

Near-Optimal Parameter Tuning of Level-1 QAOA for Ising Models

V. Vijendran^{1,2*}, Dax Enshan Koh^{1,3,4}, Eunok Bae^{5,6}, Hyukjoon Kwon⁵, Ping Koy Lam^{1,2}, and Syed M. Assad^{1,2}

¹ A*STAR Quantum Innovation Centre (Q.InC), Singapore

² Centre for Quantum Technologies, National University of Singapore, Singapore

³ A*STAR Institute of High Performance Computing, Singapore

⁴ Singapore University of Technology and Design, Singapore

⁵ Korea Institute for Advanced Study, Korea

⁶ Electronics and Telecommunications Research Institute, Korea

Abstract. The Quantum Approximate Optimisation Algorithm (QAOA) tackles combinatorial optimisation problems by encoding their solutions into the ground state of an Ising Hamiltonian prepared by a p -level parameterised circuit, with the angles tuned classically. Parameter optimisation is widely regarded as a central bottleneck, even for the shallowest circuits. Focusing on QAOA at $p = 1$ (QAOA₁), we show that tuning the two angles (γ, β) for weighted Ising models is not a black-box search but a structured signal-processing problem. We prove that the QAOA₁ expectation value is a partial Fourier series in γ whose frequencies are determined explicitly by the problem’s couplings and fields, giving instance-wise bandwidth bounds and, via the Nyquist–Shannon theorem, the sampling resolution needed to avoid the aliasing that causes coarse-grid searches to return spurious optima. We then eliminate the mixer angle analytically, computing $\beta^*(\gamma)$ in closed form to reduce the search to one dimension, and apply a subdivision algorithm that locates the globally optimal γ in polynomial time with a certificate of optimality when the weights are commensurable and bounded. For regular weighted graphs, we further prove the conventional wisdom that the globally optimal $\gamma^* \in \mathbb{R}^+$ concentrates near zero and coincides with the first local optimum, giving a rigorous account of the empirical success of small-angle initialisation and allowing gradient descent to replace exhaustive line searches. Validated within Recursive QAOA (RQAOA) on weighted instances of 128 and 256 qubits, our method consistently outperforms both coarsely optimised RQAOA and semidefinite programming.

Keywords: QAOA · Quantum optimization · Ising models · QUBO problems · Parameter tuning · Signal processing · Quantum information

1 Introduction

Training variational quantum algorithms is NP-hard in the worst case [1]. This is sometimes interpreted as evidence that reliable parameter optimization is fundamentally out of reach even for shallow QAOA [4]. We argue that this is not the right conclusion for the practically important regime of bounded-precision weighted Ising and QUBO instances. Worst-case hardness constructions rely on adversarial structure, such as exponentially separated or highly incommensurable coefficients, which forces the objective landscape to contain extremely fine features. In contrast, optimization instances arising in real-world applications usually have finite precision and moderate weight scales. This motivates a sharper question: can QAOA₁ be tuned globally, efficiently, and with a certificate of optimality in the practically relevant weighted regime?

Before answering it, we address why the question matters. QAOA₁ is not, by itself, competitive with state-of-the-art classical algorithms, and this is sometimes taken to mean that its optimization is of limited consequence. This overlooks how QAOA₁ is actually deployed: as the quantum subroutine inside non-local algorithms such as Recursive QAOA (RQAOA) [3] and iterative rounding variants [5, 2], which repeatedly query QAOA₁ correlations to fix variables and shrink the problem. These algorithms can match and surpass the Goemans–Williamson semidefinite program (SDP) [6]—but, as our benchmarks show, *only* when the QAOA₁ angles at every recursive step are tuned to near-optimality; with coarse tuning, the very same algorithms lose to the SDP. Optimal QAOA₁ tuning is therefore not a toy problem: it is the enabling ingredient that decides whether an entire family of quantum heuristics beats the classical state of the art.

We answer this positively by showing that depth-one QAOA, despite having only two variational angles γ and β , is not best understood as a generic two-dimensional black-box optimization problem. Inspired by the Fourier analysis of data encoding in variational quantum machine-learning models by Schuld et al. [9], we show that the QAOA₁ expectation value is a bandlimited Fourier signal whose frequencies are determined by the Ising coefficients. Once this bandwidth is known, the correct sampling scale follows from a Nyquist-type criterion rather than heuristic grid choice. We then analytically eliminate β , reducing the two-angle problem to a one-dimensional global optimization over γ with a certificate of optimality.

2 Main Contributions

Starting from an arbitrary weighted Ising/QUBO instance, we expose the Fourier structure of the QAOA₁ landscape, reduce the two-angle problem analytically, and solve the remaining global optimization problem with certified resolution. Our contributions are:

* Corresponding authors: vjntm@gmail.com and cqtsma@gmail.com

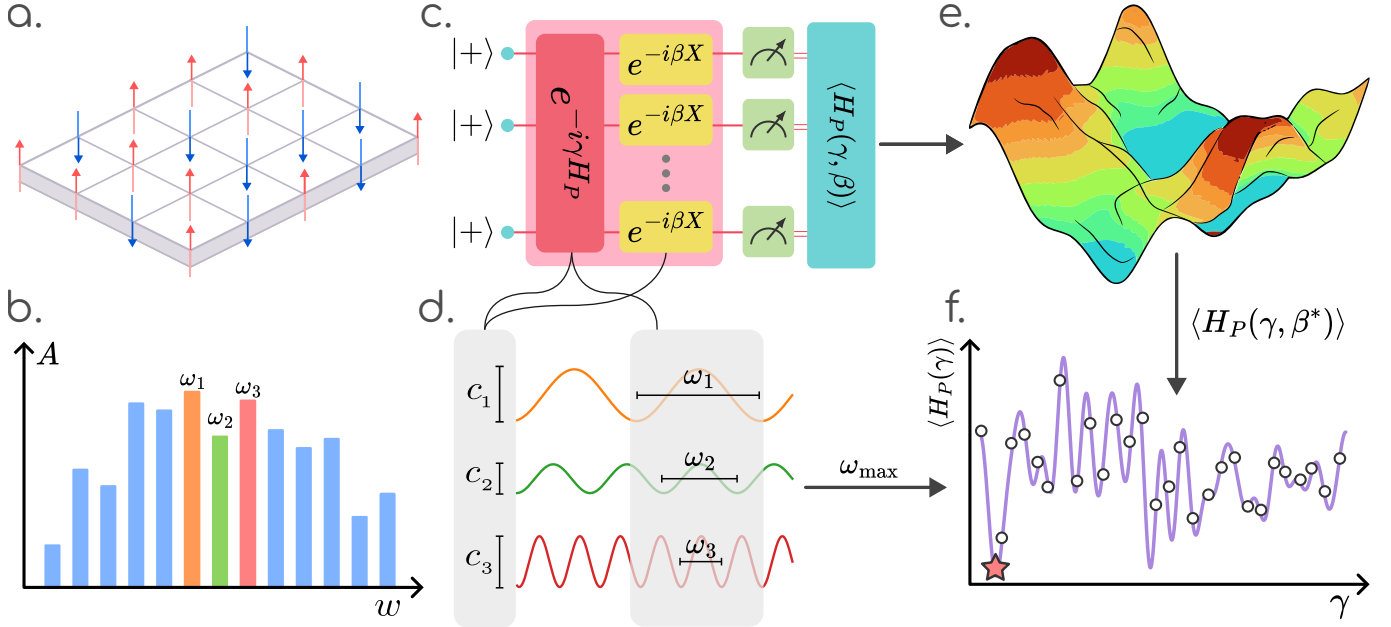


Fig. 1: **Main idea.** A weighted Ising/QUBO instance defines a problem Hamiltonian whose coefficients determine the accessible Fourier frequencies of the QAOA₁ expectation landscape. We compute an instance-wise bandwidth, use it to set a Nyquist-safe sampling scale, analytically eliminate β , and optimize the remaining one-dimensional landscape with a certificate of global optimality.

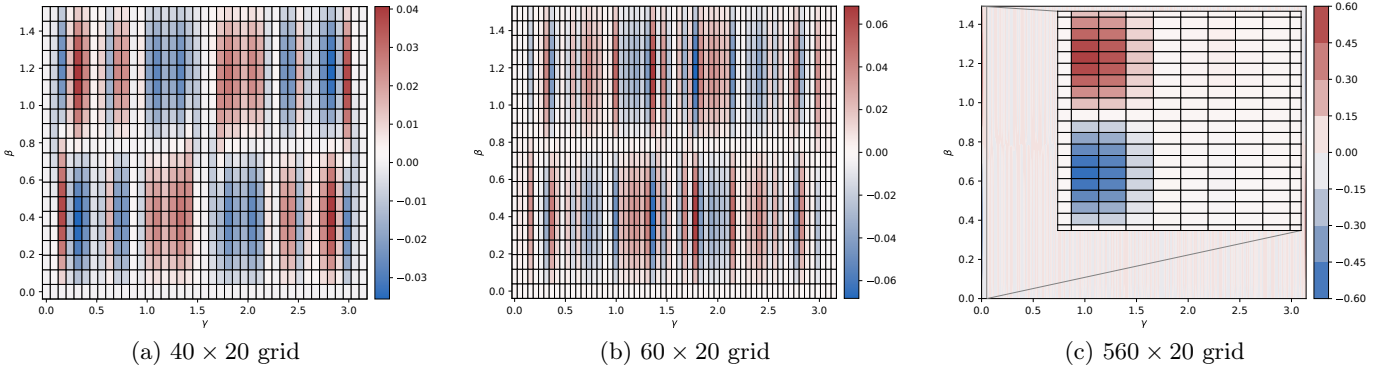


Fig. 2: Cost landscapes of QAOA₁ for a weighted regular graph under varying grid resolutions. Coarse sampling can give approximation ratios close to random guessing; only a sufficiently fine resolution resolves the true global optimum.

1. **Fourier view of QAOA₁:** We reinterpret parameter tuning for weighted Ising/QUBO models as a bandlimited Fourier optimization problem rather than a black-box search over two angles.
2. **Instance-wise bandwidth bounds:** We derive closed-form bounds on the maximum relevant γ -frequency from the problem coefficients, giving a principled sampling resolution for resolving the landscape and avoiding aliasing.
3. **Analytic elimination of β :** For each γ , we compute $\beta^*(\gamma)$ directly, reducing the original optimization over (γ, β) to a certified one-dimensional search.
4. **Certified global optimization:** We combine the reduction with a subdivision algorithm for univariate trigonometric polynomials [7], obtaining an ϵ -optimal γ with a certificate of global optimality using function evaluations linear in the Fourier bandwidth for fixed precision.
5. **Theory of small-angle initialization:** For large regular weighted instances, we prove that the global optimum concentrates near $\gamma \approx 0$ and coincides with the first local optimum on average. Our result holds for the *entire* family of regular weighted graphs—from triangle-free to complete, parameterised by the triangle density of each edge—and for Ising models *with* external fields.

Together, these results identify the relevant obstruction to QAOA₁ training: not the apparent two-dimensionality of the parameter space, but the unresolved Fourier bandwidth hidden in the weighted landscape.

3 Frequency-Certified Tuning

Figure 2 illustrates the key failure mode. The poor performance of coarse grids is not a mysterious optimizer pathology; it is an aliasing problem. The QAOA₁ objective has a finite but instance-dependent frequency spectrum. If the sampling rate is below twice the largest relevant frequency, a grid search does not resolve the function being optimized. The remedy is therefore not merely to try more optimizer restarts, but to compute the required sampling scale from the Hamiltonian itself.

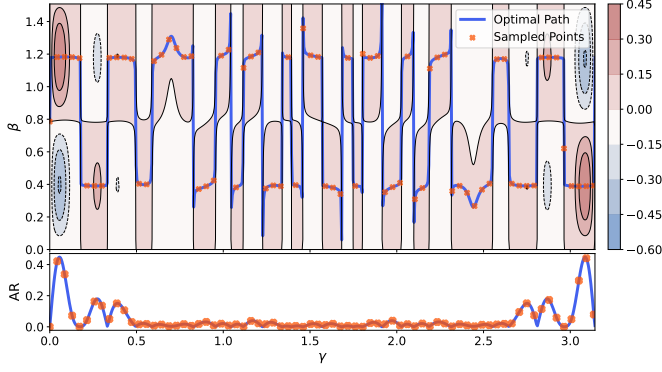


Fig. 3: Contour plot of QAOA₁'s cost landscape for a weighted graph, showing approximation ratios. The optimal univariate path (blue line) and sampled points (orange markers) are overlaid. The bottom plot shows the corresponding cost function as a function of γ .

After this bandwidth is known, the remaining question is how to locate the optimum without an expensive two-dimensional search. We use closed-form expressions for the QAOA₁ expectation value [8, 13, 11] to compute $\beta^*(\gamma)$ analytically. This maps the original landscape onto an optimal path parameterised only by γ , as shown in fig. 3. The resulting one-dimensional objective is then handled by a subdivision algorithm for trigonometric polynomials [7]. At each iteration, candidate intervals are evaluated and intervals whose Fourier upper bound is already below the best certified value are discarded. Thus, the method returns not only a good angle, but a certificate that no discarded interval contains a better one.

4 Relation to Prior Work and Validation

Among the large literature on QAOA parameter setting, the closest recent comparison is Sureshbabu et al. [10], who study weighted QAOA parameter setting, with rigorous results for triangle-free weighted MaxCut. At $p = 1$, they derive globally optimal parameters under random-weight assumptions and formalize the average-case intuition that the first local optimum near $\gamma = 0$ is globally optimal. For general p , they show that, on large-girth regular graphs with i.i.d. weights, the weighted-MaxCut landscape converges in expectation to a rescaled unweighted landscape, justifying parameter transfer from unweighted MaxCut.

Our framework addresses a different problem: instance-wise certified optimization for arbitrary weighted Ising/QUBO Hamiltonians at QAOA₁. We do not assume MaxCut structure, large girth, regularity, i.i.d. weights, an infinite-size limit, or transferability from an unweighted instance. For the actual input Hamiltonian, we compute the accessible QAOA₁ frequency content, determine the global resolution scale, analytically eliminate β , and certify the optimum of the reduced one-dimensional problem. The output is therefore not a predicted or transferred parameter setting, but a certificate-producing global tuning procedure for the given instance. This is also distinct from parameter-shift and gradient-estimation methods, which estimate local derivatives but do not determine the global Fourier bandwidth, prevent aliasing, or certify global optimality.

As a practical validation, we embed our tuning routine into RQAOA [3], tuning the QAOA₁ angles at each recursive step on weighted Erdős-Rényi Ising instances with 128 and 256 qubits. As shown in fig. 4, accurate tuning clearly improves over coarse-grid tuning and enables RQAOA₁ to outperform the SDP baseline on these instances. In subsequent work on the Binary Paint Shop Problem [12], a structured application problem, we deployed the same tuning pipeline within RQAOA₁ on instances of up to 4096 qubits, demonstrating that the method scales far beyond the regime benchmarked here and applies beyond random ensembles. Beyond depth one, the same Fourier viewpoint suggests a route to certified tuning of deeper and noisy QAOA. For QAOA_p, the objective becomes a multivariate trigonometric polynomial in the $2p$ angles, and closed-form bandwidth bounds are generally unavailable. Extending our framework therefore requires tools for estimating multi-angle bandwidths and subdivision algorithms for multivariate trigonometric polynomials, with noise-robust certification needed for hardware implementations.

In summary, we replace heuristic QAOA₁ parameter search with a principled, instance-wise, frequency-certified optimization framework. The analysis clarifies when worst-case hardness is driven by pathological numerical structure, explains why coarse tuning fails on weighted instances, and gives a certified route to globally optimal depth-one QAOA parameters in the practically relevant bounded-weight regime.

References

1. Bittel, L., Kliesch, M.: Training variational quantum algorithms is NP-Hard. *Physical Review Letters* **127**(12), 120502 (2021). <https://doi.org/10.1103/PhysRevLett.127.120502>
2. Brady, L.T., Hadfield, S.: Iterative quantum algorithms for maximum independent set. *Physical Review A* **110**, 052435 (2024). <https://doi.org/10.1103/PhysRevA.110.052435>

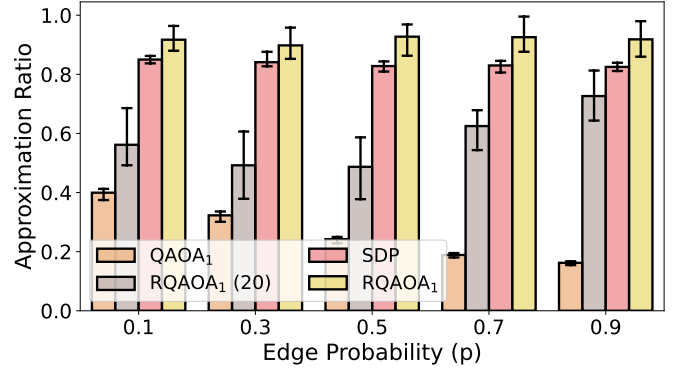


Fig. 4: Benchmark on 128-qubit Erdős-Rényi instances comparing QAOA₁, coarsely tuned RQAOA₁ (20), and optimally tuned RQAOA₁. While coarse tuning improves over QAOA₁, only optimal tuning enables RQAOA₁ to consistently outperform the SDP baseline.

3. Bravyi, S., Kliesch, A., Koenig, R., Tang, E.: Obstacles to variational quantum optimization from symmetry protection. *Physical Review Letters* **125**(26), 260505 (2020). <https://doi.org/10.1103/PhysRevLett.125.260505>
4. Farhi, E., Goldstone, J., Gutmann, S.: A quantum approximate optimization algorithm. arXiv preprint arXiv:1411.4028 (2014), <https://doi.org/10.48550/arXiv.1411.4028>
5. Finžgar, J.R., Kerschbaumer, A., Schuetz, M.J.A., Mendl, C.B., Katzgraber, H.G.: Quantum-informed recursive optimization algorithms. *PRX Quantum* **5**, 020327 (2024). <https://doi.org/10.1103/PRXQuantum.5.020327>
6. Goemans, M.X., Williamson, D.P.: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM* **42**(6), 1115–1145 (1995). <https://doi.org/10.1145/227683.227684>
7. Green, J.J.: Calculating the maximum modulus of a polynomial using Steckin’s lemma. *SIAM Journal on Numerical Analysis* **36**(4), 1022–1029 (1999). <https://doi.org/10.1137/S0036142997331335>, <https://doi.org/10.1137/S0036142997331335>
8. Hadfield, S.A.: Quantum algorithms for scientific computing and approximate optimization. Ph.D. thesis, Columbia University (2018). <https://doi.org/10.7916/D8X650C9>, <https://doi.org/10.7916/D8X650C9>
9. Schuld, M., Sweke, R., Meyer, J.J.: Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A* **103**(3), 032430 (2021)
10. Sureshbabu, S.H., Herman, D., Shaydulin, R., Basso, J., Chakrabarti, S., Sun, Y., Pistoia, M.: Parameter setting in quantum approximate optimization of weighted problems. *Quantum* **8**, 1231 (2024). <https://doi.org/10.22331/q-2024-01-18-1231>
11. Vijendran, V., Das, A., Koh, D.E., Assad, S.M., Lam, P.K.: An expressive ansatz for low-depth quantum approximate optimisation. *Quantum Science and Technology* **9**(2), 025010 (2024). <https://doi.org/10.1088/2058-9565/ad200a>
12. Vijendran, V., Koh, D.E., Lam, P.K., Assad, S.M.: Classical and quantum heuristics for the binary paint shop problem (2025)
13. Wang, Z., Hadfield, S., Jiang, Z., Rieffel, E.G.: Quantum approximate optimization algorithm for MaxCut: A fermionic view. *Physical Review A* **97**(2), 022304 (2018). <https://doi.org/10.1103/PhysRevA.97.022304>

Near-Optimal Parameter Tuning of Level-1 QAOA for Ising Models

V Vijendran^{1,2,3}, Dax Enshan Koh^{2,4,5,6}, Eunok Bae^{7,8}, Hyukjoon Kwon⁷,
Ping Koy Lam^{2,1,3}, and Syed M Assad^{2,3}

¹Centre for Quantum Technologies (CQT), National University of Singapore, Singapore 117543, Republic of Singapore

²Quantum Innovation Centre (Q.InC), Agency for Science, Technology and Research (A*STAR), 2 Fusionopolis Way, Innovis #08-03, Singapore 138634, Republic of Singapore

³Centre for Quantum Computation and Communication Technologies (CQC2T), Department of Quantum Science and Technology, Research School of Physics, Australian National University, Acton 2601, Australia

⁴Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Republic of Singapore

⁵Science, Mathematics and Technology Cluster, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372, Republic of Singapore

⁶Engineering Cluster, Singapore Institute of Technology, 1 Punggol Coast Road, Singapore 828608, Republic of Singapore

⁷School of Computational Sciences, Korea Institute for Advanced Study (KIAS), Seoul 02455, Korea

⁸Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, Korea

The Quantum Approximate Optimisation Algorithm (QAOA) tackles combinatorial optimisation problems by encoding their solutions into the ground state of an Ising Hamiltonian prepared by a p -level parameterised circuit, with the angles tuned classically. Parameter optimisation is widely regarded as a central bottleneck, even for the shallowest circuits. Focusing on QAOA at $p = 1$ (QAOA₁), we show that tuning the two angles (γ, β) for weighted Ising models is not a black-box search but a structured signal-processing problem. We prove that the QAOA₁ expectation value is a partial Fourier series in γ whose frequencies are determined explicitly by the problem’s couplings and fields, giving instance-wise bandwidth bounds and, via the Nyquist–Shannon theorem, the sampling resolution needed to avoid the aliasing that causes coarse-grid searches to return spurious optima. We then eliminate the mixer angle analytically, computing $\beta^*(\gamma)$ in closed form to reduce the search to one dimension, and apply a subdivision algorithm that locates the globally optimal γ in polynomial time with a certificate of optimality when the weights are

commensurable and bounded. For regular weighted graphs, we further prove the conventional wisdom that the globally optimal $\gamma^* \in \mathbb{R}^+$ concentrates near zero and coincides with the first local optimum, giving a rigorous account of the empirical success of small-angle initialisation and allowing gradient descent to replace exhaustive line searches. Validated within Recursive QAOA (RQAOA) on weighted instances of 128 and 256 qubits, our method consistently outperforms both coarsely optimised RQAOA and semidefinite programming.

1 Introduction

Quantum computing is poised to revolutionise problem-solving in fields ranging from optimisation and machine learning to quantum chemistry and materials science. While fault-tolerant quantum computers hold the promise of implementing groundbreaking algorithms like Shor’s for factoring [1] and Grover’s for unstructured search [2], current quantum hardware falls short of the error correction required to implement these algorithms at scale. Instead, the era of Noisy Intermediate-Scale Quantum (NISQ) devices [3, 4] has given rise to variational quantum algorithms (VQAs), which blend parameterised quantum circuits (PQCs) with classical

V Vijendran : vjqntm@gmail.com

Syed M Assad : cqtsma@gmail.com

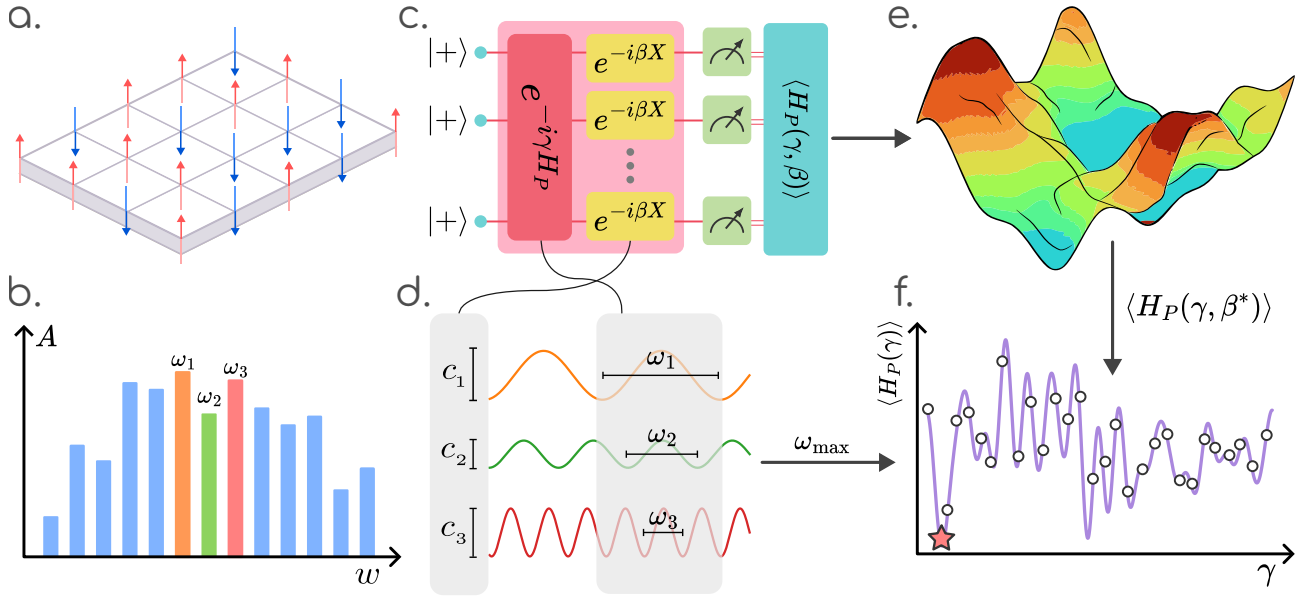


Figure 1: Illustration of the Main Results. This paper focuses on optimising the variational parameters of QAOA₁ for solving QUBO problems formulated as Ising models. (a) The Ising model is represented by the problem Hamiltonian H_P , which is diagonal in the Pauli-Z basis. Its eigenvalues define (b) the spectrum of frequencies ω for the given problem instance, where each ω_i corresponds to the difference between eigenvalues. To solve the problem exactly, the quantum circuit must express the full spectrum of frequencies. However, (c) QAOA₁ is limited to representing only a small subset of the spectrum. The expectation value of the cost function in QAOA₁ can be expressed as (d) a truncated Fourier series whose frequencies are determined by the problem unitary, while its coefficients depend on the mixing unitary. At $p = 1$, most coefficients vanish, significantly constraining the circuit's expressivity. We provide an analytical method to compute the maximum frequency of the cost function along γ , enabling an appropriate sampling rate for optimisation. Previous studies derived analytical expressions for classically computing the expectation value $\langle H_P(\gamma, \beta) \rangle$, yielding (e) the cost landscape. In this work, we show how the two-dimensional optimisation over (γ, β) can be reduced to (f) a univariate optimisation over γ , with β^* determined analytically as a function of γ . By leveraging the maximum frequency, the univariate cost function can be efficiently sampled, enabling accurate estimation of the optimal γ^* without overshooting the global optimum. Finally, we prove that, on average, the global optimum $\gamma^* \in \mathbb{R}^+$ is concentrated near 0 and that this global optimum coincides with the first local optimum.

optimisation to tackle complex problems [5, 6]. By iteratively refining circuit parameters, VQAs navigate vast quantum state spaces, leveraging their adaptability to address diverse computational challenges. However, their success depends on efficient parameter optimisation, which is complicated by the intricate optimisation landscapes of PQCs [7–13].

Among VQAs, the Quantum Approximate Optimisation Algorithm (QAOA) [14] has emerged as a leading contender for solving combinatorial optimisation problems. QAOA translates these problems into Hamiltonians and approximates their ground states using circuits parameterised by $2p$ variables, where p represents the number of alternating layers of problem and mixing unitaries. While increasing p improves solution quality, it also amplifies the computational burden of parameter optimisation and is further hindered by physical constraints such

as gate noise, limited qubit connectivity, and state-preparation-and-measurement errors, effectively limiting QAOA implementations to shallow depths on NISQ devices [15]. Even in this regime, optimising QAOA parameters remains a significant challenge, particularly for large, dense, asymmetric, or weighted problem instances. Addressing these challenges has inspired a growing body of research focused on heuristic methods to identify near-optimal parameters efficiently, with the goal of unlocking QAOA's full potential in shallow-depth applications.

The search for effective initialisation and optimisation strategies for QAOA, aimed at minimising the number of optimisation steps required for solving combinatorial problems, remains an active area of research. Effective initialisation ensures the algorithm begins close to a local or global optimum in the parameter space, while efficient optimisation enables smooth and rapid

traversal through the landscape. Existing strategies to address this challenge range from physics-based approaches to advanced machine learning techniques, including transfer learning and data-driven models. Early work focused on deriving analytical expressions for optimal parameters at $p = 1$ providing performance bounds for specific graph structures such as unweighted D -regular triangle-free graphs and large-girth regular graphs; however, these derivations are less effective for complex, weighted graphs encountered in practical settings [14, 16–22]. Parameter transfer methods, inspired by transfer learning, leverage optimised parameters from similar problem instances, particularly for uniform-structure problems like MaxCut on unweighted D -regular graphs; however, they are less reliable for graphs without well-defined distributions or regular structure [23–30]. Quantum annealing insights also guide parameter initialisation by setting parameters based on a fraction of the optimal annealing time, with additional refinements from control theory techniques like the bang-bang protocol and Lyapunov control, yet these are primarily suited to deeper circuits and struggle with capturing optimal transitions in shallow QAOA circuits [31–33]. Another approach leverages an optimised level- p circuit to initialise a level- $(p+1)$ circuit, using methods like linear interpolation or Fourier transformations to reduce optimisation steps, although full optimisation of the level- p circuit is typically required [34]. Efficiency-oriented techniques such as multilevel leapfrogging [35] and depth-progressive initialisation [36] reduce computational costs by incrementally optimising parameters, but these methods often yield suboptimal solutions and generally lack performance guarantees, particularly for real-world weighted problems where structure varies widely [37, 38]. Finally, various machine learning approaches [39–47] have been proposed to predict and generate near-optimal initial guesses for QAOA parameter optimisation. However, these methods often incur high computational costs and also lack performance guarantees.

This work addresses the challenge of optimising QAOA₁ for QUBO problems, formulated as Ising models. Despite QAOA₁ involving only two parameters, (γ, β) , the widely held belief that coarse grid searches followed by local minimisation suffice is misleading. We reveal that the optimisa-

tion landscape of QAOA₁ is highly oscillatory, with oscillation rates increasing with problem size, density, and weight variability. These oscillations necessitate high-resolution grid searches to avoid landscape distortion and suboptimal parameter estimates. To mitigate the computational cost of grid searches at fine resolutions, we propose an efficient parameter-tuning strategy that reduces the two-dimensional search over (γ, β) to a one-dimensional search over γ , with the optimal β^* derived analytically. We establish the minimum sampling frequency required to capture the oscillatory landscape accurately and provide an algorithm to estimate the optimal parameters in polynomial time. For sufficiently large instances of regular graphs, we rigorously prove the globally optimal $\gamma^* \in \mathbb{R}^+$ concentrate very close to 0 and that this global optimum coincides with the first local optimum. This insight enables the replacement of exhaustive line searches with gradient descent, dramatically simplifying the optimisation process. Our approach is provably near-optimal—if not optimal—problem-agnostic and effectively addresses the challenges associated with parameter tuning in QAOA₁ for QUBO problems.

The utility of our method extends beyond QAOA₁ itself, as it can be applied to quantum algorithms that use QAOA₁ as a subroutine for solving QUBO problems [48–60]. Additionally, the optimised parameters for $p = 1$ can serve as warm-start values for deeper QAOA circuits ($p > 1$). Instead of initialising all parameters randomly, one can utilise the near-optimal $p = 1$ parameters and only randomly initialise the additional parameters required for higher depths. This approach provides a better starting point for the optimiser, enhancing convergence efficiency compared to traditional random initialisation methods.

We validate our strategy by applying it to the Recursive QAOA (RQAOA) [48], benchmarking its performance on dense and weighted QUBO instances. Our results show that, unlike coarsely optimised RQAOA₁, which struggles to outperform semidefinite programming (SDP), our method consistently enables RQAOA₁ to surpass SDP, coarsely optimised RQAOA₁, and QAOA₁ across all tested instances. These findings highlight the critical importance of parameter optimisation in improving variational quantum algo-

rithm performance, demonstrating the potential of quantum methods to outperform state-of-the-art classical techniques in producing high-quality approximations.

The paper is structured as follows: ‘Preliminary Material’ introduces the foundational concepts, followed by ‘Theoretical Results’ and ‘Numerical Results’, which present our primary findings. The ‘Discussion’ section delves into the broader implications of these results. Comprehensive details on the benchmarked QAOA variants are available in the ‘Methods’ section, and all proofs and additional supporting information are included in the supplementary materials.

2 Preliminary Material

2.1 Ising Models

The Ising model was introduced to describe ferromagnetic interactions in materials. In this model, binary variables, called spins, take values of ± 1 to represent two possible states of a particle. Spins interact pairwise through coupling strengths and may also experience external influences, such as a magnetic field. The goal is to find the ground state configuration of spins that minimises the system’s total energy.

Mathematically, the Ising model is defined by a set of spin variables $s_j \in \{-1, +1\}$ for $j = 1, 2, \dots, n$, which interact via a symmetric coupling matrix $\mathbf{J} \in \mathbb{R}^{n \times n}$. External magnetic fields are represented by a vector $\mathbf{h} \in \mathbb{R}^n$. The energy of a configuration $\mathbf{s} = (s_1, s_2, \dots, s_n)$ is given by the Hamiltonian:

$$H(\mathbf{s}) = \sum_{j=1}^n \sum_{k=1}^n J_{jk} s_j s_k + \sum_{j=1}^n h_j s_j \quad (1)$$

$$= \mathbf{s}^T \mathbf{J} \mathbf{s} + \mathbf{h}^T \mathbf{s}.$$

The goal is to find a configuration $\mathbf{s}^* \in \{-1, +1\}^n$ that minimises $H(\mathbf{s})$. This optimisation task parallels the QUBO [61] problem, which involves minimising a quadratic function over binary variables.

At its core, a QUBO problem seeks a binary vector $\mathbf{x} \in \{0, 1\}^n$ that minimises the quadratic objective:

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x}, \quad (2)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a symmetric matrix capturing the quadratic interactions between variables,

and $\mathbf{b} \in \mathbb{R}^n$ represents the linear coefficients. For the problem to remain non-trivial, $\mathbf{A}$ must contain negative eigenvalues; otherwise, setting all binary variables to zero yields a trivial optimal solution of zero. The binary nature of the variables implies that $x_j^2 = x_j$, which allows the linear term $\mathbf{b}^T \mathbf{x}$ to be absorbed into the quadratic form by adjusting the diagonal elements of $\mathbf{A}$. Specifically, the quadratic objective function $f(\mathbf{x})$ can be rewritten as

$$f(\mathbf{x}) = \sum_{j=1}^n \sum_{k=1}^n A_{jk} x_j x_k + \sum_{j=1}^n b_j x_j \quad (3)$$

$$= \sum_{j=1}^n \sum_{k=1}^n A'_{jk} x_j x_k.$$

where $A'_{jj} = A_{jj} + b_j$ for $j = 1, 2, \dots, n$, effectively incorporating the linear coefficients into the quadratic matrix.

The equivalence between QUBO and Ising models arises from a simple variable substitution, $x_j = \frac{1}{2}(s_j + 1)$, mapping QUBO variables to Ising spins. Substituting this into the QUBO formulation gives:

$$\mathbf{J} = \frac{1}{4} \mathbf{A}, \quad (4)$$

$$h_j = \frac{1}{4} \left(2b_j + \sum_{k=1}^n (A_{jk} + A_{kj}) \right), \quad (5)$$

up to an insignificant additive constant.

