Multistart Methods for Quantum Approximate Optimization

1st Ruslan Shaydulin
School of Computing
Clemson University
Clemson, South Carolina, USA
rshaydu@g.clemson.edu

2nd Ilya Safro
School of Computing
Clemson University
Clemson, South Carolina, USA
isafro@g.clemson.edu

3rd Jeffrey Larson
Mathematics and Computer Science Division
Argonne National Laboratory
Lemont, Illinois, USA
jmlarson@mcs.anl.gov

arXiv:1905.08768v3 [quant-ph] 23 May 2019

***Abstract*—Hybrid quantum-classical algorithms such as the quantum approximate optimization algorithm (QAOA) are considered one of the most promising approaches for leveraging near-term quantum computers for practical applications. Such algorithms are often implemented in a variational form, combining classical optimization methods with a quantum machine to find parameters to maximize performance. The quality of the QAOA solution depends heavily on quality of the parameters produced by the classical optimizer. Moreover, the presence of multiple local optima in the space of parameters makes it harder for the classical optimizer. In this paper we study the use of a multistart optimization approach within a QAOA framework to improve the performance of quantum machines on important graph clustering problems. We also demonstrate that reusing the optimal parameters from similar problems can improve the performance of classical optimization methods, expanding on similar results for MAXCUT.**

***Index Terms*—quantum approximate optimization, multistart optimization, graph clustering**

I. Introduction

A number of quantum computing devices have recently become available to researchers [1], [2]. These Noisy Intermediate Scale Quantum (NISQ) devices currently have less than 100 qubits, high error rates, and a restricted set of available algorithms [3]. The famous Shor's algorithm [4] requires executing thousands of gates [5], something that is impossible to do accurately without error correction on quantum machines. At the same time, there is a growing interest in applying emerging NISQ devices to practical applications [6]–[8].

Multiple near-term algorithms have been introduced in an attempt to take advantage of NISQ devices. Among the most promising are hybrid quantum-classical algorithms, including the Variational Quantum Eigensolver (VQE) [9] and the Quantum Approximate Optimization Algorithm (QAOA) [10]. These algorithms combine a classical optimizer with a quantum machine where the quantum evolution is performed by applying gates to some initial state (in the case of QAOA, the initial state is an equal superposition of basis states), with the goal of preparing the state with desired properties. For example, in VQE the goal is to prepare the ground state (i.e., the state corresponding to the smallest eigenvalue) of a given system. An advantage of hybrid algorithms is that the quantum evolution is described by a shallow-depth circuit, enabling them to be run on NISQ computers without error correction. The shallow depth of the circuit is achieved by parameterizing the gates. An example of a parameterized gate is a rotation around the Z axis (RZ), where the parameter is the angle of the rotation; the optimal quantum evolution then can be found by varying the parameters of a shallow set of gates. In this paper we address QAOA, but our approach is applicable to other hybrid algorithms, including VQE. Although optimal parameters can be found analytically for some problems, parameters within QAOA typically are found by using a classical optimizer in a variational setting. Therefore, in this paper we consider only the variational implementation of QAOA.

In QAOA, the quantum evolution is performed by applying a series of parameterized gates, commonly referred to as the ansatz. These gates are parameterized by variational parameters, denoted by $\boldsymbol{\theta}$. At each step, a multiqubit trial state $|\psi(\boldsymbol{\theta})\rangle$ is prepared on the quantum coprocessor by applying the ansatz. The state is then measured, and the result is used by the classical optimizer to find new parameters $\boldsymbol{\theta}$, with the goal of finding the ground-state energy $E_G = \min_{\boldsymbol{\theta}} \langle\psi(\boldsymbol{\theta})| \hat{H}_C |\psi(\boldsymbol{\theta})\rangle$, where $\hat{H}_C$ is the cost Hamiltonian. The ground state encodes the global optimum of the classical optimization problem. This variational cycle continues until the classical optimizer converges or a solution of acceptable quality is found.

Such hybrid algorithms are considered the most promising path to demonstrating quantum advantage, that is, demonstrating superior performance of a quantum system on some problem when compared with state-of-the-art classical methods. Demonstrating quantum advantage is a prerequisite for quantum computers to become a valuable high-performance computing resource. Variational hybrid algorithms, including VQE and variational implementations of QAOA, require reliable classical optimization methods to obtain solutions of good quality. Moreover, the performance of classical optimization methods in terms of the number of function evaluations directly translates into an improvement in performance of a variational quantum algorithm. Therefore, it is imperative that efficient and reliable optimization methods be developed for finding optimal variational parameters. Unfortunately, the parameter space for these problems is nonconvex and contains many low-quality, nondegenerate local optima [11]. Figure 1 shows an example energy landscape of a QAOA objective function with two parameters. This landscape has many low-

quality optima that a local optimizer can get stuck in. In this paper, we address this challenge by using a multistart local optimization method. Our results are twofold. First, we explore direct optimization of QAOA parameters under realistic time constraints and show that the multistart framework APOSMM [12], [13] is able to find better parameters than single-start local search methods can (when using the same number of objective evaluations). Second, we demonstrate that the optimal QAOA parameters found for a given problem can be reused as an initial point for similar problems, both improving the quality of the solution and reducing the number of evaluations required to obtain it.

