Quantum Optimization with a Novel Gibbs Objective Function and Ansatz Architecture Search

Li Li, Minjie Fan, Marc Coram, and Patrick Riley
Google Research Applied Science, Mountain View, CA 94043, USA
Stefan Leichenauer
X, The Moonshot Factory, Mountain View, CA 94043, USA
(Dated: September 18, 2019)

The Quantum Approximate Optimization Algorithm (QAOA) is a standard method for combinatorial optimization with a gate-based quantum computer. The QAOA consists of a particular ansatz for the quantum circuit architecture, together with a prescription for choosing the variational parameters of the circuit. We propose modifications to both. First, we define the Gibbs objective function and show that it is superior to the energy expectation value for use as an objective function in tuning the variational parameters. Second, we describe an Ansatz Architecture Search (AAS) algorithm for searching the discrete space of quantum circuit architectures near the QAOA to find a better ansatz. Through the AAS we find quantum circuits with the same number of variational parameters as the QAOA but which have improved performance on certain Ising-type problems. This opens a new research field of quantum circuit architecture design for quantum optimization algorithms.

I. INTRODUCTION

Approximate combinatorial optimization via quantum computing is an active area of research. Like similar algorithms, the Quantum Approximate Optimization Algorithm (QAOA) \[1, 2\] requires optimizing variational parameters. We assert the correct way to frame the goal of this optimization is in the probably approximately correct framework \[3\]. However, the standard objective function for QAOA does not reflect this goal. We introduce a new Gibbs objective function and show its superiority in the probably approximately correct sense.

We then proceed to try and find a new circuit ansatz that is closely related to the general QAOA circuit. We define Ansatz Architecture Search (AAS) and show that for certain Ising-problems, superior circuits exist with a notable improvement for the new Gibbs objective function. Figure 1 shows two exemplary instances and the improvement of probability of low energy with ansatzes found by AAS with the Gibbs objective function.

The fact that these superior circuits exist opens a new field of research to design a search procedure for optimal ansatzes given specific problems.

A. Quantum Approximate Optimization Algorithm

The Quantum Approximate Optimization Algorithm (QAOA) \[1, 2\] is a general-purpose algorithm for finding a low-energy state of a given computational-basis Hamiltonian. This is a classical problem which can be combinatorially very difficult, but the hope is that using a quantum computer to find the solution might be more efficient than a classical method. The QAOA has performance guarantees in certain combinatorial problems \[2\] and quantum state transfer \[4\], and it has been shown that in general the output of the QAOA is not classically simulable \[5\]. The QAOA and related algorithms offer a promising avenue for near-term applications of quantum computers \[6\], though it is not yet clear if a quantum advantage is achievable in a practical use case. In pur-

FIG. 1. Particular instances of random couplings and the structures of the associated QAOA and best sparse ansatzes for grid (first row) and complete graph (second row) problems with the Gibbs objective function. On the left, each edge in the instance graph is colored by its coupling from blue \((-1)\) to red \((1)\). We show the relative improvement of the probability of low energy compared to the usual prescription of the QAOA. Sparsity is the number of two-qubit gates in the ansatz graph divided by the number in the instance graph.
suit of that goal we should be sure that we are using the quantum computer in the best possible way.

The QAOA specifies a particular quantum circuit architecture which depends on the Hamiltonian we are trying to optimize. The prescription is very similar to a discretized adiabatic algorithm. The quantum state produced by the QAOA at level \( p \) is

\[
|\psi\rangle = e^{i\beta_p X} e^{i\gamma_p E} \cdots e^{i\beta_1 X} e^{i\gamma_1 E} H^{\otimes n}|0^n\rangle
\]  

(1)

The \( 2p \) parameters \( \beta \) and \( \gamma \) are variational parameters of the model. Here \( E \) is the Hamiltonian operator we are trying to minimize, and by assumption it only depends on the Pauli-Z operators acting on the \( n \) qubits. The functional form of \( E \) determines which quantum gates are needed to construct the QAOA circuit. Below we will consider cases where only two-qubit gates are required. We will also focus on \( p = 1 \) only for simplicity.

The usual prescription for determining the continuous parameters of the QAOA is to minimize the expectation value of the energy function. However, the true goal in most approximate optimization problems is to have a high probability of finding a solution near the optimal. That goal is not necessarily achieved by a small energy expectation value. A small energy expectation value implies that the wavefunction has support on low-energy states, but this does not maximize the probability of finding a low-energy state. In fact, there are many quantum states of large energy expectation which would perform well for our objective. Consider, for example, a state which is an equal superposition of the lowest-energy state and the highest-energy state. In that situation the expectation value of the energy would not be small, but the probability of finding a low-energy state through sampling would be high. In this work we will propose a new objective function for tuning the variational parameters of the QAOA which better represents our goals—the Gibbs objective function—and demonstrate through examples that it outperforms the energy expectation value in practice.

As our second modification of the standard prescription, we will alter the ansatz Eq. (1) by removing some of the gates used to create the \( \exp(\gamma E) \) operator. We propose AAS algorithm for searching the discrete space of ansatz architectures to find a better ansatz. We find that removing approximately one quarter of the two-qubit gates of the QAOA ansatz leads to an improvement in the performance of the algorithm. The challenge is deciding which of the two-qubit gates should be removed.

We will discuss our approaches to this problem in more detail in this paper.