Linear terms in the Ising model can similarly be removed by extending the coupling matrix. Introducing an additional spin variable s_{n+1} and an extended coupling matrix¹ $\bar{\mathbf{J}} \in \mathbb{R}^{(n+1) \times (n+1)}$, the reformulated matrix is:

$$\bar{J}_{jk} = \begin{cases} J_{jk}, & \text{if } 1 \leq j, k \leq n \\ \frac{1}{2} h_j & \text{for } k = n+1, \forall j \in [n] \\ \frac{1}{2} h_k & \text{for } j = n+1, \forall k \in [n] \\ 0 & \text{for } j = k = n+1. \end{cases} \quad (6)$$

This reformulation produces an Ising model equivalent to the original but without explicit linear terms.

¹In contrast to the QUBO model, where linear terms are absorbed into the matrix diagonal, the Ising model sets diagonal entries to zero, as they add only a constant to any solution. Instead, an additional spin variable is introduced, coupling to other spins with strengths proportional to the external field.

2.2 Quantum Approximate Optimisation Algorithm

The Quantum Approximate Optimisation Algorithm (QAOA) is a hybrid quantum-classical algorithm for approximately solving combinatorial optimisation problems, including QUBO, by encoding solutions into the ground state of a Hamiltonian, approximated by a PQC. QAOA relies on two primary operators: the problem unitary and the mixing unitary, derived from the problem Hamiltonian H_P and the mixing Hamiltonian H_M , defined as:

$$H_P = \sum_{j=1}^n \sum_{k=1}^n J_{jk} Z_j Z_k + \sum_{j=1}^n h_j Z_j, \quad (7)$$

$$H_M = \sum_{j=1}^n X_j, \quad (8)$$

where Z_j and X_j are the Pauli-Z and Pauli-X operators acting on the j -th qubit. The problem Hamiltonian H_P , a quantum version of eq. (1), encodes the objective function to be minimised, while the mixing Hamiltonian H_M facilitates transitions between states.

The problem unitary is parameterised by a real-valued angle $\gamma \in \mathbb{R}$ and is given by:

$$U(H_P, \gamma) = e^{-i\gamma H_P} = \prod_{j,k} e^{-i\gamma J_{jk} Z_j Z_k} \prod_j e^{-i\gamma h_j Z_j}. \quad (9)$$

while the mixing unitary, parameterised by $\beta \in [0, \pi]$, is expressed as:

$$U(H_M, \beta) = e^{-i\beta H_M} = \prod_{j=1}^n e^{-i\beta X_j}. \quad (10)$$

For any positive integer p , the QAOA ansatz generates a parameterised quantum state using $2p$ angles, $(\gamma, \beta) = (\gamma_1, \dots, \gamma_p, \beta_1, \dots, \beta_p)$, by alternating the application of $U(H_P, \gamma)$ and $U(H_M, \beta)$, resulting in the quantum state:

$$|\gamma, \beta\rangle = \prod_{i=1}^p U(H_M, \beta_i) U(H_P, \gamma_i) |s\rangle, \quad (11)$$

where $|s\rangle$ denotes the uniform superposition over all n -bit strings:

$$|s\rangle = \frac{1}{\sqrt{2^n}} \sum_{z \in \{0,1\}^n} |z\rangle. \quad (12)$$

The goal is to optimise the parameters γ and β by minimising the expectation value of H_P with respect to the generated quantum state:

$$\langle H_P(\gamma, \beta) \rangle = \langle \gamma, \beta | H_P | \gamma, \beta \rangle, \quad (13)$$

which is computed through repeated measurements of the quantum system. The performance of QAOA is often quantified by the approximation ratio, which compares the objective value achieved at the optimised parameters, $\langle H_P(\gamma^*, \beta^*) \rangle$, with the optimal value of the underlying problem.

To identify the optimal parameters, (γ^*, β^*) , a classical optimiser iteratively adjusts γ and β to minimise $\langle H_P \rangle$:

$$(\gamma^*, \beta^*) = \arg \min_{\gamma, \beta} \langle H_P(\gamma, \beta) \rangle. \quad (14)$$

However, in general finding these parameters is NP-Hard [7], and practical implementations rely on heuristic methods such as gradient descent and grid search. Gradient descent can suffer from poor initialisation, local minima, and barren plateaus, while grid search becomes computationally infeasible as the number of parameters grows.

In this work, we focus on QAOA at depth $p = 1$ (QAOA₁), with classical parameters (γ, β) applied to an Ising model, where closed-form expressions are available for evaluating eq. (13). Before exploring these expressions, we introduce a few definitions that will be used throughout the paper.

The Ising model is depicted by an undirected graph $G = (V, E)$, where each vertex $v \in V$ signifies a spin, and each edge $\{u, v\} \in E$ signifies an interaction between spins u and v . These interactions are defined by the symmetric coupling matrix $\mathbf{J}$, with each edge $\{u, v\}$ assigned a coupling strength J_{uv} . Furthermore, the external magnetic fields are represented by node weights, where each vertex $u \in V$ carries an external field strength h_u , collectively denoted by the vector $\mathbf{h}$. For a vertex $w \in V$, let $\mathcal{N}(w) = \{x \in V \mid \{x, w\} \in E\}$ represent the set of neighbours of w , i.e., vertices adjacent to w . For an edge $\{u, v\} \in E$, we define:

- $e_{uv} = \mathcal{N}(v) \setminus \{u\}$ is the set of vertices other than u that are connected to v .
- $d_{uv} = \mathcal{N}(u) \setminus \{v\}$ is the set of vertices other than v that are connected to u .

- $F_{uv} = \mathcal{N}(u) \cap \mathcal{N}(v)$ is the set of vertices that form a triangle with the edge $\{u, v\}$. In other words, F_{uv} is the set of vertices that are neighbours of both u and v .

For compactness, we omit subscripts and use e , d , and F directly when their meanings are clear from context.

The following theorem can be used to compute the expectation value of QAOA₁ for an arbitrary Ising model with fields, thereby allowing us to assess its performance.

Theorem 1. Consider the QAOA₁ state $|\gamma, \beta\rangle$

$$\begin{aligned} \langle C_{uv} \rangle &= \frac{J_{uv}}{2} \sin 4\beta \sin \gamma'_{uv} \left(\cos \gamma'_v \prod_{w \in e} \cos \gamma'_{vw} + \cos \gamma'_u \prod_{w \in d} \cos \gamma'_{uw} \right) \\ &\quad - \frac{J_{uv}}{2} \sin^2 2\beta \prod_{\substack{w \in e \\ w \notin F}} \cos \gamma'_{vw} \prod_{\substack{w \in d \\ w \notin F}} \cos \gamma'_{uw} \left(\sum_{\chi = \pm 1} \chi \cos(\gamma'_u + \chi \gamma'_v) \prod_{f \in F} \cos(\gamma'_{uf} + \chi \gamma'_{vf}) \right) \end{aligned} \quad (17)$$

and $\gamma'_{uv} = 2J_{uv}\gamma$ and $\gamma'_i = 2h_i\gamma$.

In the absence of local fields (i.e., $h_i = 0$ for all $i \in V$), the Hamiltonian H_P of the Ising model reduces to

$$H_P = \sum_{j=1}^n \sum_{k=1}^n J_{jk} Z_j Z_k, \quad (18)$$

and the expression for the expectation value simplifies, as presented in the following corollary.

$$\begin{aligned} \langle C_{uv} \rangle &= \frac{J_{uv}}{2} \sin 4\beta \sin \gamma'_{uv} \left(\prod_{w \in e} \cos \gamma'_{vw} + \prod_{w \in d} \cos \gamma'_{uw} \right) \\ &\quad - \frac{J_{uv}}{2} \sin^2 2\beta \prod_{\substack{w \in e \\ w \notin F}} \cos \gamma'_{vw} \prod_{\substack{w \in d \\ w \notin F}} \cos \gamma'_{uw} \left(\prod_{f \in F} \cos(\gamma'_{uf} + \gamma'_{vf}) - \prod_{f \in F} \cos(\gamma'_{uf} - \gamma'_{vf}) \right) \end{aligned} \quad (20)$$

and $\gamma'_{uv} = 2J_{uv}\gamma$.

The proof of theorem 1, which concerns arbitrary Ising models with external fields, follows Ozaeta et al. [18]. In the absence of external fields, related expectation-value formulas were previously derived for unweighted Ising/MaxCut-type instances by Wang et al. [17], and for

for an arbitrary Ising model with external fields. The expectation value of the Hamiltonian H_P , as defined in eq. (7), in this state is given by:

$$\langle \gamma, \beta | H_P | \gamma, \beta \rangle = \sum_{\{u,v\} \in E} \langle C_{uv} \rangle + \sum_{i \in V} \langle C_i \rangle, \quad (15)$$

where

$$\langle C_i \rangle = h_i \sin 2\beta \sin \gamma'_i \prod_{k \in \mathcal{N}(i)} \cos \gamma'_{ik} \quad (16)$$

Corollary 1. Consider the QAOA₁ state $|\gamma, \beta\rangle$ for an arbitrary Ising model without external fields. The expectation value of the Hamiltonian H_P , as defined in eq. (18), in this state is given by:

$$\langle \gamma, \beta | H_P | \gamma, \beta \rangle = \sum_{\{u,v\} \in E} \langle C_{uv} \rangle, \quad (19)$$

where

weighted field-free graphs by Bravyi et al. [48] and Vijendran et al. [19]. Corollary 1 is the corresponding field-free specialisation in our notation.

From theorem 1 and corollary 1, it is clear that, for $p = 1$, the expectation value $\langle C_{uv} \rangle$ of any edge in a graph depends only on the nodes and edges adjacent to it. Similarly, the single-spin term $\langle C_i \rangle$

for any node depends only on the node's external field and its adjacent edges. The total expectation value for QAOA₁ is obtained by summing over all nodes and edges. For a system with n nodes, the analytical expressions for each edge and node can be computed in linear time, $\mathcal{O}(n)$. Since the graph has at most $\binom{n}{2} = \mathcal{O}(n^2)$ edges, the overall time complexity for computing the full expectation value for QAOA₁ is $\mathcal{O}(n^3)$. To obtain an approximate solution in the form of a bit string for an arbitrary problem instance, however, requires executing the QAOA circuit on a quantum computer to generate the quantum state, followed by measurements on that state.

2.3 Fourier Analysis of VQAs

Fourier analysis studies how general functions can be represented or approximated by sums of simpler trigonometric functions, allowing complex functions to be analysed in terms of their frequency components and providing insight into their structure. In VQAs, Fourier analysis is a powerful tool for examining the cost landscape [62–66]. By expressing the cost function as a Fourier series, we can analyse the frequency spectrum determined by the problem Hamiltonian and circuit parameters. This approach reveals the accessible frequencies and structural features of the optimisation landscape.

For a PQC with a cost function defined by $\mathcal{C}(\boldsymbol{\theta}) = \langle 0|U^\dagger(\boldsymbol{\theta})H_P U(\boldsymbol{\theta})|0\rangle$, where H_P is the problem Hamiltonian and $\boldsymbol{\theta} \in \mathbb{R}^M$ represents M independent continuous parameters, the Fourier series representation is given by:

$$\mathcal{C}(\boldsymbol{\theta}) = \sum_{\boldsymbol{\omega} \in \Omega} c_{\boldsymbol{\omega}} e^{i\boldsymbol{\omega} \cdot \boldsymbol{\theta}}, \quad (21)$$

where $c_{\boldsymbol{\omega}}$ are the Fourier coefficients corresponding to the frequencies $\boldsymbol{\omega}$, and $\boldsymbol{\omega} \cdot \boldsymbol{\theta}$ denotes their inner product. The frequency spectrum $\Omega \subset \mathbb{R}^M$ is determined entirely by the eigenvalues of the problem Hamiltonian, while the circuit design influences the coefficients $c_{\boldsymbol{\omega}}$ that the quantum model can realise.

By representing quantum models as partial Fourier series—where only a subset of Fourier coefficients are non-zero—we can characterise the functional families that a quantum model can learn based on two key properties. The first is the frequency spectrum Ω , which specifies the set of Fourier basis functions $e^{i\boldsymbol{\omega} \cdot \boldsymbol{\theta}}$ that may appear in

the model. The second is the expressivity of the coefficients $\{c_{\boldsymbol{\omega}}\}$, which determines which linear combinations of these basis functions can be realised. Together, Ω and $\{c_{\boldsymbol{\omega}}\}$ delineate the functional families that the quantum model can approximate and provide insight into cost-landscape features such as smoothness and the distribution of local minima.

3 Theoretical Results

3.1 QAOA as Partial Fourier Series

We show that the QAOA expectation value can be expressed as a partial Fourier series, revealing the frequency components accessible to the circuit. This decomposition links circuit depth to optimisation performance and highlights the expressivity limits of shallow QAOA (as discussed in section A). Identifying the maximum frequency also determines the grid resolution needed to characterise the QAOA₁ cost landscape effectively.

Consider the QAOA₁ with the quantum state

$$|\psi(\gamma, \beta)\rangle = e^{-i\beta H_M} e^{-i\gamma H_P} |s\rangle \quad (22)$$

for an arbitrary Ising model described by the Hamiltonian H_P . The expectation value of H_P with respect to the state $|\psi(\gamma, \beta)\rangle$ is given by:

$$f(\gamma, \beta) = \langle \psi(\gamma, \beta) | H_P | \psi(\gamma, \beta) \rangle. \quad (23)$$

Our goal is to express $f(\gamma, \beta)$ as a partial Fourier series of the form:

$$f(\gamma, \beta) = \sum_{n \in \Omega} c_n(\beta) e^{in\gamma} \quad (24)$$

with integer-valued frequencies (if $\Omega = \{-K, \dots, K\}$, then we call eq. (24) a truncated Fourier series). To proceed, let us consider the eigenstates $|j\rangle$ of the cost Hamiltonian H_P , with corresponding eigenvalues λ_j such that:

$$H_P |j\rangle = \lambda_j |j\rangle. \quad (25)$$

The unitary evolution under H_P , given by $e^{-i\gamma H_P}$, can be expanded in the eigenbasis of H_P as:

$$e^{-i\gamma H_P} = \sum_{j=1}^n e^{-i\gamma \lambda_j} |j\rangle \langle j|. \quad (26)$$

Applying the problem unitary $e^{-i\gamma H_P}$ to $|s\rangle$ yields:

$$e^{-i\gamma H_P}|s\rangle = \sum_{j=1}^n e^{-i\gamma\lambda_j} \langle j|s\rangle |j\rangle. \quad (27)$$

Thus, the expectation value $f(\gamma, \beta)$ becomes:

$$f(\gamma, \beta) = \sum_{j=1}^n \sum_{k=1}^n c_{jk}(\beta) e^{i\gamma(\lambda_j - \lambda_k)}, \quad (28)$$

where $c_{jk}(\beta)$ are the Fourier coefficients encapsulating the contributions from the overlaps of the eigenstates of H_P after evolution under the mixer Hamiltonian H_M . These coefficients are given by:

$$\begin{aligned} c_{jk}(\beta) &= \langle s|j\rangle \langle j|e^{i\beta H_M} H_P e^{-i\beta H_M}|k\rangle \langle k|s\rangle \\ &= 2^{-n} \langle j|e^{i\beta H_M} H_P e^{-i\beta H_M}|k\rangle. \end{aligned} \quad (29)$$

Since the problem Hamiltonian H_P is diagonal in the Pauli-Z basis, with diagonal entries representing cost values for each bitstring $z \in \{1, -1\}^n$, QAOA aims to identify the eigenvector corresponding to the minimum eigenvalue to solve the optimisation problem. The largest possible frequency of the expectation value $f(\gamma, \beta)$ is determined by the difference between the costs of the best and worst solutions, $\lambda_{\max} - \lambda_{\min}$. To fully capture this frequency range, all coefficients in eq. (28) must be non-zero for each pair of eigenvalues in H_P .

At shallow depths, QAOA is restricted in the frequencies it can access because many Fourier coefficients $c_{jk}(\beta)$ vanish, particularly when $|\lambda_j - \lambda_k|$ is large. This limitation arises from the dependence of accessible frequencies on the circuit depth p , which governs the range of Fourier terms representable in eq. (28). As p increases, more frequencies become available, enhancing algorithm performance. Identifying the largest non-zero frequency in QAOA₁ provides valuable insights and helps determine the optimal resolution for grid searches over (γ, β) . This avoids under- or overestimating the resolution, which could compromise search accuracy or efficiency. Although finding these non-zero coefficients can be computationally challenging, we derive the maximum frequency for QAOA₁ on Ising models analytically using theorem 1 and corollary 1. The following subsection examines the results of this derivation, shedding light on the expressivity of QAOA₁ and its optimisation landscape.

3.2 Maximum Frequency of Level-1 QAOA

To understand the maximum frequency for QAOA₁ on Ising models, it is useful to reinterpret the Fourier series in eq. (28). This representation frames the expectation value of QAOA₁, as described in theorem 1 and corollary 1, as a bandlimited two-dimensional signal. For clarity in the subsequent analysis, we begin by formally defining the concept of a bandlimited function.

Definition 1. Let $T > 0$ be the period and $K \in \mathbb{N}$ a positive integer. A periodic function $f(t)$ with period T is called *bandlimited* to bandwidth $K\omega_0$, where $\omega_0 = 2\pi/T$, if it can be expressed as

$$f(t) = \sum_{k=-K}^K c_k e^{ik\omega_0 t}, \quad (30)$$

where $c_k = 0$ for all integers k with $|k| > K$.

For any fixed value of β , we treat the β -dependent terms in theorem 1 and corollary 1 as prefactors and focus on the γ -dependence. This is because the instance-dependent oscillatory structure enters through γ , which is weighted by the couplings and external fields in the closed-form expressions, and equivalently appears through the phase factors $e^{i\gamma(\lambda_j - \lambda_k)}$ in the Fourier representation of eq. (28). The signal along the γ dimension—representing the cost landscape—is periodic only if the weights of the problem instance are commensurable².

For a bandlimited function $f(t)$ as defined in definition 1, we denote its maximum angular frequency by $\omega_{\max}[f]$, where $\omega = 2\pi\nu$ relates angular frequency to the ordinary frequency ν . The following theorem leverages the closed-form expressions in theorem 1 to compute the maximum angular frequency of the expectation value in QAOA₁, specifically for the γ -dependent terms in Ising models with fields.

Theorem 2. Consider the QAOA₁ state $|\gamma, \beta\rangle$ for an arbitrary Ising model with fields. For any fixed β , the maximum angular frequency of the γ -dependent terms in the expectation value of the Hamiltonian H_P , as defined in eq. (7), is given

²Two values are said to be commensurable if their ratio is a rational number, meaning they share a common measure or can be expressed as integer multiples of a common unit.

by:

$$\begin{aligned}\omega_{\max}[\langle H_P \rangle_\gamma] &= \max \left\{ \omega_{\max}[\langle C_i \rangle_\gamma], \right. \\ &\quad \left. \omega_{\max}[\langle C_{uv} \rangle_\gamma] \mid i \in V, \{u, v\} \in E \right\} \quad (31)\end{aligned}$$

where

$$\omega_{\max}[\langle C_i \rangle_\gamma] = 2 \times \left(|h_i| + \sum_{k \in \mathcal{N}(i)} |J_{ik}| \right) \quad (32)$$

$$\begin{aligned}\omega_{\max}[\langle C_{uv} \rangle_\gamma] &= 2 \times \max \left\{ |J_{uv}| + \max \left\{ |h_v| + \sum_{w \in e} |J_{vw}|, |h_u| + \sum_{w \in d} |J_{uw}| \right\}, \right. \\ &\quad \left. \sum_{\substack{w \in e \\ w \notin F}} |J_{vw}| + \sum_{\substack{w \in d \\ w \notin F}} |J_{uw}| + \max \left\{ |h_u \pm h_v| + \sum_{f \in F} |J_{uf} \pm J_{vf}| \right\} \right\}. \quad (33)\end{aligned}$$

The proof of this theorem is provided in section C. Notably, this theorem simplifies in the absence of external fields. The following corollary presents the maximum angular frequency of the expectation value for QAOA₁ in Ising models without fields, which can also be derived directly from corollary 1.

Corollary 2. *Consider the QAOA₁ state $|\gamma, \beta\rangle$ for an arbitrary Ising model without fields. For any fixed β , the maximum angular frequency of the γ -dependent terms in the expectation value of the Hamiltonian H_P , as defined in eq. (18), is given by:*

$$\omega_{\max}[\langle H_P \rangle_\gamma] = \max \left\{ \omega_{\max}[\langle C_{uv} \rangle_\gamma] \mid \{u, v\} \in E \right\}, \quad (34)$$

where

$$\begin{aligned}\omega_{\max}[\langle C_{uv} \rangle_\gamma] &= 2 \times \max \left\{ |J_{uv}| + \max \left\{ \sum_{w \in e} |J_{vw}|, \right. \right. \\ &\quad \left. \sum_{w \in d} |J_{uw}| \right\}, \sum_{\substack{w \in e \\ w \notin F}} |J_{vw}| + \sum_{\substack{w \in d \\ w \notin F}} |J_{uw}| \\ &\quad \left. + \max \left\{ \sum_{f \in F} |J_{uf} \pm J_{vf}| \right\} \right\}. \quad (35)\end{aligned}$$

Proof. This result follows directly from theorem 2 by setting $h_i = 0$ for all $i \in V$. $\square$

and

These results characterise the maximum frequency of QAOA₁ for arbitrary Ising models, including cases with external fields, allowing us to efficiently analyse the cost landscape and identify optimal parameters.

3.3 Characterising the Energy Landscape of Level-1 QAOA

Building on the frequency characterisation of QAOA₁, we employ the Nyquist-Shannon Sampling Theorem [67] to determine the maximum permissible sampling period required to efficiently reconstruct its cost landscape. Unlike prior studies that have utilised signal processing techniques to analyse the optimisation landscapes of PQCs [65, 68], our method is specifically tailored to QAOA₁ applied to Ising models. It relies exclusively on analytical expressions, eliminating the need to execute the quantum algorithm on actual hardware. Moreover, our approach captures the entire QAOA₁ landscape, including challenging features such as barren plateaus and narrow gorges, which earlier methods do not address.

Before proceeding, we formally state the Nyquist-Shannon Sampling Theorem [67]:

Theorem 3 (Nyquist-Shannon Sampling Theorem). *Let $x(t)$ be a bandlimited function with maximum frequency B . If $x(t)$ is sampled at uniform intervals $t_n = nT_s$ for all integers $n \in \mathbb{Z}$,*

and the sampling period T_s satisfies

$$T_s \leq \frac{1}{2B + 1}, \quad (36)$$

then $x(t)$ is uniquely determined by its samples $\{x(t_n)\}$, and can be exactly reconstructed from these samples.

The Nyquist-Shannon Sampling Theorem ensures that a bandlimited function can be perfectly reconstructed provided the sampling rate is at least twice its highest frequency. Furthermore, for a periodic bandlimited function with period T , it suffices to sample the function within the interval $[0, T]$ at this minimum rate. In the context of QAOA₁, the theorem establishes the maximum allowable distance between consecutive samples (i.e., the maximum permissible sampling period) necessary to accurately reconstruct the optimisation landscape along γ . This result is formalised in the following theorem.

Theorem 4. *Let $|\gamma, \beta\rangle$ denote the QAOA₁ state applied to an arbitrary Ising model. Suppose the expectation value $\langle C(\gamma) \rangle = \langle \gamma, \beta^* | H_P | \gamma, \beta^* \rangle$ is a bandlimited function of γ with maximum frequency $\nu_{\max}[\langle C(\gamma) \rangle]$. Then if $\langle C(\gamma) \rangle$ is sampled at uniformly spaced points with spacing Δ_γ satisfying*

$$\Delta_\gamma \leq \frac{1}{2\nu_{\max}[\langle C(\gamma) \rangle] + 1}, \quad (37)$$

then $\langle C(\gamma) \rangle$ can be uniquely determined from these samples.

Proof. The proof follows from theorem 3 by setting $x(t) = \langle C(\gamma) \rangle$ and $B = \mu_{\max}/2\pi$, where $\mu_{\max}$ is determined from theorem 2 and corollary 2. $\square$

If the problem Hamiltonian H_P consists of incommensurable weights, the expectation value of QAOA₁ along γ becomes non-periodic, necessitating sampling at uniform intervals $n\Delta_\gamma$ for all integers $n \in \mathbb{Z}$. However, if the optimal γ^* is known *a priori* to lie within a specific region, it suffices to sample only within that interval (which we will discuss in section 3.6). In contrast, when H_P has commensurable weights, the expectation value of QAOA₁ along γ exhibits periodicity with period T_γ , making it sufficient to sample within the interval $[0, T_\gamma]$. Specifically, if H_P contains only integer eigenvalues, the maximum

possible period T_γ is π . Therefore, for Hamiltonians with fractional weights, it is advantageous to convert them to integer weights by multiplying by an appropriate constant factor, allowing sampling within $[0, \pi]$ and enabling exact reconstruction with evaluations at $[\pi/\Delta_\gamma]$ equally spaced points.

Beyond providing sampling requirements, theorem 4 highlights that the shape of the QAOA₁ optimisation landscape is closely tied to graph-dependent parameters governing the maximum frequency. In particular, higher frequencies in $\langle C(\gamma) \rangle$ can lead to more rapid changes or *sharper* features across γ , whereas lower-frequency regimes yield *smoother* variations. Empirical observations in, for example, Figure 9 of Vijendran et al. [19] and related works [38, 69] illustrate how increasing graph degree or density compresses regions of large gradients while expanding flatter regions of the landscape. In the remainder of this section, we apply theorem 4 to two graph families to elucidate how these graph properties influence the QAOA₁ optimisation landscape.

We begin with the Sherrington-Kirkpatrick (SK) model, a paradigmatic example from condensed matter physics characterised by its fully connected nature. The following corollary provides an upper bound on the sampling period required for QAOA₁ applied to this model, ensuring exact reconstruction of the expectation value.

Corollary 3. *Consider the QAOA₁ state $|\gamma, \beta^*\rangle$ applied to the Sherrington-Kirkpatrick model with n spins, where the coupling and field strengths are ± 1 . If the samples are spaced by Δ_γ satisfying*

$$\Delta_\gamma \leq \frac{\pi}{\mathcal{O}(n) + \pi}, \quad (38)$$

then the expectation value $\langle \gamma, \beta^* | H_P | \gamma, \beta^* \rangle$ can be exactly reconstructed for all $\gamma \in [0, \pi]$.

The proof of this corollary is provided in section D.1. It demonstrates that the sampling interval Δ_γ must decrease inversely with the system size n to maintain exact reconstruction. This behaviour is intuitive, as the degree of the SK model scales with the number of spins, $D = n - 1$, and the complexity of interactions increases with n . Consequently, the optimisation landscape exhibits increasingly rapid variations as n grows, necessitating finer sampling to accurately capture its features.

Next, we consider triangle-free D -regular graphs, which are graphs where each vertex has exactly D neighbours and no three vertices form a triangle. The following corollary establishes a precise upper bound on the sampling period for QAOA₁ applied to such graphs with ± 1 edge weights.

Corollary 4. *Let $|\gamma, \beta^*\rangle$ denote the QAOA₁ state applied to a triangle-free D -regular graph with edge weights ± 1 . If the samples are spaced by Δ_γ satisfying*

$$\Delta_\gamma \leq \frac{\pi}{2D + \pi}, \quad (39)$$

then the expectation value $\langle \gamma, \beta^* | H_P | \gamma, \beta^* \rangle$ can be exactly reconstructed for all $\gamma \in [0, \pi]$. For triangle-free graphs with bounded degrees, this bound holds with D replaced by the maximum degree $D_{\max}$.

The proof of this corollary is presented in section D.2. Similar to the SK model, the required sampling interval Δ_γ decreases inversely with the graph's degree D . In triangle-free regular graphs, the absence of triangles simplifies interactions, yet higher degrees still induce sharper features in the optimisation landscape. For triangle-free graphs with bounded degrees, Δ_γ inversely depends on the maximum degree $D_{\max}$. Furthermore, even in generally sparse graphs with localised dense regions, the maximum frequency—and consequently the required sampling interval—remains high, making the landscape highly oscillatory. Therefore, maintaining exact reconstruction necessitates finer sampling as the degree increases, ensuring accurate capture of the optimisation landscape and preventing the loss of critical information due to under-sampling.

By applying the Nyquist-Shannon Sampling Theorem, we derived the maximum permissible spacing between consecutive samples required to accurately reconstruct the optimisation landscape along γ . These results indicate that the necessary sampling density increases inversely with the size, weights, and density of the problem. While the β frequencies are independent of the problem Hamiltonian H_P , the γ frequencies depend directly on it. Consequently, performing a full grid search over (β, γ) becomes computationally expensive as the problem's size and complexity grow, particularly for weighted instances. To address this challenge, the following section

introduces a method to streamline the process by reducing the grid search to a line search over γ while solving analytically for β^* .

3.4 Univariate Representation of the Level-1 QAOA's Objective Function

We now present one of our main results: the closed-form expressions from theorem 1 and corollary 1 allow the expectation value of QAOA₁ on Ising models to be expressed solely in terms of the γ parameter, with the optimal β^* determined analytically for any $\gamma \in \mathbb{R}$. We begin by considering models without external fields, where the following theorem demonstrates that this reduction holds, enabling an analytical solution for β^* based on the optimal γ^* .

Theorem 5 (Ising Model without Fields). *Consider the QAOA₁ state $|\gamma, \beta\rangle$ for an arbitrary Ising model without fields. The expectation value of the Hamiltonian H_P , as defined in eq. (18), is a function of γ given by:*

$$\langle \gamma | H_P | \gamma \rangle = -\sqrt{A^2(\gamma) + \frac{B^2(\gamma)}{4}} - \frac{B(\gamma)}{2}. \quad (40)$$

The optimal angles (γ^*, β^*) that minimise $\langle \gamma, \beta | H_P | \gamma, \beta \rangle$ are given by:

$$\gamma^* = \arg, \min_{\gamma \in \mathbb{R}} \langle \gamma | H_P | \gamma \rangle, \quad (41)$$

$$\beta^* = \frac{1}{4} \left[\arctan \left(2A(\gamma^*), B(\gamma^*) \right) + \pi \right]. \quad (42)$$

The functions $A(\gamma)$ and $B(\gamma)$, defined over the graph $G = (V, E)$ corresponding to the Hamiltonian H_P , are given by:

$$\begin{aligned} A(\gamma) &= \sum_{\{u,v\} \in E} \frac{J_{uv}}{2} \sin \gamma'_{uv} \left(\prod_{w \in e} \cos \gamma'_{wv} + \prod_{w \in d} \cos \gamma'_{uw} \right), \\ B(\gamma) &= \sum_{\{u,v\} \in E} \frac{J_{uv}}{2} \prod_{\substack{w \in e \\ w \notin F}} \cos \gamma'_{wv} \prod_{\substack{w \in d \\ w \notin F}} \cos \gamma'_{uw} \\ &\quad \times \left(\prod_{f \in F} \cos(\gamma'_{uf} + \gamma'_{vf}) - \prod_{f \in F} \cos(\gamma'_{uf} - \gamma'_{vf}) \right), \end{aligned}$$

with $\gamma'_{uv} = 2J_{uv}\gamma$.

The proof of this theorem is provided in section E.2. This approach reduces the grid search over (γ, β) into an efficient line search on γ using

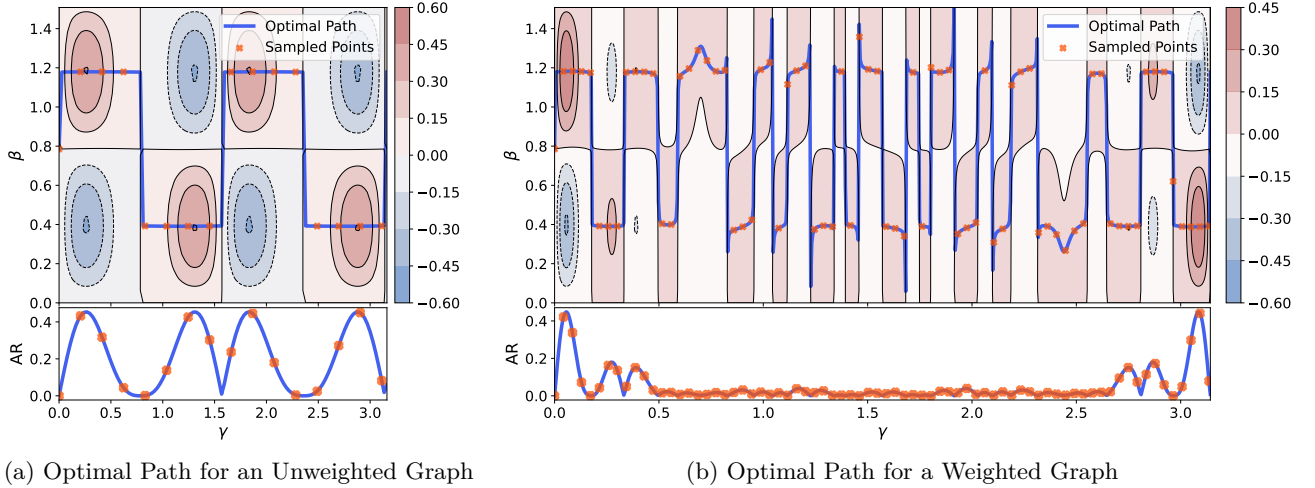


Figure 2: **Optimal (γ, β) Paths for QAOA₁ on a 4-Regular Graph.** These plots depict the optimal (γ, β) paths for QAOA₁ on a 4-regular graph with 128 vertices, calculated using theorem 5. The top contour plots show the cost landscape, with colours indicating approximation ratios (higher is better). The line plots below display the approximation ratios computed along γ using eq. (40), where the approximation ratio is defined as the objective value obtained by the algorithm divided by the optimal objective value for the same instance. Solid blue lines trace the optimal trajectory, (γ, β_γ^*) , where β_γ^* is the optimal β for each fixed γ . Orange 'x' markers indicate sampled points, spaced according to theorem 4. The left plot (unweighted graph) reveals symmetric extrema in the cost landscape, while the right plot (weighted graph with positive integer edge weights drawn from a Gaussian distribution with mean 5 and variance 1) shows extrema clustered near $\gamma \approx 0$ and π . Discontinuities in the optimal path, evident in both cases, result from abrupt shifts in β for specific γ values—an effect more pronounced in weighted graphs.

eq. (40), with sampled points spaced Δ_γ apart, and β^* determined using eq. (42). The effectiveness of theorem 5 is demonstrated in fig. 2, which presents contour plots and optimal paths on the cost landscape for both unweighted and lightly weighted 4-regular graphs with 128 vertices.

In fig. 2a, the unweighted graph exhibits clear symmetry, with large intervals along γ containing extrema. In contrast, the lightly weighted graph in fig. 2b presents a more intricate landscape. Although some symmetry remains due to the graph's regularity and small weight variance, the extrema clusters near $\gamma \approx 0$ and π , leaving the intermediate regions filled with spurious local minima and barren plateaus. A naive grid search in such scenarios would be computationally expensive, requiring extensive sampling to navigate these local traps. While exploiting symmetry can reduce computational effort, this strategy is only viable for highly structured problems, which are rare in practical applications. The method in theorem 5 overcomes these challenges by leveraging the line search on γ to compute β^* directly. This efficiency is evident in fig. 2, where the paths traced by eq. (40) show β alternating between lower and upper regions to maximise the approx-

imation ratio, with the line plot capturing the optimal trajectory along γ .

