

Parameter Setting in Quantum Approximate Optimization of Weighted Problems

Shree Hari Sureshabu^{‡1}, Dylan Herman^{‡1}, Ruslan Shaydulin^{‡1}, Joao Basso²,
Shouvanik Chakrabarti¹, Yue Sun¹, and Marco Pistoia¹

¹Global Technology Applied Research, JPMorgan Chase, New York, NY 10017

²Department of Mathematics, University of California, Berkeley, CA 94720

Quantum Approximate Optimization Algorithm (QAOA) is a leading candidate algorithm for solving combinatorial optimization problems on quantum computers. However, in many cases QAOA requires computationally intensive parameter optimization. The challenge of parameter optimization is particularly acute in the case of weighted problems, for which the eigenvalues of the phase operator are non-integer and the QAOA energy landscape is not periodic. In this work, we develop parameter setting heuristics for QAOA applied to a general class of weighted problems. First, we derive optimal parameters for QAOA with depth $p = 1$ applied to the weighted MaxCut problem under different assumptions on the weights. In particular, we rigorously prove the conventional wisdom that in the average case the first local optimum near zero gives globally-optimal QAOA parameters. Second, for $p \geq 1$ we prove that the QAOA energy landscape for weighted MaxCut approaches that for the unweighted case under a simple rescaling of parameters. Therefore, we can use parameters previously obtained for unweighted MaxCut for weighted problems. Finally, we prove that for $p = 1$ the QAOA objective sharply concentrates around its expectation, which means that our parameter setting rules hold with high probability for a random weighted instance. We numerically validate this approach on general weighted graphs and show that on average the QAOA energy with the proposed fixed parameters is only 1.1 percentage points away from that with optimized parameters. Third, we propose a general heuristic rescaling scheme inspired by the analytical results for weighted MaxCut and demonstrate its effectiveness using QAOA with the XY Hamming-weight-preserving mixer applied to the portfolio optimization problem. Our heuristic improves the convergence of local optimizers, reducing the number of iterations by 7.4x on average.

1 Introduction

Quantum computers are widely believed to be able to provide computational speedups for various problems of relevance to science and industry [1, 2]. Combinatorial optimization is a domain that is very likely to benefit from quantum computing due to the ubiquity of hard optimization problems. Quantum Approximate Optimization Algorithm (QAOA) [3, 4, 5]

[‡]Equal contribution.

is a leading candidate quantum heuristic algorithm for optimization. QAOA solves optimization problems by preparing a parameterized quantum state using a circuit consisting of layers of alternating operators, wherein each operator has a free parameter associated with it. The two operators are commonly referred to as the *phase operator* and *mixer operator*, respectively. QAOA has been shown to achieve better scaling than state-of-the-art classical solvers for finding exact solutions of k -SAT [6] and to achieve approximation ratios competitive with those of the best known classical algorithms for the unweighted MaxCut problem [7, 8]. When the mixer is different from the transverse field used in [3, 4], the algorithm is sometimes referred to as the Quantum Alternating Operator Ansatz [5]. Throughout the paper, we will use QAOA to refer to both without making a distinction.

One of the central challenges of applying QAOA to practically-relevant problems is the need to set the QAOA parameters. The parameter setting is particularly challenging for problems with objectives containing non-integer coefficients (weights) on the binary variables. The non-integer weights lead to the optimization landscape becoming non-periodic and in general hard to optimize [9, 10, 11]. While parameter setting schemes leveraging analytically-optimal QAOA parameters exist, they are only available in a limited number of cases. For example, in the infinite-size limit, optimal QAOA parameters are known for unweighted MaxCut on large-girth regular [7], Erdős-Rényi [12] and other [12] random graph ensembles, as well as for the Sherrington-Kirkpatrick (SK) model [13]. While a similar parameter setting scheme has been proposed and numerically validated for weighted MaxCut [9], no analytical results are known for weighted MaxCut or other problems with non-integer eigenvalues.

In this work, we develop parameter setting heuristics for QAOA applied to a broad class of weighted problems. Our starting point is QAOA with the transverse-field mixer applied to the weighted MaxCut problem on large-girth regular graphs. We begin with $p = 1$ and derive globally-optimal parameters for QAOA applied to graphs with edge weights drawn i.i.d. from the exponential distribution for any graph size and from an arbitrary distribution in the infinite-size limit. Our analysis rigorously proves the folklore notion that for problems with non-periodic QAOA energy landscapes, the first local optimum near zero contains globally-optimal parameters in the average case [9, 10, 11][‡]. We then analyze the case of $p \geq 1$ and connect QAOA on weighted MaxCut problems to QAOA on unweighted MaxCut by proving the following Theorem:

Theorem 3 (Informal). *Consider QAOA with depth p and a regular graph G with girth $> 2p + 1$ and i.i.d. random edge weights drawn from \mathbf{w} . Then the QAOA objective for weighted MaxCut on G at parameters $\left(\beta, \frac{\gamma}{\sqrt{\mathbb{E}_{\mathbf{w}}[w^2]}}\right)$ is, up to a global scaling factor, equal to the QAOA objective for the corresponding MaxCut problem on the unweighted version of the same graph at parameters (β, γ) on average in the infinite-size limit. Here the average is taken over the random choice of edge weights, and $\mathbb{E}_{\mathbf{w}}[w^2]$ is the second moment of the distribution from which the edge weights are drawn.*

This result proves that the parameters that are optimal for unweighted MaxCut can be rescaled to be optimal for weighted MaxCut. As a consequence, it establishes a rule for setting parameters in QAOA for weighted MaxCut using the parameters for the unweighted case obtained previously in Ref. [7]. We further prove that for $p = 1$ the QAOA objective concentrates sharply around its expectation over randomly drawn weights. This implies that the derived correspondence between optimal parameters for weighted and unweighted

[‡]We note the importance of the “average case” qualifier, as examples can be constructed where the optimal QAOA parameters are arbitrarily far from origin; see e.g. [9, Figure 2].

cases and the resulting parameter setting rule hold with a high probability for a random weighted instance.

As MaxCut is deeply connected to the SK model [14], we briefly discuss a “weighted” modification of the SK model obtained by drawing couplings in the SK model from $\mathcal{N}(\mu, \sigma^2)$ instead of $\mathcal{N}(0, 1)$. Here μ may depend on the problem size N . We call this modification “biased SK” and show that it behaves trivially in the infinite-size limit, unless $\mu = \mu(N)$ decays to zero with increasing N .

We evaluate the parameter setting rule implied by Theorem 3 numerically outside of its theoretical assumptions by applying QAOA with $p \in \{1, 2, 3\}$ to MaxCut on a dataset of 34,701 weighted regular and non-regular graphs. We observe that our scheme outperforms the previously proposed approach of Ref. [9]. On average, across all graphs, values of p and edge-weight distributions, QAOA with parameters obtained using our scheme achieves solutions that are only 1.1 percentage points (p.p.) away from optimal, improving upon the 3.5 p.p. obtained using the technique presented in the prior work [9]. Moreover, the disparity from the solutions obtained using optimized parameters is reduced by a factor of three (from 3.6 p.p. to 1.0 p.p.) when the edge weights are drawn from the exponential distribution, and by a factor of ≈ 6 with the Cauchy distribution (from 20.7 to 3.3 p.p.).

We then propose a heuristic parameter rescaling rule for QAOA on arbitrary weighted problems. The heuristic rule is inspired by the theoretical results for the weighted MaxCut problem. As an example highlighting the generality of our observations, we consider QAOA applied to a portfolio optimization problem with a budget constraint, where the constraint is enforced throughout the QAOA evolution by the XY Hamming-weight-preserving mixer. We observe that our simple rescaling procedure makes the landscape easier to optimize, reducing the number of iterations required for convergence to a fixed local optimum by a factor of 7.4x on a dataset of 280 portfolios with between 7 and 20 assets (qubits).

2 Background

We begin by briefly reviewing the Quantum Approximate Optimization Algorithm, the parameter setting schemes for it, and the weighted MaxCut problem.

2.1 Quantum Approximate Optimization Algorithm

Consider the problem of optimizing some objective function $\mathcal{C}(\mathbf{x})$ defined on the n -dimensional Boolean cube that is encoded on n qubits by a diagonal Hamiltonian $C = \text{diag}(\mathcal{C}(\mathbf{x}))$. Quantum Approximate Optimization Algorithm (QAOA)[3, 4] is a hybrid quantum-classical algorithm that approximately solves optimization problems by preparing a parameterized circuit such that upon measuring it, an approximate solution to the optimization problem is obtained. The QAOA circuit consists of layers of alternating unitaries, $e^{-i\gamma C}$ and $e^{-i\beta B}$, where C is the Hamiltonian corresponding to the optimization problem and B is the mixer Hamiltonian. Common choices of the mixer Hamiltonian B include the transverse field ($B = \sum_j X_j$) for unconstrained problems and the XY mixer ($B = \frac{1}{2} \sum_{j,k} (X_j X_k + Y_j Y_k)$) for problems with an equality constraint on the Hamming weight. The QAOA state with p layers is given by

$$|\gamma, \beta\rangle = e^{-i\beta_p B} e^{-i\gamma_p C} \dots e^{-i\beta_1 B} e^{-i\gamma_1 C} |s\rangle, \quad (1)$$

where $|s\rangle$ is the initial state and γ, β are free parameters chosen by a classical routine. We discuss the strategies for setting the parameters γ, β in Sec. 2.2 below.

The figure of merit that we use to evaluate the QAOA performance is the expected solution quality given by the “QAOA energy”:

$$\langle C(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle := \langle \boldsymbol{\gamma}, \boldsymbol{\beta} | C | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle = \sum_{\mathbf{z} \in \{0,1\}^n} \mathcal{C}(\mathbf{z}) \Pr(\mathbf{z}), \quad (2)$$

where $\Pr(\mathbf{z})$ is the probability of observing \mathbf{z} when measuring all qubits of $|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle$.

2.2 Parameter setting strategies for QAOA

Multiple techniques have been proposed for obtaining high-quality parameters for QAOA. While the parameters can be obtained by direct optimization of the objective (2) using a preferred optimization method [15, 16, 17, 18, 19, 20, 21], this procedure is typically computationally expensive [22, 23, 24, 25]. The cost of finding parameters can be significantly reduced by leveraging the apparent problem-instance independence of the optimal QAOA parameters [26, 27]. More straightforwardly, optimized parameters from one instance can be used directly as high-quality parameters for another instance from the same problem class [18, 28, 29, 30, 9]. A machine learning model can be trained that would leverage the concentration to accurately predict the parameters [31, 32, 33, 34, 35, 36]. Optimal parameters can be derived exactly in certain analytically tractable cases, such as triangle-free regular graphs at $p = 1$ [37].

In certain cases, i.e. in the infinite-size limit of a given problem, a closed-form iteration can be derived for the QAOA objective, Equation (2), at constant p . Then parameters can be optimized in the infinite-size limit and used for finite-size instances. This has been demonstrated for the Sherrington-Kirkpatrick model [13] and for MaxCut on random graphs [7, 12]. The goal of this work is to extend these results to weighted problems.

2.3 MaxCut problem

For an undirected graph $G = (V, E)$ with weights $w_{uv} = w_{\{u,v\}}$ assigned to edges $\{u, v\} \in E$, the goal of MaxCut is to partition the set of nodes V into two disjoint subsets, such that the total sum of weights of the edges spanning both partitions is maximized. We refer to this problem as weighted MaxCut in the general case and as unweighted MaxCut when $w_{uv} = 1$ for all $\{u, v\} \in E$.

For the weighted MaxCut problem the objective function is given by

$$\mathcal{C}(\mathbf{z}) = \frac{1}{2} \sum_{\{u,v\} \in E} w_{uv} (1 - z_u z_v), \quad (3)$$

where $z_u \in \{-1, 1\}$ are the variables to be optimized and w_{uv} are sampled from the desired probability distribution. The MaxCut objective is encoded on qubits by the Hamiltonian

$$C = \frac{1}{2} \sum_{\{u,v\} \in E} w_{uv} (1 - Z_u Z_v), \quad (4)$$

where Z_u and Z_v are Pauli-Z operators applied to the u th and v th qubits, respectively.

For unweighted graphs, the cut fraction is defined as the ratio between the number of edges in a cut and the total number of edges in the graph. For a random unweighted $(D + 1)$ -regular graph, the optimal cut fraction is, with high probability, given by

$$\frac{1}{2} + \frac{\Pi_*}{\sqrt{D}} + o\left(\frac{1}{\sqrt{D}}\right), \quad (5)$$

where $\Pi_* \approx 0.7632$ is the Parisi value [14].

3 Parameter setting scheme for QAOA on weighted problems

Our parameter setting scheme is motivated by the observation, formalized in Sec. 4, that in many cases the QAOA energy landscape for weighted MaxCut can be rescaled to match that of unweighted MaxCut for arbitrary p . In the case of weighted MaxCut, this gives an explicit parameter setting rule. In the case of a general objective, we use the same observation to propose a rescaling rule that makes the QAOA energy landscape easier to optimize. We validate our scheme numerically for both cases in Section 6.

3.1 Weighted MaxCut

The proposed procedure is as follows. First, rescale the edge weights in the graph following

$$w_{uv} \rightarrow \frac{w_{uv}}{\sqrt{\frac{1}{|E|} \sum_{\{u,v\} \in E} w_{uv}^2}}. \quad (6)$$

Second, use the parameter setting rule for the corresponding unweighted graph.

