Linearly simplified QAOA parameters and transferability

Ryo Sakai*, Hiromichi Matsuyama*, Wai-Hong Tam*, Yu Yamashiro*, and Keisuke Fujii†‡§

*Jij Inc., 1-4-6 Nezu, Bunkyo, Tokyo 113-0031, Japan
†School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan
‡Center for Quantum Information and Quantum Biology, Osaka University,
1-2 Machikaneyama, Toyonaka, Osaka 560-0043, Japan
§RIKEN Center for Quantum Computing (RQC), 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
r.sakai@j-ij.com, h.matsuyama@j-ij.com, w.tam@j-ij.com, y.yamashiro@j-ij.com, and fujii@qc.ee.es.osaka-u.ac.jp

Abstract—**Quantum Approximate Optimization Algorithm (QAOA) provides a way to solve combinatorial optimization problems using quantum computers. QAOA circuits consist of time evolution operators by the cost Hamiltonian and of state mixing operators, and embedded variational parameter for each operator is tuned so that the expectation value of the cost function is minimized. The optimization of the variational parameters is taken place on classical devices while the cost function is measured in the sense of quantum. To facilitate the classical optimization, there are several previous works on making decision strategies for optimal/initial parameters and on extracting similarities among instances. In our current work, we consider simplified QAOA parameters that take linear forms along with the depth in the circuit. Such a simplification, which would be suggested from an analogy to quantum annealing, leads to a drastic reduction of the parameter space from 2p to 4 dimensions with the any number of QAOA layers p. In addition, cost landscapes in the reduced parameter space have some stability on differing instances. This fact suggests that an optimal parameter set for a given instance can be transferred to other instances. In this paper we present some numerical results that are obtained for instances of the random Ising model and of the max-cut problem. The transferability of linearized parameters is demonstrated for randomly generated source and destination instances, and its dependence on features of the instances are investigated.**

Index Terms—**Optimization, Quantum Approximate Optimization Algorithm, Parameter Transferability**

I. INTRODUCTION

Quantum Approximate Optimization Algorithm (QAOA) [1], [2] is an algorithm to approach combinatorial optimization problems using quantum computers [3]. This can be regarded as a quantum-classical hybrid algorithm; one iteratively measures the cost Hamiltonian on a variationally parametrized quantum circuit and tunes the parameters on a classical computer so that the expectation value of the cost Hamiltonian is minimized. Once the optimal parameters are obtained, the solution of the problem is expressed as the lowest energy state (or the energy itself) of the cost Hamiltonian.

QAOA has some advantages compared to other variational quantum algorithms like VQE [4], *e.g.* relatively smaller number of parameters and problem dependent construction of ansatz state. Taking the optimization is, however, not a trivial task, and there is an arbitrariness in optimizing the parameters for a given Hamiltonian and circuit. As a typical situation, the expression power of QAOA circuit depends on the number of layers (whose detail is given later), but increasing layers often makes the parameters being trapped at local optimal and causes vanishing gradient. Against this fact, there are several approaches to seek for the decision strategy of the parameters in both theoretical and heuristic sense [5]–[10].

As well as the decision strategy of the parameters, there are some motivations to consider transferring a set of parameters tuned for a given instance to other instances [11]–[16]. Indeed, some of previous works report the concentration phenomenon, where optimized parameters take similar values regardless of whatever the specific instance is [17]–[19]. When considering applications on real devices, transferring the variational parameters is important as an economical approach to systems with a large number of qubits. In such systems, even taking a single sample costs much, so that it is not tolerable to iteratively calculate the energy expectation value for the classical optimization of variational parameters, and thus parameter transferring from a small instance to large instances will give tremendous benefits. Also, in Reference [20], it is pointed out that the parameter transferring is a promising approach to avoid the notorious barren plateau problem.

In another context, parameter transferring will be useful when considering relaxing constrained problems to Quadratic Unconstrained Binary Optimization problems using *e.g.* the penalty method [21]. In the penalty method, constraints are converted to weighted penalties, where the optimal weights are not always trivial; then one needs to take an optimization of the weights. Therefore, when applying QAOA to such problems, the optimization of the QAOA parameters and the optimization of the weights of penalties form a nested structure, and the cost of the optimization becomes multiplicatively severe.

Motivated both to decision strategy of the parameters and to transferability, we consider simplified QAOA parameters that take linear forms along with the depth of the QAOA

layers. Indeed, an analogy to the trotterized quantum annealing suggests the linear initial QAOA parameters [7]; on the other hand, in the current work we fix the parameters to be linear instead of using them as initial values for the optimization. Such simplification drastically reduces the dimension of parameter space, and, in the reduced parameter space, cost landscapes show a similarity among instances as shown in the following sections. Moreover, although one would expect that simplifying parameters contaminates the performance of QAOA, it surprisingly turns out that there exists a set of parameters that performs well regardless of what the instance to be solved is. This fact suggests that the optimal linear parameters have some transferability among different instances, and indeed we check the performance of a transferred parameter set that is found for a specific instance. It will be found in the numerical section that many instances are actually solvable by the transferred linear parameter set without any fine-tuning for each instance while the quality of parameter transferring depends on specific features of the transfer destination.

This paper is organized as follows. In Secs. II–III we describe QAOA and the detail of the model dealt with. The numerical results are shown in Sec. IV, and some remarks that are mainly suggested from previous other works are given in Sec. V. Finally in Sec. VI we summarize this paper and give future outlook.

II. METHOD

A. Quantum Approximate Optimization Algorithm

We assume we want to minimize a given cost function $C(z)$ that takes an n-length bit string $z = z_1 z_2 \cdots z_n$ as an argument. In principle, one can construct a Hamiltonian whose expectation value in the calculation basis denotes the cost function, and, in this sense, solving the combinatorial optimization problem is to obtain the lowest energy state of the Hamiltonian. Locating the lowest energy state in the Hilbert space is, however, not always a trivial task for generic Hamiltonians [1]. Here, and in this paper, we do not in terminology distinguish the energy and the cost function.

Quantum Adiabatic Algorithm (QAA) [23], [24] is an algorithm to get the target state by considering an adiabatic evolution of time dependent Hamiltonian that starts from a trivial Hamiltonian, whose ground state is known, to the target Hamiltonian, whose ground state is unknown. QAOA is derived from a trotterization of QAA and turns to QAA in the $p \to \infty$ limit, where p is the number of QAOA layers. Despite the equivalence in the limit, there are some known classes of problems where QAOA outperforms QAA [2]; thus QAOA is mentioned as a promising approach to combinatorial problems.

A QAOA circuit is an iteration of layers that consist of a time evolution operator by the cost Hamiltonian $e^{-i\gamma_l C}$ and of a state mixing operator $e^{-i\beta_l H_{\mathrm{mix}}}$ for each, where l

denotes the depth of the layer. H_{mix}, which is called the mixer Hamiltonian, generically takes the form $H_{\mathrm{mix}} = \sum_j X_j$ so that every qubit is flipped independently [3]. As the initial state, the superposition of all possible state $|+\rangle^{\otimes n}$ is taken. By applying $e^{-i\gamma_l C} e^{-i\beta_l H_{\mathrm{mix}}}$ repeatedly for each l to the initial state, one obtains the final state $|\gamma, \beta\rangle$ that depends on the parameters and can measure the expectation value of the cost function as $\langle \gamma, \beta | C | \gamma, \beta \rangle$. This procedure is iterated with tuning the variational parameters so that the expectation value converges to the minimum. The observation of cost function is taken place on a quantum computer, and the optimization of parameters is taken place on a classical computer, so that this algorithm is classified in quantum-classical hybrid algorithms.