While theorem 5 reduces optimisation to a line search, determining an appropriate sampling resolution for the γ dimension is crucial for balancing efficiency and accuracy when optimising the univariate cost function. Excessively fine resolutions increase computational cost, while overly coarse resolutions risk failure due to local minima and barren plateaus. To illustrate, we use a weighted 4-regular graph with 64 vertices. Figure 3 compares the two-dimensional cost landscape to the one-dimensional univariate landscape at different γ resolutions. Unlike traditional contour plots, these visualisations (excluding fig. 3d without its inset) are generated using a mesh grid to avoid interpolation, which might further distort the already inaccurate landscape. As shown in fig. 3a, fig. 3b, and fig. 3c, both grid and line searches achieve only a marginal approximation ratio of 0.04 to 0.06, barely better than random guessing. In contrast, fig. 3d, with γ sampled at the rate prescribed by theorem 4, identifies the global extremum, achieving the maximum possible approximation ratio of approximately 0.48.

Building on theorem 5, we extend the uni-

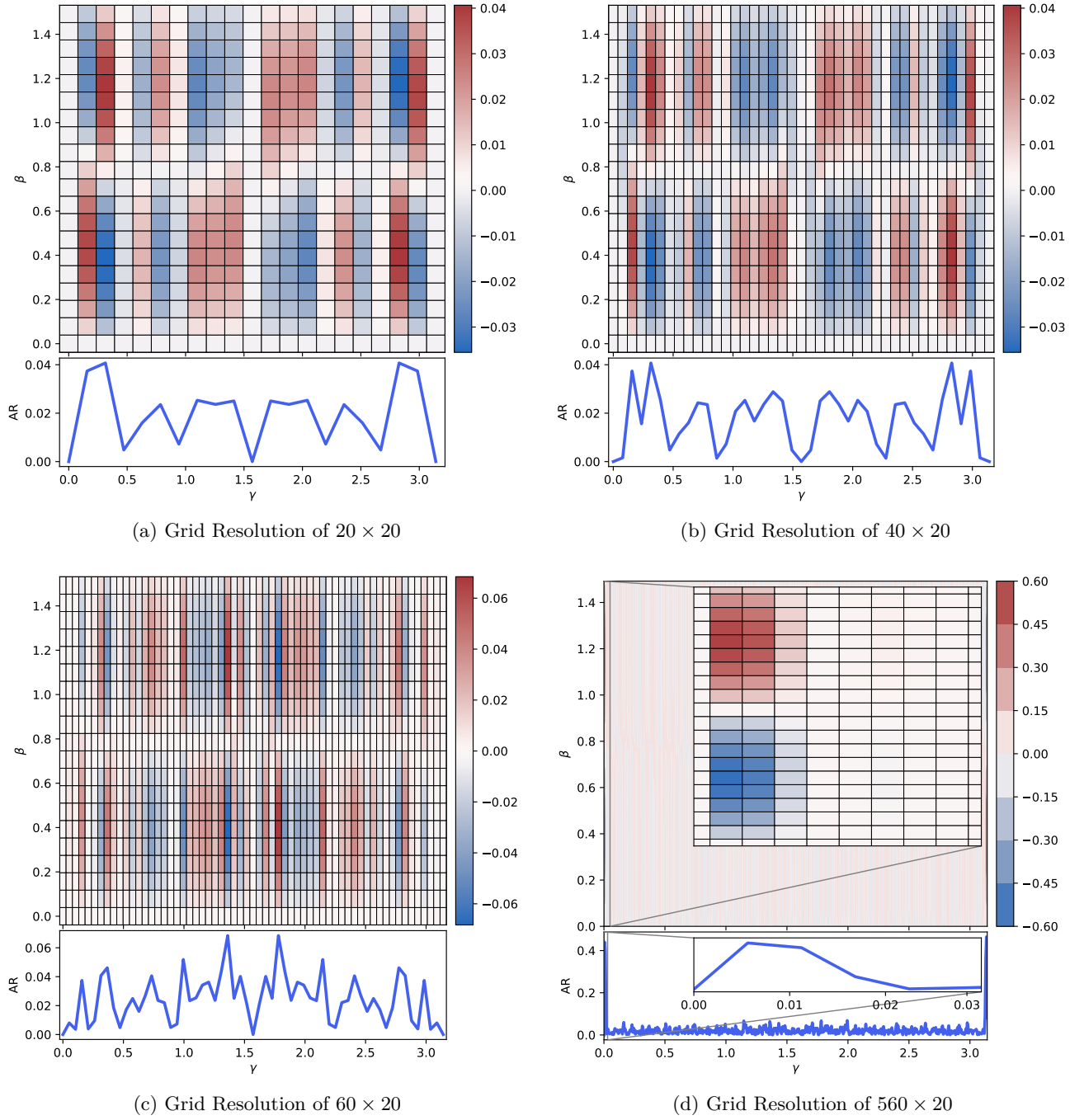


Figure 3: QAOA₁ Cost Landscape for a Weighted Regular Graph with Varying γ Resolutions. These plots illustrate the cost landscape of QAOA₁ for a weighted 4-regular graph with 64 vertices, where edge weights are sampled from a Gaussian distribution with a mean and variance of 25 and rounded to the nearest integer. The γ dimension is sampled at resolutions of 20, 40, 60, and 560, while the β dimension is fixed at 20 samples. The final plot's resolution of 560 was determined using theorem 4. Below each mesh grid, a line plot displays the cost landscape along γ , calculated using the expression in eq. (40). In plots (a) and (b), the maximum approximation ratio is 0.04, with the first optimum around $\beta < \pi/4$ and $\gamma \approx 0.15$. Plot (c) shows a higher ratio of 0.06 with γ near 1.4. In plot (d), sampling γ at the appropriate resolution yields a ratio of 0.48, with the optimum at $\beta > \pi/4$ and $\gamma \approx 0.01$. These results highlight the critical importance of appropriately sampling the γ dimension to avoid aliasing and distortion, which can lead to misleading local optima.

variate representation of the QAOA₁'s objective function to Ising models with external fields. The presence of external fields introduces additional complexity to the optimisation landscape, requiring a more general treatment. Nevertheless, the following theorem demonstrates that the QAOA₁'s objective function can still be simplified to a univariate form, enabling the reduction of the two-dimensional grid search over (γ, β) to an efficient line search over γ .

Theorem 6 (Ising Model with Fields). *Consider the QAOA₁ state $|\gamma, \beta\rangle$ for an arbitrary Ising model with fields, defined by the Hamiltonian H_P as in (7). The optimal angles γ^* and β^* that minimise the expectation value $\langle \gamma, \beta | H_P | \gamma, \beta \rangle$ are given by:*

$$(\gamma^*, \beta^*) = \arg \min_{(\gamma, \beta) \in \mathbb{R} \times \mathcal{B}_\gamma} \left[\sin(2\beta)A(\gamma) + \sin(4\beta)B(\gamma) + \sin^2(2\beta)C(\gamma) \right], \quad (43)$$

where the set $\mathcal{B}_\gamma$ is defined as:

$$\mathcal{B}_\gamma = \left\{ \beta = \pm \frac{1}{2} \arccos(x_i) \mid i = 1, 2, 3, 4 \right\}. \quad (44)$$

and x_i are the real roots (within $[-1, 1]$) of the quartic polynomial:

$$\begin{aligned} p(x) = & \left[16B^2(\gamma) + 4C^2(\gamma) \right] x^4 \\ & + 8A(\gamma)B(\gamma)x^3 \\ & + \left[A^2(\gamma) - 16B^2(\gamma) - 4C^2(\gamma) \right] x^2 \\ & - 4A(\gamma)B(\gamma)x + 4B^2(\gamma). \end{aligned} \quad (45)$$

The functions $A(\gamma)$, $B(\gamma)$, and $C(\gamma)$, defined over the graph $G = (V, E)$ corresponding to the Hamiltonian H_P , are as follows:

$$\begin{aligned} A(\gamma) &= \sum_{i \in V} h_i \sin \gamma'_i \prod_{k \in \mathcal{N}(i)} \cos \gamma'_{ik}, \\ B(\gamma) &= \sum_{\{u, v\} \in E} \frac{J_{uv}}{2} \sin \gamma'_{uv} \left(\cos \gamma'_v \prod_{w \in e} \cos \gamma'_{vw} \right. \\ &\quad \left. + \cos \gamma'_u \prod_{w \in d} \cos \gamma'_{uw} \right), \\ C(\gamma) &= - \sum_{\{u, v\} \in E} \frac{J_{uv}}{2} \prod_{\substack{w \in e \\ w \notin F}} \cos \gamma'_{vw} \prod_{\substack{w \in d \\ w \notin F}} \cos \gamma'_{uw} \\ &\quad \times \left(\sum_{\chi = \pm 1} \chi \cos(\gamma'_u + \chi \gamma'_v) \prod_{f \in F} \cos(\gamma'_{uf} + \chi \gamma'_{vf}) \right), \end{aligned}$$

where $\gamma'_{uv} = 2J_{uv}\gamma$ and $\gamma'_i = 2h_i\gamma$.

The proof of this theorem is provided in section E.2. Although the line search method in theorem 6 is slightly more computationally intensive than the approach in theorem 5, as it requires solving for β^* at each $\gamma \in \mathbb{R}$, it remains significantly more efficient than a full grid search. Specifically, β^* is determined by solving a quartic polynomial, which can be computed either using known analytical solutions [70] or efficiently with numerical solvers. Since the resulting optimal path and cost landscape closely resemble those shown in fig. 2 and fig. 3, additional visualisations are omitted for brevity.

We note that any Ising model with fields can be transformed into an equivalent model without fields using the transformation in eq. (6), enabling the use of theorem 5 for simpler optimisation. While this transformation is straightforward, it is not always efficient. In certain cases—especially for problems where the graph representation has a bounded degree, with the maximum degree much smaller than the number of vertices—transforming the model can actually increase the complexity of finding the optimal parameters. The following proposition illustrates this with an example.

Proposition 1. *Let G be a triangle-free graph with n vertices and a maximum degree $D_{\max}$, where $D_{\max} \ll n$. Assume that all edge and node weights are ± 1 . When the corresponding Ising model with external fields is transformed into an equivalent model without external fields, the sampling period Δ_γ for QAOA₁ is reduced by a factor of $\frac{n}{D_{\max}+1}$.*

The proof of this proposition is provided in section F. Proposition 1 shows that for bounded-degree graphs with $D_{\max} \ll n$, eliminating external fields can substantially worsen the sampling requirements for QAOA₁. In contrast, for models such as the Sherrington-Kirkpatrick model, the effect on optimisation complexity is minor because the maximum degree increases by only 1. Nevertheless, even Hamiltonian-equivalent formulations can yield different approximate QAOA₁ solutions due to the algorithm's locality constraints.

3.5 Optimising the Univariate Function

In this section, we focus on optimising the variable γ for the univariate functions defined in the-

orem 5 and theorem 6. Consider a univariate trigonometric polynomial of degree K defined in eq. (30). Polynomial-time algorithms exist for computing all roots of such trigonometric polynomials [71–73], enabling the efficient determination of their optimal values. However, these methods are not directly applicable to our problem for two main reasons. First, eq. (40) includes a square root, which prevents it from being expressed in the standard trigonometric polynomial form of eq. (30). Second, eq. (43) involves two variables, (γ, β) , where the optimal β_γ^* depends on γ . This dependency prevents us from applying univariate optimisation methods that assume a fixed parameter.

To address these challenges, we perform a grid search over the parameter space (γ, β) , focusing exclusively on pairs of the form (γ, β_γ^*) . Specifically, for each γ , we consider only the optimal β_γ^* that minimises the objective function. This approach effectively reduces the problem to a univariate optimisation task, enabling the application of established algorithms such as Hansen’s subdivision algorithm based on interval analysis [74, 75].

The subdivision algorithm iteratively partitions the domain into progressively smaller intervals. In each iteration, the algorithm evaluates the midpoint of every interval to identify the presence of global minima using a rejection criterion. Intervals that are unlikely to contain the global minima are discarded, while those that may still contain them are further subdivided. This process continues until the size of each interval falls below a predefined tolerance value. At this point, the midpoints of all remaining intervals are computed, and the midpoint with the maximum value is returned. For trigonometric polynomials, analogous subdivision methods have been developed to estimate their maximum modulus in both univariate and multivariate settings, utilising Steckin’s lemma as the rejection criterion [76, 77].

By analytically determining β_γ^* for each γ , we effectively transform the original multivariate optimisation problem into a univariate one. This reduction allows us to apply efficient subdivision techniques tailored for univariate trigonometric functions. The specific subdivision algorithm used for computing the optimum modulus of univariate trigonometric functions, as at-

Algorithm 1: Subdivision Algorithm for Estimating the Optimal Angle γ^*

Input: Initial Interval Spacing $\Delta_\gamma^{(0)}$,
Maximum Frequency $\nu_{\max}$,
Tolerance ϵ

Output: Near-Optimal Angle γ^*

```

1  $\Delta_\gamma \leftarrow \Delta_\gamma^{(0)}$ ;
2 Initialise list of intervals:

    $\mathcal{I} \leftarrow \{[0, \Delta_\gamma], [\Delta_\gamma, 2\Delta_\gamma], \dots, [\pi - \Delta_\gamma, \pi]\}$ ;

3 repeat
    // Initialise empty list for
    // current subdivisions
4    $\mathcal{I}_{\text{current}} \leftarrow \emptyset$ ;
5    $\tilde{q} \leftarrow -\infty$ ;
    // Evaluate the objective
    // function at midpoints of all
    // active intervals
6   foreach  $I \in \mathcal{I}$  do
7      $t_I \leftarrow$  midpoint of  $I$ ;
8     Compute  $q(t_I)$ ;
9     if  $q(t_I) > \tilde{q}$  then
10       $\tilde{q} \leftarrow q(t_I)$ ;
11    end
12  end
    // Prune intervals not meeting
    // the threshold and prepare for
    // subdivision
13  foreach  $I \in \mathcal{I}$  do
14     $t_I \leftarrow$  midpoint of  $I$ ;
15    if  $q(t_I) \geq \tilde{q} \cos(\nu_{\max} \Delta_\gamma)$  then
16      Split  $I$  into  $I_{\text{left}}$  and  $I_{\text{right}}$ ;
17      Add  $I_{\text{left}}$  and  $I_{\text{right}}$  to  $\mathcal{I}_{\text{current}}$ ;
18    end
19  end
    // Update intervals and spacing
    // for the next iteration
20   $\mathcal{I} \leftarrow \mathcal{I}_{\text{current}}$ ;
21   $\Delta_\gamma \leftarrow \frac{\Delta_\gamma}{2}$ ;
22 until  $\tilde{q}(\sec(\nu_{\max} \Delta_\gamma) - 1) < \epsilon$ ;
23  $\gamma^* \leftarrow \arg \max_{I \in \mathcal{I}} q(t_I)$ ;
24 return  $\gamma^*$ 

```

tributed to Green [76], is detailed in algorithm 1. For a detailed explanation on using Steckin’s lemma as a rejection criterion, refer to section G.

To apply algorithm 1 to our expressions

eq. (40) and eq. (43), we square the expressions. Specifically, the function q in algorithm 1 is defined as $\langle \gamma, \beta_\gamma^* | H_P | \gamma, \beta_\gamma^* \rangle^2$. For the case of eq. (43), it is crucial that $\beta_\gamma^* \in \mathcal{B}_\gamma$ is the value that minimises the objective function before squaring it. Additionally, since all γ terms have a constant prefactor of 2, the domain that the subdivision algorithm searches is effectively reduced from $[0, 2\pi]$ to $[0, \pi]$. Algorithm 1 maintains a worst-case time complexity of $\mathcal{O}(N/\sqrt{\epsilon})$ for estimating the maximum modulus and consequently the optimum of eq. (30) (see theorem 12).

We note that there exist other polynomial-time algorithms for computing the provably optimal (γ^*, β^*) angles for QAOA₁ on Ising models. Specifically, to determine the optimal (γ^*, β^*) that minimise $\langle H_P(\gamma, \beta) \rangle$, we solve the system of equations:

$$\frac{\partial \langle H_P(\gamma, \beta) \rangle}{\partial \gamma} = \frac{\partial \langle H_P(\gamma, \beta) \rangle}{\partial \beta} = 0. \quad (46)$$

This involves expressing eq. (46) as standard polynomials and applying algorithms designed to find all roots of bivariate polynomials in polynomial time [78, 79]. Although these methods are more computationally intensive than Algorithm 1, they illustrate that optimising QAOA₁ parameters need not be NP-Hard under certain conditions, such as when γ is effectively periodic in $[0, \pi]$ and the distribution of problem weights does not scale with the problem size.

This stands in contrast to the results of Bittel et al. [7], who showed that optimising the parameters of VQAs, and QAOA in particular, is NP-Hard. This hardness, however, is not a property of arbitrary Ising Hamiltonians; it is established through explicit reductions from MaxCut, in which carefully tailored problem instances are constructed so that an efficient parameter optimiser would necessarily solve an NP-Hard problem. For QAOA₁, such instances fall into one of two regimes. The first is when the problem weights are incommensurable, rendering the expectation value along γ non-periodic, so that no finite period can be exploited to confine the search to a bounded interval. The second is when the weights grow with the problem size: their single-layer reduction employs a spectrum spanning exponentially many orders of magnitude, which forces the maximum frequency of the cost landscape—and hence the sampling resolution required to reconstruct it—to scale exponen-

tially, demanding exponential precision to resolve the global optimum. A related result shows that QAOA_n remains NP-Hard even for bounded integer weights, but only at the expense of a circuit depth that grows with the number of variables. Crucially, none of these worst-case regimes apply to the setting considered here. For QAOA₁ on Ising models with commensurable weights that do not scale exponentially with the system size, the expectation value along γ is periodic on $[0, \pi]$ and bandlimited with a polynomially bounded maximum frequency, so that the optimal angles (γ^*, β^*) can be determined in polynomial time. Consequently, these instances of QAOA₁ do not fall under the NP-Hard regime identified by Bittel et al. [7].

3.6 Globally Optimal Parameters for Regular Graphs

We present our final theoretical result: in large, dense, and heavily weighted Ising model instances, the globally optimal parameter $\gamma^* \in \mathbb{R}^+$ tends to concentrate near zero, often coinciding with the first local optimum. Consequently, we propose simplifying the optimisation process by performing gradient descent in the vicinity of zero to determine γ^* . This approach is motivated by empirical observations showing that the first local optimum near zero typically yields globally optimal parameters [37, 80, 81].

We rigorously prove that for the family of regular graphs—including triangle-free regular graphs, complete graphs, and all intermediate cases—the optimal γ^* in the infinite size limit sharply concentrates near zero. This concentration depends on the graph’s degree, the second moments of the weight distribution, and the variance of the weights. The following theorem formalises this result. By assuming that $\gamma \propto D^{-1/2}$, we derive closed-form expressions for numerically evaluating the scaled expectation cost of QAOA₁ and obtain expressions for the optimal γ^* based on the leading-order terms of the expectation value.

Theorem 7. *Consider a weighted Ising model defined on a $(D + 1)$ -regular graph $G = (V, E)$. Let the field strengths $\{h_i\}_{i \in V}$ and coupling strengths $\{J_{uv}\}_{\{u,v\} \in E}$ be i.i.d. random variables with finite second moments. For any distinct vertices $u, v \in V$, let $|\overline{F}_{uv}| = aD^\lambda$ and $|F_{uv}| = bD^\mu$,*

where $a, b \geq 0$ and $0 \leq \lambda, \mu \leq 1$. Here, $\bar{F}_{uv}$ consists of neighbours of $\{u, v\}$ that do not form triangles, while F_{uv} consists of those that do. Then,

$$\frac{\mathbb{E}[\xi_G(\alpha, \beta)]}{|E|} = \begin{cases} \mathcal{C}_1(\alpha, \beta) + \mathcal{O}(D^{-1}), & b = 0, a = \lambda = 1, \\ \mathcal{C}_1(\alpha, \beta) + \mathcal{C}_2(\alpha, \beta, 1, 1) + \mathcal{O}(D^{-1}), & a = 0, b = \mu = 1, \\ \mathcal{C}_1(\alpha, \beta) + \mathcal{C}_2(\alpha, \beta, 1, b) + \mathcal{O}(D^{-1}), & a, b > 0, \lambda = \mu = 1, \\ \mathcal{C}_1(\alpha, \beta) + \mathcal{C}_2(\alpha, \beta, b, b) + \mathcal{O}(D^{-1}), & a, b > 0, \mu = 1, \lambda < \frac{1}{2}, \\ \mathcal{C}_1(\alpha, \beta) + \mathcal{C}_2(\alpha, \beta, b, b) \left(1 - 4a\alpha^2 \mathbb{E}[J^2] \frac{1}{\sqrt{D}}\right) + \mathcal{O}\left(D^{\max(2\lambda-2, -1)}\right), & a, b > 0, \mu = 1, \lambda = \frac{1}{2}, \\ \mathcal{C}_2(\alpha, \beta, b, b) \left(1 - 4a\alpha^2 \mathbb{E}[J^2] D^{\lambda-1}\right) + \mathcal{O}\left(\frac{1}{\sqrt{D}}\right), & a, b > 0, \mu = 1, \lambda > \frac{1}{2}, \\ 4b\alpha^2 \sin^2(2\beta) \mathbb{E}[J]^3 e^{-4a\alpha^2 \mathbb{E}[J^2]} D^{\mu-1} + \mathcal{O}\left(D^{\max(-\frac{1}{2}, 2\mu-2)}\right), & a, b > 0, \lambda = 1, \mu > \frac{1}{2}, \\ \mathcal{C}_1(\alpha, \beta) + 4b\alpha^2 \sin^2(2\beta) \mathbb{E}[J]^3 e^{-4a\alpha^2 \mathbb{E}[J^2]} \frac{1}{\sqrt{D}} + \mathcal{O}(D^{-1}), & a, b > 0, \lambda = 1, \mu = \frac{1}{2}, \\ \mathcal{C}_1(\alpha, \beta) + \mathcal{O}(D^{\mu-1}), & a, b > 0, \lambda = 1, \mu < \frac{1}{2}, \end{cases} \quad (47)$$

where

$$\mathcal{C}_1(\alpha, \beta) = 2\alpha \sin(4\beta) \mathbb{E}[J^2] e^{-2\alpha^2 \mathbb{E}[J^2]} \frac{1}{\sqrt{D}}, \quad (48)$$

$$\begin{aligned} \mathcal{C}_2(\alpha, \beta, \theta_1, \theta_2) &= \sin^2(2\beta) \mathbb{E}[J] e^{-4\theta_1 \alpha^2 \mathbb{E}[J^2]} \\ &\quad \times \sinh\left(4\theta_2 \alpha^2 \mathbb{E}[J]^2\right). \end{aligned} \quad (49)$$

For leading-order terms, the optimal value of $\gamma^* \in \mathbb{R}^+$ is:

$$\gamma^* = \begin{cases} \eta(4), & b = 0, a = \lambda = 1, \\ \zeta(1, 1), & a = 0, b = \mu = 1, \\ \zeta(1, b), & a, b > 0, \lambda = \mu = 1, \\ \zeta(b, b), & a, b > 0, \mu = 1, \lambda < 1, \\ \eta(2a), & a, b > 0, \lambda = 1, \mu > \frac{1}{2}, \\ \eta(4), & a, b > 0, \lambda = 1, \mu < \frac{1}{2}, \end{cases} \quad (50)$$

where

$$\eta(\theta) = \frac{1}{\sqrt{\theta D \mathbb{E}[J^2]}}, \quad (51)$$

$$\zeta(\theta_1, \theta_2) = \sqrt{\frac{1}{8\theta_2 D \mathbb{E}[J]^2} \ln\left(\frac{\theta_1 \mathbb{E}[J^2] + \theta_2 \mathbb{E}[J]^2}{\theta_1 \mathbb{E}[J^2] - \theta_2 \mathbb{E}[J]^2}\right)}. \quad (52)$$

Finally, the globally optimal parameter γ^* also corresponds to the first local optimum for $\gamma \in \mathbb{R}^+$.

The proof of this theorem is provided in section H.3. The assumption that $\gamma \in \Theta(D^{-1/2})$ in theorem 7 is motivated by numerical observations showing that the optimal γ for the unweighted MaxCut problem scales as $\Theta(D^{-1/2})$ (see Figure 1b of Boulebnane et al. [82]) and by analytical results demonstrating that, for finite-size

the scaled expected QAOA₁ cost function with respect to parameters $\beta \in \mathbb{R}$ and $\gamma = \alpha/\sqrt{D} \in \mathbb{R}^+$ can be expressed as follows:

triangle-free regular graphs with weights drawn from an exponential distribution, γ also scales as $\Theta(D^{-1/2})$ (see theorem 1 of Sureshbabu et al. [83]).

We note that Sureshbabu et al. [83] recently proved a special case of our result for triangle-free regular graphs without external fields. The increased complexity of theorem 7 compared to their result arises because, in general, the exact number of triangles an edge belongs to is not known, except in specific cases such as triangle-free graphs and complete graphs. To address this, we introduce parameters to quantify both the number of triangles an edge participates in and the number of its neighbouring edges that are not part of any triangle. This parametrisation naturally recovers triangle-free and complete graphs as limiting cases. The first two cases in eq. (47) correspond to these two extremes, while the remaining cases handle intermediate scenarios where the graph is neither triangle-free nor complete. Notably, when edge weights are drawn from a distribution with zero mean, eq. (47) simplifies significantly. Specifically, for $\mathbb{E}[J] = 0$, it reduces to $\mathcal{C}_1(\alpha, \beta)$ with the globally optimal $\gamma^* = \eta(4)$. Finally, since theorem 7 does not assume the sampled weights are commensurable, our finding that the first local optimum often coincides with the global optimum remains valid even when the γ landscape is non-periodic.

While theorem 7 establishes that, in the infinite-size limit of regular graphs, the global optimum coincides with the first local optimum, empirical evidence from finite-size problems [37, 80, 81] suggests that this alignment frequently holds

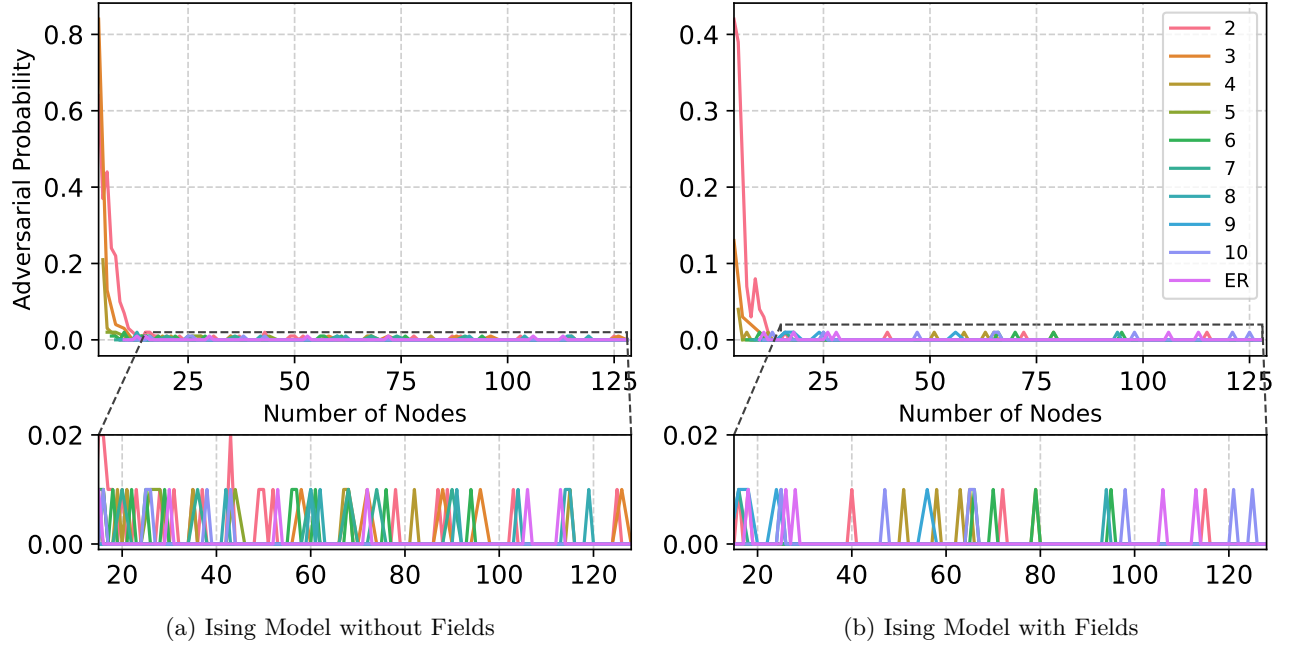


Figure 4: **Probability of Adversarial Graphs in QAOA₁**. These plots illustrate the probability of encountering adversarial instances in Ising models for D -regular graphs ($2 \leq D \leq 10$) and Erdős-Rényi (ER) graphs with edge probability $1/2$. (a) presents results for Ising models without external fields, whereas (b) illustrates results for models with external fields. Graph sizes range from 4 to 128 nodes, with edge weights sampled as integers from a Gaussian distribution (mean 0, variance 100). For each configuration, 100 random graph instances were evaluated using two methods: (1) a full line search over $\gamma \in [0, \pi]$ using algorithm 1, and (2) gradient descent initialised near $\gamma = \Delta_\gamma/2$. Adversarial instances are defined as those where the optima obtained from the two methods differ. The probability of adversarial instances stabilises around 0.01 for graphs with $n > 14$, implying that only 1 in 100 graphs exhibits adversarial behaviour at larger sizes. Smaller and sparser graphs show a higher occurrence of adversarial instances, reflecting the increased likelihood of misalignment between local and global optima in these cases.

in practice as well. We observe this phenomenon for finite-size D -regular graphs as well, as shown in fig. 2 and fig. 3d. This behaviour can be intuitively explained using the closed-form expressions in theorem 5 and theorem 6. When signals like $\cos(\alpha x)$ and $\cos(\beta x)$ with different frequencies interfere, they align constructively when in phase and destructively when out of phase, with the interference points depending on their oscillation rates. Similarly, the univariate cost function for QAOA₁ along γ comprises cosine terms with frequencies determined by the problem instance. As the instance size, density, or weights increase, the range of frequencies broadens, creating a superposition of oscillating signals. At arbitrary γ , these signals are generally out of phase, leading to destructive interference and a reduced cost function amplitude. Near $\gamma = 0$, however, all cosine terms converge to $\cos(0) = 1$, aligning perfectly in phase and maximising constructive interference, which produces the highest amplitude. As γ devi-

ates from zero, the signals fall out of phase, causing destructive interference and diminishing the amplitude.

Nevertheless, it is important to recognise that the first local optimum need not always correspond to the global optimum, as adversarial graphs can be constructed where this alignment does not hold (see Figure 2 of [37]). However, if the majority of finite instances exhibit the behaviour where the first local optimum aligns with the global optimum, the optimisation process over $\gamma \in [0, \pi]$ with interval size Δ_γ can be significantly simplified. To evaluate the prevalence of adversarial instances, we tested weighted D -regular graphs ($2 \leq D \leq 10$) and Erdős-Rényi graphs (with edge probability $1/2$) for graph sizes ranging from 4 to 128, generating 100 random instances per graph type. Edge weights were sampled from a Gaussian distribution with mean 0 and variance 100, rounded to the nearest integer. Two optimisation methods were employed: a line

search over $\gamma \in [0, \pi]$ using algorithm 1 with Δ_γ determined by theorem 4, and gradient descent initialised at $\gamma = \Delta_\gamma/2$ to provide an initial guess that avoids overshooting the first local optimum. By comparing the γ^* values obtained from both methods, we found that adversarial instances are rare. Their occurrence is concentrated in small, sparse graphs and decreases exponentially with increasing graph size or density, as illustrated in fig. 4.

These empirical findings lay the groundwork for streamlining the optimisation process of QAOA₁ on large Ising models. Specifically, the comprehensive line search over $\gamma \in [0, \pi]$ can be replaced with a gradient descent approach initialised near $\gamma \approx 0$, guided by the maximum sampling interval Δ_γ from theorem 4. This parameter serves a dual purpose: it provides an initial guess, $\gamma < \Delta_\gamma$, placing the algorithm close to the anticipated global optimum while preventing overshooting, and it establishes an upper bound on the gradient descent step size. By ensuring that updates remain smaller than Δ_γ , the algorithm operates within a region where the cost function behaves predictably, thereby minimising the risk of overshooting the first local optima.

4 Numerical Results

To validate our parameter tuning strategy for QAOA₁ on Ising models, we applied it to Recursive QAOA₁ (RQAOA₁) [48] (refer to section 6.1 for an explanation of RQAOA). We focused on large, dense, weighted problem instances to assess the effectiveness of our approach. Since any QUBO problem can be mapped to finding the ground state of an arbitrary Ising model—represented as an undirected graph with nodes as spins and edges as interaction terms—demonstrating that RQAOA₁ performs well across a broad distribution of graphs implies its applicability to most QUBO problems. Given the impracticality of testing RQAOA₁ on all possible graph structures, we utilised Erdős-Rényi graphs [84, 85] as a representative approximation. Problem instances were generated using random Erdős-Rényi graphs with 128 and 256 vertices, varying the edge probability $p \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. The edge weights were sampled from a Gaussian distribution with a mean of 50 and a variance of 30, then rounded to

the nearest integer. For Ising models that include external fields, the external field values were similarly sampled from a Gaussian distribution (mean 40, variance 20) and rounded to integers. For each combination of graph size and edge probability, we generated 20 random instances for both Ising models with and without external fields.

The numerical simulations of QAOA₁ and RQAOA₁ leveraged analytical solutions derived from theorem 5 and theorem 6. To assess performance, we computed approximation ratios by comparing the objective value achieved by each algorithm considered in our benchmark with the corresponding reference value returned by the GUROBI solver [86], an industry-standard optimisation tool renowned for its efficiency and reliability. Each problem instance was solved with GUROBI using a fixed wall-clock budget. In many cases, the incumbent objective stabilised before termination, with no further improvement in the best-found solution observed while the MIP gap³ continued to decrease slowly. When the final MIP gap was zero, the returned solution was certified optimal; otherwise, we used the best incumbent solution at termination and report the associated final MIP gap. Thus, the approximation ratios discussed in this section are computed relative to GUROBI reference values, with exact optimality certified whenever the final MIP gap is zero. In the remainder of this section, we present detailed benchmark results for the Ising models, highlighting the performance and effectiveness of our parameter tuning strategy in optimising QAOA₁ and RQAOA₁.

4.1 Ising Models without Fields

To benchmark RQAOA₁ on Ising models without fields, we compared its performance against four other algorithms: QAOA₁, two variants of RQAOA₁, and a classical semidefinite program (SDP). QAOA₁ parameters were optimised using gradient descent initialised near $\gamma \approx 0$ on eq. (40) and is labelled as QAOA₁ in the results. RQAOA₁ was tested in two configurations: RQAOA₁ (20), which used a coarse line search of 20 samples for $\gamma \in [0, \pi]$ followed by

³The MIP Gap refers to the relative difference between the best integer solution and the best possible bound on the optimal solution in mixed-integer programming. Smaller MIP Gaps indicate solutions closer to optimality.