B. Random Couplings Ising Model

The models we consider in this paper are Ising models. A given model \( \mathcal{I} \) is defined on a graph \( \mathcal{G} \) with \( n \) vertices \( \mathcal{V} = \{1, 2, \ldots, n\} \) and a set of undirected edges \( \mathcal{E} = \{e_{ij}\} \). Two types of graph instances are studied:

- **Grid**: A \( 4 \times 4 \) square lattice. Edges only exist between nearest-neighbor vertices. This graph contains 16 vertices and \(|\mathcal{E}| = 24\) edges. The average degree of the vertex in this graph is 3.

- **Complete Graph**: A complete graph with 10 vertices. Edges exist between any pair of vertices. This graph contains \(|\mathcal{E}| = 45\) edges. The degree of each vertex in this graph is 9.

We select these two graph types to cover the extreme case of sparse and dense graphs. Each instance consists of a set of couplings \( \mathcal{J} \). A coupling \( J_{ij} \) is assigned to each undirected edge \( e_{ij} \) between vertices \( i \) and \( j \). The Hamiltonian is written as a sum over edges:

\[
E = \sum_{e_{ij}} J_{ij} Z_i Z_j.
\]

(2)

The couplings are sampled independently from a uniform distribution,

\[
J_{ij} \sim U(-1, 1).
\]

(3)

![FIG. 2. Exact ground state energies per vertex of (a) grid and (b) complete graph instances. Black dashed lines indicate the medians of the exact energies per vertex.](image)

We denote a problem instance as \( \mathcal{I} = (\mathcal{G}^\mathcal{E}, \mathcal{J}) \). In this paper we analyze 1000 instances each of the grid and complete graph problems. In Figure 2 we plot histograms of the exact ground state energies per vertex for these instances.

As discussed in the introduction, the success of an approximate optimization algorithm is measured according to the probability of finding a low energy state. With that in mind, we evaluate the performance of the quantum circuits we discuss according to \( P(E < E_0) \), where \( P \) is the Born probability distribution of the output quantum state \( |\psi\rangle \) and \( E_0 \) is the cutoff for what we consider low energy. For definiteness, in this paper we use \( E_0 = 0.95 E_{\text{med}}(\mathcal{I}) \) as our definition, where \( E_{\text{med}}(\mathcal{I}) \) is the exact ground state energy of the given instance \( \mathcal{I} \) (which is always negative for the models we consider).
II. GIBBS OBJECTIVE FUNCTION

A. Theory

In this section we will discuss the problem of choosing the optimal values of the variational parameters. The standard prescription of minimizing the expectation value of the energy, $\langle E \rangle$, is just a proxy for maximizing $P(E < E_0)$ for some choice of low-energy cutoff $E_0$. Recent work [7] has explored using Conditional Value-at-Risk (CVaR) as the objective function. As an alternative, we propose choosing to minimize the Gibbs objective function, defined as follows:

$$f = -\log \langle e^{-\eta E} \rangle.$$  \hspace{1cm} (4)

Here $\eta > 0$ is a hyperparameter that we should tune based on the general properties of the class of problems we are considering. The function $f$ is very similar to the Gibbs free energy from statistical mechanics, which is the origin of the name.

The reason why $\langle e^{-\eta E} \rangle$ might be preferred over $\langle E \rangle$ is easily understood intuitively. The exponential profile rewards us for increasing the probability of low energy, and de-emphasizes the shape of the probability distribution at higher energies. Note that the Gibbs objective function just as easy to measure as the energy expectation value itself when the energy is diagonal in the computational basis. We just perform a different computation with our measurement samples.

The Gibbs objective function $f(\eta)$ is essentially the cumulant generating function of the energy [8]. The Taylor expansion reads $f(\eta) = \sum_{n=1}^{\infty} (-1)^{n+1} \kappa_n(E) \eta^n/n! = \mu_E \eta - \sigma_E^2 \eta^2/2 + 3/6 \kappa_3 \eta^3/6 + \cdots$. For small $\eta$, then, minimizing the Gibbs objective function is equivalent to minimizing $\mu_E = \langle E \rangle$. As $\eta$ increases, the higher-order cumulants become more important.

To better understand the Gibbs objective function, we can try to estimate the best value of the hyperparameter $\eta$. For any $\eta > 0$ the probability of low energy is bounded from above as follows:

$$P(E < E_0) \leq \langle e^{-\eta(E-E_0)} \rangle.$$  \hspace{1cm} (5)

Choosing $\eta$ to minimize the right-hand side gives the strongest inequality out of this one-parameter family. That value of $\eta$ is the one which satisfies the equation

$$E_0 = \frac{\langle E e^{-\eta E} \rangle}{\langle e^{-\eta E} \rangle}.$$  \hspace{1cm} (6)

Now, $\eta$ is meant to be a fixed hyperparameter that is maintained throughout parameter optimization, whereas the $\eta$ satisfying this equation depends functionally on the probability distribution itself. Our prescription for estimating $\eta$ is to find an approximate solution to this equation, valid for a large class of probability distributions that we may encounter during parameter optimization.

