

Efficient Internal Strategies in Quantum Relaxation based Branch-and-Bound

Hiromichi Matsuyama
Jij Inc.

Tokyo, Japan

Email: h.matsuyama@j-ij.com

Wei-hao Huang
Jij Inc.

Tokyo, Japan

Email: w.huang@j-ij.com

Kohji Nishimura
Jij Inc.

Tokyo, Japan

Email: knishimura@j-ij.com

Yu Yamashiro
Jij Inc.

Tokyo, Japan

Email: y.yamashiro@j-ij.com

Abstract—A combinatorial optimization problem is to find an optimal solution under the constraints. This is one of the potential applications for quantum computers. Quantum Random Access Optimization (QRAO) is the quantum optimization algorithm that encodes multiple classical variables into a single qubit to construct a quantum Hamiltonian, thereby reducing the number of qubits required. The ground energy of the QRAO Hamiltonian provides a lower bound on the original problem's optimal value before encoding. This property allows the QRAO Hamiltonian to be used as a relaxation of the original problem, and it is thus referred to as a quantum relaxed Hamiltonian. In the Branch-and-Bound method, solving the relaxation problem plays a significant role. In this study, we developed Quantum Relaxation based Branch-and-Bound (QR-BnB), a method incorporating quantum relaxation into the Branch-and-Bound framework. We solved the MaxCut Problem and the Travelling Salesman Problem in our experiments. In all instances in this study, we obtained the optimal solution whenever we successfully computed the exact lower bound through quantum relaxation. Internal strategies, such as relaxation methods and variable selection, influence the convergence of the Branch-and-Bound. Thus, we have further developed the internal strategies for QR-BnB and examined how these strategies influence its convergence. We show that our variable selection strategy via the expectation value of the Pauli operators gives better convergence than the naive random choice. QRAO deals with only unconstrained optimization problems, but QR-BnB can handle constraints more flexibly because of the Branch-and-Bound processes on the classical computing part. We demonstrate that in our experiments with the Travelling Salesman Problem, the convergence of QR-BnB became more than three times faster by using the information in the constraints.

Index Terms—Quantum Applications, Optimization, Branch-and-Bound, Quantum Relaxation, Maximum Cut, Travelling Salesman Problem, Hybrid Algorithm, Performance Evaluation

I. INTRODUCTION

Combinatorial optimization problems involve finding a solution that minimizes (or maximizes) an objective function under constraints [1]–[3]. Various problems, such as the Vehicle Routing Problem, Crew Scheduling Problem, and Facility Location Problem, can be formulated as Combinatorial optimization problems. Therefore, various classical algorithms have been proposed to obtain an optimal or practically useful solution. The branch-and-bound method is a classical exact

solution algorithm [4], [5]. Its techniques are widely used in many commercial and non-commercial software programs, such as Gurobi [6] and CBC [7].

Recent developments in both hardware and software in quantum computation have led to various potential applications. One application that draws attention is the combinatorial optimization problem [8]. Current quantum devices, which lack error correction and have fewer than a few hundred physical qubits, are called Noisy Intermediate Scale Quantum (NISQ) devices [9], [10]. The Quantum Approximate Optimization Algorithm (QAOA) [11], [12] is a NISQ algorithm for combinatorial optimization inspired by adiabatic quantum computation [13]–[15]. QAOA encodes a combinatorial optimization problem into an Ising Hamiltonian and searches its ground state as an optimal solution. QAOA and its variants have been applied to various NP-hard problems [12] such as the MaxCut problem [11], [16] and the graph coloring problem [17], [18].

QAOA has some issues as a method for solving combinatorial optimization problems. The first issue is the number of qubits required to describe the problem. We need as many qubits as the number of classical bits in the mathematical model. Several quantum optimization algorithms have been proposed to reduce the number of required qubits [19]–[22]. Quantum Random Access Optimization (QRAO) is a method to reduce the number of qubits by encoding multiple classical bits into a small number of qubits using a Quantum Random Access Code (QRAC) [19], [23], [24]. QAOA only uses the Pauli Z operators to encode the problem into the Hamiltonian. However, in QRAO, Pauli X, Y operators are also used to encode. Due to the difference between the QRAO Hamiltonian $\tilde{H}$ and the Ising Hamiltonian H , these ground states are different. The ground state of $\tilde{H}$ can be an entangled quantum state. Therefore, a rounding algorithm must be applied to obtain a classical solution from the ground state of $\tilde{H}$. Fuller *et al.* showed that QRAO provides an approximate solution with the approximation ratio under an appropriate Rounding algorithm [19]. Moreover, they showed that the ground energy of $\tilde{H}$ becomes a lower bound on the original problem.

The second issue with QAOA is the handling of constraints. The constrained optimization problem must be converted to Quadratic Unconstrained Binary Optimization (QUBO) using the relaxation method, such as the Penalty method, to encode

This work was performed for Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), “Promoting the application of advanced quantum technology platforms to social issues”(Funding agency : QST).

into the Ising Hamiltonian. Relaxation of constraint is a problem-independent technique, but it requires tuning weights of penalty terms. Since tuning weights is an optimization problem, tuning weights in problems with many constraints becomes challenging. Quantum Alternating Operator Ansatz has been proposed to overcome this problem [25]. However, it requires the design of a problem-specific quantum circuit, which tends to increase circuit depth.

The ground energy of the QRAO Hamiltonian $\tilde{H}$ gives a lower bound on the original problem. The problem is known as quantum relaxation. Solving relaxation problems is a major subroutine of the branch-and-bound method, the classical exact solution method. In [26], we proposed the concept of Quantum Relaxation based Branch-and-Bound (QR-BnB), a quantum-classical hybrid algorithm using quantum relaxation within the framework of classical branch-and-bound methods, and showed that QR-BnB works by conducting experiments on a single small instance of the MaxCut problem. QR-BnB consists of two fundamental processes: the branching process, which divides the problem into subproblems, and the bounding process, which evaluates the subproblem by solving the quantum relaxation problem and prunes the subproblem tree. The branching and bounding processes are performed on the classical device, and the quantum device is used only to solve the subproblem's quantum relaxation problem. QR-BnB requires less qubits than QAOA due to quantum relaxation.

In this paper, we developed QR-BnB to handle combinatorial optimization problems more efficiently. It is known that the performance of the branch-and-bound algorithm depends on various internal strategies, such as what kind of relaxation method is applied, how to select fixing variables, and how to branch the subproblem tree. In this study, we examined several internal strategies to investigate the appropriate strategy for QR-BnB through detailed experiments using the MaxCut, an unconstrained optimization problem, and the Travelling Salesman Problem, a constrained optimization problem. The optimal solution was obtained for all problem instances when an exact lower bound on quantum relaxation was obtained. In this case QR-BnB provides the optimal solution with guarantees. We developed variable selection strategies using the expectation value of Pauli operators. Our variable selection strategies always give better convergence than Random selection of the fixing variables. We further examine the treatment of constraints in QR-BnB by branching strategies using constraints information. In particular, QR-BnB converged faster when using a branching strategy with constraints information than a branching strategy without constraints information.

We also performed experiments where the lower bound of the relaxation solution was computed using the Variational Quantum Eigensolver (VQE) [27] to evaluate its availability for NISQ devices. In this case, QR-BnB cannot guarantee optimality due to the heuristic nature of VQE. However, we found that QR-BnB provides an optimal solution in almost all cases, even in the VQE case.