As an example of parameter setting rule for unweighted graphs to be used in the second step, one can use the parameters $\beta^{\text{inf}}, \gamma^{\text{inf}}$ optimized for large-girth regular graphs in the infinite-size limit [7, Tables 4 and 5] and follow the rescaling procedure therein, which we include here for completeness: $\beta = \beta^{\text{inf}}, \gamma = \gamma^{\text{inf}}/\sqrt{D}$. Here D is the average degree of the graph. Alternatively, the procedure from Ref. [38] can be used. For small D and p , higher quality results may be obtained by taking inspiration from the explicit formula of Ref. [37] and setting $\gamma = \gamma^{\text{inf}} \arctan \frac{1}{\sqrt{D-1}}$.

As an optional third step, the quality of the parameters can be improved further by running a local optimizer with a small initial step from the parameters obtained in the second step.

3.2 General objective

For a general objective function and QAOA with an arbitrary mixer (e.g., constraint-preserving), analytical results are not available. At the same time, we can use the intuition from MaxCut to rescale the QAOA objective to make the geometry of the landscape more amenable to optimization. Specifically, if the objective f is given by a degree- k polynomial over spins $\mathbf{z} \in \{-1, 1\}^n$:

$$f(\mathbf{z}) = \sum_{\{u_1, \dots, u_k\}} w_{u_1 \dots u_k}^{(k)} z_{u_1} \dots z_{u_k} + \dots + \sum_u w_u^{(1)} z_u, \quad (7)$$

our first step is to divide the objective by

$$\sqrt{\frac{1}{|E_k|} \sum_{\{u_1, \dots, u_k\}} (w_{u_1, \dots, u_k}^{(k)})^2 + \dots + \frac{1}{|E_1|} \sum_u (w_u^{(1)})^2}, \quad (8)$$

where E_i is the set of i -way hyperedges, i.e. the number of terms of order i . In the second step, parameter optimization is performed as usual.

This scaling is inspired by the observation that our results on weighted MaxCut generalize to problems with higher-order (higher than quadratic) terms; see Remark 3. In Section 6 we demonstrate the power of this simple procedure using the example of mean-variance portfolio optimization with a budget constraint enforced by the XY-mixer.

4 Analytical results for QAOA on weighted MaxCut

We now present the analytical results for QAOA applied to weighted MaxCut on large-girth regular graphs with i.i.d. edge weights. We begin by analyzing $p = 1$ in Section 4.1. QAOA energy for $p = 1$ is given by a simple trigonometric formula derived in [39, Theorem 7]. We use this formula to derive globally-optimal QAOA parameters. The parameters we derive are optimal in expectation, with the expectation taken over the distribution of the edge weights. We first consider weights sampled from the exponential distribution and obtain optimal parameters for any graph size (Theorem 1). We analyze the exponential distribution separately as it allows us to derive globally-optimal parameters for finite-sized graphs. Then we consider the infinite-size limit, which enables us to relax the assumption on the distribution and obtain optimal parameters for graphs with weights sampled from an arbitrary distribution (Theorem 2). We then consider $p \geq 1$ in Section 4.3. We extend the techniques of [7] to relate the QAOA objective landscape for weighted MaxCut to that for unweighted MaxCut (Theorem 5) and the SK model (Corollary 6.1).

4.1 Globally-optimal parameters for QAOA with $p = 1$

According to [39, Theorem 7], the expected QAOA performance for MaxCut on triangle-free graphs can be expressed in closed form as:

$$\begin{aligned} \langle C(\gamma, \beta) \rangle &= \frac{\sum_{\{u,v\} \in E} w_{uv}}{2} \\ &+ \frac{\sin(4\beta)}{4} \sum_{\{u,v\} \in E} w_{uv} \sin(w_{uv}\gamma) \left(\prod_{k \in \text{nbhd}(u)/\{v\}} \cos(w_{uk}\gamma) + \prod_{t \in \text{nbhd}(v)/\{u\}} \cos(w_{tv}\gamma) \right), \end{aligned} \quad (9)$$

where $\text{nbhd}(u)$ is the neighborhood function that gives the set of vertices adjacent to u . The above is always maximized at $\beta = \frac{\pi}{8}$. Thus, with a slight abuse of notation, we define $\langle C(\gamma) \rangle = \langle C(\gamma, \frac{\pi}{8}) \rangle$.

We are considering the expected QAOA energy over the edge weights, i.e. $\mathbb{E}_{\mathbf{w}}[\langle C(\gamma) \rangle]$. In the sections that follow $\mathbb{E}_{\mathbf{w}}[\cdot]$ denotes the expectation over the graph weights, w_{uv} , that are all drawn independently from the distribution \mathbf{w} . Thus, for $(D + 1)$ -regular graphs with i.i.d. edge weights, this expectation simplifies as follows

$$\begin{aligned} \mathbb{E}_{\mathbf{w}}[\langle C(\gamma) \rangle] &= \frac{\sum_{\{u,v\} \in E} \mathbb{E}_{\mathbf{w}}[w_{uv}]}{2} + \frac{1}{2} \sum_{\{u,v\} \in E} \mathbb{E}_{\mathbf{w}}[w_{uv} \sin(w_{uv}\gamma)] (\mathbb{E}_{\mathbf{w}}[\cos(w_{uv}\gamma)])^D \\ &= \frac{N(D+1)}{2} \left[\frac{\mathbb{E}_{\mathbf{w}}[w]}{2} + \frac{1}{2} \mathbb{E}_{\mathbf{w}}[w \sin(w\gamma)] (\mathbb{E}_{\mathbf{w}}[\cos(w\gamma)])^D \right], \end{aligned} \quad (10)$$

where we drop the subscript on w since the edge weights are i.i.d. and use the fact that

$$|E| = \frac{N(D+1)}{2}. \quad (11)$$

We now consider edge weights distributed identically and independently according to the exponential distribution with parameter $\lambda > 0$, which has as its probability density function $f(x) = \lambda e^{-\lambda x}$ if $x > 0$ or $f(x) = 0$ otherwise. The mean and standard deviation are $\mu = \sigma = \frac{1}{\lambda}$.

Theorem 1 ($p = 1$, exponential distribution, finite size). *Let $\mathbb{E}_{\text{exp}(\lambda)}[\langle C(\gamma) \rangle]$ denote the expected QAOA objective with $p = 1$ over instances of the weighted MaxCut problem on*

a given triangle-free $(D + 1)$ -regular graph with edge weights, w , drawn i.i.d. from an exponential distribution with parameter λ . Then $\mathbb{E}_{\exp(\lambda)}[\langle C(\gamma) \rangle]$ has a global maximum at $\gamma^* = \frac{1}{\sqrt{\mathbb{E}_{\mathbf{w}}[w^2]} \sqrt{D + \frac{3}{2}}}$.

Proof. To obtain the optimal parameters, we start with Equation (10) and use the following identities

$$\mathbb{E}_{\mathbf{w}}[w \sin(w\gamma)] = \frac{2\gamma\lambda^2}{(\lambda^2 + \gamma^2)^2}, \quad (12)$$

$$\mathbb{E}_{\mathbf{w}}[\cos(w\gamma)] = \frac{\lambda^2}{\lambda^2 + \gamma^2}, \quad (13)$$

which give

$$\begin{aligned} \mathbb{E}_{\exp(\lambda)}[\langle C(\gamma) \rangle] &= \frac{N(D+1)}{2} \left[\frac{1}{2\lambda} + \frac{1}{2} \frac{2\gamma\lambda^2}{(\lambda^2 + \gamma^2)^2} \left(\frac{\lambda^2}{\lambda^2 + \gamma^2} \right)^D \right] \\ &= \frac{N(D+1)}{2} \left[\frac{1}{2\lambda} + \frac{\gamma\lambda^{2D+2}}{(\lambda^2 + \gamma^2)^{D+2}} \right]. \end{aligned} \quad (14)$$

Taking the derivative with respect to γ , we obtain

$$\frac{d}{d\gamma} \mathbb{E}_{\exp(\lambda)}[\langle C(\gamma) \rangle] = c \frac{\lambda^2 - (2D+3)\gamma^2}{(\lambda^2 + \gamma^2)^{D+3}}, \quad (15)$$

where c is a positive and γ -independent constant. Setting the derivative to zero gives

$$\gamma = \frac{\pm\lambda}{\sqrt{2D+3}} = \frac{\pm 1}{\sqrt{\mathbb{E}_{\exp(\lambda)}[w^2]} \sqrt{D + \frac{3}{2}}}. \quad (16)$$

From Equation (14), we can see that

$$\mathbb{E}_{\exp(\lambda)}[\langle C(-\infty) \rangle] = \mathbb{E}_{\exp(\lambda)}[\langle C(\infty) \rangle] = \frac{N(D+1)}{4\lambda},$$

so the global maximum is

$$\gamma^* = \frac{1}{\sqrt{\mathbb{E}_{\exp(\lambda)}[w^2]} \sqrt{D + \frac{3}{2}}}. \quad (17)$$

□

Note that unlike the following Theorems, this proof does not rely on any assumptions on D .

We now consider a graph with edge weights drawn from an arbitrary distribution with mean value μ and standard deviation σ . To study the infinite-size limit, we define a quantity that tends to a constant as $D \rightarrow \infty$. Specifically, we consider the following quantity

$$\frac{\mathbb{E}_{\mathbf{w}}[\langle C(\gamma) \rangle]}{\mathbb{E}_{\mathbf{w}} \left[\sum_{\{u,v\} \in E} w_{uv} \right]}, \quad (18)$$

which reduces to the cut fraction if the graph is unweighted, i.e. $w_{uv} = 1; \forall \{u, v\} \in E$.

Using Equation (11) followed by Equation (10), we can write

$$\begin{aligned} \frac{\mathbb{E}_{\mathbf{w}}[\langle C(\gamma) \rangle]}{\mathbb{E}_{\mathbf{w}} \left[\sum_{\{u,v\} \in E} w_{uv} \right]} &= \frac{2}{N(D+1)\mu} \mathbb{E}_{\mathbf{w}} [\langle C(\gamma) \rangle] \\ &= \frac{1}{2} + \frac{1}{2\mu} \mathbb{E}_{\mathbf{w}} [w \sin(w\gamma)] (\mathbb{E}_{\mathbf{w}} [\cos(w\gamma)])^D \end{aligned} \quad (19)$$

$$= \frac{1}{2} + \frac{\vartheta_1(D, \gamma)}{\sqrt{D}}, \quad (20)$$

where we introduce $\vartheta_1(D, \gamma)$ to match Π_* in Equation (5). We will now show that $\vartheta_1(D, \gamma)$ tends to a D -independent quantity as $D \rightarrow \infty$ when $\gamma = \Theta(D^{-1/2})$, and use the resulting limit to derive the optimal value γ^* in the limit of infinite-sized graphs.

The assumption of $\gamma = \Theta(D^{-1/2})$ is inspired by the numerical observation that the optimal γ for unweighted MaxCut is $\Theta(D^{-1/2})$ (see, e.g. [12, Figure 1b]). Furthermore, we prove that $\vartheta_1(D, \gamma)$ has a local maximum at a value $\gamma = \Theta(D^{-1/2})$ for sufficiently large D . In the limit of $D \rightarrow \infty$, we prove that this local maximum is also the global maximum. This motivates the definition of the following limiting quantity

$$\vartheta_1(\gamma) = \lim_{D \rightarrow \infty} \vartheta_1(D, \gamma/\sqrt{D}). \quad (21)$$

Theorem 2 ($p = 1$, infinite size). *Consider weighted MaxCut on a given triangle-free $(D+1)$ -regular graph with edge weights, w , drawn i.i.d. from a distribution \mathbf{w} with finite second moment. Then for sufficiently large D , the function $\vartheta_1(D, \gamma)$ associated with QAOA for $p = 1$ has a local maximum at a γ that is $\Theta(D^{-1/2})$. Moreover, the limiting quantity $\vartheta_1(\gamma)$ attains its global maximum at $\gamma^* = \frac{1}{\sqrt{\mathbb{E}_{\mathbf{w}}[w^2]}}$.*

Proof. The assumption of finite second moment along with Jensen's inequality implies that $\mathbb{E}_{\mathbf{w}}[|w|]$ is also finite. Thus, since the derivatives of the functions inside the expectations taken in Equation (19) are dominated in γ , i.e. $|w^2 \cos(w\gamma)| \leq w^2$ and $|w \sin(w\gamma)| \leq |w|$, the dominated convergence theorem and the mean-value theorem [40, Section 7.2.2] ensure that the operations of differentiation w.r.t. γ and expectation over w can be interchanged. This gives

$$\begin{aligned} \frac{d}{d\gamma} \vartheta_1(D, \gamma) &= \frac{\sqrt{D}}{2\mu} (\mathbb{E}_{\mathbf{w}}[\cos(w\gamma)])^{D-1} \\ &\cdot \left(\mathbb{E}_{\mathbf{w}}[w^2 \cos(w\gamma)] \mathbb{E}_{\mathbf{w}}[\cos(w\gamma)] - D (\mathbb{E}_{\mathbf{w}}[w \sin(w\gamma)])^2 \right). \end{aligned} \quad (22)$$

Substituting $\gamma = \frac{\alpha}{\sqrt{D}}$, for α independent of D and using the Taylor series expansions of the trigonometric functions, we get

$$\left. \frac{d}{d\gamma} \vartheta_1(D, \gamma) \right|_{\gamma=\alpha/\sqrt{D}} = \frac{\sqrt{D}}{2\mu} e^{-\frac{\mathbb{E}_{\mathbf{w}}[w^2]\alpha^2}{2}} [\mathbb{E}_{\mathbf{w}}[w^2] - (\mathbb{E}_{\mathbf{w}}[w^2])^2 \alpha^2] + O(D^{-1/2}), \quad (23)$$

where the implicit exchange of infinite series and expectation over w is justified by the finiteness of the second moment and Fubini's theorem [41, Theorem 8.8]. Here we use the observation that

$$\begin{aligned} \cos(x/\sqrt{D})^{D-1} &= \left(1 - \frac{x^2}{2D} \right)^{D-1} + O(D^{-1}) \\ &= e^{-\frac{x^2}{2}} + O(D^{-1}), \end{aligned} \quad (24)$$

for x that is bounded by a constant independent of D .[‡]

For sufficiently large D , both

$$\frac{d}{d\gamma}\vartheta_1(D, \gamma)\Big|_{\gamma=0} = \frac{\sqrt{D}}{2\mu}\mathbb{E}_{\mathbf{w}}[w^2] + O(D^{-1/2}) > 0 \quad (25)$$

and

$$\frac{d}{d\gamma}\vartheta_1(D, \gamma)\Big|_{\gamma=\alpha^*/\sqrt{D}} \leq 0 \quad (26)$$

for some sufficiently large constant α^* independent of D . Thus, by Darboux's theorem [42, Theorem 5.12], γ has a local maxima in the interval $(0, \alpha^*/\sqrt{D})$ for each triangle-free $(D+1)$ -regular graph.

