Article

A Depth-Progressive Initialization Strategy for Quantum Approximate Optimization Algorithm

Xinwei Lee, Ningyi Xie, Dongsheng Cai, Yoshiyuki Saito and Nobuyoshi Asai

Special Issue
Advances in Quantum Computing and Applications

Edited by
Prof. Dr. Frank Phillipson, Dr. Sebastian Feld, Dr. Matthias Möller, Ward van der Schoot and Niels Neumann

Article

A Depth-Progressive Initialization Strategy for Quantum Approximate Optimization Algorithm

Xinwei Lee [1,*], Ningyi Xie [1], Dongsheng Cai [1], Yoshiyuki Saito [2] and Nobuyoshi Asai [2]

1 Faculty of Engineering, Information and Systems, University of Tsukuba, Ibaraki 305-8577, Japan
2 School of Computer Science and Engineering, University of Aizu, Fukushima 965-0006, Japan
* Correspondence: xwlee@cavelab.cs.tsukuba.ac.jp

Abstract: The quantum approximate optimization algorithm (QAOA) is known for its capability and universality in solving combinatorial optimization problems on near-term quantum devices. The results yielded by QAOA depend strongly on its initial variational parameters. Hence, parameter selection for QAOA becomes an active area of research, as bad initialization might deteriorate the quality of the results, especially at great circuit depths. We first discuss the patterns of optimal parameters in QAOA in two directions: the angle index and the circuit depth. Then, we discuss the symmetries and periodicity of the expectation that is used to determine the bounds of the search space. Based on the patterns in optimal parameters and the bounds restriction, we propose a strategy that predicts the new initial parameters by taking the difference between the previous optimal parameters. Unlike most other strategies, the strategy we propose does not require multiple trials to ensure success. It only requires one prediction when progressing to the next depth. We compare this strategy with our previously proposed strategy and the layerwise strategy for solving the Max-cut problem in terms of the approximation ratio and the optimization cost. We also address the non-optimality in previous parameters, which is seldom discussed in other works despite its importance in explaining the behavior of variational quantum algorithms.

Keywords: quantum variational algorithm; combinatorial optimization; initialization strategy

MSC: 81P68

1. Introduction

The Quantum Approximate Optimization Algorithm (QAOA) was first introduced by Farhi et al. [1] as a quantum-classical hybrid algorithm, which consists of a quantum circuit with an outer classical optimization loop, to approximate the solution of combinatorial optimization problems. Since then, many studies have been conducted to discuss its quantum advantage and its implementability on near-term Noisy Intermediate Scale Quantum (NISQ) devices [2–8]. QAOA is shown to guarantee the approximation ratio $\alpha > 0.6924$ for circuit depth $p = 1$ in the Max-cut problem on 3-regular graphs [1]. Further studies have shown a lower bound of $\alpha > 0.7559$ for $p = 2$ and $\alpha > 0.7924$ for $p = 3$ [9].

Parameter selection in QAOA has been an active area of research due to the difficulties that lie within the classical optimization of QAOA, especially the barren plateaus problem [10–12]. Patterns in optimal parameters of QAOA have constantly been studied, and various strategies are proposed to improve the quality of the solution [13–20]. Machine learning methods are also employed to improve parameter selection [18,19,21,22]. For instance, it is found that for some classes of graphs, e.g., regular graphs, the optimal parameters of a smaller graph can be reused as they are on larger graphs to approximate the solution without solving them [23]. This characteristic is defined as the 'parameter concentration' in [24] and is found in some projectors as well. Recently, the transferability of parameters has also been studied with the discovery of parameter concentration in

d-regular subgraphs with the same parity (odd or even) [25]. These works focused on the characteristics of the optimal parameters of QAOA in the direction of problem size *n*.

Another direction that is mostly concerned is the QAOA circuit depth *p*. It is discussed that we usually need a larger *p* to solve problems with a larger *n* with a higher *α*. However, as *p* grows larger, the increased occurrence of local minima makes the optimization difficult. If the QAOA parameters are initialized randomly, there is a high chance that they will converge to an undesired local optimum. This is shown in our previous work [26], and we proposed to use the previous optimal parameters as starting points for the following depths. We found out that this improves the convergence of the approximation ratio towards optimality. This implies that there is some relationship between the previous optimum and the current optimum. This motivates us to study the relationship between the optimal parameters and circuit depths.

In this work, we study the patterns in the optimal parameters of QAOA Max-cut in two directions: the angle index *j* and the circuit depth *p*. We name the pattern exhibited by the optima with respect to *j* as the *adiabatic path*, and the pattern with respect to *p* as the *non-optimality*, which we explain each of them in detail in Section 3. Also, as the expectation function of QAOA Max-cut is highly periodic and symmetric, the landscape it produces will have multiple optima. Therefore, for the adiabatic path and the non-optimality patterns to be seen explicitly, it is required to restrict the bounds of the parameter search space, so that the redundant optimal points in the full search space can be removed.

Based on the adiabatic path and the non-optimality, we propose the *bilinear initialization strategy* (or simply *bilinear strategy*) which generates initial parameters at the new depth given the optimal parameters from the previous depths. This strategy aims to reproduce the optimal patterns at the new depth so that the initial parameters generated will be close to the optimal parameters, reducing the likelihood to converge to an undesired optimum. Since the previous parameters are used to predict the new parameters, this strategy requires optimization at every depth up to the desired depth. However, unlike most other strategies [27,28] which require multiple trials to ensure success, the bilinear strategy only requires one trial at each depth.

We then demonstrate the effect of the bilinear strategy on solving the Max-cut problem for 30 non-isomorphic instances composing different classes of graphs, including the 3-regular, the 4-regular, and the Erdös-Rényi graphs with different edge probabilities. We compare the strategy with our previously proposed parameters fixing strategy, and the layerwise strategy, in terms of approximation ratio and optimization cost.

We also study the case where the strategy might fail in odd-regular graphs with wrongly specified bounds. The result is interesting as it shows that there exists an optimum in the expectation function of odd-regular graphs that does not follow the adiabatic path pattern. Instead, the *β* parameters (parameters associated with the mixer Hamiltonian) oscillate back and forth. We then explain this phenomenon using the symmetry in odd-regular graphs.

2. QAOA: Background and Notation

The objective of QAOA is to maximize the expectation of some cost Hamiltonian H_z with respect to the ansatz state $|\psi(\gamma, \beta)\rangle$ prepared by the evolution of the alternating operators:

$$|\psi_p(\gamma, \beta)\rangle = \prod_{j=1}^{p} e^{-i\beta_j H_x} e^{-i\gamma_j H_z} |+\rangle^{\otimes n}. \tag{1}$$

where $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_p)$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$ are the $2p$ variational parameters, with $\gamma \in [0, 2\pi)^p$ and $\beta \in [0, \pi)^p$. $|+\rangle^{\otimes n}$ corresponds to n qubits in the ground state of $H_x = \sum_{j=1}^{n} X_j$, where X_j is the Pauli X operator acting on the j-th qubit.