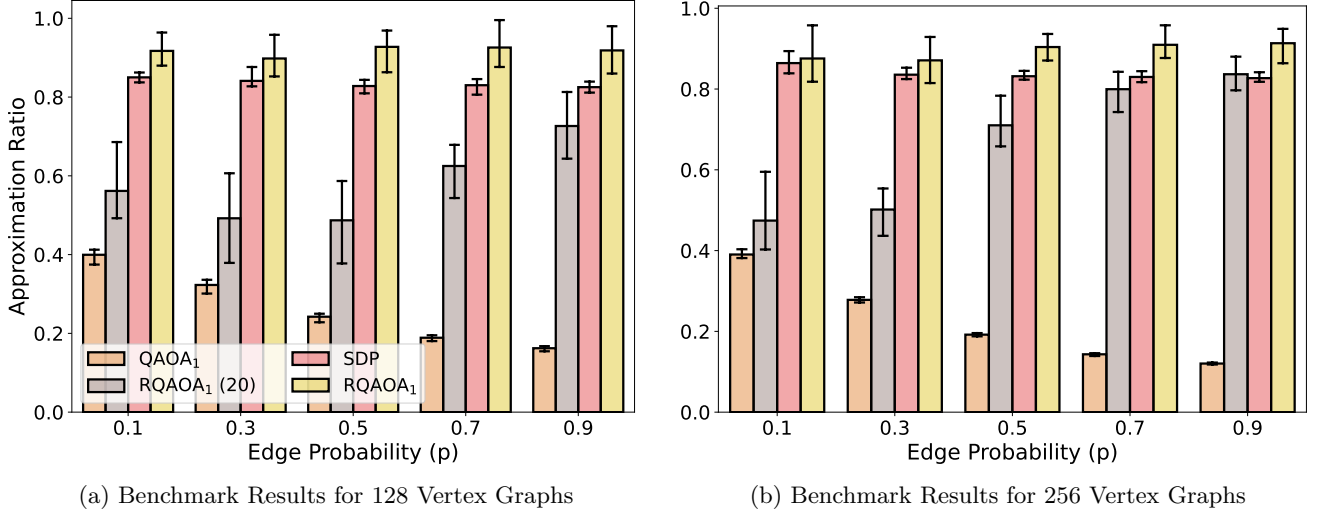


Figure 5: **Benchmark Results for Ising Models Without External Fields.** These plots compare four algorithms—QAOA₁, RQAOA₁, RQAOA₁(20), and a classical SDP—for approximating the ground state of the Ising model without external fields. We generate weighted Erdős–Rényi graphs with varying edge probabilities p for (a) 128-vertex and (b) 256-vertex instances. Each bar indicates the *mean* approximation ratio over 20 random graphs at a given p ; the whiskers do not denote standard deviations, but instead span the *minimum* and *maximum* ratios observed across those instances. QAOA₁ is optimised via gradient descent initialised near $\gamma \approx 0$; RQAOA₁(20) employs a coarse line search of 20 samples in $[0, \pi]$ followed by gradient descent on the best found point; RQAOA₁ uses gradient descent from $\gamma \approx 0$ with $\eta = 120$ recursive steps for 128-vertex graphs and $\eta = 248$ for 256-vertex graphs. The SDP adapts the Goemans–Williamson relaxation by replacing the MaxCut cost function with the Ising Hamiltonian, and each solution is rounded from 1024 random hyperplanes.

gradient descent on the best found point, and RQAOA₁, which employed gradient descent initialised near $\gamma \approx 0$ on eq. (40). Both variants were tested with $\eta = 120$ and $\eta = 248$ recursive steps for 128- and 256-vertex graphs, respectively. The final comparison was with a classical SDP based on the Goemans–Williamson algorithm [87], adapted for Ising models by replacing the MaxCut cost function with the Ising model Hamiltonian. After solving the relaxed SDP, the solution was rounded using 1024 random hyperplanes.

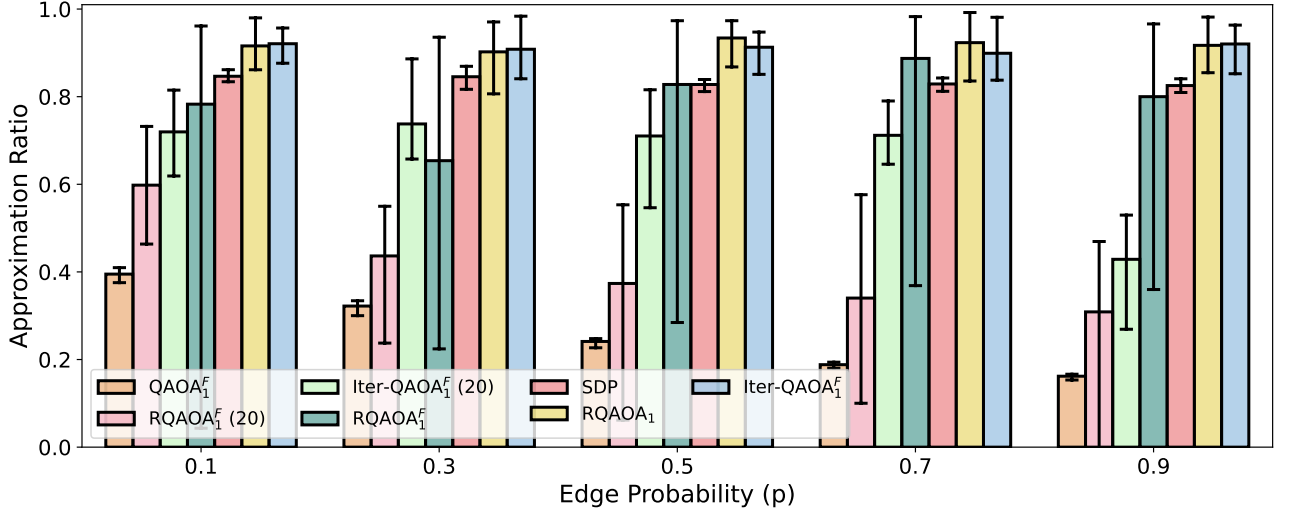
The benchmark results are summarised in fig. 5. QAOA₁ consistently yields the lowest approximation ratios among the tested algorithms. Both RQAOA₁ (20) and RQAOA₁ outperform QAOA₁ across all instances, while the SDP generally outperforms RQAOA₁ (20), often by a significant margin. One exception occurs on 256-node graphs with $p = 0.9$, where RQAOA₁ (20) slightly exceeds the performance of the SDP (fig. 5b). In contrast, RQAOA₁ consistently outperforms the SDP across all values of p for both graph sizes. These results underscore the effectiveness of the parameter tuning strategy outlined in the-

orem 5 and show that, with optimal parameters, RQAOA₁ can outperform classical SDPs, highlighting the critical role of parameter optimisation in VQAs.

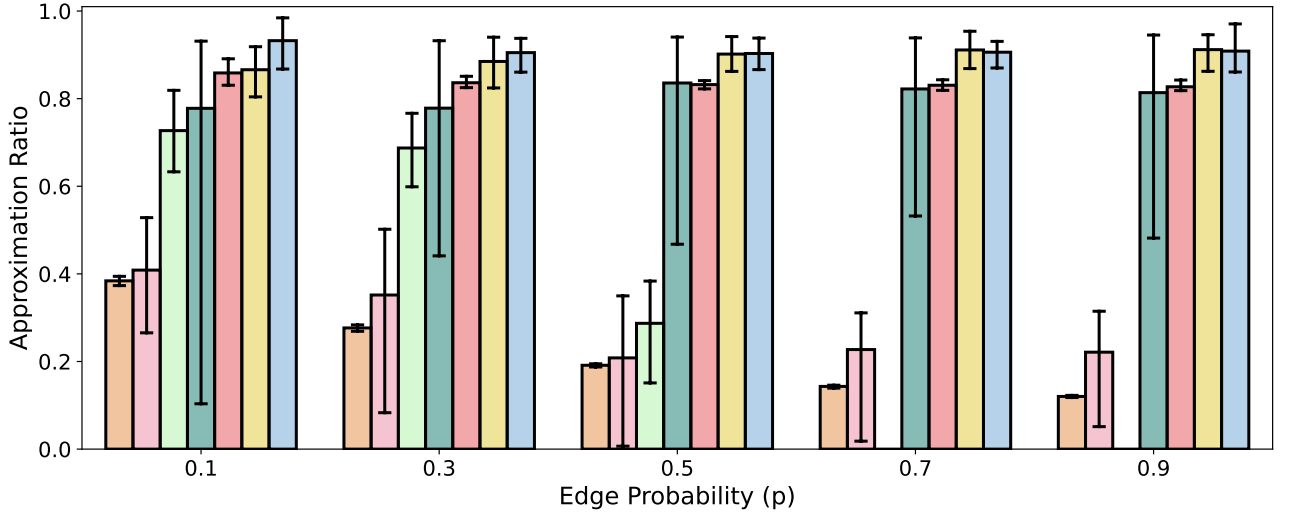
4.2 Ising Models with Fields

Benchmarking RQAOA₁ for Ising models with external fields revealed inconsistent performance, the underlying cause of which remains uncertain. To address this, we introduced a simplified iterative rounding procedure that demonstrated greater robustness and consistently produced better results. This modified approach, named Iterative-QAOA (Iter-QAOA), is detailed in section 6.2.

We benchmarked RQAOA₁ and Iter-QAOA₁ on Ising models with external fields, comparing their performance to seven algorithms: QAOA₁, three RQAOA₁ variants, two Iter-QAOA₁ variants, and a classical SDP. Algorithms labelled with a superscript F used the parameter expressions from theorem 6 for Ising models with fields, while those without F used theorem 5 for models without fields. For QAOA₁ ^{F} , parameters were optimised via gradient descent initialised near $\gamma \approx 0$



(a) Benchmark Results for 128 Vertex Graphs



(b) Benchmark Results for 256 Vertex Graphs

Figure 6: Benchmark Results for Ising Models with External Fields. These plots compare the performance of seven algorithms for approximating the ground state of the Ising model with external fields. We generate weighted Erdős–Rényi graphs with varying edge probabilities p , shown for (a) 128-vertex and (b) 256-vertex instances. Each bar indicates the *mean* approximation ratio over 20 random graphs at a given p ; the whiskers do not denote standard deviations, but instead span the *minimum* and *maximum* ratios observed across those instances. Algorithms labelled with a superscript F (e.g., QAOA_1^F) employ parameter expressions tailored to Ising models with fields; those without F (e.g., RQAOA_1) convert the model to a field-free form before parameter optimisation. QAOA_1^F is initialised near $\gamma \approx 0$ for gradient descent; $\text{RQAOA}_1^F(20)$ and $\text{Iter-QAOA}_1^F(20)$ each use a coarse line search of 20 samples in $[0, \pi]$ prior to gradient-descent refinement. RQAOA and Iter-QAOA variants are run for $\eta = 120$ steps (128-vertex) or $\eta = 248$ steps (256-vertex). The classical SDP adapts the Goemans–Williamson relaxation by replacing the MaxCut cost function with the Ising Hamiltonian, and each SDP solution is rounded from 1024 random hyperplanes.

on eq. (43). RQAOA_1^F was tested in two configurations: $\text{RQAOA}_1^F(20)$, which employed a coarse line search with 20 samples for $\gamma \in [0, \pi]$ followed by gradient descent on the best found point, and RQAOA_1^F , which directly used gradient descent initialised near $\gamma \approx 0$. RQAOA_1 , by contrast, converted the Ising model with fields to

one without fields and optimised parameters using QAOA_1 . Similarly, Iter-QAOA_1^F was tested in two configurations: $\text{Iter-QAOA}_1^F(20)$, which followed the same coarse line search strategy as $\text{RQAOA}_1^F(20)$, and Iter-QAOA_1^F , which used gradient descent initialised near $\gamma \approx 0$. All RQAOA and Iter-QAOA variants were run for

$\eta = 120$ and $\eta = 248$ recursive steps on 128- and 256-vertex graphs, respectively. Finally, the SDP was benchmarked on equivalent Ising models without fields. The results, shown in fig. 6, highlight the relative performance of each algorithm.

As shown in fig. 6, Iter-QAOA₁^F consistently outperforms RQAOA₁^F across all instances. This is reflected not only in the higher mean approximation ratios, but also in the generally narrower min-max whiskers, which indicate more stable performance across all the random instances and are consistent with reduced sensitivity to parameter tuning errors, particularly on sparser graphs. In contrast, RQAOA₁^F shows higher variability, with worst-case results occasionally falling below both its coarsely optimised counterpart and QAOA₁^F. Iter-QAOA₁^F also outperforms the SDP by a significant margin and is competitive with RQAOA₁, occasionally surpassing it. Similar trends are observed for the coarsely optimised variants: Iter-QAOA₁^F (20) generally performs better than RQAOA₁^F (20), except in 256-vertex cases with edge probabilities $p \in \{0.7, 0.9\}$. In those instances, Iter-QAOA₁^F (20) produces negative approximation ratios—worse than random guessing—likely due to the use of partial information in these cases.

These results highlight limitations in RQAOA’s performance on Ising models with fields, raising questions about its applicability to more general pseudo-Boolean optimisation problems with higher-degree terms. This motivates further exploration of whether RQAOA and related quantum algorithms [48, 52, 57, 88, 89] can effectively generalise beyond QUBO. Nonetheless, Iter-QAOA₁^F consistently outperformed the SDP across all benchmarks, underscoring the robustness and efficacy of our parameter-setting strategy outlined in theorem 6, which was the central focus of this particular study.

5 Discussion

In this work, we focused on optimising QAOA₁ for arbitrary Ising models, specifically addressing the challenges and misconceptions surrounding its parameter optimisation. Although QAOA₁ is governed by only two parameters, (γ, β) , it is commonly assumed that a simple coarse grid search followed by local minimisation is sufficient

for finding optimal values. Contrary to this belief, our analysis revealed that the expectation value landscape is highly oscillatory. By expressing it as a partial Fourier series, we demonstrated that the oscillations along γ are heavily dependent on the specific problem instance. We derived the maximum frequency of these oscillations for Ising models both with and without external fields and employed the Nyquist-Shannon sampling theorem to determine the minimum sampling resolution necessary for accurately reconstructing the landscape.

To overcome the inefficiency of traditional grid searches over (γ, β) , we streamlined the two-dimensional optimisation to a one-dimensional line search over γ and analytically determined the optimal β^* , thereby simplifying the search process. We then introduced Green’s subdivision algorithm [76] for computing the maximum modulus of univariate trigonometric polynomials, which aids in estimating the optimal γ^* angle. Furthermore, we proved that in the infinite-size limit for Ising models with fields on regular graphs, the global optimum for $\gamma \in \mathbb{R}^+$ tends to concentrate near zero and coincides with the first local optimum. Supporting this theoretical result, our empirical evidence demonstrates that the global optimum typically aligns with the first local optimum along $\gamma \in \mathbb{R}^+$ even for finite-size instances. Leveraging this insight, we refined the optimisation process further by initiating gradient descent near zero, thereby eliminating the need for a comprehensive line search over γ .

We validated our parameter tuning strategies by applying them to RQAOA₁ and benchmarking its performance on large, dense, and weighted problem instances—scenarios where naive grid searches and traditional heuristics often falter. Our benchmarks encompassed Erdős-Rényi graphs with 128 and 256 vertices, varying densities, and weight configurations. For Ising models without external fields, RQAOA₁ optimised using a naive coarse grid search followed by local minimisation did not surpass the performance of the SDP. However, when optimised using our proposed parameter tuning strategy, RQAOA₁ consistently outperformed both the coarse search approach and the SDP across all tested graph instances. In contrast, for Ising models with external fields, RQAOA₁ exhibited inconsistent performance, the underlying causes of which remain

unclear⁴. To overcome this limitation, we developed Iter-QAOA, which demonstrated greater robustness and consistently outperformed the SDP for Ising models with fields. Although the diminished performance of RQAOA₁ in the presence of fields raises questions about its applicability to more general binary optimisation problems, addressing these concerns is beyond the scope of the current work and will be explored in future research.

In summary, our work presents near-optimal—if not optimal—parameter tuning strategies for QAOA₁ applicable to arbitrary Ising models, achievable in polynomial time. We have rigorously validated these strategies on large, dense, and heavily weighted problem instances, demonstrating their effectiveness and robustness. Unlike previously proposed heuristics, our approach does not rely on assumptions about the problem’s structure, symmetry, or weight distribution. Moreover, it achieves near-optimal performance with minimal computational overhead. Consequently, we assert that our work effectively resolves the challenge of efficiently finding globally near-optimal QAOA₁ parameters for arbitrary QUBO problems.

6 Methods

6.1 Recursive QAOA (RQAOA)

Recursive QAOA (RQAOA) is a non-local variant of QAOA that uses shallow-depth QAOA as a subroutine to recursively reduce the problem size until it becomes trivial to solve via brute force. RQAOA₁, which specifically employs QAOA₁ as its subroutine, has demonstrated superior performance compared to standalone QAOA₁ and is competitive with state-of-the-art classical algorithms [48, 88, 91]. Additionally, RQAOA₁ provides performance guarantees for specific graph structures [48, 92, 93] and has been explored in various contexts [60, 89, 94–96].

At the core of RQAOA is correlation rounding, a technique that identifies the qubit or pair of qubits with the largest absolute expectation

⁴This instability may not be unique to the present setting: recent results on the Binary Paint Shop Problem likewise reported that RQAOA₁, despite using provably optimal QAOA₁ parameters at each recursive step, showed diminishing performance as the problem size increased [90].

Algorithm 2: RQAOA for Ising Models

Input: Problem Hamiltonian C_0 with n variables $V_0 = \{1, \dots, n\}$, recursion depth η

Output: Approximate ground state of C_0

```

1 for  $t \leftarrow 0$  to  $n - \eta$  do
2   Prepare QAOA state  $|\psi_t(\gamma, \beta)\rangle$  for  $C_t$ ;
3   Optimise parameters:
       $(\gamma^*, \beta^*) \leftarrow \arg, \min_{\gamma, \beta} \langle \psi_t(\gamma, \beta) | C_t | \psi_t(\gamma, \beta) \rangle$ ;
4   Compute  $M_i$  for  $i \in V_t$ :
       $M_i \leftarrow \langle \psi_t(\gamma^*, \beta^*) | Z_i | \psi_t(\gamma^*, \beta^*) \rangle$ ;
5   Identify the variable  $u \leftarrow \arg, \max_{i \in V_t} |M_i|$ ;
6   Compute  $M_{ij}$  for all  $\{i, j\} \in V_t$ :
       $M_{ij} \leftarrow \langle \psi_t(\gamma^*, \beta^*) | Z_i Z_j | \psi_t(\gamma^*, \beta^*) \rangle$ ;
7   Identify the pair  $\{u, v\} \leftarrow \arg, \max_{\{i, j\} \in V_t} |M_{ij}|$ ;
8   if  $|M_u| \geq |M_{uv}|$  then
9     Assign Value  $u : Z_u = \text{sign}(M_u)$ ;
10    Update Variable Set:  $V_{t+1} \leftarrow V_t \setminus \{u\}$ ;
11    Define  $h'_i \leftarrow h_i + \text{sign}(M_u) J_{ui}$ ;
12    Update Hamiltonian:
       $C_{t+1} \leftarrow \sum_{\{i, j\} \in V_{t+1}} J_{ij} Z_i Z_j + \sum_{i \in V_{t+1}} h'_i Z_i + \text{sign}(M_u) h_u$ ;
13  else
14    Impose Constraint:
       $Z_u = \text{sign}(M_{uv}) Z_v$ ;
15    Update Variable Set:  $V_{t+1} \leftarrow V_t \setminus \{u\}$ ;
16    Define  $J'_{uj} \leftarrow J_{uj} + \text{sign}(M_{uv}) J_{vj}$ ;
17    Define  $h''_u \leftarrow h_u + \text{sign}(M_{uv}) h_v$ ;
18    Update Hamiltonian:
       $C_{t+1} \leftarrow \sum_{\{i, j\} \in V_{t+1}} J_{ij} Z_i Z_j + \sum_{j \in V_{t+1}} J'_{uj} Z_u Z_j$ 
       $+ \sum_{i \in V_{t+1}} h_i Z_i + h''_u Z_u + \text{sign}(M_{uv}) J_{uv}$ ;
19  end
20 end
21 end
22 Solve  $C_t$  via brute-force when  $|V_t| \leq \eta$ ;
23 Solve  $C_0$  by backtracking;
24 return Approximate ground state of  $C_0$ 

```

or correlation to impose constraints (see algorithm 2). This approach shares similarities with iterative rounding [97, 98], a method commonly used in integer linear programming (ILP). In iterative rounding, the linear relaxation of an ILP

is solved, fractional variables are rounded, and the problem is updated iteratively. Typically, the variable with the largest absolute fractional value is selected, rounded to ± 1 , and the reduced problem is solved.

In contrast, correlation rounding shifts attention from single variables to the correlations between variable pairs, which are derived from the QAOA state. At each iteration, the algorithm identifies the variable u^* with the largest absolute single-qubit expectation $|M_{u^*}|$ and the pair $\{u^*, v^*\}$ with the largest absolute pairwise correlation $|M_{u^*v^*}|$. If the single-variable expectation $|M_{u^*}|$ is greater, u^* is fixed to $\text{sign}(M_{u^*})$. Otherwise, a correlation constraint is imposed: v^* is set relative to u^* based on the sign of $M_{u^*v^*}$. This process ensures that decisions are guided by the strongest available information, whether from individual variables or their correlations. Unlike iterative rounding, where variables are deterministically set immediately, correlation rounding delays assigning specific values to some variables until the reduced problem is small enough to solve via brute force. By focusing on pairwise interactions, it captures the problem's structure more effectively, especially in cases where interactions play a dominant role. This distinction is crucial for problems with strong quadratic terms, where correlations between variables carry more information than individual biases.

6.2 Iterative QAOA (Iter-QAOA)

The key distinction between Iter-QAOA and RQAOA lies in their treatment of linear and quadratic terms within the optimisation process. As outlined in algorithm 2, RQAOA evaluates the expectation values of both linear terms ($\langle Z_u \rangle$) and quadratic terms ($\langle Z_u Z_v \rangle$) at each recursive step for every node u and edge $\{u, v\}$. It selects the term with the largest absolute value and either assigns Z_u a value of ± 1 or enforces a correlation or anti-correlation constraint between Z_u and Z_v . In contrast, Iter-QAOA streamlines this process by utilising only the linear terms. It assigns values based solely on the signs of these linear terms, thereby eliminating the need to impose additional constraints.

At first glance, Iter-QAOA appears counter-intuitive and should, in theory, underperform compared to RQAOA since it relies on partial information for decision-making. For instance,

consider a scenario involving a node u and an edge $\{u, v\}$ where the magnitudes of the linear term $\langle Z_u \rangle$ and the quadratic term $\langle Z_u Z_v \rangle$ are the largest, with the condition $0 < h_u < J_{uv}$. Here, the field strength at node u is weaker than the coupling between nodes u and v . Ideally, the optimal angle should satisfy $|\langle Z_u \rangle| < |\langle Z_u Z_v \rangle|$. In such cases, RQAOA would impose the anti-correlation constraint $Z_u = -Z_v$, thereby reducing the energy by $-J_{uv}$. Conversely, Iter-QAOA, relying solely on the linear term, would assign $Z_u = -1$, achieving an energy reduction of $-h_u$ and missing the more substantial energy decrease from the coupling term.

However, our benchmark results indicate that Iter-QAOA₁ consistently outperforms RQAOA₁ across most instances where most nodes u and edges $\{u, v\}$ satisfy $h_u < J_{uv}$. In our tests, edge weights were sampled from a Gaussian distribution with a mean of 50 and a variance of 30, while node weights were sampled from a Gaussian distribution with a mean of 40 and a variance of 20, both rounded to the nearest integer. It is important to note that these problem instances were generated *a priori* and not specifically tailored for this benchmark. However, when $h_u > J_{uv}$, both RQAOA and Iter-QAOA would correctly prioritise the linear term, resulting in appropriate rounding decisions.

Finally, we note that Iter-QAOA closely resembles the iterative algorithms described in [52, 53], which are themselves extensions of RQAOA. We do not claim novelty for Iter-QAOA over these prior works; rather, it was developed to address performance issues and to validate our optimal parameter-setting strategy.

Code and Data Availability

The code used to generate the numerical results, together with the datasets, is publicly available in the accompanying [GitHub repository](#).

Acknowledgements

V.V. is thankful to Ahad N Zehmakan and Joshua Dai for the stimulating discussions and insightful comments. This research is supported by A*STAR Start Up Fund (KIMR220701aIMRSEF) and Q.InC Strategic Research and Translational Thrust. D.E.K.

is supported by the National Research Foundation, Singapore, and the Agency for Science, Technology and Research (A*STAR), Singapore, under its Quantum Engineering Programme (NRF2021-QEP2-02-P03); A*STAR C230917003; and A*STAR under the Central Research Fund (CRF) Award for Use-Inspired Basic Research (UIBR) and the Quantum Innovation Centre (Q.InC) Strategic Research and Translational Thrust. E.B. is supported by the National Research Foundation of Korea (NRF) of Korea grant funded by the Ministry of Science and ICT (MSIT) (Grant No. NRF-2022R1C1C2006396). H.K. is supported by KIAS individual grant (No. CG085302) at the Korea Institute for Advanced Study. E.B. and H.K. acknowledge support from the NRF of Korea (Grants No. 2023M3K5A109480511 and No. 2023M3K5A1094813). V.V. is supported by the National Quantum Scholarship Scheme.

Author Contributions

All authors conceived the project. V.V., D.E.K., and S.M.A. contributed to the theoretical results. V.V. contributed to the numerical results. V.V. wrote the manuscript with input from all authors. All authors contributed to discussions regarding the results in this paper. D.E.K., P.K.L., and S.M.A. supervised the project.

Competing Interests

All authors declare no financial or non-financial competing interests.

References

- [1] Peter W Shor. “Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer”. *SIAM Review* **41**, 303–332 (1999).
- [2] Lov K Grover. “Quantum mechanics helps in searching for a needle in a haystack”. *Physical Review Letters* **79**, 325 (1997).
- [3] John Preskill. “Quantum computing in the NISQ era and beyond”. *Quantum* **2**, 79 (2018).
- [4] Bin Cheng, Xiu-Hao Deng, Xiu Gu, Yu He, Guangchong Hu, Peihao Huang, Jun Li, Ben-Chuan Lin, Dawei Lu, Yao Lu, Chudan Qiu, Hui Wang, Tao Xin, Shi Yu, Man-Hong Yung, Junkai Zeng, Song Zhang, Youpeng Zhong, Xinhua Peng, Franco Nori, and Dapeng Yu. “Noisy intermediate-scale quantum computers”. *Frontiers of Physics* **18**, 21308 (2023).
- [5] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. “Variational quantum algorithms”. *Nature Reviews Physics* **3**, 625–644 (2021).
- [6] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermann Heimonen, Jakob S Kottmann, Tim Menke, et al. “Noisy intermediate-scale quantum algorithms”. *Reviews of Modern Physics* **94**, 015004 (2022).
- [7] Lennart Bittel and Martin Kliesch. “Training variational quantum algorithms is NP-Hard”. *Physical Review Letters* **127**, 120502 (2021).
- [8] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. “Barren plateaus in quantum neural network training landscapes”. *Nature communications* **9**, 4812 (2018).
- [9] Eric R Anschuetz and Bobak T Kiani. “Quantum variational algorithms are swamped with traps”. *Nature Communications* **13**, 7760 (2022).
- [10] Samson Wang, Enrico Fontana, Marco Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J Coles. “Noise-induced barren plateaus in variational quantum algorithms”. *Nature communications* **12**, 6961 (2021).
- [11] Marco Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. “Cost function dependent barren plateaus in shallow parametrized quantum circuits”. *Nature communications* **12**, 1791 (2021).
- [12] Carlos Ortiz Marrero, Mária Kieferová, and Nathan Wiebe. “Entanglement-induced barren plateaus”. *PRX Quantum* **2**, 040316 (2021).
- [13] Nikita A. Nemkov, Evgeniy O. Kiktenko, and Aleksey K. Fedorov. “Barren plateaus

- swamped with traps”. *Physical Review A* **111**, 012441 (2025).
- [14] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. “A quantum approximate optimization algorithm” (2014). [arXiv:1411.4028](#).
 - [15] Johannes Weidenfeller, Lucia C. Valor, Julien Gacon, Caroline Tornow, Luciano Bello, Stefan Woerner, and Daniel J. Egger. “Scaling of the quantum approximate optimization algorithm on superconducting qubit based hardware”. *Quantum* **6**, 870 (2022).
 - [16] Stuart Andrew Hadfield. “Quantum algorithms for scientific computing and approximate optimization”. *PhD thesis*. Columbia University. (2018).
 - [17] Zhihui Wang, Stuart Hadfield, Zhang Jiang, and Eleanor G Rieffel. “Quantum approximate optimization algorithm for Max-Cut: A fermionic view”. *Physical Review A* **97**, 022304 (2018).
 - [18] Asier Ozaeta, Wim van Dam, and Peter L McMahon. “Expectation values from the single-layer quantum approximate optimization algorithm on ising problems”. *Quantum Science and Technology* **7**, 045036 (2022).
 - [19] V Vijendran, Aritra Das, Dax Enshan Koh, Syed M Assad, and Ping Koy Lam. “An expressive ansatz for low-depth quantum approximate optimisation”. *Quantum Science and Technology* **9**, 025010 (2024).
 - [20] Joao Basso, Edward Farhi, Kunal Marwaha, Benjamin Villalonga, and Leo Zhou. “The Quantum Approximate Optimization Algorithm at High Depth for Max-Cut on Large-Girth Regular Graphs and the Sherrington-Kirkpatrick Model”. In François Le Gall and Tomoyuki Morimae, editors, 17th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2022). *Volume 232 of Leibniz International Proceedings in Informatics (LIPIcs)*, pages 7:1–7:21. Dagstuhl, Germany (2022). Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
 - [21] Truman Yu Ng, Jin Ming Koh, and Dax Enshan Koh. “Analytical expressions for the quantum approximate optimization algorithm and its variants” (2024). [arXiv:2411.09745](#).
 - [22] Jonathan Wurtz and Danylo Lykov. “Fixed-angle conjectures for the quantum approximate optimization algorithm on regular MaxCut graphs”. *Physical Review A* **104**, 052419 (2021).
 - [23] Fernando G. S. L. Brandao, Michael Broughton, Edward Farhi, Sam Gutmann, and Hartmut Neven. “For fixed control parameters the quantum approximate optimization algorithm’s objective function value concentrates for typical instances” (2018). [arXiv:1812.04170](#).
 - [24] Vishwanathan Akshay, Daniil Rabinovich, Ernesto Campos, and Jacob Biamonte. “Parameter concentrations in quantum approximate optimization”. *Physical Review A* **104**, L010401 (2021).
 - [25] Vivek Katal, Kate Smith-Miles, Charles Hill, and Lloyd Hollenberg. “On the instance dependence of parameter initialization for the quantum approximate optimization algorithm: insights via instance space analysis”. *INFORMS Journal on Computing* **37**, 146–171 (2025).
 - [26] Alexey Galda, Xiaoyuan Liu, Danylo Lykov, Yuri Alexeev, and Ilya Safro. “Transferability of optimal QAOA parameters between random graphs”. In 2021 IEEE International Conference on Quantum Computing and Engineering (QCE). *Pages 171–180*. (2021).
 - [27] Alexey Galda, Eesh Gupta, Jose Falla, Xiaoyuan Liu, Danylo Lykov, Yuri Alexeev, and Ilya Safro. “Similarity-based parameter transferability in the quantum approximate optimization algorithm”. *Frontiers in Quantum Science and Technology* **2**, 1200975 (2023).
 - [28] Isak Lyngfelt and Laura García-Álvarez. “Symmetry-informed transferability of optimal parameters in the quantum approximate optimization algorithm”. *Physical Review A* **111**, 022418 (2025).
 - [29] Quinn Langfitt, Jose Falla, Ilya Safro, and Yuri Alexeev. “Parameter transferability in QAOA under noisy conditions”. In 2023 IEEE International Conference on Quantum Computing and Engineering (QCE). *Volume 02*, pages 300–301. (2023).