We now consider the limiting value of ϑ_1 in the regime of small γ . With $\gamma = \frac{\gamma'}{\sqrt{D}}$ for some D -independent γ' , we get

$$\begin{aligned} \vartheta_1(D, \gamma'/\sqrt{D}) &= \frac{\sqrt{D}}{2\mu}\mathbb{E}_{\mathbf{w}}[w \sin(w\gamma'/\sqrt{D})](\mathbb{E}_{\mathbf{w}}[\cos(w\gamma'/\sqrt{D})])^D \\ &= \frac{\mathbb{E}_{\mathbf{w}}[w^2]}{2\mu}\gamma' e^{-\frac{\mathbb{E}_{\mathbf{w}}[w^2]\gamma'^2}{2}} + O(D^{-1}), \end{aligned}$$

where we use Equation (24), and the implicit exchange of infinite series and expectation is justified by Fubini's theorem. Now taking the limit in D , we obtain

$$\vartheta_1(\gamma') := \lim_{D \rightarrow \infty} \vartheta_1(D, \gamma'/\sqrt{D}) = \frac{\mathbb{E}_{\mathbf{w}}[w^2]}{2\mu}\gamma' e^{-\frac{\mathbb{E}_{\mathbf{w}}[w^2]\gamma'^2}{2}}. \quad (27)$$

Now, consider the derivative,

$$\frac{d}{d\gamma'}\vartheta_1(\gamma') = \frac{\mathbb{E}_{\mathbf{w}}[w^2]e^{-\frac{\mathbb{E}_{\mathbf{w}}[w^2]\gamma'^2}{2}}}{2\mu}(1 - \mathbb{E}_{\mathbf{w}}[w^2]\gamma'^2). \quad (28)$$

It can be easily seen that the function $\vartheta_1(\gamma')$ is always decreasing to the right of the local maximum at $\gamma^* = \frac{1}{\sqrt{\mathbb{E}_{\mathbf{w}}[w^2]}}$, and the function is negative to the left of zero. Thus this is in fact a global optima. \square

Remark 1. To see the correspondence between Theorem 2 and Theorem 1, i.e. that the latter is a special case of the former, rescale $\gamma \rightarrow \gamma/\sqrt{D}$ and note that the constant in the denominator in Theorem 1 has no effect on the limiting value as $D \rightarrow \infty$.

4.2 Concentration of QAOA objective at $p = 1$

We show that the QAOA objective for weighted MaxCut instances concentrates sharply around its expectation as $D \rightarrow \infty$, for triangle-free $(D+1)$ -regular graphs when $p = 1$. This indicates that our scaling rules, which are derived by investigating the expectation of the objective, can also be expected to hold with high probability for *any* weighted instance of a fixed graph. The QAOA objective in this setting is given in closed form in Equation (9), where we set $\beta = \frac{\pi}{8}$. We will use the following bound on the partial derivatives of the objective as a function of the individual edge weights.

[‡]Note that the second equality follows from $(1 - \frac{x^2}{2D})^{D-1} = (1 - \frac{x^2}{2D})^D(1 - O(D^{-1})) = \left[\sum_{k=0}^D \frac{(-x^2/2)^k}{k!}\right](1 - O(D^{-1}))^2$ and works in the limit due to absolute convergence.

Lemma 2.1. Let $\langle C(\gamma) \rangle$ be the $p = 1$ QAOA objective for weighted MaxCut on triangle-free $(D + 1)$ -regular graphs as given in Equation (9), with $\beta = \frac{\pi}{8}$. The partial derivatives of $\langle C(\gamma) \rangle$ with respect to the weights w_{ij} satisfy,

$$\left| \frac{\partial \langle C(\gamma) \rangle}{\partial w_{ij}} \right| \leq \frac{1}{2} + \frac{(1 + \gamma)|w_{ij}|}{2} + \frac{\gamma}{4} \left(\sum_{v \in \text{nbhd}(i)/\{j\}} |w_{iv}| + \sum_{u \in \text{nbhd}(j)/\{i\}} |w_{uj}| \right) \quad (29)$$

Proof. We define for convenience,

$$T_{uv} = w_{uv} \sin(w_{uv}\gamma) \left(\prod_{k \in \text{nbhd}(u)/\{v\}} \cos(w_{uk}\gamma) + \prod_{t \in \text{nbhd}(v)/\{u\}} \cos(w_{vt}\gamma) \right)$$

so that $\langle C(\gamma) \rangle = \sum_{\{u,v\} \in E} \frac{w_{uv}}{2} + \frac{T_{uv}}{4}$. Clearly, $\frac{\partial \langle C(\gamma) \rangle}{\partial w_{ij}} = \frac{1}{2} + \frac{1}{4} \sum_{\{u,v\} \in E} \frac{\partial T_{uv}}{\partial w_{ij}}$. To determine the last term, we have four cases,

1. $u = i; v = j$: In T_{uv} , the term inside the parenthesis is upper bounded by 2 and independent of w_{ij} . We have,

$$\left| \frac{\partial T_{uv}}{\partial w_{ij}} \right| \leq 2(\sin(w_{ij}) + \gamma w_{ij} \cos(w_{ij})) \quad (30)$$

$$\leq 2(1 + \gamma)|w_{ij}|. \quad (31)$$

2. $u = i; j \in \text{nbhd}(u)/\{v\}$: In T_{uv} , only the first term inside the parenthesis depends on w_{ij} . We have,

$$\left| \frac{\partial T_{uv}}{\partial w_{ij}} \right| \leq |w_{iv}| (\gamma \cos(w_{ij})) \leq \gamma |w_{iv}|. \quad (32)$$

3. $v = j; i \in \text{nbhd}(v)/\{u\}$: In T_{uv} , only the second term inside the parenthesis depends on w_{ij} . We have,

$$\left| \frac{\partial T_{uv}}{\partial w_{ij}} \right| \leq |w_{uj}| (\gamma \cos(w_{ij})) \leq \gamma |w_{uj}|. \quad (33)$$

4. $u \neq i, v \neq j$: T_{uv} is independent of w_{ij} and $\frac{\partial T_{uv}}{\partial w_{ij}} = 0$.

Combining the above four cases, the result follows. □

We shall also use the following straightforward observation.

Lemma 2.2. For any $(D + 1)$ -regular triangle free graph, it must be that the number of vertices $N \geq D(D + 1)$.

Proof. Consider any node u . By definition u has $D + 1$ unique neighbors. Each neighbor of u has D neighbors in addition to u . Additionally, none of these new neighbors can also be neighbors of u (or they will form a triangle). The observation follows. □

We are now ready to establish the concentration of $\langle C(\gamma) \rangle$ for some common edge weight distributions, at the value of γ that optimizes the expected objective.

4.2.1 Bounded Edge Weights

In this setting the edge weights are independently and identically sampled from a distribution that is supported on an interval $[a, b]$, where $|a|, |b|$ are independent of D . We can show concentration by using the well-known McDiarmid's inequality [43], which we state in the necessary form below.

Lemma 2.3 (McDiarmid's Inequality [Adapted from [44, Theorem 1]]). *Let X_1, \dots, X_n be independent random variables each with range \mathcal{X} . Let $f: \mathcal{X}^n \rightarrow \mathbb{R}$ be any function with the bounded differences property, i.e., for any co-ordinate $i \in [n]$ and points $(x_1, \dots, x_d), (x'_1, \dots, x'_d) \in \mathcal{X}^n$ such that $x_j = x'_j$ for all $j \in [n] \setminus \{i\}$, we have*

$$|f(x_1, \dots, x_d) - f(x'_1, \dots, x'_d)| \leq c_i. \quad (34)$$

Then it holds that

$$\Pr[|f(X_1, \dots, X_n) - \mathbb{E}f(X_1, \dots, X_n)| \geq \epsilon] \leq 2 \exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^n c_i^2}\right). \quad (35)$$

We now show a concentration inequality on the relative error from the mean of the QAOA objective under this setting.

Theorem 3 ($p = 1$, concentration, bounded weights). *Let the edge weights of a triangle-free $(D + 1)$ -regular graph (V, E) with $|V| = N$ vertices, be chosen from a distribution \mathbf{w} with mean μ , that is supported on the interval $[a, b]$, where $|a|, |b|, |\mu|$ are independent of D . Then, for all $\gamma = \Theta(D^{-\alpha})$ for some $\alpha \in (0, 1)$, the QAOA objective $\langle C(\gamma) \rangle$ with $p = 1$ satisfies the following concentration inequality, where K is independent of D :*

$$\Pr[|\langle C(\gamma) \rangle - \mathbb{E}_{\mathbf{w}}[\langle C(\gamma) \rangle]| \geq \epsilon |\mathbb{E}_{\mathbf{w}}[\langle C(\gamma) \rangle]|] \leq 2 \exp\left(\frac{-\epsilon^2 D^{1+2\alpha}(1 - O(D^{-\alpha}))}{K}\right). \quad (36)$$

Proof. Assume wlog that $|a| \leq |b|$. The $N(D+1)/2$ edge weights w_{ij} form a set of identical independent random variables taking values in $[a, b]$. Thus for any $\{i, j\} \in E$, $|w_{ij}| \leq |b|$. Using the result of Lemma 2.1, and noting that for any $\{i, j\} \in E$, there are exactly D vertices in each of the sets $\text{nbhd}(i) \setminus \{j\}$ and $\text{nbhd}(j) \setminus \{i\}$, we have

$$\forall \{i, j\} \in E, \quad \left| \frac{\partial \langle C(\gamma) \rangle}{\partial w_{ij}} \right| \leq \frac{1 + (1 + \gamma)|b| + \gamma D|b|}{2}. \quad (37)$$

The right hand side of the above inequality can be written as $K_1 + \gamma DK_2$ where K_1, K_2 are independent of D .

Now, we view $\langle C(\gamma) \rangle$ as a function of $N(D + 1)/2$ identical independent random variables. By the mean value theorem, Equation (37), the assumption $\gamma = \Theta(D^{-\alpha})$, and the boundedness of w_{ij} , $\langle C(\gamma) \rangle$ satisfies the bounded differences property (Equation (34)) with respect to w_{ij} with $c_{ij} \leq |b - a|(K_1 + \gamma DK_2) \leq K_3 \gamma D$ with K_3 independent of D .

Noting that $|\mathbb{E}_{\mathbf{w}}[\cos(w\gamma)]| \leq 1$ and $\mathbb{E}_{\mathbf{w}}[w \sin(w\gamma)] = O(D^{-\alpha})$, for $\gamma = \Theta(D^{-\alpha})$, from Equation (10) it follows that

$$|\mathbb{E}_{\mathbf{w}}[\langle C(\gamma) \rangle]| \geq \frac{N(D + 1)|\mu|(1 - O(D^{-\alpha}))}{2}. \quad (38)$$

Finally,

$$\begin{aligned}
\Pr \left[\left| \langle C(\gamma) \rangle - \mathbb{E}_{\mathbf{w}} [\langle C(\gamma) \rangle] \right| \geq \epsilon \left| \mathbb{E}_{\mathbf{w}} [\langle C(\gamma) \rangle] \right| \right] \\
\leq 2 \exp \left(\frac{-\epsilon^2 N^2 (D+1)^2 |\mu|^2 (1 - O(D^{-\alpha}))}{2 \sum_{\{i,j\} \in E} c_{ij}^2} \right) \\
\leq 2 \exp \left(\frac{-\epsilon^2 N (D+1) |\mu|^2 (1 - O(D^{-\alpha}))}{K_3^2 \gamma^2 D^2} \right) \\
\leq 2 \exp \left(\frac{-\epsilon^2 (D+1)^2 D^{2\alpha-1} |\mu|^2 (1 - O(D^{-\alpha}))}{K_4^2} \right), \tag{39}
\end{aligned}$$

where K_4 is a constant independent of D . Here, the first inequality follows from Lemma 2.3, and the final inequality follows from Lemma 2.2. Setting $K = \frac{K_3^2}{|\mu|^2}$, the result of the theorem follows. \square

Theorem 3 applies immediately to the case where the parameter γ is set to $\frac{1}{\sqrt{\mathbb{E}_{\mathbf{w}}[w^2]}D}$. Such a parameter setting is shown in Theorem 2 to maximize the expected objective in the limit $D \rightarrow \infty$.

4.2.2 Gaussian Edge Weights

In order to show concentration when the weight distribution of each edge is Gaussian, we need an additional assumption that the number of vertices N is such that $\log(N) = o(D \log(D))$. We use the following result on the sub-gaussian concentration of Lipschitz continuous functions of standard normal variables.