Note that the elements of γ and β are independent of those for each other layer, and then a p-layer QAOA circuit has $2p$ parameters. The sufficient number of layers is not trivial, and also there is an arbitrariness in how to choose initial parameters and how to take the optimization. In principle, the more the number of layers is, the richer expression power the circuit has; however increasing the number of layers makes the numerical complexity more demanding in both quantum and classical side. Also, when having a deep circuit on a real device, there should be a large number of noise sources in the system. Thus, in practical uses of QAOA, one needs to fix the minimum sufficient number of layers and the suitable decision strategy for the parameters γ, β; this point is a motivation to consider the simplified decision strategy and transferability at once in our current work.

B. Linearly simplification of QAOA parameters

As well as analytical studies conducted to find out the optimal parameters [5], [6], there are also heuristic approaches like layer by layer determination [28] and inter(extra)polation strategy [29]. While some previous works attempt to reduce the $2p$ dimensional (γ, β) parameter space by imposing some constraints or modifications [30], the simplest assumption would be a linear form

$$\begin{aligned} \gamma_l &= \gamma_{\mathrm{slope}} \frac{l}{p} + \gamma_{\mathrm{intcp.}}, \\ \beta_l &= \beta_{\mathrm{slope}} \frac{l}{p} + \beta_{\mathrm{intcp.}}, \end{aligned} \tag{1}$$

where l (0-based) is an index that denotes the depth of the layer. In Reference [7], the authors consider a linear choice of the initial parameters in an analogy to time scheduling for quantum annealing, and then they take an optimization where the resulting parameters are not necessarily linear; on the other hand, in our current work, we constrain the depth dependence of the parameters to be strictly linear. Under this constraint,

---

[1] Rigorously speaking, locating the ground state for a k-local Hamiltonian belongs to QMA-hard [22].

[2] For example, QAA does not improve the accuracy always by extending the adiabatic transition time [25].

[3] The mixer Hamiltonian H_{mix} in the main body of text does not depend on the detail of the problem to be solved and is designed to explore the solution space blindly. Indeed an improvement of this part is proposed in several contexts. One of them is Reference [26], where a problem dependent mixer Hamiltonian is designed so that the constraints in the problem are always satisfied; this algorithm is called Quantum Alternating Operator Ansatz. Another example is the Grover mixers [27].

the dimension of the parameter space is reduced from $2p$ to 4 for the any number of layers p.

In Sec. IV, with taking the advantage of the reduced parameter space, we show cost landscapes, find the optimal slopes and intercepts, and investigate the transferability among randomly generated instances.

III. MODEL

For a given undirected graph $G = (V, E)$, an Ising model [31], [32] where the weights ± 1 are randomly assigned on each edge is defined by

$$H = \sum_{(i,j) \in E} J_{ij} s_i s_j. \tag{2}$$

Throughout this paper we consider only connected graphs. J_{ij} denotes the weight that is assigned on the edge (i,j). For the random Ising model we do not assume any biased distribution of the signs in J, so that the positive and the negative signs appear in the equal probability. Here we do not include external fields to the Hamiltonian for simplicity. $s_i = \pm 1$ is a spin variable on the node i; therefore, the system that is described by this Hamiltonian takes $2^{|V|}$ states.

Standing away from the purely random case above, one can impose other types of condition to the distribution of signs (*cf.* the Sherrington–Kirkpatrick model [33]), and indeed if we limit the signs in J to be positive, Eq. (2) coincides to the Hamiltonian of max-cut problem

$$H_{\text{maxcut}} = -\frac{1}{2} \sum_{(i,j) \in E} (1 - s_i s_j) \tag{3}$$

up to the constant term and factor. Thus we can consider the max-cut problem as a special case of Eq. (2).

In later of this paper we consider finding the minimum energy state of the random Ising model by QAOA. We simply assign each spin variable to each qubit, so that the number of qubits n_{qubits} in later sections is identical to the number of nodes $|V|$ of the given graph. Also, for later use, we define the density of edges by the number of edges over the maximum possible number of edges in the graph: $d_{\text{edges}} = |E| / {}_{|V|}C_2$.

IV. RESULT

In this section, the measurements of energy (cost function) are taken place on an ideal simulator without any noise source. As the ideal simulator we use Qulacs [34]. Throughout this (and next) section, the number of QAOA layers is taken to be $p = 8$ unless otherwise declared.

A. Search for optimal linear parameters

If we assume the variational parameters of a p-layer QAOA circuit take the form in Eq. (1), the parameter space to be explored reduces to the four dimensional $(\gamma_{\text{slope}}, \gamma_{\text{intcp.}}, \beta_{\text{slope}}, \beta_{\text{intcp.}})$-space regardless of the number of layers p. Then, we consider finding the optimal slopes and intercepts by searching the four dimensional parameter space for a fixed instance of the random Ising model. Here we adopt a Bayesian estimation by Optuna [35] (rather than the grid

search with a fixed resolution or any other optimizer) to find the parameter set that gives the minimum energy [4].

The important features that characterize an instance of random Ising model are the number of nodes $|V|$ ($= n_{\text{qubits}}$) and the density of edges d_{edges} [5]. As the fixed instance that is used for the parameter search, we generate a graph with $(n_{\text{qubits}}, d_{\text{edges}}) = (16, 0.6)$. With fixing the number of layers p to 8 and the number of shots to 2^{14}, the obtained values are

$$\gamma_l = -0.376 \frac{l}{p} - 0.165,$$
$$\beta_l = -0.881 \frac{l}{p} + 0.913. \tag{4}$$

One may notice that β_{slope} takes a close value to $-\beta_{\text{intcp.}}$. In the analogy to the annealing time schedule, it is quite natural to set $\boldsymbol{\beta}$ so that the effect of the mixer Hamiltonian vanishes at the end of circuit.

We have conducted the same experiment for instances with other choices of $(n_{\text{qubits}}, d_{\text{edges}})$. As a result, while n_{qubits}-dependence is not so obvious, we observe that the convergence of the Bayesian estimation gets worse for d_{edges}s that are close to 1 (complete graph). For instances with $d_{\text{edges}} = 1$ the ground states have some degeneracy, so a possible scenario is that in such a situation there would be multiple quasi-optimal parameters.

B. Cost landscapes in linearized parameter space

Here we try to discover a similarity among instances of the model by seeing cost landscapes in the linearized QAOA parameter space. With the linearization, each of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ can be expressed by the two parameters (slope and intercept) regardless of the number of layers, respectively, so that one can draw a cost landscape as a heatmap that contains all layer information for $\boldsymbol{\gamma}$ with fixing $\boldsymbol{\beta}$ and *vice versa* [6].

Figure 1 shows cost landscapes in the $(\gamma_{\text{slope}}, \gamma_{\text{intcp.}})$-space. For all panels $\boldsymbol{\beta}$ is fixed to that in Eq. (4) [7]. Even though each instance is generated at random, we can observe a common structure among the instances. Also, in Fig. 2 that shows the landscapes in the $(\beta_{\text{slope}}, \beta_{\text{intcp.}})$-space in the same manner, we can find a common structure that is independent of instance as in the case of $(\gamma_{\text{slope}}, \gamma_{\text{intcp.}})$-space. The landscapes are shown for the instances with different three choices of $(n_{\text{qubits}}, d_{\text{edges}})$; of course the structure is common

---

[4]Note that one has to be careful to use the terms like "optimal" parameters and "minimum" energy since constraining the parameters to take the form (1) may limit the expression power of the circuit. Thus, we emphasize that the minimum energy obtained by the linearized parameters is not necessarily identical to the minimum energy that would be found by the original QAOA without any constraint.

