To what extent does circuit cutting improve the results of QAOA

Background

QAOA enables solving combinatorial optimization problems like Max Cut and maximum independent set problems. The goal is to find the bitstring $z = (z_1, z_2, \dots, z_n) \in \{0,1\}^n$ that maximizes an objective function $C(z)$

$$H_c|z\rangle = C(z)|z\rangle$$

since the cost function is encoded in the diagonal of H_c

Thus the optimization problem can be solved by finding the eigenvector $|z\rangle$ of H_c with maximal eigenvalue.

$$\max_{|z\rangle} \langle z|H_c|z\rangle = C_{max}$$

QAOA employs an ansatz that alternately applies p times the cost operator $U(H_c, \gamma_l) = e^{-i\gamma_l H_c}$ and mixing operator $U(H_B, \beta_l) = e^{-i\beta_l H_B}$ to the initial state of $|+\rangle^{\otimes n}$.

$$H_B = \sum_i X_i$$

$$|\psi(\beta,\gamma)\rangle = \left(\prod_{l=1}^{p} U(H_B, \beta_l)\, U(H_c, \gamma_l)\right)|+\rangle^{\otimes n}$$

β, γ are variational parameters, they are optimized.

$$\langle H_c\rangle_{\beta,\gamma} = \langle \psi(\beta,\gamma)|H_c|\psi(\beta,\gamma)\rangle$$

Solving max cut on QAOA

for an n-node undirected graph $G=(V,E)$ with edge weights $w_{ij}>0$. The cost function for max cut becomes:

$$\tilde{C}(x) = \sum_{i,j} w_{ij}\, x_i(1-x_j)$$

$$x_i \rightarrow \frac{(1-Z_i)}{2}$$ where Z_i is the Pauli operator.

$$C(z) = \sum_{i,j} \frac{w_{ij}}{4}(1-Z_i)(1+Z_j) + \sum_i \frac{w_i}{2}(1-Z_i)$$

$$H \Rightarrow \sum_i w_i Z_i + \sum_{i<j} w_{ij} Z_i Z_j$$

For an unweighted max cut, we get

$$H_{C_{\text{max cut}}} = \frac{1}{2}\sum_{i,j\in E}(1 - Z_i Z_j)$$

where Z_i is the Pauli matrix applied to i^{th} qubit.

The corresponding unitary becomes

$$U(H_{C_{\text{max cut}}}, \gamma_l) = e^{-i\gamma_l |E|/2)} \prod e^{i\gamma_l Z_i Z_j/2}$$

which can be implented as a product of $R_{ZZ}(-\gamma_l) = e^{i\gamma_l Z_i Z_j/2}$ upto a global factor $e^{i\gamma_l |E|/2}$