- [30] Ryo Sakai, Hiromichi Matsuyama, Wai-Hong Tam, Yu Yamashiro, and Keisuke Fujii. “Linearly simplified QAOA parameters and transferability” (2024). [arXiv:2405.00655](#).
- [31] Stefan H Sack and Maksym Serbyn. “Quantum annealing initialization of the quantum approximate optimization algorithm”. *Quantum* **5**, 491 (2021).
- [32] Daniel Liang, Li Li, and Stefan Leichenauer. “Investigating quantum approximate optimization algorithms under bang-bang protocols”. *Physical Review Research* **2**, 033402 (2020).
- [33] Alicia B Magann, Kenneth M Rudinger, Matthew D Grace, and Mohan Sarovar. “Lyapunov-control-inspired strategies for quantum combinatorial optimization”. *Physical Review A* **106**, 062414 (2022).
- [34] Leo Zhou, Sheng-Tao Wang, Soonwon Choi, Hannes Pichler, and Mikhail D Lukin. “Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices”. *Physical Review X* **10**, 021067 (2020).
- [35] Xiao-Hui Ni, Bin-Bin Cai, Hai-Ling Liu, Su-Juan Qin, Fei Gao, and Qiao-Yan Wen. “Multilevel leapfrogging initialization strategy for quantum approximate optimization algorithm”. *Advanced Quantum Technologies* **7**, 2300419 (2024).
- [36] Xinwei Lee, Ningyi Xie, Dongsheng Cai, Yoshiyuki Saito, and Nobuyoshi Asai. “A depth-progressive initialization strategy for quantum approximate optimization algorithm”. *Mathematics* **11**, 2176 (2023).
- [37] Ruslan Shaydulin, Phillip C Lotshaw, Jeffrey Larson, James Ostrowski, and Travis S Humble. “Parameter transfer for quantum approximate optimization of weighted MaxCut”. *ACM Transactions on Quantum Computing* **4**, 1–15 (2023).
- [38] Michał Stęchły, Lanruo Gao, Boniface Yogendran, Enrico Fontana, and Manuel Rudolph. “Connecting the Hamiltonian structure to the QAOA energy and Fourier landscape structure” (2024). [arXiv:2305.13594](#).
- [39] Mahabubul Alam, Abdullah Ash-Saki, and Swaroop Ghosh. “Accelerating quantum approximate optimization algorithm using machine learning”. In 2020 Design, Automation & Test in Europe Conference & Exhibition (DATE). Pages 686–689. IEEE (2020).
- [40] Ningyi Xie, Xinwei Lee, Dongsheng Cai, Yoshiyuki Saito, and Nobuyoshi Asai. “Quantum approximate optimization algorithm parameter prediction using a convolutional neural network”. *Journal of Physics: Conference Series* **2595**, 012001 (2023).
- [41] Zhiding Liang, Gang Liu, Zheyuan Liu, Jinglei Cheng, Tianyi Hao, Kecheng Liu, Hang Ren, Zhixin Song, Ji Liu, Fanny Ye, and Yiyu Shi. “Invited: Graph learning for parameter prediction of quantum approximate optimization algorithm”. In Proceedings of the 61st ACM/IEEE Design Automation Conference. DAC ’24 New York, NY, USA (2024). Association for Computing Machinery.
- [42] Fanxu Meng, Xiangzhen Zhou, Pengcheng Zhu, and Yu Luo. “Conditional diffusion-based parameter generation for quantum approximate optimization algorithm”. *EPJ Quantum Technology* **12**, 100 (2025).
- [43] Ohad Amosy, Tamuz Danzig, Ohad Lev, Ely Porat, Gal Chechik, and Adi Makmal. “Iteration-free quantum approximate optimization algorithm using neural networks”. *Quantum Machine Intelligence* **6**, 38 (2024).
- [44] Sami Khairy, Ruslan Shaydulin, Lukasz Cincio, Yuri Alexeev, and Prasanna Balaprakash. “Reinforcement-learning-based variational quantum circuits optimization for combinatorial problems” (2019). [arXiv:1911.04574](#).
- [45] J. A. Montañez-Barrera, Dennis Willsch, and Kristel Michielsen. “Transfer learning of optimal QAOA parameters in combinatorial optimization”. *Quantum Information Processing* **24**, 129 (2025).
- [46] Jose Falla, Quinn Langfitt, Yuri Alexeev, and Ilya Safro. “Graph representation learning for parameter transferability in quantum approximate optimization algorithm”. *Quantum Machine Intelligence* **6**, 46 (2024).
- [47] Lixue Cheng, Yu-Qin Chen, Shi-Xin Zhang, and Shengyu Zhang. “Quantum ap-

- proximate optimization via learning-based adaptive optimization”. *Communications Physics* **7**, 83 (2024).
- [48] Sergey Bravyi, Alexander Kliesch, Robert Koenig, and Eugene Tang. “Obstacles to variational quantum optimization from symmetry protection”. *Physical Review Letters* **125**, 260505 (2020).
 - [49] Seongmin Kim, Vincent R. Pascuzzi, Zhihao Xu, Tengfei Luo, Eungkyu Lee, and In-Saeng Suh. “Distributed quantum approximate optimization algorithm on a quantum-centric supercomputing architecture”. *npj Quantum Information* (2026).
 - [50] Bo Yue, Shibei Xue, Yu Pan, Min Jiang, and Daoyi Dong. “Local to global: A distributed quantum approximate optimization algorithm for pseudo-Boolean optimization problems” (2023). [arXiv:2310.05062](#).
 - [51] Kuan-Cheng Chen, Xiaotian Xu, Felix Burt, Chen-Yu Liu, Shang Yu, and Kin K. Leung. “Noise-aware distributed quantum approximate optimization algorithm on near-term quantum hardware”. In 2024 IEEE International Conference on Quantum Computing and Engineering (QCE). *Volume 02*, pages 144–149. (2024).
 - [52] Jernej Rudi Finžgar, Aron Kerschbaumer, Martin JA Schuetz, Christian B Mendl, and Helmut G Katzgraber. “Quantum-informed recursive optimization algorithms”. *PRX Quantum* **5**, 020327 (2024).
 - [53] Lucas T Brady and Stuart Hadfield. “Iterative quantum algorithms for maximum independent set”. *Physical Review A* **110**, 052435 (2024).
 - [54] Bao Bach, Jose Falla, and Ilya Safro. “MLQAOA: Graph Learning Accelerated Hybrid Quantum-Classical Multilevel QAOA”. In 2024 IEEE International Conference on Quantum Computing and Engineering (QCE). *Volume 01*, pages 1–12. (2024).
 - [55] Zeqiao Zhou, Yuxuan Du, Xinmei Tian, and Dacheng Tao. “QAOA-in-QAOA: solving large-scale MaxCut problems on small quantum machines”. *Physical Review Applied* **19**, 024027 (2023).
 - [56] Moises Ponce, Rebekah Herrman, Phillip C. Lotshaw, Sarah Powers, George Siopsis, Travis Humble, and James Ostrowski. “Graph decomposition techniques for solving combinatorial optimization problems with variational quantum algorithms”. *Quantum Information Processing* **24**, 60 (2025).
 - [57] Maxime Dupont, Bram Evert, Mark J Hodson, Bhuvanesh Sundar, Stephen Jeffrey, Yuki Yamaguchi, Dennis Feng, Filip B Maciejewski, Stuart Hadfield, M Sohaib Alam, et al. “Quantum-enhanced greedy combinatorial optimization solver”. *Science Advances* **9**, eadi0487 (2023).
 - [58] Maxime Dupont and Bhuvanesh Sundar. “Extending relax-and-round combinatorial optimization solvers with quantum correlations”. *Physical Review A* **109**, 012429 (2024).
 - [59] Libor Caha, Alexander Kliesch, and Robert Koenig. “Twisted hybrid algorithms for combinatorial optimization”. *Quantum Science and Technology* **7**, 045013 (2022).
 - [60] Victor Fischer, Maximilian Passek, Friedrich Wagner, Jernej Rudi Finžgar, Lilly Palackal, and Christian B. Mendl. “Quantum and classical correlations in shrinking algorithms for optimization” (2024). [arXiv:2404.17242](#).
 - [61] Abraham P Punnen. “The quadratic unconstrained binary optimization problem”. *Springer International Publishing* **10**, 978–3 (2022).
 - [62] Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. “Effect of data encoding on the expressive power of variational quantum-machine-learning models”. *Physical Review A* **103**, 032430 (2021).
 - [63] Francisco Javier Gil Vidal and Dirk Oliver Theis. “Input redundancy for parameterized quantum circuits”. *Frontiers in Physics* **8**, 297 (2020).
 - [64] Enrico Fontana, Ivan Rungger, Ross Duncan, and Cristina Cîrstoiu. “Spectral analysis for noise diagnostics and filter-based digital error mitigation” (2022). [arXiv:2206.08811](#).
 - [65] Enrico Fontana, Ivan Rungger, Ross Duncan, and Cristina Cîrstoiu. “Efficient recovery of variational quantum algorithms landscapes using classical signal processing” (2022). [arXiv:2208.05958](#).

- [66] Nikita A. Nemkov, Evgeniy O. Kiktenko, and Aleksey K. Fedorov. “Fourier expansion in variational quantum algorithms”. *Physical Review A* **108**, 032406 (2023).
- [67] Claude Elwood Shannon. “Communication in the presence of noise”. *Proceedings of the IRE* **37**, 10–21 (1949).
- [68] Tianyi Hao, Kun Liu, and Swamit Tannu. “Enabling high performance debugging for variational quantum algorithms using compressed sensing”. In Proceedings of the 50th Annual International Symposium on Computer Architecture. Pages 1–13. (2023).
- [69] Daniel Müssig, Markus Wappler, Steve Lenk, and Jörg Lässig. “Connecting the Hamiltonian structure to the QAOA performance and energy landscape”. In INFOR-MATIK 2024. Pages 595–604. Gesellschaft für Informatik e.V., Bonn (2024).
- [70] Mauricio Chávez-Pichardo, Miguel A Martínez-Cruz, Alfredo Trejo-Martínez, Daniel Martínez-Carbajal, and Tanya Arenas-Resendiz. “A complete review of the general quartic equation with real coefficients and multiple roots”. *Mathematics* **10**, 2377 (2022).
- [71] John P Boyd. “Computing the zeros, maxima and inflection points of Chebyshev, Legendre and Fourier series: solving transcendental equations by spectral interpolation and polynomial rootfinding”. *Journal of Engineering Mathematics* **56**, 203–219 (2006).
- [72] John P. Boyd. “A comparison of companion matrix methods to find roots of a trigonometric polynomial”. *Journal of Computational Physics* **246**, 96–112 (2013).
- [73] Bahman Kalantari, Fedor Andreev, and Chun Lau. “Characterization of local optima of polynomial modulus over a disc”. *Numerical Algorithms* **90**, 773–787 (2022).
- [74] Eldon R Hansen. “Global optimization using interval analysis: the one-dimensional case”. *Journal of Optimization Theory and Applications* **29**, 331–344 (1979).
- [75] Eldon Hansen and G William Walster. “Global optimization using interval analysis: revised and expanded”. Volume 264. CRC press. (2003).
- [76] J. J. Green. “Calculating the maximum modulus of a polynomial using Steckin’s lemma”. *SIAM Journal on Numerical Analysis* **36**, 1022–1029 (1999).
- [77] Gabriel De La Chevrotiere. “Finding the maximum modulus of a polynomial on the polydisk using a generalization of Steckin’s lemma”. *SIAM Undergraduate Research Online* **2**, 2–2 (2009).
- [78] Pavel Emeliyanenko and Michael Sagraloff. “On the complexity of solving a bivariate polynomial system”. In Proceedings of the 37th International Symposium on Symbolic and Algebraic Computation. Pages 154–161. (2012).
- [79] Bor Plestenjak and Michiel E Hochstenbach. “Roots of bivariate polynomial systems via determinantal representations”. *SIAM Journal on Scientific Computing* **38**, A765–A788 (2016).
- [80] Sami Boulebnane, Xavier Lucas, Agnes Meyder, Stanislaw Adaszewski, and Ashley Montanaro. “Peptide conformational sampling using the quantum approximate optimization algorithm”. *npj Quantum Information* **9**, 70 (2023).
- [81] Sebastian Brandhofer, Daniel Braun, Vanessa Dehn, Gerhard Hellstern, Matthias Hüls, Yanjun Ji, Ilia Polian, Amandeep Singh Bhatia, and Thomas Wellens. “Benchmarking the performance of portfolio optimization with QAOA”. *Quantum Information Processing* **22**, 25 (2022).
- [82] Sami Boulebnane and Ashley Montanaro. “Predicting parameters for the quantum approximate optimization algorithm for max-cut from the infinite-size limit” (2021). [arXiv:2110.10685](https://arxiv.org/abs/2110.10685).
- [83] Shree Hari Sureshbabu, Dylan Herman, Ruslan Shaydulin, Joao Basso, Shouvanik Chakrabarti, Yue Sun, and Marco Pistoia. “Parameter setting in quantum approximate optimization of weighted problems”. *Quantum* **8**, 1231 (2024).
- [84] P. Erdős and A. Rényi. “On random graphs I”. *Publ. math. debrecen* **6**, 290–297 (1959).
- [85] E. N. Gilbert. “Random Graphs”. *The Annals of Mathematical Statistics* **30**, 1141 – 1144 (1959).
- [86] Gurobi Optimization, LLC. “Gurobi Optimizer Reference Manual”. <https://www.gurobi.com> (2022).

- [87] Michel X Goemans and David P Williamson. “Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming”. *Journal of the ACM (JACM)* **42**, 1115–1145 (1995).
- [88] Sergey Bravyi, Alexander Kliesch, Robert Koenig, and Eugene Tang. “Hybrid quantum-classical algorithms for approximate graph coloring”. *Quantum* **6**, 678 (2022).
- [89] Yash J Patel, Sofiene Jerbi, Thomas Bäck, and Vedran Dunjko. “Reinforcement learning assisted recursive QAOA”. *EPJ Quantum Technology* **11**, 6 (2024).
- [90] V Vijendran, Dax Enshan Koh, Ping Koy Lam, and Syed M Assad. “Classical and quantum heuristics for the binary paint shop problem” (2025). [arXiv:2509.15294](#).
- [91] Sergey Bravyi, David Gosset, Daniel Grier, and Luke Schaeffer. “Classical algorithms for forrelation” (2021). [arXiv:2102.06963](#).
- [92] Eunok Bae and Soojoon Lee. “Recursive QAOA outperforms the original QAOA for the MAX-CUT problem on complete graphs”. *Quantum Information Processing* **23**, 78 (2024).
- [93] Eunok Bae, Hyukjoon Kwon, V Vijendran, and Soojoon Lee. “Modified recursive QAOA for exact MAX-CUT solutions on bipartite graphs: closing the gap beyond QAOA limit”. *Journal of Physics A: Mathematical and Theoretical* **58**, 485304 (2025).
- [94] Charles Moussa, Hao Wang, Thomas Bäck, and Vedran Dunjko. “Unsupervised strategies for identifying optimal parameters in quantum approximate optimization algorithm”. *EPJ Quantum Technology* **9**, 11 (2022).
- [95] Simon Thelen, Hila Safi, and Wolfgang Maurer. “Approximating under the Influence of Quantum Noise and Compute Power”. In 2024 IEEE International Conference on Quantum Computing and Engineering (QCE). *Pages 274–279*. Los Alamitos, CA, USA (2024). IEEE Computer Society.
- [96] Daniel J Egger, Jakub Mareček, and Stefan Woerner. “Warm-starting quantum optimization”. *Quantum* **5**, 479 (2021).
- [97] David P Williamson and David B Shmoys. “The design of approximation algorithms”. *Cambridge University Press*. (2011).
- [98] Lap Chi Lau, Ramamoorthi Ravi, and Mohit Singh. “Iterative methods in combinatorial optimization”. *Volume 46*. Cambridge University Press. (2011).
- [99] Michael H Freedman and Matthew B Hastings. “Quantum systems on non- k -hyperfinite complexes: A generalization of classical statistical mechanics on expander graphs”. *Quantum Information & Computation* **14**, 144–180 (2014).
- [100] Edward Farhi, David Gamarnik, and Sam Gutmann. “The quantum approximate optimization algorithm needs to see the whole graph: A typical case” (2020). [arXiv:2004.09002](#).
- [101] Edward Farhi, David Gamarnik, and Sam Gutmann. “The quantum approximate optimization algorithm needs to see the whole graph: Worst case examples” (2020). [arXiv:2005.08747](#).
- [102] Joao Basso, David Gamarnik, Song Mei, and Leo Zhou. “Performance and limitations of the QAOA at constant levels on large sparse hypergraphs and spin glass models”. In 2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS). *Pages 335–343*. IEEE (2022).
- [103] Anurag Anshu and Tony Metger. “Concentration bounds for quantum states and limitations on the QAOA from polynomial approximations”. *Quantum* **7**, 999 (2023).
- [104] Boaz Barak and Kunal Marwaha. “Classical Algorithms and Quantum Limitations for Maximum Cut on High-Girth Graphs”. In Mark Braverman, editor, 13th Innovations in Theoretical Computer Science Conference (ITCS 2022). *Volume 215 of Leibniz International Proceedings in Informatics (LIPIcs)*, pages 14:1–14:21. Dagstuhl, Germany (2022). Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- [105] Mark Goh. “The overlap gap property limits limit swapping in the QAOA”. *Quantum Information & Computation* **25**, 329–343 (2025).
- [106] Vishwanathan Akshay, Hariphan Philathong, Mauro ES Morales, and Jacob D Biamonte. “Reachability deficits in quan-

- tum approximate optimization”. *Physical Review Letters* **124**, 090504 (2020).
- [107] Vishwanathan Akshay, H Philathong, Igor Zacharov, and J Biamonte. “Reachability deficits in quantum approximate optimization of graph problems”. *Quantum* **5**, 532 (2021).
 - [108] Antares Chen, Neng Huang, and Kunal Marwaha. “Local algorithms and the failure of log-depth quantum advantage on sparse random CSPs” (2023). [arXiv:2310.01563](#).
 - [109] Chi-Ning Chou, Peter J. Love, Jusepreet Singh Sandhu, and Jonathan Shi. “Limitations of Local Quantum Algorithms on Random MAX-k-XOR and Beyond”. In Mikołaj Bojańczyk, Emanuela Merelli, and David P. Woodruff, editors, 49th International Colloquium on Automata, Languages, and Programming (ICALP 2022). *Volume 229 of Leibniz International Proceedings in Informatics (LIPIcs)*, pages 41:1–41:20. Dagstuhl, Germany (2022). Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
 - [110] Kunal Marwaha and Stuart Hadfield. “Bounds on approximating Max k XOR with quantum and classical local algorithms”. *Quantum* **6**, 757 (2022).
 - [111] R. E. Edwards. “Fourier series: A modern introduction volume 1”. *Volume 2 of Graduate Texts in Mathematics*, pages XII, 228. Springer-Verlag New York, NY. (1979). 2 edition.

A Fourier Insights into the Expressivity and Landscape Limitations of QAOA₁

In section 3.2, we derived analytical expressions to compute the maximum frequencies achievable by QAOA₁ for arbitrary Ising models. These expressions were utilised in section 3.3 to characterise the energy landscape and estimate the minimum sampling resolution required for accurately reconstructing the optimisation landscape. In this section, we compare the maximum frequencies of the Hamiltonian H_P with those accessible via QAOA₁ for two sets of field-free Ising models: (i) 4-regular graphs with node counts ranging from 5 to 20 and (ii) D -regular graphs with 20 nodes, where D varies from 2 to 19. For each case, 100 random graphs were generated with edge weights drawn from a Gaussian distribution (mean 50, variance 25). To facilitate this comparison, we employed specific computational methods to determine the maximum frequencies of both QAOA₁ and H_P .

The maximum frequency of QAOA₁ was determined analytically using corollary 2. In contrast, the maximum frequency of H_P was obtained by enumerating all possible solutions and calculating the difference between the highest and lowest costs. This is because constructing the matrix representation of H_P for n variables (qubits) becomes impractical for large n due to its exponential scaling as $2^n \times 2^n$. Specifically, for $n = 20$, we leverage the diagonal structure of H_P in the computational basis, where the diagonal entries correspond to the costs of all possible configurations in $\{0, 1\}^n$. Consequently, we enumerate all configurations in $\{0, 1\}^{20}$, compute their respective costs, and determine the maximum frequency as the difference between the minimum and maximum costs.

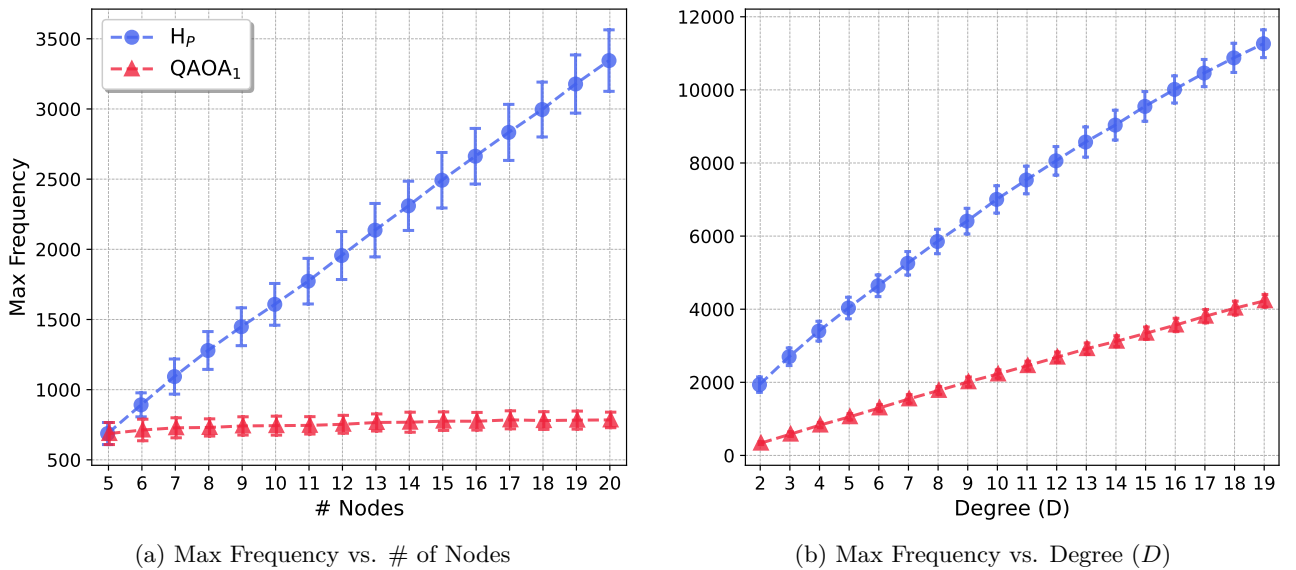


Figure 7: **Maximum Frequency Comparison between Problem Hamiltonian and QAOA₁**. The plots show the average maximum frequencies for 100 random graphs. **(a)** Maximum frequencies for 4-regular graphs with varying node counts from 5 to 20. **(b)** Maximum frequencies for 20-node D -regular graphs with degrees D ranging from 2 to 19. Error bars represent the standard deviation across the random instances.

As shown in fig. 7, the maximum frequency attainable by QAOA₁ is significantly lower than that of H_P , primarily due to the locality constraints inherent in QAOA. In a p -level implementation, each qubit interacts only with others within p edges; for $p = 1$, this means interactions are limited to immediate neighbours. This limitation is evident in fig. 7a, where the maximum reachable frequency for 4-regular graphs remains nearly constant as the number of nodes increases. In smaller graphs, qubits can effectively access the entire graph, but in larger graphs with a fixed degree $d = 4$, qubits become increasingly isolated, restricting their influence. Conversely, fig. 7b demonstrates that for 20-node D -regular graphs, the maximum frequency increases linearly with D . As the degree grows, each

qubit interacts with more neighbours, increasing the range of local information incorporated into the QAOA₁ expectation value and thereby yielding higher maximum frequencies.

Despite these improvements, even in complete graphs ($d = 19$), where every qubit is fully connected to every other qubit, a gap persists between the maximum frequencies achievable by QAOA₁ and those of H_P . This gap arises not only from the locality constraints but also from fundamental limitations in QAOA's Fourier series representation at level 1, where the finite number of terms restricts the circuit's ability to capture complex, high-frequency features of the energy landscape. Additionally, concepts such as the No Low Trivial Energy States (NLTS) [99] provide further insights into this persistent frequency gap. It is important to note that the performance limitations of QAOA have been extensively explored in prior studies, which have focused on factors like locality constraints, the Overlap Gap Property (OGP) [100–105], reachability deficits related to constraint-to-variable ratios [106–108], and symmetry protection [48, 109, 110]. The novelty of this analysis, as illustrated in fig. 7, lies in interpreting QAOA's expressivity and optimisation-landscape limitations through the lens of Fourier theory.

B Some Useful Lemmas

Before deriving the analytical formulas, we first state and prove some key identities that will be used in section C. In particular, for functions that involve a product of cosine terms, each with its own fundamental frequency, the following lemma provides a method to compute the maximum frequency of the entire function.

Lemma 8. *Let $h(\mathbf{w})$ be a product of cosine functions defined by*

$$h(\mathbf{w}) = \prod_{i=1}^n \cos(w_i \theta), \quad (53)$$

where $\mathbf{w} = (w_1, \dots, w_n)$ is a vector of real coefficients. The maximum angular frequency of $h(\mathbf{w})$ is given by

$$\omega_{\max}[h(\mathbf{w})] = \sum_{i=1}^n |w_i|. \quad (54)$$

Proof. We start by expressing the product of cosine functions $h(\mathbf{w}) = \prod_{i=1}^n \cos(w_i \theta)$ using the product-to-sum identity for cosine functions. The product-to-sum identity for n cosine terms is given by:

$$\prod_{k=1}^n \cos \theta_k = \frac{1}{2^n} \sum_{\mathbf{e} \in \{-1, 1\}^n} \cos \left(\sum_{i=1}^n e_i \theta_i \right), \quad (55)$$

where $\mathbf{e} = (e_1, \dots, e_n)$ is an n -tuple with each e_i taking values in $\{-1, 1\}$. Applying this identity to $h(\mathbf{w})$, we have:

$$h(\mathbf{w}) = \prod_{i=1}^n \cos(w_i \theta) = \frac{1}{2^n} \sum_{\mathbf{e} \in \{-1, 1\}^n} \cos \left(\sum_{i=1}^n e_i w_i \theta \right). \quad (56)$$

The maximum angular frequency component of $h(\mathbf{w})$ corresponds to the term in the sum where the argument of the cosine function is maximised. This maximum occurs when each e_i is chosen to match the sign of the corresponding w_i , i.e., $e_i = \text{sgn}(w_i)$. Therefore, the term with the maximum angular frequency is:

$$\cos \left(\sum_{i=1}^n \text{sgn}(w_i) w_i \theta \right) = \cos \left(\sum_{i=1}^n |w_i| \theta \right). \quad (57)$$

Thus, the maximum angular frequency of $h(\mathbf{w})$ is given by $\sum_{i=1}^n |w_i|$. $\square$

The following lemma extends lemma 8 to compute the maximum angular frequency of the product of a single sine function and n cosine functions, each with its own fundamental angular frequency.

Lemma 9. *Let $g(\mathbf{w})$ be a product of one sine function and n cosine functions defined by*

$$g(\mathbf{w}) = \sin(w_0\theta) \prod_{i=1}^n \cos(w_i\theta), \quad (58)$$

where $\mathbf{w} = (w_0, w_1, \dots, w_n)$ is a vector of real coefficients. The maximum angular frequency of $g(\mathbf{w})$ is given by

$$\omega_{\max}[g(\mathbf{w})] = |w_0| + \sum_{i=1}^n |w_i|. \quad (59)$$

Proof. From lemma 8, we know that the maximum angular frequency of the product $\prod_{i=1}^n \cos(w_i\theta)$ is $\sum_{i=1}^n |w_i|$, corresponding to the term $\cos((\sum_{i=1}^n |w_i|)\theta)$.

Next, consider the product of this cosine term with $\sin(w_0\theta)$. To analyse the frequency content of this product, we use the product-to-sum identity for sine and cosine functions, which states:

$$\sin\theta \cos\varphi = \frac{1}{2}[\sin(\theta + \varphi) + \sin(\theta - \varphi)]. \quad (60)$$

Applying this identity with $\theta = w_0\theta$ and $\varphi = \sum_{i=1}^n |w_i|\theta$, we obtain:

$$\sin(w_0\theta) \cos\left(\left(\sum_{i=1}^n |w_i|\right)\theta\right) = \frac{1}{2} \left[\sin\left(\left(w_0 + \sum_{i=1}^n |w_i|\right)\theta\right) + \sin\left(\left(w_0 - \sum_{i=1}^n |w_i|\right)\theta\right) \right]. \quad (61)$$

From the above expression, we observe that the maximum angular frequency of the product $g(\mathbf{w}) = \sin(w_0\theta) \prod_{i=1}^n \cos(w_i\theta)$ depends on the sign of w_0 . If $w_0 > 0$, the first sine term on the RHS dominates with the maximum angular frequency, while for $w_0 < 0$, the second sine term attains this maximum angular frequency. In both cases, the maximum angular frequency of $g(\mathbf{w})$ is given by $|w_0| + \sum_{i=1}^n |w_i|$. $\square$

C Maximum Angular Frequency of Level-1 QAOA for Ising Models

We now present the proof of theorem 2, which follows directly from the application of lemmas 8 and 9 to the analytical expressions in theorem 1. This proof demonstrates how to compute the maximum angular frequency exhibited by QAOA₁ for an arbitrary Ising model with external fields.

Proof of Theorem 2. For the single spin variable i , the terms involving γ are given by eq. (16). From lemma 9, the maximum frequency is:

$$\omega_{\max}[\langle C_i \rangle_\gamma] = 2 \left(|h_i| + \sum_{k \in \mathcal{N}(i)} |J_{ik}| \right). \quad (62)$$

For the two spin variables u and v , the terms involving γ are:

$$\langle C_{uv} \rangle_\gamma = \frac{J_{uv}}{2} \sin(4\beta) f_{uv}(\mathbf{J}, \mathbf{h}, \gamma) - \frac{J_{uv}}{2} \sin^2(2\beta) g_{uv}(\mathbf{J}, \mathbf{h}, \gamma), \quad (63)$$

where

$$f_{uv}(\mathbf{J}, \mathbf{h}, \gamma) = \sin(2J_{uv}\gamma) \underbrace{\left(\cos(2h_v\gamma) \prod_{w \in e} \cos(2J_{vw}\gamma) \right)}_{\textcircled{1}} + \underbrace{\cos(2h_u\gamma) \prod_{w \in d} \cos(2J_{uw}\gamma)}_{\textcircled{2}}, \quad (64)$$

$$g_{uv}(\mathbf{J}, \mathbf{h}, \gamma) = \underbrace{\prod_{\substack{w \in e \\ w \notin F}} \cos(2J_{vw}\gamma) \prod_{\substack{w \in d \\ w \notin F}} \cos(2J_{uw}\gamma)}_{\textcircled{3}} \underbrace{\left(\cos(2\gamma(h_u + h_v)) \prod_{f \in F} \cos(2\gamma(J_{uf} + J_{vf})) \right)}_{\textcircled{4}} \\ - \underbrace{\cos(2\gamma(h_u - h_v)) \prod_{f \in F} \cos(2\gamma(J_{uf} - J_{vf}))}_{\textcircled{5}}. \quad (65)$$

Using lemma 9, we have:

$$\omega_{\max} \left[\sin(2J_{uv}\gamma) \times \textcircled{1} \right] = 2 \left(|J_{uv}| + |h_v| + \sum_{w \in e} |J_{vw}| \right), \quad (66)$$

$$\omega_{\max} \left[\sin(2J_{uv}\gamma) \times \textcircled{2} \right] = 2 \left(|J_{uv}| + |h_u| + \sum_{w \in d} |J_{uw}| \right). \quad (67)$$

Using lemma 8, we have:

$$\omega_{\max} \left[\textcircled{3} \times \textcircled{4} \right] = 2 \left(\sum_{\substack{w \in e \\ w \notin F}} |J_{vw}| + \sum_{\substack{w \in d \\ w \notin F}} |J_{uw}| + |h_u + h_v| + \sum_{f \in F} |J_{uf} + J_{vf}| \right), \quad (68)$$

$$\omega_{\max} \left[\textcircled{3} \times \textcircled{5} \right] = 2 \left(\sum_{\substack{w \in e \\ w \notin F}} |J_{vw}| + \sum_{\substack{w \in d \\ w \notin F}} |J_{uw}| + |h_u - h_v| + \sum_{f \in F} |J_{uf} - J_{vf}| \right). \quad (69)$$

Thus, the maximum frequencies of $f_{uv}(\mathbf{h}, \gamma)$ and $g_{uv}(\mathbf{h}, \gamma)$ are:

$$\omega_{\max} [f_{uv}(\mathbf{J}, \mathbf{h}, \gamma)] = 2 \left(|J_{uv}| + \max \left\{ |h_v| + \sum_{w \in e} |J_{vw}|, |h_u| + \sum_{w \in d} |J_{uw}| \right\} \right), \quad (70)$$

$$\omega_{\max} [g_{uv}(\mathbf{J}, \mathbf{h}, \gamma)] = 2 \left(\sum_{\substack{w \in e \\ w \notin F}} |J_{vw}| + \sum_{\substack{w \in d \\ w \notin F}} |J_{uw}| + \max \left\{ |h_u \pm h_v| + \sum_{f \in F} |J_{uf} \pm J_{vf}| \right\} \right). \quad (71)$$

Therefore, the maximum angular frequency of $\langle C_{uv} \rangle_\gamma$ is the maximum of these two frequencies. $\square$

D Upper Bounds on the Sampling Periods

In this section, we establish upper bounds on the sampling period necessary to reconstruct the optimisation landscape of QAOA₁ for the SK model, D -regular triangle-free graphs, and triangle-free graphs with bounded maximum degrees. These bounds are proven by integrating theorem 4 and theorem 2 and leveraging the specific properties of each case.

D.1 Sherrington-Kirkpatrick Model

Proof of Corollary 3. Consider the case without external fields. The maximum angular frequency of the γ terms in the expectation value is

$$\omega_{\max}[\langle C_{uv} \rangle_{\gamma}] = 2 \times \max \left\{ |J_{uv}| + \max \left\{ \sum_{w \in e} |J_{wv}|, \sum_{w \in d} |J_{uw}| \right\}, \max \left\{ \sum_{f \in F} |J_{uf} \pm J_{vf}| \right\} \right\}. \quad (72)$$

Since the graph is complete, the number of neighbours excluding vertices u and v is $|e| = |d| = n - 2$, and the number of triangles is $|F| = n - 2$. Therefore, we have

$$\max \left\{ \sum_{w \in e} |J_{wv}|, \sum_{w \in d} |J_{uw}| \right\} = n - 2, \quad (73)$$

$$\max \left\{ \sum_{f \in F} |J_{uf} \pm J_{vf}| \right\} = 2n - 4. \quad (74)$$

The maximum in the summation of the triangle terms in eq. (74) is achieved when all edges forming a triangle with edge $\{u, v\}$ have the same weights. Specifically, this occurs when $\text{sgn}(J_{uf}) = -\text{sgn}(J_{vf})$ for all $f \in F$. Under this condition, the maximum angular frequency of the γ terms scales as $\mathcal{O}(n)$.

This scaling remains valid for models with external fields of strength ± 1 , where the maximum angular frequency may differ by a constant factor compared to the case without external fields. Consequently, the maximum frequency is given by

$$\nu_{\max} = \frac{\mathcal{O}(n)}{2\pi}. \quad (75)$$

Since all weights are ± 1 and each γ term is prefactored by 2, the period of the optimisation function is $T_{\gamma} = \pi$. Utilising the maximum frequency and the period of the function, sampling period is determined from theorem 4. $\square$

D.2 Triangle-Free Graphs

Proof of Corollary 4. For a triangle-free graph, the maximum angular frequency of the γ terms in the expectation value is given by:

$$\omega_{\max}[\langle C_{uv} \rangle_{\gamma}] = 2|J_{uv}| + 2 \times \max \left\{ \sum_{w \in e} |J_{wv}|, \sum_{w \in d} |J_{uw}| \right\}. \quad (76)$$

Since the graph is D -regular, each vertex has $D - 1$ neighbours excluding u and v , so $|e| = |d| = D - 1$. Consequently,

$$\max \left\{ \sum_{w \in e} |J_{wv}|, \sum_{w \in d} |J_{uw}| \right\} = D - 1, \quad (77)$$

which gives the maximum angular frequency as

$$\omega_{\max}[\langle C_{uv} \rangle_{\gamma}] = 2 + 2(D - 1) = 2D. \quad (78)$$

The corresponding maximum frequency is $\omega_{\max}/2\pi = D/\pi$. Since each γ term is prefactored by 2 and the weights are ± 1 , the optimisation function has a period of $T_{\gamma} = \pi$. Using this period and the maximum frequency, the sampling period is derived from theorem 4.

For triangle-free graphs with varying degrees, the maximum angular frequency is determined by the edge u, v where at least one vertex has the maximum degree $D_{\max}$. $\square$

E Univariate Representation of Level-1 QAOA's Objective Function

In this section, we present the proofs of the linearised closed-form expressions for computing the expectation value of QAOA₁, specifically for Ising models with and without external fields, as described in theorem 5 and theorem 2. These proofs rely on identities involving linear combinations of trigonometric functions and simple algebraic techniques.