[5]Other important features would be the regularity and the parity of graph [11]; however, in this section we do not limit the graphs to be regular and do not fix the parity.

[6]In original QAOA without any simplification of parameters, one can draw a heatmap layer by layer with assigning γ to an axis and β to the other. Note that, on the other hand, the heatmaps in this paper contains all layer information.

[7]Of course one can retake the Bayesian estimation for $\boldsymbol{\beta}$ for each instance instead of adopting Eq. (4). Even in that case the rough structure of landscapes is not affected much.
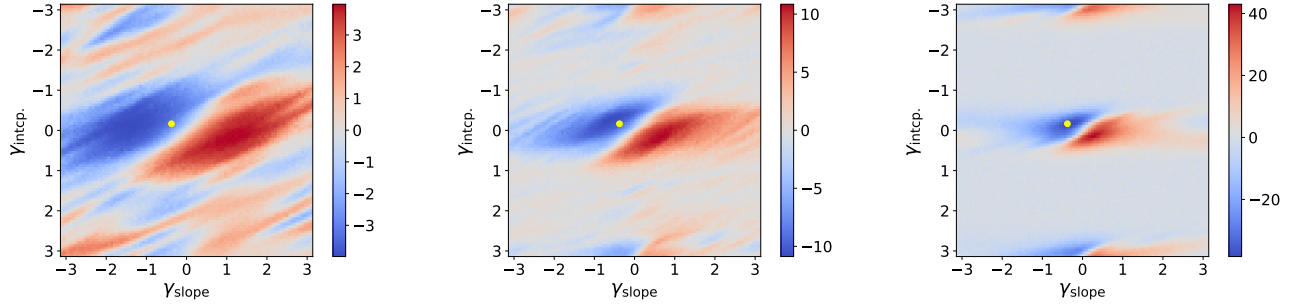
Fig. 1. Cost landscapes in $(\gamma_{\mathrm{slope}}, \gamma_{\mathrm{intcp.}})$-space. Each panel shows the cost landscape of the Ising model on a graph that is randomly generated for $(n_{\mathrm{qubits}}, d_{\mathrm{edges}}) = (5, 0.42)$ (left), $(9, 0.61)$ (center), $(21, 0.73)$ (right). We point a position that corresponds to Eq. (4) for a reference.
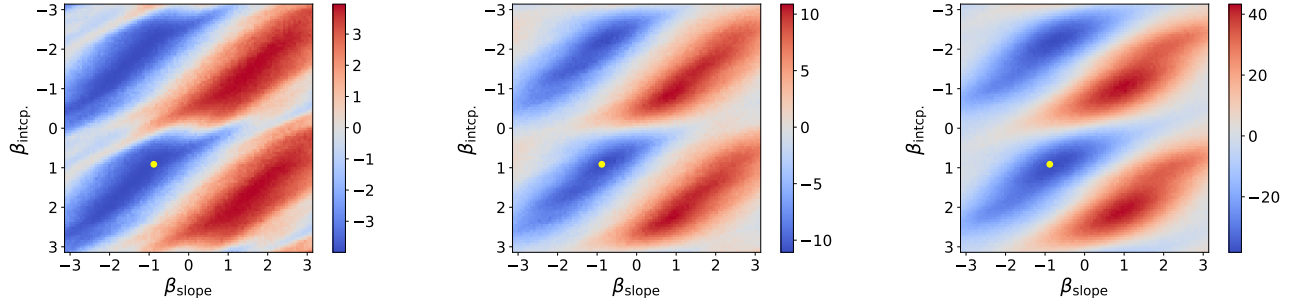

Fig. 2. Cost landscapes in $(\beta_{\mathrm{slope}}, \beta_{\mathrm{intcp.}})$-space. Each panel shows the cost landscape of the Ising model on a graph that is randomly generated for $(n_{\mathrm{qubits}}, d_{\mathrm{edges}}) = (5, 0.42)$ (left), $(9, 0.61)$ (center), $(21, 0.73)$ (right). We point a position that corresponds to Eq. (4) for a reference.

for randomly generated instances with fixed $(n_{\mathrm{qubits}}, d_{\mathrm{edges}})$ even though they are not shown here.

Thus one can expect that the optimal set of slopes and intercepts for a instance stands close to those for other instances. This result suggests that one can transfer the optimal set of parameters from and to each other instance. Indeed in the next subsection we examine the transferability of the set of linear parameters in Eq. (4) to other instances.

C. Parameter transfer to other instances

Figure 3 shows the distribution of the expectation value of energy for randomly generated Ising models with some choices of $(n_{\mathrm{qubits}}, d_{\mathrm{edges}})$. The energies are measured on 8-layer QAOA circuits where the variational parameters are fixed to those in Eq. (4). When one considers transferring the optimal parameters for an instance to another instance, there are two ways to do that: (i) reusing the completely same parameters at the destination instance and (ii) using the parameters as initial values for rerunning QAOA at the destination (imagine the transfer learning in terms of machine learning), where "destination" means an instance to where a given parameter set is transferred. On the other hand, we call an instance where an optimal set of parameters is searched "source". In this paper we define the parameter transferring as just reusing the parameters without any fine-tuning for destination instances (*i.e.* (i) in the above). Thus, for the destination instance we just measure the expectation value of energy. From the figure it is found out that there are solvable,

in the sense that the minimum energy state is most frequently observed and that the ratio of the expectation to the exact values of energy $\langle E \rangle / E_{\mathrm{exact}}$ is greater than 0.8, instances by the set of linear parameters in Eq. (4) that is found for an instance where $(n_{\mathrm{qubits}}, d_{\mathrm{edges}}) = (16, 0.6)$.

D. Dependence on features of destination instance

Figure 4 shows n_{qubits}- and d_{edges}-dependence of $\langle E \rangle / E_{\mathrm{exact}}$. They are scattering plots of $\langle E \rangle / E_{\mathrm{exact}}$ for 1024 samples of destination instances for each panel. From Fig. 4 (a), one can observe that the variance grows in the small n_{qubits} region. In Fig. 4 (b), $\langle E \rangle / E_{\mathrm{exact}}$ is greater than 0.6 for all samples in $0.1 \leq d_{\mathrm{edges}} \leq 1$ although the accuracy gets slightly worse around $d_{\mathrm{edges}} = 0.1$.

