The Quantum Approximation Optimization Algorithm for MaxCut: A Fermionic View

Zhihui Wang,[1,2] Stuart Hadfield,[3] Zhang Jiang,[1,4] and Eleanor G. Rieffel[1]

[1] *Quantum Artificial Intelligence Laboratory (QuAIL), NASA Ames Research Center, California 94035*

[2] *Universities Space Research Association, Columbia 21046*

[3] *Columbia University, New York 10027*

[4] *Stinger Ghaffarian Technologies Inc., Maryland 20770*

(Dated: March 16, 2017)

Farhi *et al.* recently proposed a class of quantum algorithms, the Quantum Approximate Optimization Algorithm (QAOA), for approximately solving combinatorial optimization problems. A level-p QAOA circuit consists of $2p$ steps; in each step a classical Hamiltonian, derived from the cost function, is applied followed by a mixing Hamiltonian. The $2p$ times for which these two Hamiltonians are applied are the parameters of the algorithm. As p increases, however, parameter optimization becomes inefficient due to the curse of dimensionality. The success of the QAOA approach will depend, in part, on finding effective parameter-setting strategies. Here, we analytically and numerically study parameter setting for QAOA applied to MaxCut . For level-1 QAOA, we derive an analytical expression for a general graph. In principle, expressions for higher p could be derived, but the number of terms quickly becomes prohibitive. For a special case of MaxCut , the Ring of Disagrees, or the 1D antiferromagnetic ring, we provide an analysis for arbitrarily high level. Using a fermionic representation, the evolution of the system under QAOA translates into quantum control of an ensemble of independent spins. This treatment enables us to obtain analytical expressions for the performance of QAOA for any p, and give prove to a previous conjecture on the best performance. It also greatly simplifies numerical search for the optimal values of the parameters. By exploring symmetries, we identify a lower-dimensional sub-manifold of interest; the search effort can be accordingly reduced. This analysis also explains an observed symmetry in the optimal parameter values. Further, we numerically investigate the parameter landscape and show that it is a simple one in the sense of having no local optima.

PACS numbers:

I. INTRODUCTION

Recently, Farhi *et al.* [1] proposed a new class of quantum algorithm, the Quantum Approximate Optimization Algorithm (QAOA)to tackle challenging approximate optimization problems on a gate model quantum computer. In QAOA, the problem Hamiltonian which encodes the cost function of the optimization problem, and a mixing Hamiltonian are applied alternately. A handful of recent papers suggest the power of such circuits [2–5]. Once the problem and mixing Hamiltonians have been chosen, the parameters of the algorihtm are the times for which each Hamiltonian is applied at each stage. With an optimized time sequence for each piece, the optimal output of the problem Hamiltonian is approximated.

The success of QAOA relies on the properly chosen time sequence. For a fixed level QAOA, straight-forward sampling of search space was proposed [1], but it is practical only for small p; as the level increases the parameter optimization becomes inefficient due to the curse of dimensionality. Elegant analytical tools designed for specific problem class can provide parameter values for $p \gg 1$ that give excellent performance, e.g., search an unstructured database [5], but for a general problem, practical search strategies are needed. Here, we analystically and numerically study the parameter setting problem, with a focus on the MaxCut problem. We demonstrate how analyzing parameter symmetries and the landscape of the expectation value over the space of the parameter values can aid in finding optimal parameter values.

In Ref. [1], Farhi *et al.* investigated MaxCut for specific (triangle-free) graphs, and provided numerical results for a special case, the *ring of disagrees*, the one-dimensional chain of spin-1/2's with nearest-neighbored antiferromagnetic couplings. We first extend the results of MaxCut in Ref. [1] to derive analytical expressions which can be solved to obtain the optimal parameters for level-one QAOA for MaxCut arbitrary graphs. Direct analysis through operator reduction quickly becomes cumbersome as the level p of the algorithm increases. We then focus on the ring of disagrees where we are able to advance the analysis to arbitrary levels.

Using a Fermionic representation, we show that the evolution of the system under QAOA translates into quantum optimal control of an ensemble of independent spins, significantly simplifying the analysis. In the new representation, the analytical expression for the expectation value as a trigonometric polynomial of the parameters can be efficiently derived for arbitrary level p. Furthermore, the reduction to independent spins simplifies the numerical search greatly because evaluation involves only $\sim 2p$ matrix multiplication of 2-by-2 matrices and is linear in problem size, the number of spins in the original problem, N. Further, by exploring symmetries, we

identify a lower-dimensional sub-manifold whose critical points are also critical points in the full manifold. We numerically confirm that all optimal parameters lie in this sub-manifold. The search effort can be accordingly reduced, and it also explains an observed symmetry in the optimal parameter values. Finally, a numerical investigation of the parameter landscape shows that it is a simple one in the sense of having no local optima.

II. RECAP OF THE ALGORITHM

Given an objective function $C(\mathbf{x})$ in the range $[C_{\min}, C_{\max}]$ to minimize, QAOA aims to find a bit string $\mathbf{x}$ whose cost is r-approximately optimal in the sense that the *approximation ratio*

$$r \equiv \frac{C_{\max} - C(\mathbf{x})}{C_{\max} - C_{\min}} \geqslant r^* \tag{1}$$

where the constant $r^* \in [0, 1]$ is the desired approximation ratio.

A Hamiltonian H_C can be constructed accordingly. Since it comes from a classical cost function, H_C is diagonal in the σ^z basis, taking the form

$$H_C \;=\; \sum_{\mathcal{C}\subset\{1,\dots N\}} \alpha_{\mathcal{C}} \bigotimes_{j\in\mathcal{C}} \sigma_j^z \;, \tag{2}$$

where σ^z is the Pauli operator, $\mathcal{C}$ is a subset of all spins, and $\alpha_{\mathcal{C}}$ is a real coefficient for the many-body coupling between spins in the subset $\mathcal{C}$.

A level-p QAOA circuit consists of p steps, in each step H_C and a mixing Hamiltonian H_B are applied sequentially. The mixing operator is often chosen to be $H_B = \sum_j \sigma_j^x$, which is also the setup in this study. For many problems, alternative mixing Hamiltonians that incorporate some problem constraints can reduce resource requirements and improve performance over the standard setup [6].

Once the mixing Hamiltonian and the classical Hamiltonian have been chosen, the parameters of a p-level QAOA circuit are the $2p$ real numbers (γ_i, β_i), for $1 \leqslant i \leqslant p$, which determine how long each operator is applied in iteration i:

$$U_C(\gamma_i) \;=\; \exp[-i\gamma_i H_C] \tag{3}$$
$$U_B(\beta_i) \;=\; \exp[-i\beta_i H_B]. \tag{4}$$

The initial state is prepared as the ground state of $-H_B$. The density matrix reads

$$\rho_0 = \bigotimes_j \frac{1}{2}(\mathbb{1} + \sigma_j^x) \tag{5}$$

The circuit

$$U = U_B(\beta_p)U_C(\gamma_p)\cdots U_B(\beta_2)U_C(\gamma_2)U_B(\beta_1)U_C(\gamma_1) \tag{6}$$

applied to the initial state creates a final state in which the expectation value of H_C is

$$F(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \mathrm{Tr}[H_C U \rho_0 U^\dagger]. \tag{7}$$

With the optimal $F^* = \min_{\boldsymbol{\gamma},\boldsymbol{\beta}} F(\boldsymbol{\gamma}, \boldsymbol{\beta})$, the approximation ratio for QAOA is a generalization of Eq. (1) to the final quantum state as

$$r = \frac{F^* - C_{\max}}{C_{\min} - C_{\max}} \tag{8}$$

The goal of the circuit U is to drive the system into a quantum state which, upon measuring in the computational basis, yields with high probability a classical bit string that is r^*-approximately optimal. Equivalently, one requires the expectation value F^* in the final state is of r^*-approximately optimal, and the distribution of bit strings from measuring this state in the computational basis is concentrated on bit strings with costs close to this expectation value.