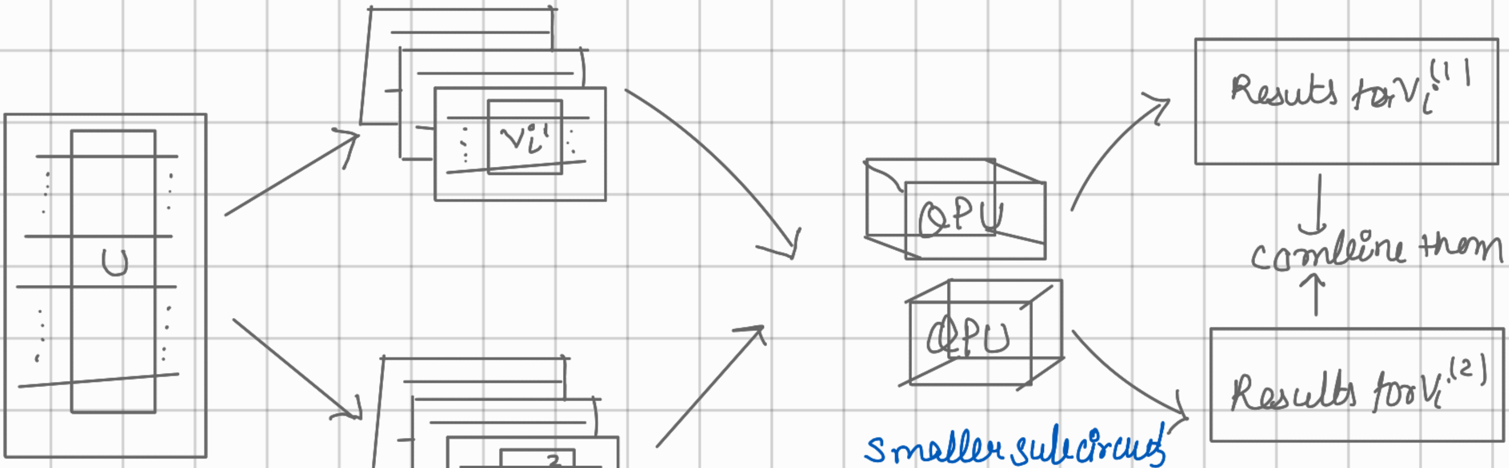
for mixer unitary we have

$$U(H_B, \beta_l) = \prod_{i\in E} e^{-i\beta_l X_i}$$ which can be implemented using R_x gate.

Quantum circuit cutting with gate cuts

The primary focus of quantum circuit cutting is on reducing the number of required qubits, but the subcircuits also consist of fewer gates such that circuit cutting may also reduce the depth of the executed circuits.

- cutting the circuit into smaller subcircuits
- execution of these subcircuits
- classical recombination of these subcircuits



A large quantum circuit gets split into smaller subcircuits. Smaller subcircuits get executed on one or more quantum devices. Results we got are then recombined.

Let $\mathcal{H}^{(1)} \otimes \mathcal{H}^{(2)}$ denote a bipartite n-qubit quantum system. The quantum state of this system can be described by a density operator ρ and there's an n-qubit gate U.

The evolution of the quantum state (ρ) under the action of gate (U) can be described by a superoperator $S(U)$, which is a linear operator that acts on the space of density operator.

$$S(U)\rho = U\rho U^\dagger$$

A gate cut of the unitary U is its quasiprobability decomposition into a set of operators $V_i = (V_i^{(1)} \otimes V_i^{(2)})$ with associated complex factors c_i given by $\{(V_i, c_i)\}$ such that

$$S(U) = \sum_i c_i S(V_i) = \sum_i c_i S(V_i^{(1)}) \otimes S(V_i^{(2)})$$

where $V_i^{(1)}$ and $V_i^{(2)}$ are operators acting on subsystem $\mathcal{H}^{(1)}$ and $\mathcal{H}^{(2)}$ respectively.

> Each operator $V_i^{(j)}$ has to physically realizable, i.e, its superoperator $S(V_i^{(j)})$ is a completely positive linear map

$$0 \leq tr[S(V_i^{(j)})\rho] \leq 1$$

hence its also include non unitary transformations such as projections.

Consider for the partition of quantum system $\mathcal{H}^{(1)} \otimes \mathcal{H}^{(2)}$ an observable $O = O^{(1)} \otimes O^{(2)}$ that acts independently on each subsystem, and a separable initial state $\rho_0 = \rho_0^{(1)} \otimes \rho_0^{(2)}$.
Thus, the subsystems are independent of each other concerning their initialization and measurement.
Then the evolution of density operator ρ_0 according to the unitary U and subsequent measurement with observable O can now be reproduced by applying the operations V_i and weighing their measurement with O according to c_i

$$tr[O S(U)\rho_0] = \sum c_i\, tr[O S(V_i)\rho_0]$$

$$\Rightarrow \sum_i c_i \underbrace{tr\left[O^{(1)} S(V_i^{(1)})\rho_0^{(1)}\right]}_{:= R_i^{(1)}} \; \underbrace{tr\left[O^{(2)} S(V_i^{(2)})\rho_0^{(2)}\right]}_{:= R_i^{(2)}}$$

This allows us to evaluate each of the expectation values $R_i^{(1)}$ and $R_i^{(2)}$ individually and then recombine them to produce the expectation value of the original circuit

$$\sum_i c_i R_i^{(1)} R_i^{(2)}$$

The computational overhead of this procedure can be measured by the number of additional shots needed to approximate the expectation value of original circuit.