In this paper, we consider the Max-cut problem, which aims to divide a graph into two parts, with the maximum number of edges between them. The Max-cut problem is

an NP-complete problem due to its reducibility to the MAX-2-SAT problem [29]. The cost Hamiltonian H_z for the Max-cut problem for an unweighted graph $G = (V, E)$ is given as

$$H_z = \frac{1}{2} \sum_{(j,k) \in E} (\mathbb{1} - Z_j Z_k), \tag{2}$$

where Z_j is the Pauli Z operator acting on the j-th qubit. The $Z_j Z_k$ operators are applied to qubits j and k for every edge (j, k) in the graph. We define the expectation of H_z with respect to the ansatz state in Equation (1):

$$F_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) \equiv \langle \psi_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) | H_z | \psi_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle, \tag{3}$$

where p is known as the circuit depth of QAOA. Solving the problem with QAOA is equivalent to maximizing Equation (3) with respect to the variational parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$. This can be completed with a classical optimizer that searches for the maximum F and the parameters that maximize it:

$$(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*) \equiv \arg \max_{\boldsymbol{\gamma}, \boldsymbol{\beta}} F(\boldsymbol{\gamma}, \boldsymbol{\beta}), \tag{4}$$

where the superscript * denotes optimal parameters. We also define the *approximation ratio* α as

$$\alpha \equiv \frac{F(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)}{C_{\max}}, \tag{5}$$

where $C_{\max}$ is the maximum cut value for the graph. The approximation ratio is a typical evaluation metric indicating how close the solution given by QAOA is to the true solution, $0 \le \alpha \le 1$, with the value of 1 nearer to the true solution.

Throughout the paper, we use the symbol ϕ to generally denote either γ or β in situations where the distinction of both is not required. Also, we sometimes use $\boldsymbol{\Phi}_p$ to denote the entire parameter vector at circuit depth p:

$$\boldsymbol{\Phi}_p \equiv (\boldsymbol{\gamma}, \boldsymbol{\beta})_p = (\gamma_1, \ldots, \gamma_p, \beta_1, \ldots, \beta_p). \tag{6}$$

For a single parameter, we use ϕ_j^p to denote the parameter at circuit depth p with index j.

The maximum of F_p in the p-level search space will approach $C_{\max}$ as $p \to \infty$, thus the approximation ratio α will approach 1 [1]. However, due to the increased occurrence of local maxima in larger p, it gets more difficult to find the maximum F_p [3,27]. If the parameters were initialized randomly, the optimizer is more likely to be trapped in local maxima for larger p [26]. The choice of initial points for the optimizer determines whether the optimizer converges to a global maximum. Hence, we would prefer to have "good" initial points for the optimizer to converge to the desired maximum.

3. Patterns in the Optimal Parameters of QAOA
3.1. Resemblance to the Quantum Adiabatic Evolution

It has been repeatedly reported that the optimal parameters resemble the adiabatic quantum computation (AQC) process [30], where the mixer Hamiltonian H_x is gradually turned off (decreasing β) and the cost Hamiltonian H_z is slowly turned on (increasing γ) [27,31]. This comes from the fact that the angles γ and β is related to the discrete time step of the adiabatic process [1,15,30]. Consider the time-dependent Hamiltonian $H(t) = (1 - t/T)H_x + (t/T)H_z$ going through a simple adiabatic evolution with total run time T. Discretizing the evolution gives

$$e^{-i \int_0^T H(t) dt} \approx \prod_{j=1}^{p} e^{-iH(j\Delta t)\Delta t}, \tag{7}$$

with $t = j\Delta t$. Applying the first order Lie-Suzuki-Trotter decomposition to Equation (7) gives

$$(7) \approx \prod_{j=1}^{p} e^{-i(1-j\Delta t/T)H_x\Delta t} e^{-i(j\Delta t/T)H_z\Delta t}. \tag{8}$$

We can then substitute $\gamma_j = (j\Delta t/T)\Delta t$ and $\beta_j = (1 - j\Delta t/T)\Delta t$ into Equation (8), and it leads to the QAOA form in Equation (1). Note that the discretization in Equation (7) divides the total run time T into p steps, i.e., $\Delta t = T/p$. Hence, we obtain the *parameter-index-depth* relation:

$$\gamma_j^p = \frac{j}{p}\Delta t; \quad \beta_j^p = \left(1 - \frac{j}{p}\right)\Delta t. \tag{9}$$

It is obvious that γ_j increases linearly with j and β_j decreases linearly with j. Previous works [2,15,27,31,32] have shown that the optimal parameters of QAOA tend to follow this linear-like adiabatic path, and the pattern becomes nearer to linearity as p increases. Consequently, this pattern is exploited in devising various strategies. Figure 1a shows the adiabatic path taken by the optimal parameters at different p. Note that the patterns are not completely linear. This might be due to the discretization error and the Trotter error in our process of approximating the continuous evolution, and it is expected to approach linear as $p \to \infty$ [32].
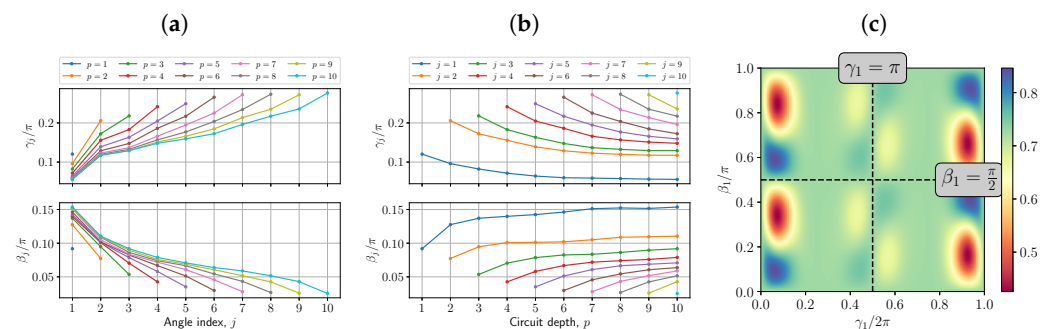
Figure 1. Optimal parameters variation of a 10-node Erdös-Rényi graph with edge probability of 0.7. (**a**) The variation of the optimal parameters at fixed circuit depth p against the angle index j. It shows the adiabatic path of the parameters with increasing γ and decreasing β. (**b**) The variation of the optimal parameters at fixed angle index j against the circuit depth p. It shows the non-optimality of the parameters with decreasing γ and increasing β. (**c**) The landscape of $p = 1$ normalized expectation (i.e., α) against γ_1 and β_1. The symmetry is shown by the $\gamma_1 = \pi$ axis and the periodicity is shown by the $\beta_1 = \pi/2$ axis. It can be seen that the landscape in $\gamma_1 \in [\pi, 2\pi)$ is the landscape in $\gamma_1 \in [0, \pi)$ rotated by $180°$, and the landscape just repeats itself beyond $\beta_1 = \pi/2$.