Upon the proposal of QAOA, Farhi *et al.* [2] were able to briefly beat the approximation for the best classical algorithm for the optimization problem of E3Ln2 with a level-1 QAOA circuit until their algorithm inspired a better classical one [7] that improved mildly upon the level-1 QAOA approach. QAOA circuits have also been applied for exact optimization [8] and sampling [3] as well as approximate optimization. Harrow and Farhi [3] proved that, under reasonable complexity theoretic assumptions, the output distribution of even level-1 QAOA cannot be efficiently sampled classically. QAOA circuits are therefore among the most promising candidates for early demonstrations of "quantum supremacy" [9, 10]. It remains an open question whether QAOA circuits provide a quantum advantage for approximate optimization.

QAOA has close connection with the Variational Quantum Algorithm (VQA), classical optimization of parameters a quantum evolution is performed. The result of evaluation of the final state is fed back to the parameter optimization, forming a closed-loop learning process. Yang *et al.* [4] proved that for evolution under a Hamiltonian that is the weighted sum of Hamiltonian terms, with the weights allowed to vary in time, the optimal control is bang-bang, i.e. constant magnitude, of either the maximum or minimum allowed weight, for each of the terms in the Hamiltonian at any given time. Their work implies that QAOA circuits with the right parameters are optimal among Hamiltonians of the form $H(s) = \big(1 - f(s)\big)H_B + f(s)H_C$, where $f(s)$ is a real function in the range $[0, 1]$.

The ultimate success of the QAOA approach will depend on finding effective parameter-setting strategies. Farhi *et al.* [1] show that, for fixed p, the optimal parameters can be computed in time polynomial in the number of qubits N. With increasing p, however, exhaustive search of the QAOA parameters becomes inefficient due to the curse of dimensionality. If we discretize so that each parameter can take on m values, exhaustive search of the

optimum takes exponential steps in p as m^{2p}. Here, we analytically and numerically study parameter setting for QAOA applied to MaxCut .

III. LEVEL-1 QAOA FOR MAXCUT

In this section, we derive an analytical expression for F for level-1 QAOA for MaxCut on general graph, furthering the analysis in [1]. In principle, we could similarly derive expressions for higher p, but the workload quickly becomes prohibitive.

MaxCut Problem: Given a graph $G(V, E)$ with $n = |V|$ vertices and $m = |E|$ edges. The objective is to partition the graph vertices into two sets such that the number of edges connecting vertices in different sets is maximized.

The objective function for MaxCut can be represented by the Hamiltonian

$$H_C = -\sum_{\langle uv\rangle\in E} C_{uv}, \qquad C_{uv} = \frac{1}{2}(I - \sigma_u^z\sigma_v^z) \tag{9}$$

The expectation value of QAOA decomposes as

$$F(\boldsymbol{\gamma}, \boldsymbol{\beta}) \;=\; -\sum_{\langle uv\rangle\in E} \langle C_{uv}\rangle \tag{10}$$

where $\langle C_{uv}\rangle := \mathrm{Tr}[C_{uv}U\rho_0U^\dagger]$.

Theorem 1. *For QAOA with $p = 1$, for each edge $\langle uv\rangle$,*

$$\begin{aligned}\langle C_{uv}\rangle(d, e, f) \;=\; &\frac{1}{2} + \frac{1}{4}\sin 4\beta \sin\gamma(\cos^d\gamma + cos^e\gamma)\\ &-\frac{1}{4}\sin^2\beta\cos^{d+e-2f}\gamma(1-\cos^f 2\gamma)\end{aligned} \tag{11}$$

where $d + 1$ and $e + 1$ are the degrees of vertices u and v, respectively, and f is the number of triangles in the graph containing edge $\langle uv\rangle$.

See Appendix. A for a proof. Here we showed that for $p = 1$ the expectation value of any edge $\langle C_{uv}\rangle$ depends only on the parameters (d, e, f). Then the overall expectation value $F(\gamma, \beta) = -\sum_{(d,e,f)}\langle C_{uv}\rangle(d, e, f)\chi(d, e, f)$ where the summation is taken over distinct subgraphs (d, e, f) and $\chi(d, e, f)$ is the multiplicity of the subgraph (d, e, f), i.e. the number of times the subgraph appears in G. Thus for an arbitrary graph $F(\gamma, \beta)$ may be efficiently computed classically.

For a triangle-free graph of fixed vertex degree $d + 1$, i.e., $f = 0$ in Eq. (11). The expectation value reduces to

$$F(\gamma, \beta) = -\frac{|E|}{2}\big(1 + \sin 4\beta\sin\gamma\cos^d\gamma\big). \tag{12}$$

with optimal value

$$F^* = -\frac{|E|}{2}\Big(1 + \frac{1}{\sqrt{d}}\Big(\frac{d}{D}\Big)^{D/2}\Big) \tag{13}$$

For any such graph, one optimal pair of angles is $(\beta, \gamma) = (\pi/8, \arctan(1/\sqrt{d}))$. MaxCut in the case in which all vertices have degree 2, so the graph is a ring, is called the ring of disagrees. In this case, minimizing (12) yields the approximation ratio 0.75 at $(\beta, \gamma) = (\pi/8, \pi/4)$, and for $d = 2$, the ratio is 0.692, both reproducing the results in [1].

For an arbitrary triangle-free graph with maximum vertex degree D, the right-hand side of Eq. (13) gives a upper bound to F^*. We see that even for $p = 1$, QAOA always beats random guessing.

It is straightforward to extend the analysis to QAOA of higher levels. The number of terms quickly becomes prohibitive for direct calculation, however; many more non-commuting terms coming from the U_C's and U_B's must be retained and carried through the calculation. The expectation value of a given edge will also depend on its local graph topology (within p hops), which becomes difficult to succinctly characterize as p increases.

IV. ANALYSIS OF THE PROBLEM OF RING OF DISAGREES (ANTI-FERROMAGNETIC CHAIN)

We now study in detail QAOA for the ring of disagrees, where the symmetry and simplicity of the problem means that analysis can be done for QAOA of arbitrary level p. Numerical results for small p, and a conjecture for the approximation ratio for arbitrary p were given in Ref. [1].

A. Formulation of the problem

The Hamiltonian for MaxCut on a ring, or the Ring of Disagrees, with N vertices is $\tilde{H}_C = -\frac{1}{2}\sum_{j=1}^N(1-\sigma_j^z\sigma_{j+1}^z)$ where $\sigma_{N+1}^z = \sigma_1^z$. For convenience, we consider only even N, in which case the ground state of $\tilde{H}_C$ is trivial with every pair of neighboring spins aligned in anti-parallel fashion, corresponding to $C_{\max} = -N$. The maximally excited state is also trival with all spins aligned in parallel, corresponding to $C_{\min} = 0$. The optimization ratio is then $r = -F^*/N$.

To simplify the derivation, and without losing generality, we drop the constant and rescale $\tilde{H}_C$ to be

$$H_C = \sum_j \sigma_j^z\sigma_{j+1}^z \tag{14}$$

which is used in the evolution operator Eq. (3). The initial state of the system is prepared as Eq. (5), and the algorithm follows the evolution Eq. (6).

The relation between angles, expectation value used in this paper and the ones in Ref. [1] (notations with tilde) is $\gamma = -\tilde{\gamma}/2$, $\beta = \tilde{\beta}$ and $\tilde{F}(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\beta}}) = \big(N - F(\boldsymbol{\gamma}, \boldsymbol{\beta})\big)/2$

B. Fermionic representation

We show that using a fermionic representation, the parameter setting of QAOA reduces to finding the optimal quantum control of an ensemble of independent spins (spin-1/2), significantly simplifying the analysis.