If $E_0$ is meant to be close to $E_{gs}$, then it’s clear that the interesting limit of Eq. (6) is the large-$\eta$ limit $[\text{6}]$. The first correction at large-$\eta$ to the RHS is equal to $\eta^{-1}$:

$$\frac{\langle E e^{-\eta E} \rangle}{\langle e^{-\eta E} \rangle} \approx E_{gs} + \eta^{-1}.$$  \hspace{1cm} (7)

Combined with Eq. (5), this suggests that we should set $\eta = (E_0 - E_{gs})^{-1}$. We may only able to estimate values for $E_0$ and $E_{gs}$ based on the specification of our problem, but in practice these estimates are good enough. For the problems we consider, $E_0 = 0.95E_{gs}$ and $E_{gs} \approx -1$ gives $\eta \approx 20$ as an estimate, which we use for the majority of our numerical experiments below.

Note that in the large-$\eta$/small-$(E_0 - E_{gs})$ regime one can make much stronger statements about the relationship between $\langle e^{-\eta E} \rangle$ and $P(E < E_0)$. We will sketch some of them here. In taking the large-$\eta$ limit above, we effectively approximated the probability density function for the energy, $p(E)$, by its constant term $p(E_{gs})$ near the ground-state energy. If $p(E)$ is treated as a constant, then $P(E < E_0)$ and $\langle e^{-\eta(E-E_{gs})} \rangle$ are actually equal when $\eta = (E_0 - E_{gs})^{-1}$. More generally, if $p(E)$ is well-approximated by a finite-degree polynomial in $E - E_{gs}$ with bounded coefficients, then we have the slightly weaker condition $P(E < E_0) \sim \langle e^{-\eta(E-E_{gs})} \rangle$, meaning that either quantity is bounded from above and below by constant multiples of the other. This further motivates the use of the Gibbs objective function.

![Comparison of Gibbs objective function with different η to the energy expectation objective function on QAOA ansatz.](image)

FIG. 3. Comparison of Gibbs objective function with different $\eta$ to the energy expectation objective function on QAOA ansatz. For every instance given $\eta$, we measure the probability of low energy of QAOA+Gibbs divided by QAOA+energy. The bars show the range from 5% to 95% and the horizontal segments are median. For small values of $\eta$, the Gibbs objective function is equivalent to the energy expectation value for purposes of optimization, while for large values of $\eta$ it is equivalent to maximizing the probability of finding the ground state.

---

1 We are assuming that $\eta$ is large compared to the inverse of the energy scale of the Hamiltonian, but not large compared to the gap. In other words, even when $\eta$ is large there should still be many states between $E_{gs}$ and $E_{gs} + \eta^{-1}$.
B. Numerical Experiments

To evaluate the performance of the Gibbs objective function, as well as the ansatz optimization described in the next section, we analyze 1000 instances each of the grid and complete graph Ising models described in Section III. For each instance, we optimize the variational parameters $\beta$ and $\gamma$ using the Nelder-Mead algorithm [9] to minimize either the expectation value of the energy or the Gibbs objective function, i.e., Eq. (1). The underlying circuit ansatz is either the QAOA or an optimized sparse ansatz as described in the next section. In all cases, we evaluate the algorithm performance according to the probability of finding a low energy state, $P(E < 0.95E_{gs}(I))$.

In Figure 3, we show the effect of changing the hyperparameter $\eta$ in the Gibbs objective function using the QAOA circuit ansatz. As $\eta$ increases, the probability of low energy increases before finally converging. We take this as the evidence that the large-$\eta$ regime has been reached. For these problems, our estimated value of $\eta = 20$ falls within this large-$\eta$ range, and from now on we will stick to that value. From the plot, it is clear that using the Gibbs objective function is always superior to the energy expectation value, and that a large value of the hyperparameter $\eta$ improves the performance. However, we observe that in the extreme-$\eta$ regime (not shown in the figure), e.g., $10^5$, the optimization does not converge due to the fact that the objective function is always approximately zero except when the exact ground state is sampled. This is an obstacle to efficient parameter optimization. Similar results are found using the sparse ansatz described in Section III.3 below.

Figure 4 displays the probability of finding a low energy state for each quantum optimization algorithm, denoted as $\{\text{ansatz type}$+$\{\text{objective}\}$. QAOA + energy is the original QAOA prescription and provides the baseline for comparison. The Gibbs objective function is always with $\eta = 20$, and the sparse ansatz is the subject of the next section. As shown in the scatter plots of QAOA + Gibbs vs. QAOA + energy, using the Gibbs objective function improves the solution. More significant improvement can be achieved using a sparse ansatz in addition to the Gibbs objective function, especially for complete graph instances.

III. OPTIMIZING THE ANSATZ

In this section, we will discuss alternatives to the QAOA circuit ansatz. In particular, for the Ising Hamiltonians in Eq. (2), the operator $e^{i\gamma E}$ of Eq. (1) involves a two-qubit operator for each edge in the instance graph $G^I$. We denote by $G^A$ the ansatz graph, which is obtained from $G^I$ by removing some of the edges. The associated circuit ansatz $A$ is obtained by removing from $e^{i\gamma E}$ those two-qubit operators corresponding to the edges which were removed from $G^I$. The rest of the quantum circuit remains the same as in the QAOA. This is clearly not the most general possible prescription for $G^A$, but makes use of the intuition that the QAOA ansatz $G^A = G^I$ is a good starting point for the architecture search.

