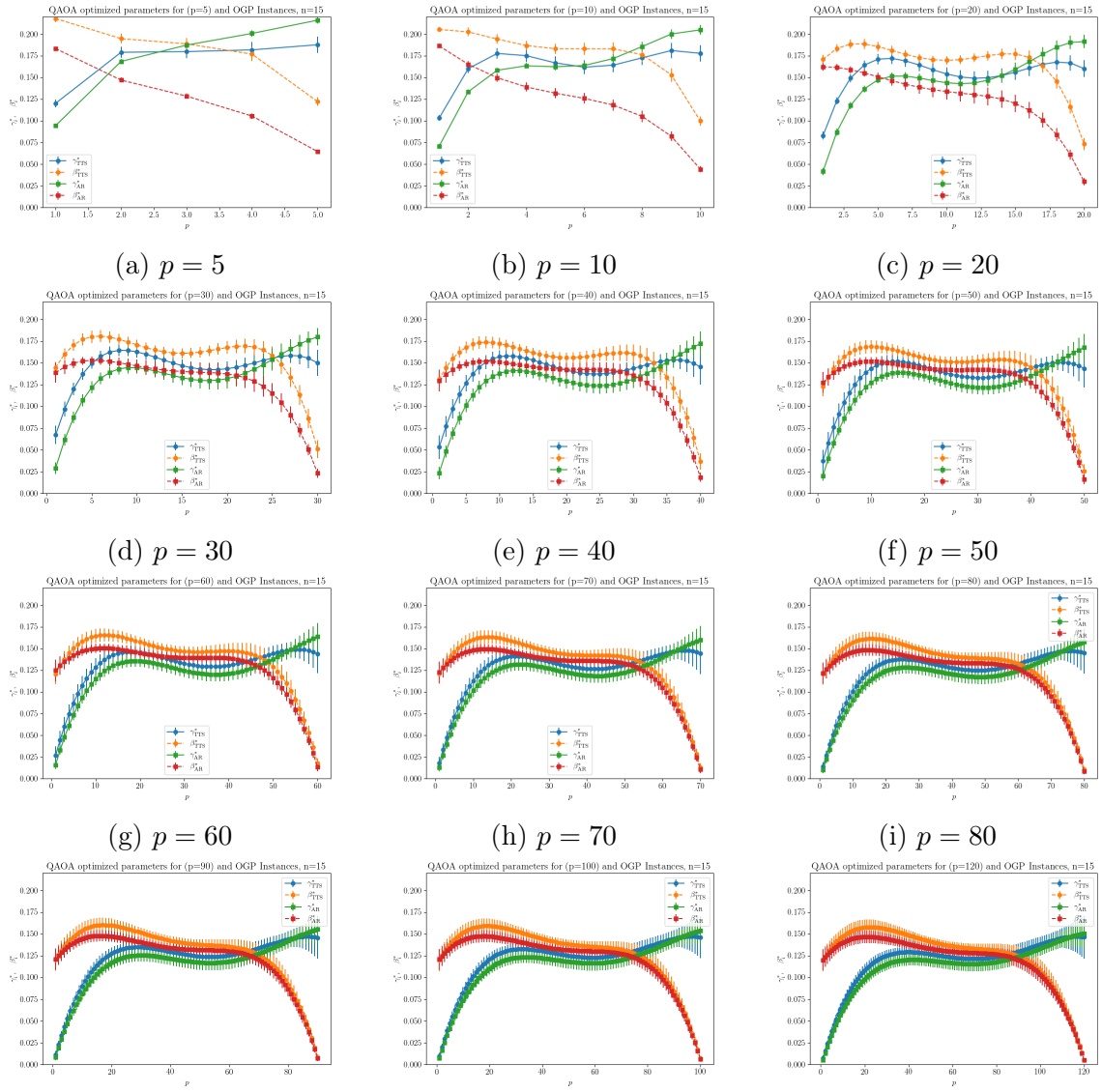


(a) $p = 5$ (b) $p = 10$ (c) $p = 20$

(d) $p = 30$ (e) $p = 40$ (f) $p = 50$

Figure B.1: The optimal parameters of $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ and $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ at various depths for Max-2-XORSAT instances.

While $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ is consistently larger than $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ for both the 2-XOR and 4-XOR problems, we note that the difference is considerably small for the 2-XOR case even at low depth while for the 4-XOR case, the gap is rather significant particularly in $\boldsymbol{\beta}$ at low depth consistent with the findings of [87]. Note that the $p = 5$ QAOA for the 4-XORSAT exhibits an adiabatic behaviour consistent with the theory that at low depth, the variational parameters are universal [80]. Another noticeable feature



(a) $p = 5$ (b) $p = 10$ (c) $p = 20$

(d) $p = 30$ (e) $p = 40$ (f) $p = 50$

(g) $p = 60$ (h) $p = 70$ (i) $p = 80$

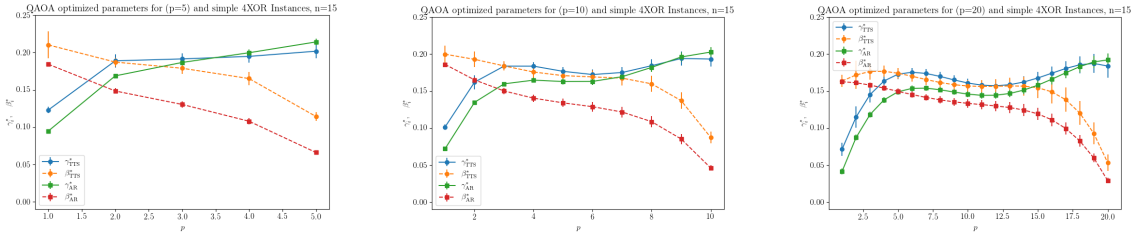
(j) $p = 90$ (k) $p = 100$ (l) $p = 120$

Figure B.2: The optimal parameters of $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ and $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ at various depths for Max-4-XORSAT instances exhibiting the OGP.

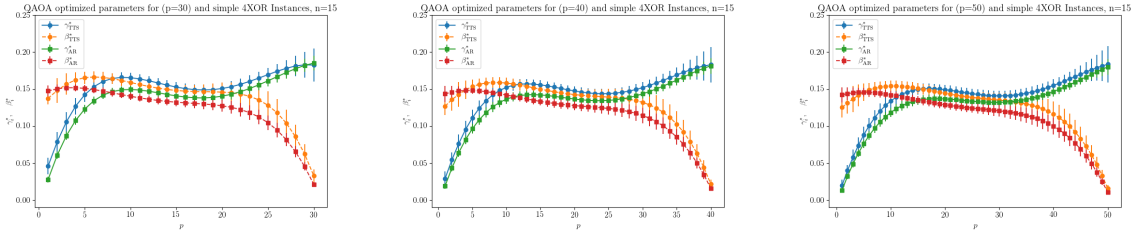
is that while $\boldsymbol{\gamma}_{AR}$ has consistently lower values than $\boldsymbol{\gamma}_{TTS}$, the situation is reversed in the last few layers. This overestimation in the last few layers of $\boldsymbol{\gamma}_{AR}$ is consistently across all p.