The structure of this paper is as follows: In Sec. II, we explain QRAO which is the essential component of this

study. Sec. III explains QR-BnB and its development in this study. We discuss the experiment setting at Sec. IV and the experimental results using QR-BnB for the Maxcut problem and TSP at Sec. V. Finally, in Sec. VI, we summarize our results and discuss future works.

II. QUANTUM RANDOM ACCESS OPTIMIZATION

Quantum Random Access Optimization (QRAO) [19], [23], [24] is a quantum optimization algorithm that utilizes Quantum Random Access Codes (QRAC) [28], [29]. The number of qubits can be reduced compared to the conventional quantum optimization algorithm using the Ising Hamiltonian [19]. QRAO consists of two steps. In the first step, we construct a quantum relaxed Hamiltonian $\tilde{H}$ using QRAC and search for its ground state. In the second step, we apply the rounding algorithm to the quantum state $\tilde{\rho}$ to obtain the classical solution. QRAO has been proposed for the (weighted) MaxCut problem [19]. However, its construction can be applied to the general Ising Hamiltonian, so we use the general Ising Hamiltonian for the explanation.

A. Quantum Relaxed Hamiltonian

This subsection explains the construction of the relaxed Hamiltonian using $(3, 1, p)$ -QRAO as an example. $(3, 1, p)$ -QRAC encodes three classical bits $x_1, x_2, x_3 \in \{0, 1\}$ into a single qubit as

$$\rho(x_1, x_2, x_3) = \frac{1}{2} \left(I + \frac{1}{\sqrt{3}} ((-1)^{x_1} X + (-1)^{x_2} Y + (-1)^{x_3} Z) \right). \quad (1)$$

Three classical bits are encoded by mapping them to the Pauli X, Y, Z operators of a single qubit. The success probability of decoding for $(3, 1, p)$ -QRAC is $p = \frac{1}{2} \left(1 + \frac{1}{\sqrt{3}} \right) \approx 0.79$. At most $4^n - 1$ classical bits can be encoded in n qubits using QRAC with $p \geq 1/2$, so $(3, 1, p)$ -QRAC is the maximum encoding for a single qubit [30].

Next, we describe the construction of the relaxed Hamiltonian. In QAOA, optimization problems are encoded into the Ising Hamiltonian as below:

$$H = \sum_{ij} J_{ij} Z_i Z_j + \sum_i h_i Z_i. \quad (2)$$

Since encoding uses only the Pauli Z_i operators, H is diagonal, and its ground state corresponds to the classical optimal solution. In QRAO, not only Pauli Z_i operators but also Pauli X_i, Y_i operators are used to construct a relaxed Hamiltonian $\tilde{H}$. Since QRAO encodes multiple classical bits into a single qubit to construct $\tilde{H}$, we must first determine which classical bits to map to which qubit. We define the interaction graph $G(E, V)$ as an undirected graph composed of nodes $v_i, v_j \in V$ connected by edges such that $J_{ij} \neq 0$. The QRAO mapping requires that adjacent classical bits on $G(E, V)$ be encoded in different qubits. We can determine the mappings by solving the vertex coloring problem on $G(E, V)$. We can employ the greedy method to solve the vertex coloring problem since obtaining a feasible solution rather than an optimal one is

sufficient. By solving the coloring problem, nodes v_i, v_j at both ends of edge $(i, j) \in E$ have different colors $c(i) \neq c(j)$. Then, we assign up to three nodes $v_i \in V_c$ with each color $c \in C$ to each qubit. Consequently, the quantum relaxed Hamiltonian $\tilde{H}$ corresponding to eq. (2) becomes

$$\tilde{H} = \sum_{i,j} 3J_{ij}P_iP_j + \sum_i \sqrt{3}h_iP_i, \quad (3)$$

where $P_i \in \{X_i, Y_i, Z_i\}$ is the Pauli operator corresponding to node v_i . The coefficient $3, \sqrt{3}$ appears for $\tilde{\rho}_{\text{opt}}$, the state in which the optimal solution is encoded using $(3, 1, p)$ -QRAC, to satisfy $\text{Tr}(\tilde{H}\tilde{\rho}_{\text{opt}}) = \text{OPT}$ where OPT is the optimal solution to the original problem. Since $\tilde{H}$ is not diagonal, the ground state $\tilde{\rho}$ can be an entangled quantum state. Thus, $\tilde{\rho}$ does not coincide with the optimal solution of the original problem. There is also no guarantee that $\tilde{\rho}$ becomes $\tilde{\rho}_{\text{opt}}$. Interestingly, however, $\text{Tr}(\tilde{H}\tilde{\rho})$ gives a lower bound of the optimal value on the original problem.

The number of qubits required for each color is $n_c = \lceil |V_c|/3 \rceil$, which means that up to three times more compression ratio can be achieved than in eq. (2). Note that the compression ratio depends on the given Ising Hamiltonian. In particular, if the given Ising Hamiltonian is fully connected, no qubits are reduced.

The relaxed Hamiltonian for $(2, 1, p)$ -QRAO can be constructed similarly. The only difference is the number of qubits to be encoded in a single qubit and the coefficient of each term, and the relaxed Hamiltonian for $(2, 1, p)$ -QRAO is

$$\tilde{H} = \sum_{i,j} 2J_{ij}P_iP_j + \sum_i \sqrt{2}h_iP_i, \quad (4)$$

where $P_i \in \{X_i, Z_i\}$. Recently, other QRAOs have also been proposed [23].

B. Rounding Algorithms

Since the ground state $\tilde{\rho}$ of $\tilde{H}$ can be an entangled quantum state, we need rounding algorithms to decode a classical solution from $\tilde{\rho}$. Fuller *et al.* proposed two rounding algorithms, Magic State Rounding and Pauli Rounding [19]. The Magic State Rounding is a method that decodes information from three classical bits at once. This method has been proven to have an approximation ratio of $5/9$ for $(3, 1, p)$ -QRAC and $5/8$ for $(2, 1, p)$ -QRAC in the MaxCut problem.

Pauli Rounding is a method that decodes classical information by measuring the expected value $\text{Tr}[P_i\tilde{\rho}]$ of the Pauli operator P_i used to encode each classical variable for all classical variables $i \in V$. Then, the sign is taken as the value corresponding to the variable. Pauli Rounding has been reported to yield better solutions in most cases compared to Magic State Rounding by the numerical experiments [24]. However, unlike Magic State Rounding, Pauli Rounding does not have approximation guarantees, even in the case of the MaxCut problem.

Algorithm 1 Quantum Relaxation based Branch-and-Bound with strategies

Input: Original Problem S

Output: optimal solution $\mathbf{x}_{\text{inc}}$

```

1:  $L \leftarrow \{S\}, z_{\text{inc}} \leftarrow \infty$ 
2: while  $L \neq \emptyset$  do
3:   Take  $S_{\text{sub}}$  from  $L$  using tree search strategy
4:   Encode  $S_{\text{sub}}$  into Relaxed Hamiltonian  $\tilde{H}$ 
5:    $z_{\text{QR}}^{\text{sub}} \leftarrow$  ground energy of  $\tilde{H}$ 
6:    $z, \mathbf{x} \leftarrow$  objective value and state by Pauli Rounding
7:   if  $z < z_{\text{fes}}$  and  $\mathbf{x}$  is feasible then
8:      $z_{\text{inc}} \leftarrow z, \mathbf{x}_{\text{inc}} \leftarrow \mathbf{x}$ 
9:   end if
10:  if  $z_{\text{fes}} < z_{\text{QR}}^{\text{sub}}$  then
11:    Go to line 2
12:  end if
13:  Choose index  $i$  by using variable selection strategy
14:  Branch  $S_{\text{sub}}$  into set of subproblems using branching strategy and add to  $L$ 
15: end while
16: return  $\mathbf{x}_{\text{inc}}$ 
```

III. QUANTUM RELAXATION BASED BRANCH-AND-BOUND ALGORITHM

In this section, we describe the Quantum Relaxation based Branch-and-Bound (QR-BnB) algorithm and the strategy of the algorithms examined in this study. The branch-and-bound method is an algorithm for obtaining optimal solutions to the combinatorial optimization problems [2], [4], [5]. Solving the relaxation problem is a major subroutine of the branch-and-bound method. QR-BnB uses quantum relaxation as its relaxation problem. In QR-BnB, quantum devices are used only to solve relaxation problems, while classical devices are used for constraint-dependent processing and solution construction.