In total, an ansatz $A(G^A, \beta, \gamma)$ is determined by its graph architecture $G^A$ and continuous parameters $\beta, \gamma$. The optimal ansatz graph and variational parameters for a given instance are denoted by $G^A$, $\beta, \gamma$, and they are the ones that minimize the objective function:

$$\hat{G}^A, \hat{\beta}, \hat{\gamma} = \arg\min_{G^A, \beta, \gamma} f(A(G^A, \beta, \gamma), I).$$

For each $G^A$, $A(G^A, \beta, \gamma)$ represents a family of ansatzes differing by $\beta, \gamma$. We can optimize Eq. (6) in a nested manner,

$$\hat{G}^A = \arg\min_{G^A} f(A(G^A, \hat{\beta}, \hat{\gamma}), I)$$

with $\hat{\beta}, \hat{\gamma} = \arg\min_{\beta, \gamma} f(A(G^A, \beta, \gamma), I)$.

The outer step (Eq. (7)) searches the space of the architectures $\{G^A\}$. For a fixed architecture $G^A$, the inner step (Eq. (8)) returns the optimal ansatz $A(G^A, \hat{\beta}, \hat{\gamma})$ in the family of $A(G^A, \beta, \gamma)$. Even though we will not make it explicit in the notation, we should understand that $\hat{\beta}$ and $\hat{\gamma}$ are implicit functions of the ansatz graph $G^A$ through Eq. (8). We denote the outer step as AAS and the inner step as parameter optimization.
A. Ansatz Architecture Search

It has been shown that a good design of the search space is essential in discrete structure optimization problems, e.g., neural architecture search [10, 12], molecule optimization [13], composite design [14] and symbolic regression [15, 16]. Since the QAOA is a well-recognized ansatz for combinatorial problems, we have designed the search space for \( G^A \) based on gradual modifications of the QAOA ansatz. The QAOA prescription is to take \( G^A = G^2 \), and our search through architectures is a search through graphs obtained by removing edges from \( G^2 \).

Denote by \( G_k \) a graph containing \( k \) edges. If \( m \) is the number of edges in \( G^2 \), then there is only one \( G_m \) in our search space, namely \( G^2 \) itself. Thus we say \( \{G_m\} = 1 \). If we remove up to \( n \) edges from the graph, then the total search space is

\[
\bigcup_{l=0}^{n} \{G_{m-l}\} = \{G_m\} \cup \{G_{m-1}\} \cup \ldots \cup \{G_{m-n}\}. \tag{9}
\]

The size of this space is \( \sum_{l=0}^{n}(\binom{m}{l}) \). As we increase \( n \), this size quickly grows too large to enumerate effectively. For example, considering a complete graph with 10 vertices, \( \sum_{l=0}^{10}(\binom{10}{l}) \sim 1 \times 10^6 \) and \( \sum_{l=0}^{15}(\binom{15}{l}) \sim 6 \times 10^{11} \). Thus a brute-force enumerative search is impractical.

In this section, we propose greedy search as an affordable strategy for AAS. Comparing to enumerative search, greedy search largely reduces the search space by only expanding the most promising node in the search tree. Given an instance \( I \), the search starts with \( G^A = G_m \) at level 0. Then the search is conducted level by level. The output architectures at level \( l \) have \( l \) two-qubit gates (i.e., edges of the graph) removed. The following three steps are performed at level \( l \) in the search:

**Expansion:** Generate all the unique \( \{G_{m-l}\} \) by removing one two-qubit gate from the output of the previous level.

**Scoring:** Evaluate a scoring function \( S \) on each of the architectures \( \{G_{m-l}\} \) generated by the previous step. Ideally, the scoring function would exactly match the final target function. However, that can be expensive to compute so we will examine alternative scoring functions below. In particular, we will consider different methods for specifying variational parameters \( \beta^*, \gamma^* \) for each circuit, and then evaluating the Gibbs objective function by simulation using those parameters:

\[
S(G_{m-l}, I) = \{f(A(G_{m-l}, \beta^*, \gamma^*), I)\}. \tag{10}
\]

**Selection:** Select the architecture with the best score as the output of this level.

At the \( l \)-th level, \( |\{G_{m-l}\}| \leq (m - l) \). The total number of architectures visited in the greedy search is

\[
N \leq \sum_{l=0}^{n}(m - l) = (n + 1)(m - n/2).
\]

B. Scoring an Ansatz

In the scoring step of AAS, we are presented with a number of architectures \( \{G_{m-l}\} \), and we need to decide which one is the most promising for further expansion. The strategy of Eq. (10) is to specify parameters \( \beta^* \) and \( \gamma^* \), which may depend on the ansatz in question, and evaluate the objective function using those parameters to obtain the score. In this section we will describe three different approaches for obtaining \( \beta^*, \gamma^* \).

1. **Nelder-Mead**

One solution is to let \( \beta^* \) and \( \gamma^* \) simply be the optimal values of \( \beta \) and \( \gamma \) for minimizing the objective function with a given architecture:

\[
\beta^*, \gamma^* = \arg \min_{\beta, \gamma} f(A(G_{m-l}, \beta, \gamma), I)
\]

In practice we use the Nelder-Mead algorithm [9] to perform this minimization. In other words, we take \( \beta^* \) and \( \gamma^* \) to be close approximation to the optimal values \( \beta \) and \( \gamma \) for the given ansatz graph. Nelder-Mead is a blackbox optimization algorithm popular in the quantum variational circuit literature [17, 18].