As an extreme example for comparison, Fig. 5 shows a scattering plot with varying d_{edges}, where $\langle E \rangle$s are measured with a set of linear parameters $\gamma_l = -0.790l/p - 0.259$, $\beta_l = -0.697l/p + 0.792$ that is found for an instance where $(n_{\mathrm{qubits}}, d_{\mathrm{edges}}) = (16, 0.1)$. From this result, the worsening of transferability seems to be associated with the distance of d_{edges}s between the source and destination instances. In References [15], [16], where transferring of generic (not linearized) QAOA parameters is studied for the weighted max-cut problem, it is claimed that one needs a shift of parameters according to the difference of energy scales between the source and destination instances. Even in the case of random Ising model, the typical energy of a system depends on the number of edges $|E|$, so the behavior in Fig. 5 seems to reflect this
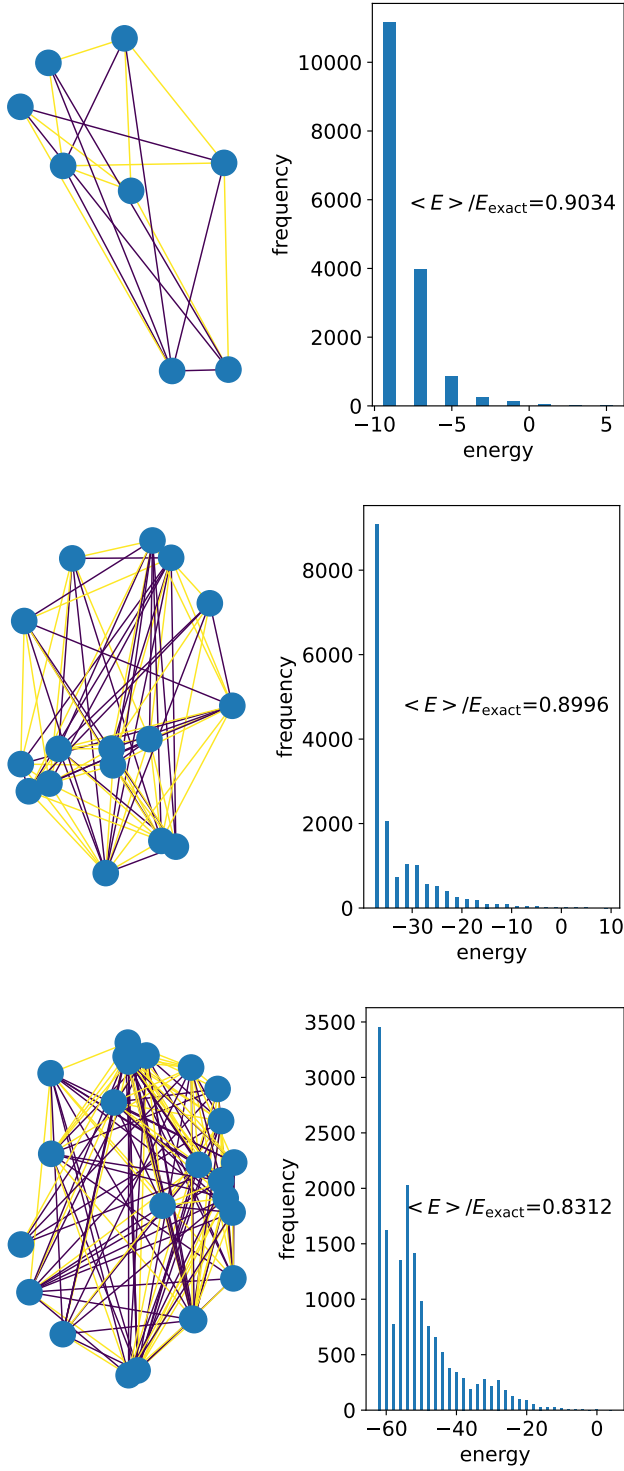
point. In conclusion, the linear parameters do not always show high transferability, and the transferability in terms of $\langle E \rangle / E_{\text{exact}}$ can get worse depending on the features of the source and destination instances. In such cases, cost landscapes like in Figs. 1–2 would show some shifts of bottoms and distortions. Indeed, one can see that zooming in and out of the $(\gamma_{\text{slope}}, \gamma_{\text{intcp.}})$-space changes with decreasing/increasing energy scale that would roughly correspond to the difference between the highest and lowest energies [8]. This point will be seen a bit more concretely in later section V-A.
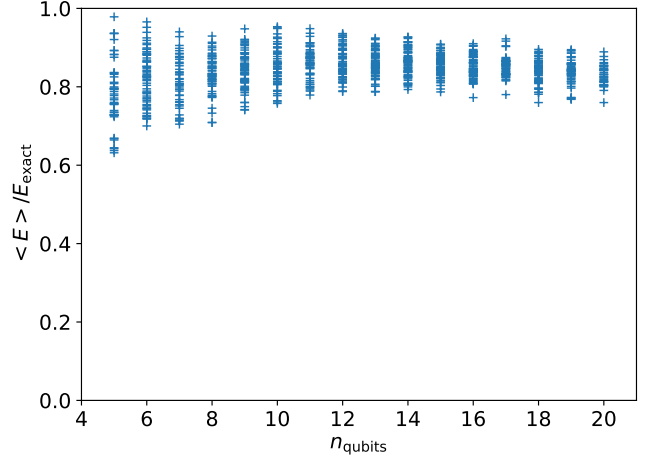
(a) n_{qubits}-dependence. 1024 samples are generated with fixed $d_{\text{edges}} = 0.6$.
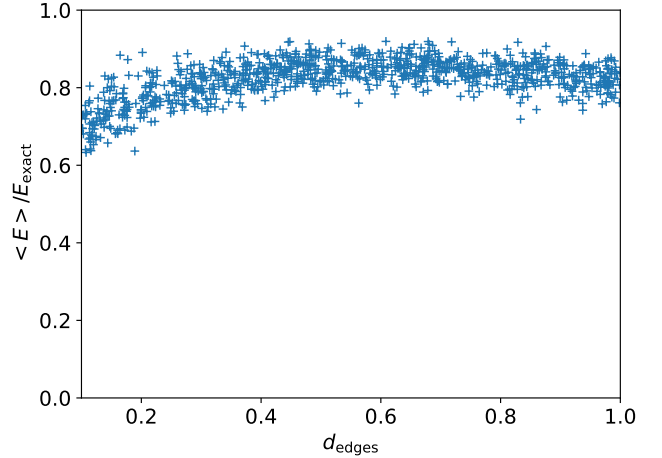
(b) d_{edges}-dependence. 1024 samples are generated with fixed $n_{\text{qubits}} = 16$.

Fig. 4. (a) n_{qubits}- and (b) d_{edges}-dependence of $\langle E \rangle / E_{\text{exact}}$. In both cases Eq. (4) is used as the variational parameters in the QAOA circuits.

Fig. 3. Results of transferring the parameter set in Eq. (4) to randomly generated instances. $(n_{\text{qubits}}, d_{\text{edges}}) = (8, 0.69)$ (top), $(16, 0.67)$ (middle), $(24, 0.59)$ (bottom). For each instance, the left panel shows the generated graph, and the right panel shows the frequency of energy value observed. In the graphs, the difference of edge colors corresponds to the difference of signs of weights ($+1$ or -1).