Lemma 3.1 (Adapted from [45, Theorem 5.2.2]). *Consider an n -dimensional random vector X drawn from the n -dimension standard normal, i.e. $X \sim \mathcal{N}(0, I_n)$, and a differentiable function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ that is L -Lipschitz continuous, i.e. $\|\nabla f(x)\|_2 \leq L$ for all $x \in \mathbb{R}^n$. Then,*

$$\Pr [|f(X) - \mathbb{E}f(X)| \geq t] \leq \exp \left(\frac{-t^2}{K_\psi L^2} \right), \tag{40}$$

where K_ψ is a universal constant.

The Lipschitz condition of Lemma 3.1 does not directly apply to the QAOA objective because the weights may be unbounded. Our argument will therefore address the concentration in two cases, using Lemma 3.1 to show concentration inside a suitably chosen interval, and Gaussian tail bounds to show that the probability of lying outside this interval is sufficiently small. We formalize this in the following concentration inequality on the relative error from the mean of the QAOA objective.

Theorem 4 ($p = 1$, concentration, Gaussian weights). *Let the edge weights of a $(D+1)$ -regular triangle-free graph (V, E) with $|V| = N$ vertices such that $\log(N) = o(D \log(D))$, be chosen from the normal distribution $\mathcal{N}(\mu, \sigma^2)$, where μ, σ are independent of D . Then, for all $\gamma = \Theta(D^{-1/2})$, the QAOA objective $\langle C(\gamma) \rangle$ with $p = 1$ satisfies the following concentration inequality, where K is independent of D :*

$$\Pr \left[\left| \langle C(\gamma) \rangle - \mathbb{E}_{\mathbf{w}} [\langle C(\gamma) \rangle] \right| \geq \epsilon \left| \mathbb{E}_{\mathbf{w}} [\langle C(\gamma) \rangle] \right| \right] \leq 2 \exp \left(\frac{-\epsilon D (1 - o(1))}{K \log(D)} \right). \tag{41}$$

Proof. For any $\epsilon > 0$, define the interval

$$\mathcal{I}_\eta = [\mu - t_\eta\sigma, \mu + t_\eta\sigma], \quad (42)$$

where

$$t_\eta = \sqrt{2(\epsilon^\eta D \log(D) + \log(N) + \log(D + 1))}, \quad (43)$$

with η to be chosen later. Let \mathcal{B}_η denote the event that $w_{ij} \in \mathcal{I}_\eta$ for all $\{i, j\} \in E$, and the event \mathcal{E}_ϵ as $[|\langle C(\gamma) \rangle - \mathbb{E}_\mathbf{w}[\langle C(\gamma) \rangle]| \geq \epsilon |\mathbb{E}_\mathbf{w}[\langle C(\gamma) \rangle]|]$. From the definition of conditional probability,

$$\begin{aligned} \Pr[\mathcal{E}_\epsilon] &= \Pr[\neg\mathcal{B}_\eta \cap \mathcal{E}_\epsilon] + \Pr[\mathcal{B}_\eta \cap \mathcal{E}_\epsilon] \\ &\leq \Pr[\neg\mathcal{B}_\eta] + \Pr[\mathcal{E}_\epsilon | \mathcal{B}_\eta]. \end{aligned} \quad (44)$$

We bound $\Pr[\neg\mathcal{B}_\eta]$ by the following consideration:

$$\begin{aligned} \Pr[\neg\mathcal{B}_\eta] &= \Pr[\exists\{i, j\} \in E \text{ s.t. } w_{ij} \notin \mathcal{I}_\eta] \\ &\leq \frac{N(D+1)}{2} \exp(-t_\eta^2/2) \\ &\leq \exp(-t_\eta^2/2 + \log(N) + \log(D+1)) \\ &= \exp(-\epsilon^\eta D \log(D)), \end{aligned} \quad (45)$$

where the first inequality follows by the union bound and the final equality by definition of t_η .

To bound the second term on the RHS of (44), note that in the event of \mathcal{B}_η , the weights are all bounded in the interval $[\mu - t_\eta\sigma, \mu + t_\eta\sigma]$. The absolute value of any weight in this interval is bounded above by $b_\eta = |\mu + \text{sgn}(\mu)t_\eta\sigma|$. We now consider the QAOA objective $\langle C(\gamma) \rangle$ as a function of a random $N(D+1)/2$ -dimensional vector $Z = (z_{ij})_{\{i,j\} \in E}$ whose entries are each drawn from the standard normal distribution $\mathcal{N}(0, 1)$ by defining $z_{ij} = \frac{w_{ij} - \mu}{\sigma}$. From Lemma 2.1 and an application of the chain rule,

$$\begin{aligned} \forall\{i, j\} \in E, \quad \left| \frac{\partial \langle C(\gamma) \rangle}{\partial z_{ij}} \right| &= \left| \frac{\partial w_{ij}}{\partial z_{ij}} \right| \cdot \left| \frac{\partial \langle C(\gamma) \rangle}{\partial w_{ij}} \right| \\ &\leq \frac{\sigma(1 + (1 + \gamma)b_\eta + \gamma D b_\eta)}{2} \\ &= \frac{\sigma \gamma D b_\eta (1 + O(D^{-1}))}{2}. \end{aligned} \quad (46)$$

We can now bound the Lipschitz constant L of $\langle C(\gamma) \rangle$ as

$$L \leq \sqrt{\frac{N(D+1)}{2}} \max_{\{i,j\} \in E} \left| \frac{\partial \langle C(\gamma) \rangle}{\partial z_{ij}} \right|, \quad (47)$$

which leads to the following sequence of deductions:

$$\begin{aligned} L^2 &\leq \frac{(ND^2(D+1)\sigma^2\gamma^2|\mu + \text{sgn}(\mu)t_\eta\sigma|^2)(1 + O(D^{-1}))}{8} \\ &\leq \frac{(ND^2(D+1)\sigma^4\gamma^2t_\eta^2)(1 + O(D^{-1}))}{8} \\ &\leq \frac{(ND^3(D+1)\log(D)\sigma^4\gamma^2\epsilon^\eta)(1 + o(1))}{4} \\ &\leq K_1 ND^2(D+1)\log(D)\sigma^4\epsilon^\eta(1 + o(1)), \end{aligned} \quad (48)$$

where K_1 is a constant independent of D . The second inequality follows because $\frac{\mu}{\sigma t_\eta} = O(D^{-1/2})$, the third because $\log(D+1), \log(N) = o(D \log(D))$ by assumption, and the fourth because $\gamma = \Theta(D^{-1/2})$ by assumption.

Finally, we can return to Equation (10) to see that

$$|\mathbb{E}_{\mathbf{w}}[\langle C(\gamma) \rangle]| \geq \frac{N(D+1)|\mu|(1 - O(D^{-1/2}))}{2}. \quad (49)$$

Using Lemma 3.1 we have that,

$$\begin{aligned} \Pr[\mathcal{E}_\epsilon \mid \mathcal{B}_\eta] &\leq \exp\left(-\frac{\epsilon^{2-\eta} N(D+1) \mu^2 (1 - o(1))}{4K_\psi K_1 D^2 \log(D) \sigma^4}\right) \\ &\leq \exp\left(-\frac{\epsilon^{2-\eta} D(D+1)^2 \mu^2 (1 - o(1))}{4K_\psi K_1 D^2 \log(D) \sigma^4}\right) \\ &\leq \exp\left(-\frac{\epsilon^{2-\eta} D(1 - o(1))}{K_2 \log(D)}\right), \end{aligned} \quad (50)$$

where K_2 is a constant independent of D . Finally, setting $\eta = 1$, the result follows from Equations (44,45,50). \square

4.2.3 Summary of concentration results

In view of our concentration results for graphs with weight distributions that have bounded support (Theorem 3), or for normally distributed weights (Theorem 4), we see that, in the limit $D \rightarrow \infty$, and when $\gamma = \Theta(D^{-1/2})$, the typical relative deviations from the mean for the QAOA objective $\langle C(\gamma) \rangle$ are of the order $\tilde{O}(D^{-1})$, which ignores terms that are $O(\text{polylog}(D))$. Particularly, the probability of a relative deviation greater than ϵ are $\approx \exp(-\epsilon^2 D^2)$ and $\approx \exp(-\epsilon D / \log D)$ for bounded and Gaussian weights, respectively. Crucially, this shows that the relative dimensions in the quantity $\vartheta_1(D, \gamma)$ have relative deviations of the order $\tilde{O}(D^{-1/2})$. This is notable as $\vartheta_1(D, \gamma)$ tends to a constant as $D \rightarrow \infty$ when $\gamma = \Theta(D^{-1/2})$ (Theorem 2), and is the primary quantity of interest when investigating the cut fraction (tending to the Parisi value in the unweighted case [14]).

4.3 Correspondence between QAOA on weighted and unweighted graphs with $p \geq 1$

To derive a parameter scaling rule for arbitrary p , we extend the techniques developed in Ref. [7] for MaxCut on large-girth, regular unweighted graphs to large-girth, regular graphs with i.i.d. edge weights. Without a subscript, $|\gamma, \beta\rangle$ will refer to the p -layer QAOA state for a random weighted instance of MaxCut on a given $(D+1)$ -regular graph with weights, w , drawn from a distribution \mathbf{w} .

Note that

$$\langle C(\gamma, \beta) \rangle = \frac{1}{2} \sum_{\{u,v\} \in E} w_{uv} (1 - \langle \gamma, \beta | z_u z_v | \gamma, \beta \rangle). \quad (51)$$

We start by proving the following result about the above quantity that is valid for any p .

Lemma 4.1. *If a $(D+1)$ regular graph has girth $> 2p+1$ and i.i.d. edge weights, w , drawn from \mathbf{w} , then the QAOA objective for weighted MaxCut on this graph satisfies*

$$\mathbb{E}_{\mathbf{w}}[\langle C(\gamma, \beta) \rangle] = \frac{N(D+1)\mu}{4} - \frac{N(D+1)}{4} \mathbb{E}_{\mathbf{w}}[w \langle \gamma, \beta | z_L z_R | \gamma, \beta \rangle], \quad (52)$$

for any edge $\{L, R\}$.

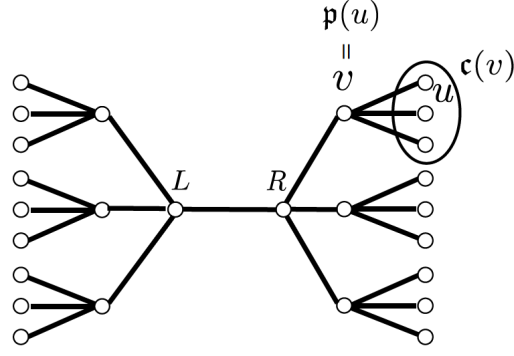


Figure 1: Unweighted tree subgraph seen by QAOA from the edge $\{L, R\}$ with $p = 2$ on a four-regular graph. The operation \mathfrak{p} produces the parent of a node, and the operation \mathfrak{c} produces the set of a node's immediate children.

Proof. The locality of the p -layer QAOA combined with the regularity and girth $> 2p + 1$ assumptions implies that the quantity $w_{LR}\langle\gamma, \beta|z_L z_R|\gamma, \beta\rangle$ can only depend on vertices that lie in a D -ary tree of diameter $2p + 1$. This is the QAOA “light cone” for the term $\langle\gamma, \beta|z_L z_R|\gamma, \beta\rangle$. The tree subgraph, without weights, for $p = 2$ and $D + 1 = 4$ is shown in Figure 1. Thus, it follows from [7, Equations A.2-A.5] that

$$w_{LR}\langle\gamma, \beta|z_L z_R|\gamma, \beta\rangle = \sum_{\{z_u\}} w_{LR} z_L^{[0]} z_R^{[0]} \cdot \left[\exp \left(-i \sum_{\{u,v\} \in E_{LR}} w_{uv} \mathbf{\Gamma} \cdot (z_u z_v) \right) \prod_{v \in V_{LR}} g(z_v) \right], \quad (53)$$

where (V_{LR}, E_{LR}) denotes the vertex and edge sets corresponding to the tree subgraph seen from the edge $\{L, R\}$. Note that Ref. [7] has an extra factor of $\frac{1}{\sqrt{D}}$ in the exponential since the cost function in Equation (51) is therein defined with that extra factor. In addition, $\mathbf{\Gamma}$ is a $(2p + 1)$ -component vector with entries $\mathbf{\Gamma}_r = \gamma_r$, $\mathbf{\Gamma}_{-r} = -\gamma_r$, $\mathbf{\Gamma}_0 = 0$ and $1 \leq r \leq p$. Also for node u , the vector $\mathbf{z}_u = (z_u^{[1]}, \dots, z_u^{[r]}, z_u^{[0]}, z_u^{[-r]}, \dots, z_u^{[-1]}) \in \{-1, 1\}^{2p+1}$. The notation $\{z_u\}$ is short for the collection $\{z_u | u \in V_{LR}, z_u \in \{-1, 1\}^{2p+1}\}$. Lastly $(z_u z_v)$ denotes an element-wise product, and

$$g(\mathbf{z}_u) = \left[\langle z_u^{[p]} | e^{i\beta_p x} | z_u^{[0]} \rangle \langle z_u^{[0]} | e^{-i\beta_p x} | z_u^{[-p]} \rangle \right] \cdot \frac{1}{2} \prod_{r=1}^{p-1} \langle z_u^{[-(r+1)]} | e^{-i\beta_r x} | z_u^{[-r]} \rangle \langle z_u^{[r]} | e^{i\beta_r x} | z_u^{[r+1]} \rangle. \quad (54)$$

The authors of [7] noted that Equation (53) can be computed recursively by traversing, from leaves to roots, the left and right branches simultaneously. This effectively “factors” the right-hand side of Equation (53). For simplicity we will only do this for $p = 2$, however, the generalization is straightforward. Define $H_D^{(0)} = 1$. We start by summing over the configuration of an arbitrary leaf in either branch of Figure 1:

$$\sum_{z_u} g(z_u) \exp \left(-i w_{\mathfrak{p}(u)} \mathbf{\Gamma} \cdot (z_u z_{\mathfrak{p}(u)}) \right), \quad (55)$$

where $\mathfrak{p}(u)$ is the parent of u in the tree subgraph. Since the left and right branches are D -ary trees, the expression for any node v in the second level of either branch is

$$\sum_{\mathbf{z}_v} g(\mathbf{z}_v) H_D^{(1)}(\mathbf{z}_v) \exp\left(-i w_{v\mathbf{p}(v)} \mathbf{\Gamma} \cdot (\mathbf{z}_v \mathbf{z}_{\mathbf{p}(v)})\right), \quad (56)$$

where

$$\begin{aligned} H_D^{(1)}(\mathbf{z}_v) &:= \prod_{u \in \mathfrak{c}(v)} \sum_{\mathbf{z}_u} H_D^{(0)}(\mathbf{z}_u) g(\mathbf{z}_u) \exp(-i w_{uv} \mathbf{\Gamma} \cdot (\mathbf{z}_u \mathbf{z}_v)) \\ &= \prod_{u \in \mathfrak{c}(v)} \sum_{\mathbf{z}_u} H_D^{(0)}(\mathbf{z}_u) g(\mathbf{z}_u) \cos(w_{uv} \mathbf{\Gamma} \cdot (\mathbf{z}_u \mathbf{z}_v)). \end{aligned} \quad (57)$$

The second equality follows from the even parity of $g \cdot H_D^{(r)}$, implied by [7, Claims A.14 and A.15] and $\mathfrak{c}(v)$ denotes the set of D immediate children of v .