QAOA has attracted considerable attention as a candidate algorithm for NISQ devices. When QAOA was originally introduced in 2014, it was shown to outperform the state-of-the-art classical solver for the combinatorial problem of bounded occurrence Max E3LIN2 [19]. (Thereafter, an improved classical algorithm was introduced that outperformed QAOA on this problem [20].) A recent paper [21] shows that QAOA (using a circuit with modest depth) can exceed the performance of Goemans-Williamson [22] algorithm for MAXCUT. In addition to these empirical results, theoretical results demonstrate that QAOA for MAXCUT improves on the best-known classical approximation algorithms for certain graphs [23], [24]. Although there is an active discussion about exactly how many qubits are required for meaningful quantum speedups [6], [25], the future of QAOA looks bright.

II. Problem Definition

Consider a cost Hamiltonian $\hat{H}_C$ encoding the classical optimization problem (later in this section we present a cost Hamiltonian for network community detection). Because the underlying optimization problem we are solving is *maximization*, we construct the cost Hamiltonian $\hat{H}_C$ such that its highest-energy eigenstate encodes the solution, as opposed to the ground or lowest-energy state commonly used in VQE.[1] The goal of the hybrid algorithm is to prepare this eigenstate. In hybrid quantum-classical algorithms, the evolution is performed by applying a set of parameterized gates (ansatz). The goal then is to find a set of parameters that describe the evolution that prepares the desired state.

In QAOA, the quantum evolution starts in the initial state $|+\rangle^{\otimes n}$. Then the evolution is performed by applying two alternating operators based on the cost Hamiltonian $\hat{H}_C$ and mixing Hamiltonian $\hat{H}_M = \sum_i \hat{\sigma}_i^x$:

$$\begin{aligned} |\psi(\boldsymbol{\theta})\rangle &= |\psi(\boldsymbol{\beta},\boldsymbol{\gamma})\rangle \\ &= e^{-i\beta_p \hat{H}_M} e^{-i\gamma_p \hat{H}_C} \cdots e^{-i\beta_1 \hat{H}_M} e^{-i\gamma_1 \hat{H}_C} |+\rangle^{\otimes n} . \end{aligned} \tag{1}$$

Here p is the number of alternating operators or QAOA "steps." Then the objective function f (i.e., the energy of $\hat{H}_C$ in the state $|\psi(\boldsymbol{\beta},\boldsymbol{\gamma})\rangle$) is

$$f(\boldsymbol{\beta},\boldsymbol{\gamma}) = -\langle\psi(\boldsymbol{\beta},\boldsymbol{\gamma})| \hat{H}_C |\psi(\boldsymbol{\beta},\boldsymbol{\gamma})\rangle . \tag{2}$$

[1]Note that in our case it is just a matter of convention, since introducing a minus sign changes a maximization problem into a minimization problem.

Based on the value $f(\boldsymbol{\beta},\boldsymbol{\gamma})$, the classical optimizer chooses the next set of parameters $\boldsymbol{\beta},\boldsymbol{\gamma}$ with the goal of finding parameters that minimize f:

$$\begin{aligned} \boldsymbol{\beta}_*, \boldsymbol{\gamma}_* &= \arg\min_{\boldsymbol{\beta},\boldsymbol{\gamma}} f(\boldsymbol{\beta},\boldsymbol{\gamma}) \\ &= \arg\min_{\boldsymbol{\beta},\boldsymbol{\gamma}} (-\langle\psi(\boldsymbol{\beta},\boldsymbol{\gamma})| \hat{H}_C |\psi(\boldsymbol{\beta},\boldsymbol{\gamma})\rangle). \end{aligned} \tag{3}$$

The objective function f is periodic with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, allowing the parameters to be restricted to $\beta_i \in [0,\pi]$, $\gamma_i \in [0,2\pi]$. Therefore the optimization domain is compact: $(\boldsymbol{\beta},\boldsymbol{\gamma}) \in \mathcal{D} = ([0,\pi]\times[0,2\pi])^p$.

We explore QAOA applied to the modularity maximization problem for the network community detection. Also known as graph clustering, network community detection aims to group vertices of the graph so that they are nontrivially connected compared with a random graph model. Modularity maximization often (but not necessarily) groups vertices so that there are as many edges as possible within the groups and as few as possible between the groups. Formally, for an undirected graph $G = (V,E)$ with two communities, modularity is defined as in [26]:

$$C = \frac{1}{4|E|}\sum_{ij}(A_{ij} - \frac{k_i k_j}{2|E|})s_i s_j = \frac{1}{4|E|}\sum_{ij} B_{ij} s_i s_j, \tag{4}$$

where A is the adjacency matrix of G, k_i is the degree of vertex $i \in V$, and the variables $s_i \in \{-1,+1\}$ indicate community assignment of vertex i. That is, $s_i = -1$ denotes vertex i as being assigned to the first community, and $s_j = +1$ denotes that vertex j is assigned to the second community. Modularity maximization for general graphs is NP-hard [27] and has a variety of applications in complex systems [28]–[32].

The modularity maximization problem can be mapped onto QAOA by promoting variables s_i in (4) to Pauli spin operators $\hat{\sigma}^z$ [6]–[8], resulting in the Hamiltonian

$$\hat{H}_C = \frac{1}{4|E|}\sum_{ij} B_{ij}\hat{\sigma}_i^z\hat{\sigma}_j^z. \tag{5}$$

Multiple ansatzes (sets of gates used to produce trial state $|\psi(\boldsymbol{\theta})\rangle$) have been explored for QAOA, with the hardware-efficient ansatz [33] (originally proposed for VQE) being one of the most successful [6]. A similar ansatz leveraging nearest-neighbor interactions available on the device has been shown to achieve a better-than-random-guess approximation ratio for the MAXCUT problem on 3-regular graphs [34]. In this work we do not consider these ansatzes, however, because at the time of writing there is no evidence that QAOA with such ansatzes can beat the best classical algorithms. Instead, we focus on the alternating operator ansatz in (1).