3.2. Non-Optimality of Previous Parameters

Besides the adiabatic path, we also discovered the non-optimality of optimal parameters from previous depths, i.e., the optimal parameters for p are not optimal for $p + 1$. The optimal parameters are shifted a little as the depth increases, as observed in Figure 1b. This phenomenon can also be inferred from Equation (9). As p increases, at the same index j, γ_j will decrease and β_j will increase. Also, we noticed that as p gets larger, the parameters at smaller indices have fewer changes compared with those with larger indices, e.g., it can be observed in Figure 1b that $|\phi_8^{10} - \phi_8^9|$ (rightmost two points of the gray line) is greater than $|\phi_1^{10} - \phi_1^9|$ (rightmost two of the blue line). We emphasize that the non-optimality is just the counterpart of the adiabatic path as they can be explained with the same relation, but it is seldom discussed in previous works. The patterns in the optimal parameters seem to be inherited from the time steps in the discrete adiabatic evolution. Note that Figure 1a,b are essentially plotted from the same set of parameters, viewed from different perspectives (j and p). They are both plotted for better visualization.

The results of layerwise training of QAOA also imply this non-optimality [28]. Layerwise training is an optimization strategy in which only the parameters of the current layer are optimized, the rest of the parameters are taken from the previous optimal parameters. The layerwise training has a relatively low training cost in exchange for a lower approximation ratio, as it suffers from premature saturation (saturation before the approximation ratio reaches 1). In [28], the authors discussed the premature saturation at $p = n$ for the rank-1 projector Hamiltonian $H_z = |0^n\rangle\langle 0^n|$, where n is the number of qubits. Since the previous parameters in layerwise training are held constant and not allowed to move throughout the optimization, they will not reach the global minimum because of the non-optimality.

3.3. Bounded Optimization of QAOA

The adiabatic path and the non-optimality show that the optimal parameters exhibit some trends related to AQC. However, in the QAOA parameter space, not only the adiabatic path leads to the solution of the problem. There are redundant optimal points in the search space whose patterns do not follow the adiabatic path. This leads to the optimization of the bounded parameter space, where no redundancy exists. Therefore, the properties of the problem and its parameter space need to be studied beforehand to ensure only one optimum exists in the parameter space.

For instance, the bounds of the unweighted Max-cut problem were originally taken as $\gamma \in [0, 2\pi)^p$ and $\beta \in [0, \pi)^p$ because of their periodicity [1]. However, it is further discussed in [27] that the operator $e^{-i(\pi/2)H_x} = X^{\otimes n}$ commutes through the operators in Equation (1), and due to the symmetry of the solutions, the period of β becomes $\pi/2$. Also, QAOA has a time-reversal symmetry:

$$F_p(\gamma, \beta) = F_p(-\gamma, -\beta) = F_p\left(2\pi - \gamma, \frac{\pi}{2} - \beta\right). \tag{10}$$

The second equality is due to the fact that γ has a period of 2π and β has a period of $\pi/2$. From Equation (10), one would expect that the landscape of F_p beyond $\gamma = \pi$ is the image of the rotation by 180° of the landscape within $\gamma = \pi$ (corresponding to the reflection of both $\gamma = \pi$ and $\beta = \pi/4$). Therefore, in general, the optimization can be completed in the bounds $\Phi_p \in [0, \pi)^p \times [0, \pi/2)^p$ due to the redundancies in the landscape, i.e., one part of the landscape being the image of another. Figure 1c shows the visualization of the $p = 1$ expectation landscape for a 10-node Erdös-Rényi graph. It is observed that the maximum point (colored in blue) is redundant beyond $\gamma_1 = \pi$ and $\beta_1 = \pi/2$. Moreover, in regular graphs, there are symmetries in $e^{-i\pi H_z} = Z^{\otimes n}$ for odd-degree regular graphs and $e^{-i\pi H_z} = \mathbb{1}$ for even-degree regular graphs. Thus, the optimization bounds can be further restricted to $[0, \pi/2)^p \times [0, \pi/2)^p$ for unweighted regular graphs. The details for the periodicity and symmetries are mainly discussed in [27,33,34], and we include the derivations in Appendix A.

4. Bilinear Strategy

Using the properties discussed in Section 3, we devise a strategy that is depth-progressive, i.e., the optimization is completed depth-by-depth up to the desired depth p. We utilize the fact that in the bounded search space, the optimal parameters undergo smooth changes, as shown in the patterns of the adiabatic path and the non-optimality. Our strategy tries to reproduce the adiabatic path and the non-optimality patterns so that we can generate initial parameters that are near optimal. Therefore, we use the difference in previous optimal parameters to predict the initial points for the new parameters, i.e., the parameters for the next depth. Following the adiabatic path, we can predict ϕ_{j+1}^p using $\Delta_{j,j-1}^p \equiv \phi_j^p - \phi_{j-1}^p$. For the non-optimality, we can predict ϕ_j^{p+1} using $\Delta_j^{p,p-1} \equiv \phi_j^p - \phi_j^{p-1}$. We define $\Delta_{i,j}$ as the difference between the parameters ϕ_i and ϕ_j, and this works the same way for the superscript. We call this the bilinear strategy as it involves the linear differences of two directions: j and p.

We explain the mechanism of our strategy. First, we can use any exhaustion method to find the optima for $p = 1$ and $p = 2$ within the specified bounds $\Phi_p \in [\gamma_{\min}, \gamma_{\max})^p \times [\beta_{\min}, \beta_{\max})^p$. This is to establish the base for our strategy, where we can take the difference between two sets of optimal parameters. The bounds are chosen such that there are no redundant optimals in the search space, as mentioned in Section 3.3, so that we can capture the pattern. We start applying the strategy from $p = 3$. The parameters with indices up to $j = p - 2$ are extrapolated using the pattern of non-optimality:

$$\forall j \leq p - 2, \quad \phi_j^p = \phi_j^{p-1} + \Delta_j^{p-1,p-2} \qquad (11)$$
$$= 2\phi_j^{p-1} - \phi_j^{p-2}.$$

The current parameter ϕ_j^p is extended from the previous parameter ϕ_j^{p-1} by adding the difference between the previous two parameters $\Delta_j^{p-1,p-2} = \phi_j^{p-1} - \phi_j^{p-2}$. Note that Δ can be either positive or negative, which determines the direction of the extrapolation. This agrees with the monotonous change in optimal parameters. For the parameters with index $j = p - 1$, we want to use a relation similar to Equation (11). However, the parameter ϕ_{p-1}^{p-2} does not exist, so we take the difference from the previous index $j = p - 2$ instead:

$$\phi_{p-1}^p = \phi_{p-1}^{p-1} + \Delta_{p-2}^{p-1,p-2}. \qquad (12)$$

For the newly added parameter $j = p$, it is predicted using the adiabatic path pattern:

$$\phi_p^p = \phi_{p-1}^p + \Delta_{p-1,p-2}^p \qquad (13)$$
$$= 2\phi_{p-1}^p - \phi_{p-2}^p.$$