This can again be done for the next level, i.e. the roots L or R , producing the quantity $H_D^{(2)}$. Lastly, we combine the results from the two branches by summing over the configurations of the left and right roots:

$$\begin{aligned} w_{LR} \langle \gamma, \beta | Z_u Z_v | \gamma, \beta \rangle &= \sum_{\mathbf{z}_L, \mathbf{z}_R} w_{LR} z_L^{[0]} z_R^{[0]} \\ &\quad \cdot \left[g(\mathbf{z}_L) g(\mathbf{z}_R) H_D^{(2)}(\mathbf{z}_L) H_D^{(2)}(\mathbf{z}_R) \exp(-i w_{LR} \mathbf{\Gamma} \cdot (\mathbf{z}_L \mathbf{z}_R)) \right] \\ &= -i \sum_{\mathbf{z}_L, \mathbf{z}_R} z_L^{[0]} z_R^{[0]} g(\mathbf{z}_L) g(\mathbf{z}_R) \\ &\quad \cdot \left[H_D^{(2)}(\mathbf{z}_L) H_D^{(2)}(\mathbf{z}_R) w_{LR} \sin(w_{LR} \mathbf{\Gamma} \cdot (\mathbf{z}_L \mathbf{z}_R)) \right], \end{aligned} \quad (58)$$

where the last equality follows again from the even parity of $g \cdot H_D^{(r)}$. The general iteration for the random quantity $w_{LR} \langle \gamma, \beta | Z_L Z_R | \gamma, \beta \rangle$ follows by induction.

Note that in the previous recursion, each edge was only counted once. Since all of the edges are i.i.d., the expectation operation commutes with all products that appear in the right-hand side of Equation (58). More specifically, we have the following for general p :

$$\begin{aligned} \mathbb{E}_{\mathbf{w}}[w_{LR} \langle \gamma, \beta | Z_L Z_R | \gamma, \beta \rangle] &= -i \sum_{\mathbf{z}_L, \mathbf{z}_R} z_L^{[0]} z_R^{[0]} g(\mathbf{z}_L) g(\mathbf{z}_R) \\ &\quad \cdot \left[\bar{H}_D^{(p)}(\mathbf{z}_L) \bar{H}_D^{(p)}(\mathbf{z}_R) \mathbb{E}_{\mathbf{w}}[w \sin(w \mathbf{\Gamma} \cdot (\mathbf{z}_L \mathbf{z}_R))] \right], \end{aligned} \quad (59)$$

where subscripts have been dropped from the weights due to the i.i.d. assumption, and

$$\bar{H}_D^{(r)}(\mathbf{z}_v) := \mathbb{E}_{\mathbf{w}}[H_D^{(r)}(\mathbf{z}_v)] = \prod_{u \in \mathfrak{c}(v)} \sum_{\mathbf{z}_u} \bar{H}_D^{(r-1)}(\mathbf{z}_u) g(\mathbf{z}_u) \mathbb{E}_{\mathbf{w}}[\cos(w \mathbf{\Gamma} \cdot (\mathbf{z}_u \mathbf{z}_v))] \quad (60)$$

for $1 \leq r \leq p$.

Since all randomness has been removed by the expectation operation, it is evident that this quantity only depends on the graph structure and not the sampled weights, since they are i.i.d. The non-random quantities in the iteration for $w_{LR} \langle \gamma, \beta | Z_L Z_R | \gamma, \beta \rangle$ only depend on the local graph structure that QAOA sees from a given edge. As argued earlier this graph structure is always two D -ary trees joined at their roots. Thus, the quantity $\mathbb{E}_{\mathbf{w}}[w \langle \gamma, \beta | Z_L Z_R | \gamma, \beta \rangle]$ is independent of the chosen edge $\{L, R\}$, which is analogous to the unweighted case.

The result of the lemma follows from Equation (51) and the linearity of expectation. \square

By the previous lemma, we can, analogously to the $p = 1$ case, define $\vartheta_p(D, \gamma, \beta)$ as follows:

$$\begin{aligned} \frac{\mathbb{E}_{\mathbf{w}}[\langle C(\gamma, \beta) \rangle]}{\mathbb{E}_{\mathbf{w}} \left[\sum_{\{u,v\} \in E} w_{uv} \right]} &= \frac{\mathbb{E}_{\mathbf{w}}[\langle C(\gamma, \beta) \rangle]}{\frac{N(D+1)\mu}{2}} \\ &= \frac{1}{2} - \frac{1}{2\mu} \mathbb{E}_{\mathbf{w}}[w \langle \gamma, \beta |_{ZZ} | \gamma, \beta \rangle] \\ &= \frac{1}{2} + \frac{\vartheta_p(D, \gamma, \beta)}{\sqrt{D}}, \end{aligned} \quad (61)$$

and

$$\begin{aligned} \vartheta_p(\gamma, \beta) &:= \lim_{D \rightarrow \infty} \vartheta_p(D, \gamma/\sqrt{D}, \beta) \\ &= \lim_{D \rightarrow \infty} -\frac{\sqrt{D}}{2\mu} \mathbb{E}_{\mathbf{w}} \left[w \left\langle \frac{\gamma}{\sqrt{D}}, \beta \middle|_{ZZ} \middle| \frac{\gamma}{\sqrt{D}}, \beta \right\rangle \right], \end{aligned} \quad (62)$$

where the subscripts on w and z have been dropped since they can be arbitrary by the previous lemma.

The quantity considered in [7, Equation A.19] for the unweighted case is the following

$$\nu_p(\gamma, \beta) = \lim_{D \rightarrow \infty} -\frac{\sqrt{D}}{2} \left\langle \frac{\gamma}{\sqrt{D}}, \beta \middle|_{ZZ} \middle| \frac{\gamma}{\sqrt{D}}, \beta \right\rangle_{\mathbf{u}}, \quad (63)$$

where the subscript ‘‘u’’ indicates that the parameterized state is prepared by a p -layer QAOA for the corresponding unweighted problem on the same graph. Our main result below shows that γ can be scaled to make these two quantities equal up to a global scaling factor.

Theorem 5 ($p \geq 1$, infinite size). *If the girth $> 2p + 1$, and the edge-weight distribution, \mathbf{w} , has finite second moment, then for all parameters γ, β the following holds*

$$\nu_p(\gamma, \beta) = \frac{\mu}{\sqrt{\mathbb{E}_{\mathbf{w}}[w^2]}} \vartheta_p \left(\frac{\gamma}{\sqrt{\mathbb{E}_{\mathbf{w}}[w^2]}}, \beta \right). \quad (64)$$

Proof. We implicitly assume that $\gamma \mapsto \frac{\gamma}{\sqrt{D}}$ and thus $\Gamma \mapsto \frac{\Gamma}{\sqrt{D}}$ in Equation (59). By the product rule for limits, we can evaluate the limits of the terms $\bar{H}_D^{(p)}(z_L)$, $\bar{H}_D^{(p)}(z_R)$, and the one involving the sin separately, since we will show they individually exist. Note that for any sum inside of the product of Equation (60):

$$\begin{aligned} \sum_{z_u} g(z_u) \bar{H}_D^{(p-1)}(z_u) \mathbb{E}_{\mathbf{w}} \left[\cos \left(w \frac{\Gamma}{\sqrt{D}} \cdot (z_u z_v) \right) \right] \\ = \left(1 - \frac{\mathbb{E}_{\mathbf{w}}[w^2]}{2D} \sum_{z_u} g(z_u) \bar{H}_D^{(p-1)}(z_u) (\Gamma \cdot (z_u z_v))^2 + O(D^{-2}) \right), \end{aligned} \quad (65)$$

where the implicit exchange of the expectation operator and infinite series expansion of trig functions is justified by Fubini’s theorem and the assumption of the weight distribution having finite second moment, like in Section 4.1. In addition, we have used the following generalization of [7, Equation (A.23)], where for any r :

$$\sum_{z_u} g(z_u) H_D^{(r-1)}(z_u) = 1. \quad (66)$$

After taking expectations, it follows that $\sum_{z_u} g(z_u) \bar{H}_D^{(r-1)}(z_u) = 1$. By the i.i.d. assumption, Equation (65) is the same for every $u \in \mathfrak{c}(v)$, and thus

$$\begin{aligned} \bar{H}_D^{(p)}(z_v) &= \left(1 - \frac{\mathbb{E}_{\mathbf{w}}[w^2]}{2D} \sum_{z_u} g(z_u) \bar{H}_D^{(p-1)}(z_u) (\mathbf{\Gamma} \cdot (z_u z_v))^2\right)^D + O(D^{-1}) \\ &= \exp\left(-\frac{\mathbb{E}[w^2]}{2} \sum_{z_u} g(z_u) \bar{H}_D^{(p-1)}(z_u) (\mathbf{\Gamma} \cdot (z_u z_v))^2\right) + O(D^{-1}), \end{aligned}$$

where we use Equation (24). Along with continuity, the previous result implies

$$\begin{aligned} \bar{H}^{(p)}(z_v) &:= \lim_{D \rightarrow \infty} \bar{H}_D^{(p)}(z_v) \\ &= \exp\left(-\frac{\mathbb{E}_{\mathbf{w}}[w^2]}{2} \sum_{z_u} g(z_u) \bar{H}^{(p-1)}(z_u) (\mathbf{\Gamma} \cdot (z_u z_v))^2\right). \end{aligned} \quad (67)$$

The limit can then be propagated down to the lowest level of the recursion. Similarly for the term involving:

$$\lim_{D \rightarrow \infty} \sqrt{D} \mathbb{E}_{\mathbf{w}} \left[w \sin\left(w \frac{\mathbf{\Gamma}}{\sqrt{D}} \cdot (z_L z_R)\right) \right] = \mathbb{E}_{\mathbf{w}}[w^2] \mathbf{\Gamma} \cdot (z_L z_R) \quad (68)$$

Putting this altogether, for arbitrary $\{L, R\} \in E$, we have

$$\vartheta_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \frac{i\sqrt{\mathbb{E}_{\mathbf{w}}[w^2]}}{2\mu} \sum_{z_L, z_R} z_L^{[0]} z_R^{[0]} g(z_L) g(z_R) \left[\bar{H}^{(p)}(z_L) \bar{H}^{(p)}(z_R) \sqrt{\mathbb{E}_{\mathbf{w}}[w^2]} \mathbf{\Gamma} \cdot (z_L z_R) \right] \quad (69)$$

and comparing with the unweighted case [7, Equation A.26] reveals that the equality in Equation (64) holds. \square

Remark 2. Note that in the i.i.d. case the square of the denominator of Equation (6) is an unbiased estimator of $\mathbb{E}_{\mathbf{w}}[w^2]$. The i.i.d. case is the inspiration for the more general scaling rule.

Remark 3. By [7, Section 6], which extends the iteration presented earlier for unweighted regular graphs to unweighted regular k -uniform hypergraphs, Theorem 5 can be trivially extended to weighted regular k -uniform hypergraphs.

This result implies a relationship in the infinite-size limit between QAOA's objective value for weighted MaxCut and the SK model. Let

$$C^{\text{SK}} = \sum_{1 \leq u < v \leq N} \frac{J_{uv}}{\sqrt{N}} z_u z_v, \quad (70)$$

where $J_{uv} \sim \mathcal{N}(0, 1)$, and

$$V_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) := \lim_{n \rightarrow \infty} \mathbb{E}_J[\langle \boldsymbol{\gamma}, \boldsymbol{\beta} | C^{\text{SK}}/N | \boldsymbol{\gamma}, \boldsymbol{\beta} \rangle_J], \quad (71)$$

where the subscript J of the state signifies that the state was prepared by a p -layer QAOA with the SK objective as the phase operator.