B Occurrence of phase transition

We provide here a sample of instances with varying n to verify that the 'phase transition' occurring in the QAOA's performance is not unique to any individual instance.

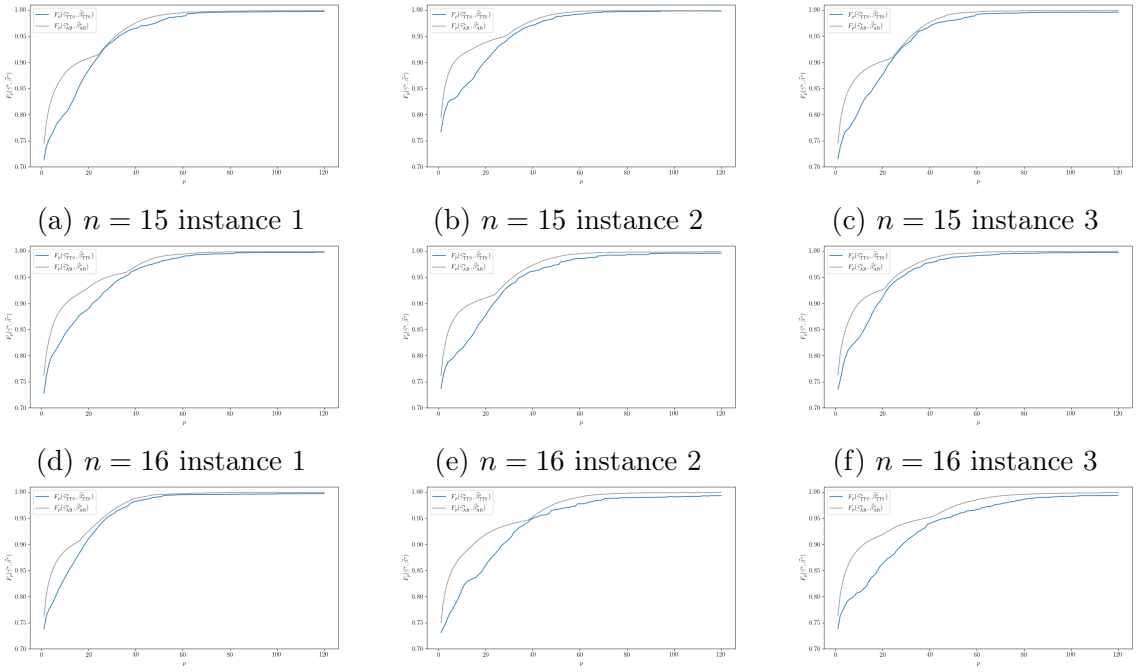


(a) $p = 5$ (b) $p = 10$ (c) $p = 20$



(d) $p = 30$ (e) $p = 40$ (f) $p = 50$

Figure B.3: The optimal parameters of $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{AR}$ and $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$ at various depths for Max-4-XORSAT instances without the OGP.



(a) $n = 15$ instance 1 (b) $n = 15$ instance 2 (c) $n = 15$ instance 3

(d) $n = 16$ instance 1 (e) $n = 16$ instance 2 (f) $n = 16$ instance 3

(g) $n = 17$ instance 1 (h) $n = 17$ instance 2 (i) $n = 17$ instance 3

Figure B.4: A sample of instances of varying n. Note that while the expectation values of the QAOA is always lower when using $(\boldsymbol{\gamma}, \boldsymbol{\beta})_{TTS}$, the qualitative behaviour of the curve is approximately the same at some critical depth p^*.

Appendix C

Details of the BPSP Ising formulation

A Simplification of BPSP Hamiltonian

Starting with the BPSP Ising Hamiltonian, we now provide a step by step proof to simplify it into the standard Ising model following the procedure of [110]. Let b_i be the car body type associated with spin/qubit i to distinguish it from s_i (the i-th car in the sequence).

$$H_{\text{BPSP}} = \frac{1}{2}\sum_{i=1}^{2n-1} J'_{s_i s_{i+1}} Z_{\Omega(s_i)} Z_{\Omega(s_{i+1})} + \frac{2n-1}{2} \tag{A.1}$$

$$= \frac{1}{2}\sum_{k=1}^{n}\sum_{j=1}^{k}\left(\sum_{i=1}^{2n-1} J'_{s_i s_{i+1}}[\Omega(s_i), \Omega(s_{i+1}) = (b_j, b_k)]\right) Z_{\Omega(b_j)} Z_{\Omega(b_k)} + \frac{2n-1}{2} \tag{A.2}$$

$$= \frac{1}{2}\left(\sum_{i,j\in E} J_{ij} Z_i Z_j + \sum_{i=1}^{2n-1}[\Omega(s_i), \Omega(s_{i+1}) = (b_j, b_j)]\right) + \frac{2n-1}{2} \tag{A.3}$$

$$= \frac{1}{2}\sum_{i,j\in E} J_{ij} Z_i Z_j + \frac{2n + C(\mathcal{I}) - 1}{2} \tag{A.4}$$

where $[\dots]$ in eq. (A.2) denotes the Iverson bracket and $C(\mathcal{I}) \in [0, n]$ denotes the number of adjacent pairs of cars in the sequence with the same body type for a specific problem instance $\mathcal{I}$. Equation (A.3) is obtained by summing up the set of edges in the Ising graph while excluding the self-interacting $i = j$ term. Note that because we are summing up over the sequence, the J matrix is a triangular matrix and not a symmetric one. If one wishes to use a symmetric formulation, the prefactor of 1/2 has to be modified to 1/4.

B Probability Distribution of coupling term and constants

In this section, we prove several properties of the Ising formulation of the BPSP.

B.1 Adjacent car body types

We prove that the term $C(\mathcal{I})$ goes to 0 for a random sequence in the large n limit. To prove this, Let A_j be the probability that a specific car body type b_j is adjacent to it self in a sequence. Fixed any car body type at site s_i. The probability that it is next to its pair is given by

$$A_j = \frac{1}{2n}\frac{2}{2n-1} \times 2 = \frac{2}{n(2n-1)}, \tag{B.1}$$

where the first term comes from the probability that $\Omega(s_i) = b_j$, the next term for the second occurrence of body type b_j being at site $s_i - 1$ or s_{i+1}, and the factor of 2 due to symmetry. By a union bound, we have

$$P\left(\bigcup_{j=1}^{n} A_j\right) \leq \frac{2}{(2n-1)}. \tag{B.2}$$

Thus, we see that in the limit $n \to \infty$, the probability of a car with body type b_j being adjacent to itself goes to 0. As such, for any arbitrary instance $\mathcal{I}$ of the BPSP, we have $C(\mathcal{I}) = 0$ with high probability in the limit $n \to \infty$.