---

[8]On the other hand, zooming in/out of the $(\beta_{\text{slope}}, \beta_{\text{intcp.}})$-space is stable. This would be because the elements of $\boldsymbol{\beta}$ are the coefficients of H_{mix} and do not depend on the problem in a sense.
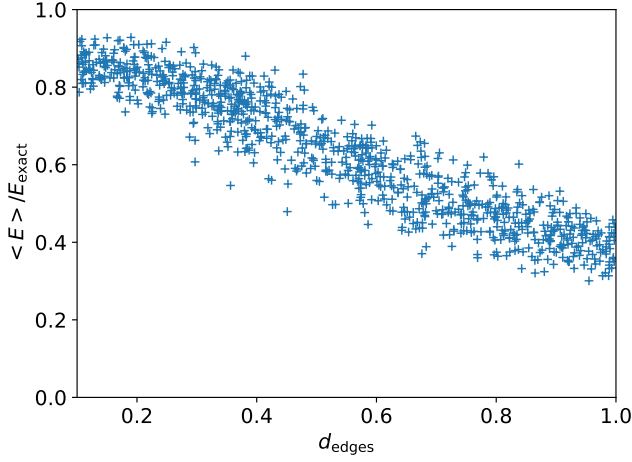
Fig. 5. d_{edges}-dependence of $\langle E \rangle / E_{\mathrm{exact}}$. 1024 samples are generated with fixed $n_{\mathrm{qubits}} = 16$. The optimal set of linear parameters found for an instance where $(n_{\mathrm{qubits}}, d_{\mathrm{edges}}) = (16, 0.1)$ is used as the variational parameters in the QAOA circuits.
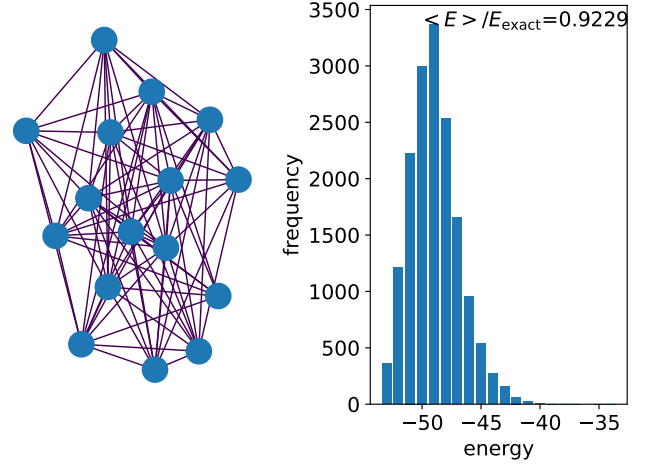
Fig. 6. The result of transferring Eq. (4) to an instance of the max-cut problem. $(n_{\mathrm{qubits}}, d_{\mathrm{edges}}) = (16, 0.68)$.

E. Parameter transferring from random Ising model to max-cut problem

Even though there is some instance dependent tendency, high transferability of the set of linear parameters in Eq. (4) is demonstrated in previous subsection. In this subsection, to check the capability of Eq. (4) a bit more radically, we show an experiment where parameter transferring from the random Ising model to the max-cut problem is taken place. Figure 6 shows the result of transferring Eq. (4) to a randomly generated instance of the max-cut problem. In this case, although the minimum energy state is not most frequently observed, $\langle E \rangle / E_{\mathrm{exact}}$ marks a value greater than 0.9, and it is suggested that there may be some transferability between problems even under different Hamiltonians.

To be a bit more concrete, we transferred Eq. (4) to randomly generated 128 instances of the max-cut problem where $d_{\mathrm{edges}} \in [0.4, 0.8]$ with fixed $n_{\mathrm{qubits}} = 16$ and calculated the average of $\langle E \rangle / E_{\mathrm{exact}}$; the result is 0.893 ± 0.023.

V. REMARKS

A. Dependence on energy scale

As seen in the previous section, there is a d_{edges}-dependence of the transferability of linear parameters, and this would be related to a dependence on the difference of energy scales between the source and destination instances. To confirm this point, in this subsection we turn to the weighted max-cut problem $H_{\mathrm{weight}} = -(1/2) \sum_{(i,j) \in E} w J_{ij} (1 - s_i s_j)$, where each element of J is uniformly sampled in $(0.1, 1)$ and where the energy scale is tunable via the weight factor w.

We randomly generate 1024 instances of the weighted max-cut problem with $|V| \in [5, 12]$, $d_{\mathrm{edges}} \in (0.1, 1)$, $J_{ij} \in (0.1, 1) \ \forall \ (i,j) \in E$, and $w \in [0.1, 1, 10, 100, 1000]$, and for each instance we estimate the optimal slopes and intercepts of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ by Optuna as done in the previous section. The

optimal $(\gamma_{\mathrm{slope}}, \gamma_{\mathrm{intcp.}})$ and $(\beta_{\mathrm{slope}}, \beta_{\mathrm{intcp.}})$ for each instance are shown in Fig. 7. While $(\beta_{\mathrm{slope}}, \beta_{\mathrm{intcp.}})$ does not depend on the weight factor w, there is a clear w-dependence of $(\gamma_{\mathrm{slope}}, \gamma_{\mathrm{intcp.}})$. This tendency is somewhat natural since the elements of $\boldsymbol{\gamma}$ are the coefficients of problem Hamiltonian. Thus a consistency of the QAOA circuit over drastic changes of the energy scale is retained by absorbing the overall factor in the Hamiltonian into $\boldsymbol{\gamma}$. Indeed, the success ratio of QAOA $\langle E \rangle / E_{\mathrm{exact}}$ is not so affected and keeps being better than 0.8 for most cases as seen in Fig. 8.

Concerning parameter transferring, one can conclude that the shift of $\boldsymbol{\gamma}$ is required in inverse proportion to the energy scale as is also discussed in References [15], [16] for the generic (not linearized) QAOA. For general models apart from the weighted max-cut problem, determining the energy scale of given system is not usually a trivial task, so that in such cases d_{edges} or the most significant weight among the edges would be used alternatively.

B. Dependence on the number of layers p

The results shown in Sec. IV are provided with the fixed number of layers $p = 8$ that might be seen to be somewhat large. In this subsection, we show cost landscapes for instances of the random Ising model with smaller choices $p = 1$ and 4.

Indeed, the parameter transferability has been intensively studied for $p = 1$ QAOA in Reference [11] with putting the focus on the parity of (sub)graph. In the paper, it is conjectured and verified for the max-cut problem that (i) transferring between regular graphs with the same degree is successful, (ii) transferring between regular graphs with same parity is successful, and (iii) transferring between regular graphs with different parity is NOT successful.

Figures 9–10 show the cost landscapes measured on $p = 1$ and 4 QAOA circuits, respectively. To make a connection to the previous work by others, the instance graphs are generated with fixed degrees for each in this subsection. In
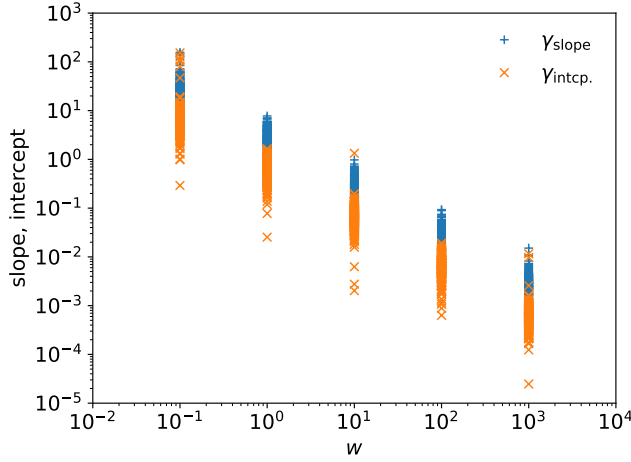
(a) Distribution of γ_{slope} and $\gamma_{\text{intcp.}}$.
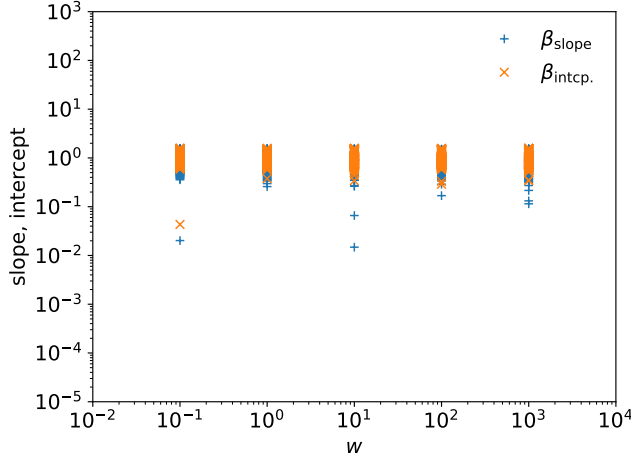
(b) Distribution of β_{slope} and $\beta_{\text{intcp.}}$.