Although the expectation value of the original circuit computed from the subcircuits results, remains unchanged, additional shots are necessary due to increased variance resulting from sampling the subcircuits.

$$K = \sum_i |c_i|$$

Gate cutting for QAOA maxcut ansatz

Since gate cutting works only on multi qubit gates, and the only multi qubit gate in QAOA are the Rzz gates. Thus the ansatz for a given graph can be cut by partitioning all Rzz gates between these qubit partitions. To cut a single Rzz gate, we use the QPD for cutting arbitrary two qubit gate.

$$S(R_{zz}(\gamma)) = \cos^2\left(\frac{\gamma}{2}\right) S(I \otimes I) + \sin^2\left(\frac{\gamma}{2}\right) S(Z \otimes Z) +$$

$$\cos\left(\frac{\gamma}{2}\right) \sin\left(\frac{\gamma}{2}\right) (A \otimes B + B \otimes A)$$

where we know

$$A_i := S\left(R_z\left(-\frac{\pi}{\ldots}\right)\right) - S\left(R_z\left(\frac{\pi}{\ldots}\right)\right)$$

$$B := S\left(\frac{I - Z}{2}\right) - S\left(\frac{I+Z}{2}\right)$$

the two operations forming B are the projections on the state $|1\rangle$ and $|0\rangle$ respectively.

$$\frac{I - Z}{2} = |1\rangle\langle 1| \qquad \frac{I+Z}{2} = |0\rangle\langle 0|$$

Although these operations are not unitary and thus cannot be directly implemented as a gate of a quantum circuit, they can be realized by a post-selective measurement.

So, we replace a two qubit gate R_{ZZ} by five operations $I, Z, R_Z\left(-\frac{\pi}{2}\right), R_Z\left(\frac{\pi}{2}\right)$ and a measurement for projection

Since the qubits are now separable, ten different subcircuits consisting of the five operations for the upper, and five for the lower must be executed.

more details on eq. (ii)

The operation of $R_{ZZ}(\gamma)$ can be expressed as

$$R_{ZZ}(\gamma) = e^{-i\gamma\frac{Z\otimes Z}{2}} \Rightarrow \cos\left(\frac{\gamma}{2}\right) I\otimes I - i\sin\left(\frac{\gamma}{2}\right) Z\otimes Z$$

Apply $R_{ZZ}(\gamma)$ to an arbitrary state $|\varphi, \varphi_0\rangle$ with $\rho = |\varphi, \varphi_0\rangle\langle \varphi, \varphi_0|$

$$R_{ZZ}(\gamma)\,\gamma\, R_{ZZ}(\gamma)^{\dagger} \Rightarrow \cos^2\left(\frac{\gamma}{2}\right) I^{\otimes 2}\rho\, I^{\otimes 2} + \sin^2\left(\frac{\gamma}{2}\right) Z^{\otimes 2}\rho\, Z^{\otimes 2} +$$

$$i\cos\left(\frac{\gamma}{2}\right)\sin\left(\frac{\gamma}{2}\right)\left(I^{\otimes 2}\rho Z^{\otimes 2} - Z^{\otimes 2}\rho I^{\otimes 2}\right)$$

the operation $\left(I^{\otimes 2}\rho Z^{\otimes 2} - Z^{\otimes 2}\rho I^{\otimes 2}\right)$ can be implemented on a NISQ device as the following.

Let $\rho_0 = |\varphi_0\rangle\langle\varphi_0|$ and $\rho_1 = |\varphi_1\rangle\langle\varphi_1|$. Thus $\rho = \rho_0 \otimes \rho_1$.

Let $P_0 = \frac{I+Z}{2}$ and $P_1 = \frac{I-Z}{2}$, be the projection on $|0\rangle$ and $|1\rangle$

$$A_i := R_Z\left(-\frac{\pi}{2}\right)\rho_i \cdot R_Z\left(-\frac{\pi}{2}\right)^{\dagger} - R_Z\left(\frac{\pi}{2}\right)\rho_i \cdot R_Z\left(\frac{\pi}{2}\right)^{\dagger}$$

$$B_i = P_1 \rho_i P_1^{\dagger} - P_0 \rho_i P_0^{\dagger}$$

Putting it together:

$$A_0 \otimes B_1 + B_0 \otimes A_1 = i\left(I^{\otimes 2}\rho Z^{\otimes 2} - Z^{\otimes 2}\rho I^{\otimes 2}\right)$$

This results.

$$R_{ZZ}(\gamma)\rho R_{ZZ}(\gamma)^{\dagger} = \cos^2\left(\frac{\gamma}{2}\right) I^{\otimes 2}\rho I^{\otimes 2} + \sin^2\left(\frac{\gamma}{2}\right) Z^{\otimes 2}\rho Z^{\otimes 2} + \cos\left(\frac{\gamma}{2}\right)\sin\left(\frac{\gamma}{2}\right)\left(A_0 \otimes B_1 + B_0 \otimes A_1\right)$$

Note:- Noise free barren plateaus allows the global minima to reside inside a deep narrow valley, while Noise induced Barren plateaus exponentially flatten the entire landscape.