Since spin operators do not obey canonical commutation relations, transforming them into bosonic or fermionic opearators are commonly used techniques. Such transformation allows for the application of standard techniques in condensed matter physics like diagrammatic perturbation. The algebra of the original spin operators need to be preserved in the mappings. The Jordanj-Wigner transformation [11, 12] maps the spin operators into fermions with a long-ranged phase factor.

We apply the Jordan-Wigner transformation [11, 12],

$$a_j = S_j^- e^{-i\phi_j} \tag{15}$$
$$a_j^\dagger = S_j^+ e^{i\phi_j} \tag{16}$$

where $S_j^+ = (\sigma_j^y + i\sigma_j^z)/2$, $S_j^- = (\sigma_j^y - i\sigma_j^z)/2$, and the phase factor $\phi_j = \pi \sum_{j'<j} (\sigma_j^x + 1)/2$ is long-ranged involving all operators for $j' < j$. The new operators a_j, $a_j^\dagger$ can be verified to obey the fermion anticommutation relations, $\{a_j, a_{j'}^\dagger\} = a_j a_{j'}^\dagger + a_{j'}^\dagger a_j = \delta_{j,j'}$, and $\{a_j, a_{j'}\} = \{a_j^\dagger, a_{j'}^\dagger\} = 0$. The inverse transformation reads

$$S_j^+ = a_j e^{i\phi_j} \tag{17}$$
$$S_j^- = a_j^\dagger e^{-i\phi_j} \tag{18}$$
$$\sigma_j^x = 2a_j^\dagger a_j - 1 \tag{19}$$

and the phase factor in the fermionic representation is $\phi_j = \pi \sum_{j'<j} a_j^\dagger a_j$. The Jordan-Wigner transformation is a convenient tool for one-dimentional spin systems, particularly for nearest-neighbored couplings because in product of the neighboring spin operators like $S_j^+ S_{j+1}^-$, the phase factors drop out, leaving a concise expression with short-ranged coupling.

Apply the transformation to our problem, we get

$$H_B = \sum_{j=1}^{N} \left(2a_j^\dagger a_j - 1\right) \tag{20}$$
$$H_C = \sum_{j=1}^{N-1} a_j^\dagger a_{j+1} + a_j a_{j+1} - (a_N^\dagger a_1 + a_N a_1) G + \text{h.c.} \tag{21}$$

where we introduce the gauge operator $G = \exp[i\pi \sum_{l=1}^{N} a_l^\dagger a_l] = (-1)^N \prod_{j=1}^{N} \sigma_j^x$, which is a necessary treatment for periodic boundary condition. In the current (standard) QAOA settings, for even N, the initial state is an eigenstate of G with eigenvalue 1. The operator G is a constant of motion since it commutes with both H_B and H_C, so the value of G remains 1 throughout the evolution. The sign of the $j = N$ terms in H_C therefore are different from the others. We further introduce a phase factor to unify the expression, $b_j = a_j e^{-ij\pi/N}$. The Hamiltonians then read

$$H_B = 2\sum_{j=1}^{N} b_j^\dagger b_j - N \tag{22}$$
$$H_C = e^{i\pi/N} \sum_{j=1}^{N} \left(b_j^\dagger b_{j+1} + e^{2ij\pi/N} b_j b_{j+1}\right) + \text{h.c.} \tag{23}$$

Upon applying a Fourier transformation to b_j,

$$c_j = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} e^{\omega jk} b_k\,, \quad \omega = 2i\pi/N\,, \tag{24}$$

the driver and the problem Hamiltonians in the momentum space take the form

$$H_B = 2\sum_{k=1}^{N} c_k^\dagger c_k - N \tag{25}$$
$$H_C = 2\sum_{k=1}^{N/2} \cos\theta_k \left(c_k^\dagger c_k + c_{-k}^\dagger c_{-k}\right) + \sin\theta_k \left(c_{-k} c_k + c_k^\dagger c_{-k}^\dagger\right), \tag{26}$$

where $\theta_k = (2k-1)\pi/N$ and $c_{-k} \equiv c_{N-k+1}$. Since c_k and $c_k^\dagger$ are only coupled to c_{-k} and $c_{-k}^\dagger$, we only need to solve a set of 2-fermion problems. Because both H_B and H_C preserve the parity of the fermionic excitations, we need to consider only the ground state and the double excited state of the two fermions. For each k, in this two-dimensional subspace the driver and the problem Hamiltonian become $2\sigma^z$ and $2\sigma^z \cos\theta_k + 2\sigma^x \sin\theta_k$, respectively.

In summary, after transforming the problem to a fermionic representation, the original many-body Hamiltonian of a ring of N spins reduces to an ensemble of $N/2$ non-interacting spins,

$$H_B = \sum_{k=1}^{N/2} H_{B,k}$$
$$H_C = \sum_{k=1}^{N/2} H_{C,k} \tag{27}$$

each taking the form

$$H_{B,k} = 2\sigma^z \tag{28}$$
$$H_{C,k} = 2\left(\cos\theta_k \sigma^z + \sin\theta_k \sigma^x\right) = 2\hat{k}\cdot\hat{\sigma} \tag{29}$$

where $k = 1, 2, \ldots, N/2$, and the unit vector $\hat{k} = (\sin\theta_k, 0, \cos\theta_k)$

The initial state for each spin is the ground state of $-H_{B,k}$, i.e., $\rho_0 = |1\rangle\langle 1| = (\mathbb{1} + \sigma^z)/2$ and the optimization

reduces to minimize

$$F(\boldsymbol{\gamma},\boldsymbol{\beta}) \;=\; \sum_{k=1}^{N/2} F_k(\boldsymbol{\gamma},\boldsymbol{\beta}) \tag{30}$$

where

$$F_k(\boldsymbol{\gamma},\boldsymbol{\beta}) \;=\; \frac{1}{2}\big[H_{C,k} U_k \sigma^z U_k^\dagger\big] \tag{31}$$

$$\;=\; \mathrm{Tr}\left[\hat{k}\cdot\hat{\sigma} U_k \sigma^z U_k^\dagger\right] \tag{32}$$

Hereafter, for notation simplicity, we drop the subscript for U_k and use U to refer to the evolution operator for the single spin. $U = U_B(\beta_p)U_C(\gamma_p)\cdots U_B(\beta_1)U_C(\gamma_1)$ now consists of only single-spin operators

$$U_B(\beta_l) \;=\; \exp[-i2\beta_l\sigma^z] \tag{33}$$

$$U_C(\gamma_l) \;=\; \exp[-i2\gamma_l\hat{k}\cdot\hat{\sigma}] \tag{34}$$

for $l = 1, 2, \ldots, p$.

C. Analysis of the approximation ratio

For sufficiently large problem size, the approximation ratio of QAOA on the problem of ring of disagrees is independent of the problem size. This property has been shown in Ref. [1] based on argument on the operator reductions. The specific value of the approximation ratio for level-p QAOA was conjectured to be $(2p+1)/(2p+2)$ therein. In this section we give an analytical proof of the approximation ratio based on the spin ensemble picture.