Fig. 7. w-dependence of (a) $(\gamma_{\text{slope}}, \gamma_{\text{intcp.}})$ and (b) $(\beta_{\text{slope}}, \beta_{\text{intcp.}})$. Note that the optimal values are searched for each instance; in other words any parameter is not transferred among them. $p = 8$.

the $p = 1$ case, one can find parity dependent patterns in the $(\gamma_{\text{slope}}, \gamma_{\text{intcp.}})$-space as pointed out in Reference [11] while the landscapes in the $(\beta_{\text{slope}}, \beta_{\text{intcp.}})$-space is stable. On the other hand, in the $p = 4$ case, such a parity dependent behavior disappears, and actually the landscapes show the similar structure to those in the $p = 8$ case shown in Sec. IV-B.

C. Instance dependence of final states

When transferring (fixing) parameters, a possible concern is that similar superposition states are produced regardless of the detail of instance. To visualize the tendency of resulting superposition states of QAOA, we check the fidelity $|\langle \text{src.} | \text{dest.} \rangle|$, where $|\text{src.}\rangle$ and $|\text{dest.}\rangle$ are the QAOA states for the source and destination instances, respectively. 1024 samples are generated
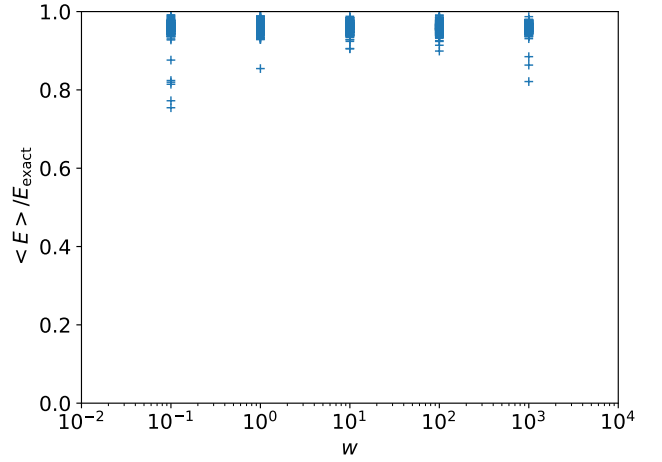

Fig. 8. w-dependence of success ratio of QAOA $\langle E \rangle / E_{\text{exact}}$. $p = 8$.

as the destination instances, where n_{qubits} is fixed to be 8 [9]. On the other hand, d_{edges} is taken from $(0.1, 1)$ at uniformly random for the destination instances while $d_{\text{edges}} = 0.6$ for the source instance.

First, for a reference, Fig. 11 shows the d_{edges}-dependence of the fidelity. There is not a clear d_{edges}-dependence of the fidelity. Note that there is a point with $|\langle \text{src.} | \text{dest.} \rangle| = 1$; this corresponds to transferring from the source to the source instance, which is shown as a reference level.

Figure 12 shows the fidelity dependence of the transferability in terms of $\langle E \rangle / E_{\text{exact}}$. As seen from the figure, high transferability is not necessarily associated with high fidelity between the QAOA states for the source and destination instances. Indeed, one can find the instances that mark $\langle E \rangle / E_{\text{exact}}$ greater than that for the source instance (see the point with $|\langle \text{src.} | \text{dest.} \rangle| = 1$). To answer the concern given in the beginning of this subsection, the features of problem Hamiltonian is embedded to the QAOA circuit, and thus the amplitude of the final state depends on instance even with parameters fixed to a certain choice.

VI. SUMMARY

In this paper, the behavior of linearized QAOA parameters and their transferability are investigated. Surprisingly, cost landscapes in the (slope, intercept)-spaces take a similar shape regardless of the detail of instance. From this fact, it is suggested that the optimal set of linear parameters found for an instance can be transferred to other instances, and indeed we demonstrated that many instances are solvable, in terms of the distribution of observed energy states and the ratio of the expectation to the exact values of energy, without instance dependent fine-tuning.
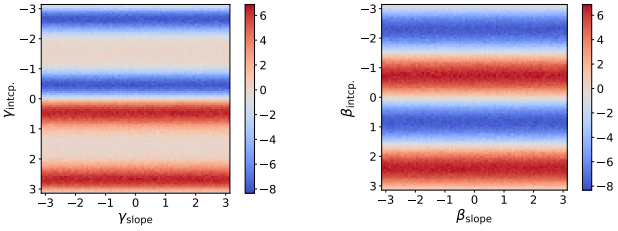
[9]Note that this means we use another parameter set instead of Eq. (4). Indeed, we have observed clear n_{qubits}-dependence of the fidelity; if we take larger $n_{\text{qubits}} = 16$, the fidelity decreases in an order of magnitude. Thus we take $n_{\text{qubits}} = 8$ in this subsection to roughly fit the range of fidelity to $(0, 1)$.
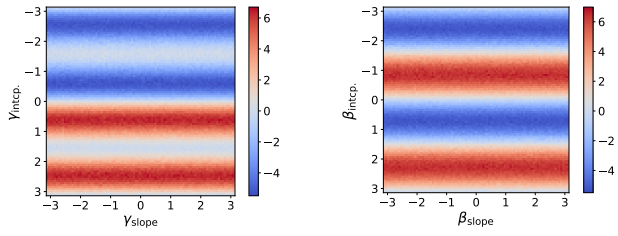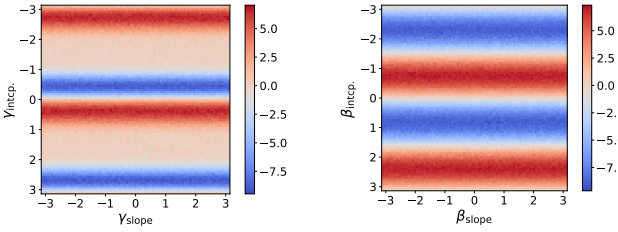
(a) Landscape for 3-regular graph.
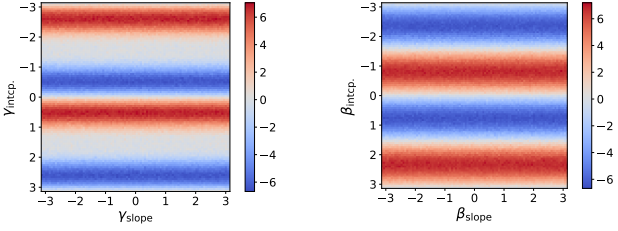
(b) Landscape for 4-regular graph.
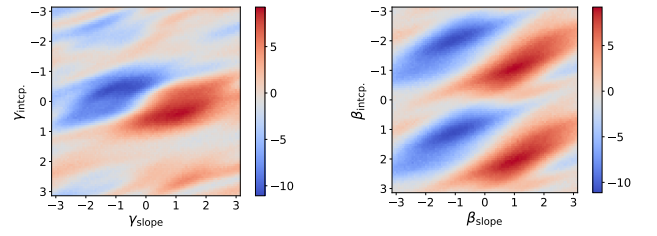

(c) Landscape for 5-regular graph.

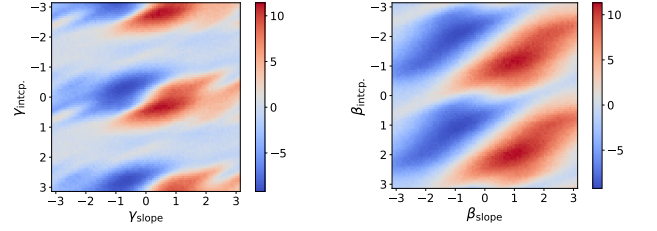(d) Landscape for 6-regular graph.