If the parameters produced in Equations (11)–(13) are out of the bounds specified, we will take the boundary value of $\phi_{\min}$ or $\phi_{\max}$ (whichever is nearer) to replace them. After the process, the initial parameters $\Phi_p = (\gamma_1, \ldots, \gamma_p, \beta_1, \ldots, \beta_p)$ for p will be obtained, and it is optimized to find the optimal at p. This entire process is summarized in Algorithm 1. A visualization diagram of the strategy is also shown in Figure 2.
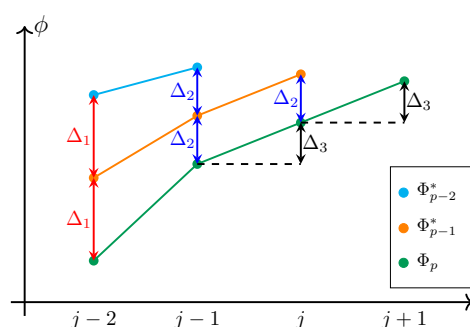

Algorithm 1 Bilinear initialization

1: **Input:** Φ_1^* and Φ_2^*. $\Phi_p \in [\gamma_{\min}, \gamma_{\max})^p \times [\beta_{\min}, \beta_{\max})^p$.
2: **for** $p := 3 \ldots q$ **do**
3:     Build the initial parameters Φ_p:
4:     **for** $j := 1 \ldots p$ **do**
5:         **if** $j \leq p - 2$ **then**
6:             $\phi_j^p \leftarrow \phi_j^{p-1} + \Delta_j^{p-1,p-2}$
7:         **else if** $j = p - 1$ **then**
8:             $\phi_j^p \leftarrow \phi_j^{p-1} + \Delta_{j-1}^{p-1,p-2}$
9:         **else if** $j = p$ **then**
10:            $\phi_j^p \leftarrow \phi_{j-1}^p + \Delta_{j-1,j-2}^p$
11:         **end if**
12:     **end for**
13:     Initialize QAOA with Φ_p and perform bounded optimization.
14: **end for**
15: **Output:** Φ_p^* and $F_p(\Phi_p^*)$ for each p up to q.

Figure 2. Visualization of the bilinear strategy. Δ_1, Δ_2, and Δ_3 correspond to the values calculated in Equations (11), (12), and (13) respectively. Δ_1 and Δ_2 represent the change due to the non-optimality. Δ_3 represents the change due to the adiabatic path. $\mathbf{\Phi}_p$ is the new initial parameters extrapolated from $\mathbf{\Phi}_{p-1}^*$ and $\mathbf{\Phi}_{p-2}^*$.
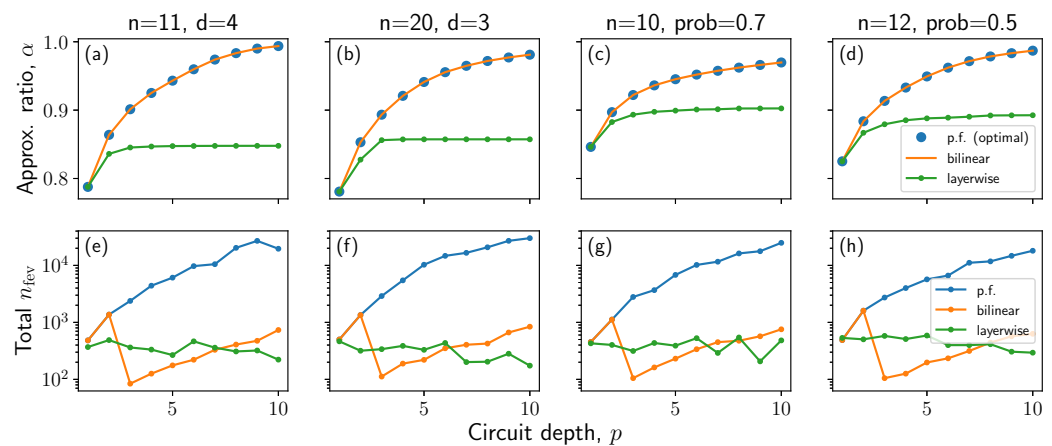
5. Results

We use statevector simulation and apply the bilinear strategy to solving the Max-cut problem for regular graphs and Erdös-Rényi graphs. The performance of the strategy is evaluated on 30 non-isomorphic instances of different classes of graphs up to the number of nodes $n = 20$, which include the 3-regular, 4-regular graphs, and Erdös-Rényi graphs with different edge probabilities. For the regular graphs, we optimize the parameters within the bound $[0, \pi/2)^p \times [0, \pi/2)^p$, whereas the Erdös-Rényi graphs are optimized within $[0, \pi)^p \times [0, \pi/2)^p$. Here, we only show the results for 4 of the instances in Figure 3, but the trends discussed also apply to all other instances unless otherwise stated.

We compare the approximation ratio α obtained from the bilinear strategy with our previously proposed parameter fixing strategy [26]. From Figure 3a–d, it can be observed that the α produced by the bilinear strategy traces the optimal α (found by parameters fixing) with minimal error. The results of the layerwise strategy are also plotted. Besides the projectors completed in the previous work [28], we found out that for the Max-cut Hamiltonian, α also saturates at a certain p due to the non-optimality of the parameters.

In Figure 3e–h, we compare the number of function evaluations n_{fev} before the convergence of the Limited-memory BFGS Bounded (L-BFGS-B) [35] optimizer for the strategies. It is the number of function calls to the quantum circuit to compute the expectation in Equation (3), and less n_{fev} usually means less quantum and classical resources used. For parameter fixing and layerwise, which need multiple trials to ensure success, we consider the total n_{fev} for 20 trials. The results show that the n_{fev} required by the bilinear initial points is always less than that of the parameters fixed by an order of 10^2 to 10^3 for $p \geq 3$. This clearly shows the advantage of the bilinear strategy on the optimization cost, as only a single trial is required. For $p = 1$ and $p = 2$, the n_{fev}'s are the same, as we used parameter fixing to search for the optima. As for layerwise, the n_{fev}'s are relatively small, as only two parameters are optimized for each p. It is observed that for small depths up to $p = 6$, even the bilinear strategy costs less than layerwise, and the cost grows with p as the number of optimization variables increases.

On the other hand, we also consider the case where the bilinear strategy fails, where the initial points do not follow the monotonous trend of the adiabatic path. One of the examples is the odd-regular graph. We mentioned in Section 3.3 that one should take the bound $[0, \pi/2)^p \times [0, \pi/2)^p$ for regular graphs to avoid redundancies. However, if one tries to take the bound $[0, \pi)^p \times [0, \pi/2)^p$, which is considered the general bound for unweighted Max-cut, one has a chance to fall into the starting point in $\gamma_1 \in [\pi/2, \pi)$ for $p = 1$ (shown in Figure 4a). In this case, the optimal parameters will not follow the adiabatic path as shown in Figure 1a. Figure 4b shows that for this non-adiabatic starting point, the optimal β's oscillate back and forth instead. In fact, this point is symmetric to the adiabatic start $\gamma_1 \in [0, \pi/2)$. We explain this odd-regular symmetry, including the β oscillation in Appendix B.

Figure 3. Comparison of the results for parameters fixing, layerwise, and the newly proposed bilinear strategy, for 4 graph instances extracted from the 30 graph instances evaluated. *n* is the number of nodes/vertices of the graph, *d* is the degree for regular graphs, 'prob' is the edge probability for Erdös-Rényi graphs. (**a–d**) show the changes in the approximation ratio α against *p*. (**e–h**) show the n_{fev} required before convergence at different *p*'s for the L-BFGS-B optimizer (log scale). For parameters fixing and layerwise, the n_{fev} is the total of 20 trials.