III. Related Work

While the most commonly used strategy for identifying optimal QAOA parameters is using a classical optimizer in a variational loop, QAOA is not necessarily a variational algorithm. For example, Parekh et al. show that for one-step ($p = 1$) QAOA for MAXCUT on k-regular triangle-free graphs, parameters can be derived analytically [23]. Wang et

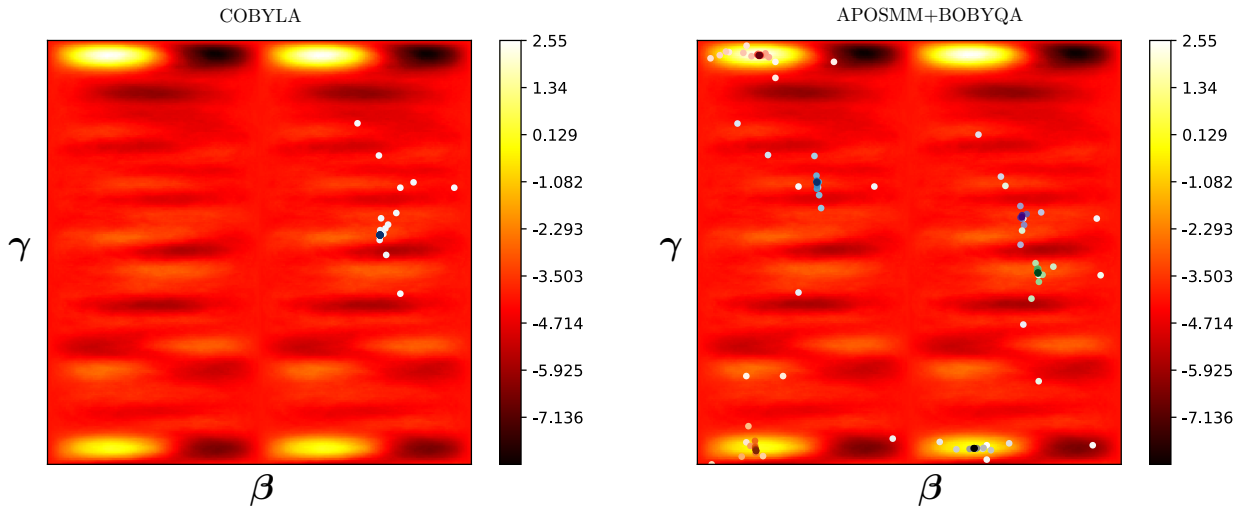

Fig. 1. Energy landscape of QAOA objective function $\langle\psi(\beta,\gamma)|\,\hat{H}_C\,|\psi(\beta,\gamma)\rangle$ for modularity maximization community detection on connected caveman graph [14], [15] with 4 cliques of 4 vertices. Higher (white) is better. Left: the points evaluated by a single run of COBYLA [16]–[18]; each point corresponds to a pair (β,γ) that the local optimizer queried. Right: trace of APOSMM [12], [13] coordinating multiple COBYLA instances. Both methods were given a budget of 200 function evaluations.

al. show a similar result for one-dimensional antiferromagnetic rings [24]. More generally, Farhi et al. [10] proposed discretizing parameters into a grid. For N-qubit QAOA, however, this approach requires $N^{O(p)}$ evaluations, making it impractical even for small p. Finding good QAOA parameters remains a challenging problem, which motivates this work.

A. Parameter optimization in hybrid algorithms

Despite the recent advances in gradient-based methods [11], [21], [35]–[38], gradient-free black-box methods remain the most common approach for optimizing parameters in hybrid quantum-classical algorithms. A variety of methods have been used, including the Nelder-Mead method [39] (for both QAOA parameter optimization [25], [35] and training quantum Boltzmann machines [40]), Bayesian methods [41], Powell's method [42], and an interior-point minimization method [43]. Researchers resort to derivative-free methods because analytic gradients for quantum circuits may not be available and approximating gradients can be computationally expensive [21]. (In some cases, algorithmic differentiation techniques may provide gradient information [36].) Since gradient-based methods can be sensitive to noise [44], they may be less suitable for noisy intermediate-scale quantum hardware.

A number of recent advances in finding good parameters have been made in the recent years, potentially making their optimization simpler. For QAOA, multiple papers have shown connections between adiabatic schedule and QAOA parameters [11], [21], [45].

Zhou et al. [11] show that even at small depth p the schedule defined by optimal QAOA parameters is reminiscent of adiabatic quantum annealing, where $\hat{H}_C$ is gradually turned on while $\hat{H}_M$ is gradually turned off (see Sec. II). Similar results were found by Crooks [21]. Additionally, Zhou et al. [11] demonstrate that the optimal values $\boldsymbol{\beta}_*, \boldsymbol{\gamma}_*$ have small variation between similar problem instances, a finding that we confirm in this work for a different graph problem (see Sec. V). Zhou et al. use these insights to introduce a novel parameterization of QAOA and a heuristic optimization scheme based on it.

Brandao et al. [45] show that for MAXCUT on 3-regular graphs, the objective function value is concentrated; that is, typical instances have nearly the same value of objective function. They make a case that the same holds for any combinatorial search problem where the number of clauses with a given variable is bounded (e.g., MAXCUT on a bounded-degree graph). They propose reusing optimal parameters between problems that come from the same distribution and refining them using a local optimization heuristic. In this work, we successfully apply this strategy to modularity clustering, a problem where the number of clauses in which a variable can appear grows with n (see Sec. V).