A. Algorithm Flow

QR-BnB employs quantum relaxation within the branch-and-bound method. The pseudocode for QR-BnB is shown in Fig. 1. The algorithm consists of a branch process that generates a subproblem tree and a bound process that prunes the subproblem tree. We consider an optimization problem with n binary variables and m constraints:

$$S : z^* = \min\{\mathbf{x}^T Q \mathbf{x} : \mathbf{x} \in X\}, \quad (5)$$

where $X = \{\mathbf{x} \in \{0, 1\}^n : A\mathbf{x} \leq \mathbf{b}\}$ is the feasible set of the problem, $\mathbf{b} \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}, Q \in \mathbb{R}^{n \times n}$.

First, we explain the branching process. The branching process divides the feasible set into smaller subsets that do not

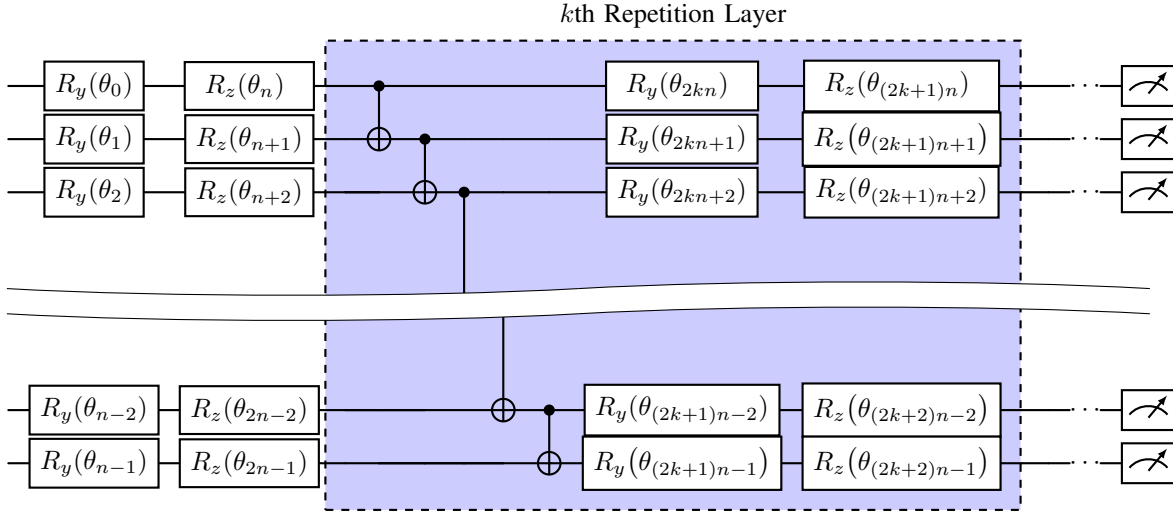


Fig. 1: Variational quantum circuit of the Hardware-Efficient Ansatz used in the experiments.

overlap. In particular, if the problem has only binary variables, this process generates two subproblems by fixing a variable:

$$\begin{aligned} S_0 : z_0^* &= \min\{\mathbf{x}^T Q \mathbf{x} : \mathbf{x} \in X_0\}, \\ S_1 : z_1^* &= \min\{\mathbf{x}^T Q \mathbf{x} : \mathbf{x} \in X_1\}, \end{aligned} \quad (6)$$

where $X_0 = X \cap \{\mathbf{x} : x_i = 0\}$, $X_1 = X \cap \{\mathbf{x} : x_i = 1\}$. We can repeat this process to generate smaller subproblems by fixing additional variables from these X_0, X_1 . The relation between subproblems can be drawn as a tree. The root node of this subproblem tree is the original problem S , and the leaves correspond to the solutions to the problem. During the search process, the solution to the problem S is found, and it is stored as an incumbent solution $\mathbf{x}_{\text{inc}}$ for the algorithm. The final incumbent solution is the output of the algorithm.

The branching process expands the subproblem tree, which is no different from a brute-force search. The bound process reduces the size of the subproblem tree by evaluating and pruning the tree. The subproblem S_{sub} is evaluated by the optimal value of quantum relaxation $\tilde{z}_{\text{QR}}^{\text{sub}}$. We denote the objective value of the feasible and optimal solution of the original problem S as z_{fes} , z^* respectively, and the optimal value of the relaxation problem as $\tilde{z}_{\text{QR}}$. As mentioned in section II-A, $\tilde{z}_{\text{QR}}$ gives a lower bound on the optimal value of S . z_{fes} gives an upper bound on the optimal value of S . Therefore, relation $\tilde{z}_{\text{QR}} \leq z^* \leq z_{\text{fes}}$ holds. On the other hand, since the feasible set of the subproblem is more bounded than S , the objective value of the relaxed subproblem $\tilde{z}_{\text{QR}}^{\text{sub}}$ does not necessarily satisfy $\tilde{z}_{\text{QR}}^{\text{sub}} \leq z_{\text{inc}}$ where z_{inc} is the objective value of the incumbent solution. This implies that no solution in the feasible set of the relaxation problem for this subproblem is better than the incumbent solution. Thus, we prune the branch below this subproblem.

One advantage of using quantum relaxation is that it can solve problems involving second-order objective functions, such as the eq. (5), which cannot be solved by simple linear relaxation. Moreover, although the subproblems are evaluated

by quantum relaxation, the search for the solution itself is performed using the branching process, so the difficulty of handling constraints in the quantum algorithm can be indirectly avoided by pruning infeasible solutions from the subproblem tree when searching for solutions. A detailed discussion about the feasibility evaluations of the subproblem is in the Appendix.

B. Strategies in Branch-and-Bound

In branch-and-bound methods, convergence can be very slow if we cannot prune the subproblem tree efficiently. It is known that convergence speed depends on internal operations. For this purpose, various methods have been proposed for internal operation strategies [5]. In this study, we examined several strategies to determine how changes in these strategies affect QR-BnB's convergence. We investigate three strategies: the variable selection strategy, the branching strategy, and the tree search strategy. In this subsection, we describe these strategies and improvements specific to QR-BnB.

First, we describe the variable selection strategy for choosing the variables to be fixed. In the case of linear relaxation, the solution is obtained as continuous values. There are two simple variable selection methods that utilize continuous values. In the Least Fractional rule, the variable closest to 1 or 0 is fixed. On the other hand, the Most Fractional rule fixes the variable farthest from 1 or 0. Unlike these methods, the Random rule selects variables to be fixed regardless of continuous values. In the case of quantum relaxation, since the relaxed solution is in a quantum state, only the Random rule can be applied. To overcome this issue, in this study, we developed a variable selection method through the expectation value $\text{Tr}[P_i \tilde{\rho}]$ of the Pauli operator P_i in the obtained quantum state $\tilde{\rho}$. We use $\text{Tr}[P_i \tilde{\rho}]$ as the continuous value in the linear relaxation

¹. These expectation values allow QR-BnB to use the Least Fractional and Most Fractional rules as variable selection strategies.