1. *Size-dependence of the approximation ratio*

First we show that the size-independence feature comes naturally out of the picture of single spin rotations. Each $U_C(\gamma) = \cos(2\gamma) - i\sin(2\gamma)k\cdot\hat{\sigma}$ would contribute a trigonometric function of θ_k, therefore F_k takes a form

$$F_k \;=\; \sum_{l=0}^{2p+1} f_l(\boldsymbol{\gamma},\boldsymbol{\beta})\sin^l\theta_k + g_l(\boldsymbol{\gamma},\boldsymbol{\beta})\cos^l\theta_k \tag{35}$$

where f_l, g_l are coefficients independent of θ_k. Since each $\sin\theta_k$ accompanies one σ^x, noticing the properties of Pauli matrices, $\mathrm{Tr}[\sigma_\alpha\sigma_{\alpha'}] = 2\delta_{\alpha,\alpha'}$, the coefficient $f_l(\boldsymbol{\gamma},\boldsymbol{\beta})$ is zero for odd l. Furthermore, when we consider F,

$$F = \sum_{l=0}^{2p+1}\left(f_l(\boldsymbol{\gamma},\boldsymbol{\beta})\sum_{k=1}^{N/2}\sin^l\theta_k + g_l(\boldsymbol{\gamma},\boldsymbol{\beta})\sum_{k=1}^{N/2}\cos^l\theta_k\right), \tag{36}$$

recall that $\theta_k = (2k-1)\pi/N$, we have $\sum_{k=1}^{N/2}\cos^l\theta_k = 0$ for odd l. Therefore we can consider only terms of even l in Eq. (35), equivalent to a trigonometric polynomial of $2\theta_k$ of degree p,

$$F_k \;=\; \sum_{l=0}^{p} d_{2l}(\boldsymbol{\gamma},\boldsymbol{\beta})\cos(2l\theta_k)\,, \tag{37}$$

where $d_{2l}(\boldsymbol{\gamma},\boldsymbol{\beta})$ is a coefficient independent of k. See the analysis for $p = 1$ and 2 in Sec. IV E for example.

Eq. (37) takes the form of the Fourier transformation of series d_{2l} with a cutoff at order p. Therefore for $N \geqslant 2p+2$, the component $d_0 = \sum_k F_k/N$, in other words, F is the DC component of F_k:

$$F \;=\; \frac{N}{2}\cdot d_0(\boldsymbol{\gamma},\boldsymbol{\beta}) \tag{38}$$

Since the N-dependence of F_k lies in θ_k and d_0 is θ_k-independent, the expectation value F and furthermore the approximation ratio of QAOA is independent of N.

Note that d_0 here is an abstract function whose dependence on $(\boldsymbol{\gamma},\boldsymbol{\beta})$ is not provided for general level p. For an arbitrary level p, simplifying Eq. (32) to get the specific trigonometrical function form can be easily done.

2. *The approximation ratio*

There are $N/2$ non-interacting spins to be controlled through the unitary evolutions dictated by $2p$ rotation angles $(\boldsymbol{\gamma},\boldsymbol{\beta})$; each spin k is initialized along $\hat{k}$ and (ideally) driven to be line with $-\hat{z}$. The rotation axes are constrained to be $\hat{z}$ and $\hat{k}$ alternately.

For $N \leqslant 2p$, the number of rotations is equal or more than twice the number of the pseudo-spins. As for each spin two rotations are sufficent to lign up the spin to $\hat{z}$, $2p$ angles are sufficient to align $N \leqslant 2p$ spins, hence the approximation ratio is 1.

For $N \geqslant 2p+2$, as F is size-independent, we will take $N = 2p+2$ to show that the approximation ratio is $(2p+1)/(2p+2)$.

From Eq. (36) one can see that the optimal solution should be independent on specific choice of angles θ_k as long as they are uniformly spaced in $[0,\pi]$. Without losing generality instead of consider $\theta_k = (2k-1)\pi/N$ we consider $\phi_{k'} = 2k'\pi/N$ with $k' = 0, 2, \ldots, N/2$. Note the count of k' is one more than the count of k. By further assigning the two end spins $k' = 0$ and $k' = N/2$ each a weight $1/2$ compared to the rest, one can verify that

$$\sum_{k=1}^{N/2}\sin^2\theta_k \;=\; \sum_{k'=0}^{N/2}\sin^2\phi_{k'}\,,$$

$$\sum_{k=1}^{N/2}\cos^2\theta_k \;=\; \sum_{k'=0}^{N/2}\cos^2\phi_{k'}\,, \tag{39}$$

Hence F as in Eq. (36) is maintained the same. For $k' = 0$ and $k' = N/2$, the spin vector is aligned with $\hat{z}$ and $-\hat{z}$, respectively, and are both invariant under U for

any rotation angles. The corresponding $F_{k'}$, weighted, is $-1/2$ and $1/2$, respectively. We can then deploy the $2p$ angles, which are sufficient to align the other p spins with $\hat{z}$. This immediately gives the total F^*

$$\frac{F^*}{N} = \frac{-1/2 + p + 1/2}{N/2} = \frac{-p}{p+1}$$

and the approximation ratio

$$\text{approx. ratio} \;=\; \frac{1 - F^*/N}{2} = \frac{2p+1}{2p+2}$$

D. Symmetry and criticality-constrained manifolds

In this section, we show that, based on symmetries in the spin rotations, we can identify sub-manifolds in the search space that admits extrema. In later sections, we provide numerical evidence that the global minima always lie in these sub-manifolds. It is then suggestive that one can focus the search within the identified sub-manifolds and thus reduce the search effort.

1. *Physics: rotations of the Bloch vectors*

For each spin, Eq. (32) can be expressed as

$$F_k(\boldsymbol{\gamma}, \boldsymbol{\beta}) \;=\; 4\mathcal{F}_k - 2 \tag{40}$$

where

$$\mathcal{F}_k \;\equiv\; \mathrm{Tr}[\rho_{\hat{k}} U \rho_z U^\dagger] \tag{41}$$

and $\rho_{\hat{k}} = \frac{1}{2}(1 + \hat{k}\cdot\hat{\sigma})$ and $\rho_z = \frac{1}{2}(1+\sigma^z)$.

On the Bloch sphere, $\rho_{\hat{k}}$ and $\rho_{\hat{z}}$ can be interpreted as the density matrices for the Bloch vector along the $\hat{k}$-direction, and that along the $\hat{z}$, respectively. Equation. (41) represents a single spin-$\frac{1}{2}$, initialized along $\hat{z}$-direction, rotate about $\hat{k}$-axis for angle $4\gamma_1$, rotate about $\hat{z}$ for $4\beta_1$, ..., rotate about $\hat{k}$ for $4\gamma_p$, rotate about $\hat{z}$ for $4\beta_p$, and measured along $\hat{k}$. The fidelity $\mathcal{F}_k$ measures the overlap between the final state and the state $\rho_{\hat{k}}$, whose Bloch vector is along direction $\hat{k}$.

Due to the periodicity in rotation, $F(4\boldsymbol{\gamma} + 2\boldsymbol{l}\pi, 4\boldsymbol{\beta} + 2\boldsymbol{l}'\pi) = F(\boldsymbol{\gamma}, \boldsymbol{\beta}) \Rightarrow F(\boldsymbol{\gamma} + \boldsymbol{l}\cdot\pi/2, \boldsymbol{\beta} + \boldsymbol{l}'\cdot\pi/2) = F(\boldsymbol{\gamma}, \boldsymbol{\beta})$, where $\boldsymbol{l},\ \boldsymbol{l}' \in \mathbb{Z}^p$. The search space can be limited to $\beta_k, \gamma_k \in [0, \pi/2]$ for $k = 1, 2, \ldots, p$.

QAOA on the ring of disagrees thus has a correspondence of physical picture in optimal quantum control. Noting the final average over k to get F, we can think of the system as an ensemble of spins, each spin k experiences a constant magnetic field along $\hat{k}$, (the quantization field), and the system is controlled by applying a strong uniform magnetic field along $\hat{z}$ in the "bang-bang" style. Specifically, when the field along $\hat{z}$, $\vec{B}_z$, is on, it dominates the quantization field and all spins rotate about $\hat{z}$ by the same angle $4\gamma_p$; when $\vec{B}_z$ is paused, each spin evolves freely, i.e., rotates about its own quantization axis $\hat{k}$ to pick up an angle $4\beta_p$, so on and so forth. After the whole control sequence is applied, the overall magnetization along $\hat{z}$, $F = \sum_k \langle\sigma_k^z\rangle$, is measured. The goal of the quantum control is to design a time sequence $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ so that F is minimized.