Periodicity of the objective function with respect to QAOA parameters, visible on heatmaps in Fig. 1, has been demonstrated for MAXCUT [11], [24]. The periodicity was also observed for quasi-maximum-likelihood decoding of classical channel codes [46]. This can potentially allow for further restriction of the domain, eliminating some of the local optima and making the optimization problem easier. However, the theoretical results so far are problem specific. Therefore, we restrict our optimization domain to $\beta_i \in [0,\pi]$, $\gamma_i \in [0,2\pi]$, following [10]. Note that this differs from the approach in [21], where the values of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ were not constrained. A recent result shows that exploiting the periodicity of variational parameters of certain ansatzes for QAOA and other variational algorithms can improve optimization performance [47].

B. Derivative-free optimization methods

Selecting $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ values that maximize the objective function in (3) is a central optimization problem in variational algorithms. Since the gradient of the objective function with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ is unavailable on real quantum computers, researchers usually resort to so-called derivative-free optimization (DFO) methods: those that work only with observations of the objective function. Classical derivative-free direct-search methods are commonly applied to such problems: for example, Nelder-Mead is the default method for VQE problems in Grove [48]. Yet McClean et al. [49] shows that modern DFO methods achieve considerable benefits in terms of the number of function evaluations required. The BOBYQA method [50] is one such method for bound-constrained derivative-free optimization that builds quadratic models of the objective and optimizes them over a trust region in order to produce candidate points.

In the numerical optimization community, one commonly starts local optimization methods from different initial conditions in an attempt to identify better optima. While such an approach may be easy to implement, it may result in unnecessary objective function evaluations. Assuming there are a finite number of local optima, the ideal approach would identify each using only a single local run.

The multilevel single linkage method (MLSL) [51], [52], uniformly samples points over the domain $\mathcal{D}$ and starts runs from those points that do not have a better point within a ball of some radius. They show favorable results for a specific approach for updating the radius as the number of sampled points increases, although such results are only asymptotic. MLSL was generalized by APOSMM [12], [13] to consider all points generated by an ensemble of local optimization runs, and not just those sampled from the domain.

IV. Difficulty of Optimizing QAOA Parameters

In this section we present the results from using DFO methods to find optimal QAOA parameters. We use the high-performance simulator Qiskit Aer [53] to perform noiseless simulations of QAOA circuits. We measure the quality of the solution found by six derivative-free local optimization methods as implemented in the NLopt nonlinear-optimization package [18]: BOBYQA [54], COBYLA [16], [17], NEWUOA [55], Nelder-Mead [39], PRAXIS [56] and SBPLX [57]. We compare their performance to the implementation of APOSMM from the libEnsemble library [58]. APOSMM coordinates multiple local optimization runs in an attempt to identify better local optima. In this work, we use BOBYQA as the local optimization method within APOSMM (we denote this method as APOSMM+BOBYQA in figures). The performance of all methods is evaluated using two-way modularity maximization community detection problem on six synthetic graphs with community structure: three instances of connected caveman graph [14] and three instances of random partition graph [59]. All graphs have between 10 and 12 vertices and were generated with NetworkX [15]. The code used to perform the experiments is available [60].

We performed two sets of experiments. First, we set the tolerances of local solvers to zero and allow them to run until convergence. The quality of the obtained solutions was then compared with the solutions found by APOSMM using the same number of evaluations. We observe that APOSMM finds solutions with a much higher value of the objective function (see Fig. 2). Since APOSMM is allowed another local optimization run after one has converged, a local method may not take full advantage of the function evaluations budget. To allow for a more equal comparison, we performed a second set of experiments. In the second set, we set the tolerances of local solvers to be equal to the tolerance of BOBYQA within APOSMM and if the method convergence before exhausting function evaluations, it is restarted at a different random point. This restart scheme is essentially a naive version of MLSL. These results are also compared with APOSMM (see Fig. 3).

For both sets of the experiments, we limit the number of evaluations to 1,000. We choose this number as the realistic number of evaluations based on the estimates in [25]. We use the same realistic if aggressive assumption of 1 millisecond per single run. Estimating objective function in Eq. 2 requires thousands to tens of thousands measurements in practice [25], [33], [41]; we use an optimistic assumption of 1,000 measurements per run for obtaining the statistics to estimate the objective function. This gives an estimate on the time cost of performing the optimization equal to (time per single run) $\times$ (1,000 measurements per run) $\times$ (1,000 evaluations) $\approx$ 16 min. Note that this runtime is still orders of magnitude greater than the runtime of classical state-of-the-art MAXSAT solvers applied to the same problem [25]. Additionally, as the hardware is rapidly evolving, it is not possible to project these estimates into the future with certainty. However, it provides a useful estimate on the reasonable number of calls to the quantum device in a QAOA run.

Results show that a single run of a local optimization method cannot identify parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ corresponding to a high-quality solution of underlying problem (i.e., a high value of objective function). Fig. 2 shows that APOSMM is capable of finding parameters corresponding to values of objective function much larger than just the local solvers. This is partly due to local solvers converging before exhausting the limit on number of function evaluations (1,000).