B.2 Coupling strength

It has been previously shown that coupling strength takes on the values of -1 with probability 2/3 and +1 with probability 1/3 in the limit $n \to \infty$ [74]. We reproduce those results here along side the asymptotic behaviour of the coupling strength -2 which vanishes in probability.

A coupling strength of $J = -2$, occurs only when a car body type has the same neighbour where both body types occur for the first and second time in the sequence. Let A_i be the event that an arbitrary car body type b_i has a coupling of $J = -2$.

The probability of A_i occurring is

$$\begin{aligned} P(A_i) &= P(A_i \mid s_1 = s_{2n} = b_i)P(s_1 = s_{2n} = b_i) \\ &\quad + P(A_i \mid s_1 \neq b_i \neq s_{2n})P(s_1 \neq b_i \neq s_{2n}) \\ &\quad + 2P(A_i \mid s_1 = b_i, s_j = b_i, j \neq 2n)P(s_1 = b_i, s_j = b_i, j \neq 2n) \qquad \text{(B.3)} \\ &\approx \frac{1}{2n-3}\frac{2}{2n(2n-1)} + \frac{2n-4}{2n-3}\frac{4}{(2n-4)}\frac{2n-2}{2n} + 2\frac{2}{2n-3}\frac{1}{2n(2n-3)} \qquad \text{(B.4)} \\ &\sim \mathcal{O}(1/n) \qquad \text{(B.5)} \end{aligned}$$

where the approximation comes from only taking the dominant terms in each possible case (e.g. for the second term, we only consider the case where the two instances of body type b_i are not adjacent to each other). Thus, as $n \to \infty$, we see that the probability of obtaining a coupling strength of -2 goes to 0.

For the coupling strength of $J = -1$, this occurs when the adjacent cars both occur for the first time or both have already appeared in the sequence before. Thus, The probability of a coupling being -1 is given by

$$P(J = -1) = \frac{1}{2n-1}\sum_{ij} P_{ij}(00) + P_{ij}(11), \qquad \text{(B.6)}$$

where $P_{ij}(00)$ denotes the probability that the cars in position i, j are both the first occurrence and $P_{ij}(11)$ for the case when both cars appear the second time. For a car in position i, the probability that it has occurred before is given by $(i-1)/2n$. Thus, we can express eq. (B.6) as

$$\frac{1}{2n-1}\sum_{i=1}^{2n-1}\left[\left(1 - \frac{i-1}{2n-1}\right)\left(1 - \frac{i}{2n-1}\right) + \frac{i-1}{2n-1}\frac{i}{2n-1}\right] = \frac{1-6n+8n^2}{6n(2n-1)}. \qquad \text{(B.7)}$$

In the limit $n \to \infty$, this leads to a probability of $2/3$. A similar calculation can be performed for the coupling strength $J = +1$ where we have

$$\begin{aligned} P(J = +1) &= \frac{1}{2n-1}\sum_{ij} P_{ij}(01) + P_{ij}(10) \qquad \text{(B.8)} \\ &= \frac{1}{2n-1}\sum_{i=1}^{2n-1}\left[\left(1 - \frac{i-1}{2n-1}\right)\frac{i}{2n-1} + \left(1 - \frac{i-1}{2n-1}\right)\frac{i}{2n-1}\right] \qquad \text{(B.9)} \\ &= \frac{4n^2-1}{6n(2n-1)}, \qquad \text{(B.10)} \end{aligned}$$

for which, in the limit $n \to \infty$, gives us a probability of $1/3$.

One might have noticed that for finite n, adding $P(J = +1)$ and $P(J = -1)$

gives us unity which doesn't seem right since there must be a non-zero probability to obtain $J = -2$. This is not a mistake and $P(J = \pm 1)$ is the probability of assigning a coupling strength for a given adjacent pair in the sequence. Thus, if the same adjacent pair occurs twice in the sequence, the coupling strength of -1 is additive and would lead to an effective coupling of -2. However, as argued above, the probability of the same adjacent pair occurring twice is on the order of $\mathcal{O}(1/n)$ and thus, the probability of having a coupling strength $J = -2$ goes to 0 in the limit $n \to \infty$.

C Average Degree

We provide a short but simple proof that the average degree equals to 4 with high probability in the limit $n \to \infty$.

First, we note that by construction, the degree for any Ising spin is upper bounded by 4 since a car body type is connected to at most 4 other distinct car body type. As for the lower bound, we note that the Ising spin has less than degree 4 if and only if in the sequence, the car body type associated with the Ising spin is i) adjacent to itself or ii) adjacent to the same car body type twice or iii) is at the start or the end of the sequence. Let A be the event that a body type is adjacent to itself in the sequence and B be the event that a body type is adjacent to a different body type twice and C be the event it is at the start or the end of the sequence. Thus, the probability that a spin has less than degree 4 is given by

$$P(A) + P(B) + P(C) = \frac{2}{2n-1} + \left(\frac{2n-2}{(2n-1)(2n-2)} + \frac{2n-2}{2n-1}\frac{4}{2n-3} \right) + \frac{1}{n}. \tag{C.1}$$

Since the leading order is $\mathcal{O}(1/n)$, the probability that an arbitrary spin having degree less than 4 goes to 0 with high probability as $n \to \infty$. Note that this does not mean that such an event does not exists, it merely states that the event has a measure of 0.

Appendix D

Tree angles for the BPSP

We include the angles used to evaluate the performance of the QAOA for the binary paint shop problem in the limit $n \to \infty$. Note that $\boldsymbol{\gamma}$ was used with the scaled unitary operator $\exp\Big\{-i\gamma\frac{1}{\sqrt{D}}\sum Z_jZ_k\Big\}$ for a $D+1$-regular graph instead of the usual QAOA unitary operator $\exp\big\{-i\gamma\frac{1}{2}\sum Z_jZ_k\big\}$.

p	Angles for MaxCut with $D=3$
1	$\boldsymbol{\gamma} = (0.4535)$ $\boldsymbol{\beta} = (0.3927)$
2	$\boldsymbol{\gamma} = (0.3532, 0.6406)$ $\boldsymbol{\beta} = (0.5341, 0.283)$
3	$\boldsymbol{\gamma} = (0.3070, 0.5641, 0.6532)$ $\boldsymbol{\beta} = (0.5879, 0.4232, 0.2230)$
4	$\boldsymbol{\gamma} = (0.2728, 0.5089, 0.5830, 0.6679)$ $\boldsymbol{\beta} = (0.6050, 0.4778, 0.3613, 0.1875)$
5	$\boldsymbol{\gamma} = (0.2519, 0.4735, 0.5225, 0.5951, 0.6793)$ $\boldsymbol{\beta} = (0.6225, 0.5051, 0.4167, 0.3253, 0.1628)$
6	$\boldsymbol{\gamma} = (0.2327, 0.4441, 0.4881, 0.5318, 0.6025, 0.6813)$ $\boldsymbol{\beta} = (0.6293, 0.5232, 0.4528, 0.3883, 0.2981, 0.1459)$
7	$\boldsymbol{\gamma} = (0.2198, 0.4235, 0.4605, 0.4986, 0.5382, 0.6043, 0.6830)$ $\boldsymbol{\beta} = (0.6378, 0.5327, 0.4719, 0.4325, 0.3632, 0.2778, 0.1339)$
8	$\boldsymbol{\gamma} = (0.2083, 0.4062, 0.4428, 0.0.4746, 0.5067, 0.5419, 0.6254, 0.7222)$ $\boldsymbol{\beta} = (0.6405, 0.5385, 0.4817, 0.4526, 0.4101, 0.3452, 0.2605, 0.1198)$

Table D.1: optimised parameters that can be found in [74], [120] and converted to the convention used in this thesis.