Fig. 9. Landscapes measured on $p = 1$ QAOA circuits. The instance graphs are chosen to be (a) 3-, (b) 4-, (c) 5-, and (d) 6-regular. To draw a landscape in $(\gamma_{\text{slope}}, \gamma_{\text{intcp.}})$-space, the fixed β is determined by Optuna for each instance, and *vice versa*. Note that, in $p = 1$ circuits, we cannot define the slopes of γ and β, so that the horizontal axis is indeed irrelevant. However we show the landscapes in the same format as those with other ps for comparison.

(a) Landscape for 3-regular graph.

(b) Landscape for 4-regular graph.

(c) Landscape for 5-regular graph.

(d) Landscape for 6-regular graph.

Fig. 10. Landscapes measured on $p = 4$ QAOA circuits. The instance graphs are chosen to be (a) 3-, (b) 4-, (c) 5-, and (d) 6-regular. To draw a landscape in $(\gamma_{\text{slope}}, \gamma_{\text{intcp.}})$-space, the fixed β is determined by Optuna for each instance, and *vice versa*.

The set of linear parameters in Eq. (4) that is used for the experiments is found by a Bayesian estimation for an instance where $(n_{\text{qubits}}, d_{\text{edges}}) = (16, 0.6)$. This parameter set performs high transferability so that $\langle E \rangle / E_{\text{exact}}$ is over 0.6 for every destination instance, while the transferability gets worse when the difference of d_{edges}s between the source and destination instances is large. For example, when we transfer an optimal set of linear parameters that is found for an instance where $(n_{\text{qubits}}, d_{\text{edges}}) = (16, 0.1)$ to an instance on the complete graph *i.e.* $d_{\text{edges}} = 1$, $\langle E \rangle / E_{\text{exact}}$ gets as bad as 0.3.
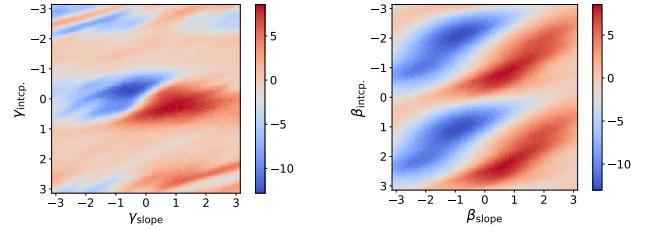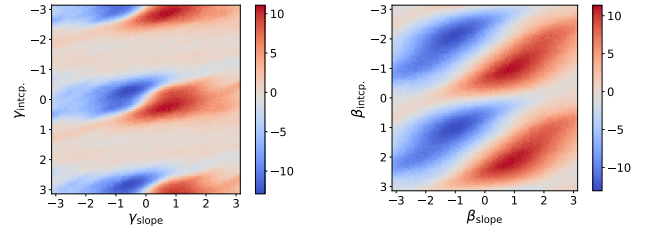
While most of the experiments in this paper are taken place

for the random Ising model, we have also tried transferring to the max-cut problem that can be regarded as a special case of the former model. As a result, there are instances of the max-cut problem where the set of parameters in Eq. (4) achieves $\langle E \rangle / E_{\text{exact}} > 0.9$ even though Eq. (4) is found for an instance of the random Ising model.

In this paper we exclusively use a noiseless simulator, so that the verification on real devices is remaining as a future work. Simplifying parameters and not taking instance dependent fine-tuning, we, in a sense, ignore the detail of instances. Thus, if the rough structure of cost landscapes like in Figs. 1–2 has some stability against the strength of noises, we can expect a transferability from ideal simulators to noisy real devices. This point should depend on the detail of devices,
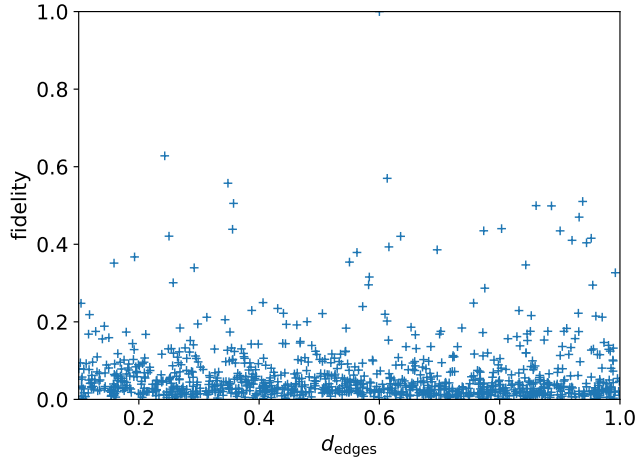
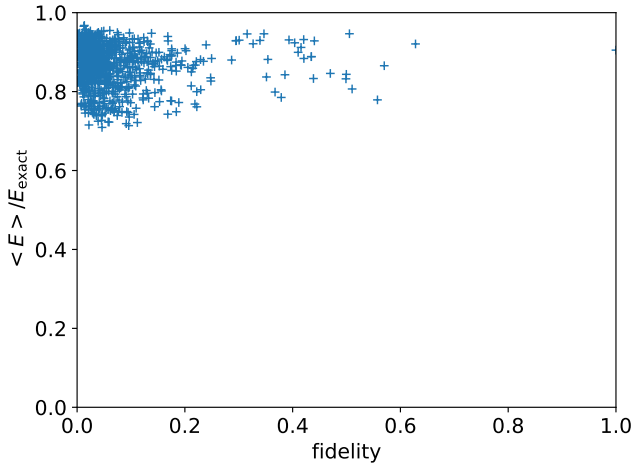Fig. 11. Fidelity as a function of d_{edges}. $p = 8$.

Fig. 12. Fidelity dependence of transferability. $p = 8$.

so extensive trials and errors on many types of devices will provide some insight towards practical uses.

Most results we report in this paper are provided with fixed $p = 8$ QAOA, but in actual use cases the number of layers should be determined in response to the accuracy, complexity, and noises (on real devices). The smaller numbers of layers would be interesting towards practical uses. Indeed, for $p \leq 2$, the linearized QAOA is equivalent to the original QAOA; thus studying how it transits between $p = 2$ and 3 would be the most important.

In practical point of view, solving penalized constrained problems with parameter transferring QAOA and checking how it is tolerable would be an interesting direction. Thanks to the transferring, one can omit (or reduce) the optimization for QAOA parameters, so that the total cost gets significantly milder. Thus we believe that transferring linearized parameters (with combined to a shift technique) contributes to actual use cases of QAOA.

REFERENCES

[1] E. Farhi, J. Goldstone, and S. Gutmann, "A Quantum Approximate Optimization Algorithm," Nov. 2014. arXiv: 1411.4028 [quant-ph].

[2] K. Blekos, D. Brand, A. Ceschini, *et al.*, "A Review on Quantum Approximate Optimization Algorithm and its Variants," Jun. 2023. arXiv: 2306.09198 [quant-ph].

[3] A. Abbas *et al.*, "Quantum Optimization: Potential, Challenges, and the Path Forward," Dec. 2023. arXiv: 2312.02279 [quant-ph].

[4] A. Peruzzo, J. McClean, P. Shadbolt, *et al.*, "A variational eigenvalue solver on a photonic quantum processor," *Nat. Commun.*, vol. 5, no. 1, p. 4213, Jul. 2014. arXiv: 1304.3061 [quant-ph].