If we set tolerances for local methods to the same values as in APOSMM (the tolerances on change in the function value to 10^{-3} and on the change in optimization parameters to 10^{-2}) and restart local methods after convergence, we observe that APOSMM is still solving more problems within the same budget of function evaluations. This is measured in the data profiles in Fig. 3; these data profiles track the fraction of problems solved to some level τ after a given number of function evaluations. Explicitly, if $t_{p,s}$ is the number of function evaluations required for each optimization method s to solve problem p in the set of problems P, then the data profile is

$$d_s(\alpha) = \frac{|\{p : t_{p,s} \leq \alpha\}|}{|P|}.$$

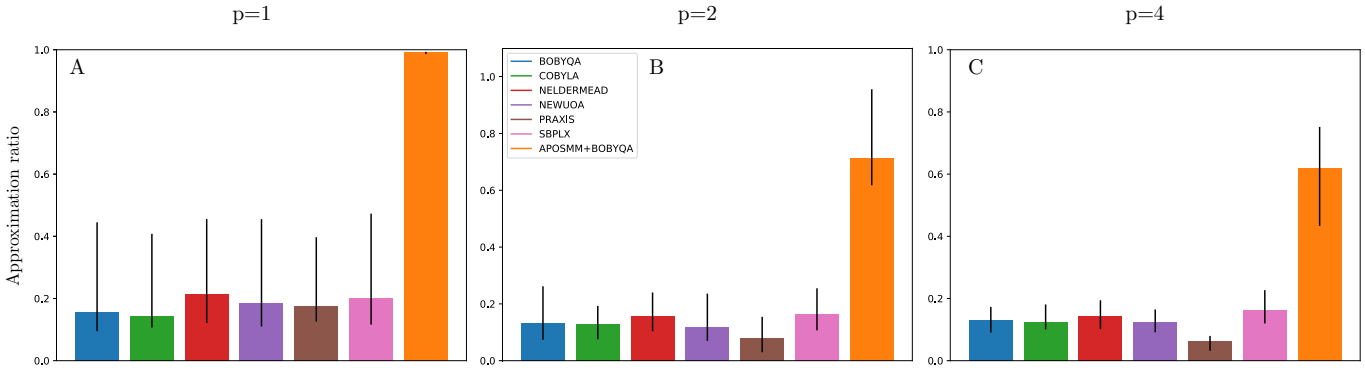

Fig. 2. Ratio between the value of the objective function found by an optimization method and the best-found value. All local methods are run with *no* restart and zero tolerances. Heights of bars represent median over (10 seeds per problem) $\times$ (6 problems) $=$ 60 runs. Error bars represent quartiles (25th and 75th percentiles). When compared with local methods without restarting, APOSMM finds solutions with much higher objective function values. This is due to local methods converging before exhausting the budget on number of function evaluations. p is number of QAOA steps ($p = 1$ corresponds to 2-dimensional domain $\mathcal{D}$ (A), $p = 2$ corresponds to $dim(\mathcal{D}) = 4$ (B), and $p = 4$ corresponds to $dim(\mathcal{D}) = 8$ (C).) Note that the approximation ratio 1.0 corresponds to the maximum value observed for a given problem and given value of p. The maximum absolute values of the objective function vary between the different numbers of QAOA steps.

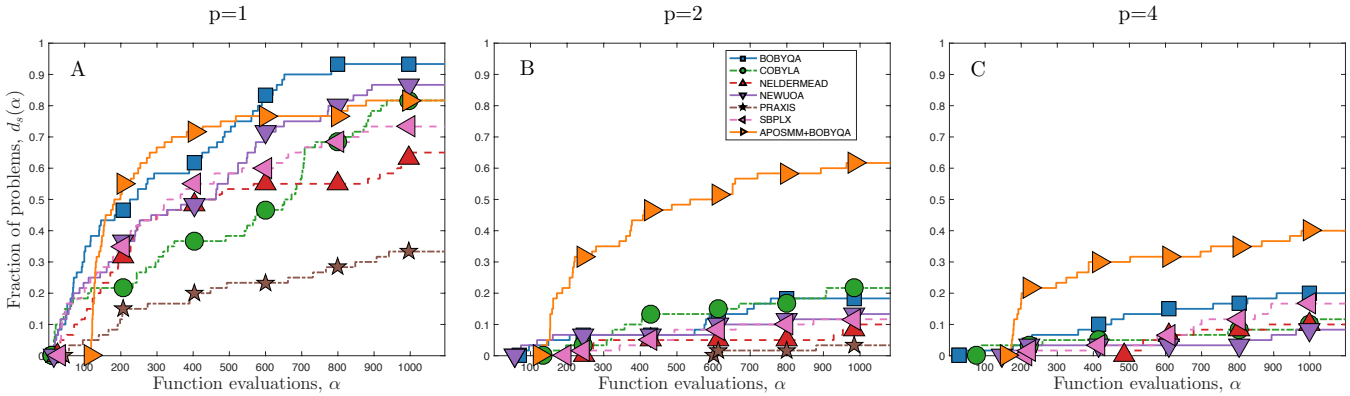

Fig. 3. Data profiles for seven optimization methods on the $p = 1$ (A), $p = 2$ (B), and $p = 4$ (C) benchmark problems with $\tau = 0.01$. For the $p = 1$ (i.e., two-dimensional) problems, most methods are competitive; but as the number of parameters (i.e., circuit depth) increases, all methods have difficulty in identifying high-quality solutions on a large fraction of the test problems. Yet, we see that APOSMM+BOBYQA performs noticeably better.

where α is the number of function evaluations. Data profiles require some definition of solving a problem to a level τ. For these problems, an optimization method s is determined to have solved problem p to a level τ after j evaluations if

$$f(x^0) - f(x^j) \geq (1 - \tau)(f(x^0) - \tilde{f}_p), \tag{6}$$

where x^0 is the problem's starting point, x^j is the jth point evaluated by the method, and $\tilde{f}_p$ is the best-found function value by any optimization method on problem p. For example, if $\tau = 0.01$, the convergence test in (6) determines a method to solve problem p when a point is evaluated with 99% of the possible decrease on the problem (among the implementations being compared).