Figure 4c shows the result of a bilinear strategy with a non-adiabatic start for a 10-node 3-regular graph. The α produced by the bilinear strategy traces the optimal until $p = 7$, where it deviates after $p = 8$. This shows that the bilinear strategy is still effective to some extent, even for non-adiabatic starts.
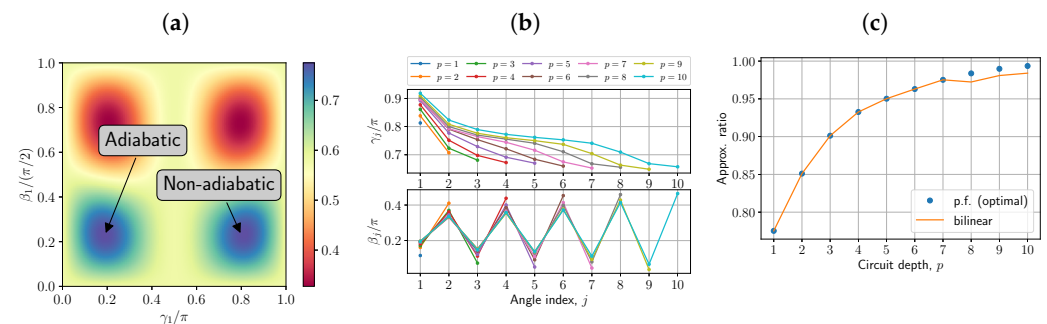
Figure 4. (**a**) $p = 1$ normalized expectation (i.e., α) landscape for a 10-node 3-regular graph showing multiple maxima in $\gamma_1 \in [0, \pi)$. When used as a starting point in the bilinear strategy, the maximum on the left follows the adiabatic path, whereas the maximum on the right does not follow the adiabatic path. (**b**) The variation of the optimal parameters with a non-adiabatic start. Unlike the adiabatic start, the β's oscillate back and forth. (**c**) The effect of the bilinear strategy under the non-adiabatic start.

6. Conclusions

To conclude, we have studied the patterns in the optimal parameters of QAOA for the unweighted Max-cut problem in two directions, namely the angle index *j* and the circuit depth *p*. We call the variation against *j* and *p* the *adiabatic path* and the *non-optimality* respectively. By leveraging these properties, we devise the depth-progressive bilinear strategy, in which the optimization is completed for each depth until the desired depth. The bilinear strategy utilizes the optimal parameters from the previous two depths, Φ^*_{p-2} and Φ^*_{p-1}, to initialize the parameters for the current depth Φ_p.

We have demonstrated the effectiveness of the bilinear strategy by comparing it with the parameter fixing strategy [26] and the layerwise [28] strategy on 30 non-isomorphic random regular and Erdös-Rényi graphs. The results show bilinear is able to trace the optimal approximation ratio α found by parameter fixing. While we have also observed premature saturation occurring in the Max-cut Hamiltonian for layerwise. It is also found

out that the number of function evaluations n_{fev} of bilinear is less than that of parameter fixing due to its single prediction.

The bilinear strategy is advantageous against most other strategies [27,28] that usually require multiple trials to ensure success, including the parameter fixing strategy. It only requires the optimization of one set of initial parameters at each circuit depth. The bilinear strategy also requires knowledge of the bounds for the optimization to avoid redundancies in the search space, thereby ensuring success. However, we have considered the case where it fails when initialized from a "non-adiabatic" point. Numerically, for a particular 3-regular graph, it is still capable of tracing the optimal until circuit depth $p = 7$.

We suggest some potential work that can be pursued in the future. Since the new prediction is extrapolated from the change in optimal parameters, we can increase the depth step of the bilinear strategy for less optimization cost. In this work, we have shown using the depth step of 1. One can, for example, increase the depth step to 2 ($p = 2, 4, 6, \ldots$) while progressing to larger circuit depths. Although not tested, the bilinear strategy is expected to perform on different kinds of problems (different H_z) in which their optimal parameters follow the adiabatic path and non-optimality, which is believed to be true for QAOA. This is also a good future work to explore.

Author Contributions: Conceptualization, X.L. and N.A.; methodology, X.L. and Y.S.; software, X.L. and N.X.; validation, X.L. and N.X.; formal analysis, X.L. and N.A.; investigation, N.X. and Y.S.; resources, N.X. and Y.S.; data curation, N.X.; writing—original draft preparation, X.L., N.X. and Y.S.; writing—review and editing, X.L., D.C. and N.A.; visualization, X.L. and N.X.; supervision, D.C. and N.A.; project administration, D.C. and N.A. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are available on reasonable request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Properties of QAOA Max-Cut

The QAOA for the Max-cut problem is highly periodic and symmetric. This is addressed in several works [27,33,34]. In this section, we derive the properties that help us to avoid global optima redundancies and to determine the bounds for the optimization.

Theorem A1 (Angle-reversal symmetry of QAOA). *The expectation of QAOA stays the same when its angles (parameters) are negated.*

$$F_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = F_p(-\boldsymbol{\gamma}, -\boldsymbol{\beta}), \tag{A1}$$

for any circuit depth p. This is true for any Hermitian mixer and problem Hamiltonian H_x and H_z.

Proof. We use the fact that the expectation $F(\boldsymbol{\gamma}, \boldsymbol{\beta})$ is real, so its complex conjugate is just itself: $F(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \overline{F(\boldsymbol{\gamma}, \boldsymbol{\beta})}$. Hence, for any Hermitian matrices H_x and H_z,

$$F_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \overline{F_p(\boldsymbol{\gamma}, \boldsymbol{\beta})} \tag{A2}$$

$$\langle +|^{\otimes n} e^{i\beta_1 H_x} e^{i\gamma_1 H_z} \cdots H_z \cdots e^{-i\beta_1 H_x} e^{-i\gamma_1 H_z} |+\rangle^{\otimes n}$$
$$= \langle +|^{\otimes n} e^{-i\beta_1 H_x} e^{-i\gamma_1 H_z} \cdots H_z \cdots e^{i\beta_1 H_x} e^{i\gamma_1 H_z} |+\rangle^{\otimes n} \tag{A3}$$

$$F_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = F_p(-\boldsymbol{\gamma}, -\boldsymbol{\beta}). \tag{A4}$$

$\square$

Theorem A2 (General periodicity and symmetry for unweighted Max-cut). *The expectation function of the unweighted Max-cut problem for any graph has a period of 2π w.r.t. the parameter(s) $\boldsymbol{\gamma}$, and a period of $\pi/2$ w.r.t. the parameter(s) $\boldsymbol{\beta}$.*

$$F_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = F_p(\boldsymbol{\gamma} + 2\pi, \boldsymbol{\beta}) \tag{A5}$$

$$= F_p\left(\boldsymbol{\gamma}, \boldsymbol{\beta} + \frac{\pi}{2}\right) \tag{A6}$$

$$= F_p\left(\boldsymbol{\gamma} + 2\pi, \boldsymbol{\beta} + \frac{\pi}{2}\right), \tag{A7}$$

where $\boldsymbol{\phi} + c$ means a shift of every element in the parameter vector $\boldsymbol{\phi}$ by a scalar c, $\boldsymbol{\phi} + c = (\phi_1 + c, \phi_2 + c, \ldots, \phi_p + c)$.