2. *Criticality-constrained sub-manifolds*

Since the trace operator preserves cycling, and the role of $\hat{z}$ and $\hat{k}$ in Eq. (41) are interchangeable, the evolution can also be viewed, after initializing the spin along the $\hat{k}$-direction, as rotate about $\hat{z}$-axis for angle $-4\beta_p$, rotate about $\hat{k}$ for $-4\gamma_p$,..., rotate about $\hat{z}$ for $-4\beta_1$, rotate about $\hat{k}$ for $-4\gamma_1$, and measured along $\hat{z}$.

Manifold 1 Due to equivalence in the two views, it must hold that

$$F_k(\boldsymbol{\gamma}, \boldsymbol{\beta}) = F_k(-\boldsymbol{\beta}', -\boldsymbol{\gamma}') \tag{42}$$

where

$$\boldsymbol{\gamma} \;=\; (\gamma_1, \gamma_2, \ldots, \gamma_{p-1}, \gamma_p) \tag{43}$$
$$\boldsymbol{\beta} \;=\; (\beta_1, \beta_2, \ldots, \beta_{p-1}, \beta_p) \tag{44}$$
$$\boldsymbol{\gamma}' \;=\; (\gamma_p, \gamma_{p-1}, \ldots, \gamma_2, \gamma_1) \tag{45}$$
$$\boldsymbol{\beta}' \;=\; (\beta_p, \beta_{p-1}, \ldots, \beta_2, \beta_1) \tag{46}$$

This can be verified with the help of a π-rotation about the axis $\hat{z} + \hat{k}$, see Appendix. B for a proof. Consider the manifold that satisfies

$$\gamma_i + \beta_{p+1-i} = 0 \text{ for } i = 1, 2, \ldots, p. \tag{47}$$

It has a special property: the gradient of the function $F_k(\boldsymbol{\gamma}, \boldsymbol{\beta})$ is constrained to lie tangent to the manifold. Therefore critical points in the manifold are critical points of the whole function.

Manifold 2 Eq. (41) actually holds for $\rho_{\hat{k}} = \frac{1}{2}(1 \pm \hat{k}\cdot\hat{\sigma})$ and $\rho_z = \frac{1}{2}(1 \pm \sigma^z)$. The $+$ $(-)$ sign correspond to the picture when the initial and final states are parallel (anti-parallel) with respect to the rotation axes, respectively. Comparing these two pictures, notice that rotation about any axis $-\hat{v}$ and $+\hat{v}$ by the same angle are Hermitian to each other, i.e., $\hat{R}_{(\hat{v})}(\alpha) = \hat{R}^\dagger_{(-\hat{v})}(\alpha)$, it must hold that

$$F_k(\boldsymbol{\gamma}, \boldsymbol{\beta}) = F_k(\boldsymbol{\beta}', \boldsymbol{\gamma}') \tag{48}$$

Eq. (48) defines another manifold

$$\gamma_i - \beta_{p+1-i} = 0 \text{ for } i = 1, 2, \ldots, p \tag{49}$$

with constrained gradient.

Eqs. (42) and (48) further indicate that F_k and is an even function of the angle sequence:

$$F_k(\boldsymbol{\gamma}, \boldsymbol{\beta}) = F_k(-\boldsymbol{\gamma}, -\boldsymbol{\beta}) \tag{50}$$

and accordingly so is F, $F(\boldsymbol{\gamma},\boldsymbol{\beta}) = F(-\boldsymbol{\gamma},-\boldsymbol{\beta})$.

Global extrema lie in the submanifolds In our numerical search, the minima of F were always contained in the manifold defined by Eq. (47) while the maxima of F always lie in the manifold Eq. (49).

E. Illustration of level-1 and level-2 QAOA

We use level-1 and level-2 QAOA to illustrate the results of symmetry and size-dependence of the optimization discussed above. Results on higher levels are shown in Appendix. C

For $p = 1$, the unitary evolution operator $U = e^{-i2\beta_1\sigma^z}e^{-i2\gamma_1\hat{k}\cdot\hat{\sigma}}$ Note that if a term $f(k)$ in F_k satisfis $f(N/2+1-k) = -f(k)$, then $f(k)$ would vanish in F through the summation over k; and note the properties of Pauli matrices, $\mathrm{Tr}[\sigma_\alpha\sigma_{\alpha'}] = 2\delta_{\alpha,\alpha'}$, one comes to

$$F = 2\sin(4\beta)\sin(4\gamma)\sum_k \sin^2\theta_k \tag{51}$$

$$= \begin{cases} N\sin(4\beta)\sin(4\gamma) & \text{for } N=2 \\ \frac{N}{2}\sin(4\beta)\sin(4\gamma) & \text{for } N>2 \end{cases} \tag{52}$$

The optimal angles are $(\gamma_1^*,\beta_1^*) = (3\pi/8,\pi/8)$ or $(\pi/8,3\pi/8)$.

For $N = 2$, the optimal angles correspond to $F^* = -N$ while for larger problem size, $F^* = -N/2$. This reflects the property that QAOA of level p suffices to perfectly optimize the ring for $N \leqslant 2p$ but for $N \geqslant 2p+2$ the optimization ratio is a fixed constant smaller than 1.

Eq. (52) is plotted in Fig. 1. Along the symmetry line $\beta_1+\gamma_1 = 0$, the critical points are global minima and saddle points. While along the symmetry line $\beta_1-\gamma_1 = 0$, the critical points are global maxima and saddle points.

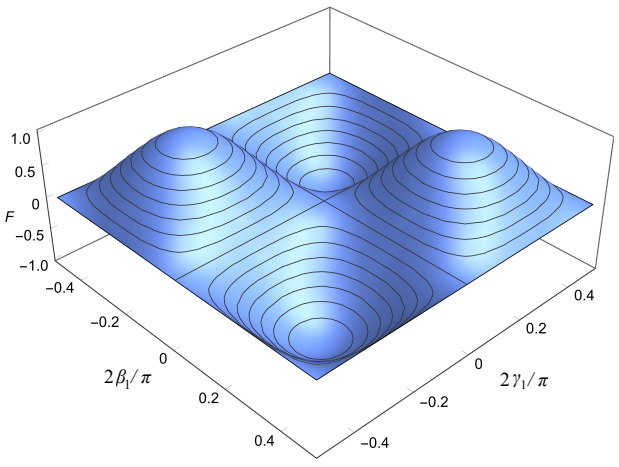

Fig. 1: level-1 QAOA. The expectation value F as a function of γ_1 and β_1.

For level-2, the evolution operator reads

$$U = e^{-i2\beta_2\sigma^z}e^{-i2\gamma_2\hat{\sigma}\cdot\hat{k}}e^{-i2\beta_1\sigma^z}e^{-i2\gamma_1\hat{\sigma}\cdot\hat{k}} \tag{53}$$

The expectation value F as a trigonometric function of $(\boldsymbol{\gamma},\boldsymbol{\beta})$ is shown in Appendix C. Numerically found optimal angles are $(\gamma_1^*,\beta_1^*,\gamma_2^*,\beta_2^*) = \pi\cdot(0.3956,\ 0.1978,\ 0.3022,\ 0.1044)$ or $\pi\cdot(0.2052,\ 0.1026,\ 0.3974,\ 0.2948)$

In both optimal angle sets, $4(\gamma_1^*+\beta_2^*)$ and $4(\gamma_2^*+\beta_1^*)$ are integer multipliers of 2π, thus both optima lie in the manifold defined by Eq. (47).