Figures 2 and 3 demonstrate that finding optimal parameters becomes increasingly harder as the dimension of the domain $\mathcal{D}$ (i.e., the number of QAOA steps p) increases. For $p = 1$, BOBYQA and APOSMM solve most of the problems (Fig. 3A) within 1,000 function evaluations, for $p = 2$ and $p = 4$ the best-performing method (APOSMM) solves only 60% and 40% of the problems, respectively. These results indicate that even for small number of QAOA steps ($p = 4$) direct optimization of variational parameters is hard under realistic time constraints.

V. Reusing Optimal QAOA Parameters

Sec. IV presents results demonstrating the complexity of finding good QAOA parameters under realistic time constraints. Recently a number of researchers proposed amortizing the cost of finding good QAOA parameters for MAXCUT by reusing optimal parameters found for a given problem on similar problems [11], [21], [45]. We confirm and extend these findings by reusing optimal QAOA parameters found by exhaustive search. Optimal parameters for QAOA for modularity maximization on a given graph are used as an initial guess for the local solver on a similar graph constructed by removing an edge from the original graph. This simulates a realistic scenario of solving community detection on a

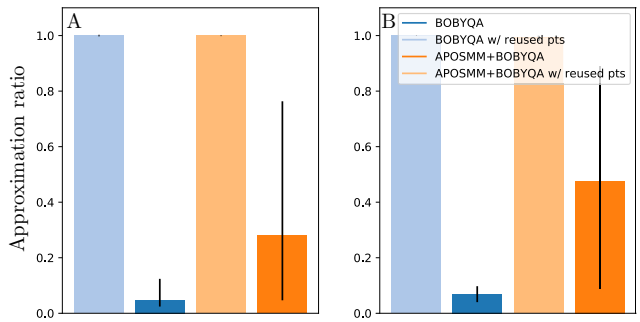

Fig. 4. Ratio between the value of the objective function found by an optimization method and the best-found value. Left (A): we compare the best-performing local method and APOSMM with optimal points from similar problems ("w/ reused pts") and with random initial points. Heights of bars represent median over (10 seeds per problem) $\times$ (6 problems) $\times$ (5 different random edges removed) $= 300$ runs. Right (B): for each problem we remove only one "worst-case" edge. Error bars represent quartiles (25th and 75th percentiles). Reusing precomputed optimal points allows optimization methods to find better solutions (corresponding to higher objective values) within the same budget of function evaluations.

dynamical graph, for example, a social network where new friend connections are dynamically added and removed.

We estimate true optimal parameters by setting the tolerance on the change in the function value to 10^{-3} and the tolerances on the change in the parameters to 10^{-2} and restarting BOBYQA after each convergence until 100,000 function evaluations have been used. We observe that this exhaustive approach identifies multiple high-quality local optima. We then use these high-quality QAOA parameters as initial guesses for local methods and APOSMM. After a local method converges, it is restarted from the next-best local optima.

Our contribution extends previous work in two ways. First, we consider a different optimization problem, namely, modularity community detection. Second, in addition to random similar problems, we consider "worst-case" small changes. To simulate a "worst-case" scenario, we remove an edge from the graph that has the greatest impact on its spectrum. Concretely, we compute the spectrum of the graph Laplacian before and after removing an edge. The change in the spectrum is measured by computing the Euclidean distance between the eigenvectors of the graph Laplacians. The graph spectrum has deep connections to many optimization problems on graphs, including graph partitioning and community detection [61], [62].

Figure 4 presents the results. We observe that using optimal parameters from similar problems allows optimization methods to find high-quality solutions under realistic time constraints. Thus, we are hopeful that the high cost of finding good QAOA parameters can be amortized by reusing the parameters from similar problems.

VI. Discussion

This paper present results on finding optimal QAOA parameters to improve the performance of quantum optimization solvers. We show that multistart methods such as APOSMM can utilize a fixed number of function evaluations more efficiently by interleaving multiple local optimization runs and considering all $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ parameters generated by them. We observe that as the number of QAOA steps and the dimension of the corresponding optimization domain $\mathcal{D}$ is increased, the optimization problem becomes increasingly hard. These results highlight the need to develop more efficient approaches to finding optimal parameters to accelerate and improve the performance of QAOA—a challenge because, in order to compete with state-of-the-art classical solvers on problems with fewer than 200 variables, QAOA has to run in no more than a minute [6], [25]. An additional challenge is presented by the high levels of noise on near-term hardware.

We show that the obstacles can be partially addressed by reusing optimal parameters found for a similar problem. We observe that parameters can be reused both for similar problems with a random change introduced and in "worst-case" scenarios, where the change in the underlying problem has the greatest impact on its structure. For example, reusing optimal parameters found for $p = 1$ using BOBYQA or APOSMM for a dynamic graph over 1,000 changes and allowing local methods a realistic 10–30 iterations in order to refine reused optimal points at each iteration, would bring amortized cost down from (1 ms per run) $\times$ (1,000 measurements per run) $\times$ (1,000 evaluations) $\approx$ 16 min minutes to a more competitive $\approx$ 1 second. Reusing parameters and employing heuristic techniques such as FOURIER proposed in [11] could bring down amortized costs of QAOA run even further. We believe this could make quantum optimization solvers a valuable extreme-computing resource.

The limited connectivity between qubits in many hardware implementations presents an additional challenge. For example, superconducting qubit technology, developed by, among others, IBM, Rigetti, and Google, provides only nearest-neighbor connectivity with qubits arranged on a two-dimensional lattice. The modularity maximization graph clustering problem discussed in this paper requires all-to-all connectivity. The connectivity limitation can be addressed by a SWAP network [21], [63], [64] with only $O(N)$ overhead (where N is the number of qubits). Additionally, ion-trap architectures (the most famous implementation developed by IonQ) do not have the same connectivity limitations, because they allow the application of gates between any pair of qubits.