[5] Z. Wang, S. Hadfield, Z. Jiang, and E. G. Rieffel, "Quantum approximate optimization algorithm for maxcut: A fermionic view," *Phys. Rev. A*, vol. 97, p. 022 304, 2 Feb. 2018. arXiv: 1706 . 02998 [quant-ph].

[6] S. Hadfield, T. Hogg, and E. G. Rieffel, "Analytical framework for quantum alternating operator ansätze," *Quantum Sci. Technol.*, vol. 8, no. 1, p. 015 017, Dec. 2022. arXiv: 2105.06996 [quant-ph].

[7] S. H. Sack and M. Serbyn, "Quantum annealing initialization of the quantum approximate optimization algorithm," *Quantum*, vol. 5, p. 491, Jul. 2021. arXiv: 2101.05742 [quant-ph].

[8] R. Shaydulin, I. Safro, and J. Larson, "Multistart methods for quantum approximate optimization," in *2019 IEEE High Performance Extreme Computing Conference (HPEC)*, Sep. 2019, pp. 1–8. arXiv: 1905.08768 [quant-ph].

[9] M. Streif and M. Leib, "Training the quantum approximate optimization algorithm without access to a quantum processing unit," *Quantum Sci. Technol.*, vol. 5, no. 3, p. 034 008, May 2020.

[10] X. Lee, Y. Saito, D. Cai, and N. Asai, "Parameters Fixing Strategy for Quantum Approximate Optimization Algorithm," in *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*, Aug. 2021. arXiv: 2108.05288 [quant-ph].

[11] A. Galda, X. Liu, D. Lykov, Y. Alexeev, and I. Safro, "Transferability of optimal QAOA parameters between random graphs," in *2021 IEEE International Conference on Quantum Computing and Engineering*, Jun. 2021. arXiv: 2106.07531 [quant-ph].

[12] A. Galda, E. Gupta, J. Falla, *et al.*, "Similarity-based parameter transferability in the quantum approximate optimization algorithm," *Front. Quantum Sci. Technol.*, vol. 2, Jul. 2023. arXiv: 2307.05420 [quant-ph].

[13] J. Falla, Q. Langfitt, Y. Alexeev, and I. Safro, "Graph Representation Learning for Parameter Transferability in Quantum Approximate Optimization Algorithm," Jan. 2024. arXiv: 2401.06655 [quant-ph].

[14] J. Wurtz and D. Lykov, "Fixed-angle conjectures for the quantum approximate optimization algorithm on regular maxcut graphs," *Phys. Rev. A*, vol. 104, p. 052 419, 5 Nov. 2021. arXiv: 2107.00677 [quant-ph].

[15] R. Shaydulin, P. C. Lotshaw, J. Larson, J. Ostrowski, and T. S. Humble, "Parameter Transfer for Quantum Approximate Optimization of Weighted MaxCut," *ACM Trans. Quant. Comput.*, vol. 4, no. 3, p. 19, Apr. 2023. arXiv: 2201.11785 [quant-ph].

[16] S. H. Sureshbabu, D. Herman, R. Shaydulin, *et al.*, "Parameter Setting in Quantum Approximate Optimization of Weighted Problems," *Quantum*, vol. 8, p. 1231, Jan. 2024. arXiv: 2305.15201 [quant-ph].

[17] V. Akshay, D. Rabinovich, E. Campos, and J. Biamonte, "Parameter concentrations in quantum approximate optimization," *Phys. Rev. A*, vol. 104, no. 1, p. L010401, Jul. 2021. arXiv: 2103.11976 [quant-ph].

[18] F. G. S. L. 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," Dec. 2018. arXiv: 1812.04170 [quant-ph].

[19] E. Farhi, J. Goldstone, S. Gutmann, and L. Zhou, "The Quantum Approximate Optimization Algorithm and the Sherrington-Kirkpatrick Model at Infinite Size," *Quantum*, vol. 6, p. 759, Jul. 2022. arXiv: 1910.08187 [quant-ph].

[20] A. A. Mele, G. B. Mbeng, G. E. Santoro, M. Collura, and P. Torta, "Avoiding barren plateaus via transferability of smooth solutions in a Hamiltonian variational ansatz," *Phys. Rev. A*, vol. 106, no. 6, p. L060401, Dec. 2022. arXiv: 2206.01982 [quant-ph].

[21] A. Lucas, "Ising formulations of many NP problems," *Front. Phys.*, vol. 2, p. 5, Feb. 2014. arXiv: 1302.5843 [cond-mat.stat-mech].

[22] J. Kempe, A. Kitaev, and O. Regev, "The Complexity of the Local Hamiltonian Problem," *SIAM J. Comput.*, vol. 35, no. 5, pp. 1070–1097, Jan. 2006. arXiv: quant-ph/0406180.

[23] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, "Quantum Computation by Adiabatic Evolution," Jan. 2000. arXiv: quant-ph/0001106.

[24] T. Kadowaki and H. Nishimori, "Quantum annealing in the transverse Ising model," *Phys. Rev. E*, vol. 58, no. 5, p. 5355, Nov. 1998. arXiv: cond-mat/9804280.

[25] E. Crosson, E. Farhi, C. Y.-Y. Lin, H.-H. Lin, and P. Shor, "Different Strategies for Optimization Using the Quantum Adiabatic Algorithm," Jan. 2014. arXiv: 1401.7320 [quant-ph].

[26] S. Hadfield, Z. Wang, B. O'Gorman, E. G. Rieffel, D. Venturelli, and R. Biswas, "From the Quantum Approximate Optimization Algorithm to a Quantum Alternating Operator Ansatz," Sep. 2017. arXiv: 1709.03489 [quant-ph].

[27] A. Bärtschi and S. Eidenbenz, "Grover Mixers for QAOA: Shifting Complexity from Mixer Design to State Preparation," in *2020 IEEE International Conference on Quantum Computing and Engineering*, May 2020. arXiv: 2006.00354 [quant-ph].

[28] X. Lee, X. Yan, N. Xie, Y. Saito, D. Cai, and N. Asai, "Iterative Layerwise Training for Quantum Approximate Optimization Algorithm," Sep. 2023. arXiv: 2309.13552 [quant-ph].

[29] X. Lee, N. Xie, D. Cai, Y. Saito, and N. Asai, "A depth-progressive initialization strategy for quantum approximate optimization algorithm," *Mathematics*, vol. 11, no. 9, May 2023. arXiv: 2209.11348 [quant-ph].

[30] M. Wu and H. Chen, "Adiabatic-Passage-Based Parameter Setting for Quantum Approximate Optimization Algorithm," Nov. 2023. arXiv: 2312.00077 [quant-ph].

[31] W. Lenz, "Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körpern," *Z. Phys.*, vol. 21, pp. 613–615, 1920.

[32] E. Ising, "Beitrag zur Theorie des Ferromagnetismus," *Z. Phys.*, vol. 31, pp. 253–258, 1925.

[33] D. Sherrington and S. Kirkpatrick, "Solvable Model of a Spin-Glass," *Phys. Rev. Lett.*, vol. 35, pp. 1792–1796, Dec. 1975.

[34] Y. Suzuki, Y. Kawase, Y. Masumura, *et al.*, "Qulacs: A fast and versatile quantum circuit simulator for research purpose," *Quantum*, vol. 5, p. 559, Oct. 2021. arXiv: 2011.13524 [quant-ph].

[35] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '19, Anchorage, AK, USA: Association for Computing Machinery, Jul. 2019, pp. 2623–2631. arXiv: 1907.10902 [cs.LG].