Combining the angle-reversal and the periodicity creates a symmetry on the expectation function:

$$F_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = F_p\left(2\pi - \boldsymbol{\gamma}, \frac{\pi}{2} - \boldsymbol{\beta}\right). \tag{A8}$$

Proof. It is known that $e^{-i(2\pi)H_z} = \mathbb{1}$. Consider the ansatz with γ shifted by 2π:

$$\left|\psi_p(\boldsymbol{\gamma} + 2\pi, \boldsymbol{\beta})\right\rangle = \prod_{j=1}^{p} e^{-i\beta_j H_x} e^{-i\gamma_j H_z} e^{-i(2\pi)H_z} |+\rangle^{\otimes n} \tag{A9}$$

$$= \prod_{j=1}^{p} e^{-i\beta_j H_x} e^{-i\gamma_j H_z} \mathbb{1} |+\rangle^{\otimes n} \tag{A10}$$

$$= \prod_{j=1}^{p} e^{-i\beta_j H_x} e^{-i\gamma_j H_z} |+\rangle^{\otimes n} \tag{A11}$$

$$= \left|\psi_p(\boldsymbol{\gamma}, \boldsymbol{\beta})\right\rangle. \tag{A12}$$

Since the ansatz stays the same under the shift, so is the expectation, hence proving Equation (A5).

It is known that $e^{i(\pi/2)H_x} = X^{\otimes n}$ commutes with the QAOA operators $e^{-i\beta H_x}$ and $e^{-i\gamma H_z}$. The former commutation is due to the rotation of the same Pauli. The latter is due to the symmetry of the eigenstates of H_z, e.g., the eigenvalue (or cut-value) of $|0110\rangle$ is equal to the eigenvalue of $|1001\rangle$. The eigenvalues are invariant under the bit-flip operation $X^{\otimes n}$. Consider the ansatz with β shifted by $\pi/2$:

$$\left|\psi_p\left(\boldsymbol{\gamma}, \boldsymbol{\beta} + \frac{\pi}{2}\right)\right\rangle = \prod_{j=1}^{p} e^{-i\beta_j H_x} e^{-i(\pi/2)H_x} e^{-i\gamma_j H_z} |+\rangle^{\otimes n} \tag{A13}$$

$$= \prod_{j=1}^{p} e^{-i\beta_j H_x} X^{\otimes n} e^{-i\gamma_j H_z} |+\rangle^{\otimes n}. \tag{A14}$$

Since $X^{\otimes n}$ commutes through the operators, we can move all of them to the rightmost before the initial state $|+\rangle^{\otimes n}$. Note that since $|+\rangle^{\otimes n}$ is the eigenstate of $X^{\otimes n}$, it will not change the initial state. Therefore, we have

$$\left|\psi_p\left(\boldsymbol{\gamma}, \boldsymbol{\beta} + \frac{\pi}{2}\right)\right\rangle = \prod_{j=1}^{p} e^{-i\beta_j H_x} e^{-i\gamma_j H_z} X^{\otimes n} |+\rangle^{\otimes n} \tag{A15}$$

$$= \prod_{j=1}^{p} e^{-i\beta_j H_x} e^{-i\gamma_j H_z} |+\rangle^{\otimes n} \tag{A16}$$

$$= \left|\psi_p(\boldsymbol{\gamma}, \boldsymbol{\beta})\right\rangle. \tag{A17}$$

Hence proving Equation (A6).

Combining Equations (A1) and (A7), Equation (A8) can be derived.

$$F_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = F_p(-\boldsymbol{\gamma}, -\boldsymbol{\beta}) = F_p\left(2\pi - \boldsymbol{\gamma}, \frac{\pi}{2} - \boldsymbol{\beta}\right), \tag{A18}$$

showing the expectation function reflects over the axis $\gamma = \pi$, and then the axis $\beta = \pi/4$. This is equivalent to a 180° rotation about the point $(\vec{\pi}, \vec{\pi}/4)$, where $\vec{\pi} = (\pi, \pi, \ldots, \pi)$. $\square$

Remark A1. *The proof of Theorem A2 shows that the periodicity is not only true for the shift for the parameter vector $\boldsymbol{\phi}$, but also true for any arbitrary number of ϕ_j shift by its corresponding period, as any number of the operator $\mathbb{1}$ (or $X^{\otimes n}$) will still be canceled out.*

Theorem A3 (Periodcity and symmetry for even-regular graphs). *For the Max-cut of even-regular graphs, the expectation has a period of π w.r.t. the parameter(s) $\boldsymbol{\gamma}$, which is shorter than the general period.*

$$F_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = F_p(\boldsymbol{\gamma} + \pi, \boldsymbol{\beta}) \tag{A19}$$

$$= F_p\left(\boldsymbol{\gamma} + \pi, \boldsymbol{\beta} + \frac{\pi}{2}\right). \tag{A20}$$

Due to the angle-reversal, the expectation of even-regular graphs has the symmetry

$$F_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = F_p\left(\pi - \boldsymbol{\gamma}, \frac{\pi}{2} - \boldsymbol{\beta}\right). \tag{A21}$$

Proof. For the H_z of even-regular graphs, $e^{-i\pi H_z} = \mathbb{1}$. Consider the ansatz with γ shifted by π:

$$\left|\psi_p(\boldsymbol{\gamma} + \pi, \boldsymbol{\beta})\right\rangle = \prod_{j=1}^{p} e^{-i\beta_j H_x} e^{-i\gamma_j H_z} e^{-i\pi H_z} |+\rangle^{\otimes n} \tag{A22}$$

$$= \prod_{j=1}^{p} e^{-i\beta_j H_x} e^{-i\gamma_j H_z} |+\rangle^{\otimes n} \tag{A23}$$

$$= \left|\psi_p(\boldsymbol{\gamma}, \boldsymbol{\beta})\right\rangle. \tag{A24}$$

Also, by combining Equations (A1) and (A20), we can derive the symmetry.

$$F_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = F_p(-\boldsymbol{\gamma}, -\boldsymbol{\beta}) = F_p\left(\pi - \boldsymbol{\gamma}, \frac{\pi}{2} - \boldsymbol{\beta}\right). \tag{A25}$$

$\square$

From Equation (A8), if the expectation F_p has a global optimum at the point (γ^*, β^*), then it will also have a global optimum at $(2\pi - \gamma^*, \frac{\pi}{2} - \beta^*)$. Therefore, to avoid redundancies in general graphs, it is suitable to set the optimization bounds as $\boldsymbol{\Phi}_p \in [0, \pi)^p \times [0, \pi/2)^p$. On the other hand, for even-regular graphs, Equation (A21) shows that an extra symmetry exists at $(\pi - \gamma^*, \frac{\pi}{2} - \beta^*)$ due to the shortened period, so the suitable bounds are $\boldsymbol{\Phi}_p \in [0, \pi/2)^p \times [0, \pi/2)^p$.