All these factors strengthen the potential of QAOA. As hardware continues to improve and more advanced techniques for parameter optimization are developed, QAOA has the potential to outperform classical state-of-the-art solvers.

Acknowledgments

The authors would like to acknowledge Yuri Alexeev for insightful discussions. This material was based upon work supported by the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357. Clemson University is acknowledged for generous allotment of compute time on Palmetto cluster.

References

[1] C. Ballance, T. Harty, N. Linke, M. Sepiol, and D. Lucas, "High-fidelity quantum logic gates using trapped-ion hyperfine qubits," *Physical Review Letters*, vol. 117, no. 6, p. 060504, 2016.

[2] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank *et al.*, "Superconducting quantum circuits at the surface code threshold for fault tolerance," *Nature*, vol. 508, no. 7497, p. 500, 2014.

[3] J. Preskill, "Quantum computing in the NISQ era and beyond," *arXiv:1801.00862*, 2018.

[4] P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in *Proceedings 35th Annual Symposium on Foundations of Computer Science*. IEEE, 1994, pp. 124–134.

[5] M. Roetteler, M. Naehrig, K. M. Svore, and K. Lauter, "Quantum resource estimates for computing elliptic curve discrete logarithms," in *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 2017, pp. 241–270.

[6] R. Shaydulin, H. Ushijima-Mwesigwa, I. Safro, S. Mniszewski, and Y. Alexeev, "Network community detection on small quantum computers," *Advanced Quantum Technologies (to appear), arXiv:1810.12484*, 2019.

[7] ——, "Community detection across emerging quantum architectures," *Proceedings of the 3rd International Workshop on Post Moore's Era Supercomputing*, 2018.

[8] H. Ushijima-Mwesigwa, C. F. Negre, and S. M. Mniszewski, "Graph partitioning using quantum annealing on the D-Wave system," in *Proceedings of the Second International Workshop on Post Moore's Era Supercomputing*. ACM, 2017, pp. 22–29.

[9] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou *et al.*, "A variational eigenvalue solver on a photonic quantum processor," *Nature Communications*, vol. 5, p. 4213, 2014.

[10] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," *arXiv:1411.4028*, 2014.

[11] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, "Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices," *arXiv:1812.01041*, 2018.

[12] J. Larson and S. M. Wild, "A batch, derivative-free algorithm for finding multiple local minima," *Optimization and Engineering*, vol. 17, no. 1, pp. 205–228, 2016.

[13] ——, "Asynchronously parallel optimization solver for finding multiple minima," *Mathematical Programming Computation*, vol. 10, no. 3, pp. 303–332, 2018.

[14] D. J. Watts, "Networks, dynamics, and the small-world phenomenon," *American Journal of Sociology*, vol. 105, no. 2, pp. 493–527, 1999.

[15] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using NetworkX," in *Proceedings of the 7th Python in Science Conference (SciPy 2008)*, G. Varoquaux, T. Vaught, and J. Millman, Eds., Pasadena, CA USA, 2008, pp. 11–15.

[16] M. J. Powell, "A direct search optimization method that models the objective and constraint functions by linear interpolation," in *Advances in Optimization and Numerical Analysis*. Springer, 1994, pp. 51–67.

[17] M. Powell, "Direct search algorithms for optimization calculations," *Acta Numerica*, vol. 7, pp. 287–336, 1998.

[18] S. G. Johnson, "The NLopt nonlinear-optimization package," 2019. [Online]. Available: http://github.com/stevengj/nlopt

[19] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm applied to a bounded occurrence constraint problem," *arXiv:1412.6062*, 2014.

[20] B. Barak, A. Moitra, R. O'Donnell, P. Raghavendra, O. Regev *et al.*, "Beating the random assignment on constraint satisfaction problems of bounded degree," *arXiv:1505.03424*, 2015.

[21] G. E. Crooks, "Performance of the quantum approximate optimization algorithm on the maximum cut problem," *arXiv:1811.08419*, 2018.

[22] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *Journal of the ACM*, vol. 42, no. 6, pp. 1115–1145, 1995.

[23] O. D. Parekh, C. Ryan-Anderson, and S. Gharibian, "Quantum optimization and approximation algorithms." Tech. Rep., 2019. [Online]. Available: https://doi.org/10.2172%2F1492737

[24] Z. Wang, S. Hadfield, Z. Jiang, and E. G. Rieffel, "Quantum approximate optimization algorithm for maxcut: A fermionic view," *Physical Review A*, vol. 97, p. 022304, 2018.

[25] G. G. Guerreschi and A. Y. Matsuura, "QAOA for max-cut requires hundreds of qubits for quantum speed-up," *Scientific Reports*, vol. 9, no. 1, May 2019. [Online]. Available: https://doi.org/10.1038/s41598-019-43176-9

[26] M. E. J. Newman, "From the cover: Modularity and community structure in networks," *Proceedings of the National Academy of Science*, vol. 103, pp. 8577–8582, 2006.

[27] U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hoefer *et al.*, "Maximizing modularity is hard," *arXiv:physics/0608255*, 2006.

[28] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, p. 814, 2005.

[29] G. Su, A. Kuchinsky, J. H. Morris, D. J. States, and F. Meng, "GLay: community structure analysis of biological networks," *Bioinformatics*, vol. 26, no. 24, pp. 3135–3137, 2010.

[30] G. Bardella, A. Bifone, A. Gabrielli, A. Gozzi, and T. Squartini, "Hierarchical organization of functional connectivity in the mouse brain: A complex network approach," *Scientific Reports*, vol. 6, p. 32060, 2016.