Obtaining a better incumbent solution with z_{inc} close to the optimal value at an early stage of the algorithm is crucial for reducing the branches to be explored. In the naive application of QR-BnB, the solution to the original problem can only be obtained when we arrive at the leaves of the subproblem tree. As we mentioned above, we computed the expectation value $\text{Tr}[P_i \tilde{\rho}]$ for variable selection strategies. Since taking the sign of $\text{Tr}[P_i \tilde{\rho}]$ is Pauli Rounding, we apply Pauli Rounding to each subproblem of QR-BnB to obtain one solution for the original problem to increase the frequency of z_{inc} updates.

The branching strategy is a strategy for branching the subproblem tree. The most basic branching strategy is the Binary Branch, which generates two subproblems by fixing one variable, as discussed in section III-A. However, for problems with specific constraints, we can branch the subproblem tree more efficiently with information about the constraints. Let us consider the problem with the onehot constraint $\sum_{k \in K} x_k = 1$. Once the variable x_i in the onehot constraint is fixed to 1, the other variables must be 0. Therefore, we can fix multiple variables simultaneously as $x_i = 1, x_k = 0$ ($k \in K \setminus \{i\}$). In this study, we call this method the Onehot Branch.

The tree search strategy is a method for determining the search order of a subproblem tree. A general tree search algorithm can be used for this. In this study, we examined three search methods: Depth First Search (DFS), Breath First Search (BrFS), and Best First Search (BFS).

IV. EXPERIMENT SETTINGS

In this study, we solve the MaxCut problem and Travelling Salesman Problems (TSP) as examples of unconstrained and constrained optimization problems. First, we describe these problems and their setup below.

MaxCut problem divides the node set V of an undirected graph $G(V, E)$ into two subsets. The problem is to maximize the sum of the number of edges where the nodes at both ends of the edges are assigned to different subsets. The MaxCut problem on an undirected graph $G(V, E)$ is formulated by the following mathematical model:

$$\max_s \frac{1}{2} \sum_{(i,j) \in E} (1 - s_i s_j), \quad (7)$$

with the spin variable $s_i \in \{-1, 1\}$ indicating the assignment to each subset. Since this problem is a maximization problem, we reversed the sign of the problem and solved it as a minimization problem. In this study, we evaluated the performance of QR-BnB by randomly generating 3-regular graphs with 100 samples, each from 16 to 24 nodes.

¹ $\text{Tr}[P_i \tilde{\rho}]$ satisfies $-1 \leq \text{Tr}[P_i \tilde{\rho}] \leq 1$. The value closest to 0 is fixed for the Most Fractional rule, and the values closest to 1 and -1 are fixed for the Least Fractional rule.

TSP is the problem in finding the shortest tour of N cities. This problem can be formulated as

$$\begin{aligned} \min_x \quad & \sum_t \sum_{ij} d_{ij} x_{i,t} x_{j,t+1 \bmod N} \\ \text{s.t.} \quad & \sum_{i=0}^{N-1} x_{i,t} = 1 \quad \forall t \in \{0, 1, \dots, N-1\}, \\ & \sum_{t=0}^{N-1} x_{i,t} = 1 \quad \forall i \in \{0, 1, \dots, N-1\}, \end{aligned} \quad (8)$$

using the variable $x_{i,t} \in \{0, 1\}$, which is 1 when a salesman in a city i at time t . The problem has two onehot constraints for time and city, so that Onehot Branch can be applied. QRAO requires transforming constrained optimization problems into the Ising Hamiltonians. Therefore, we transformed eq. (8) into the following QUBO:

$$\begin{aligned} \min \quad & \sum_t \sum_{ij} d_{ij} x_{i,t} x_{j,t+1 \bmod N} \\ & + A \sum_t \left(\sum_{i=0}^{N-1} x_{i,t} - 1 \right)^2 + B \sum_i \left(\sum_{t=0}^{N-1} x_{i,t} - 1 \right)^2, \end{aligned} \quad (9)$$

using the Penalty method. We then generated the Ising Hamiltonian by transforming the binary variables into spin variables. We set $A = B = 1$ in our experiment. We evaluated the performance of QR-BnB using 100 instances of 4-city problems randomly distributed in $[0, 1]^2$ as problem instances.

In this study, we changed the variable selection strategy, the branching strategy, the tree search strategy, and the quantum relaxation method to evaluate the convergence speed of QR-BnB. We used two methods to calculate the lower bound by quantum relaxation in QR-BnB. One is the computation of optimal lower bounds by exact diagonalization, and the other is by Variational Quantum Eigensolver (VQE) [27]. We refer to the QR-BnB employing exact lower bound calculations as QR-BnB-Exact and the version using VQE as QR-BnB-VQE. We used hardware-efficient ansatz [31] for the variational quantum circuit of VQE. It consists of parameterized RY, RZ gates and $CNOT$ entangling layers. In this experiment, we used a linear entangler for the entangling layer (Fig. 1) and varied the number of layers k from 1 to 3. For the optimization of the parameter in the variational quantum circuit, we used the Nakanishi-Fuji-Todo (NFT) algorithm [32] which is the gradient free-type optimizer designed for hardware-efficient ansatz. When the unfixed variables in QR-BnB were less than three bits, we performed a brute-force search without encoding into the quantum state. We used Qamomile², a quantum optimization library, to generate the quantum relaxed Hamiltonian, and qiskit [33] and QURI Parts³, an open source library for quantum computation to implement the algorithm. To obtain the optimal solution of the problems for comparison,

²<https://github.com/Jij-Inc/Qamomile>

³<https://github.com/QunaSys/quri-parts>

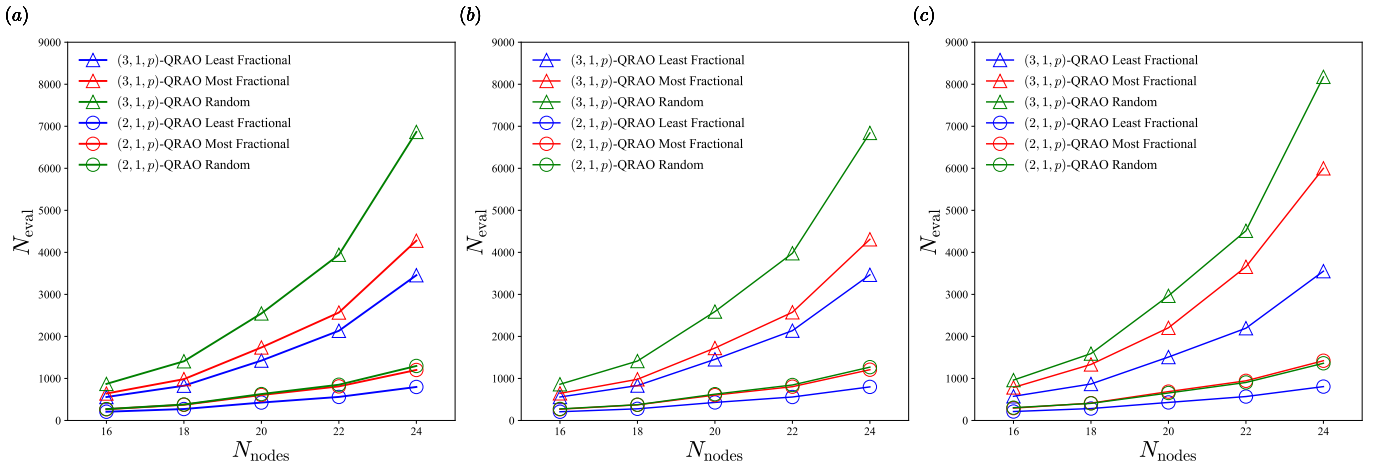


Fig. 2: N_{nodes} dependence of N_{eval} for the Maxcut problem with QR-BnB-Exact. (a) Best First Search. (b) Breath First Search. (c) Depth First Search, showing changes in convergence with different tree search strategies. Each line color corresponds to the variable selection strategies, and each mark corresponds to the quantum relaxation methods. $(2, 1, p)$ -QRAO with the Least Fractional rule converges fastest regardless of the search strategy chosen.