E.1 Ising Models without Fields

Proof of Theorem 5. Using the definitions of the coefficients $A(\gamma)$ and $B(\gamma)$, the expectation value for an arbitrary Ising model without external fields can be rewritten as:

$$\langle \gamma, \beta | H_P | \gamma, \beta \rangle = A(\gamma) \sin 4\beta - B(\gamma) \sin^2 2\beta \quad (79)$$

$$= A(\gamma) \sin 4\beta - B(\gamma) \left(\frac{1 - \cos 4\beta}{2} \right) \quad (80)$$

$$= A(\gamma) \sin 4\beta + \frac{B(\gamma)}{2} \cos 4\beta - \frac{B(\gamma)}{2}. \quad (81)$$

Using the trigonometric identity for a linear combination of sine and cosine, specifically:

$$R \cos(x - \alpha) = A \sin x + B \cos x, \quad (82)$$

the expression simplifies to:

$$\langle \gamma, \beta | H_P | \gamma, \beta \rangle = R(\gamma) \cos(4\beta - \alpha(\gamma)) - \frac{B(\gamma)}{2}, \quad (83)$$

where

$$R(\gamma) = \text{sgn}[B(\gamma)] \sqrt{A^2(\gamma) + \frac{B^2(\gamma)}{4}}, \quad \alpha(\gamma) = \arctan(2A(\gamma), B(\gamma)). \quad (84)$$

The minimum value of $\langle \gamma, \beta | H_P | \gamma, \beta \rangle$ is thus:

$$\min_{\gamma \in \mathbb{R}} \left(-\sqrt{A^2(\gamma) + \frac{B^2(\gamma)}{4}} - \frac{B(\gamma)}{2} \right), \quad (85)$$

which is achieved by setting $\beta^* = \frac{\alpha(\gamma^*) + \pi}{4}$. □

E.2 Ising Models with Fields

Proof of Theorem 6. Using the definitions of the coefficients $A(\gamma)$, $B(\gamma)$, and $C(\gamma)$, the expectation value for an arbitrary 2-local Ising model with external fields can be rewritten as:

$$\langle \gamma, \beta | H_P | \gamma, \beta \rangle = \sin(2\beta)A(\gamma) + \sin(4\beta)B(\gamma) + \sin^2(2\beta)C(\gamma). \quad (86)$$

To minimise this expression with respect to β for a fixed γ , consider:

$$\langle \beta | H_\gamma | \beta \rangle = A \sin 2\beta + B \sin 4\beta + C \sin^2 2\beta. \quad (87)$$

This function is bounded with endpoints $\langle 0 | H_\gamma | 0 \rangle = \langle \pi | H_\gamma | \pi \rangle = 0$ and has a period of π . To find the global minimum in $(0, \pi)$, let $b = 2\beta$:

$$\langle \beta | H_\gamma | \beta \rangle = A \sin b + B \sin 2b + C \sin^2 b. \quad (88)$$

Taking the derivative with respect to β :

$$\frac{d \langle \beta | H_\gamma | \beta \rangle}{d\beta} = 2 [A \cos b + 2B \cos 2b + 2C \sin b \cos b] \quad (89)$$

$$= 2 [A \cos b + 2B(2 \cos^2 b - 1) + 2C \sin b \cos b] \quad (90)$$

$$= 2 [A \cos b + 4B \cos^2 b - 2B + 2C \sin b \cos b]. \quad (91)$$

Setting $\frac{d \langle \beta | H_\gamma | \beta \rangle}{d\beta} = 0$ to find the minima:

$$0 = A \cos b + 4B \cos^2 b - 2B + 2C \sin b \cos b \quad (92)$$

$$4C^2 \sin^2 b \cos^2 b = (A \cos b + 4B \cos^2 b - 2B)^2 \quad (93)$$

$$4C^2(1 - \cos^2 b) \cos^2 b = (A \cos b + 4B \cos^2 b - 2B)^2. \quad (94)$$

Letting $x = \cos b$, we get:

$$4C^2 (1 - x^2) x^2 = (Ax + 4Bx^2 - 2B)^2, \quad (95)$$

which simplifies to the quartic equation:

$$0 = p(x) = 4B^2 - 4ABx + (A^2 - 16B^2 - 4C^2)x^2 + 8ABx^3 + (16B^2 + 4C^2)x^4. \quad (96)$$

Let x_1, x_2, x_3, x_4 be the roots of $p(x)$. The minimum of $\langle \beta | H_\gamma | \beta \rangle$ occurs at one of the β values where $x = \cos(2\beta)$ is a root of $p(x)$. Therefore:

$$\beta_\gamma^* = \operatorname{argmin}_{\beta \in \mathcal{B}_\gamma} \langle \beta | H_\gamma | \beta \rangle, \quad (97)$$

where

$$\mathcal{B}_\gamma = \left\{ \beta = \pm \frac{1}{2} \arccos(x_i) \mid i = 1, 2, 3, 4 \right\}. \quad (98)$$

Finally, the global optimal angles (γ^*, β^*) are given by:

$$(\gamma^*, \beta^*) = \operatorname{argmin}_{\gamma \in \mathbb{R}} \langle \gamma, \beta_\gamma^* | H_P | \gamma, \beta_\gamma^* \rangle. \quad (99)$$

□

F Proof of Proposition 1

We present the proof of proposition 1, demonstrating that transforming an Ising model with fields into one without them can actually increase the complexity of finding optimal parameters, particularly in bounded-degree graphs where the maximum degree is much smaller than the number of vertices.

Proof of proposition 1. From theorem 2, the maximum angular frequency for an Ising model with external fields is given by:

$$\omega_{\max} [\langle H_P \rangle_\gamma] = \max \left\{ \omega_{\max} [\langle C_i \rangle_\gamma], \omega_{\max} [\langle C_{uv} \rangle_\gamma] \mid \forall i \in V, \forall \{u, v\} \in E \right\}. \quad (100)$$

For a triangle-free graph with bounded maximum degree $D_{\max}$, it follows that:

$$\omega_{\max} [\langle C_i \rangle_\gamma] = \omega_{\max} [\langle C_{uv} \rangle_\gamma] = 2(D_{\max} + 1). \quad (101)$$

Applying theorem 4, we derive the upper bound on the permissible spacing between consecutive samples:

$$\Delta_\gamma < \frac{\pi}{2(D_{\max} + 1)}. \quad (102)$$

When converting an Ising model with external fields of size n (i.e., n spins or vertices) to an equivalent Ising model without external fields, the maximum degree of the resulting graph increases from $D_{\max}$ to n . According to corollary 2, the maximum angular frequency for an Ising model without external fields can be computed using eq. (35). Although the resulting graph may contain triangles, the maximum contribution to the angular frequency arises from the non-triangle terms. Therefore, the maximum angular frequency simplifies to:

$$\omega_{\max} [\langle C_{uv} \rangle_{\gamma}] = 2 \left(|J_{uv}| + \max \left\{ \sum_{w \in e} |J_{wv}|, \sum_{w \in d} |J_{uw}| \right\} \right) \quad (103)$$

$$= 2 \left(1 + \sum_{w \in d} |J_{uw}| \right) \quad (104)$$

$$= 2(1 + (n - 1)) \quad (105)$$

$$= 2n. \quad (106)$$

Applying theorem 4 again, we obtain the bound on the maximum permissible spacing between consecutive samples for the transformed model:

$$\Delta'_{\gamma} < \frac{\pi}{2n}. \quad (107)$$

Taking the ratio of Δ_{γ} to Δ'_{γ} , we have:

$$\frac{\Delta_{\gamma}}{\Delta'_{\gamma}} < \frac{n}{D_{\max} + 1}. \quad (108)$$

Thus, when an Ising model with external fields, whose underlying graph is triangle-free with maximum degree $D_{\max}$, is converted to an equivalent model without external fields, the maximum permissible spacing between consecutive samples decreases by a factor of $\frac{n}{D_{\max}+1}$. $\square$

G Steckin's Lemma for the Subdivision Algorithm

In this section, we introduce Steckin's Lemma and demonstrate its application as a rejection criterion in the subdivision algorithm for estimating the maximum modulus of trigonometric polynomials and identifying their optima. We begin by presenting Steckin's Lemma for univariate real-valued trigonometric polynomials:

Lemma 10 (Steckin's Lemma). *Suppose that the polynomial*

$$f(t) = \sum_{|n| \leq N} \alpha_n e^{int} \quad (\alpha_n \in \mathbf{C}) \quad (109)$$

is real-valued and that $t_0 \in [0, 2\pi]$ satisfies $f(t_0) = \|f\|_{\infty}$. Then

$$f(t_0 + s) \geq \|f\|_{\infty} \cos Ns \quad (|s| \leq \pi/N). \quad (110)$$

The proof of theorem 10 is outlined in Exercise 1.8 of [111]. Stecklin's Lemma offers valuable insight into the behaviour of real-valued trigonometric polynomials at their maximum points. Specifically, it states that if a polynomial $f(t)$ attains its maximum value at t_0 , then within a neighbourhood around t_0 (where the distance s satisfies $|s| \leq \pi/N$), the polynomial does not decrease below $\|f\|_{\infty} \cos(Ns)$. This ensures that $f(t)$ retains a significant portion of its maximum value near t_0 . Such a property is crucial in applications like signal processing and approximation theory, where understanding the local behaviour of trigonometric polynomials is essential for accurate analysis and optimisation.

Now, given an analytic polynomial p , we can define its maximum modulus on the unit disk as

$$\|p\|_\infty := \sup\{|p(z)| : z \in \mathbb{C}, |z| \leq 1\} \quad (111)$$

Since $\|p\|_\infty$ is attained by $|p|$ on the boundary of the unit disc, our objective is to find bounds on

$$\sup \left\{ \left| p \left(e^{it} \right) \right| : t \in [0, 2\pi] \right\}, \quad (112)$$

and since our calculations are simplified if we work with the square of $|p|$, we write

$$q(t) := \left| p \left(e^{it} \right) \right|^2 = p \left(e^{it} \right) \overline{p \left(e^{it} \right)}. \quad (113)$$

Suppose q has been evaluated at t_k . If $h < \pi/N$ and the global maximiser t_0 lies within the interval $[t_k - h, t_k + h]$, then

$$q(t_k) \geq \|q\|_\infty \cos N(t_k - t_0) \geq \|q\|_\infty \cos Nh \quad (114)$$

as shown by theorem 10. Consequently, if the global maximiser is known to lie within one of the intervals $I_1, \dots, I_n$ (each of size h), and q has been evaluated at their midpoints $t_1, \dots, t_n$, then there exists some k such that $\|q\|_\infty \leq q(t_k) \sec Nh$. Define $\tilde{q} := \max \{q(t_i) : i = 1, \dots, n\}$. Then, the following bounds hold:

$$\tilde{q} \leq \|q\|_\infty \leq \tilde{q} \sec Nh. \quad (115)$$

Moreover, if $q(t_i) < \tilde{q} \cos Nh$, the global maximiser cannot lie within the interval I_i , allowing us to discard it. These observations lead to a straightforward algorithm for calculating $\|p\|_\infty$ and determining the location of the global optimum, as outlined in algorithm 1.

Chevrotiere [77] extended Green's [76] work by generalising theorem 10 to multivariate polynomials and expanding the subdivision algorithm to estimate the maximum modulus of multivariate trigonometric polynomials on polydisks. The Multivariate Stecklin's Lemma is formalised as follows:

Lemma 11 (Multivariate Stecklin's Lemma). *Let $g(t)$ be a real-valued multivariate trigonometric polynomial defined by*

$$g(t) = \sum_{\substack{|n_j| \leq d_j \\ j=1, \dots, k}} c_{n_1, \dots, n_k} e^{i(n_1 t_1 + \dots + n_k t_k)}, \quad (116)$$

where $t = (t_1, \dots, t_k) \in [0, 2\pi]^k$ and $c_{n_1, \dots, n_k} \in \mathbb{C}$. Suppose there exists a point $t_0 \in [0, 2\pi]^k$ such that $g(t_0) = \|g\|_\infty$. Then, for any $s = (s_1, \dots, s_k) \in \mathbb{R}^k$ satisfying $d_1 |s_1| + \dots + d_k |s_k| \leq \pi$, the polynomial g satisfies

$$g(t_0 + s) \geq \|g\|_\infty \cos(d_1 |s_1| + \dots + d_k |s_k|). \quad (117)$$

The proof of theorem 11 is provided in Chevrotiere [77]. Chevrotiere demonstrates that by using theorem 11 as a rejection criterion, Green's [76] subdivision algorithm can be effectively extended to arbitrary dimensions.

Let us now address the optimisation of the univariate functions defined in theorem 5 and theorem 6. Suppose we optimise over both parameters (γ, β) using the expressions from theorem 1 and corollary 1. In this scenario, theorem 11 provides the following bound:

$$\langle C(\gamma^* + s_1, \beta^* + s_2) \rangle^2 \geq \langle C(\gamma^*, \beta^*) \rangle^2 \cos \left(\nu_{\max}^\gamma |s_1| + \nu_{\max}^\beta |s_2| \right), \quad (118)$$

where $\nu_{\max}^\gamma$ and $\nu_{\max}^\beta$ are the maximum frequencies of the γ and β terms, respectively, and $|s_1|$ and $|s_2|$ denote the distances from the global maximum satisfying $\nu_{\max}^\gamma |s_1| + \nu_{\max}^\beta |s_2| < \pi$.

However, if we restrict our search to (γ, β^*) as in theorem 5, then $|s_2| = 0$, and the bound simplifies to the same form as in eq. (114). In the context of theorem 6, we only know a locally optimal β_γ^* for each γ ; if $\beta_\gamma^* = \beta^*$ for any $\gamma \in [0, \pi]$, then $|s_2| = 0$ again, and the bound remains unchanged. If instead $|s_2| \neq 0$ (i.e., $\beta_\gamma^* \neq \beta^*$), then using the fact that $\nu_{\max}^\gamma |s_1| + \nu_{\max}^\beta |s_2| < \pi$, it follows that

$$\cos(\nu_{\max}^\gamma |s_1| + \nu_{\max}^\beta |s_2|) < \cos(\nu_{\max}^\gamma |s_1|), \quad (119)$$

which leads to the rejection of that particular interval. Therefore, in optimising the univariate function of γ , it suffices to treat both expressions in theorem 5 and theorem 6 as univariate trigonometric polynomials. Consequently, we can apply Green's [76] subdivision algorithm to efficiently estimate the optimal value γ^* .

Theorem 12. *Let $p : [0, 2\pi] \rightarrow \mathbb{R}$ be a real trigonometric polynomial of degree N such that $\|p\|_\infty = 1$. The subdivision algorithm requires $\mathcal{O}(N/\sqrt{\varepsilon})$ function evaluations to approximate $\max |p(x)|^2$ to within an additive tolerance $\varepsilon > 0$.*

Proof. The algorithm begins by subdividing the interval $[0, 2\pi]$ into M subintervals, each of length $\Delta_\gamma = \pi/M$, where $M > 2N$. Consequently, the initial number of subintervals satisfies $M = \mathcal{O}(N)$. Although the algorithm prunes many subintervals in each iteration—specifically those whose midpoint values are below a certain threshold—it is possible, in the worst case, to retain up to $\mathcal{O}(N)$ subintervals.

With each subdivision step, the number of subintervals doubles. Therefore, after i iterations, the number of subintervals can grow to $\mathcal{O}(N \cdot 2^i)$. Since the algorithm only evaluates the midpoint of each subinterval during subdivision, the total number of function evaluations after i iterations remains bounded by $\mathcal{O}(N \cdot 2^i)$.

The stopping criterion for Green's algorithm is that the quantity $\tilde{q}(\sec(N\Delta_\gamma) - 1)$ becomes sufficiently small. Here, $\tilde{q}$ represents the current estimate of the polynomial's maximum modulus, and Δ_γ is the current width of the subintervals. Given that $\|p\|_\infty = 1$, we approximate $\tilde{q} \approx 1$. For small Δ_γ , we can approximate:

$$\sec(N\Delta_\gamma) - 1 \approx \frac{(N\Delta_\gamma)^2}{2}. \quad (120)$$

To ensure that the algorithm's error is below a tolerance ε , we require:

$$\tilde{q}(\sec(N\Delta_\gamma) - 1) \lesssim \varepsilon \quad \Rightarrow \quad (N\Delta_\gamma)^2 \lesssim \varepsilon \quad \Rightarrow \quad \Delta_\gamma \lesssim \frac{\sqrt{\varepsilon}}{N}. \quad (121)$$

Since each iteration halves Δ_γ , starting from $\Delta_\gamma^{(0)} = \mathcal{O}\left(\frac{1}{N}\right)$, after i halvings we have:

$$\Delta_\gamma \approx \frac{1}{N} \frac{1}{2^i}. \quad (122)$$

To satisfy the error tolerance, we set:

$$\frac{1}{N \cdot 2^i} \lesssim \frac{\sqrt{\varepsilon}}{N} \quad \Rightarrow \quad 2^{-i} \lesssim \sqrt{\varepsilon} \quad \Rightarrow \quad 2^i \approx \frac{1}{\sqrt{\varepsilon}}. \quad (123)$$

Substituting $2^i \approx \frac{1}{\sqrt{\varepsilon}}$ into the total number of function evaluations, we obtain $\mathcal{O}(N/\sqrt{\varepsilon})$. $\square$

H Globally Optimal Parameters for Regular Graphs

H.1 Mathematical Preliminaries

H.1.1 Probability Density Functions

In this subsection, we introduce the necessary definitions and notations related to probability density functions (PDFs) that are essential for our proof. We begin by defining the expectation operator for functions of random variables and extend these definitions to multiple variables and arbitrary index sets.

Single-Variable Expectation

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a probability density function (PDF). For a measurable function $g : \mathbb{R} \rightarrow \mathbb{R}$, the expectation of $g(x)$ with respect to f is denoted and defined as

$$\mathbb{E}_{x \sim f}[g(x)] = \int_{\mathbb{R}} f(x)g(x)dx. \quad (124)$$

This integral represents the average value of $g(x)$ when x is distributed according to the PDF f .

Multi-Variable Expectation for n Independent Variables

Consider n independent random variables $x_1, x_2, \dots, x_n$, each distributed according to the same PDF f . We denote the joint expectation over these n variables as $\mathbb{E}_{x \sim f^n}$. Formally, for a function $g : \mathbb{R}^n \rightarrow \mathbb{R}$, the expectation is defined by

$$\mathbb{E}_{x \sim f^n}[g(x)] = \mathbb{E}_{x_1 \sim f} \cdots \mathbb{E}_{x_n \sim f}[g(x_1, \dots, x_n)] \quad (125)$$

$$= \int_{\mathbb{R}^n} f(x_1) f(x_2) \cdots f(x_n) g(x_1, \dots, x_n) dx_1 dx_2 \cdots dx_n. \quad (126)$$

This expression computes the expectation of g over the product measure f^n , reflecting the independence of each x_i .

Expectation over an Arbitrary Index Set

Let E be an arbitrary index set, and for each $e \in E$, let x_e be a random variable distributed according to the PDF f . We define the expectation operator over the collection of variables indexed by E as

$$\mathbb{E}_{x \sim f^E} = \prod_{e \in E} \mathbb{E}_{x_e \sim f}. \quad (127)$$

For a function $g : \mathbb{R}^E \rightarrow \mathbb{R}$, the expectation is given by

$$\mathbb{E}_{x \sim f^E}[g(x)] = \prod_{e \in E} \mathbb{E}_{x_e \sim f}[g((e)_{e \in E})] \quad (128)$$

$$= \mathbb{E}_{x \sim f^E}[g(x)] = \int_{\mathbb{R}^E} \prod_{e \in E} f(x_e) g((x_e)_{e \in E}) \prod_{e \in E} dx_e. \quad (129)$$

This generalises the expectation to any index set E , allowing for flexibility in handling functions of numerous variables.

H.1.2 Interchanging Summation and Integration in Expectations

Throughout this section, we assume that the probability density functions p are chosen such that the infinite sum arising from the Taylor series expansion and the integrals involved in taking expectations can be interchanged. This assumption is crucial for the validity of the subsequent manipulations.

Consider a function f expressed as its Taylor series expansion:

$$f(x) = \sum_{i=0}^{\infty} a_i x^i \quad (130)$$

Under our assumption, we can compute the expectation of $f(x)$ with respect to the probability distri-

bution p by interchanging the summation and the integral. The steps are as follows:

$$\mathbb{E}_{x \sim p} [f(x)] = \mathbb{E}_{x \sim p} \left[\sum_{i=0}^{\infty} a_i x^i \right] \quad (131)$$

$$= \int dx p(x) \sum_{i=0}^{\infty} a_i x^i \quad (132)$$

$$= \int dx \underbrace{\sum_{i=0}^{\infty} p(x) a_i x^i}_{=\sum_{i=0}^{\infty} \int dx} \quad (133)$$

$$= \sum_{i=0}^{\infty} a_i \int dx p(x) x^i \quad (134)$$

$$= \sum_{i=0}^{\infty} a_i \mathbb{E}_{x \sim p} [x^i]. \quad (135)$$

H.2 Some Useful Lemmas

Lemma 13. *Let x be a random variable distributed according to $p(x)$, and let D be a large parameter. Then the expectation of $x \sin\left(\frac{ax}{\sqrt{D}}\right)$ can be approximated by:*

$$\begin{aligned} \mathbb{E}_{x \sim p} \left[x \sin\left(\frac{ax}{\sqrt{D}}\right) \right] &= \frac{a}{\sqrt{D}} \mathbb{E} [h^2] - \frac{a^3}{6D^{3/2}} \mathbb{E} [h^4] + \mathcal{O}\left(D^{-\frac{5}{2}}\right), \\ &= \frac{a}{\sqrt{D}} \mathbb{E} [h^2] + \mathcal{O}\left(D^{-\frac{3}{2}}\right). \end{aligned} \quad (136)$$

Proof.

$$\mathbb{E}_{x \sim p} \left[x \sin \frac{ax}{\sqrt{D}} \right] = \mathbb{E}_{x \sim p} \left[x \sum_{\substack{k=1 \\ \text{odd}}}^{\infty} \frac{(-1)^{\frac{k-1}{2}}}{k!} \left(\frac{ax}{\sqrt{D}} \right)^k \right] \quad \left(\text{using } \sin x = \sum_{\substack{k=1 \\ \text{odd}}}^{\infty} \frac{(-1)^{\frac{k-1}{2}}}{k!} x^k \right) \quad (137)$$

$$= \mathbb{E}_{x \sim p} \left[\sum_{\substack{k=1 \\ \text{odd}}}^{\infty} \frac{(-1)^{\frac{k-1}{2}}}{k!} \frac{a^k}{D^{\frac{k}{2}}} x^{k+1} \right] \quad (138)$$

$$= \sum_{\substack{k=1 \\ \text{odd}}}^{\infty} \mathbb{E}_{x \sim p} \left[\frac{(-1)^{\frac{k-1}{2}}}{k!} \frac{a^k}{D^{\frac{k}{2}}} x^{k+1} \right] \quad (139)$$

$$= \sum_{\substack{k=1 \\ \text{odd}}}^{\infty} \frac{(-1)^{\frac{k-1}{2}}}{k!} \frac{a^k}{D^{\frac{k}{2}}} \mathbb{E}_{x \sim p} [x^{k+1}] \quad (140)$$

$$= \frac{a}{\sqrt{D}} \mathbb{E}_{x \sim p} [h^2] - \frac{1}{3!} \frac{a^3}{D^{\frac{3}{2}}} \mathbb{E}_{x \sim p} [h^4] + \mathcal{O}\left(\frac{1}{D^{\frac{5}{2}}}\right) \quad (141)$$

$$= \frac{a}{\sqrt{D}} \mathbb{E}_{x \sim p} [h^2] + \mathcal{O}\left(\frac{1}{D^{\frac{3}{2}}}\right). \quad (142)$$

□

Lemma 14. *Let a and b be parameters independent of D , and let D be a large parameter. Then the following approximations hold:*

$$\left[1 + \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right) \right]^D = e^a \left(1 + \left(b - \frac{a^2}{2} \right) \frac{1}{D} + \mathcal{O}\left(\frac{1}{D^2}\right) \right), \quad (143)$$

and:

$$\left[1 + \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right)\right]^{D+1} = e^a \left(1 + \frac{a + b - \frac{a^2}{2}}{D} + \mathcal{O}\left(\frac{1}{D^2}\right)\right). \quad (144)$$

Proof. Using the Taylor expansion:

$$\ln(1+x) = x - \frac{x^2}{2} + \mathcal{O}(x^3), \quad (145)$$

we expand:

$$\ln\left(1 + \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right)\right) = \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right) - \frac{1}{2}\left(\frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right)\right)^2 + \mathcal{O}\left(\frac{1}{D^3}\right) \quad (146)$$

$$= \frac{a}{D} + \frac{b}{D^2} - \frac{a^2}{2D^2} + \mathcal{O}\left(\frac{1}{D^3}\right) \quad (147)$$

$$= \frac{a}{D} + \left(b - \frac{a^2}{2}\right) \frac{1}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right). \quad (148)$$

Multiplying by D , we find:

$$D \ln\left(1 + \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right)\right) = a + \left(b - \frac{a^2}{2}\right) \frac{1}{D} + \mathcal{O}\left(\frac{1}{D^2}\right). \quad (149)$$

1. For the first result, we compute:

$$\left[1 + \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right)\right]^D = e^{D \ln\left(1 + \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right)\right)} \quad (150)$$

$$= e^{a + \left(b - \frac{a^2}{2}\right) \frac{1}{D} + \mathcal{O}\left(\frac{1}{D^2}\right)} \quad (151)$$

$$= e^a \cdot e^{\left(b - \frac{a^2}{2}\right) \frac{1}{D} + \mathcal{O}\left(\frac{1}{D^2}\right)} \quad (152)$$

$$= e^a \left(1 + \left(b - \frac{a^2}{2}\right) \frac{1}{D} + \mathcal{O}\left(\frac{1}{D^2}\right)\right). \quad (153)$$

where the Taylor expansion:

$$e^x = 1 + x + \frac{x^2}{2} + \mathcal{O}(x^3)$$

is applied in the last step.

2. For the second result:

$$\left[1 + \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right)\right]^{D+1} = e^a \left(1 + \left(b - \frac{a^2}{2}\right) \frac{1}{D} + \mathcal{O}\left(\frac{1}{D^2}\right)\right) \left[1 + \frac{a}{D} + \mathcal{O}\left(\frac{1}{D^2}\right)\right] \quad (154)$$

$$= e^a \left(1 + \left(b - \frac{a^2}{2}\right) \frac{1}{D} + \frac{a}{D} + \mathcal{O}\left(\frac{1}{D^2}\right)\right) \quad (155)$$

$$= e^a \left(1 + \frac{a + b - \frac{a^2}{2}}{D} + \mathcal{O}\left(\frac{1}{D^2}\right)\right). \quad (156)$$

□

Lemma 15. *Let x be a random variable distributed according to $p(x)$, and let D be a large parameter. Then the following approximations hold:*

1. The expectation of $\cos\left(\frac{ax}{\sqrt{D}}\right)$ is given by:

$$\mathbb{E}_{x \sim p} \left[\cos \frac{ax}{\sqrt{D}} \right] = 1 - \frac{1}{2} \frac{a^2}{D} \mathbb{E}_{x \sim p} [x^2] + \frac{1}{4!} \frac{a^4}{D^2} \mathbb{E}_{x \sim p} [x^4] + \mathcal{O}\left(\frac{1}{D^3}\right). \quad (157)$$

2. For the D -th power of expectation:

$$\begin{aligned} \left(\mathbb{E}_{x \sim p} \left[\cos \frac{ax}{\sqrt{D}} \right] \right)^D &= e^{-\frac{1}{2} a^2 \mathbb{E}_{x \sim p} [x^2]} \left[1 + \frac{1}{8} a^4 \left(\frac{1}{3} \mathbb{E}_{x \sim p} [x^4] - \left(\mathbb{E}_{x \sim p} [x^2] \right)^2 \right) \frac{1}{D} + \mathcal{O}\left(\frac{1}{D^2}\right) \right] \\ &= e^{-\frac{1}{2} a^2 \mathbb{E}_{x \sim p} [x^2]} \left(1 + \mathcal{O}\left(\frac{1}{D}\right) \right). \end{aligned} \quad (158)$$

3. For the $(D+1)$ -th power of expectation:

$$\begin{aligned} \left(\mathbb{E}_{x \sim p} \left[\cos \frac{ax}{\sqrt{D}} \right] \right)^{D+1} &= e^{-\frac{1}{2} a^2 \mathbb{E}_{x \sim p} [x^2]} \left[1 + \left(-\frac{1}{2} a^2 \mathbb{E}_{x \sim p} [x^2] + \frac{1}{4!} a^4 \mathbb{E}_{x \sim p} [x^4] \right. \right. \\ &\quad \left. \left. - \frac{1}{8} a^4 \left(\mathbb{E}_{x \sim p} [x^2] \right)^2 \right) \frac{1}{D} + \mathcal{O}\left(\frac{1}{D^2}\right) \right] \\ &= e^{-\frac{1}{2} a^2 \mathbb{E}_{x \sim p} [x^2]} \left(1 + \mathcal{O}\left(\frac{1}{D}\right) \right). \end{aligned} \quad (159)$$

Proof. 1. For the first result, we compute:

$$\mathbb{E}_{x \sim p} \left[\cos \frac{ax}{\sqrt{D}} \right] = \mathbb{E}_{x \sim p} \left[\sum_{\substack{k=0 \\ k \text{ even}}}^{\infty} \frac{(-1)^{k/2}}{k!} \left(\frac{ax}{\sqrt{D}} \right)^k \right] \quad \left(\text{using } \cos x = \sum_{\substack{k=0 \\ k \text{ even}}}^{\infty} \frac{(-1)^{k/2}}{k!} x^k \right) \quad (160)$$

$$= \sum_{\substack{k=0 \\ k \text{ even}}}^{\infty} \mathbb{E}_{x \sim p} \left[\frac{(-1)^{k/2}}{k!} \left(\frac{ax}{\sqrt{D}} \right)^k \right] \quad (161)$$

$$= \sum_{\substack{k=0 \\ k \text{ even}}}^{\infty} \frac{(-1)^{k/2}}{k!} \frac{a^k}{D^{k/2}} \mathbb{E}_{x \sim p} [x^k] \quad (162)$$

$$= 1 - \frac{1}{2} \frac{a^2}{D} \mathbb{E}_{x \sim p} [x^2] + \frac{1}{4!} \frac{a^4}{D^2} \mathbb{E}_{x \sim p} [x^4] + \mathcal{O}\left(\frac{1}{D^3}\right). \quad (163)$$

2. Using the eq. (157), we expand its D -th power:

$$\left(\mathbb{E}_{x \sim p} \left[\cos \frac{ax}{\sqrt{D}} \right] \right)^D = \left[1 - \frac{1}{2} \frac{a^2}{D} \mathbb{E}_{x \sim p} [x^2] + \frac{1}{4!} \frac{a^4}{D^2} \mathbb{E}_{x \sim p} [x^4] + \mathcal{O}\left(\frac{1}{D^3}\right) \right]^D \quad (164)$$

Let:

$$s = -\frac{1}{2} a^2 \mathbb{E}_{x \sim p} [x^2], \quad (165)$$

$$t = \frac{1}{4!} a^4 \mathbb{E}_{x \sim p} [x^4]. \quad (166)$$

Then:

$$t - \frac{s^2}{2} = \frac{1}{4!} a^4 \mathbb{E}_{x \sim p} [x^4] - \frac{1}{2} \left(-\frac{1}{2} a^2 \mathbb{E}_{x \sim p} [x^2] \right)^2 \quad (167)$$

$$= \frac{1}{4!} a^4 \mathbb{E}_{x \sim p} [x^4] - \frac{1}{8} a^4 \left(\mathbb{E}_{x \sim p} [x^2] \right)^2 \quad (168)$$

$$= \frac{1}{8} a^4 \left[\frac{1}{3} \mathbb{E}_{x \sim p} [x^4] - \left(\mathbb{E}_{x \sim p} [x^2] \right)^2 \right]. \quad (169)$$

Using eq. (143) of lemma 14, we find:

$$\left(\mathbb{E}_{x \sim p} \left[\cos \frac{ax}{\sqrt{D}} \right] \right)^D = e^{-\frac{1}{2}a^2 \mathbb{E}_{x \sim p}[x^2]} \left[1 + \frac{1}{8}a^4 \left(\frac{1}{3} \mathbb{E}_{x \sim p}[x^4] - \left(\mathbb{E}_{x \sim p}[x^2] \right)^2 \right) \frac{1}{D} + \mathcal{O}\left(\frac{1}{D^2}\right) \right]. \quad (170)$$

3. Similarly, for the $(D+1)$ -th power:

$$\left(\mathbb{E}_{x \sim p} \left[\cos \frac{ax}{\sqrt{D}} \right] \right)^{D+1} = \left[1 - \frac{1}{2} \frac{a^2}{D} \mathbb{E}_{x \sim p}[x^2] + \frac{1}{4!} \frac{a^4}{D^2} \mathbb{E}_{x \sim p}[x^4] + \mathcal{O}\left(\frac{1}{D^3}\right) \right]^{D+1}. \quad (171)$$

Using the definitions of s and t from eq. (165) and eq. (166), we compute:

$$s + t - \frac{s^2}{2} = -\frac{1}{2}a^2 \mathbb{E}_{x \sim p}[x^2] + \frac{1}{8}a^4 \left[\frac{1}{3} \mathbb{E}_{x \sim p}[x^4] - \left(\mathbb{E}_{x \sim p}[x^2] \right)^2 \right]. \quad (172)$$

Using the eq. (144) of lemma 14, we get:

$$\begin{aligned} \left(\mathbb{E}_{x \sim p} \left[\cos \frac{ax}{\sqrt{D}} \right] \right)^{D+1} &= e^{-\frac{1}{2}a^2 \mathbb{E}_{x \sim p}[x^2]} \left[1 + \left(-\frac{1}{2}a^2 \mathbb{E}_{x \sim p}[x^2] + \frac{1}{4!}a^4 \mathbb{E}_{x \sim p}[x^4] \right. \right. \\ &\quad \left. \left. - \frac{1}{8}a^4 \left(\mathbb{E}_{x \sim p}[x^2] \right)^2 \right) \frac{1}{D} + \mathcal{O}\left(\frac{1}{D^2}\right) \right]. \end{aligned} \quad (173)$$

□

Lemma 16. *Let x and x' be independent random variables distributed according to $p(x)$, and let D be a large parameter. Then the following approximation holds:*

$$\mathbb{E}_{x, x' \sim p} \left[\cos \left((x \pm x') \frac{a}{\sqrt{D}} \right) \right] = 1 - \frac{a^2}{D} \left[\mathbb{E}_{x \sim p}[x^2] \pm \left(\mathbb{E}_{x \sim p}[x] \right)^2 \right] + \mathcal{O}\left(\frac{1}{D^2}\right). \quad (174)$$

Proof. Using eq. (157) of lemma 15, we have:

$$\mathbb{E}_{x, x' \sim p} \left[\cos \left((x \pm x') \frac{a}{\sqrt{D}} \right) \right] = 1 - \frac{1}{2} \frac{a^2}{D} \underbrace{\mathbb{E}_{x, x' \sim p} \left[(x \pm x')^2 \right]}_{\textcircled{1}} + \mathcal{O}\left(\frac{1}{D^2}\right). \quad (175)$$

Expanding $\textcircled{1}$, we get:

$$\mathbb{E}_{x, x' \sim p} \left[(x \pm x')^2 \right] = \mathbb{E}_{x, x' \sim p} \left[x^2 + x'^2 \pm 2xx' \right] \quad (176)$$

$$= \mathbb{E}_{x \sim p}[x^2] + \mathbb{E}_{x' \sim p}[x'^2] \pm 2 \left(\mathbb{E}_{x \sim p}[x] \right) \left(\mathbb{E}_{x' \sim p}[x'] \right) \quad (177)$$

$$= 2 \mathbb{E}_{x \sim p}[x^2] \pm 2 \left(\mathbb{E}_{x \sim p}[x] \right)^2. \quad (178)$$

Substituting $\textcircled{1}$ back, we obtain:

$$\mathbb{E}_{x, x' \sim p} \left[\cos \left((x \pm x') \frac{a}{\sqrt{D}} \right) \right] = 1 - \frac{1}{2} \frac{a^2}{D} \left[2 \mathbb{E}_{x \sim p}[x^2] \pm 2 \left(\mathbb{E}_{x \sim p}[x] \right)^2 \right] + \mathcal{O}\left(\frac{1}{D^2}\right) \quad (179)$$

$$= 1 - \frac{a^2}{D} \left[\mathbb{E}_{x \sim p}[x^2] \pm \left(\mathbb{E}_{x \sim p}[x] \right)^2 \right] + \mathcal{O}\left(\frac{1}{D^2}\right). \quad (180)$$

□

Lemma 17. Let $a, b, c \in \mathbb{R}$, and let $\lambda \in [0, 1]$. For a large parameter D , the following approximation holds:

$$\left[1 + \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right)\right]^{cD^\lambda} = \begin{cases} e^{ac} \left(1 + c \left(b - \frac{a^2}{2}\right) D^{-1} + \mathcal{O}(D^{-2})\right) & \text{if } \lambda = 1, \\ 1 + acD^{\lambda-1} + \mathcal{O}(D^{2\lambda-2}) & \text{if } 0 \leq \lambda < 1. \end{cases} \quad (181)$$

Proof. Recall from the proof of lemma 14, we know that:

$$\ln \left(1 + \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right)\right) = \frac{a}{D} + \left(b - \frac{a^2}{2}\right) \frac{1}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right). \quad (182)$$

Multiplying by cD^λ , we obtain:

$$cD^\lambda \ln \left(1 + \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^2}\right)\right) = c \left[aD^{\lambda-1} + \left(b - \frac{a^2}{2}\right) D^{\lambda-2} + \mathcal{O}(D^{\lambda-3}) \right]. \quad (183)$$

Expanding the LHS of eq. (181), we get:

$$\left[1 + \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right)\right]^{cD^\lambda} = \left[1 + \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right)\right]^{cD^\lambda} \quad (184)$$

$$= e^{cD^\lambda \ln \left(1 + \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right)\right)} \quad (185)$$

$$= e^{c \left[aD^{\lambda-1} + \left(b - \frac{a^2}{2}\right) D^{\lambda-2} + \mathcal{O}(D^{\lambda-3}) \right]}. \quad (186)$$

We now handle two cases based on the value of λ .