Bibliography

[1] M. Goh, "The overlap gap property limits limit swapping in the QAOA," *Quantum Information & Computation*, vol. 25, no. 4, pp. 329–343, 2025. DOI: 10.2478/qic-2025-0018 [Online]. Available: https://doi.org/10.2478/qic-2025-0018

[2] M. Goh and T. Müller, *Mind the gap: The QAOA outperforms the MF-AOA in surpassing the OGP barrier*, forthcoming, 2026.

[3] M. Goh and T. Müller, *Watch that gap: Parameter optimisation in the QAOA depends on the OGP*, forthcoming, 2026.

[4] M. Goh, L. C. P. dos Santos, and M. Sperl, *No quantum advantage implies improved bounds and classical algorithms for the binary paint shop problem*, 2026. arXiv: 2604.00607 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2604.00607

[5] D. Gamarnik, "The overlap gap property: A topological barrier to optimizing over random structures," *Proceedings of the National Academy of Sciences*, vol. 118, no. 41, e2108492118, 2021. DOI: 10.1073/pnas.2108492118 eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.2108492118. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.2108492118

[6] D. Gamarnik, C. Moore, and L. Zdeborová, "Disordered systems insights on computational hardness," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2022, no. 11, p. 114 015, Nov. 2022. DOI: 10.1088/1742-5468/ac9cc8 [Online]. Available: https://dx.doi.org/10.1088/1742-5468/ac9cc8

[7] S. Bhamidi, D. Gamarnik, and S. Gong, *Finding a dense submatrix of a random matrix. sharp bounds for online algorithms*, 2025. arXiv: 2507.19259 [math.PR]. [Online]. Available: https://arxiv.org/abs/2507.19259

[8] D. Gamarnik, E. C. Kızıldağ, and L. Warnke, *Optimal hardness of online algorithms for large independent sets*, 2025. arXiv: 2504.11450 [cs.DS]. [Online]. Available: https://arxiv.org/abs/2504.11450

[9] B. Huang and M. Sellke, "Tight lipschitz hardness for optimizing mean field spin glasses," in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, 2022, pp. 312–322. DOI: 10.1109/FOCS54457.2022.00037

[10] E. R. Anschuetz, *Efficient learning implies quantum glassiness*, 2025. arXiv: 2505.00087 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2505.00087

[11] A. Chen, N. Huang, and K. Marwaha, *Local algorithms and the failure of log-depth quantum advantage on sparse random csps*, 2023. arXiv: 2310.01563 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2310.01563

[12] P. Benioff, "The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines," *Journal of Statistical Physics*, vol. 22, no. 5, pp. 563–591, May 1980. DOI: 10.1007/BF01011339

[13] P. Benioff, "Quantum mechanical hamiltonian models of turing machines," *Journal of Statistical Physics*, vol. 29, pp. 515–546, 1982. [Online]. Available: https://api.semanticscholar.org/CorpusID:14956017

[14] R. P. Feynman, "Simulating physics with quantum computers," *International Journal of Theoretical Physics*, vol. 21, pp. 467–488, 1982. [Online]. Available: https://link.springer.com/article/10.1007/BF02650179

[15] D. Deutsch and R. Jozsa, "Rapid solution of problems by quantum computation," *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, vol. 439, no. 1907, pp. 553–558, 1992. DOI: 10.1098/rspa.1992.0167 eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1992.0167. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1992.0167

[16] E. Bernstein and U. Vazirani, "Quantum complexity theory," *SIAM Journal on Computing*, vol. 26, no. 5, pp. 1411–1473, 1997. DOI: 10.1137/S0097539796300921 eprint: https://doi.org/10.1137/S0097539796300921. [Online]. Available: https://doi.org/10.1137/S0097539796300921

[17] D. R. Simon, "On the power of quantum computation," *SIAM Journal on Computing*, vol. 26, no. 5, pp. 1474–1483, 1997. DOI: 10.1137/S0097539796298637 eprint: https://doi.org/10.1137/S0097539796298637. [Online]. Available: https://doi.org/10.1137/S0097539796298637

[18] P. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in *Proceedings 35th Annual Symposium on Foundations of Computer Science*, 1994, pp. 124–134. DOI: 10.1109/SFCS.1994.365700

[19] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM Journal on Computing*, vol. 26, no. 5, pp. 1484–1509, 1997. DOI: 10.1137/S0097539795293172 eprint: https://doi.org/10.1137/S0097539795293172. [Online]. Available: https://doi.org/10.1137/S0097539795293172

[20] L. K. Grover, "A fast quantum mechanical algorithm for database search," in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, ser. STOC '96, Philadelphia, Pennsylvania, USA: Association for Computing Machinery, 1996, pp. 212–219, ISBN: 0897917855. DOI: 10.1145/237814.237866 [Online]. Available: https://doi.org/10.1145/237814.237866

[21] A. W. Harrow, A. Hassidim, and S. Lloyd, "Quantum algorithm for linear systems of equations," *Phys. Rev. Lett.*, vol. 103, p. 150 502, 15 Oct. 2009. DOI: 10.1103/PhysRevLett.103.150502 [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.103.150502

[22] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, "Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics," in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, ser. STOC '19, ACM, Jun. 2019, pp. 193–204. DOI: 10.1145/3313276.3316366 [Online]. Available: http://dx.doi.org/10.1145/3313276.3316366

[23] C. Cade, M. Folkertsma, I. Niesen, and J. Weggemans, "Quantifying grover speed-ups beyond asymptotic analysis," *Quantum*, vol. 7, p. 1133, Oct. 2023, ISSN: 2521-327X. DOI: 10.22331/q-2023-10-10-1133 [Online]. Available: http://dx.doi.org/10.22331/q-2023-10-10-1133

[24] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018, ISSN: 2521-327X. DOI: 10.22331/q-2018-08-06-79 [Online]. Available: https://doi.org/10.22331/q-2018-08-06-79

[25] F. Arute et al., "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, pp. 505–510, 2019. [Online]. Available: https://www.nature.com/articles/s41586-019-1666-5

[26] H. Zhang, K. Boothby, and A. Kamenev, "Cyclic quantum annealing: Searching for deep low-energy states in 5000-qubit spin glass," *Scientific Reports*, vol. 14, no. 1, Dec. 2024, ISSN: 2045-2322. DOI: 10.1038/s41598-024-80761-z [Online]. Available: http://dx.doi.org/10.1038/s41598-024-80761-z