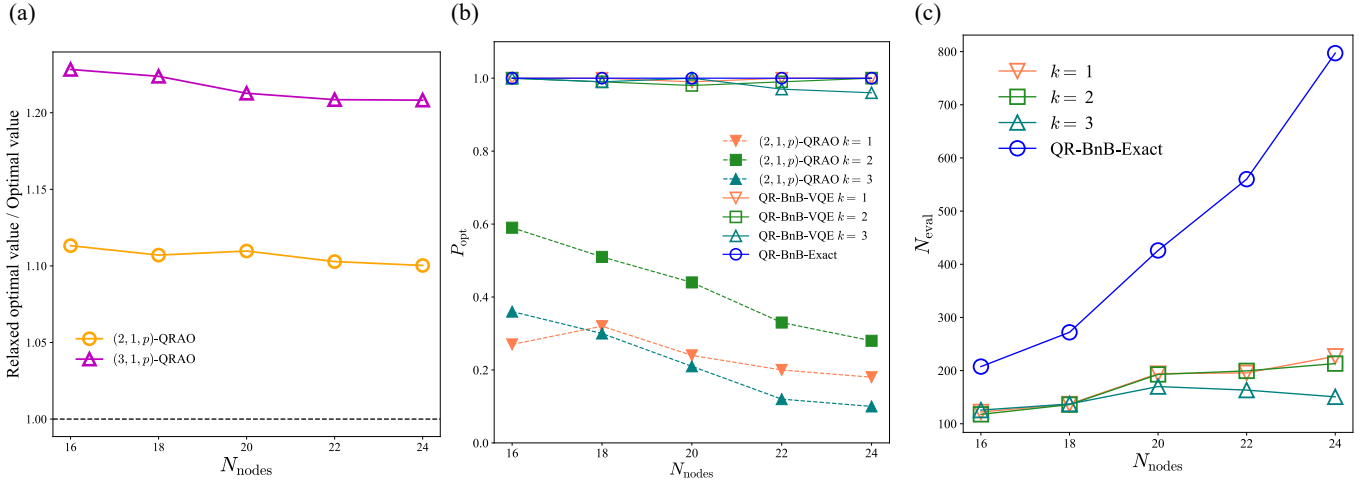


Fig. 3: (a) Comparison of the size of the quantumness gap by quantum relaxation methods. The quantumness gap of $(2, 1, p)$ -QRAO is always smaller than that of $(3, 1, p)$ -QRAO. (b) N_{nodes} dependences of P_{opt} for MaxCut problem with QR-BnB-VQE and naive $(2, 1, p)$ -QRAO. QR-BnB-VQE always gives better success probability than the naive QRAO. (c) N_{nodes} dependences of N_{eval} for MaxCut problem with QR-BnB-VQE. QR-BnB-VQE always provides better convergence than QR-BnB-Exact.

we linearized the problem employing the Glover-Woolsey method [34] and solved it with CBC [7].

V. RESULTS

A. MaxCut Problem

First, we investigate the performance of QR-BnB on unconstrained optimization problems using the MaxCut problem. We analyze the average number of evaluations of the quantum relaxation problems N_{eval} until converging the branch-and-bound process. Figure 2 (a)-(c) show the dependence of N_{eval} on the problem size N_{nodes} when using QR-BnB-Exact. QR-BnB-Exact gave optimal solutions for all problem instances for all methods. Note that the optimality of the solutions is

guaranteed by QR-BnB-Exact due to the branch-and-bound process. (a), (b) and (c) are the results when BFS, BrFS, and DFS are employed as the tree search strategies, respectively. The color indicates the variable selection method, and the mark indicates the quantum relaxation method. For $(2, 1, p)$ -QRAO, the dependence on the tree search strategy is small, but for $(3, 1, p)$ -QRAO, BFS converges faster than DFS. Therefore, in the following, we will focus on the case of BFS (Fig. 2(a)). Next, we consider the dependence on the variable selection method. Regardless of the quantum relaxation method, the Least Fractional rule converges the fastest. Similar results have been reported in the previous study of classical branch-and-bound algorithm [35].

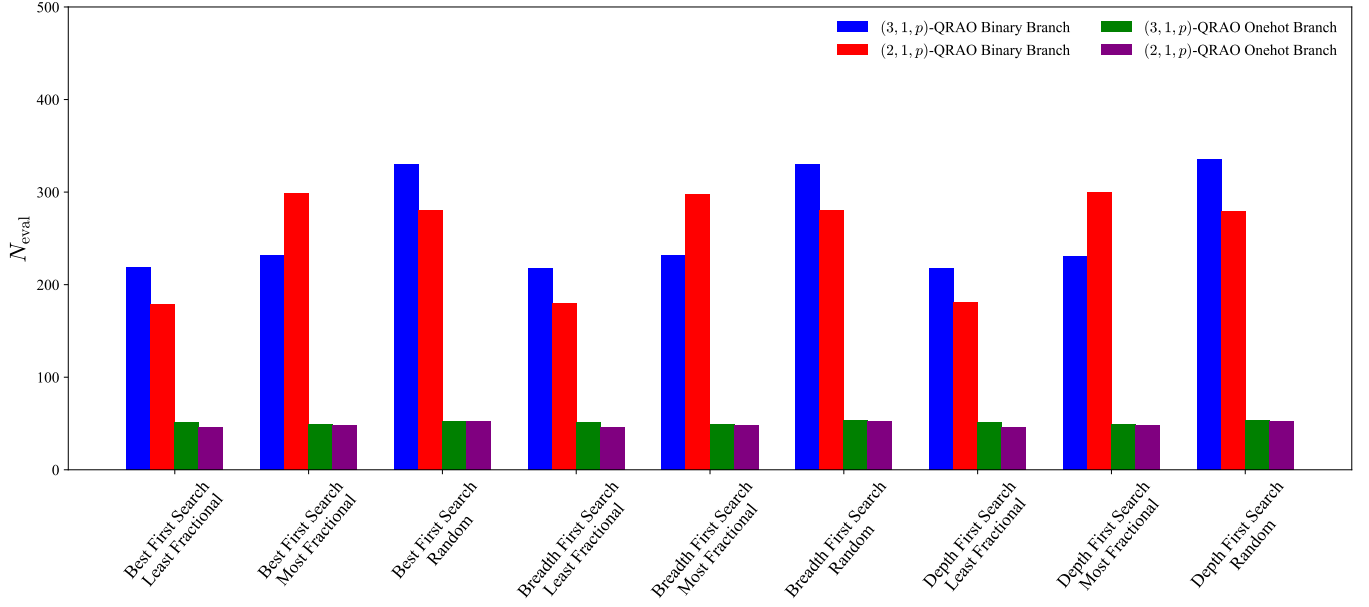


Fig. 4: Comparison of N_{eval} for various strategies in TSP. Using the Onehot Branch always provides faster convergence than the Binary Branch.