The initial values of \( \beta \) and \( \gamma \) are sampled independently from the uniform distribution \( U(0, 0.1) \). Using this algorithm requires running quantum circuit simulations at each iteration and reporting the objective function value to the optimizer until convergence. Thus it is extremely expensive in terms of calls to the (simulated) quantum computer. Since we want to limit the number of such calls, we are motivated to consider other strategies for finding \( \beta^*, \gamma^* \).

2. **Estimated \( \beta, \gamma \)**

Rather than use an optimization algorithm like Nelder-Mead to minimize the objective function and thereby obtain \( \beta^* \) and \( \gamma^* \), we can use analytical estimates to approximate the parameters instead. This saves the computation time required to evaluate the quantum circuits for parameter optimization during scoring. The \( \beta^* \) and \( \gamma^* \) we find will not necessarily be close to the optimal values \( \beta \) and \( \gamma \), but the idea is that this may not be important for the purposes of scoring. We may still wish to use Nelder-Mead for evaluation of the final ansatz at the conclusion of the AAS.

The estimates for \( \beta^* \) and \( \gamma^* \) we use in this section come from making several simplifying assumptions that are not necessarily valid. The first assumption is that it is reasonable to use \( \beta^* \) and \( \gamma^* \) values obtained by minimizing the energy expectation value instead of the Gibbs objective function. Note that we will still use the Gibbs objective function in Eq. (10); we are merely using the
expectation value to find estimates of $\beta^*$ and $\gamma^*$. Focusing the grid Ising model, we can write down an exact formula for $\langle E \rangle$ in a $p = 1$ QAOA as follows:

$$\langle E \rangle = \sin 4\beta \sum_{i,j} \left( \prod_k \cos (2\gamma J_{k,j}) \right) J_{ij} \tan (2\gamma J_{ij}). \quad (11)$$

To use this formula for an ansatz graph $G^A$ other than $G^2$, one just sets to 0 the $J_{ij}$ associated to the missing edges.

Eq. (11) determines $\beta^* = \pi/8$. To find a formula for $\gamma^*$ we make another simplifying assumption, namely that $\gamma^* \ll 1^2$. Expanding Eq. (11) to third order in $\gamma$ and minimizing the resulting cubic polynomial gives

$$\gamma^* = -\sqrt{\frac{\sum_{ij} J_{ij}^2}{6 \left( \sum_{i,j,k} J_{ijk}^2 - \frac{1}{3} \sum_{ij} J_{ij}^4 \right)}}. \quad (12)$$

All of this was in the context of grid instances, and in particular in deriving Eq. (11) we made use of the fact that two neighboring vertices in the graph do not have any neighbors in common. Nevertheless, as our final simplifying assumption we will insist on using Eq. (12) for the complete graph as well. We will see in the numerical experiments below that these simplifying assumptions are good enough for scoring.

3. Fixed $\beta, \gamma$

Our third method for specifying $\beta^*$ and $\gamma^*$ is to fix them in a way that is independent of the particular instance $I$. The estimated $\beta^*$ above is already instance-independent, and we saw the estimated $\gamma^*$ above performed well despite the approximations involved not being fully justified. This suggests that the precise value of $\gamma^*$ used in scoring is not crucially important. This is similar to the observation in Ref. [19] that the behavior of the QAOA tends to concentrate across instances.

With that as motivation, we generated the distribution of estimated $\gamma^*$ values according to the formula Eq. (12) for both the grid and complete graph models by looking at $10^5$ choices of couplings drawn independently from the uniform distribution, $J \sim U(-1,1)$, for each model with $G^A = G^2$. The associated histograms are shown in Figure 5. The “fixed-parameter” prescription for $\gamma^*$ is defined by using the medians of these distributions for all instances of the associated model. These values are listed in Figure 5.

C. Numerical Experiments

We apply the AAS procedure from Section III A with the three approaches to choosing the scoring parameters $\beta^*$ and $\gamma^*$ described in Section III B on grid and complete graph Ising models.

The performance of an ansatz $A$ produced by AAS is measured by the scaled probability of low energy,

$$\tilde{P}(A,I) = \frac{P_A(E < 0.95E_{gs})}{P_I(E < 0.95E_{gs})}. \quad (13)$$

The probability in the numerator is the one associated to the ansatz graph $G^A$ with parameters equal to their optimal values $\beta^*, \gamma^*$ obtained by minimizing the Gibbs objective function for that ansatz. The probability in the denominator is similar, except using the instance graph $G^I$ as the ansatz. In other words, the prescription for computing the denominator probability is similar to the standard QAOA, except the parameters are optimized using the Gibbs objective function rather than the energy expectation value. We could have compared to the standard QAOA prescription itself, but we chose to use the Gibbs objective function for both probabilities in order to isolate the effects of the AAS. In both cases, the optimization is done using Nelder-Mead.

Figure 6 shows the scaled probabilities of low energy of the optimal ansatz at each level for both grid and complete graph instances. Each column corresponds to a different prescription for the scoring function of AAS.

We first discuss the results of grid instances. In (a), the scoring is done using parameters that are optimized by Nelder-Mead as described in Section III B 1. The scaled probabilities of low energy increase as more two-qubit gates are removed. But they start to decrease when more than 5 two-qubit gates are removed. In (b) and (c) the scoring was performed according to the estimated parameter prescription and fixed parameter prescription of Sections III B 2 and III B 3, respectively. The dark curves in each case represent the performance of the final circuit.