Theorem 6 (Restated from [7]). *For all p and all parameters $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ the following holds*

$$V_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \nu_p(\boldsymbol{\gamma}, \boldsymbol{\beta}). \quad (72)$$

Trivially, this in combination with Theorem 5 leads to the following corollary.

Corollary 6.1. *If the edge-weight distribution has finite second moment, then for all p and all parameters $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ the following holds*

$$V_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \frac{\mu}{\sqrt{\mathbb{E}_{\mathbf{w}}[w^2]}} \vartheta_p \left(\frac{\boldsymbol{\gamma}}{\sqrt{\mathbb{E}_{\mathbf{w}}[w^2]}}, \boldsymbol{\beta} \right). \quad (73)$$

Thus the performance of QAOA on SK, MaxCut on large-girth, regular graphs and weighted MaxCut on large-girth, regular graphs are equivalent in the infinite-size limit.

Remark 4. *By [46, Theorem 3] and Remark 3, one can trivially extend Corollary 6.1 to connect QAOA's performance on weighted MaxCut on regular k -uniform hypergraphs to its performance on pure k -spin models, generalizing SK.*

5 Observations about biased SK model

As presented in Corollary 6.1, there is a deep connection between arbitrarily-weighted MaxCut and the SK model. The SK model is given in Equation (70) and has couplings $J_{uv} \sim \mathcal{N}(0, 1)$. A natural generalization to consider is a model which has couplings $J_{uv} \sim \mathcal{N}(\mu, \sigma^2)$ with arbitrary μ and σ . More generally, we can allow for the bias to be a function of the number of spins, i.e. $\mu(N)$. When $\mu(N) \neq 0$, we call this the biased SK model, and when $\mu(N) = 0$, we call it the standard SK model. Unfortunately, this natural generalization does not lead to interesting behavior. Specifically, we show that unless $\mu(N) \rightarrow 0$, the biased SK problem is trivial in the thermodynamic limit.

The performance of QAOA for arbitrary p on standard SK, specifically an iteration for the quantity V_p , was originally established in [13] using different techniques than those of Section 4.3. However, it is not clear how these techniques can be generalized to non-symmetric distributions. In this section, we use a different set of elementary techniques to determine the limiting optimal value of different versions of the biased SK model. Our goal is to find an analog to the Parisi value [47, 48] for the biased model. The following is based on Ref. [49].

For $\mathbf{z} \in \{-1, 1\}^N$, let

$$G(\mathbf{z}) = - \sum_{1 \leq i < j \leq N} J_{ij}^{(N)} z_i z_j, \quad (74)$$

where $J_{ij}^{(N)} \sim \mathcal{N}(\mu(N), \sigma^2)$. The optimization problem is $\max_{\mathbf{z}} G(\mathbf{z})$. Note that unlike the standard SK, which is symmetric around zero, here we must keep track of the signs of the couplings.

When $\mu(N) = 0$, we know that $\mathbb{E}_J[\max_{\mathbf{z}} G(\mathbf{z})] = O(N^{3/2})$. In the standard SK model, the weights are scaled by $N^{-1/2}$ to ensure that expected maximum instead grows linearly with N . This is the reasoning for the scaling of the standard SK objective presented in Equation (70). For even N , let

$$h(N) = \begin{cases} \frac{N^2}{4}, & \text{if } \mu(N) > 0, \\ -\binom{N}{2}, & \text{if } \mu(N) < 0, \\ 0, & \text{if } \mu(N) = 0, \end{cases} \quad (75)$$

which equals $\frac{\max_{\mathbf{z}} \mathbb{E}_J[G(\mathbf{z})]}{\mu(N)}$ when $\mu(N) \neq 0$. More specifically, when $\mu(N) > 0$, the problem $\max_{\mathbf{z}} \mathbb{E}_J[G(\mathbf{z})]$ reduces to MaxCut on a complete graph with all edge weights equal to $\mu(N)$, and thus the optimal cut value is $\mu(N) \frac{N^2}{4}$, i.e. set half of the $z_i = 1$. When $\mu(N) < 0$, the optimal value is obtained when all $z_i = 1$, and results in an objective function value of $\mu(N) \binom{N}{2}$. Note that when N is odd, the factor in the denominator is the same for all cases, and thus we can restrict to even N , wlog.

We have the following by the convexity of max:

$$\mu(N)h(N) = \max_{\mathbf{z}} \mathbb{E}_J[G(\mathbf{z})] \leq \mathbb{E}_J[\max_{\mathbf{z}} G(\mathbf{z})] \quad (76)$$

and

$$\begin{aligned} \mathbb{E}_J[\max_{\mathbf{z}} G(\mathbf{z})] &= \frac{1}{\alpha} \log \exp \mathbb{E}_J[\alpha \max_{\mathbf{z}} G(\mathbf{z})] \\ &\leq \frac{1}{\alpha} \log \mathbb{E}_J[\exp(\alpha \max_{\mathbf{z}} G(\mathbf{z}))] \\ &\leq \frac{1}{\alpha} \log \sum_{\mathbf{z}} \mathbb{E}_J[\exp(G(\mathbf{z}))] \\ &= \frac{1}{\alpha} \log \sum_{\mathbf{z}} \exp \left(\frac{\alpha^2 \sigma^2}{2} \binom{N}{2} - \alpha \mu(N) \sum_{i < j} z_i z_j \right) \\ &\leq \frac{1}{\alpha} \log 2^N \exp \left(\frac{\alpha^2 \sigma^2}{2} \binom{N}{2} + \alpha \mu(N) h(N) \right) \\ &= \frac{N \log(2)}{\alpha} + \frac{\alpha \sigma^2}{2} \binom{N}{2} + \mu(N) h(N). \end{aligned} \quad (77)$$

Minimizing over $\alpha > 0$, gives

$$\alpha^2 = \frac{N \log(4)}{\sigma^2 \binom{N}{2}}, \quad (78)$$

and thus

$$\mu(N)h(N) \leq \mathbb{E}_J[\max_{\mathbf{z}} G(\mathbf{z})] \leq \sigma \sqrt{\log(4) N \binom{N}{2}} + \mu(N)h(N). \quad (79)$$

Note, the right-hand side of the last inequality now involves only σ instead of σ^2 , and thus is invariant under any scaling before or after the expectation in $\mathbb{E}_J[\max_{\mathbf{z}} G(\mathbf{z})]$.

When $\mu(N) = 0$, we have $h(N) = 0$, and we recover the SK scaling, i.e.,

$$\mathbb{E}_J[\max_{\mathbf{z}} G(\mathbf{z})] \leq \sigma \sqrt{\log(4) N \binom{N}{2}} = O(N^{3/2}). \quad (80)$$

Also for standard SK with $J_{ij} \sim \mathcal{N}(0, 1)$, which is a special case of $\mu(N) = 0$, it is known that

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E}_J[\max_{\mathbf{z}} G(\mathbf{z})]}{N^{3/2}} = \Pi^*, \quad (81)$$

where Π^* is the Parisi value [49]. Thus for $J_{ij} \sim \mathcal{N}(0, \sigma^2)$, it is easily seen that

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E}_J[\max_{\mathbf{z}} G(\mathbf{z})]}{N^{3/2}} = \sigma \Pi^*. \quad (82)$$

To summarize, the limiting behavior for SK with a distribution with any $\sigma \neq 1$ can be obtained by simple rescaling. To connect this to the quantities discussed at the end of Section 4.3, for $\mathcal{N}(0, \sigma^2)$, the quantity V_p in Equation (71) scales as $V_p(\gamma, \beta) = V_{p,1}(\gamma, \beta) = V_{p,\sigma}(\frac{\gamma}{\sigma}, \beta)$. Therefore for the remainder of this Section we focus on the case where $\mu(N) \neq 0$.

If $\mu(N) \rightarrow \mu$ as $N \rightarrow \infty$ for some nonzero constant μ , the term involving $\mu(N)$ dominates, and the expected maximum is $\Theta(N^2)$. More specifically, using the definition of $h(N)$ in (75) for nonzero μ , we get that $\lim_{N \rightarrow \infty} \mu(N)h(N) = \Theta(N^2)$, and both sides of Equation (79) are of the same order in N . Thus the squeeze theorem implies that the limiting quantity is exactly

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E}_J[\max_{\mathbf{z}} G(\mathbf{z})]}{N^2} = \frac{\mu}{2(1 + \text{sign}(\mu))}, \quad (83)$$

where we define $\text{sign}(x)$ to be 1 when $x \geq 0$ and 0 when $x < 0$. This accounts for the constant factor differences in $h(N)$ for $\mu \neq 0$. The solution to the above problem is trivial in the infinite limit: set all $z_i = 1$, when $\mu < 0$, and set half of the z_i to 1 when $\mu > 0$.

However, to compensate for the maximum growing $\Theta(\sqrt{N})$ faster (than standard SK) when $\mu(N)$ is not always zero, we could consider defining the biased SK model to have $\mu(N) = O(N^{-1/2})$. Then letting $\mu = \lim_{N \rightarrow \infty} \mu(N)\sqrt{N}$ we get, from Equation (79),

$$0 \leq \lim_{N \rightarrow \infty} \frac{\mathbb{E}_J[\max_{\mathbf{z}} G(\mathbf{z})]}{N^{3/2}} - \frac{\mu}{2(1 + \text{sign}(\mu))} \leq \frac{\sigma}{\sqrt{2}} \log^{1/2}(4). \quad (84)$$

Thus the growth of the maximum is again $O(N^{3/2})$, like in the original SK. Recall that the original SK satisfies

$$0 \leq \lim_{N \rightarrow \infty} \frac{\mathbb{E}_J[\max_{\mathbf{z}} G(\mathbf{z})]}{N^{3/2}} \leq \frac{\sigma}{\sqrt{2}} \log^{1/2}(4), \quad (85)$$

with the limiting value being the Parisi value.

In this regime, the ‘‘biased SK’’ model appears non-trivial and the limiting value and solution are not obvious. We are unable to find any mention of such a model in existing literature, and we leave the study of its properties to future work.

6 Numerical results

6.1 Weighted MaxCut

Numerical investigation of the proposed parameter setting rule has been performed on a dataset of weighted graphs from Ref. [9], available through QAOAKit [50]. The dataset consists of a total of 34,701 weighted graphs with up to 20 nodes and contains both regular and non-regular graphs. The graphs have edge weights drawn i.i.d. from four different distributions, namely Uniform over $[0, 1]$ (‘‘Uniform+’’), Uniform over $[-1, 1]$ (‘‘Uniform \pm ’’), Exponential (with $\lambda = 0.2$), and Cauchy.

For the numerical study, we investigate two proposed parameter setting rules, which are variants of Equation (6):

$$(i) \quad \gamma^* = \frac{\gamma^{\text{inf}}}{\sqrt{\frac{D-1}{|E|} \sum_{\{u,v\} \in E} w_{uv}^2}}, \quad (86)$$

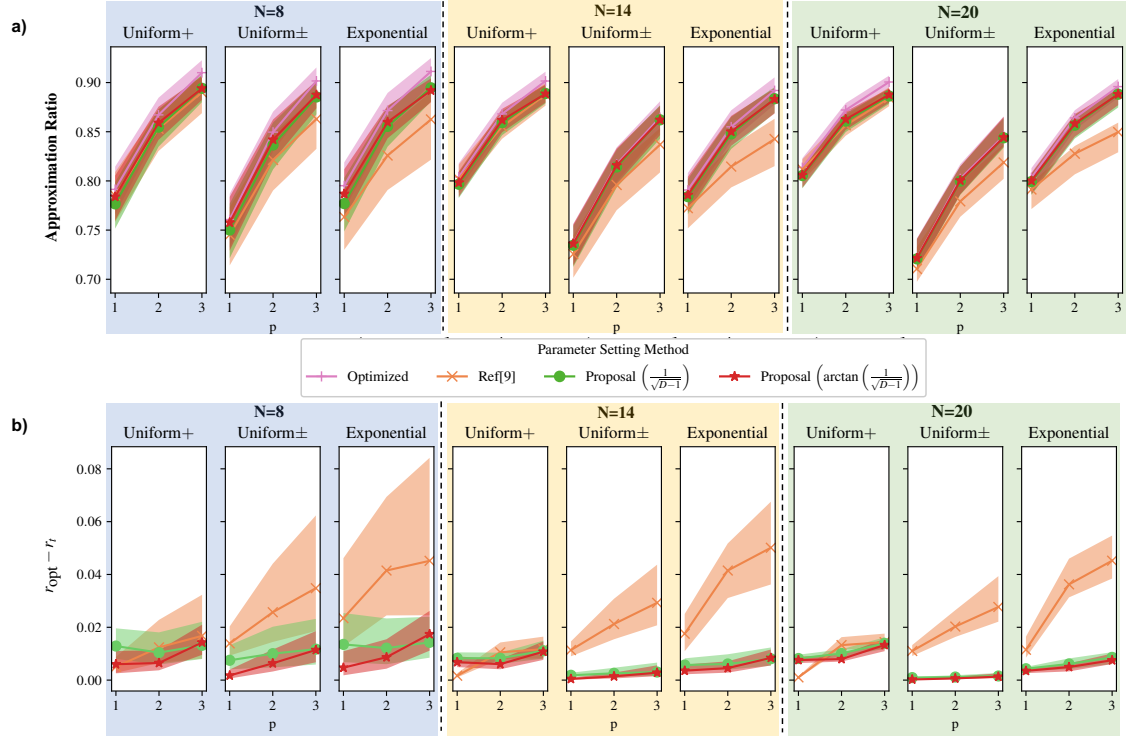


Figure 2: a) The approximation ratios obtained with directly optimized parameters, parameter setting method of Ref. [9], and parameter setting methods presented in this work. b) The gap between the approximation ratios with optimized parameters and with parameter setting methods of Ref. [9], (i) and (ii). The proposed parameter setting methods perform better when compared to the prior work, as indicated by the reduced gap from the objective obtained with the optimized parameters.

$$(ii) \quad \gamma^* = \frac{\gamma^{\text{inf}}}{\sqrt{\frac{1}{|E|} \sum_{\{u,v\} \in E} w_{uv}^2}} \arctan\left(\frac{1}{\sqrt{D-1}}\right), \quad (87)$$

where the parameters $\beta^{\text{inf}}, \gamma^{\text{inf}}$ are the optimized parameters for large-girth, regular graphs in infinite-size limit from [51, Table 4], and D is the average degree. Our baseline is the parameter scheme of Ref. [9], given by:

$$\gamma^* = \frac{\gamma^{\text{median}}}{\frac{1}{|E|} \sum_{\{u,v\} \in E} |w_{uv}|} \arctan\left(\frac{1}{\sqrt{D-1}}\right), \quad (88)$$

where γ^{median} is a median taken over optimized parameters for all 261,080 nonisomorphic connected 9-node graphs. The key difference between our scaling and that of Ref. [9] is the choice of the denominator. Since γ^{median} is close in value to γ^{inf} , the nominator is similar in both schemes.