Case 1: $\lambda = 1$

$$\left[1 + \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right)\right]^{cD^\lambda} = e^{c \left[a + \left(b - \frac{a^2}{2}\right) D^{-1} + \mathcal{O}(D^{-2}) \right]} \quad (187)$$

$$= e^{ac} e^{c \left(b - \frac{a^2}{2}\right) D^{-1} + \mathcal{O}(D^{-2})} \quad (188)$$

$$= e^{ac} \left(1 + c \left(b - \frac{a^2}{2}\right) D^{-1} + \mathcal{O}(D^{-2})\right). \quad (189)$$

Case 2: $\lambda < 1$

$$\left[1 + \frac{a}{D} + \frac{b}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right)\right]^{cD^\lambda} = e^{c \left[aD^{\lambda-1} + \left(b - \frac{a^2}{2}\right) D^{\lambda-2} + \mathcal{O}(D^{\lambda-3}) \right]} \quad (190)$$

$$= 1 + c \left[aD^{\lambda-1} + \left(b - \frac{a^2}{2}\right) D^{\lambda-2} + \mathcal{O}(D^{\lambda-3}) \right] \quad (191)$$

$$+ c^2 \left[aD^{\lambda-1} + \left(b - \frac{a^2}{2}\right) D^{\lambda-2} + \mathcal{O}(D^{\lambda-3}) \right]^2 + \mathcal{O}(D^{3(\lambda-1)}) \quad (192)$$

$$= 1 + acD^{\lambda-1} + a^2c^2D^{2\lambda-2} + \mathcal{O}(D^{\lambda-2}) \quad (193)$$

$$= 1 + acD^{\lambda-1} + \mathcal{O}(D^{2\lambda-2}). \quad (194)$$

□

Lemma 18. Let x be a random variable distributed according to $p(x)$, with $c \in \mathbb{R}$, a independent of D , and $\lambda \in [0, 1]$. For large D , the following approximation holds:

$$\left(\mathbb{E}_{x \sim p} \left[\cos \frac{ax}{\sqrt{D}} \right] \right)^{cD^\lambda} = \begin{cases} e^{-\frac{c}{2}a^2 \mathbb{E}_{x \sim p}[x^2]} \left[1 + \frac{1}{8}a^4c \left(\frac{1}{3} \mathbb{E}_{x \sim p}[x^4] - \left(\mathbb{E}_{x \sim p}[x^2] \right)^2 \right) \frac{1}{D} + \mathcal{O}\left(\frac{1}{D^2}\right) \right] & \lambda = 1, \\ 1 - \frac{c}{2}a^2 \mathbb{E}_{x \sim p}[x^2] D^{\lambda-1} + \mathcal{O}\left(D^{2\lambda-2}\right) & 0 \leq \lambda < 1. \end{cases} \quad (195)$$

Proof. Raising eq. (157) of lemma 15 to the power cD^λ , we have:

$$\left(\mathbb{E}_{x \sim p} \left[\cos \frac{ax}{\sqrt{D}} \right] \right)^{cD^\lambda} = \left[1 - \frac{1}{2} \frac{a^2}{D} \mathbb{E}_{x \sim p}[x^2] + \frac{1}{4!} \frac{a^4}{D^2} \mathbb{E}_{x \sim p}[x^4] + \mathcal{O}\left(\frac{1}{D^3}\right) \right]^{cD^\lambda}. \quad (196)$$

Using lemma 17, we evaluate the expression for the two cases:

Case 1: $\lambda = 1$

$$\left(\mathbb{E}_{x \sim p} \left[\cos \frac{ax}{\sqrt{D}} \right] \right)^{cD^\lambda} = e^{-\frac{c}{2}a^2 \mathbb{E}_{x \sim p}[x^2]} \left[1 + \frac{1}{8}a^4c \left(\frac{1}{3} \mathbb{E}_{x \sim p}[x^4] - \left(\mathbb{E}_{x \sim p}[x^2] \right)^2 \right) \frac{1}{D} + \mathcal{O}\left(\frac{1}{D^2}\right) \right]. \quad (197)$$

Case 2: $0 \leq \lambda < 1$

$$\left(\mathbb{E}_{x \sim p} \left[\cos \frac{ax}{\sqrt{D}} \right] \right)^{cD^\lambda} = 1 - \frac{c}{2}a^2 \mathbb{E}_{x \sim p}[x^2] D^{\lambda-1} + \mathcal{O}\left(D^{2\lambda-2}\right). \quad (198)$$

□

Lemma 19. Let x and x' be independent random variables distributed according to $p(x)$, with $c \in \mathbb{R}$, a independent of D , and $\lambda \in [0, 1]$. For large D , the following approximation holds:

$$\left[\mathbb{E}_{x, x' \sim p} \left[\cos \left((x \pm x') \frac{a}{\sqrt{D}} \right) \right] \right]^{cD^\lambda} = \begin{cases} e^{-a^2c \left[\mathbb{E}_{x \sim p}[x^2] \pm \left(\mathbb{E}_{x \sim p}[x] \right)^2 \right]} (1 + \mathcal{O}(D^{-1})) & \text{if } \lambda = 1, \\ 1 - a^2c \left[\mathbb{E}_{x \sim p}[x^2] \pm \left(\mathbb{E}_{x \sim p}[x] \right)^2 \right] D^{\lambda-1} + \mathcal{O}(D^{2\lambda-2}) & \text{if } 0 \leq \lambda < 1. \end{cases} \quad (199)$$

Proof. Raising eq. (174) of lemma 16 to the power cD^λ , we get:

$$\left[\mathbb{E}_{x, x' \sim p} \left[\cos \left((x \pm x') \frac{a}{\sqrt{D}} \right) \right] \right]^{cD^\lambda} = \left[1 - \frac{a^2}{D} \left[\mathbb{E}_{x \sim p}[x^2] \pm \left(\mathbb{E}_{x \sim p}[x] \right)^2 \right] + \mathcal{O}\left(\frac{1}{D^2}\right) \right]^{cD^\lambda}. \quad (200)$$

Using lemma 17, we obtain the stated result. □

H.3 Proof of Theorem 7

Before proceeding with the proof of theorem 7, we redefine some notations introduced earlier in the manuscript for improved clarity and compactness. These adjustments facilitate a more streamlined presentation of the proof. Specifically, for an edge $\{u, v\} \in E$ in a graph $G = (V, E)$, we define:

- $\mathcal{N}_{v \setminus u} = \mathcal{N}(v) \setminus \{u\}$: the set of vertices connected to v excluding u .
- $\mathcal{N}_{u \setminus v} = \mathcal{N}(u) \setminus \{v\}$: the set of vertices connected to u excluding v .
- $\mathcal{N}_{v \setminus \setminus u} = \{w \in \mathcal{N}(v) \mid w \neq u \text{ and } w \notin \mathcal{N}(u)\}$: the neighbours of v that are not neighbours of u .
- $\mathcal{N}_{u \setminus \setminus v} = \{w \in \mathcal{N}(u) \mid w \neq v \text{ and } w \notin \mathcal{N}(v)\}$: the neighbours of u that are not neighbours of v .
- $\mathcal{N}_{uv} = \mathcal{N}_{u \setminus v} \cap \mathcal{N}_{v \setminus u} = \mathcal{N}(u) \cap \mathcal{N}(v)$: the set of vertices that form a triangle with the edge $\{u, v\}$.

Here, $\mathcal{N}(w) = \{x \in V \mid \{x, w\} \in E\}$ denotes the set of neighbours of a vertex w .

Proof. Recall that the QAOA₁ cost function for an Ising model defined on a graph $G = (V, E)$ with respect to parameters $\beta, \gamma \in \mathbb{R}$ is

$$\xi_G(\gamma, \beta) = \sum_{i \in V} C_i(\gamma, \beta; G) + \sum_{\{u, v\} \in E} C_{uv}^{(1)}(\gamma, \beta; G) + \sum_{\{u, v\} \in E} C_{uv}^{(2)}(\gamma, \beta; G). \quad (201)$$

where the functions $C_i, C_{uv}^{(1)}$, and $C_{uv}^{(2)}$ are defined as follows:

$$C_i(\gamma, \beta; G) = h_i \sin(2\beta) \sin(2h_i\gamma) \prod_{k \in \mathcal{N}(i)} \cos(2J_{ik}\gamma), \quad (202)$$

$$C_{uv}^{(1)}(\gamma, \beta; G) = \frac{J_{uv}}{2} \sin(4\beta) \sin(2J_{uv}\gamma) \left\{ \cos(2h_v\gamma) \prod_{w \in \mathcal{N}_v \setminus u} \cos(2J_{vw}\gamma) + \cos(2h_u\gamma) \prod_{w \in \mathcal{N}_u \setminus v} \cos(2J_{uw}\gamma) \right\}, \quad (203)$$

$$C_{uv}^{(2)}(\gamma, \beta; G) = -\frac{J_{uv}}{2} \sin^2(2\beta) \prod_{w \in \mathcal{N}_v \setminus u} \cos(2J_{vw}\gamma) \prod_{w \in \mathcal{N}_u \setminus v} \cos(2J_{uw}\gamma) \\ \times \left\{ \cos(2(h_u + h_v)\gamma) \prod_{f \in \mathcal{N}_{uv}} \cos(2(J_{uf} + J_{vf})\gamma) - \cos(2(h_u - h_v)\gamma) \prod_{f \in \mathcal{N}_{uv}} \cos(2(J_{uf} - J_{vf})\gamma) \right\}. \quad (204)$$

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ be probability density functions. Consider a graph $G = (V, E)$ where the edge weights J_{uv} are independently drawn from f and the vertex weights h_i are independently drawn from g . The expected QAOA₁ cost function for G with parameters β and γ is then

$$\mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} [\xi_G(\gamma, \beta)] = \sum_{i \in V} \mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} [C_i(\gamma, \beta; G)] + \sum_{\{u, v\} \in E} \mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} \left\{ C_{uv}^{(1)}(\gamma, \beta; G) + \sum_{\{u, v\} \in E} C_{uv}^{(2)}(\gamma, \beta; G) \right\}. \quad (205)$$

To facilitate the analysis, we decompose the expectation into separate terms:

$$T_1 = \mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} [C_i(\gamma, \beta; G)], \quad (206)$$

$$T_2 = \mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} [C_{uv}^{(1)}(\gamma, \beta; G)], \quad (207)$$

$$T_3 = \mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} [C_{uv}^{(2)}(\gamma, \beta; G)]. \quad (208)$$

We will compute each of these terms individually to understand their contributions to the overall expected cost function.

Step 1: Computing the Term T_1

We begin by evaluating the expectation T_1 of the term $C_i(\gamma, \beta; G)$ over the distributions of

the edge weights $J \sim f^E$ and vertex weights $h \sim g^V$:

$$T_1 = \mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} [C_i(\gamma, \beta : G)] \quad (209)$$

$$= \mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} \left[h_i \sin(2\beta) \sin(2h_i\gamma) \prod_{k \in \mathcal{N}(i)} \cos(2J_{ik}\gamma) \right] \quad (210)$$

$$= \sin(4\beta) \mathbb{E}_{h \sim g} [h \sin(2h\gamma)] \prod_{k \in \mathcal{N}(i)} \mathbb{E}_{J \sim f} [\cos(2J\gamma)] \quad (211)$$

$$= \sin(4\beta) \mathbb{E}_{h \sim g} [h \sin(2h\gamma)] \left(\mathbb{E}_{J \sim f} [\cos(2J\gamma)] \right)^{|\mathcal{N}(i)|}. \quad (212)$$

Since the graph G is $(D+1)$ -regular, each vertex i has $|\mathcal{N}(i)| = D+1$ neighbors. By setting $\gamma = \frac{\alpha}{\sqrt{D}}$, the expression for T_1 simplifies to:

$$T_1 = \underbrace{\sin(4\beta) \mathbb{E}_{h \sim g} [h \sin(2h\gamma)]}_{\textcircled{1}} \underbrace{\left(\mathbb{E}_{J \sim f} [\cos(2J\gamma)] \right)^{D+1}}_{\textcircled{2}}. \quad (213)$$

Next, we apply theorem 13 and theorem 15 to approximate the terms $\textcircled{1}$ and $\textcircled{2}$ as follows:

$$\textcircled{1} = \mathbb{E}_{h \sim g} \left[h \sin \frac{2h\alpha}{\sqrt{D}} \right] = \frac{2\alpha}{\sqrt{D}} \mathbb{E}_{h \sim g} [h^2] + \mathcal{O} \left(\frac{1}{D^{3/2}} \right), \quad (214)$$

$$\textcircled{2} = \left(\mathbb{E}_{J \sim f} [\cos(2J\gamma)] \right)^{D+1} = e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \left(1 + \mathcal{O} \left(\frac{1}{D} \right) \right). \quad (215)$$

Substituting $\textcircled{1}$ and $\textcircled{2}$ back into the expression for T_1 , we obtain:

$$T_1 = \sin(4\beta) \left[\frac{2\alpha}{\sqrt{D}} \mathbb{E}_{h \sim g} [h^2] + \mathcal{O} \left(\frac{1}{D^{3/2}} \right) \right] e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \left(1 + \mathcal{O} \left(\frac{1}{D} \right) \right) \quad (216)$$

$$= \sin(4\beta) \mathbb{E}_{h \sim g} [h^2] \frac{2\alpha}{\sqrt{D}} e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} + \mathcal{O} \left(\frac{1}{D} \right). \quad (217)$$

Step 2: Computing the Term T_2

Next, we evaluate the expectation T_2 of the term $C_{uv}^{(1)}(\gamma, \beta; G)$ over the distributions of the edge weights $J \sim f^E$ and vertex weights $h \sim g^V$:

$$T_2 = \mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} [C_{uv}^{(1)}(\gamma, \beta; G)] \quad (218)$$

$$= \mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} \left[\frac{J_{uv}}{2} \sin(4\beta) \sin(2J_{uv}\gamma) \left\{ \cos(2h_v\gamma) \prod_{w \in \mathcal{N}_v \setminus u} \cos(2J_{vw}\gamma) + \cos(2h_u\gamma) \prod_{w \in \mathcal{N}_u \setminus v} \cos(2J_{uw}\gamma) \right\} \right] \quad (219)$$

$$= \frac{1}{2} \sin(4\beta) \mathbb{E}_{J \sim f} [J \sin(2J\gamma)] \left\{ \mathbb{E}_{h \sim g} [\cos(2h\gamma)] \prod_{w \in \mathcal{N}_v \setminus u} \mathbb{E}_{J \sim f} [\cos(2J\gamma)] + \mathbb{E}_{h \sim g} [\cos(2h\gamma)] \prod_{w \in \mathcal{N}_u \setminus v} \mathbb{E}_{J \sim f} [\cos(2J\gamma)] \right\} \quad (220)$$

$$= \frac{1}{2} \sin(4\beta) \mathbb{E}_{J \sim f} [J \sin(2J\gamma)] \mathbb{E}_{h \sim g} [\cos(2h\gamma)] \left\{ \left(\mathbb{E}_{J \sim f} [\cos(2J\gamma)] \right)^{|\mathcal{N}_v \setminus u|} + \left(\mathbb{E}_{J \sim f} [\cos(2J\gamma)] \right)^{|\mathcal{N}_u \setminus v|} \right\}. \quad (221)$$

Given that the graph G is $(D + 1)$ -regular, each edge $\{u, v\}$ has neighbourhoods satisfying:

$$|\mathcal{N}_{u \setminus v}| = |\mathcal{N}_{v \setminus u}| = (D + 1) - 1 = D. \quad (222)$$

Substituting this into the expression and setting $\gamma = \frac{\alpha}{\sqrt{D}}$, we further simplify T_2 as:

$$T_2 = \sin(4\beta) \underbrace{\mathbb{E}_{J \sim f} \left[J \sin \left(2J \frac{\alpha}{\sqrt{D}} \right) \right]}_{\textcircled{1}} \underbrace{\mathbb{E}_{h \sim g} \left[\cos \left(2h \frac{\alpha}{\sqrt{D}} \right) \right]}_{\textcircled{2}} \underbrace{\left(\mathbb{E}_{J \sim f} \left[\cos \left(2J \frac{\alpha}{\sqrt{D}} \right) \right] \right)^D}_{\textcircled{3}}. \quad (223)$$

Applying theorem 13 and theorem 15, we approximate the terms $\textcircled{1}$, $\textcircled{2}$, and $\textcircled{3}$ as follows:

$$\textcircled{1} = \mathbb{E}_{J \sim f} \left[J \sin \left(2J \frac{\alpha}{\sqrt{D}} \right) \right] = \frac{2\alpha}{\sqrt{D}} \mathbb{E}_{J \sim f} [J^2] \mathcal{O} \left(\frac{1}{D^{\frac{3}{2}}} \right), \quad (224)$$

$$\textcircled{2} = \mathbb{E}_{h \sim g} \left[\cos \left(2h \frac{\alpha}{\sqrt{D}} \right) \right] = 1 + \mathcal{O} \left(\frac{1}{D} \right), \quad (225)$$

$$\textcircled{3} = \left(\mathbb{E}_{J \sim f} \left[\cos \left(2J \frac{\alpha}{\sqrt{D}} \right) \right] \right)^D = e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \left(1 + \mathcal{O} \left(\frac{1}{D} \right) \right). \quad (226)$$

Substituting $\textcircled{1}$, $\textcircled{2}$, and $\textcircled{3}$ back into the expression for T_2 , we obtain:

$$T_2 = \sin(4\beta) \left[\frac{2\alpha}{\sqrt{D}} \mathbb{E}_{J \sim f} [J^2] + \mathcal{O} \left(\frac{1}{D^{3/2}} \right) \right] e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \left(1 + \mathcal{O} \left(\frac{1}{D} \right) \right) \quad (227)$$

$$= \sin(4\beta) \mathbb{E}_{J \sim f} [J^2] \frac{2\alpha}{\sqrt{D}} e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} + \mathcal{O} \left(\frac{1}{D} \right). \quad (228)$$

Step 3: Computing the Term T_3

We now evaluate the expectation T_3 of the term $C_{uv}^{(2)}(\gamma, \beta; G)$ over the distributions of the edge weights $J \sim f^E$ and vertex weights $h \sim g^V$ as follows:

$$T_3 = \mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} [C_{uv}^{(2)}(\gamma, \beta; G)] \quad (229)$$

$$= \mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} \left\{ -\frac{J_{uv}}{2} \sin^2(2\beta) \prod_{w \in N_v \setminus u} \cos(2J_{wv}\gamma) \prod_{w \in N_u \setminus v} \cos(2J_{uw}\gamma) \right. \quad (230)$$

$$\times \left\{ \cos(2(h_u + h_v)\gamma) \prod_{f \in N_{uv}} \cos(2(J_{uf} + J_{vf})\gamma) - \cos(2(h_u - h_v)\gamma) \prod_{f \in N_{uv}} \cos(2(J_{uf} - J_{vf})\gamma) \right\} \\ = -\frac{1}{2} \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] \prod_{w \in N_v \setminus u} \mathbb{E}_{J \sim f} [\cos(2J\gamma)] \prod_{w \in N_u \setminus v} \mathbb{E}_{J \sim f} [\cos(2J\gamma)] \quad (231)$$

$$\times \left\{ \mathbb{E}_{h, h' \sim g} [\cos(2(h + h')\gamma)] \prod_{f \in N_{uv}} \mathbb{E}_{J, J' \sim f} [\cos(2(J + J')\gamma)] \right. \\ \left. - \mathbb{E}_{h, h' \sim g} [\cos(2(h - h')\gamma)] \prod_{f \in N_{uv}} \mathbb{E}_{J, J' \sim f} [\cos(2(J - J')\gamma)] \right\} \\ = -\frac{1}{2} \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] \left(\mathbb{E}_{J \sim f} [\cos(2J\gamma)] \right)^{|N_v \setminus u| + |N_u \setminus v|} \quad (232)$$

$$\times \left\{ \mathbb{E}_{h,h' \sim g} [\cos(2(h+h')\gamma)] \left(\mathbb{E}_{J,J' \sim f} [\cos(2(J+J')\gamma)] \right)^{|\mathcal{N}_{uv}|} \right. \\ \left. - \mathbb{E}_{h,h' \sim g} [\cos(2(h-h')\gamma)] \left(\mathbb{E}_{J,J' \sim f} [\cos(2(J-J')\gamma)] \right)^{|\mathcal{N}_{uv}|} \right\}.$$

Since G is $(D+1)$ -regular graph, we have:

$$|\mathcal{N}(u)| = |\mathcal{N}(v)| = D+1 \quad (233)$$

$$\mathcal{N}(u) = \mathcal{N}_{u \setminus v} \cup \mathcal{N}_{uv} \cup \{v\}, \quad (234)$$

$$\mathcal{N}(v) = \mathcal{N}_{v \setminus u} \cup \mathcal{N}_{uv} \cup \{u\}, \quad (235)$$

which implies:

$$D+1 = |\mathcal{N}(u)| = |\mathcal{N}_{u \setminus v}| + |\mathcal{N}_{uv}| + 1, \quad (236)$$

$$D+1 = |\mathcal{N}(v)| = |\mathcal{N}_{v \setminus u}| + |\mathcal{N}_{uv}| + 1. \quad (237)$$

Hence, we get:

$$|\mathcal{N}_{u \setminus v}| + |\mathcal{N}_{uv}| = |\mathcal{N}_{v \setminus u}| + |\mathcal{N}_{uv}| = D, \quad (238)$$

$$\Rightarrow |\mathcal{N}_{u \setminus v}| = |\mathcal{N}_{v \setminus u}|, |\mathcal{N}_{uv}| \leq D. \quad (239)$$

Setting $\gamma = \frac{\alpha}{\sqrt{D}}$ and using $|\mathcal{N}_{u \setminus v}| = |\mathcal{N}_{v \setminus u}|$, T_3 can be written as follows:

$$T_3 = -\frac{1}{2} \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] \left(\mathbb{E}_{J \sim f} \left[\cos \left(2J \frac{\alpha}{\sqrt{D}} \right) \right] \right)^{2|\mathcal{N}_{u \setminus v}|} \\ \times \left\{ \mathbb{E}_{h,h' \sim g} \left[\cos \left(2(h+h') \frac{\alpha}{\sqrt{D}} \right) \right] \left(\mathbb{E}_{J,J' \sim f} \left[\cos \left(2(J+J') \frac{\alpha}{\sqrt{D}} \right) \right] \right)^{|\mathcal{N}_{uv}|} \right. \\ \left. - \mathbb{E}_{h,h' \sim g} \left[\cos \left(2(h-h') \frac{\alpha}{\sqrt{D}} \right) \right] \left(\mathbb{E}_{J,J' \sim f} \left[\cos \left(2(J-J') \frac{\alpha}{\sqrt{D}} \right) \right] \right)^{|\mathcal{N}_{uv}|} \right\}. \quad (240)$$

For some $a, b \geq 0$ and $0 \leq \lambda, \mu \leq 1$, we let

$$|\mathcal{N}_{u \setminus v}| = aD^\lambda, \quad (241)$$

$$|\mathcal{N}_{uv}| = bD^\mu. \quad (242)$$

such that

$$|\mathcal{N}_{u \setminus v}| + |\mathcal{N}_{uv}| = D \Rightarrow aD^\lambda + bD^\mu = D. \quad (243)$$

Using the above, T_3 can be rewritten as follows:

$$T_3 = -\frac{1}{2} \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] \underbrace{\left(\mathbb{E}_{J \sim f} \left[\cos \left(2J \frac{\alpha}{\sqrt{D}} \right) \right] \right)^{2aD^\lambda}}_{(1)} \\ \times \left\{ \underbrace{\mathbb{E}_{h,h' \sim g} \left[\cos \left(2(h+h') \frac{\alpha}{\sqrt{D}} \right) \right]}_{(2)} \underbrace{\left(\mathbb{E}_{J,J' \sim f} \left[\cos \left(2(J+J') \frac{\alpha}{\sqrt{D}} \right) \right] \right)^{bD^\mu}}_{(3)} \right. \\ \left. - \underbrace{\mathbb{E}_{h,h' \sim g} \left[\cos \left(2(h-h') \frac{\alpha}{\sqrt{D}} \right) \right]}_{(4)} \underbrace{\left(\mathbb{E}_{J,J' \sim f} \left[\cos \left(2(J-J') \frac{\alpha}{\sqrt{D}} \right) \right] \right)^{bD^\mu}}_{(5)} \right\}. \quad (244)$$

We now compute the individual terms of T_3 below.

1. Computation of ①

- **Case 1:** $a = 0$ (i.e., $|\mathcal{N}_{u \setminus v}| = 0$)

$$\textcircled{1} = \left(\mathbb{E}_{J \sim f} \left[\cos \left(2J \frac{\alpha}{\sqrt{D}} \right) \right] \right)^0 = 1. \quad (245)$$

- **Case 2:** $a > 0$ (i.e., $|\mathcal{N}_{u \setminus v}| \neq 0$)

Using theorem 18 with substitutions $x \rightarrow J$, $p \rightarrow f$, $a \rightarrow 2\alpha$, and $c \rightarrow 2a$, we obtain:

$$\textcircled{1} = \begin{cases} e^{-4a\alpha^2 \mathbb{E}_{J \sim f} [J^2]} (1 + \mathcal{O}(D^{-1})) & \text{if } \lambda = 1, \\ 1 - 4a\alpha^2 \mathbb{E}_{J \sim f} [J^2] D^{\lambda-1} \mathcal{O}(D^{2\lambda-2}) & \text{if } 0 \leq \lambda < 1. \end{cases} \quad (246)$$

Hence, ① can be written as follows:

$$\textcircled{1} = \begin{cases} 1 & \text{if } a = 0, \\ e^{-4a\alpha^2 \mathbb{E}_{J \sim f} [J^2]} (1 + \mathcal{O}(D^{-1})) & \text{if } a > 0, \lambda = 1, \\ 1 - 4a\alpha^2 \mathbb{E}_{J \sim f} [J^2] D^{\lambda-1} \mathcal{O}(D^{2\lambda-2}) & \text{if } a > 0, 0 \leq \lambda < 1. \end{cases} \quad (247)$$

2. Computation of ② and ④

$$\textcircled{2,4} = \mathbb{E}_{h, h' \sim g} \left[\cos \left(2(h \pm h') \frac{\alpha}{\sqrt{D}} \right) \right] \quad (248)$$

$$= 1 - \frac{4\alpha^2}{D} \left[\mathbb{E}_{h \sim g} [h^2] \pm (\mathbb{E}_{h \sim g} [h])^2 \right] + \mathcal{O}\left(\frac{1}{D^2}\right) \quad (249)$$

$$= 1 + \mathcal{O}\left(\frac{1}{D}\right) \quad (250)$$

where the second line was obtained using lemma 16.

3. Computation of ③ and ⑤

- **Case 1:** $b = 0$ ($|\mathcal{N}_{uv}| = 0$)

$$\textcircled{3,5} = \left(\mathbb{E}_{J, J' \sim f} \left[\cos \left(2(J \pm J') \frac{\alpha}{\sqrt{D}} \right) \right] \right)^0 = 1. \quad (251)$$

- **Case 2:** $b > 0$ ($|\mathcal{N}_{uv}| > 0$)

Using lemma 19, and setting $x \rightarrow J$, $p \rightarrow f$, $a \rightarrow 2\alpha$, $c \rightarrow b$, and $\lambda \rightarrow \mu$, we get:

$$\textcircled{3,5} = \left(\mathbb{E}_{J, J' \sim f} \left[\cos \left(2(J \pm J') \frac{\alpha}{\sqrt{D}} \right) \right] \right)^{bD^\mu} \quad (252)$$

$$= \begin{cases} e^{-4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] \pm \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right)} (1 + \mathcal{O}(D^{-1})) & \text{if } \mu = 1, \\ 1 - 4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] \pm \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) D^{\mu-1} + \mathcal{O}(D^{2\mu-2}) & \text{if } 0 \leq \mu < 1. \end{cases} \quad (253)$$

Hence, $\textcircled{3,5}$ can be written as follows:

$$\textcircled{3,5} = \begin{cases} 1 & \text{if } b = 0, \\ e^{-4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] \pm \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right)} (1 + \mathcal{O}(D^{-1})) & \text{if } b > 0, \mu = 1, \\ 1 - 4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] \pm \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) D^{\mu-1} + \mathcal{O}(D^{2\mu-2}) & \text{if } b > 0, 0 \leq \mu < 1. \end{cases} \quad (254)$$

4. Computation of $\textcircled{6} = \textcircled{2}\textcircled{3} - \textcircled{4}\textcircled{5}$

- Case 1: $b = 0$ (i.e., $|\mathcal{N}_{uv}| = 0$)

$$\textcircled{6} = \left(1 - \frac{4\alpha^2}{D} \left[\mathbb{E}_{h \sim g} [h^2] + \left(\mathbb{E}_{h \sim g} [h] \right)^2 \right] + \mathcal{O}\left(\frac{1}{D^2}\right) \right) \cdot 1 \quad (255)$$

$$\begin{aligned} & - \left(1 - \frac{4\alpha^2}{D} \left[\mathbb{E}_{h \sim g} [h^2] - \left(\mathbb{E}_{h \sim g} [h] \right)^2 \right] + \mathcal{O}\left(\frac{1}{D^2}\right) \right) \cdot 1 \\ & = -\frac{8\alpha^2}{D} \left(\mathbb{E}_{h \sim g} [h] \right)^2 + \mathcal{O}\left(\frac{1}{D^2}\right). \end{aligned} \quad (256)$$

- Case 2: $b > 0$ (i.e., $|\mathcal{N}_{uv}| > 0$)

$$\textcircled{6} = \mathbb{E}_{h, h' \sim g} \left[\cos \left(2(h + h') \frac{\alpha}{\sqrt{D}} \right) \right] \left(\mathbb{E}_{J, J' \sim f} \left[\cos \left(2(J + J') \frac{\alpha}{\sqrt{D}} \right) \right] \right)^{bD^\mu} \quad (257)$$

$$- \mathbb{E}_{h, h' \sim g} \left[\cos \left(2(h - h') \frac{\alpha}{\sqrt{D}} \right) \right] \left(\mathbb{E}_{J, J' \sim f} \left[\cos \left(2(J - J') \frac{\alpha}{\sqrt{D}} \right) \right] \right)^{bD^\mu} \quad (258)$$

$$\begin{aligned} & = \left[1 + \mathcal{O}(D^{-1}) \right] \times \begin{cases} e^{-4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] + \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right)} (1 + \mathcal{O}(D^{-1})) & \text{if } \mu = 1 \\ 1 - 4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] + \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) D^{\mu-1} + \mathcal{O}(D^{2\mu-2}) & \text{if } 0 \leq \mu < 1 \end{cases} \\ & \quad (259) \end{aligned}$$

$$\begin{aligned} & - \left[1 + \mathcal{O}(D^{-1}) \right] \times \begin{cases} e^{-4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] - \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right)} (1 + \mathcal{O}(D^{-1})) & \text{if } \mu = 1 \\ 1 - 4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] - \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) D^{\mu-1} + \mathcal{O}(D^{2\mu-2}) & \text{if } 0 \leq \mu < 1. \end{cases} \\ & \quad (260) \end{aligned}$$

When $\mu = 1$, (6) can be further simplified as follows:

$$\textcircled{6} = e^{-4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] + \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right)} \left(1 + \mathcal{O}(D^{-1}) \right)^2 - e^{-4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] - \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right)} \left(1 + \mathcal{O}(D^{-1}) \right)^2 \quad (261)$$

$$= e^{-4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] + \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right)} - e^{-4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] - \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right)} + \mathcal{O}(D^{-1}) \quad (262)$$

$$= e^{-4\alpha^2 b \mathbb{E}_{J \sim f} [J^2]} \left[e^{-4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J] \right)^2} - e^{4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J] \right)^2} \right] + \mathcal{O}(D^{-1}) \quad (263)$$

$$= -2e^{-4\alpha^2 b \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) + \mathcal{O}(D^{-1}), \quad (264)$$

where in the last line, we have used the identity $\sinh x = \frac{1}{2} (e^x - e^{-x})$.

When $0 \leq \mu < 1$, (6) can be simplified as follows:

$$\textcircled{6} = \left[1 + \mathcal{O}(D^{-1}) \right] \left[1 - 4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] + \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) D^{\mu-1} + \mathcal{O}(D^{2\mu-2}) \right] \quad (265)$$

$$- \left[1 + \mathcal{O}(D^{-1}) \right] \left[1 - 4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] - \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) D^{\mu-1} + \mathcal{O}(D^{2\mu-2}) \right] \quad (266)$$

$$= 1 - 4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] + \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) D^{\mu-1} + \mathcal{O}(D^{-1}) + \mathcal{O}(D^{2\mu-2}) \quad (267)$$

$$- 1 - 4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J^2] - \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) D^{\mu-1} + \mathcal{O}(D^{-1}) + \mathcal{O}(D^{2\mu-2}) \quad (268)$$

$$= -8\alpha^2 b \left(\mathbb{E}_{J \sim f} [J] \right)^2 + \mathcal{O}(D^{\max(-1, 2\mu-2)}). \quad (269)$$

In summary, (6) can be expressed as follows:

$$\textcircled{6} = \begin{cases} -\frac{8\alpha^2}{D} \left(\mathbb{E}_{h \sim g} [h] \right)^2 + \mathcal{O}\left(\frac{1}{D^2}\right) & \text{if } b = 0, \\ -2e^{-4\alpha^2 b \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) + \mathcal{O}(D^{-1}) & \text{if } b > 0, \mu = 1, \\ -8\alpha^2 b \left(\mathbb{E}_{J \sim f} [J] \right)^2 + \mathcal{O}(D^{\max(-1, 2\mu-2)}) & \text{if } b > 0, 0 \leq \mu < 1. \end{cases} \quad (270)$$

Having computed all the terms in T_3 , we now begin to evaluate the possible expressions for T_3 for

different possible values of a , b , λ , and μ .

$$T_3 = -\frac{1}{2} \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] \times \begin{cases} 1 & \text{if } a = 0, \\ e^{-4a\alpha^2 \mathbb{E}_{J \sim f} [J^2]} (1 + \mathcal{O}(D^{-1})) & \text{if } a > 0, \lambda = 1, \\ 1 - 4a\alpha^2 \mathbb{E}_{J \sim f} [J^2] D^{\lambda-1} \mathcal{O}(D^{2\lambda-2}) & \text{if } a > 0, 0 \leq \lambda < 1, \end{cases} \quad (271)$$

$$\times \begin{cases} -\frac{8\alpha^2}{D} \left(\mathbb{E}_{h \sim g} [h] \right)^2 + \mathcal{O}\left(\frac{1}{D^2}\right) & \text{if } b = 0, \\ -2e^{-4\alpha^2 b \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4\alpha^2 b \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) + \mathcal{O}(D^{-1}) & \text{if } b > 0, \mu = 1, \\ -8\alpha^2 b \left(\mathbb{E}_{J \sim f} [J] \right)^2 + \mathcal{O}(D^{\max(-1, 2\mu-2)}) & \text{if } b > 0, 0 \leq \mu < 1. \end{cases}$$

Recall that $aD^\lambda + bD^\mu = D$. We can divide into four cases, the possible values of a , b , λ , and μ .