---

2 Rather than finding an explicit formula, one could also choose to minimize Eq. (11) numerically to find $\gamma^*$. This does not affect the results.
found by AAS using the optimal \( \hat{\beta}, \hat{\gamma} \) obtained by minimizing the Gibbs objective function. We see that there is not a strong dependence on the scoring prescription, though the “fixed” procedure is slightly worse. However, the light curves in (b) and (c) represent the performance of those same output circuits if, rather than using \( \hat{\beta} \) and \( \hat{\gamma} \), we use the \( \beta^* \) and \( \gamma^* \) values used in the scoring step of AAS. Then we see that there is a significant decrease in performance, especially for the fixed method in (c). The lesson here is that for scoring in AAS, which only cares about relative performance for ranking, the circuit parameter values are less important. In fact, good relative performance from these two prescriptions suggests that it is possible to construct inexpensive heuristic functions for scoring without calls to the quantum computer. We explore this further in Appendix B. On the other hand, it is crucial to get the parameters right when considering absolute performance.

The story for complete graphs in (d)(e)(f) is very similar. The main qualitative change is that the performance does not drop off as steeply as a function of the number of removed gates. This is easily understood from the fact that the complete graphs have far more edges than the grid (45 vs 24). We also see that the “fixed” procedure is closer in performance to the others for complete graphs.

**IV. CONCLUSION**

We have proposed two modifications to the QAOA method for improved approximate optimization. The Gibbs objective function is an alternative to the energy expectation value for optimizing the variational parameters, and AAS is a method for searching the discrete space of quantum circuit architectures for superior gate layouts.

There are several potential follow-ups and opportunities for further developments:

The Gibbs objective function is easy to implement for the combinatorial optimization problems we considered here, but may be useful more broadly for quantum optimization problems, such as variational approaches to molecular ground states \[17, 20\]. In those cases, where the energy is not diagonal in the computational basis, it will be more challenging to evaluate \( e^{-\eta E} \) by sampling, but may still be worthwhile for increased performance.

Even within combinatorial optimization, it is not yet clear if AAS is worth doing because of the requirement that the quantum circuit be simulated (or run on a real quantum computer) during the scoring step of the search. Any improvements one finds in the ansatz could be offset by this extra cost. That is the motivation for the alternative heuristic methods we explore in Appendix B and it remains an open problem to find an effective scoring method that does not rely on quantum simulation. Our estimated parameter and fixed parameter scoring methods show that it is possible to capture relative performance without reproducing the absolute performance. This leaves open the possibility that a good heuristic scoring function exists without needing to do quantum simulation.

In the present work, for the sake of simplicity, we computed probabilities and expectation values directly from the wavefunction. On a real quantum computer, of course, this is impossible. Instead, one estimates expectation values based on a finite number of samples drawn from the Born distribution. The number of samples is another hyperparameter of the model, and it directly affects the cost of running the algorithm on a quantum computer. An interesting open question is whether the scoring step in AAS can work with a very small number of samples. This would mitigate the cost of the search, and might serve to make AAS worthwhile even without solving the problem of finding effective simulation-free heuristics. Finally, on a real device one may want to include other effects in the scoring, e.g., the fidelity of the two-qubit gates in the circuit, and search for the Pareto optimum \[21\] for multi-objective optimization.

---

3 We thank Edward Farhi for discussion of this point.
ACKNOWLEDGMENTS

The authors thank Steven Kearnes for code review and discussions, Zan Armstrong, Nathan Neibauer for suggestions and help in data visualization, and Dave Bacon, Edward Farhi, John Platt for their review and comments. X, formerly known as Google[x], is part of the Alphabet family of companies, which includes Google, Verily, Waymo, and others (www.x.company). Quantum simulation and AAS in this paper were implemented using Cirq [22] and Apache Beam [23].

APPENDIX A: BEAM SEARCH

We introduce beam search, a generalized search algorithm of greedy search introduced in Section III A. It has been used in combinatorial optimization [24], program synthesis [25] and machine translation [26]. Beam search differs from greedy search in the selection step:

**Selection:** Select architectures with the best scores, where the integer \( w \) is called the beam width. These best-performing architectures are the output of this level. At early stages in the beam search we may have fewer than \( w \) candidates available, in which case all candidates are returned.

At the \( l \)-th level, \( |\{ \tilde{G}_{m-l} \}| \leq w \times (m-l) \). The total number of architectures visited in the beam search is \( N \leq w \sum_{l=0}^{n} (m-l) = w(n+1)(m-n/2) \). As special cases, we recover enumerative search as \( w \to \infty \) and greedy search as \( w = 1 \).

Figure 7 illustrates the procedure of AAS for a complete graph with 4 vertices.

![Diagram of AAS Procedure](image)

**FIG. 7.** Illustration of Ansatz Architecture Search (AAS) by removing up to 3 two-qubit gates with beam width \( w = 2 \).

For experiments in Section III C we found that there was not much improvement in performance from increasing the beam width \( w \geq 1 \).