Appendix B. Non-Adiabatic Path for Odd-Regular Graphs

It is known from the previous section, the QAOA for Max-cut has symmetric properties in its expectation function. In other words, we know that, generally, if the expectation F_p has a global optimum at the point (γ^*, β^*), then it will also have a global optimum at $(2\pi - \gamma^*, \frac{\pi}{2} - \beta^*)$. For even-regular graphs, they have an extra symmetry at $(\pi - \gamma^*, \frac{\pi}{2} - \beta^*)$ due to the shortened period. Understanding this allows us to predict, for example, the other symmetric optimum will also have the adiabatic path pattern shown in Figure 1a,

except that γ_j will decrease and β_j will increase, as the original optimum and the symmetric optimum are negatively related.

However, things are a bit different in odd-regular graphs. We found out that in odd-regular graphs, the other "symmetric" optimum follows the pattern shown in Figure 4b, with smooth decrease in γ_j and oscillating β_j. In this section, we will derive the symmetry for odd-regular graphs.

Theorem A4 (Symmetry for odd-regular graphs)**.** *For odd-regular graphs, the expectation has a symmetry that follows*

$$F_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = F_p(\pi - \boldsymbol{\gamma}, \tilde{\boldsymbol{\beta}}), \tag{A26}$$

where $\tilde{\boldsymbol{\beta}}$ has elements β_j if j is odd and $(\pi/2 - \beta_j)$ if j is even:

$$\tilde{\boldsymbol{\beta}} \equiv \begin{pmatrix} \beta_1 \\ \frac{\pi}{2} - \beta_2 \\ \beta_3 \\ \frac{\pi}{2} - \beta_4 \\ \vdots \end{pmatrix}. \tag{A27}$$

Proof. For the H_z of odd-regular graphs, $e^{-i\pi H_z} = Z^{\otimes n}$. Consider the ansatz with $\boldsymbol{\gamma}$ shifted by π:

$$\left|\psi_p(\boldsymbol{\gamma} + \pi, \boldsymbol{\beta})\right\rangle = \prod_{j=1}^{p} e^{-i\beta_j H_x} e^{-i\pi H_z} e^{-i\gamma_j H_z} |+\rangle^{\otimes n} \tag{A28}$$

$$= \prod_{j=1}^{p} e^{-i\beta_j H_x} Z^{\otimes n} e^{-i\gamma_j H_z} |+\rangle^{\otimes n}. \tag{A29}$$

The expectation is thus

$$F_p(\boldsymbol{\gamma} + \pi, \boldsymbol{\beta})$$
$$= \langle +|^{\otimes n} (e^{i\gamma_1 H_z} Z^{\otimes n} e^{i\beta_1 H_x} \cdots e^{i\gamma_p H_z} Z^{\otimes n} e^{i\beta_p H_x}) H_z$$
$$(e^{-i\beta_p H_x} Z^{\otimes n} e^{-i\gamma_p H_z} \cdots e^{-i\beta_1 H_x} Z^{\otimes n} e^{-i\gamma_1 H_z}) |+\rangle^{\otimes n}. \tag{A30}$$

We know that $Z^{\otimes n}$ commutes with $e^{-i\gamma H_z}$, but does not commute with $e^{-i\beta H_x}$. However, both $Z^{\otimes n}$ and $e^{-i\beta H_x}$ are Hermitian, so

$$Z^{\otimes n} e^{-i\beta H_x} = (Z^{\otimes n} e^{-i\beta H_x})^{\dagger} \tag{A31}$$

$$= e^{i\beta H_x} Z^{\otimes n}. \tag{A32}$$

As we can see from the above, as $Z^{\otimes n}$ moves through $e^{-i\beta H_x}$, the sign of β changes. The goal here is to use this property to move the Z's in Equation (A30) towards the center H_z, so that the Z's from the left and right cancels out at the center, since they commute with H_z. As $Z^{\otimes n}$ moves through the equation, the sign of β's that the operator passes through will be toggled. It is not difficult to see that after all the Z's have been canceled out, the resulting vector of $\boldsymbol{\beta}$ will have elements with alternating signs. We define the resulting vector $\boldsymbol{\beta}'$:

$$\boldsymbol{\beta}' \equiv \begin{pmatrix} -\beta_1 \\ \beta_2 \\ -\beta_3 \\ \beta_4 \\ \vdots \end{pmatrix}. \tag{A33}$$

Hence, we have

$$F_p(\gamma + \pi, \boldsymbol{\beta}) = F_p(\gamma, \boldsymbol{\beta}'). \tag{A34}$$

Shifting γ on both sides of the equation by $-\pi$ and applying the angle-reversal on the RHS yields

$$F_p(\gamma, \boldsymbol{\beta}) = F_p(\gamma - \pi, \boldsymbol{\beta}') \tag{A35}$$

$$= F_p(\pi - \gamma, -\boldsymbol{\beta}'). \tag{A36}$$

To tidy up this, we want to get the values of $-\boldsymbol{\beta}'$ in the range of $[0, \pi/2)^p$. As mentioned in Remark A1, the expectation is left unchanged if we shift any number of the parameter(s) by its period. Thus, we shift the negative elements (even indices) in $-\boldsymbol{\beta}'$ by $\pi/2$, so if $\beta_j \in [0, \pi/2)$, then $(\pi/2 - \beta_j)$ will also be within the range. We use a new symbol $\tilde{\boldsymbol{\beta}}$ to denote this parameter vector, arriving at Equation (A27). Hence,

$$F_p(\gamma, \boldsymbol{\beta}) = F_p(\pi - \gamma, -\boldsymbol{\beta}') = F_p(\pi - \gamma, \tilde{\boldsymbol{\beta}}). \tag{A37}$$

□

Thus, Equations (A26) and (A27) explain the oscillation of β values shown in Figure 4b. By restricting the optimization bounds to $\boldsymbol{\Phi}_p \in [0, \pi/2)^p \times [0, \pi/2)^p$, the non-adiabatic path for odd-regular graphs can be avoided.