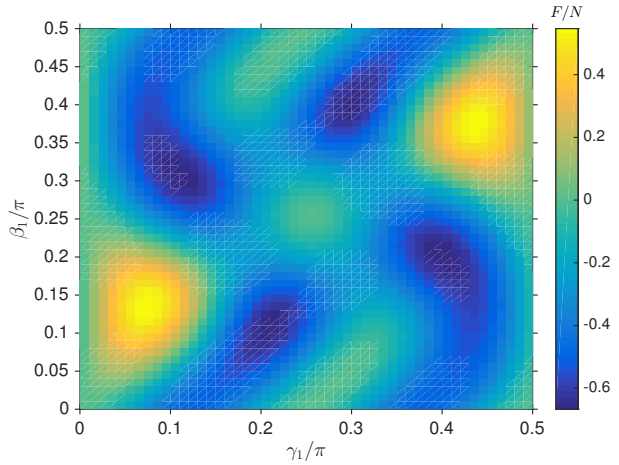

Fig. 2: The landscape of F/N for level-2 QAOA, in the submanifold Eq. (47). The four darkest spots indicate the global minima $F^*/N = -2/3$. The origin (0, 0) is a saddle point. No local minima is observed. The contour is symmetric w.r.t. $(\gamma_1,\beta_1) = (\pi/4,\pi/4)$, reflecting the symmetry in Eq. (50), (and the period $\pi/2$).

F. Discussion: controllability and optimality

The optima of F for QAOA level $p = 1$ to 10 are tabulated in Appendix. C. The optimal angles were obtained through numerical gradient descent search in the submanifold Eq. (47) . The evaluation for F_k in each step is realized as Eq. (32), which only involves $2p$ multiplications of 2-by-2 matrices, and sum over k gives F. The numerical evaluation therefore can be efficiently realized.

Start with a random initial guess of $(\boldsymbol{\gamma},\boldsymbol{\beta})$, (with sufficiently small steps) the search always converges to a global minimum. This indicates that at least within the sub-manifold, all local minima are global minima. For example, for $p = 2$, there are two free parameters in the submanifold, which we choose as γ_1 and β_1. Figure. 2 shows the landscape of the expectation value F. Four minima (darkest spots) observed in one period $(\gamma_1,\beta_1,\in[0,\pi/2])$ are all global minima.

This result calls for extended understanding of landscapes of quantum control. In the quantum control the-

ory, it has been shown that assuming *controllability*, i.e., evolution between any two states is achievable via the set of controls given, the landscape of the infidelity F over the space of parameter values $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ generically has only global minima [13–15]. Without controllability, the quantum control landscape in general is rugged and admits local minima [16].

In the case of QAOA, the controls are constrained in a specific way: if an infinite number of controls are allowed, i.e., $p \to \infty$, then the system is controllable. The finite number of control steps dictated by the level p limits the controllability. For $N \leqslant 2p$, the number of controls, i.e., free independent parameters in control, $2p$, is equal or larger than the number of spins in the chain, and twice the number of the pseudo-spins. Since arbitrary rotation of single spin can always be achieved with any two-axis controls, full controllability is guaranteed. The no local minima observation thus aligns with the theory in quantum control. For large problem size, $N \geqslant 2p + 2$, the system is not controllable. Nevertheless, our numerical results indicate that, at least within the sub-manifold Eq. (47), all local minima are global minima.

V. SUMMARY AND DISCUSSIONS

We studied parameter setting for QAOA on MaxCut. For level-1 QAOA, we extended the results in Ref. [1], providing an analytical expression for general graphs. As a corollary, for triangle-free graphs with fixed vertex degree, the optimal angles for the driver Hamiltonian can be directly read off while the optimal angles for the problem Hamiltonian show a dependence on the vertex degree. For higher p, this approach becomes cumbersome, providing further evidence that more advanced parameter setting techniques need to be developed.

For a special case of MaxCut, the ring of disagrees, using a fermionic representations, we provided analysis of QAOA for arbitrary p. Applying the Jordan-Wigner transformation to obtain a fermionic representation, transforms the evolution of the system under QAOA to a set of independent evolutions within two-dimentional subspaces, showing that the parameter setting problem corresponds to finding optimal control parameters for an ensemble of non-interacting spins. From this treatment we obtain analytical expression for any p, and an easy numerical search for the corresponding optimal angles.

The fermionic picture also enables us to explain symmetries in the optimal parameters, suggesting a means to further reduce the effort required to find optimal parameters by restricting to manifolds defined by this symmetry. The specific symmetry in the problem of ring of disagrees has its roots in the equal footing of the action of the driver and the problem Hamiltonians – each corresponds to a single spin rotation. We observed numerically that within the parameter space, all minima are global minima. While such a "no-trap" character of a quantum control landscape can be explained given controllability, the structure of QAOA for finite p often does not guarantee controllability. Future research that reveals the underlying theory may shed further light on the structure of QAOA and inspire ways to simplify or improve the algorithm.

VI. ACKNOWLEDGEMENTS

We thank Sergey Knysh for enlightening discussions and in particular for the proof of the approximation ratio. ZW thank Jason Dominy for helpful discussions on the landscape of quantum control. We thank Salvatore Mandrà, Brian O'Gorman, and Davide Venturelli for useful discussions. The authors would like to acknowledge support from the NASA Advanced Exploration Systems program and NASA Ames Research Center. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright annotation thereon.

Appendix A: Proof of Theorem 1

Proof of Theorem 1. For $p = 1$, only terms corresponding to neighbors of u and v can contribute to the expectation of C_{uv}. [1] We thus partition our objective function as

$$C = \frac{1}{2}(I - \sigma_u^z\sigma_v^z) + C_u + C_v + \bar{C},$$

where C_u is the d-many constraints involving only vertex u but not v, and C_v is the e-many constraints involving only v. The remaining constraints $\bar{C}$ do not contribute to the expectation value.

Let $c = \cos 2\beta$ and $s = \sin 2\beta$. We have

$$\begin{aligned} & e^{i\beta B}\sigma_u^z\sigma_v^z e^{-i\beta B} \\ = \; & c^2\sigma_u^z\sigma_v^z + sc(\sigma_u^y\sigma_v^z + \sigma_u^z\sigma_v^y) + s^2\sigma_u^y\sigma_v^y \end{aligned} \tag{A1}$$

The first term $\sigma_u^z\sigma_v^z$ commutes with C and does not contribute to $\langle C_{uv}\rangle$. We conjugate each remaining term separately by $e^{i\gamma C}$. Let $c' = \cos\gamma$ and $s' = \sin\gamma$. We have

$$\begin{aligned} & \mathrm{Tr}[\rho_0 e^{i\gamma C}\sigma_u^y\sigma_v^z e^{-i\gamma C}] \\ = \; & \mathrm{Tr}\left[\rho_0(Ic' - is'\sigma_u^z\sigma_v^z)\prod_{i=1}^{d}(Ic' - is'\sigma_u^z\sigma_{w_i}^z)\sigma_u^y\sigma_v^z\right]. \end{aligned} \tag{A2}$$

Expanding the product on the right gives a sum of tensor products of Pauli operators. Clearly, the only term that can contribute is proportional to $\sigma_u^z\sigma_v^z * I^{\otimes d} * \sigma_u^y\sigma_v^z = -i\sigma_u^x$. Thus we have

$$\mathrm{Tr}\left[\rho_0 e^{i\gamma C}\sigma_u^y\sigma_v^z e^{-i\gamma C}\right] = \mathrm{Tr}\left[\rho_0(-i)s'c'^d(-i\sigma_u^x)\right] = -s'c'^d \tag{A3}$$