APPENDIX B: CHALLENGE: SEARCH WITHOUT QUANTUM SIMULATION

We demonstrated that with AAS and parameter optimization, a circuit ansatz that significantly improves the
probability of low energy can be found. However, all of the methods in Section III-B made use of quantum circuit simulation at each stage in the search. While we were able to show the method with the most quantum circuit simulation (Nelder-Mead) does not improve significantly on cheaper scoring methods, all the methods require some quantum circuit simulation at each level. In this section, we investigate some heuristic functions for replacement of quantum simulation for the purpose of the scoring step of AAS. Our results are mixed, and fully solving this problem remains an open challenge for the community.

a. Random For each ansatz in the scoring step, we assign a random number to replace \( f(A, I) \) in Eq. (10). This baseline does not use any information from the ansatz architecture and problem instance, and amounts to removing edges from the graph randomly during AAS.

b. Energy Approximation Our second heuristic uses the estimated energy expectation value as the scoring function. That is, we plug \( \beta \) the estimated energy expectation value as the scoring to removing edges from the graph randomly during AAS.

c. Neural Network We use a neural network to approximate \( f(A, I) \) in Eq. (10). It contains two dense layers with 128 hidden units and ReLU activation functions. The instance is represented by 2rd and 4th powers of couplings on edges. The ansatz graph \( G^A \) is represented by booleans indicating whether a two-qubit gate is placed on an edge of the instance. We concatenate these features as input of the network. We take all the ansatzes generated by AAS and Nelder-Mead in Section III-C and split them randomly by their instances into a training set and test set. For grid instances, the training set contains 800 instances with 232,800 ansatzes. For complete graph instances, the training set contains 800 instances with 568,800 ansatzes. Both test sets contain 200 instances not seen in the training set. To fix normalization we use the scaled objective function value \( f(A, I)/f(QAOA, I) \) as the label.

Although a training set is not required for the random and energy approximation heuristics, for a fair comparison we restrict each heuristic to the same 200 test instances for each instance type in Figure 8. We find optimal sparse ansatzes by removing 5 two-qubit gates for grid instances and 15 two-qubit gates for complete graph instances. (a)(e) show the results of AAS and Nelder-Mead. They are the best ansatzes we can find for the test instances. Since random, energy approximation and neural network are inexpensive to compute compared to quantum simulation, we give these an advantage in the search by setting the beam width \( w \) to 100. At the end of AAS, we sort the ansatzes from the last level by their objective function values. Then we run quantum simulations for the top candidates and report the best scaled probability of low energy. As more candidates are taken into consideration, the performance of all three heuristic functions improves but the cost of quantum simulation for evaluation also increases. The number of top candidates chosen in each case is listed along the horizontal axis, and the candidates with the lowest Gibbs objective function value is selected for the plot. The solid lines are the mean performance across instances, the dashed lines are at 1, and the shadows show the range from 5% to 95%.

Appendix C: The Effect of Noise

The purpose of this appendix is to analyze the effects of a simple noise model on the Gibbs objective function of Eq. (4). In the absence of noise, the ideal Gibbs objective function is given by

\[
 f_{\text{ideal}} = -\log \langle e^{-\eta E} \rangle_{\psi},
\]

where \( E \) is the Hamiltonian we are optimizing and the angled brackets represent the quantum expectation value

\[
 \langle \cdot \rangle_{\psi} = \frac{\langle \cdot | \psi \rangle \langle \psi | \cdot \rangle}{\langle \psi | \psi \rangle}.
\]

4 Really, we first expanded Eq. (11) to third order in \( \gamma \) before plugging in \( \gamma^* \). This is for consistency, but does not make a large difference. We also experimented with numerically minimizing Eq. (11) rather than using any estimates, and this, too, does not make much difference.
in the output state $\psi$ of the quantum circuit ansatz.

The noise model we consider is a simple depolarizing channel. With probability $1 - p$ the quantum circuit is executed perfectly and the output state is the one we expect. With probability $p$, there is some error in the execution and the output state is the maximally mixed one. In other words, with probability $p$ we sample from the uniform distribution on bit strings instead of the desired Born distribution. In the language of density operators, we can say that the effective density operator describing the quantum state is

$$(1 - p)|\psi\rangle\langle\psi| + \frac{p}{2^n} I,$$

where $n$ is the number of qubits. Using this error model, we can ask what the noise does to the Gibbs objective function $f$. We simply replace the expectation value in the ideal state $\psi$ with an expectation value in the noisy state $(1 - p)|\psi\rangle\langle\psi| + pI/2^n$. Equivalently, we can take a weighted average of the $\langle \cdot \rangle_\psi$ expectation value with an expectation value according the uniform distribution over bit strings. We find

$$f_{\text{noisy}} = -\log \left( (1 - p) \langle e^{-\eta E} \rangle_\psi + \frac{p}{2^n} \text{Tr} e^{-\eta E} \right)$$

$$= f_{\text{ideal}} - \log \left( 1 - p \frac{\langle e^{-\eta E} \rangle_\psi - \text{Tr} e^{-\eta E}/2^n}{\langle e^{-\eta E} \rangle_\psi} \right).$$

Note that one expects $\langle e^{-\eta E} \rangle_\psi \geq \text{Tr} e^{-\eta E}/2^n$ if the circuit is properly trained, and so the correction makes the objective function larger (worse), as it should. We also have the following bound on the change in the objective function, coming from the positivity of $e^{-\eta E}$:

$$f_{\text{noisy}} - f_{\text{ideal}} \leq -\log(1 - p). \quad (C1)$$

For small $p$ the right-hand-side is just $\approx p$. It’s reasonable to expect that $\langle e^{-\eta E} \rangle_\psi \gg \text{Tr} e^{-\eta E}/2^n$—in other words, the trained ansatz should have a much better Gibbs objective function value than the uniform distribution over bit strings—which means that the bound in Eq. (C1) will be approximately saturated.

