Seminar: Advanced Topics in Quantum Computing
On efficient encodings for QAOA solutions to vehicle routing problems

Eben Jowie Haezer

October 19, 2023

Abstract

This report details recent advances in the optimisation of computing resources for quantum approaches to solving the vehicle routing problem (VRP) and its variants. This set of problems is of significant importance with regard to logistical applications in industry. In accordance with the input constraints of the quantum hardware, the problem is formulated as a quadratic unconstrained binary optimisation (QUBO). A simple approach known as the full encoding results in each solution represented by a unique basis state, thereby requiring one qubit per classical variable. Due to this inefficiency in resource allocation when considering the at worst factorial search space, a more optimised minimal encoding is suggested that offers a logarithmic reduction in necessary computing power. In spite of certain drawbacks incurred by employing this minimal encoding, experiments have shown that the solution quality is not heavily impacted.

I. Introduction

The Vehicle Routing Problem (VRP) is concerned with finding optimal routes for vehicles to deliver goods to a set of customers with some geographical distance between them. It is obvious that this type of problem has far reaching practical applications in various contexts, in fact the need to find a reasonable solution is almost ubiquitous when dealing with logistical planning, for example to determine efficient road, rail, shipping, and air routes for commercial or public interest.

VRP, being itself a more general version of the Travelling Salesman Problem (TSP), is similarly an NP hard combinatorial optimisation problem with a solution space scaling factorially in the number of customers, and thus finding the definitive optimal solution already becomes intractable at two digit customer counts. The most optimal classical algorithms known to date and used in practice employ heuristics and greedy methods to construct routes within single digit percentage tolerances.

More recently, the noisy intermediate scale quantum (NISQ) era has paved the way for the development of variational quantum algorithms alongside the potent quantum annealing method as a contender for computing reliable and near optimal solutions to combinatorial optimisation problems, exploiting the unique inherent property of quantum algorithms to operate on the entire solution space at once, albeit incrementally and probabilistically.

Initial tests have shown some promise in applying these quantum algorithms to the VRP against classical solvers, achieving comparable accuracy on small problem instances. As quantum hardware continues to improve, it makes sense to also look at utilising it as efficiently as possible. Solving VRP naively using quantum methods maps one classical variable to one whole qubit, quickly exhausting the already limited computing resources when introducing additional constraints such as vehicle capacity or time windows on larger problem scales more commonly encountered in practical scenarios.

This report discusses recent research into an idea to mitigate this uneconomical use of computing power through a more clever and refined encoding of the input problem, in order to achieve a logarithmic correlation between problem size and resource consumption. Section II outlines a formal description of the VRP and some of its variants. Section III explains the QUBO and its application to solving the VRP on quantum hardware, along with a brief mention of the closely related Ising model and in particular its Hamiltonian function. Section IV details the encoding approaches onto the quantum system and lists benefits and drawbacks for each. Published experimental observations describing the impact of the encodings are discussed in Section V, and the report concludes in Section VI.

II. Vehicle Routing Problem

The vehicle routing problem (VRP) is a generalisation of the perhaps more well known NP hard Travelling Salesman Problem (TSP). VRP seeks the optimal route or routes for a number of vehicles to traverse in order to deliver certain goods to customers in various locations.

The problem is modelled intuitively by a graph $G = (V, E)$ where each node or vertex $v_i \in V$ represents the location of a customer and each edge $(i, j) \in E$ connecting two vertices corresponds to a path traversible by a delivery vehicle. $\|V\| = N$ is then the number of customers considered. Traversing an edge (i, j) typically incurs a cost represented by $c_{ij} \in \mathbb{R}_{\geq 0}$, of which the total value summed up across the entire journey is to be optimised, ie. made as small as possible. This cost may be set based on travel time, distance, or other concerns with economic consequences.