We refer to the parameter setting procedure described in Equation (86) as method (i) and that in Equation (87) as method (ii). The first method is inspired directly by the analytical results described in Section 4. We observe that in the case of small p , better results are obtained when the formula for $p = 1$ from [37] is considered, which motivates the second rule. We note that while we have derived the exact formula for graphs with weights sampled from the exponential distribution, we do not use it in the numerical experiments. Our goal for numerics is to simulate the practical setting, wherein one does not know the distribution from which the weights are sampled.

We analyze the performance of the proposed parameter setting rules across multiple

Weight Distribution	p	$r_{opt} - r_{[9]}$	$r_{opt} - r_D$	$r_{opt} - r_{\arctan}$
Uniform+	1	0.5	1.3	0.6
	2	1.2	1.0	0.6
	3	1.6	1.3	1.4
Uniform \pm	1	1.4	0.7	0.2
	2	2.6	1.1	0.6
	3	3.5	1.3	1.1
Exponential	1	2.3	1.3	0.5
	2	4.1	1.2	0.8
	3	4.5	1.4	1.7
Cauchy	1	17.5	4.8	3.3
	2	20.8	4.1	3.9
	3	24.0	3.2	3.0

Table 1: The optimality gaps achieved by the proposed parameter setting rules and the rule in [9]. Proposed method leads to lower optimality gaps in all cases except $p = 1$ for the Uniform+ distribution.

weight distributions, values of p and values of N . Herein, we denote the median approximation ratio with directly optimized parameters by r_{opt} , with the parameter setting scheme from Ref. [9] by $r_{[9]}$ and with the two proposed methods as r_D and r_{\arctan} respectively. We refer to the difference between the approximation ratio of a given parameter setting scheme and r_{opt} as the optimality gap. The results are presented in Figure 2.

Our techniques lead to lower optimality gaps as compared to Ref. [9] in all cases except $p = 1$ with weights sampled uniformly from $[0, 1]$. We note that the gap between the methods (i) and (ii) reduces as p increases. For example, for $N = 8$ the optimality gap drops from 0.0111 on average for $p = 1$, to 0.0062 for $p = 2$, and eventually to 0.0005 for $p = 3$.

The median difference in approximation ratios for all considered p and weight distributions is 1.8 p.p. for method (i) and 1.45 p.p. for method (ii). Specifically, for the cases of exponential and Cauchy distributions, the median differences in approximation ratios from our method (i) are 1.3 p.p. and 3.8 p.p. respectively, and those from method (ii) remain a mere 1.0 p.p. and 3.3 p.p. respectively. For comparison, the previous proposal [9] obtains median differences of 3.6 p.p. and 20.7 p.p. for the weights drawn from exponential and Cauchy distributions respectively. As can be seen in Figure 3, for the case when the edge weights are drawn from a Cauchy distribution, the improvement over Ref. [9] is the largest, with an $8\times$ reduction in optimality gap at $p = 3$.

s^2	p	$r_{opt} - r_{[9]}$	$r_{opt} - r_D$	$r_{opt} - r_{\arctan}$
0.5697 [Uniform+]	1	0.00166	0.00828	0.0068
	2	0.01053	0.00827	0.00604
	3	0.01172	0.01112	0.01077
0.9557 [Exponential]	1	0.00572	0.00265	0.00357
	2	0.04149	0.00611	0.00455
	3	0.05012	0.0081	0.00843
1.1364 [Uniform±]	1	0.00183	0.00160	0.00045
	2	0.02124	0.00258	0.00142
	3	0.02928	0.00306	0.00282
2.4811 [Cauchy]	1	0.17524	0.04775	0.03344
	2	0.20788	0.04098	0.03867
	3	0.23958	0.03238	0.02967

Table 2: The median quality of solution for median variance across different distributions and for multiple p achieved through our introduced methods when compared with the previous study[9] for $n = 14$. As variance, s^2 , grows, the optimality gaps achieved by our presented techniques are almost an order of magnitude better in comparison to the method in Ref. [9].

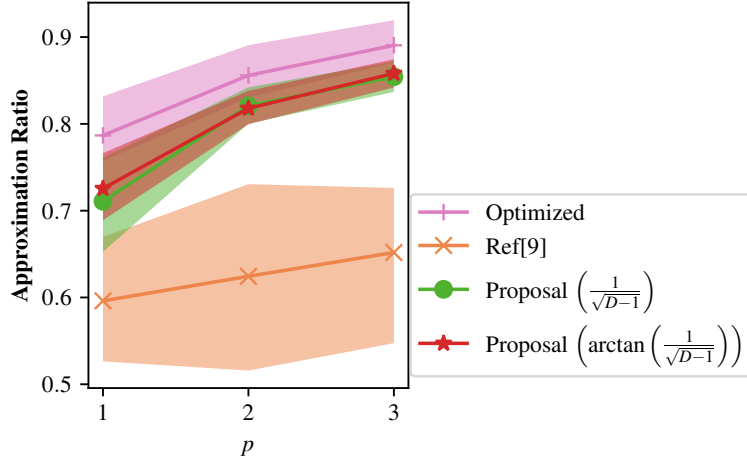


Figure 3: Approximation ratio for the graphs with edge weights drawn from a Cauchy distribution for $N = 14$. The proximity to the optimized parameter scenario, especially for large p , indicates the power of the suggested parameter setting strategies and shows a clear improvement over the earlier work. Our methods reduce the optimality gap by a factor of 8 for $p = 3$ as compared to Ref. [9].

The optimality gaps obtained by considering our presented parameters setting rules are of comparable values for each p . However, the performance of the method of Ref. [9] deteriorates as p increases.

From the values shown in Table. 2, it can be observed that when the variance is small,

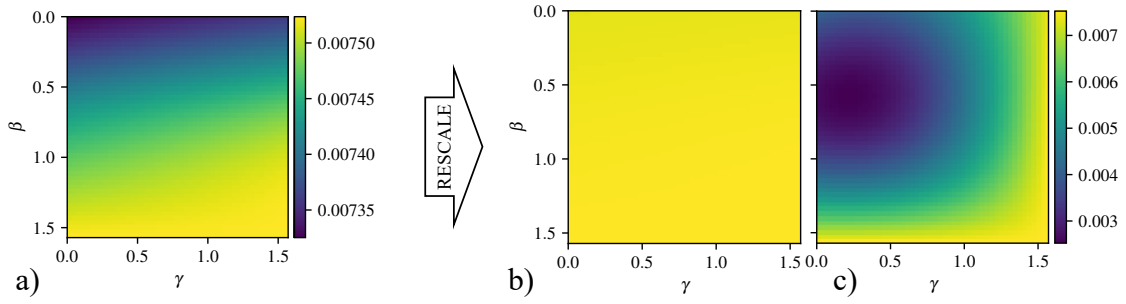


Figure 4: Energy landscape of χ_Y -QAOA applied to a random portfolio optimization problem with (a,b) original and (c) rescaled objective. Rescaling improves the geometry of the landscape. To highlight the flatness of the unrescaled landscape, (b) plots the same landscape as (a), but using the color scheme of (c).

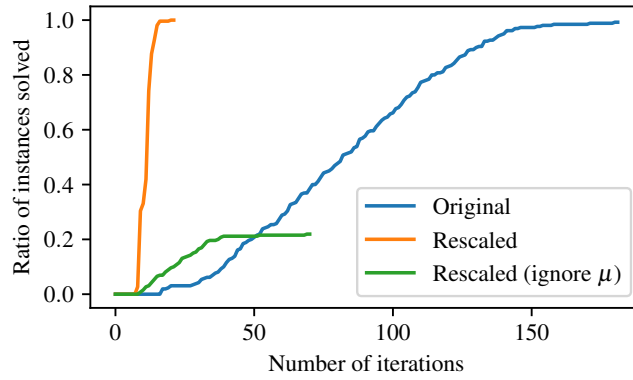


Figure 5: Performance profiles for BOBYQA run from a fixed initial point on 280 instances between 7 and 20 qubits. A line plots the ratio of instances solved with a given number of iterations. Rescaling the objective reduces the number of iterations until convergence to the target local optimum by 7.4x on average. Ignoring weights for linear terms, i.e. μ , leads to incorrect rescaling and to optimizer failing to recover the optimum.

the performance obtained using the method in the prior work [9] is comparable to that of our methods. However, as variance increases the solution qualities achieved by the methods introduced in this work beat those of the previous work [9].

6.2 Portfolio optimization

While the analytical results of Section 4 only apply to MaxCut and QAOA with transverse-field mixer $B = \sum_j x_j$, the intuition applies more broadly. To illustrate this point, in this Section we consider a portfolio optimization problem with a budget constraint, given by

$$\min_{\mathbf{x} \in \{0,1\}^N} q\mathbf{x}^\top \Sigma \mathbf{x} - \boldsymbol{\mu}^\top \mathbf{x}, \quad (89)$$

$$\text{s.t. } \sum_j x_j = k. \quad (90)$$

This problem is commonly considered as a target for QAOA [52, 11, 53, 54, 55, 56]. We use random instances generated by `RandomDataProvider` in `qiskit_finance` [57] and set $q = 0.5$ and $k = \lfloor \frac{n}{2} \rfloor$. We consider 20 instances for each number of qubits between 7 and 20, for a total of 280 instances. For each instance, we reformulate the problem in

terms of spin variables and rescale the objective following Equation (8), with the scaling coefficient computed separately for each instance. We use QAOA with $p = 1$ and the XY mixer given by $B = \frac{1}{2} \sum_{j=1}^N (X_j X_{j+1} + Y_j Y_{j+1})$ and set the initial state $|s\rangle$ to be the uniform superposition over all Hamming weight k states (Dicke state). We denote this variant XY-QAOA.

We first observe that the rescaling improves the optimization landscape. The QAOA energy at $p = 1$ is presented in Figure 4. The landscape with the original objective function (Figure 4a,b) is flat, making the optimization difficult. After rescaling (Figure 4c), a clear local minimum is visible. Moreover, the geometry of the landscape is similar with regards to both parameters. This suggests that the optimization of QAOA parameters with the rescaled objective should be easier.

To quantify this improvement, we optimize the parameters using the NLOpt [58] implementation of the BOBYQA [59] gradient-free optimizer. We use BOBYQA as it has been shown to be an effective optimizer for QAOA parameters [18]. For each instance, we run one local optimization starting from the optimal QAOA parameters for the Sherrington-Kirkpatrick model obtained from [51, Table 4] and given by $\gamma^{\text{init}} = 1$, $\beta^{\text{init}} = \frac{\pi}{4}^\ddagger$. We set the stopping criteria to be $\text{xtol} = \text{ftol} = 10^{-8}$. We present the performance profiles in Figure 5. The optimizer finds the same optimum for the original and rescaled objective in 91% of the cases. We consider that an optimizer “solves” the instance when it recovers this optimum. When the objective is rescaled, the optimizer takes on average 7.4 times fewer iterations to find the same optimum.

To contrast portfolio optimization with the weighted MaxCut problem and to highlight the importance of including the weight for the linear term in Eq. 8, we additionally plot the performance profile for results with rescaling rule where the coefficients for linear terms μ_j are ignored during the rescaling and only Σ_{ij} are used. This profile is marked “ignore μ ” in Figure 5. As can be seen from the plot, ignoring weights for linear terms leads to incorrect rescaling, with the optimizer failing to recover the optimum in the majority of cases. We note that the weights μ_j act analogously to bias in the “biased SK” model of Section 5 and can dominate the objective for some instances.

7 Discussion

In this work, we propose heuristic parameter setting rules for QAOA, inspired by a formal connection between weighted and unweighted MaxCut on regular graphs. For $p = 1$, we derive explicit expressions for the parameter γ that maximizes the cost function in the weighted case. Our analysis of MaxCut at $p = 1$ rigorously proves that the globally-optimal γ are small, providing additional justification for this commonly used assumption [7, 10, 6, 12]. For $p \geq 1$, we show explicitly how the energy landscape and, consequently, the optimal parameters scale between the weighted and unweighted cases. As we prove the concentration of the QAOA objective, our results apply with high probability to any random weighted MaxCut instance. An important limitation of our analysis is the high-girth assumption, which we inherit from Ref. [51]. Surprisingly, the simple rules we derive for MaxCut on high-girth, regular graphs apply broadly, which we demonstrate by extensive numerical experiments on MaxCut on a general class of random graphs and on random instances of a constrained portfolio optimization problem.