This means that we can directly translate improvements to the objective function into resilience against depolarizing noise. An improvement of size $\Delta f$ in the objective function can counteract the effect of depolarizing noise with size $p \approx \Delta f$ (assuming $p < 1$). It is also worth noting the effects of noise on the probability of finding a low-energy bit string, $P(E < E_0)$. Using the same depolarizing noise model,

$$P_{\text{noisy}} = P_{\text{ideal}} - p(P_{\text{ideal}} - P_{\text{uni}}).$$

Here $P_{\text{uni}}$ is just the probability for success by random guessing using the uniform distribution on bit strings. Then, taking logs, we find

$$\log P_{\text{noisy}} = \log P_{\text{ideal}} + \log \left( 1 - p \frac{P_{\text{ideal}} - P_{\text{uni}}}{P_{\text{ideal}}} \right)$$

This is very similar to what we saw in for the behavior of the Gibbs objective function. This is not a coincidence: part of the reason why the Gibbs objective function was chosen is that the operator $e^{-\eta E}$ for appropriate values of $\eta$ behaves very similarly to the projection operator one would use to define $P$. We already noted in Section II A that for large $\eta$ and $E_0$ close to $E_{gs}$, $P$ and $\langle e^{-\eta E} \rangle_\psi$ become equal up to a state-independent multiplicative factor.

1. E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv preprint arXiv:1411.4028 (2014).
2. E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm applied to a bounded occurrence constraint problem, arXiv preprint arXiv:1412.6062 (2014).
3. L. G. Valiant, A theory of the learnable, Commun. ACM 27, 1134 (1984).
4. M. Y. Niu, S. Lu, and I. L. Chuang, Optimizing qao: Success probability and runtime dependence on circuit depth, arXiv preprint arXiv:1905.12134 (2019).
5. E. Farhi and A. W. Harrow, Quantum supremacy through the quantum approximate optimization algorithm, arXiv preprint arXiv:1602.07674 (2016).
6. J. Preskill, Quantum computing in the nisq era and beyond, Quantum 2, 79 (2018).
7. P. K. Barkoutsos, G. Nannicini, A. Robert, I. Tavvernelli, and S. Woerner, Improving variational quantum optimization using cva, arXiv preprint arXiv:1907.04769 (2019).
8. P. McCullagh and J. Kolassa, Cumulants, Scholarpedia 4, 4699 (2009) revision #137322.
9. J. A. Nelder and R. Mead, A simplex method for function minimization, The computer journal 7, 308 (1965).
10. B. Zoph and Q. V. Le, Neural architecture search with reinforcement learning, The International Conference on Learning Representations (2017)
11. S. Xie, A. Kirillov, R. Girshick, and K. He, Exploring randomly wired neural networks for image recognition, arXiv preprint arXiv:1904.01569 (2019).
12. L. Li and A. Talwalkar, Random search and reproducibility for neural architecture search, arXiv preprint arXiv:1902.07638 (2019).
13. Z. Zhou, S. Kearnes, L. Li, R. N. Zare, and P. Riley, Optimization of molecules via deep reinforcement learning, Scientific reports 9, 10752 (2019).
14. G. X. Gu, C.-T. Chen, D. J. Richmond, and M. J. Buehler, Bioinspired hierarchical composite design using machine learning: simulation, additive manufacturing,
and experiment, Materials Horizons 5, 939 (2018).

[15] M. Schmidt and H. Lipson, Distilling free-form natural
laws from experimental data, science 324, 81 (2009).

[16] S.-M. Udrescu and M. Tegmark, AI Feynman: a
Physics-Inspired method for symbolic regression, (2019),
arXiv:1905.11481 [physics.comp-ph]

[17] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q.
Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. Obrien,
A variational eigenvalue solver on a photonic quantum
processor, Nature communications 5, 4213 (2014).

[18] G. Verdon, J. Pye, and M. Broughton, A universal training
algorithm for quantum deep learning, arXiv preprint
arXiv:1806.09729 (2018).

[19] F. G. Brandao, M. Broughton, E. Farhi, S. Gutmann,
and H. Neven, For fixed control parameters the quantum
approximate optimization algorithm’s objective function
value concentrates for typical instances, arXiv preprint
arXiv:1812.04170 (2018).

[20] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-
Guzik, The theory of variational hybrid quantum-
classical algorithms, New Journal of Physics 18, 023023
(2016).

[21] K. Deb, Multi-objective optimization, in
Search methodologies (Springer, 2014) pp. 403–449.

[22] Cirq: A python framework for creating, editing, and in-
voking noisy intermediate scale quantum (nisq) circuits.,
https://github.com/quantumlib/Cirq

[23] Apache beam, https://beam.apache.org/

[24] Z. Li, Q. Chen, and V. Koltun, Combinatorial optimization with graph convolutional networks and guided tree search, in
Advances in Neural Information Processing Systems
(2018) pp. 539–548.

[25] R. Bunel, M. Hausknecht, J. Devlin, R. Singh,
and P. Kohli, Leveraging grammar and reinforcement
learning for neural program synthesis, arXiv preprint
arXiv:1805.04276 (2018).

[26] I. Sutskever, O. Vinyals, and Q. V. Le, Sequence
to sequence learning with neural networks, in
Advances in neural information processing systems
(2014) pp. 3104–3112.