By symmetry, we have $\mathrm{Tr}[\rho_0 e^{i\gamma C}\sigma_u^z\sigma_v^y e^{-i\gamma C}] = -s'c'^e$. Observe that these terms are independent of the number of mutual neighbours (triangles) of u and v. The next term is

$$\begin{aligned} & \mathrm{Tr}\left[\rho_0(e^{i\gamma C}\sigma_u^y\sigma_v^y e^{-i\gamma C}\right]\mathrm{Tr}\left[e^{2i\gamma C_u}e^{2i\gamma C_v}\sigma_u^y\sigma_v^y\right] \\ = \; & \mathrm{Tr}\left[\prod_{i=1}^{d}(c'I - is'\sigma_u^z\sigma_{w_i}^z)\prod_{j=1}^{e}(c'I - is'\sigma_v^z\sigma_{w_j}^z)\sigma_u^y\sigma_v^y\right] \end{aligned} \tag{A4}$$

The simplest term that contributes in this case is $\mathrm{Tr}\left[\rho_0 fc'^{d+e-2}(-is')^2(-i\sigma_u^x)(-i\sigma_v^x)\right] = fc'^{d+e-2}s'^2$. Corresponding to the triangles of $\langle uv\rangle$, in the above product we have f-many distinct values i such that $w_i = w_j$. As $\sigma_u^z\sigma_{w_i}^z * \sigma_u^z\sigma_{w_i}^z = I$, if $f > 2$ then higher order terms depending on the number of triangles f will contribute. For example, the next order terms will result from three pairs of $(\sigma_u^z\sigma_{w_i}^z, \sigma_u^z\sigma_{w_i}^z)$ and hence be proportional to s'^6. Thus we have

$$\begin{aligned} & \mathrm{Tr}\left[\rho_0(e^{i\gamma C}\sigma_u^y\sigma_v^y e^{-i\gamma C})\right] \\ = \; & \binom{f}{1}c'^{d+e-2}s'^2 + \binom{f}{3}c'^{d+e-6}s'^6 \\ & + \binom{f}{5}c'^{d+e-10}s'^{10} + \ldots \\ = \; & c'^{d+e-2f}\sum_{i=1,3,5,\ldots}^{f}\binom{f}{i}(c'^2)^{f-i}(s'^2)^i. \end{aligned} \tag{A5}$$

To sum this series, recall the binomial theorem, which we may split into even and odd sums as

$$\begin{aligned} & \sum_{i=0,2,\ldots}^{f}\binom{f}{i}a^{f-i}b^i + \sum_{i=1,3,\ldots}^{f}\binom{f}{i}a^{f-i}b^i \\ = \; & \sum_{i=0}^{f}\binom{f}{i}a^{f-i}b^i = (a+b)^f \end{aligned} \tag{A6}$$

which also gives

$$\begin{aligned} & \sum_{i=0,2,\ldots}^{f}\binom{f}{i}a^{f-i}b^i - \sum_{i=1,3,\ldots}^{f}\binom{f}{i}a^{f-i}b^i \\ = & \sum_{i=0}^{f}(-1)^i\binom{f}{i}a^{f-i}b^i = (a-b)^f, \end{aligned} \tag{A7}$$

and hence

$$\sum_{i=1,3,\ldots}^{f}\binom{f}{i}a^{f-i}b^i = \frac{1}{2}\left((a+b)^f - (a-b)^f\right). \tag{A8}$$

Thus the above sum becomes

$$\sum_{i=1,3,\ldots}^{f}\binom{f}{i}(c'^2)^{f-i}(s'^2)^i = \frac{1}{2}(1-\cos^f 2\gamma) \tag{A9}$$

which yields

$$\mathrm{Tr}\left[\rho_0(e^{i\gamma C}\sigma_u^y\sigma_v^y e^{i\gamma C})\right] = \frac{1}{2}c'^{d+e-2f}(1-\cos^f 2\gamma) \tag{A10}$$

Putting this all together, we have

$$\begin{aligned} \langle C_{uv}\rangle \; &= \; \mathrm{Tr}\left[\rho_0 e^{i\gamma C}e^{i\beta B}C_{uv}e^{-i\beta B}e^{-i\gamma C}\right] \\ &= \; \frac{1}{2} + \frac{sc}{2}\mathrm{Tr}\left[\rho_0 e^{i\gamma C}(\sigma_u^y\sigma_v^z + \sigma_u^z\sigma_v^y)e^{-i\gamma C})\right] \\ &\quad -\frac{s^2}{2}\mathrm{Tr}\left[\rho_0 e^{i\gamma C}\sigma_u^y\sigma_v^y e^{-i\gamma C})\right] \\ &= \; \frac{1}{2} + \frac{1}{2}scs'(c'^d + c'^e) - \frac{1}{4}s^2c'^{d+e-2f}(1-\cos^f 2\gamma). \end{aligned} \tag{A11}$$

Thus we see that the expectation of each edge term $\langle C_{uv}\rangle$ depends only on the parameters (d, e, f) which characterize the edge $\langle uv\rangle$. The desired expectation value of C is given by the sum of the expectation of each edge term C_{uv}, which reduces to the sum over possible (d, e, f) values, weighted by the multiplicity of each edge type for the given graph. □

Appendix B: Proof of symmetry relation Eq. (42)

We prove Eq. (42)

$$F_k(\boldsymbol{\gamma},\boldsymbol{\beta}) = F_k(-\boldsymbol{\beta}',-\boldsymbol{\gamma}') \tag{B1}$$

where

$$\begin{aligned}
\boldsymbol{\gamma} &= (\gamma_1,\gamma_2,\ldots,\gamma_{p-1},\gamma_p) && \text{(B2)}\\
\boldsymbol{\beta} &= (\beta_1,\beta_2,\ldots,\beta_{p-1},\beta_p) && \text{(B3)}\\
\boldsymbol{\gamma}' &= (\gamma_p,\gamma_{p-1},\ldots,\gamma_2,\gamma_1) && \text{(B4)}\\
\boldsymbol{\beta}' &= (\beta_p,\beta_{p-1},\ldots,\beta_2,\beta_1) && \text{(B5)}
\end{aligned}$$

Proof. We consider a unitary operator $R = \cos\frac{\theta}{2}\sigma^z + \sin\frac{\theta}{2}\sigma^x$ which rotates a Bloch vector about axis $(\hat{k}+\hat{z})$ by π. Note that $R^\dagger = R$, $R^2 = 1$ and

$$\begin{aligned}
R\sigma^z R &= \hat{k}\cdot\hat{\sigma}\\
R\hat{k}\cdot\hat{\sigma}R &= \sigma^z\\
RU_B(\beta)R &= U_C(\beta)\\
RU_C(\gamma)R &= U_B(\gamma) && \text{(B6)}
\end{aligned}$$

we have

$$\begin{aligned}
RUR &= RU_B(\beta_p)RRU_C(\gamma_p)R\ldots RU_B(\beta_1)RRU_C(\gamma_1)R\\
&= U_C(\beta_p)U_B(\gamma_p)\cdots U_C(\beta_1)U_B(\gamma_1)\\
&= \left[U_B(-\gamma_1)U_C(-\beta_1)\cdots U_B(-\gamma_p)U_C(-\beta_p)\right]^\dagger\\
&= U'^\dagger && \text{(B7)}
\end{aligned}$$

where $U' \equiv U(-\boldsymbol{\beta}',-\boldsymbol{\gamma}')$.