[27] C. J. Ballance, T. P. Harty, N. M. Linke, M. A. Sepiol, and D. M. Lucas, "High-fidelity quantum logic gates using trapped-ion hyperfine qubits," *Phys. Rev. Lett.*, vol. 117, p. 060 504, 6 Aug. 2016. DOI: 10.1103/PhysRevLett.117.060504 [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.117.060504

[28] J. J. Burnett et al., "Decoherence benchmarking of superconducting qubits," *npj Quantum Information*, vol. 5, no. 1, Jun. 2019, ISSN: 2056-6387. DOI: 10.1038/s41534-019-0168-5 [Online]. Available: http://dx.doi.org/10.1038/s41534-019-0168-5

[29] C. Gidney, *How to factor 2048 bit rsa integers with less than a million noisy qubits*, 2025. arXiv: 2505.15917 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2505.15917

[30] M. Cerezo et al., "Variational quantum algorithms," *Nature Reviews Physics*, vol. 3, no. 9, pp. 625–644, Aug. 2021, ISSN: 2522-5820. DOI: 10.1038/s42254-021-00348-9 [Online]. Available: http://dx.doi.org/10.1038/s42254-021-00348-9

[31] K. Blekos et al., "A review on quantum approximate optimization algorithm and its variants," *Physics Reports*, vol. 1068, pp. 1–66, Jun. 2024, ISSN: 0370-1573. DOI: 10.1016/j.physrep.2024.03.002 [Online]. Available: http://dx.doi.org/10.1016/j.physrep.2024.03.002

[32] A. Montanaro and L. Zhou, *Quantum speedups in solving near-symmetric optimization problems by low-depth QAOA*, 2025. arXiv: 2411.04979 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2411.04979

[33] C.-N. Chou, P. J. Love, J. S. Sandhu, and J. Shi, "Limitations of Local Quantum Algorithms on Random MAX-k-XOR and Beyond," in *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*, M. Bojańczyk, E. Merelli, and D. P. Woodruff, Eds., ser. Leibniz International Proceedings in Informatics (LIPIcs), vol. 229, Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, 41:1–41:20, ISBN: 978-3-95977-235-8. DOI: 10.4230/LIPIcs.ICALP.2022.41 [Online]. Available: https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ICALP.2022.41

[34] S. Arora and S. Safra, "Probabilistic checking of proofs: A new characterization of np," *J. ACM*, vol. 45, no. 1, pp. 70–122, Jan. 1998, ISSN: 0004-5411. DOI: 10.1145/273865.273901 [Online]. Available: https://doi.org/10.1145/273865.273901

[35] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy, "Proof verification and the hardness of approximation problems," *J. ACM*, vol. 45, no. 3, pp. 501–555, May 1998, ISSN: 0004-5411. DOI: 10.1145/278298.278306 [Online]. Available: https://doi.org/10.1145/278298.278306

[36] S. Arora and B. Barak, "Pcp theorem and hardness of approximation: An introduction," in *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009, pp. 237–256.

[37] A. M. Frieze and B. Reed, "Probabilistic analysis of algorithms," in *Probabilistic Methods for Algorithmic Discrete Mathematics*, M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 36–92, ISBN: 978-3-662-12788-9. DOI: 10.1007/978-3-662-12788-9_2 [Online]. Available: https://doi.org/10.1007/978-3-662-12788-9_2

[38] D. A. Spielman and S.-H. Teng, "Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time," *J. ACM*, vol. 51, no. 3, pp. 385–463, May 2004, ISSN: 0004-5411. DOI: 10.1145/990308.990310 [Online]. Available: https://doi.org/10.1145/990308.990310

[39] B. Bollobás, "A probabilistic proof of an asymptotic formula for the number of labelled regular graphs," *European Journal of Combinatorics*, vol. 1, no. 4, pp. 311–316, 1980, ISSN: 0195-6698. DOI: https://doi.org/10.1016/S0195-6698(80)80030-8 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0195669880800308

[40] D. Sherrington and S. Kirkpatrick, "Solvable model of a spin-glass," *Phys. Rev. Lett.*, vol. 35, pp. 1792–1796, 26 Dec. 1975. DOI: 10.1103/PhysRevLett.35.1792 [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.35.1792

[41] T. Castellani and A. Cavagna, "Spin-glass theory for pedestrians," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 05, P05012, May 2005, ISSN: 1742-5468. DOI: 10.1088/1742-5468/2005/05/p05012 [Online]. Available: http://dx.doi.org/10.1088/1742-5468/2005/05/P05012

[42] G. Parisi, P. Urbani, and F. Zamponi, "Replica symmetry breaking and hierarchical free energy landscapes," in *Theory of Simple Glasses: Exact Solutions in Infinite Dimensions*. Cambridge University Press, 2020, pp. 140–179.

[43] G. Parisi, "A Sequence of Approximated Solutions to the S-K Model for Spin Glasses," *J. Phys. A*, vol. 13, p. L115, 1980. DOI: 10.1088/0305-4470/13/4/009

[44] M. Talagrand, "The Parisi formula," *Annals of Mathematics*, vol. 163, no. 1, pp. 221–263, 2006, ISSN: 0003486X. Accessed: Nov. 20, 2023. [Online]. Available: http://www.jstor.org/stable/20159953

[45] D. Panchenko, "The Parisi formula for mixed p-spin models," *The Annals of Probability*, vol. 42, no. 3, May 2014, ISSN: 0091-1798. DOI: 10.1214/12-aop800 [Online]. Available: http://dx.doi.org/10.1214/12-AOP800

[46] A. Auffinger and W.-K. Chen, "Parisi formula for the ground state energy in the mixed p-spin model," *The Annals of Probability*, vol. 45, no. 6B, pp. 4617–4631, 2017. DOI: 10.1214/16-AOP1173 [Online]. Available: https://doi.org/10.1214/16-AOP1173

[47] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond*. WORLD SCIENTIFIC, 1986. DOI: 10.1142/0271 eprint: https://www.worldscientific.com/doi/pdf/10.1142/0271. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/0271

[48] W.-K. Chen, D. Gamarnik, D. Panchenko, and M. Rahman, "Suboptimality of local algorithms for a class of max-cut problems," *The Annals of Probability*, vol. 47, no. 3, pp. 1587–1618, May 2019, ISSN: 0091-1798. DOI: 10.1214/18-aop1291 [Online]. Available: http://dx.doi.org/10.1214/18-AOP1291

[49] F. Guerra, "Sum rules for the free energy in mean fields spin glass models," *Fields Institute Communications*, vol. 30, Nov. 2001. DOI: 10.1090/fic/030/10

[50] F. Guerra and F. L. Toninelli, "Some comments on the connection between disordered long range spin glass models and their mean field version," *Journal of Physics A: Mathematical and General*, vol. 36, no. 43, p. 10 987, Oct. 2003. DOI: 10.1088/0305-4470/36/43/022 [Online]. Available: https://dx.doi.org/10.1088/0305-4470/36/43/022