Vehicles may only travel across edges. Furthermore, each graph contains a designated node $v_0 \in V$ known as the depot. Valid routes must always begin and end at the depot. A valid route is then a tuple $(v_n, v_{n+1}, ..., v_m)$ s.t. $v_n = v_m = v_0 \wedge \forall n.\ (v_n, v_{n+1}) \in E$.

For a problem to be classified as VRP, it should fulfil at least the above minimal constraints. However, additional constraints may be imposed as needed to better reflect a practical use case at the expense of slightly complicating the model. For example, the capacitated VRP or CVRP stipulates a fixed upper bound on the carrying capacity of each vehicle leaving the depot, where in most cases this value is consistent across all vehicles. Customers v_i are then assigned a score $d_i \in \mathbb{R}_{\geq 0}$ reflecting their demand quantity. This introduces the complication of optimising for capacity and demand as well as edge cost, and the ideal solution for a given graph will most likely differ from the simple VRP.

On the other hand, the VRP with time windows (VRPTW) introduces a secondary time parameter. Customers v_i are assigned a certain time window $[t_i^0, t_i^f] \subseteq \mathbb{R}_{\geq 0}$ in which they expect a delivery. For the depot v_0 this interval is $[0, \infty)$ for simplicity. Hence the edge costs c_{ij} represent travel time between two nodes in this formulation, and the objective shifts to finding the optimal route that serves all customers whilst respecting these time windows, or failing this attempting to maximise the number of customers or total goods supplied.

III. Problem Formulation

The quadratic unconstrained binary optimisation (QUBO) is concerned with finding a binary vector $|x^*\rangle \in \{0, 1\}^n$, $n \in \mathbb{N}$ that fulfils the following optimal condition:

$$|x^*\rangle = \underset{|x\rangle \in \{0,1\}^n}{\operatorname{argmin}} \langle x|Q|x\rangle \tag{1}$$

where the linear operator $Q \in \mathbb{R}^{n \times n}$ is a symmetric matrix. This can be interpreted as an objective function:

$$f_Q(x) = \langle x|Q|x\rangle = \sum_{i=1}^{n} \sum_{j=i}^{n} Q_{ij} x_i x_j \tag{2}$$

In general, QUBO is also NP hard due to the exponential scaling of the solution space in $\|\{0, 1\}^n\| \in \mathcal{O}(2^n)$ with respect to the number of dimensions n. Many combinatorial optimisation problems have conversions into QUBO, not least the VRP and its variants. These conversions are useful to establish a uniform problem description for solvers to work with, however in the context of quantum solvers, they are further motivated by the equivalence between QUBO and the Hamiltonian of the Ising model for ferromagnetism in particle physics.

The Ising model describes a lattice structure Λ in which each lattice site houses a particle. The spin of each particle is represented by a discrete variable $\sigma_i \in \{-1, 1\}$, $i \in \Lambda$. This spin value governs the local magnetic moment of the particle according to the shell model (appendix). Neighbouring lattice sites $\langle i\ j\rangle$. $i, j \in \Lambda$ influence each other, termed nearest neighbour interactions, whose interaction strength is represented by $J_{ij} \in \mathbb{R}$. Furthermore, one may consider the influence of an external field h_i at site $i \in \Lambda$, such that the spin wants to align with this field.

Thus the Hamiltonian reads:

$$H = -\sum_{\langle i\ j\rangle} J_{ij} \sigma_i \sigma_j - \mu \sum_i h_i \sigma_i \tag{3}$$

where μ is the magnetic moment.

Replacing σ with the Pauli operators yields the quantum mechanical description:

$$H = -\sum_{\langle i\ j\rangle} J_{ij} \sigma_i^z \sigma_j^z - \mu \sum_i h_i^z \sigma_i^z - \mu \sum_i h_i^x \sigma_i^x \tag{4}$$

where the second term with σ^z describes the external longitudinal field, and the final term with σ^x describes the transverse field per lattice site.