Insert $R^2 = \mathbf{1}$ to F_k we get

$$\begin{aligned}
F_k(\boldsymbol{\gamma},\boldsymbol{\beta}) &= \mathrm{Tr}[(\hat{k}\cdot\hat{\sigma})U\sigma^z U^\dagger]\\
&= \mathrm{Tr}[R(\hat{k}\cdot\hat{\sigma})RRURR\sigma^z RRU^\dagger R]\\
&= \mathrm{Tr}[\sigma^z RUR(\hat{k}\cdot\hat{\sigma})RU^\dagger R]\\
&= \mathrm{Tr}[\sigma^z U'^\dagger(\hat{k}\cdot\hat{\sigma})U']\\
&= \mathrm{Tr}[(\hat{k}\cdot\hat{\sigma})U'\sigma^z U'^\dagger]\\
&= F_k(-\boldsymbol{\beta}',-\boldsymbol{\gamma}') && \text{(B8)}
\end{aligned}$$

□

Appendix C: Detailed results for $p = 2$ and higher

For $p = 2$, terms in that are non-vanishing to F is

$$\begin{aligned}
\frac{F}{N} = \frac{1}{64}\big[&-7\cos(4\beta_1+4\beta_2+4\gamma_1+4\gamma_2)-6\cos(4\beta_1+4\beta_2+4\gamma_1)\\
&+3\cos(4\beta_1+4\beta_2-4\gamma_1+4\gamma_2)+4\cos(4\beta_1+4\beta_2+4\gamma_2)\\
&+3\cos(4\beta_1-4\beta_2+4\gamma_1+4\gamma_2)-6\cos(4\beta_1-4\beta_2+4\gamma_1)-\\
&3\cos(4\beta_1-4\beta_2-4\gamma_1+4\gamma_2)+4\cos(4\beta_1+4\gamma_1+4\gamma_2)-\\
&4\cos(4\beta_1+4\gamma_1)-4\cos(4\beta_1+4\gamma_2)-3\cos(-4\beta_1+4\beta_2+4\gamma_1+4\gamma_2)\\
&+6\cos(-4\beta_1+4\beta_2+4\gamma_1)+3\cos(-4\beta_1+4\beta_2-4\gamma_1+4\gamma_2)+\\
&7\cos(-4\beta_1-4\beta_2+4\gamma_1+4\gamma_2)+6\cos(-4\beta_1-4\beta_2+4\gamma_1)-\\
&3\cos(-4\beta_1-4\beta_2-4\gamma_1+4\gamma_2)-4\cos(-4\beta_1-4\beta_2+4\gamma_2)-\\
&4\cos(-4\beta_1+4\gamma_1+4\gamma_2)+4\cos(-4\beta_1+4\gamma_1)+4\cos(-4\beta_1+4\gamma_2)-\\
&6\cos(4\beta_2+4\gamma_1+4\gamma_2)-6\cos(4\beta_2-4\gamma_1+4\gamma_2)-4\cos(4\beta_2+4\gamma_2)\\
&+6\cos(-4\beta_2+4\gamma_1+4\gamma_2)+6\cos(-4\beta_2-4\gamma_1+4\gamma_2)\\
&+4\cos(-4\beta_2+4\gamma_2)\big] && \text{(C1)}
\end{aligned}$$

If limited in the sub-manifold Eq. (47),

p	$-F^*/N$	γ_1	β_1	γ_2	β_2	γ_3	β_3	γ_4	β_4	γ_5	β_5
1	1/2	0.1250									
2	2/3	0.2052	0.1026								
3	3/4	0.2268	0.1888	0.0918							
4	4/5	0.2357	0.2161	0.1791	0.0850						
5	5/6	0.2403	0.2282	0.2094	0.1724	0.0802					
6	6/7	0.3035	0.1639	0.2506	0.2835	0.0794	0.2409				
7	7/8	0.2303	0.1623	0.3468	0.2690	0.1042	0.2397	0.1599			
8	8/9	0.2445	0.1638	0.2839	0.3484	0.1539	0.1530	0.2581	0.1291		
9	9/10	0.1929	0.1648	0.3307	0.3016	0.1551	0.2538	0.2174	0.1089	0.3117	
10	10/11	0.2208	0.1374	0.3098	0.2974	0.2702	0.1205	0.3148	0.1904	0.1423	0.2572

TABLE I: Optimal angles for different levels of QAOA. Angles are in unit π. Gradient descend search is implemented with the optimal angles for level p is set to be the initial guess for level p. Arbitrary initial guess also always converges to a global minimum. Multiple sets of optimal angles exist for $p \geqslant 2$, only one of them is shown for each level.

$$\begin{aligned}\frac{F}{N} \;=\; \frac{1}{64}\Big(& -2\cos(8\beta_1) + 3\cos(8\beta_1+8\gamma_1) - 12\cos(4\beta_1+8\gamma_1) \\ & -8\cos(4\beta_1+4\gamma_1) + 12\cos(4\beta_1-8\gamma_1) + 8\cos(4\beta_1-4\gamma_1) \\ & +7\cos(8\beta_1-8\gamma_1) - 8\cos(8\beta_1-4\gamma_1) + 6\cos(8\gamma_1) \\ & +8\cos(4\gamma_1) - 14\Big) \end{aligned} \tag{C2}$$

In Table. I we show numerical data for higher QAOA levels (multiple optima were found for $p \geqslant 2$, we show only one for each p).

[1] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A Quantum Approximate Optimization Algorithm. *arXiv:1411.4028 [quant-ph]*, November 2014.

[2] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A Quantum Approximate Optimization Algorithm Applied to a Bounded Occurrence Constraint Problem. *arXiv:1412.6062 [quant-ph]*, December 2014.

[3] Edward Farhi and Aram W. Harrow. Quantum Supremacy through the Quantum Approximate Optimization Algorithm. *arXiv:1602.07674 [quant-ph]*, February 2016.

[4] Zhi-Cheng Yang, Armin Rahmani, Alireza Shabani, Hartmut Neven, and Claudio Chamon. Optimizing variational quantum algorithms using pontryagin's minimum principle. *arXiv preprint arXiv:1607.06473*, 2016.

[5] Zhang Jiang, Eleanor Rieffel, and Zhihui Wang. A QAOA-inspired circuit for grover's unstructured search using a transverse field. *arXiv preprint arXiv:1702.02577*, 2017.

[6] Stuart Hadfield and Eleanor Rieffel. Quantum approximate constrained optimization algorithm. *in preparation*, 2017.

[7] Boaz Barak, Ankur Moitra, Ryan O'Donnell, Prasad Raghavendra, Oded Regev, David Steurer, Luca Trevisan, Aravindan Vijayaraghavan, David Witmer, and John Wright. Beating the random assignment on constraint satisfaction problems of bounded degree. *arXiv:1505.03424*, 2015.

[8] Dave Wecker, Matthew B. Hastings, and Matthias Troyer. Progress towards practical quantum variational algorithms. *Phys. Rev. A*, 92:042303, Oct 2015.

[9] John Preskill. Quantum computing and the entanglement frontier. *arXiv:1203.5813*, March 2012.

[10] Sergio Boixo, Sergei V. Isakov, Vadim N. Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, John M. Martinis, and Hartmut Neven. Characterizing Quantum Supremacy in Near-Term Devices. *arXiv:1608.00263*, July 2016.

[11] Elliott Lieb, Theodore Schultz, and Daniel Mattis. Two soluble models of an antiferromagnetic chain. *Annals of Physics*, 16(3):407–466, December 1961.

[12] Eytan Barouch, Barry M. McCoy, and Max Dresden. Statistical Mechanics of the XY Model. I. *Physical Review A*, 2(3):1075–1092, September 1970.

[13] Herschel A. Rabitz, Michael M. Hsieh, and Carey M. Rosenthal. Quantum optimally controlled transition landscapes. *Science*, 303(5666):1998–2001, 2004.

[14] Re-Bing Wu, Ruixing Long, Jason Dominy, Tak-San Ho, and Herschel Rabitz. Singularities of quantum control landscapes. *Phys. Rev. A*, 86:013405, Jul 2012.

[15] Benjamin Russell, Herschel Rabitz, and Rebing Wu. Quantum control landscapes are almost always trap free. *arXiv preprint arXiv:1608.06198*, 2016.

[16] Re-Bing Wu, Michael A. Hsieh, and Herschel Rabitz. Role of controllability in optimizing quantum dynamics. *Phys. Rev. A*, 83:062306, Jun 2011.