Additionally, we consider the biased SK problem and rigorously show that it has a

[‡]The parameters in [7, Table 4] differ from these by a factor of 2. The difference is due to the constant factors in the QAOA simulator implementation that we use.

trivial solution in the infinite-size limit, unless the mean of the weight distribution falls sufficiently fast with the number of vertices. This investigation was inspired by the connection between SK and MaxCut on regular graphs, and the observation that the closed-form iterations that we use for QAOA do not apply to complete graphs. However, it appears that, unlike standard SK, the analysis of QAOA performance is unlikely to lead to significant insights when the weights are biased.

Our observation that QAOA parameters γ have to decrease with problem size is an instantiation of a broader principle, namely that parameterized quantum circuits are not scale-independent. Similar results have been observed for quantum kernel methods [60, 61] and quantum neural network initialization [62]. A unification of these observations into a general theory of parameterized quantum circuits is a tempting prospect, though it would require the development of novel mathematical techniques.

Acknowledgements

The authors thank Kunal Marwaha for early exploratory discussions, Aram Harrow for insightful discussions on biased SK, and their colleagues at the Global Technology Applied Research center of JPMorgan Chase for support and helpful discussions. JB was supported by a grant from the Simons Foundation under Award No. 825053.

References

- [1] Michael A Nielsen and Isaac L Chuang. “Quantum computation and quantum information”. Cambridge university press. (2010).
- [2] Dylan Herman, Cody Googin, Xiaoyuan Liu, Alexey Galda, Ilya Safro, Yue Sun, Marco Pistoia, and Yuri Alexeev. “A survey of quantum computing for finance” (2022). url: <https://doi.org/10.48550/arXiv.2201.02773>.
- [3] Tad Hogg and Dmitriy Portnov. “Quantum optimization”. *Information Sciences* **128**, 181–197 (2000).
- [4] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. “A quantum approximate optimization algorithm” (2014). url: <https://doi.org/10.48550/arXiv.1411.4028>.
- [5] Stuart Hadfield, Zihui Wang, Bryan O’Gorman, Eleanor G Rieffel, Davide Venturelli, and Rupak Biswas. “From the quantum approximate optimization algorithm to a quantum alternating operator ansatz”. *Algorithms* **12**, 34 (2019). url: <https://doi.org/10.3390/a12020034>.
- [6] Sami Boulebnane and Ashley Montanaro. “Solving boolean satisfiability problems with the quantum approximate optimization algorithm” (2022). url: <https://doi.org/10.48550/arXiv.2208.06909>.
- [7] Joao Basso, Edward Farhi, Kunal Marwaha, Benjamin Villalonga, and Leo Zhou. “The quantum approximate optimization algorithm at high depth for maxcut on large-girth regular graphs and the sherrington-kirkpatrick model”. *Proceedings of the Conference on the Theory of Quantum Computation, Communication and Cryptography* **7**, 1–21 (2022).
- [8] Matthew B. Hastings. “A classical algorithm which also beats $\frac{1}{2} + \frac{2}{\pi} \frac{1}{\sqrt{d}}$ for high girth max-cut” (2021). url: <https://doi.org/10.48550/arXiv.2111.12641>.

- [9] 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).
- [10] 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). url: <https://doi.org/10.1038/s41534-023-00733-5>.
- [11] Sebastian Brandhofer, Daniel Braun, Vanessa Dehn, Gerhard Hellstern, Matthias Hüls, Yanjun Ji, Ilya Polian, Amandeep Singh Bhatia, and Thomas Wellens. “Benchmarking the performance of portfolio optimization with qaoa”. *Quantum Information Processing* **22**, 25 (2022).
- [12] Sami Boulebnane and Ashley Montanaro. “Predicting parameters for the quantum approximate optimization algorithm for max-cut from the infinite-size limit” (2021). url: <https://doi.org/10.48550/arXiv.2110.10685>.
- [13] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Leo Zhou. “The quantum approximate optimization algorithm and the Sherrington-Kirkpatrick model at infinite size”. *Quantum* **6**, 759 (2022).
- [14] Amir Dembo, Andrea Montanari, and Subhabrata Sen. “Extremal cuts of sparse random graphs”. *The Annals of Probability* **45** (2017).
- [15] Gavin E Crooks. “Performance of the quantum approximate optimization algorithm on the maximum cut problem” (2018). url: <https://doi.org/10.48550/arXiv.1811.08419>.
- [16] Michael Streif and Martin Leib. “Training the quantum approximate optimization algorithm without access to a quantum processing unit”. *Quantum Science and Technology* **5**, 034008 (2020).
- [17] 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).
- [18] Ruslan Shaydulin, Ilya Safro, and Jeffrey Larson. “Multistart methods for quantum approximate optimization”. In *IEEE High Performance Extreme Computing Conference*. Pages 1–8. (2019).
- [19] Xinwei Lee, Yoshiyuki Saito, Dongsheng Cai, and Nobuyoshi Asai. “Parameters fixing strategy for quantum approximate optimization algorithm”. *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)* (2021).
- [20] Stefan H. Sack and Maksym Serbyn. “Quantum annealing initialization of the quantum approximate optimization algorithm”. *Quantum* **5**, 491 (2021).
- [21] Ohad Amosy, Tamuz Danzig, Ely Porat, Gal Chechik, and Adi Makmal. “Iterative-free quantum approximate optimization algorithm using neural networks” (2022). url: <https://doi.org/10.48550/arXiv.2208.09888>.
- [22] Danylo Lykov, Roman Schutski, Alexey Galda, Valeri Vinokur, and Yuri Alexeev. “Tensor network quantum simulator with step-dependent parallelization”. In *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)*. Pages 582–593. (2022).
- [23] Matija Medvidović and Giuseppe Carleo. “Classical variational simulation of the quantum approximate optimization algorithm”. *npj Quantum Information* **7** (2021).

- [24] Ruslan Shaydulin and Stefan M. Wild. “Exploiting symmetry reduces the cost of training QAOA”. *IEEE Transactions on Quantum Engineering* **2**, 1–9 (2021).
- [25] Ruslan Shaydulin and Yuri Alexeev. “Evaluating quantum approximate optimization algorithm: A case study”. *Tenth International Green and Sustainable Computing Conference* (2019).
- [26] Fernando G. S. L. Brandão, 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). url: <https://doi.org/10.48550/arXiv.1812.04170>.
- [27] V. Akshay, D. Rabinovich, E. Campos, and J. Biamonte. “Parameter concentrations in quantum approximate optimization”. *Physical Review A* **104** (2021).
- [28] Phillip C. Lotshaw, Travis S. Humble, Rebekah Herrman, James Ostrowski, and George Siopsis. “Empirical performance bounds for quantum approximate optimization”. *Quantum Information Processing* **20**, 403 (2021).
- [29] 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).
- [30] 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).
- [31] Sami Khairy, Ruslan Shaydulin, Lukasz Cincio, Yuri Alexeev, and Prasanna Balaprakash. “Learning to optimize variational quantum circuits to solve combinatorial problems”. *Proceedings of the AAAI Conference on Artificial Intelligence* **34**, 2367–2375 (2020).
- [32] Guillaume Verdon, Michael Broughton, Jarrod R. McClean, Kevin J. Sung, Ryan Babbush, Zhang Jiang, Hartmut Neven, and Masoud Mohseni. “Learning to learn with quantum neural networks via classical neural networks” (2019). url: <https://doi.org/10.48550/arXiv.1907.05415>.
- [33] Sami Khairy, Ruslan Shaydulin, Lukasz Cincio, Yuri Alexeev, and Prasanna Balaprakash. “Reinforcement-learning-based variational quantum circuits optimization for combinatorial problems” (2019). url: <https://doi.org/10.48550/arXiv.1911.04574>.
- [34] Matteo M. Wauters, Emanuele Panizon, Glen B. Mbeng, and Giuseppe E. Santoro. “Reinforcement-learning-assisted quantum optimization”. *Physical Review Research* **2** (2020).
- [35] Mahabubul Alam, Abdullah Ash-Saki, and Swaroop Ghosh. “Accelerating quantum approximate optimization algorithm using machine learning”. *2020 Design, Automation & Test in Europe Conference & Exhibition (DATE)* (2020).
- [36] Jiahao Yao, Lin Lin, and Marin Bukov. “Reinforcement learning for many-body ground-state preparation inspired by counterdiabatic driving”. *Physical Review X* **11** (2021).
- [37] Zhihui Wang, Stuart Hadfield, Zhang Jiang, and Eleanor G. Rieffel. “Quantum approximate optimization algorithm for MaxCut: A fermionic view”. *Physical Review A* **97** (2018).

- [38] Jonathan Wurtz and Danylo Lykov. “The fixed angle conjecture for QAOA on regular MaxCut graphs” (2021). url: <https://doi.org/10.48550/arXiv.2107.00677>.
- [39] Stuart Hadfield. “Quantum algorithms for scientific computing and approximate optimization” (2018). url: <https://doi.org/10.48550/1805.03265>.
- [40] Paul Glasserman. “Monte carlo methods in financial engineering”. Volume 53. Springer. (2004).
- [41] Walter Rudin. “Real and complex analysis”. McGraw-Hill. (1974).
- [42] Walter Rudin. “Principles of mathematical analysis”. McGraw-hill. (1976).
- [43] Colin McDiarmid. “On the method of bounded differences”. Page 148–188. London Mathematical Society Lecture Note Series. Cambridge University Press. (1989).
- [44] Lutz Warnke. “On the method of typical bounded differences”. *Combinatorics, Probability and Computing* **25**, 269–299 (2016).
- [45] Roman Vershynin. “High-dimensional probability: An introduction with applications in data science”. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press. (2018).
- [46] 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”. *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* (2022).
- [47] G Parisi. “A sequence of approximated solutions to the s-k model for spin glasses”. *Journal of Physics A: Mathematical and General* **13**, L115 (1980).
- [48] Michel Talagrand. “The Parisi formula”. *Annals of Mathematics* (2006).
- [49] Dmitry Panchenko. “The Sherrington-Kirkpatrick model”. Springer Science & Business Media. (2013).
- [50] Ruslan Shaydulin, Kunal Marwaha, Jonathan Wurtz, and Phillip C Lotshaw. “QAOAKit: A toolkit for reproducible study, application, and verification of QAOA”. *Second International Workshop on Quantum Computing Software* (2021).
- [51] Joao Basso, Edward Farhi, Kunal Marwaha, Benjamin Villalonga, and Leo Zhou. “The quantum approximate optimization algorithm at high depth for maxcut on large-girth regular graphs and the sherrington-kirkpatrick model” (2021). url: <https://doi.org/10.48550/arXiv.2110.14206>.
- [52] Dylan Herman, Ruslan Shaydulin, Yue Sun, Shouvanik Chakrabarti, Shaohan Hu, Pierre Minssen, Arthur Rattew, Romina Yalovetzky, and Marco Pistoia. “Constrained optimization via quantum zeno dynamics”. *Communications Physics* **6**, 219 (2023).
- [53] N. Slate, E. Matwiejew, S. Marsh, and J. B. Wang. “Quantum walk-based portfolio optimisation”. *Quantum* **5**, 513 (2021).
- [54] Mark Hodson, Brendan Ruck, Hugh Ong, David Garvin, and Stefan Dulman. “Portfolio rebalancing experiments using the quantum alternating operator ansatz” (2019). url: <https://doi.org/10.48550/arXiv.1911.05296>.
- [55] Tianyi Hao, Ruslan Shaydulin, Marco Pistoia, and Jeffrey Larson. “Exploiting in-constraint energy in constrained variational quantum optimization”. *2022 IEEE/ACM Third International Workshop on Quantum Computing Software (QCS)* (2022).

- [56] Zichang He, Ruslan Shaydulin, Shouvanik Chakrabarti, Dylan Herman, Changhao Li, Yue Sun, and Marco Pistoia. “Alignment between initial state and mixer improves qaoa performance for constrained optimization”. *npj Quantum Information* **9**, 121 (2023).
- [57] “Qiskit finance”. <https://qiskit.org/documentation/finance/>.
- [58] Steven G. Johnson. “The NLOpt nonlinear-optimization package” (2022). <http://github.com/stevengj/nlopt>.
- [59] Michael JD Powell. “The BOBYQA algorithm for bound constrained optimization without derivatives”. Cambridge NA Report NA2009/06**26** (2009).
- [60] Ruslan Shaydulin and Stefan M. Wild. “Importance of kernel bandwidth in quantum machine learning”. *Physical Review A* **106** (2022).
- [61] Abdulkadir Canatar, Evan Peters, Cengiz Pehlevan, Stefan M. Wild, and Ruslan Shaydulin. “Bandwidth enables generalization in quantum kernel models” (2022). url: <https://doi.org/10.48550/arXiv.2206.06686>.
- [62] Kaining Zhang, Liu Liu, Min-Hsiu Hsieh, and Dacheng Tao. “Escaping from the barren plateau via gaussian initializations in deep variational quantum circuits”. In *Advances in Neural Information Processing Systems*. Volume 35, pages 18612–18627. Curran Associates, Inc. (2022).

Disclaimer

This paper was prepared for informational purposes with contributions from the Global Technology Applied Research center of JPMorgan Chase & Co. This paper is not a product of the Research Department of JPMorgan Chase & Co. or its affiliates. Neither JPMorgan Chase & Co. nor any of its affiliates makes any explicit or implied representation or warranty and none of them accept any liability in connection with this paper, including, without limitation, with respect to the completeness, accuracy, or reliability of the information contained herein and the potential legal, compliance, tax, or accounting effects thereof. This document is not intended as investment research or investment advice, or as a recommendation, offer, or solicitation for the purchase or sale of any security, financial instrument, financial product or service, or to be used in any way for evaluating the merits of participating in any transaction.