Notably, the Ising model is typically simplified to exclude the transverse field, resulting in a classical Hamiltonian where the constituent terms commute ie. a diagonal operator in the Z basis. This means the ground state is described by a basis vector $|\phi\rangle$ with $\phi \in \{0,1\}^n$, allowing for simple measurement to obtain a bit string solution.

Through the reversible transformation $\sigma \mapsto 2x-1$, where $x \in \{0,1\}$ s.t. $-1 \mapsto 0$ and $1 \mapsto 1$, one obtains the equivalent QUBO formulation for free, and hence optimising for the ground state of the Ising Hamiltonian is equivalent to optimising the QUBO objective function.

With this knowledge, one can formulate the VRP as follows.

IV. Quantum Algorithms

Traditionally, quantum solvers for combinatorial optimisation problems have been implemented using the quantum annealing approach. This method relies on the theory of adiabatic quantum computation to evolve a prepared system state gradually towards the problem Hamiltonian. With sufficiently slow evolution, the system is guaranteed to arrive in the ground state by the adiabatic theorem.

More precisely, the system evolves according to

$$H(t) = (1 - \frac{t}{T}) \cdot H_0 + \frac{t}{T} \cdot H_p \tag{5}$$

with $0 \le t \le T$, where T is bound by $T \in \mathcal{O}(\frac{1}{g^2})$. Here g denotes the spectral gap of H, the minimal energy difference between the ground state and first excited state throughout the evolution. Consequently, Hamiltonians with narrower spectral gaps have to be evolved more slowly for the sake of precision.

On the other hand, the more common circuit model processor operates based on a series of quantum gates connected in a circuit in order to manipulate the quantum state to a desired outcome. This principle of operation is founded upon the discretised solution to the Schrödinger equation for negligibly small time intervals:

$$i\hbar \frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi(t)\rangle \Longleftrightarrow U(t) = e^{\frac{-i}{\hbar}Ht} \tag{6}$$

where U denotes the unitary time evolution operator. This unitary can be decomposed into several building blocks termed quantum gates, which are implemented in the hardware and operate on a small subset of qubits. The quantum circuit thus realises the Hamiltonian.

The advances in circuit model quantum processors that have brought about the NISQ era have offered an alternative approach to tackle optimisation problems, namely that of supplementing a quantum algorithm with classically tuned parameters to iteratively arrive at a good approximation of the ground state, known as variational algorithms. NISQ era processors are characterised by a primary limitation of restricted circuit depth due to their sensitivity to noise as well as the inability to correct the faults introduced. By offloading some of the computation to a classical algorithm and proceeding iteratively, this limitation can be worked around to a degree that allows one to obtain reasonable estimations of the ideal solutions to optimisation problems.

V. Encoding Approaches

We assume the use of a conventional NISQ processor to tackle a VRP involving n_c classical binary variables. A simple mapping from the VRP to QUBO uses what is known as the complete encoding, where each classical binary variable is represented by a single qubit. The quantum state takes the form:

$$|\psi(\theta)\rangle = U(\theta)|\psi_0\rangle \tag{7}$$

$$= \sum_{x \in \{0,\,1\}^{n_c}} \alpha_x |x\rangle \tag{8}$$

In this case, sampling the quantum state yields any given bitstring solution $x \in \{0,\,1\}^{n_c}$ with probability $\|\alpha_x\|^2$, where α is parametrised by a set of angles θ. In variational algorithms, θ are heuristically determined values that serve to tune arrays of rotation unitaries, which are adjusted after each pass to attain the optimised state $|\psi^*\rangle = U(\theta^*)|\psi_0\rangle$.

The QUBO objective function in the complete encoding is described by the expectation of the equivalent Ising Hamiltonian:

$$C_{cpl}(\theta) = \langle H \rangle = \langle \psi(\theta)|H|\psi(\theta)\rangle \tag{9}$$

The advantage of the complete encoding is its ability to capture all possible correlations between classical variables. This however sacrifices scalability, requiring computational resources that scale proportionally to the problem size as each binary classical variable is mapped to a single qubit, hence $n_q = n_c$. As a result, feasible applications of the complete encoding are limited to toy examples. Even as quantum hardware becomes more powerful, it is prudent to examine encoding schemes that can more efficiently take advantage of the available technology.