- (C1) $a = 0 \Rightarrow D = bD^\mu \Rightarrow b = \mu = 1$
(C2) $b = 0 \Rightarrow D = aD^\lambda \Rightarrow a = \lambda = 1$
(C3) $\lambda < 1 \Rightarrow \mu = 1$ (otherwise $aD^\lambda + bD^\mu < D$ for some D)
(C4) $\mu = 1 \Rightarrow \lambda = 1$ (otherwise $aD^\lambda + bD^\mu < D$ for some D)

The above gives rise to the following cases:

$a \backslash b$	$a = 0$	$a > 0, \lambda = 1$	$a > 0, 0 \leq \lambda < 1$
$b = 0$	Contradicts (C1) $\because a = 0 \Rightarrow b = 1$	Case 1 $b = 0, a = \lambda = 1$	Contradicts (C2) $\because b = 0 \Rightarrow \lambda = 1$
$b > 0, \mu = 1$	Case 2 $a = 0, b = \mu = 1$	Case 3 $a, b > 0, a + b = 1, \lambda = \mu = 1$	Case 4 $a, b > 0, \mu = 1, 0 \leq \lambda < 1$
$b > 0, 0 \leq \mu < 1$	Contradicts (C1) $\because a = 0 \Rightarrow \mu = 1$	Case 5 $a, b > 0, \lambda = 1, 0 \leq \mu < 1$	Contradicts (C3) $\because \lambda < 1 \Rightarrow \mu = 1$

As indicated above, four of the above cases result in contradictions. We will consider the remaining 5 cases separately.

- **Case 1:** $b = 0, a = \lambda = 1$ ($|\mathcal{N}_{u \setminus v}| = D, |\mathcal{N}_{uv}| = 0$, i.e. no triangles)

$$T_3 = -\frac{1}{2} \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4\alpha^2 \mathbb{E}_{J \sim f} [J^2]} (1 + \mathcal{O}(D^{-1})) \left[-\frac{8\alpha^2}{D} \left(\mathbb{E}_{h \sim g} [h] \right)^2 + \mathcal{O}(D^{-2}) \right] \quad (272)$$

$$= 4\alpha^2 \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \left(\mathbb{E}_{h \sim g} [h] \right)^2 \frac{1}{D} + \mathcal{O}(D^{-2}). \quad (273)$$

- **Case 2:** $a = 0, b = \mu = 1$ ($|\mathcal{N}_{u \setminus v}| = 0, |\mathcal{N}_{uv}| = D$, i.e. Complete Graph)

$$T_3 = -\frac{1}{2} \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] \left[-2e^{-4\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) + \mathcal{O}(D^{-1}) \right] \quad (274)$$

$$= \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) + \mathcal{O}(D^{-1}). \quad (275)$$

- **Case 3:** $a, b > 0, \lambda = \mu = 1, a + b = 1$

$$T_3 = -\frac{1}{2} \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4a\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \left(1 + \mathcal{O}(D^{-1})\right) \quad (276)$$

$$\times \left[-2e^{-4b\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) + \mathcal{O}(D^{-1}) \right] \\ = \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4a\alpha^2 \mathbb{E}_{J \sim f} [J^2]} e^{-4b\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) + \mathcal{O}(D^{-1}) \quad (277)$$

$$= \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4(a+b)\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) + \mathcal{O}(D^{-1}) \quad (278)$$

$$= \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) + \mathcal{O}(D^{-1}). \quad (279)$$

- **Case 4:** $a, b > 0, \mu = 1, 0 \leq \lambda < 1$

$$T_3 = -\frac{1}{2} \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] \left(1 - 4a\alpha^2 \mathbb{E}_{J \sim f} [J^2] D^{\lambda-1} + \mathcal{O}(D^{2\lambda-2}) \right) \quad (280)$$

$$\times \left[-2e^{-4b\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) + \mathcal{O}(D^{-1}) \right] \\ = \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4b\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) \quad (281)$$

$$- 4a\alpha^2 \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] \mathbb{E}_{J \sim f} [J^2] e^{-4b\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) D^{\lambda-1} \\ + \mathcal{O}(D^{\max(2\lambda-2, -1)}).$$

- **Case 5:** $a, b > 0, \lambda = 1, 0 \leq \mu < 1$

$$T_3 = -\frac{1}{2} \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4a\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \left(1 + \mathcal{O}(D^{-1}) \right) \quad (282)$$

$$\times \left[-8b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 D^{\mu-1} + \mathcal{O}(D^{\max(-1, 2\mu-2)}) \right] \\ = 4b\alpha^2 \sin^2(2\beta) \left(\mathbb{E}_{J \sim f} [J] \right)^3 e^{-4a\alpha^2 \mathbb{E}_{J \sim f} [J^2]} D^{\mu-1} + \mathcal{O}(D^{\max(-1, 2\mu-2)}). \quad (283)$$

Step 4: Computing the Sum of Terms T_2 and T_3

Recall that T_2 is computed as follows:

$$T_2 = \sin(4\beta) \mathbb{E}_{J \sim f} [J^2] \frac{2\alpha}{\sqrt{D}} e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} + \mathcal{O}(D^{-1}). \quad (284)$$

- **Case 1:** $b = 0, a = \lambda = 1$

$$T_2 + T_3 = \sin(4\beta) \mathbb{E}_{J \sim f} [J^2] \frac{2\alpha}{\sqrt{D}} e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} + \mathcal{O}(D^{-1}) \quad (285)$$

$$+ 4\alpha^2 \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \left(\mathbb{E}_{h \sim g} [h] \right)^2 \frac{1}{D} + \mathcal{O}(D^{-2}) \\ = \sin(4\beta) \mathbb{E}_{J \sim f} [J^2] \frac{2\alpha}{\sqrt{D}} e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} + \mathcal{O}(D^{-1}). \quad (286)$$

- Case 2: $a = 0, \quad b = \mu = 1$

$$\begin{aligned} T_2 + T_3 &= \sin(4\beta) \mathbb{E}_{J \sim f} [J^2] \frac{2\alpha}{\sqrt{D}} e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \\ &\quad + \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) + \mathcal{O} \left(D^{-1} \right). \end{aligned} \quad (287)$$

- Case 3: $a, b > 0, \quad \lambda = \mu = 1$, i.e. $a + b = 1$

$$\begin{aligned} T_2 + T_3 &= \sin(4\beta) \mathbb{E}_{J \sim f} [J^2] \frac{2\alpha}{\sqrt{D}} e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \\ &\quad + \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) + \mathcal{O} \left(D^{-1} \right). \end{aligned} \quad (288)$$

- Case 4: $a, b > 0, \quad \mu = 1, \quad 0 \leq \lambda < 1$

$$\begin{aligned} T_2 + T_3 &= \sin(4\beta) \mathbb{E}_{J \sim f} [J^2] \frac{2\alpha}{\sqrt{D}} e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} + \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4b\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) \\ &\quad - 4a\alpha^2 \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] \mathbb{E}_{J \sim f} [J^2] e^{-4b\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) D^{\lambda-1} \\ &\quad + \mathcal{O} \left(D^{\max(2\lambda-2, -1)} \right). \end{aligned} \quad (289)$$

Case 4a: $\lambda < \frac{1}{2}$

Note that $\lambda < \frac{1}{2} \Leftrightarrow -\frac{1}{2} > \lambda - 1$. Hence, $T_2 + T_3$ can be further simplified as follows:

$$\begin{aligned} T_2 + T_3 &= \sin(4\beta) \mathbb{E}_{J \sim f} [J^2] \frac{2\alpha}{\sqrt{D}} e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \\ &\quad + \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4b\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) + \mathcal{O} \left(D^{\lambda-1} \right). \end{aligned} \quad (290)$$

Case 4b: $\lambda = \frac{1}{2}$

Note that $\lambda = \frac{1}{2} \Leftrightarrow -\frac{1}{2} = \lambda - 1$. Hence, $T_2 + T_3$ can be further simplified as follows:

$$\begin{aligned} T_2 + T_3 &= \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4b\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) \\ &\quad + \mathbb{E}_{J \sim f} [J^2] \left\{ 2\alpha \sin(4\beta) e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \right. \\ &\quad \left. - 4a\alpha^2 \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4b\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) \right\} \frac{1}{\sqrt{D}} \\ &\quad + \mathcal{O} \left(D^{\max(2\lambda-2, -1)} \right). \end{aligned} \quad (291)$$

Case 4c: $\lambda > \frac{1}{2}$

Note that $\lambda > \frac{1}{2} \Leftrightarrow -\frac{1}{2} < \lambda - 1$. Hence, $T_2 + T_3$ can be further simplified as follows:

$$\begin{aligned} T_2 + T_3 &= \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] e^{-4b\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) \\ &\quad - 4a\alpha^2 \sin^2(2\beta) \mathbb{E}_{J \sim f} [J] \mathbb{E}_{J \sim f} [J^2] e^{-4b\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \sinh \left(4b\alpha^2 \left(\mathbb{E}_{J \sim f} [J] \right)^2 \right) D^{\lambda-1} \\ &\quad + \mathcal{O} \left(\frac{1}{\sqrt{D}} \right). \end{aligned} \quad (292)$$

- **Case 5a:** $a, b > 0, \quad \lambda = 1, \quad 0 \leq \mu < 1$

$$\begin{aligned} T_2 + T_3 &= \sin(4\beta) \mathbb{E}_{J \sim f} [J^2] \frac{2\alpha}{\sqrt{D}} e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} + \mathcal{O} \left(D^{-1} \right) \\ &\quad + 4b\alpha^2 \sin^2(2\beta) \left(\mathbb{E}_{J \sim f} [J] \right)^3 e^{-4a\alpha^2 \mathbb{E}_{J \sim f} [J^2]} D^{\mu-1} + \mathcal{O} \left(D^{\max(-1, 2\mu-2)} \right). \end{aligned} \quad (293)$$

Case 5a: $\mu > \frac{1}{2}$

Note that $\mu > \frac{1}{2} \Rightarrow \mu - 1 > -\frac{1}{2}$. Now, $-\frac{1}{2} > 2\mu - 2 \Leftrightarrow \mu < \frac{3}{4}$. Hence, $T_2 + T_3$ can be further simplified as follows:

$$T_2 + T_3 = 4b\alpha^2 \sin^2(2\beta) \left(\mathbb{E}_{J \sim f} [J] \right)^3 e^{-4a\alpha^2 \mathbb{E}_{J \sim f} [J^2]} D^{\mu-1} + \mathcal{O} \left(D^{\max(-\frac{1}{2}, 2\mu-2)} \right). \quad (294)$$

Case 5b: $\mu = \frac{1}{2}$

$$T_2 + T_3 = \left(2\alpha \sin(4\beta) \mathbb{E}_{J \sim f} [J^2] e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} + 4b\alpha^2 \sin^2(2\beta) \left(\mathbb{E}_{J \sim f} [J] \right)^3 e^{-4a\alpha^2 \mathbb{E}_{J \sim f} [J^2]} \right) \frac{1}{\sqrt{D}} + \mathcal{O} \left(D^{-1} \right). \quad (295)$$

Case 5c: $\mu < \frac{1}{2}$

Note that $\mu < \frac{1}{2} \Rightarrow -1 \leq \mu - 1 < -\frac{1}{2}$. Hence, $T_2 + T_3$ can be further simplified as follows:

$$T_2 + T_3 = \sin(4\beta) \mathbb{E}_{J \sim f} [J^2] \frac{2\alpha}{\sqrt{D}} e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} + \mathcal{O} \left(D^{\mu-1} \right). \quad (296)$$

Step 5: Computing the Scaled Expected QAOA₁ Cost Function

Recall that the expected QAOA₁ cost function for an Ising model on a graph $G = (V, E)$ with parameters $\beta, \gamma \in \mathbb{R}$ is given by:

$$\mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} [\xi_G(\gamma, \beta)] = \sum_{i \in V} \mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} [C_i(\gamma, \beta; G)] + \sum_{\{u, v\} \in E} \mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} \left\{ C_{uv}^{(1)}(\gamma, \beta; G) + \sum_{\{u, v\} \in E} C_{uv}^{(2)}(\gamma, \beta; G) \right\} \quad (297)$$

$$= \sum_{i \in V} T_1 + \sum_{\{u, v\} \in E} (T_2 + T_3). \quad (298)$$

For a $(D+1)$ -regular graph, the expectation terms T_1, T_2 , and T_3 are identical for all vertices i and edges $\{u, v\}$. This allows us to simplify the expression as follows:

$$\mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} [\xi_G(\gamma, \beta)] = T_1 |V| + (T_2 + T_3) |E|, \quad (299)$$

where $n = |V|$ is the number of vertices and $|E|$ is the number of edges in the graph G . Since the graph is $(D + 1)$ -regular, the number of edges satisfies $|E| = \frac{n}{2}(D + 1)$. Substituting this into the equation, we obtain:

$$\mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} [\xi_G(\gamma, \beta)] = nT_1 + \frac{n}{2}(D + 1)(T_2 + T_3). \quad (300)$$

To facilitate analysis, we introduce a scaled version of the cost function by dividing the expected cost by the number of edges $|E|$:

$$\frac{1}{|E|} \mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} [\xi_G(\gamma, \beta)] = \frac{|V|}{|E|} T_1 + (T_2 + T_3). \quad (301)$$

Calculating the ratio $|V|/|E|$ for a $(D + 1)$ -regular graph, we have:

$$\frac{|V|}{|E|} = \frac{n}{\frac{n}{2}(D + 1)} = \frac{2}{D + 1}. \quad (302)$$

Next, we evaluate the term $(|V|/|E|)T_1$ to get the following:

$$\frac{|V|}{|E|} T_1 = \frac{2}{D + 1} \left[\sin(4\beta) \mathbb{E}_{h \sim g} [h^2] \frac{2\alpha}{\sqrt{D}} e^{-2\alpha^2 \mathbb{E}_{J \sim f} [J^2]} + \mathcal{O}\left(\frac{1}{D}\right) \right] \quad (303)$$

$$= \mathcal{O}\left(D^{-\frac{3}{2}}\right). \quad (304)$$

Finally, substituting this result into the scaled cost function yields:

$$\frac{1}{|E|} \mathbb{E}_{\substack{J \sim f^E \\ h \sim g^V}} [\xi_G(\gamma, \beta)] = T_2 + T_3 + \mathcal{O}\left(D^{-\frac{3}{2}}\right). \quad (305)$$

This scaling demonstrates that as the degree D of the graph increases, the contribution from T_1 becomes negligible, and the expected scaled cost function is primarily determined by the terms T_2 and T_3 , with corrections of order $D^{-\frac{3}{2}}$. Therefore, the expressions derived in step 4 accurately characterise the scaled expected QAOA₁ cost function.

Step 6: Optimal γ^* Values for Leading-Order Terms

Before proceeding, we note that the subscripts for the expectation value are omitted, as the expressions in the remainder of this proof involve only the terms where $J \sim f$.

Not all expressions derived for the different cases yield closed-form solutions for the optimal α^* value. To address this, we focus on the leading-order terms, which simplify the expressions and allow for closed-form solutions for α^* . The leading-order expressions for the scaled QAOA₁ objective are given by:

$$\frac{1}{|E|} \mathbb{E} [\xi_G(\alpha, \beta)] = \begin{cases} \mathcal{C}_1(\alpha, \beta) + \mathcal{O}(D^{-1}), & b = 0, a = \lambda = 1, \\ \mathcal{C}_2(\alpha, \beta, 1, 1) + \mathcal{O}(1), & a = 0, b = \mu = 1, \\ \mathcal{C}_2(\alpha, \beta, 1, b) + \mathcal{O}(1), & a, b > 0, \lambda = \mu = 1, \\ \mathcal{C}_2(\alpha, \beta, b, b) + \mathcal{O}(1), & a, b > 0, \mu = 1, \lambda < 1, \\ \mathcal{C}_3(\alpha, \beta) + \mathcal{O}(D^{\mu-1}), & a, b > 0, \lambda = 1, \mu > \frac{1}{2}, \\ \mathcal{C}_1(\alpha, \beta) + \mathcal{O}(D^{\mu-1}), & a, b > 0, \lambda = 1, \mu < \frac{1}{2}, \end{cases} \quad (306)$$

where the functions $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$ are defined as:

$$\mathcal{C}_1(\alpha, \beta) = 2\alpha \sin(4\beta) \mathbb{E} [J^2] e^{-2\alpha^2 \mathbb{E} [J^2]} \frac{1}{\sqrt{D}}, \quad (307)$$

$$\mathcal{C}_2(\alpha, \beta, \theta_1, \theta_2) = \sin^2(2\beta) \mathbb{E} [J] e^{-4\theta_1 \alpha^2 \mathbb{E} [J^2]} \sinh\left(4\theta_2 \alpha^2 \mathbb{E} [J]^2\right), \quad (308)$$

$$\mathcal{C}_3(\alpha, \beta) = 4b\alpha^2 \sin^2(2\beta) \mathbb{E} [J]^3 e^{-4a\alpha^2 \mathbb{E} [J^2]} D^{\mu-1}. \quad (309)$$

For the case $(a, b > 0, \lambda = 1, \mu = \frac{1}{2})$, the leading-order terms coincide with the original expression, but they do not admit a closed-form solution for α^* . In contrast, for the cases $(b = 0, a = \lambda = 1)$, $(a, b > 0, \lambda = 1, \mu > \frac{1}{2})$, and $(a, b > 0, \lambda = 1, \mu < \frac{1}{2})$, the leading-order terms also coincide with the original expressions, but they do allow for closed-form solutions for α^* .

We now proceed to compute the optimal closed-form values of the three functions $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$.

Optimising $\mathcal{C}_1$

To determine the optimal α that minimises $\mathcal{C}_1$, we fix $\beta = 3\pi/8$, simplifying $\mathcal{C}_1$ to:

$$\mathcal{C}_1(\alpha) = -2\alpha \mathbb{E}[J^2] e^{-2\alpha^2 \mathbb{E}[J^2]} \frac{1}{\sqrt{D}}. \quad (310)$$

The optimal α is found by differentiating $\mathcal{C}_1$ with respect to α and setting the derivative to zero:

$$\frac{d\mathcal{C}_1(\alpha)}{d\alpha} = -2\mathbb{E}[J^2] \left(e^{-2\alpha^2 \mathbb{E}[J^2]} - 4\alpha^2 \mathbb{E}[J^2] e^{-2\alpha^2 \mathbb{E}[J^2]} \right) \frac{1}{\sqrt{D}} \quad (311)$$

$$= -2\mathbb{E}[J^2] \left(1 - 4\alpha^2 \mathbb{E}[J^2] \right) e^{-2\alpha^2 \mathbb{E}[J^2]} \frac{1}{\sqrt{D}}. \quad (312)$$

Setting the derivative to zero yields:

$$1 - 4\alpha^2 \mathbb{E}[J^2] = 0 \quad \Rightarrow \quad \alpha^2 = \frac{1}{4\mathbb{E}[J^2]} \quad \Rightarrow \quad \alpha = \pm \frac{1}{2\sqrt{\mathbb{E}[J^2]}}. \quad (313)$$

Since $\alpha \geq 0$, the optimal value is:

$$\alpha^* = \frac{1}{2\sqrt{\mathbb{E}[J^2]}}. \quad (314)$$

Optimising $\mathcal{C}_2$

For $\mathcal{C}_2$, we set $\beta = \pi/4$, reducing the expression to:

$$\mathcal{C}_2(\alpha, \theta_1, \theta_2) = \mathbb{E}[J] e^{-4\theta_1 \alpha^2 \mathbb{E}[J^2]} \sinh \left(4\theta_2 \alpha^2 \mathbb{E}[J^2] \right). \quad (315)$$

To facilitate differentiation, we take the natural logarithm:

$$\ln [\mathcal{C}_2(\alpha, \theta_1, \theta_2)] = \ln [\mathbb{E}[J]] - 4\theta_1 \alpha^2 \mathbb{E}[J^2] + \ln \left[\sinh \left(4\theta_2 \alpha^2 \mathbb{E}[J^2] \right) \right]. \quad (316)$$

Differentiating with respect to α gives:

$$\frac{d \ln [\mathcal{C}_2(\alpha, \theta_1, \theta_2)]}{d\alpha} = -8\theta_1 \alpha \mathbb{E}[J^2] + 8\theta_2 \alpha \mathbb{E}[J^2]^2 \coth \left(4\theta_2 \alpha^2 \mathbb{E}[J^2] \right). \quad (317)$$

Setting this derivative to zero results in:

$$-8\theta_1 \alpha \mathbb{E}[J^2] + 8\theta_2 \alpha \mathbb{E}[J^2]^2 \coth \left(4\theta_2 \alpha^2 \mathbb{E}[J^2] \right) = 0. \quad (318)$$

It is evident that $\alpha = 0$ is a critical point. To identify other critical points, we divide through by 8α ,

assuming $\alpha \neq 0$, and solve for α^2 :

$$-\theta_1 \mathbb{E}[J^2] + \theta_2 \mathbb{E}[J]^2 \coth(4\theta_2 \alpha^2 \mathbb{E}[J]^2) = 0 \quad (319)$$

$$\coth(4\theta_2 \alpha^2 \mathbb{E}[J]^2) - \frac{\theta_1 \mathbb{E}[J^2]}{\theta_2 \mathbb{E}[J]^2} = 0 \quad (320)$$

$$4\theta_2 \alpha^2 \mathbb{E}[J]^2 = \coth^{-1}\left(\frac{\theta_1 \mathbb{E}[J^2]}{\theta_2 \mathbb{E}[J]^2}\right) \quad (321)$$

$$4\theta_2 \alpha^2 \mathbb{E}[J]^2 = \frac{1}{2} \ln \left(\frac{\frac{\theta_1 \mathbb{E}[J^2]}{\theta_2 \mathbb{E}[J]^2} + 1}{\frac{\theta_1 \mathbb{E}[J^2]}{\theta_2 \mathbb{E}[J]^2} - 1} \right) \quad (322)$$

$$4\theta_2 \alpha^2 \mathbb{E}[J]^2 = \frac{1}{2} \ln \left(\frac{\theta_1 \mathbb{E}[J^2] + \theta_2 \mathbb{E}[J]^2}{\theta_1 \mathbb{E}[J^2] - \theta_2 \mathbb{E}[J]^2} \right) \quad (323)$$

$$\alpha^2 = \frac{1}{8\theta_2 \mathbb{E}[J]^2} \ln \left(\frac{\theta_1 \mathbb{E}[J^2] + \theta_2 \mathbb{E}[J]^2}{\theta_1 \mathbb{E}[J^2] - \theta_2 \mathbb{E}[J]^2} \right). \quad (324)$$

As $\alpha \geq 0$, the optimal value for $\alpha > 0$ is:

$$\alpha^* = \sqrt{\frac{1}{8\theta_2 \mathbb{E}[J]^2} \ln \left(\frac{\theta_1 \mathbb{E}[J^2] + \theta_2 \mathbb{E}[J]^2}{\theta_1 \mathbb{E}[J^2] - \theta_2 \mathbb{E}[J]^2} \right)}. \quad (325)$$

Optimising $\mathcal{C}_3$

For $\mathcal{C}_3$, we set $\beta = \pi/4$, reducing the expression to:

$$\mathcal{C}_3(\alpha) = 4b\alpha^2 \mathbb{E}[J]^3 e^{-4a\alpha^2 \mathbb{E}[J^2]} D^{\mu-1}. \quad (326)$$

Differentiating with respect to α gives:

$$\frac{d\mathcal{C}_3(\alpha)}{d\alpha} = 4b\mathbb{E}[J]^3 \left[2\alpha - 4a\alpha^3 \mathbb{E}[J^2] \right] e^{-4a\alpha^2 \mathbb{E}[J^2]} D^{\mu-1}. \quad (327)$$

Setting the derivative to zero yields:

$$2\alpha - 4a\alpha^3 \mathbb{E}[J^2] = 0. \quad (328)$$

It is evident that $\alpha = 0$ is a critical point. To identify other critical points, we divide through by 2α , assuming $\alpha \neq 0$, and solve for α :

$$1 - 2a\alpha^2 \mathbb{E}[J^2] = 0 \quad (329)$$

$$\alpha = \pm \frac{1}{\sqrt{2a\mathbb{E}[J^2]}}. \quad (330)$$

As $\alpha \geq 0$, the optimal value for $\alpha > 0$ is:

$$\alpha = \frac{1}{\sqrt{2a\mathbb{E}[J^2]}}. \quad (331)$$

Optimal α^* Values

The closed-form expressions for the optimal $\alpha^* \in \mathbb{R}^+$ values for the different cases are given by:

$$\alpha^* = \begin{cases} \frac{1}{2\sqrt{\mathbb{E}[J^2]}}, & b = 0, a = \lambda = 1, \\ \sqrt{\frac{1}{8\mathbb{E}[J]^2} \ln \left(\frac{\mathbb{E}[J^2] + \mathbb{E}[J]^2}{\mathbb{E}[J^2] - \mathbb{E}[J]^2} \right)}, & a = 0, b = \mu = 1, \\ \sqrt{\frac{1}{8b\mathbb{E}[J]^2} \ln \left(\frac{\mathbb{E}[J^2] + b\mathbb{E}[J]^2}{\mathbb{E}[J^2] - b\mathbb{E}[J]^2} \right)}, & a, b > 0, \lambda = \mu = 1, \\ \sqrt{\frac{1}{8b\mathbb{E}[J]^2} \ln \left(\frac{\mathbb{E}[J^2] + \mathbb{E}[J]^2}{\mathbb{E}[J^2] - \mathbb{E}[J]^2} \right)}, & a, b > 0, \mu = 1, \lambda < 1, \\ \frac{1}{\sqrt{2a\mathbb{E}[J^2]}}, & a, b > 0, \lambda = 1, \mu > \frac{1}{2}, \\ \frac{1}{2\sqrt{\mathbb{E}[J^2]}}, & a, b > 0, \lambda = 1, \mu < \frac{1}{2}. \end{cases} \quad (332)$$

Using the relation $\gamma = \alpha/\sqrt{D}$, the closed-form expressions for the optimal $\gamma^* \in \mathbb{R}^+$ values can be readily obtained.

Step 7: Globally Optimal α^* Coincides with First Local Optimum

In this final part of the proof, we demonstrate that the derived globally optimal value α^* coincides with the first local optimum of the functions $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$.

Demonstrating Global Optimum Coincides with First Local Optimum for $\mathcal{C}_1$

We want to show that the globally optimal parameter eq. (314) for $\mathcal{C}_1$ coincides with the first local optimum. Recall that the derivative of $\mathcal{C}_1$ is given by:

$$\frac{d\mathcal{C}_1(\alpha)}{d\alpha} = -2\mathbb{E}[J^2] \left(1 - 4\alpha^2\mathbb{E}[J^2]\right) e^{-2\alpha^2\mathbb{E}[J^2]} \frac{1}{\sqrt{D}}. \quad (333)$$

We begin by analysing the behaviour of $\mathcal{C}'_1(\alpha)$ at $\alpha = 0$. Substituting $\alpha = 0$ into the derivative, we find that

$$\mathcal{C}'_1(0) = -2\mathbb{E}[J^2] \frac{1}{\sqrt{D}} < 0, \quad (334)$$

which implies that $\mathcal{C}_1(\alpha)$ is decreasing at $\alpha = 0$. Next, we consider the interval $\alpha \in (0, \alpha^*)$. In this range, the term $1 - 4\alpha^2\mathbb{E}_J[J^2]$ remains positive because $\alpha^2 < (\alpha^*)^2$. Consequently, the derivative satisfies:

$$-2\mathbb{E}[J^2] \left(1 - 4\alpha^2\mathbb{E}[J^2]\right) e^{-2\alpha^2\mathbb{E}[J^2]} \frac{1}{\sqrt{D}} < 0. \quad (335)$$

This implies that $\mathcal{C}_1(\alpha)$ continues to decrease throughout the interval $(0, \alpha^*)$. At $\alpha = \alpha^*$, the term $1 - 4\alpha^2\mathbb{E}_J[J^2]$ becomes zero and so $\mathcal{C}'_1(\alpha^*)$ evaluates to 0 thus confirming that α^* is a stationary point.

Finally, we examine the behaviour of $\mathcal{C}'_1(\alpha)$ for $\alpha > \alpha^*$. In this region, the term $1 - 4\alpha^2\mathbb{E}_J[J^2]$ becomes negative because $\alpha^2 > (\alpha^*)^2$. As a result, the derivative satisfies:

$$-2\mathbb{E}[J^2] \left(1 - 4\alpha^2\mathbb{E}[J^2]\right) e^{-2\alpha^2\mathbb{E}[J^2]} \frac{1}{\sqrt{D}} > 0. \quad (336)$$

This indicates that $\mathcal{C}_1(\alpha)$ is increasing for $\alpha > \alpha^*$.

From this analysis, we observe that $\mathcal{C}_1(\alpha)$ decreases for $\alpha \in [0, \alpha^*)$, reaches a stationary point at $\alpha = \alpha^*$, and increases for $\alpha > \alpha^*$. This behaviour establishes that α^* is the first local minimum of $\mathcal{C}_1(\alpha)$ for $\alpha \geq 0$.

Demonstrating Global Optimum Coincides with First Local Optimum for $\mathcal{C}_2$

We want to show that the globally optimal parameter eq. (325) for $\mathcal{C}_2$ coincides with the first local optimum. Recall that the derivative of $\mathcal{C}_2$ with respect to α is:

$$\mathcal{C}'_2(\alpha, \theta_1, \theta_2) = 8\alpha\mathbb{E}[J]e^{-4\theta_1\alpha^2\mathbb{E}[J^2]} \left[\theta_2\mathbb{E}[J]^2 \cosh\left(4\theta_2\alpha^2\mathbb{E}[J]^2\right) - \theta_1\mathbb{E}[J^2] \sinh\left(4\theta_2\alpha^2\mathbb{E}[J]^2\right) \right], \quad (337)$$

where (θ_1, θ_2) can be $(1, 1)$, $(1, b)$, or (b, b) with $0 \leq b \leq 1$. Additionally, since the variance is non-negative, we have $\mathbb{E}[J^2] \geq \mathbb{E}[J]^2$.

The case $\mathbb{E}[J] > 0$ is trivial because the global optimum that minimises $\mathcal{C}_2$ is $\alpha^* = 0$ which clearly is the first local optimum. Thus, we focus on $\mathbb{E}[J] < 0$. Consider $\alpha \in (0, \alpha^*)$. Since $\alpha^2 < (\alpha^*)^2$, we can approximate the hyperbolic functions for small arguments:

$$\cosh(x) \approx 1 + \frac{x^2}{2}, \quad \sinh(x) \approx x + \frac{x^3}{6}. \quad (338)$$

Applying these approximations, the derivative becomes:

$$\mathcal{C}'_2 \approx 8\alpha\mathbb{E}[J]e^{-4\theta_1\alpha^2\mathbb{E}[J^2]} \left[\theta_2\mathbb{E}[J]^2 \left(1 + \frac{(4\theta_2\alpha^2\mathbb{E}[J]^2)^2}{2} \right) - \theta_1\mathbb{E}[J^2] \left(4\theta_2\alpha^2\mathbb{E}[J]^2 + \frac{(4\theta_2\alpha^2\mathbb{E}[J]^2)^3}{6} \right) \right] \quad (339)$$

$$\approx 8\alpha\mathbb{E}[J]e^{-4\theta_1\alpha^2\mathbb{E}[J^2]} \left[\theta_2\mathbb{E}[J]^2 - 4\theta_1\theta_2\alpha^2\mathbb{E}[J^2]\mathbb{E}[J]^2 \right] \quad (340)$$

$$= 8\alpha\theta_2\mathbb{E}[J]^3e^{-4\theta_1\alpha^2\mathbb{E}[J^2]} \left[1 - 4\theta_2\alpha^2\mathbb{E}[J^2] \right]. \quad (341)$$

Since $\alpha^2 < (\alpha^*)^2$, the term $1 - 4\theta_2\alpha^2\mathbb{E}[J^2]$ remains positive. Given that $\mathbb{E}[J] < 0$ and $\theta_2 \geq 0$, the derivative $\mathcal{C}'_2 < 0$. This implies that $\mathcal{C}_2$ is decreasing on the interval $(0, \alpha^*)$.

For $\alpha > \alpha^*$, assuming α is still small enough for the above approximations to hold, the term $1 - 4\theta_2\alpha^2\mathbb{E}[J^2]$ becomes negative because $\alpha^2 > (\alpha^*)^2$. Consequently, $\mathcal{C}'_2 > 0$, indicating that $\mathcal{C}_2$ is increasing for $\alpha > \alpha^*$.

For sufficiently large $\alpha \gg \alpha^*$, we approximate the hyperbolic functions as:

$$\cosh(x) \approx \sinh(x) \approx \frac{e^x}{2}. \quad (342)$$

Substituting these into the derivative gives:

$$\mathcal{C}'_2 \approx 4\alpha\mathbb{E}[J]e^{-4\theta_1\alpha^2\mathbb{E}[J^2]}e^{4\theta_2\alpha^2\mathbb{E}[J]^2} \left(\theta_2\mathbb{E}[J]^2 - \theta_1\mathbb{E}[J^2] \right). \quad (343)$$

Given that $\theta_2 \leq \theta_1$ and $\mathbb{E}[J^2] \geq \mathbb{E}[J]^2$, we analyse the expression $\theta_2\mathbb{E}[J]^2 - \theta_1\mathbb{E}[J^2]$:

- If $\theta_1\mathbb{E}[J^2] > \theta_2\mathbb{E}[J]^2$, then $\theta_2\mathbb{E}[J]^2 - \theta_1\mathbb{E}[J^2] < 0$.
- If $\theta_1\mathbb{E}[J^2] = \theta_2\mathbb{E}[J]^2$, then $\theta_2\mathbb{E}[J]^2 - \theta_1\mathbb{E}[J^2] = 0$.

Thus, for $\alpha \gg \alpha^*$, the derivative $\mathcal{C}'_2 > 0$, confirming that $\mathcal{C}_2$ continues to increase.

The analysis shows that $\mathcal{C}_2(\alpha)$ is decreasing for $\alpha \in (0, \alpha^*)$, attains a stationary point at $\alpha = \alpha^*$, and increases for $\alpha > \alpha^*$. This behaviour confirms that α^* is the first local minimum of $\mathcal{C}_2(\alpha)$ for $\alpha \geq 0$. Therefore, the globally optimal parameter eq. (325) coincides with this first local optimum.

Demonstrating Global Optimum Coincides with First Local Optimum for $\mathcal{C}_3$

Since the proof that the globally optimal value α^* , defined in eq. (331), coincides with the first local optimum closely parallels the argument presented for $\mathcal{C}_1$, we omit it here for the sake of brevity. $\square$