[51] F. Guerra, "Broken replica symmetry bounds in the mean field spin glass model," *Communications in Mathematical Physics*, vol. 233, no. 1, pp. 1–12, Feb. 2003, ISSN: 1432-0916. DOI: 10.1007/s00220-002-0773-5 [Online]. Available: http://dx.doi.org/10.1007/s00220-002-0773-5

[52] F. Guerra and F. L. Toninelli, "The high temperature region of the viana–bray diluted spin glass model," *Journal of Statistical Physics*, vol. 115, no. 1/2, pp. 531–555, Apr. 2004, ISSN: 0022-4715. DOI: 10.1023/b:joss.0000019815.11115.54 [Online]. Available: http://dx.doi.org/10.1023/B:JOSS.0000019815.11115.54

[53] A. Dembo, A. Montanari, and S. Sen, "Extremal cuts of sparse random graphs," *The Annals of Probability*, vol. 45, no. 2, Mar. 2017, ISSN: 0091-1798. DOI: 10.1214/15-aop1084 [Online]. Available: http://dx.doi.org/10.1214/15-AOP1084

[54] S. Sen, "Optimization on sparse random hypergraphs and spin glasses," *Random Structures & Algorithms*, vol. 53, no. 3, pp. 504–536, 2018. DOI: https://doi.org/10.1002/rsa.20774 eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.20774. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20774

[55] C. Jones, K. Marwaha, J. S. Sandhu, and J. Shi, "Random max-csps inherit algorithmic hardness from spin glasses," en, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. DOI: 10.4230/LIPICS.ITCS.2023.77 [Online]. Available: https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2023.77

[56] R. O'Donnell, "Boolean functions and the fourier expansion," in *Analysis of Boolean Functions*. Cambridge University Press, 2014, pp. 1–25.

[57] D. Gamarnik and Q. Li, "Finding a large submatrix of a Gaussian random matrix," *The Annals of Statistics*, vol. 46, no. 6A, pp. 2511–2561, 2018. DOI: 10.1214/17-AOS1628 [Online]. Available: https://doi.org/10.1214/17-AOS1628

[58] D. Achlioptas and F. 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*, ser. STOC '06, Seattle, WA, USA: Association for Computing Machinery, 2006, pp. 130–139, ISBN: 1595931341. DOI: 10.1145/1132516.1132537 [Online]. Available: https://doi.org/10.1145/1132516.1132537

[59] M. Mézard, T. Mora, and R. Zecchina, "Clustering of solutions in the random satisfiability problem," *Physical Review Letters*, vol. 94, no. 19, May 2005, ISSN: 1079-7114. DOI: 10.1103/physrevlett.94.197205 [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.94.197205

[60] D. Gamarnik and M. Sudan, "Limits of local algorithms over sparse random graphs," *The Annals of Probability*, vol. 45, no. 4, pp. 2353–2376, 2017. DOI: 10.1214/16-AOP1114 [Online]. Available: https://doi.org/10.1214/16-AOP1114

[61] M. Rahman and B. Virág, "Local algorithms for independent sets are half-optimal," *The Annals of Probability*, vol. 45, no. 3, May 2017, ISSN: 0091-1798. DOI: 10.1214/16-aop1094 [Online]. Available: http://dx.doi.org/10.1214/16-AOP1094

[62] A. S. Wein, "Optimal low-degree hardness of maximum independent set," *ArXiv*, vol. abs/2010.06563, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:222310479

[63] D. Gamarnik, A. Jagannath, and A. S. Wein, "Hardness of random optimization problems for boolean circuits, low-degree polynomials, and langevin dynamics," *SIAM Journal on Computing*, vol. 53, no. 1, pp. 1–46, 2024. DOI: 10.1137/22M150263X eprint: https://doi.org/10.1137/22M150263X. [Online]. Available: https://doi.org/10.1137/22M150263X

[64] D. Gamarnik and A. Jagannath, "The overlap gap property and approximate message passing algorithms for p-spin models," *The Annals of Probability*, vol. 49, no. 1, pp. 180–205, 2021. DOI: 10.1214/20-AOP1448 [Online]. Available: https://doi.org/10.1214/20-AOP1448

[65] E. Farhi, D. Gamarnik, and S. Gutmann, *The quantum approximate optimization algorithm needs to see the whole graph: A typical case*, 2020. arXiv: 2004.09002 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2004.09002

[66] J. Basso, D. Gamarnik, S. Mei, and L. 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, Oct. 2022, pp. 335–343. DOI: 10.1109/focs54457.2022.00039 [Online]. Available: http://dx.doi.org/10.1109/FOCS54457.2022.00039

[67] A. Anshu and T. Metger, "Concentration bounds for quantum states and limitations on the QAOA from polynomial approximations," *Quantum*, vol. 7, p. 999, May 2023, ISSN: 2521-327X. DOI: 10.22331/q-2023-05-11-999 [Online]. Available: http://dx.doi.org/10.22331/q-2023-05-11-999

[68] S. Li and T. Schramm, *Some easy optimization problems have the overlap-gap property*, 2024. arXiv: 2411.01836 [cs.CC]. [Online]. Available: https://arxiv.org/abs/2411.01836

[69] B. Huang and M. Sellke, *Strong low degree hardness for stable local optima in spin glasses*, 2025. arXiv: 2501.06427 [cond-mat.dis-nn]. [Online]. Available: https://arxiv.org/abs/2501.06427

[70] F. Koehler and J. Shin, *Overlap analysis of the shortest path problem: Local search, landscapes, and franz–parisi potential*, 2025. arXiv: 2511.18666 [cs.DS]. [Online]. Available: https://arxiv.org/abs/2511.18666

[71] E. Farhi, J. Goldstone, and S. Gutmann, *A quantum approximate optimization algorithm*, 2014. arXiv: 1411.4028 [quant-ph]. [Online]. Available: https://arxiv.org/abs/1411.4028

[72] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, "Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices," *Phys. Rev. X*, vol. 10, p. 021 067, 2 Jun. 2020. DOI: 10.1103/PhysRevX.10.021067 [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevX.10.021067

[73] A. Apte et al., *Iterative interpolation schedules for quantum approximate optimization algorithm*, 2025. arXiv: 2504.01694 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2504.01694

[74] M. Streif, S. Yarkoni, A. Skolik, F. Neukart, and M. Leib, "Beating classical heuristics for the binary paint shop problem with the quantum approximate optimization algorithm," *Phys. Rev. A*, vol. 104, p. 012 403, 1 Jul. 2021. DOI: 10.1103/PhysRevA.104.012403 [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.104.012403

[75] E. Farhi, J. Goldstone, S. Gutmann, and L. Zhou, "The quantum approximate optimization algorithm and the sherrington-kirkpatrick model at infinite size," *Quantum*, vol. 6, p. 759, Jul. 2022, ISSN: 2521-327X. DOI: 10.22331/q-2022-07-07-759 [Online]. Available: http://dx.doi.org/10.22331/q-2022-07-07-759