We consider the minimal encoding as presented in [1]. This scheme allows for n_c classical variables to be mapped to $n_q = 1 + \log_2 n_c$ qubits, using a $n_r = \log_2 n_c$ wide register and $n_a = 1$ auxiliary qubit.

Let $\{|\phi_i\rangle\}$ and $\{|0\rangle, |1\rangle\}$ denote a basis for the register and auxiliary qubits respectively. The system state $|\psi\rangle$ is then defined as:

$$|\psi(\theta)\rangle = \sum_{i=1}^{n_c} \beta_i (a_i|0\rangle + b_i|1\rangle)\, |\phi_i\rangle \tag{10}$$

where the coefficients β_i, a_i, b_i are parametrised by θ, as in the complete encoding.

This encodes in the register the probability β_i of measuring each register state $|\phi_i\rangle$ that corresponds to an individual classical variable x_i, and in the auxiliary the probability of each of these binary variables taking on the value 0 or 1 according to the Born rule: Pr $(x_i = 0) = \|a_i\|^2$ and Pr $(x_i = 1) = \|b_i\|^2$. Taken together, the probability of measuring a certain bitstring solution x for a given θ is expressed by:

$$\Pr x = \prod_{i=1}^{n_c} \Pr x_i = \prod_{i=1}^{n_c} \|b_i\|^2 \tag{11}$$

As an example, consider $n_c = 4$, $n_r = 2$. The normalised system state representing a particular bitstring $x^* = (1,1,1,0)^\dagger$ would be $|\psi^*\rangle = \frac{1}{2}(|1\rangle|00\rangle + |1\rangle|01\rangle + |1\rangle|10\rangle + |0\rangle|11\rangle)$.

The variational algorithm is run for several trials to obtain an estimate of the coefficients a_i and b_i, from which the probability distribution for bitstring solutions as given in (11) may be constructed.

As stated previously, with the minimal encoding a logarithmic scaling of qubits to classical variables is achieved. However, this comes at the expense of a more limited expressiveness in the quantum state. As seen in (11), the probability distribution is sufficient to encode only statistically independent classical variables, ie. those s.t. $\forall i\ j$. Pr $(x_i \wedge x_j) = \Pr x_i \Pr x_j$, which tend not to be the case for highly complex and intertwined problems as in the VRP.

Furthermore, the number of trials required to obtain a solution of sufficient quality is necessarily increased. Depending on the exact problem configuration, particularly small coefficient values in the quantum state may result in certain bitstring outcomes being sampled too infrequently if at all while evaluating the objective function, leading to inaccuracies.

The objective function in the minimal encoding has the form:

$$C_{min}(\theta) = \sum_{i\neq j}^{n_c} Q_{ij} \|b_i\|^2 \|b_j\|^2 + \sum_{i=1}^{n_c} Q_{ii} \|b_i\|^2 \tag{12}$$

$$= \sum_{i\neq j}^{n_c} Q_{ij} \frac{\langle P_i^1\rangle}{\langle P_i\rangle} \frac{\langle P_j^1\rangle}{\langle P_j\rangle} + \sum_{i=1}^{n_c} Q_{ii} \frac{\langle P_i^1\rangle}{P_i} \tag{13}$$

expressed as projectors $P_i = |\phi_i\rangle\langle\phi_i|$ and $P_i^1 = |1\rangle\langle 1| \otimes P_i$.

VI. Discussion

VII. Conclusion

References

[1] B. Tan, M.-A. Lemonde, S. Thanasilp, J. Tangpanitanon, and D. G. Angelakis, "Qubit-efficient encoding schemes for binary optimisation problems," *Quantum*, 2021.

VIII. Appendix