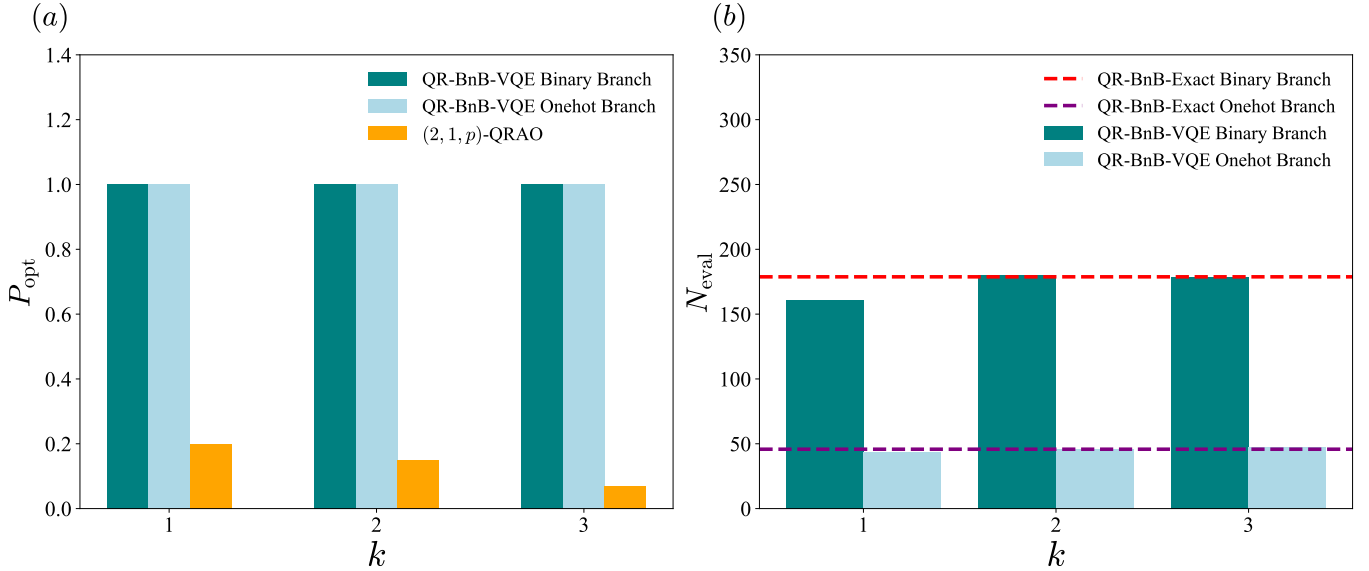


Fig. 5: (a) k dependences for P_{opt} for TSP compared to QR-BnB-VQE and naive $(2, 1, p)$ -QRAO. QR-BnB-VQEs always provide an optimal solution. (b) k dependences for N_{eval} for TSP with QR-BnB-VQE. The Onehot Branch provides faster convergence than the Binary Branch.

Next, we focus on N_{eval} dependence on the quantum relaxation method. N_{eval} for $(2, 1, p)$ -QRAO is always smaller than for $(3, 1, p)$ -QRAO, indicating that $(2, 1, p)$ -QRAO is always faster converging. In the classical branch-and-bound method, the integrality gap, the ratio of the optimal values of the original problem and its linear relaxation problem, is related to convergence speed. In particular, convergence can be faster when the integrality gap is small. Therefore,

we consider the quantumness gap, which is the ratio of the optimal value of the original optimization problem and the quantum relaxation problem, to be similar in the case of linear relaxation. As shown in Fig. 3 (a), the quantumness gap is larger for $(3, 1, p)$ -QRAO than for $(2, 1, p)$ -QRAO. As in the case of the integrality gap, the convergence speed varies with the size of the quantumness gap. In the case of both quantum relaxation methods, the quantumness gap is independent of

N_{nodes} .

Next, we investigated the behavior of QR-BnB-VQE. The optimal value of quantum relaxation $\tilde{z}_{\text{QR}}$ and the expectation value obtained by VQE $\langle \psi(\theta) | \tilde{H} | \psi(\theta) \rangle$ have the relation $\langle \psi(\theta) | \tilde{H} | \psi(\theta) \rangle \geq \tilde{z}_{\text{QR}}$. Here, θ denotes the parameters of the variational quantum circuit, and the equal sign is held when the optimal parameter θ^* is obtained. Therefore, if θ^* could not be obtained, there is a possibility that the relaxed solution of the subproblem which should be $\tilde{z}_{\text{QR}} \leq z_{\text{fes}}$ is evaluated as $\langle \psi(\theta) | \tilde{H} | \psi(\theta) \rangle \geq z_{\text{fes}}$. This implies that the subproblems that should not be pruned are pruned. Therefore, we evaluated the success probability P_{opt} of obtaining the optimal solution in QR-BnB-VQE in Fig. 3 (b). In this experiment, we used the fastest convergence setting in QR-BnB-Exact, the Least Fractional rule and BFS, and $(2, 1, p)$ -QRAO as quantum relaxation methods. Because the convergence of QR-BnB-VQE may be slowed down when the optimal solution branches were pruned by mistake, we set the upper limit of the number of evaluations at 2000 for $(2, 1, p)$ -QRAO terminated QR-BnB-VQE and evaluated the solutions obtained up to this point. In all cases, QR-BnB-VQE has a higher success probability than naive $(2, 1, p)$ -QRAO. P_{opt} for QRAO decreases with increasing N_{nodes} , while that for QR-BnB-VQE is almost N_{nodes} independent. In particular, for $k = 1, 2$, $P_{\text{opt}} \geq 0.98$ in all problem sizes. On the other hand, for $k = 3$, P_{opt} tends to decrease as N_{nodes} increases.

Figure 3 (c) shows the system size dependence of the number of evaluations required when an optimal solution was obtained using VQE. The N_{eval} in QR-BnB-VQE is always less than in QR-BnB-Exact. It is because $\langle \psi(\theta) | \tilde{H} | \psi(\theta) \rangle \geq \tilde{z}_{\text{QR}}$, so the quantum relaxation gap is estimated to be small even when the lower bound is obtained. Moreover, in QR-BnB-VQE, the dependence on the number of nodes is smaller than in QR-BnB-Exact.

B. Travelling Salesman Problem

We also solve TSP using QR-BnB as an example of a constrained optimization problem. Since TSP has onehot constraints, we can apply Onehot Branch and Binary Branch. We first investigated how these branching methods affect the performance of QR-BnB-Exact in Fig. 4. As with the MaxCut problem, optimal solutions were obtained for all problem instances in this experiment. Figure 4 compares convergence between Binary Branch and Onehot Branch. The horizontal axis indicates the combination of tree search strategies and variable selection strategies. Each bar represents N_{eval} when using Binary Branch and Onehot Branch for $(3, 1, p)$ -QRAO and $(2, 1, p)$ -QRAO, respectively. In all settings, convergence was improved when using Onehot Branch compared to Binary Branch. In particular, even in the case of BFS and Least Fractional rule, where Binary Branch was the fastest, using Onehot Branch was more than three times faster. It is attributed to the fact that more variables can be fixed in a single branching operation, reducing the size of the branches that need to be explored. In the Binary Branch, as in the MaxCut problem,

$(2, 1, p)$ -QRAO requires fewer evaluations for convergence than $(3, 1, p)$ -QRAO in most cases. However, interestingly, in the case of the Most Fractional rule, $(3, 1, p)$ -QRAO provides faster convergence. Therefore, N_{eval} does not depend on the quantumness gap but rather on the interplay between the selected internal strategies and the gap. On the other hand, in Onehot Branch, N_{eval} is almost the same, indicating that the influence of the branching method is more significant than that of the quantum relaxation method.