[76] L. Binkowski, G. Koßmann, T. Ziegler, and R. Schwonnek, "Elementary proof of QAOA convergence," *New Journal of Physics*, vol. 26, no. 7, p. 073 001, Jul. 2024, ISSN: 1367-2630. DOI: 10.1088/1367-2630/ad59bb [Online]. Available: http://dx.doi.org/10.1088/1367-2630/ad59bb

[77] A. Misra-Spieldenner, T. Bode, P. K. Schuhmacher, T. Stollenwerk, D. Bagrets, and F. K. Wilhelm, "Mean-field approximate optimization algorithm," *PRX Quantum*, vol. 4, no. 3, Sep. 2023, ISSN: 2691-3399. DOI: 10.1103/prxquantum.4.030335 [Online]. Available: http://dx.doi.org/10.1103/PRXQuantum.4.030335

[78] R. Ghosh, L. A. Nutricati, N. Feinstein, P. A. Warburton, and S. Bose, *Exponential speed-up of quantum annealing via n-local catalysts*, 2024. arXiv: 2409.13029 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2409.13029

[79] M. Dietzfelbinger, A. Goerdt, M. Mitzenmacher, A. Montanari, R. Pagh, and M. Rink, "Tight thresholds for cuckoo hashing via xorsat," in *Automata, Languages and Programming*, S. Abramsky, C. Gavoille, C. Kirchner, F. Meyer auf der Heide, and P. G. Spirakis, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 213–225, ISBN: 978-3-642-14165-2.

[80] J. Basso, E. Farhi, K. Marwaha, B. Villalonga, and L. Zhou, "The quantum approximate optimization algorithm at high depth for maxcut on large-girth regular graphs and the sherrington-kirkpatrick model," en, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. DOI: 10.4230/LIPICS.TQC.2022.7 [Online]. Available: https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.TQC.2022.7

[81] E. Farhi, S. Gutmann, D. Ranard, and B. Villalonga, *Lower bounding the maxcut of high girth 3-regular graphs using the qaoa*, 2025. arXiv: 2503.12789 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2503.12789

[82] E. Farhi, D. Gamarnik, and S. Gutmann, *The quantum approximate optimization algorithm needs to see the whole graph: Worst case examples*, 2020. arXiv: 2005.08747 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2005.08747

[83] A. Frieze and M. Karoński, *Introduction to Random Graphs*. Cambridge University Press, 2015. DOI: 10.1017/CBO9781316339831

[84] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo, "Diagnosing barren plateaus with tools from quantum optimal control," *Quantum*, vol. 6, p. 824, Sep. 2022, ISSN: 2521-327X. DOI: 10.22331/q-2022-09-29-824 [Online]. Available: http://dx.doi.org/10.22331/q-2022-09-29-824

[85] M. Larocca et al., "Barren plateaus in variational quantum computing," *Nature Reviews Physics*, vol. 7, no. 4, pp. 174–189, Mar. 2025, ISSN: 2522-5820. DOI: 10.1038/s42254-025-00813-9 [Online]. Available: http://dx.doi.org/10.1038/s42254-025-00813-9

[86] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," *Nature Communications*, vol. 9, no. 1, Nov. 2018, ISSN: 2041-1723. DOI: 10.1038/s41467-018-07090-4 [Online]. Available: http://dx.doi.org/10.1038/s41467-018-07090-4

[87] R. Shaydulin et al., "Evidence of scaling advantage for the quantum approximate optimization algorithm on a classically intractable problem," *Science Advances*, vol. 10, no. 22, May 2024, ISSN: 2375-2548. DOI: 10.1126/sciadv.adm6761 [Online]. Available: http://dx.doi.org/10.1126/sciadv.adm6761

[88] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, *Quantum computation by adiabatic evolution*, 2000. arXiv: quant-ph/0001106 [quant-ph]. [Online]. Available: https://arxiv.org/abs/quant-ph/0001106

[89] C. G. BROYDEN, "The convergence of a class of double-rank minimization algorithms 1. general considerations," *IMA Journal of Applied Mathematics*, vol. 6, no. 1, pp. 76–90, Mar. 1970, ISSN: 0272-4960. DOI: 10.1093/imamat/6.1.76 eprint: https://academic.oup.com/imamat/article-pdf/6/1/76/2233756/6-1-76.pdf. [Online]. Available: https://doi.org/10.1093/imamat/6.1.76

[90] R. Fletcher, "A new approach to variable metric algorithms," *The Computer Journal*, vol. 13, no. 3, pp. 317–322, Jan. 1970, ISSN: 0010-4620. DOI: 10.1093/comjnl/13.3.317 eprint: https://academic.oup.com/comjnl/article-pdf/13/3/317/988678/130317.pdf. [Online]. Available: https://doi.org/10.1093/comjnl/13.3.317

[91] D. Goldfarb, "A family of variable-metric methods derived by variational means," *Mathematics of Computation*, vol. 24, no. 109, pp. 23–26, 1970, ISSN: 00255718, 10886842. [Online]. Available: http://www.jstor.org/stable/2004873

[92] D. F. Shanno, "Conditioning of quasi-newton methods for function minimization," *Mathematics of Computation*, vol. 24, no. 111, pp. 647–656, 1970, ISSN: 00255718, 10886842. [Online]. Available: http://www.jstor.org/stable/2004840

[93] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009, ISSN: 1091-6490. DOI: 10.1073/pnas.0909892106 [Online]. Available: http://dx.doi.org/10.1073/pnas.0909892106

[94] D. J. Thouless, P. W. Anderson, and R. G. Palmer, "Solution of 'solvable model of a spin glass'," *The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics*, vol. 35, no. 3, pp. 593–601, 1977. DOI: 10.1080/14786437708235992 eprint: https://doi.org/10.1080/

14786437708235992. [Online]. Available: https://doi.org/10.1080/14786437708235992

[95] Q. Zou and H. Yang, *A concise tutorial on approximate message passing*, 2022. arXiv: 2201.07487 [cs.IT]. [Online]. Available: https://arxiv.org/abs/2201.07487

[96] A. E. Alaoui, A. Montanari, and M. Sellke, "Optimization of mean-field spin glasses," *The Annals of Probability*, vol. 49, no. 6, pp. 2922–2960, 2021. DOI: 10.1214/21-AOP1519 [Online]. Available: https://doi.org/10.1214/21-AOP1519

[97] A. Auffinger, A. Montanari, and E. Subag, "Optimization of random high-dimensional functions: Structure and algorithms," in *Spin Glass Theory and Far Beyond*. WORLD SCIENTIFIC, 2023, ch. Chapter 29, pp. 609–633. DOI: 10.1142/9789811273926_0029 eprint: https://www.worldscientific.com/doi/pdf/10.1142/9789811273926_0029. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/9789811273926_0029

[98] S. Boulebnane and A. Montanaro, "Solving boolean satisfiability problems with the quantum approximate optimization algorithm," *PRX Quantum*, vol. 5, p. 030 348, 3 Sep. 2024. DOI: 10.1103/PRXQuantum.5.030348 [Online]. Available: https://link.aps.org/doi/10.1103/PRXQuantum.5.030348