[31] C. Nicolini, C. Bordier, and A. Bifone, "Community detection in weighted brain connectivity networks beyond the resolution limit," *Neuroimage*, vol. 146, pp. 28–39, 2017.

[32] C. F. Negre, H. Ushijima-Mwesigwa, and S. M. Mniszewski, "Detecting multiple communities using quantum annealing on the D-Wave system," *arXiv:1901.09756*, 2019.

[33] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink *et al.*, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," *Nature*, vol. 549, no. 7671, p. 242, 2017.

[34] E. Farhi, J. Goldkanstone, S. Gutmann, and H. Neven, "Quantum algorithms for fixed qubit architectures," *arXiv:1703.06199*, 2017.

[35] G. G. Guerreschi and M. Smelyanskiy, "Practical optimization for hybrid quantum-classical algorithms," *arXiv:1701.01450*, 2017.

[36] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, and N. Killoran, "Pennylane: Automatic differentiation of hybrid quantum-classical computations," *arXiv:1811.04968*, 2018.

[37] J. Romero, R. Babbush, J. McClean, C. Hempel, P. Love, and A. Aspuru-Guzik, "Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz," *Quantum Science and Technology*, 2018.

[38] A. Harrow and J. Napp, "Low-depth gradient measurements can improve convergence in variational hybrid quantum-classical algorithms," *arXiv:1901.05374*, 2019.

[39] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.

[40] G. Verdon, M. Broughton, and J. Biamonte, "A quantum algorithm to train neural networks using low-depth circuits," *arXiv:1712.05304*, 2017.

[41] J. Otterbach, R. Manenti, N. Alidoust, A. Bestwick, M. Block *et al.*, "Unsupervised machine learning on a hybrid quantum computer," *arXiv:1712.05771*, 2017.

[42] D. Wecker, M. B. Hastings, and M. Troyer, "Training a quantum optimizer," *Physical Review A*, vol. 94, no. 2, p. 022309, 2016.

[43] Z.-C. Yang, A. Rahmani, A. Shabani, H. Neven, and C. Chamon, "Optimizing variational quantum algorithms using Pontryagin's minimum principle," *Physical Review X*, vol. 7, no. 2, p. 021027, 2017.

[44] D. Zhu, N. Linke, M. Benedetti, K. Landsman, N. Nguyen *et al.*, "Training of quantum circuits on a hybrid quantum computer," *arXiv:1812.08862*, 2018.

[45] 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:1812.04170*, 2018.

[46] T. Matsumine, T. Koike-Akino, and Y. Wang, "Channel decoding with quantum approximate optimization algorithm," *arXiv:1903.02537*, 2019.

[47] K. M. Nakanishi, K. Fujii, and S. Todo, "Sequential minimal optimization for quantum-classical hybrid algorithms," *arXiv:1903.12166*, 2019.

[48] Rigetti, "Grove," 2019. [Online]. Available: https://github.com/rigetti/grove

[49] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, "The theory of variational hybrid quantum-classical algorithms," *New Journal of Physics*, vol. 18, no. 2, p. 023023, 2016.

[50] M. J. D. Powell, "The BOBYQA algorithm for bound constrained optimization without derivatives," University of Cambridge, Tech. Rep. DAMTP 2009/NA06, 2009. [Online]. Available: http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf

[51] A. H. G. Rinnooy Kan and G. T. Timmer, "Stochastic global optimization methods, part I: Clustering methods," *Mathematical Programming*, vol. 39, no. 1, pp. 27–56, 1987.

[52] ——, "Stochastic global optimization methods, part II: Multi level methods," *Mathematical Programming*, vol. 39, no. 1, pp. 57–78, 1987.

[53] G. Aleksandrowicz, T. Alexander, P. Barkoutsos, L. Bello, Y. Ben-Haim *et al.*, "Qiskit: An open-source framework for quantum computing," 2019.

[54] M. J. Powell, "The BOBYQA algorithm for bound constrained optimization without derivatives," *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge*, pp. 26–46, 2009.

[55] M. J. D. Powell, "The NEWUOA software for unconstrained optimization without derivatives," in *Large-scale Nonlinear Optimization*. Springer, 2006, pp. 255–297.

[56] R. P. Brent, *Algorithms for minimization without derivatives*. Courier Corporation, 2013.

[57] T. H. Rowan, "Functional stability analysis of numerical algorithms." Ph.D. dissertation, University of Texas at Austin, 1990.

[58] S. Hudson, J. Larson, S. M. Wild, and D. Bindel, "libEnsemble users manual," 2019. [Online]. Available: https://buildmedia.readthedocs.org/media/pdf/libensemble/latest/libensemble.pdf

[59] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

[60] [Online]. Available: https://github.com/rsln-s/Multistart-Methods-for-Quantum-Approximate-Optimization

[61] F. R. Chung and F. C. Graham, *Spectral graph theory*. American Mathematical Soc., 1997, no. 92.

[62] R. R. Nadakuditi and M. E. Newman, "Graph spectra and the detectability of community structure in networks," *Physical Review Letters*, vol. 108, no. 18, p. 188701, 2012.

[63] E. Anschuetz, J. Olson, A. Aspuru-Guzik, and Y. Cao, "Variational quantum factoring," in *International Workshop on Quantum Technology and Optimization Problems*. Springer, 2019, pp. 74–85.

[64] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K.-L. Chan, "Low-depth quantum simulation of materials," *Physical Review X*, vol. 8, no. 1, Mar. 2018. [Online]. Available: https://doi.org/10.1103/physrevx.8.011044