Finally, we present the results of solving the TSP using QR-BnB-VQE. Here, the experiment was conducted using the Least Fractional rule, BFS, and $(2, 1, p)$ -QRAO as a quantum relaxation method, which had the fastest convergence of the Binary Branch and Onehot Branch in QR-BnB-Exact. Figure 5 (a) shows the dependence of the success probability P_{opt} of QR-BnB-VQE and VQE for naive $(2, 1, p)$ -QRAO on the number of layers k . In the case of QR-BnB-VQE, we obtain the optimal solution for both Onehot and binary branches in all instances. However, in the case of naive $(2, 1, p)$ -QRAO, P_{opt} decreases with increasing k . Next, we show N_{eval} for QR-BnB-VQE in Fig. 5 (b). As with QR-BnB-Exact, the Onehot Branch provides faster convergence than the Binary Branch in QR-BnB-VQE, indicating that the Onehot Branch improves convergence speed in both cases. Moreover, N_{eval} for QR-BnB-VQE is similar to that of QR-BnB-Exact, suggesting that QR-BnB-Exact and QR-BnB-VQE exhibit similar behavior when solving TSP instances.

VI. SUMMARY

QR-BnB is an algorithm that combines quantum relaxation and branch-and-bound methods. QR-BnB provides a guaranteed optimal solution if we can obtain the exact lower bound of the subproblem through the quantum relaxation problem. The quantum device only solves the quantum relaxation problem, so we only need fewer qubits compared to QAOA. In this study, we developed QR-BnB by examining several internal strategies, particularly branching methods using constraints and variable selection methods through the expectation value of Pauli operators. Moreover, we performed detailed experiments on these internal strategies to investigate the dependence of QR-BnB convergence on internal strategies. In Sec. V, we solved the MaxCut problem and TSP using QR-BnB. Using QR-BnB-Exact, which evaluates the exact lower bounds for each subproblem, optimal solutions were obtained in all the experiments in this study. In our experiment, $(2, 1, p)$ -QRAO with a smaller quantumness gap converges faster than $(3, 1, p)$ -QRAO in most settings. We have examined several variable selection strategies, and the Least Fractional rule we developed in this study provides better convergence than the Random rule. In a comparison of branching strategies, the Onehot Branch, which uses constraint information, is more than three times faster than the Binary Branch in solving TSP. In this study, we have also examined QR-BnB-VQE, which solves the quantum relaxation problem using VQE. Due to the heuristic nature of VQE, QR-BnB-VQE cannot guarantee the optimality of the solution. However, even in QR-BnB-VQE, we have

obtained the optimal solutions in most instances. Moreover, the convergence of QR-BnB-VQE is faster than that of QR-BnB-Exact.

We only considered relatively simple internal strategies of the classical side to evaluate the performance of QR-BnB. On the other hand, for practical use, efficiency can be improved by incorporating more complicated strategies on the classical side of branch-and-bound methods and combining continuous and quantum relaxation.

One major direction for future study is to realize a method for systematically reducing the quantumness gap, similar to the cutting plane [36], [37], which can reduce the integrality gap. The classical solver is based on the Branch-and-Cut algorithm[38], which combines the cutting plane method and Branch-and-Bound. Therefore, a more efficient branch-and-cut solver can be established if a more efficient cutting plane method can be developed for quantum relaxation rather than linear relaxation. Another major direction is to realize a more suitable quantum relaxation method. In this study, we have used a Hamiltonian constructed of QRAOs as the quantum relaxation. It should be noted that a sufficient condition for quantum relaxation is to provide a lower bound on the original problem and that QRAO is not the only option for quantum relaxation. Therefore, it is expected to develop a new quantum relaxation that can encode more qubits while keeping the quantumness gap small.

ACKNOWLEDGMENT

H.M. thanks for the fruitful discussion with Kosei Teramoto and Rudy Raymond.

APPENDIX FEASIBILITY EVALUATION

In the branch-and-bound method, the constraints are evaluated at the branch process, and subproblems that do not satisfy the constraints can be pruned. In this appendix, we discuss how to evaluate constraint violation. We concentrate our discussion on linear equality constraints $\sum_i A_{ri}x_i = b_r$ or linear inequality constraints $\sum_i A_{ri}x_i \leq b_r$ based on eq. (5).

Let I be the set of indices of all variables in the problem, and let $J \subset I$ be the set of indices of fixed variables. The linear constraint $\sum_{i \in I} A_{ri}x_i = b_r$ can then be rewritten as $\sum_{i \in I \setminus J} A_{ri}x_i = \tilde{b}_r - \sum_{j \in J} A_{rj}x_j = \tilde{b}_r$. Here, $\tilde{b}_r$ is the constant part, including the constants created by fixing the variables. Thus, the linear constraint retains its form for fixed variables. Therefore, in the following, all constraints are assumed to be after the variables are fixed, and the sums are taken for $I \setminus J$. We also define the upper and lower bounds of $\sum_i A_{ri}x_i$ for each constraint as $\sup(\sum_i A_{ri}x_i) = \sum_{i|A_{ri}>0} A_{ri}$ and $\inf(\sum_i A_{ri}x_i) = \sum_{i|A_{ri}<0} A_{ri}$ respectively.

First, consider the equality constraint $\sum_i A_{ri}x_i = \tilde{b}_r$. $\tilde{b}_r$ must be between the upper and lower bounds that can be made

with unfixed variables to satisfy the constraint condition. This implies

$$\inf \left(\sum_i A_{ri}x_i \right) \leq \tilde{b}_r \leq \sup \left(\sum_i A_{ri}x_i \right). \quad (10)$$

If this condition is not satisfied, there is no feasible solution under the current fixed variables. Thus, we can prune this subproblem. This condition is a necessary condition for the continuous relaxation of the constraint. Therefore, even if this condition is satisfied, there is a possibility that a feasible solution does not exist.

Next, consider the inequality constraint $\sum_i A_{ri}x_i \leq \tilde{b}_r$. We can consider two cases. The first case is when there is no feasible solution. This case is achieved when the lower bound exceeds $\tilde{b}$. We can write it as follows:

$$\tilde{b}_r < \inf \left(\sum_i A_{ri}x_i \right). \quad (11)$$

In this case, the subproblem can be pruned. Another case is that the constraint is always satisfied no matter what choice is made in the remaining variables. This case is when the upper bound is less than $\tilde{b}$. It can be written as follows:

$$\sup \left(\sum_i A_{ri}x_i \right) \leq \tilde{b}_r. \quad (12)$$

This constraint can be ignored in this and subsequent subproblems.