[99] P. C. Lotshaw et al., "Approximate boltzmann distributions in quantum approximate optimization," *Phys. Rev. A*, vol. 108, p. 042 411, 4 Oct. 2023. DOI: 10.1103/PhysRevA.108.042411 [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.108.042411

[100] V. Akshay et al., "Circuit depth scaling for quantum approximate optimization," *Phys. Rev. A*, vol. 106, p. 042 438, 4 Oct. 2022. DOI: 10.1103/PhysRevA.106.042438 [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.106.042438

[101] S. Boulebnane et al., *Evidence that the quantum approximate optimization algorithm optimizes the sherrington-kirkpatrick model efficiently in the average case*, 2025. arXiv: 2505.07929 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2505.07929

[102] I. Čepaitė, N. Vaishnav, L. Zhou, and A. Montanaro, *Quantum-enhanced optimization by warm starts*, 2025. arXiv: 2508.16309 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2508.16309

[103] S. M. Perez-Garcia and A. Montanaro, *Quantum-enhanced belief propagation for ldpc decoding*, 2024. arXiv: 2412.08596 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2412.08596

[104] T. Epping, W. Hochstättler, and P. Oertel, "Complexity results on a paint shop problem," *Discrete Applied Mathematics*, vol. 136, no. 2, pp. 217–226, 2004, The 1st Cologne-Twente Workshop on Graphs and Combinatorial Optimization, ISSN: 0166-218X. DOI: https://doi.org/10.1016/S0166-218X(03)00442-6 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166218X03004426

[105] P. Bonsma, T. Epping, and W. Hochstättler, "Complexity results on restricted instances of a paint shop problem for words," *Discrete Applied Mathematics*, vol. 154, no. 9, pp. 1335–1343, 2006, 2nd Cologne/Twente Workshop on Graphs and Combinatorial Optimization (CTW 2003), ISSN: 0166-218X. DOI: https://doi.org/10.1016/j.dam.2005.05.033 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166218X0500377X

[106] S. D. Andres and W. Hochstättler, "Some heuristics for the binary paint shop problem and their expected number of colour changes," *Journal of Discrete Algorithms*, vol. 9, no. 2, pp. 203–211, 2011, ISSN: 1570-8667. DOI: https://doi.org/10.1016/j.jda.2010.12.003 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570866710000559

[107] F. Meunier and B. Neveu, "Computing solutions of the paintshop–necklace problem," *Computers & Operations Research*, vol. 39, no. 11, pp. 2666–2678, 2012, ISSN: 0305-0548. DOI: https://doi.org/10.1016/j.cor.2012.01.014 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0305054812000263

[108] N. Alon, D. Elboim, J. Pach, and G. Tardos, "Random necklaces require fewer cuts," *SIAM Journal on Discrete Mathematics*, vol. 38, no. 2, pp. 1381–1408, 2024. DOI: 10.1137/22M1506699 eprint: https://doi.org/10.1137/22M1506699. [Online]. Available: https://doi.org/10.1137/22M1506699

[109] J. Hančl, A. Kabela, M. Opler, J. Sosnovec, R. Šámal, and P. Valtr, "Improved bounds for the binary paint shop problem," in *Computing and Combinatorics*, W. Wu and G. Tong, Eds., Cham: Springer Nature Switzerland, 2024, pp. 210–221, ISBN: 978-3-031-49193-1.

[110] G. J. Mooney, J. Villanueva, B. R. Radhan, J. Ghosh, C. D. Hill, and L. C. L. Hollenberg, *An optimization-free recursive qaoa for the binary paint shop problem*, 2025. arXiv: 2507.10908 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2507.10908

[111] V. Vijendran, D. E. Koh, P. K. Lam, and S. M. Assad, *Classical and quantum heuristics for the binary paint shop problem*, 2025. arXiv: 2509.15294 [quant-ph]. [Online]. Available: https://arxiv.org/abs/2509.15294

[112] A. Misra-Spieldenner, T. Bode, P. K. Schuhmacher, T. Stollenwerk, D. Bagrets, and F. K. Wilhelm. "Mean-field approximate optimization algorithm github repository. "[Online]. Available: https://github.com/FZJ-PGI-12/Mean-Field-Approximate-Optimization-Algorithm

[113] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell, "Optimal inapproximability results for max-cut and other 2-variable csps?" *SIAM Journal on Computing*, vol. 37, no. 1, pp. 319–357, 2007. DOI: 10.1137/S0097539705447372 eprint: https://doi.org/10.1137/S0097539705447372. [Online]. Available: https://doi.org/10.1137/S0097539705447372

[114] A. El Alaoui, A. Montanari, and M. Sellke, "Local algorithms for maximum cut and minimum bisection on locally treelike regular graphs of large degree," *Random Structures & Algorithms*, vol. 63, no. 3, pp. 689–715, 2023. DOI: https://doi.org/10.1002/rsa.21149 eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.21149. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.21149

[115] A. Coja-Oghlan, P. Loick, B. F. Mezei, and G. B. Sorkin, "The ising antiferromagnet and max cut on random regular graphs," *SIAM Journal on Discrete Mathematics*, vol. 36, no. 2, pp. 1306–1342, 2022. DOI: 10.1137/20M137999X eprint: https://doi.org/10.1137/20M137999X. [Online]. Available: https://doi.org/10.1137/20M137999X

[116] D. Gamarnik and I. Zadik, "The landscape of the planted clique problem: Dense subgraphs and the overlap gap property," *The Annals of Applied Probability*, vol. 34, no. 4, pp. 3375–3434, 2024. DOI: 10.1214/23-AAP2003 [Online]. Available: https://doi.org/10.1214/23-AAP2003

[117] A. Dudek, A. Frieze, A. Ruciński, and M. Šileikis, "Embedding the erdős–rényi hypergraph into the random regular hypergraph and hamiltonicity," *Journal of Combinatorial Theory, Series B*, vol. 122, pp. 719–740, Jan. 2017, ISSN: 0095-8956. DOI: 10.1016/j.jctb.2016.09.003 [Online]. Available: http://dx.doi.org/10.1016/j.jctb.2016.09.003

[118] D. Ellis and N. Linial, "On regular hypergraphs of high girth," *Electron. J. Comb.*, vol. 21, p. 1, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:961773

[119] D. J. Poole, "On the strength of connectedness of a random hypergraph," *Electron. J. Comb.*, vol. 22, no. 1, p. 1, 2015. DOI: 10.37236/4666 [Online]. Available: https://doi.org/10.37236/4666

[120] E. Wybo and M. Leib, "Missing Puzzle Pieces in the Performance Landscape of the Quantum Approximate Optimization Algorithm," *Quantum*, vol. 9, p. 1892, Oct. 2025, ISSN: 2521-327X. DOI: 10.22331/q-2025-10-22-1892 [Online]. Available: https://doi.org/10.22331/q-2025-10-22-1892