Main Research Question

To what extent can circuit cutting improve the results of QAOA when executing on NISQ devices?

How does cutting the QAOA ansatz influence its corresponding objective function on NISQ devices?

Can QAOA with circuit cutting obtain better solution on NISQ devices when the entire algorithm, including the parameter optimization is executed?

Research Design

Max cut problem with unweighted graph.

$$S(R_{zz}(\gamma)) = \cos^2\left(\frac{\gamma}{2}\right) S(I \otimes I) + \sin^2\left(\frac{\gamma}{2}\right) S(Z \otimes Z) +$$

$$\cos\left(\frac{\gamma}{2}\right) \sin\left(\frac{\gamma}{2}\right) (A \otimes B + B \otimes A)$$

where we know

$$A := S\left(R_z\left(-\frac{\pi}{2}\right)\right) - S\left(R_z\left(\frac{\pi}{2}\right)\right)$$

$$B := S\left(\frac{I - Z}{2}\right) - S\left(\frac{I+Z}{2}\right)$$

this was the decomposition of R_{zz} gate, we make little adjustments to that

we make use of

$$R_z(\gamma) = e^{\frac{-i\gamma Z}{2}} \Rightarrow \cos\left(\frac{\gamma}{2}\right) I - i \sin\left(\frac{\gamma}{2}\right) Z$$

the corresponding superoperator is

$$S(R_z(\gamma))\rho = \cos^2\left(\frac{\gamma}{2}\right) S(I)\rho + \sin^2\left(\frac{\gamma}{2}\right) S(Z)\rho \quad +$$

$$i\cos\left(\frac{\gamma}{2}\right)\sin\left(\frac{\gamma}{2}\right)(I\rho Z - Z\rho I)$$

For angle $-\gamma$ it holds:

$$S(R_Z(-\gamma))\rho = \cos^2\left(\frac{\gamma}{2}\right)S(I)\rho + \sin^2\left(\frac{\gamma}{2}\right)S(Z)\rho -$$
$$i\cos\left(\frac{\gamma}{2}\right)\sin\left(\frac{\gamma}{2}\right)(I\rho Z - Z\rho I)$$

Together

$$S(R_Z(-\gamma))\rho - S(R_Z(\gamma))\rho = -2i\cos\left(\frac{\gamma}{2}\right)\sin\left(\frac{\gamma}{2}\right)(I\rho Z - Z\rho I)$$

it follows

$$S(R_Z(-\gamma)) - S(R_Z(\gamma)) = 2\cos^2\left(\frac{\gamma}{2}\right)S(I) + 2\sin^2\left(\frac{\gamma}{2}\right)S(Z) -$$
$$2S(R_Z(\gamma))$$

so we get

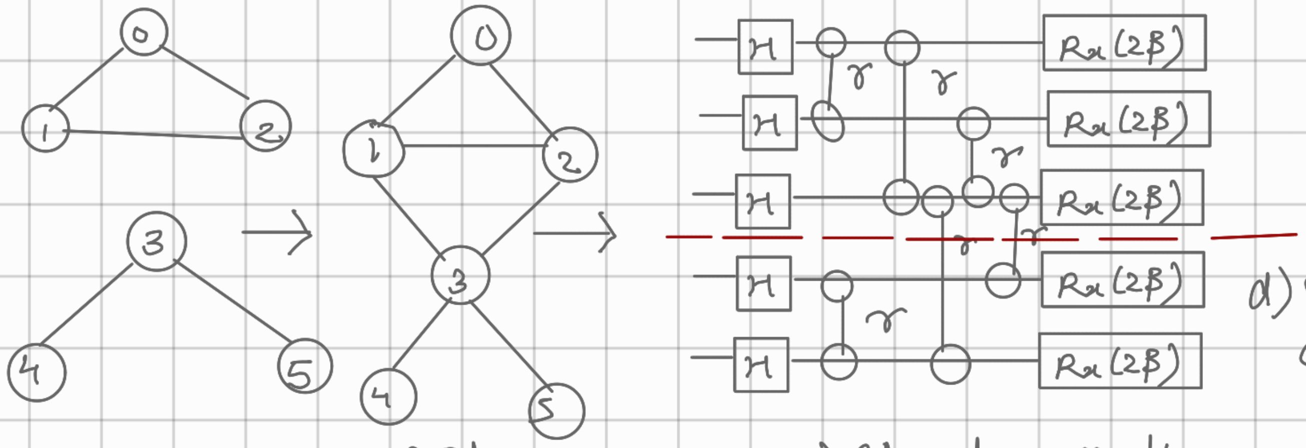
$$A = S(I) + S(Z) - 2S\left(R_Z\left(\frac{\pi}{2}\right)\right)$$