REFERENCES

- [1] B. Korte and J. Vygen, *Combinatorial Optimization: Theory and Algorithms*, 5th. Springer Berlin, Heidelberg, 2012, ISBN: 3642244874.
- [2] L. Wolsey, *Integer Programming*, en. Wiley, Sep. 2020.
- [3] M. Conforti, G. Cornuéjols, and G. Zambelli, *Integer Programming* (Graduate Texts in Mathematics). Springer International Publishing, 2014, ISBN: 3319110071.
- [4] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems," *Econometrica*, vol. 28, no. 3, pp. 497–520, 1960.
- [5] D. R. Morrison, S. H. Jacobson, J. J. Sauppe, and E. C. Sewell, "Branch-and-bound algorithms: A survey of recent advances in searching, branching, and pruning," *Discrete Optim.*, vol. 19, pp. 79–102, Feb. 2016.
- [6] Gurobi Optimization, LLC, *Gurobi Optimizer Reference Manual*, 2023. [Online]. Available: <https://www.gurobi.com>.
- [7] J. Forrest, T. Ralphs, S. Vigerske, *et al.*, *Coin-or/cbc: Release releases/2.10.11*, version releases/2.10.11, Oct. 2023. DOI: 10.5281/zenodo.10041724. [Online]. Available: <https://doi.org/10.5281/zenodo.10041724>.
- [8] A. Abbas, A. Ambainis, B. Augustino, *et al.*, "Quantum optimization: Potential, challenges, and the path forward," Dec. 2023. arXiv: 2312.02279 [quant-ph].

- [9] J. Preskill, “Quantum computing in the NISQ era and beyond,” en, *Quantum*, vol. 2, no. 79, p. 79, Aug. 2018.
- [10] K. Bharti, A. Cervera-Liarta, T. H. Kyaw, *et al.*, “Noisy intermediate-scale quantum algorithms,” *Rev. Mod. Phys.*, vol. 94, no. 1, p. 015004, Feb. 2022.
- [11] E. Farhi, J. Goldstone, and S. Gutmann, “A quantum approximate optimization algorithm,” Nov. 2014. arXiv: 1411.4028 [quant-ph].
- [12] 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].
- [13] T. Kadowaki and H. Nishimori, “Quantum annealing in the transverse ising model,” *Phys. Rev. E*, vol. 58, pp. 5355–5363, 5 Nov. 1998. DOI: 10.1103/PhysRevE.58.5355. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevE.58.5355>.
- [14] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, “Quantum computation by adiabatic evolution,” Jan. 2000. arXiv: quant-ph/0001106 [quant-ph].
- [15] T. Albash and D. A. Lidar, “Adiabatic quantum computation,” *Rev. Mod. Phys.*, vol. 90, no. 1, p. 015002, Jan. 2018.
- [16] M. Chalupnik, H. Melo, Y. Alexeev, and A. Galda, “Augmenting qaoa ansatz with multiparameter problem-independent layer,” in *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)*, 2022, pp. 97–103. DOI: 10.1109/QCE53715.2022.00028.
- [17] Z. Tabi, K. H. El-Safty, Z. Kallus, *et al.*, “Quantum optimization for the graph coloring problem with space-efficient embedding,” in *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*, 2020, pp. 56–62. DOI: 10.1109/QCE49297.2020.00018.
- [18] S. Bravyi, A. Kliesch, R. Koenig, and E. Tang, “Hybrid quantum-classical algorithms for approximate graph coloring,” *Quantum*, vol. 6, p. 678, Mar. 2022, ISSN: 2521-327X. DOI: 10.22331/q-2022-03-30-678. [Online]. Available: <https://doi.org/10.22331/q-2022-03-30-678>.
- [19] B. Fuller, C. Hadfield, J. R. Glick, *et al.*, “Approximate solutions of combinatorial problems via quantum relaxations,” Nov. 2021. arXiv: 2111.03167 [quant-ph].
- [20] B. Tan, M.-A. Lemonde, S. Thanasilp, J. Tangpanitanon, and D. G. Angelakis, “Qubit-efficient encoding schemes for binary optimisation problems,” en, *Quantum*, vol. 5, no. 454, p. 454, May 2021.
- [21] M. J. Rančić, “Noisy intermediate-scale quantum computing algorithm for solving an n -vertex MaxCut problem with $\log(n)$ qubits,” *Phys. Rev. Res.*, vol. 5, no. 1, p. L012021, Feb. 2023.
- [22] M. Sciorilli, L. Borges, T. L. Patti, *et al.*, *Towards large-scale quantum optimization solvers with few qubits*, 2024. arXiv: 2401.09421 [quant-ph].
- [23] K. Teramoto, R. Raymond, E. Wakakuwa, and H. Imai, “Quantum-Relaxation based optimization algorithms: Theoretical extensions,” Feb. 2023. arXiv: 2302.09481 [quant-ph].
- [24] K. Teramoto, R. Raymond, and H. Imai, “The role of entanglement in quantum-relaxation based optimization algorithms,” in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 01, 2023, pp. 543–553. DOI: 10.1109/QCE57702.2023.00068.
- [25] 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,” en, *Algorithms*, vol. 12, no. 2, p. 34, Feb. 2019.
- [26] W.-H. Huang, H. Matsuyama, K. Nishimura, and Y. Yamashiro, “Quantum relaxation based branch-and-bound algorithm,” in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 02, 2023, pp. 282–283. DOI: 10.1109/QCE57702.2023.10243.
- [27] A. Peruzzo, J. McClean, P. Shadbolt, *et al.*, “A variational eigenvalue solver on a photonic quantum processor,” en, *Nat. Commun.*, vol. 5, p. 4213, Jul. 2014.
- [28] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, “Dense quantum coding and a lower bound for 1-way quantum automata,” in *Proceedings of the thirty-first annual ACM symposium on Theory of Computing*, ser. STOC ’99, Atlanta, Georgia, USA: Association for Computing Machinery, May 1999, pp. 376–383.
- [29] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, “Dense quantum coding and quantum finite automata,” *J. ACM*, vol. 49, no. 4, pp. 496–511, Jul. 2002.
- [30] M. Hayashi, K. Iwama, H. Nishimura, R. Raymond, and S. Yamashita, “(4,1)-quantum random access coding does not exist—one qubit is not enough to recover one of four bits,” en, *New J. Phys.*, vol. 8, no. 8, p. 129, Aug. 2006.
- [31] A. Kandala, A. Mezzacapo, K. Temme, *et al.*, “Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets,” en, *Nature*, vol. 549, no. 7671, pp. 242–246, Sep. 2017.
- [32] K. M. Nakanishi, K. Fujii, and S. Todo, “Sequential minimal optimization for quantum-classical hybrid algorithms,” *Phys. Rev. Res.*, vol. 2, no. 4, p. 043158, Oct. 2020.
- [33] Qiskit contributors, *Qiskit: An open-source framework for quantum computing*, 2023. DOI: 10.5281/zenodo.2573505.
- [34] F. Glover and E. Woolsey, “Further reduction of Zero-One polynomial programming problems to Zero-One linear programming problems,” *Oper. Res.*, vol. 21, no. 1, pp. 156–161, Feb. 1973.
- [35] F. Ortega and L. A. Wolsey, “A branch-and-cut algorithm for the single-commodity, uncappeditated, fixed-charge network flow problem,” en, *Networks*, vol. 41, no. 3, pp. 143–158, May 2003.

- [36] G. Dantzig, R. Fulkerson, and S. Johnson, "Solution of a large-scale traveling-salesman problem," *Journal of the Operations Research Society of America*, vol. 2, no. 4, pp. 393–410, 1954. DOI: 10.1287/opre.2.4.393. [Online]. Available: <https://doi.org/10.1287/opre.2.4.393>.
- [37] R. E. Gomory, "Outline of an algorithm for integer solutions to linear programs," *Bulletin of the American Mathematical Society*, vol. 64, no. 5, pp. 275–278, 1958.
- [38] M. Padberg and G. Rinaldi, "A branch-and-cut algorithm for the resolution of large-scale symmetric traveling salesman problems," *SIAM Review*, vol. 33, no. 1, pp. 60–100, 1991. DOI: 10.1137/1033004. [Online]. Available: <https://doi.org/10.1137/1033004>.