References

1. Farhi, E.; Goldstone, J.; Gutmann, S. A Quantum Approximate Optimization Algorithm. *arXiv* **2014**, arXiv:1411.4028.
2. Crooks, G. Performance of the Quantum Approximate Optimization Algorithm on the Maximum Cut Problem. *arXiv* **2018**, arXiv:1811.08419.
3. Guerreschi, G.G.; Matsuura, A.Y. QAOA for Max-Cut requires hundreds of qubits for quantum speed-up. *Sci. Rep.* **2019**, *9*, 6903. [CrossRef]
4. Farhi, E.; Harrow, A.W. Quantum Supremacy through the Quantum Approximate Optimization Algorithm. *arXiv* **2019**, arXiv:1602.07674.
5. Moussa, C.; Calandra, H.; Dunjko, V. To quantum or not to quantum: Towards algorithm selection in near-term quantum optimization. *Quantum Sci. Technol.* **2020**, *5*, 044009. [CrossRef]
6. Marwaha, K. Local classical MAX-CUT algorithm outperforms $p = 2$ QAOA on high-girth regular graphs. *Quantum* **2021**, *5*, 437. [CrossRef]
7. Basso, J.; Farhi, E.; Marwaha, K.; Villalonga, B.; Zhou, L. *The Quantum Approximate Optimization Algorithm at High Depth for MaxCut on Large-Girth Regular Graphs and the Sherrington-Kirkpatrick Model*; Schloss Dagstuhl—Leibniz-Zentrum für Informatik: Wadern, Germany, 2022. [CrossRef]
8. Akshay, V.; Philathong, H.; Campos, E.; Rabinovich, D.; Zacharov, I.; Zhang, X.M.; Biamonte, J. On Circuit Depth Scaling For Quantum Approximate Optimization. *arXiv* **2022**, arXiv:2205.01698. [CrossRef]
9. Wurtz, J.; Love, P. MaxCut quantum approximate optimization algorithm performance guarantees for $p > 1$. *Phys. Rev. A* **2021**, *103*, 042612. [CrossRef]
10. Uvarov, A.V.; Biamonte, J.D. On barren plateaus and cost function locality in variational quantum algorithms. *J. Phys. A Math. Theor.* **2021**, *54*, 245301. [CrossRef]
11. Cerezo, M.; Sone, A.; Volkoff, T.; Cincio, L.; Coles, P.J. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nat. Commun.* **2021**, *12*, 1791. [CrossRef]
12. Wang, S.; Fontana, E.; Cerezo, M.; Sharma, K.; Sone, A.; Cincio, L.; Coles, P.J. Noise-induced barren plateaus in variational quantum algorithms. *Nat. Commun.* **2021**, *12*, 6961. [CrossRef] [PubMed]
13. Grant, E.; Wossnig, L.; Ostaszewski, M.; Benedetti, M. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum* **2019**, *3*, 214. [CrossRef]
14. Zhu, L.; Tang, H.L.; Barron, G.S.; Calderon-Vargas, F.A.; Mayhall, N.J.; Barnes, E.; Economou, S.E. An adaptive quantum approximate optimization algorithm for solving combinatorial problems on a quantum computer. *arXiv* **2020**, arXiv:2005.10258. [CrossRef]
15. Sack, S.H.; Serbyn, M. Quantum annealing initialization of the quantum approximate optimization algorithm. *arXiv* **2021**, arXiv:2101.05742. [CrossRef]
16. Shaydulin, R.; Safro, I.; Larson, J. Multistart Methods for Quantum Approximate optimization. In Proceedings of the 2019 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA, 24–26 September 2019. [CrossRef]

17. Shaydulin, R.; Wild, S.M. Exploiting Symmetry Reduces the Cost of Training QAOA. *IEEE Trans. Quantum Eng.* **2021**, *2*, 1–9. [CrossRef]

18. Alam, M.; Ash-Saki, A.; Ghosh, S. Accelerating Quantum Approximate Optimization Algorithm Using Machine Learning. In Proceedings of the 23rd Conference on Design, Automation and Test in Europe, DATE '20, Grenoble, France, 9–13 March 2020; EDA Consortium: San Jose, CA, USA, 2020; pp. 686–689.

19. Moussa, C.; Wang, H.; Bäck, T.; Dunjko, V. Unsupervised strategies for identifying optimal parameters in Quantum Approximate Optimization Algorithm. *EPJ Quantum Technol.* **2022**, *9*, 11. [CrossRef]

20. Amosy, O.; Danzig, T.; Porat, E.; Chechik, G.; Makmal, A. Iterative-Free Quantum Approximate Optimization Algorithm Using Neural Networks. *arXiv* **2022**, arXiv:2208.09888. [CrossRef]

21. Khairy, S.; Shaydulin, R.; Cincio, L.; Alexeev, Y.; Balaprakash, P. Learning to Optimize Variational Quantum Circuits to Solve Combinatorial Problems. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 2367–2375. [CrossRef]

22. Deshpande, A.; Melnikov, A. Capturing Symmetries of Quantum Optimization Algorithms Using Graph Neural Networks. *Symmetry* **2022**, *14*, 2593. [CrossRef]

23. Brandao, F.G.S.L.; Broughton, M.; Farhi, E.; Gutmann, S.; Neven, H. For Fixed Control Parameters the Quantum Approximate Optimization Algorithm's Objective Function Value Concentrates for Typical Instances. *arXiv* **2018**, arXiv:1812.04170.

24. Akshay, V.; Rabinovich, D.; Campos, E.; Biamonte, J. Parameter Concentration in Quantum Approximate Optimization. *arXiv* **2021**, arXiv:2103.11976. [CrossRef]

25. Galda, A.; Liu, X.; Lykov, D.; Alexeev, Y.; Safro, I. Transferability of optimal QAOA parameters between random graphs. *arXiv* **2021**, arXiv:2106.07531.

26. Lee, X.; Saito, Y.; Cai, D.; Asai, N. Parameters Fixing Strategy for Quantum Approximate Optimization Algorithm. In Proceedings of the 2021 IEEE International Conference on Quantum Computing and Engineering (QCE), Broomfield, CO, USA, 17–22 October 2021. [CrossRef]

27. Zhou, L.; Wang, S.T.; Choi, S.; Pichler, H.; Lukin, M.D. Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices. *Phys. Rev. X* **2020**, *10*, 021067. [CrossRef]

28. Campos, E.; Rabinovich, D.; Akshay, V.; Biamonte, J. Training saturation in layerwise quantum approximate optimization. *Phys. Rev. A* **2021**, *104*, L030401. [CrossRef]

29. Karp, R. Reducibility among combinatorial problems. In *Complexity of Computer Computations*; Miller, R.; Thatcher, J., Eds.; Plenum Press: New York, NY, USA, 1972; pp. 85–103.

30. Farhi, E.; Goldstone, J.; Gutmann, S.; Sipser, M. Quantum Computation by Adiabatic Evolution. *arXiv* **2000**, arXiv:0001106.

31. Cook, J.; Eidenbenz, S.; Bärtschi, A. The Quantum Alternating Operator Ansatz on Maximum k-Vertex Cover. In Proceedings of the 2020 IEEE International Conference on Quantum Computing and Engineering (QCE), Denver, CO, USA, 12–16 October 2020; pp. 83–92.

32. Willsch, M.; Willsch, D.; Jin, F.; De Raedt, H.; Michielsen, K. Benchmarking the quantum approximate optimization algorithm. *Quantum Inf. Process.* **2020**, *19*, 197. [CrossRef]

33. Lotshaw, P.C.; Humble, T.S.; Herrman, R.; Ostrowski, J.; Siopsis, G. Empirical performance bounds for quantum approximate optimization. *Quantum Inf. Process.* **2021**, *20*, 403. [CrossRef]

34. Shaydulin, R.; Lotshaw, P.C.; Larson, J.; Ostrowski, J.; Humble, T.S. Parameter Transfer for Quantum Approximate Optimization of Weighted MaxCut. *arXiv* **2022**, arXiv:2201.11785. [CrossRef]

35. Morales, J.L.; Nocedal, J. Remark on "Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound Constrained Optimization". *ACM Trans. Math. Softw.* **2011**, *38*, 550–560. [CrossRef]