using this further decreases the number of subcircuits to eight from 10. Since two local $R_Z\left(-\frac{\pi}{2}\right)$ gates disappear.

The results for the I and Z operations can be reused as they are stored in classical memory.

Although this substitution reduces the number of subcircuits, it introduces large factors c_i in the QPD. It leads to higher variance, and therefore, more shots are needed for convergence. However this can be a worthwhile trade-off for a small number of cuts, as the factors for them remain relatively small.

EXPERIMENTS

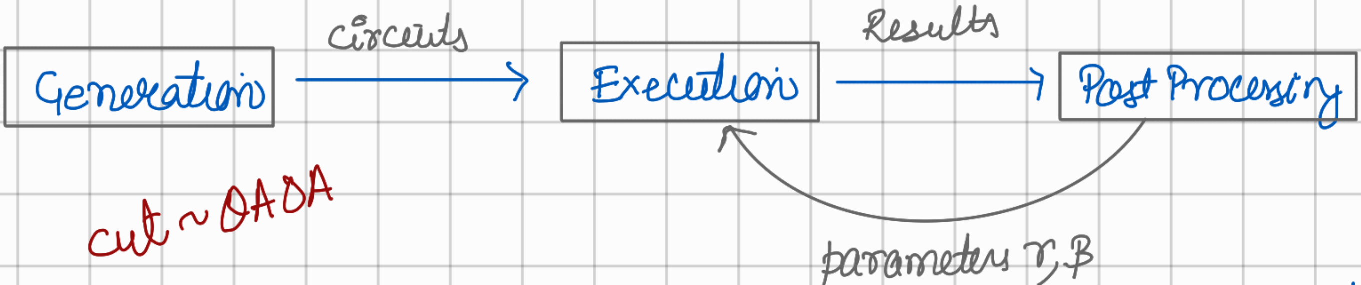


a) Subgraph generation

b) Edge insertion

c) Circuit generation

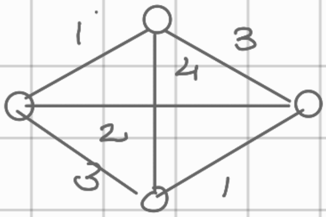
d) Circuit cutting



cut ~ QAOA

these are repeated in a variational optimization loop to refine the circuit parameters.

Max cut as QUBO



Weight matrix

$$w = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 3 & 4 \\ 2 & 3 & 0 & 1 \\ 3 & 4 & 1 & 0 \end{pmatrix}$$

cost function

$$c(x) = \sum_{i,j=1}^{n} w_{ij} x_i (1 - x_j).$$

QUBO

$$c_i = \sum_{j=1}^{n} w_{ij} \qquad Q_{ij} = -w_{ij}$$

cost function

$$C(x) = \sum_{i,j=1}^{n} x_i \theta_{ij} x_j + \sum_{i=1}^{n} c_i x_i$$

$$x^T \theta x + c^T x$$

from QUBO to Hamiltonian

$$H_c |x\rangle = C(x) |x\rangle$$

$$\sum x_i \theta_{ij} x_j + \sum c_i x_i \longrightarrow H_c = \sum \frac{1}{4} \theta_{ij} Z_i Z_j - \sum \frac{1}{2}\left(c_i + \sum \theta_{ij}\right) Z_i + \left(\frac{\sum \theta_{ij}}{4} + \frac{\sum c_i}{2}\right)$$

qubo cost function $\longrightarrow$ Hamiltonian operator