

QAOA: The Good, the Bad, and the Math

Boris Tselikhovskiy
University of California, Riverside

Joint work with

Ilya Safro (University of Delaware), Yuri Alexeev (NVIDIA)
and Bojko Bakalov (North Carolina State University)

References to our preprints:

<https://arxiv.org/abs/2509.10424>

<https://arxiv.org/abs/2309.13787>

<https://arxiv.org/abs/2405.07211>

Combinatorial Optimization Problems

Setup

Let X be a finite search space (e.g., all labelings or configurations of a system), and let

$$F : X \rightarrow \mathbb{R}$$

be an objective (cost) function.

Goal: find one or more elements $x^* \in X$ such that

$$F(x^*) = \min_{x \in X} F(x).$$

Examples

- Traveling Salesman Problem (TSP)
- Max-Cut and Graph Coloring
- Boolean Satisfiability Problems (e.g., m -SAT)

The Symmetries

Let \mathcal{S} denote the group of all permutations of the elements of X . If a permutation $g \in \mathcal{S}$ is '*undetectable*' by F , i.e.

$$F(g(x)) = F(x)$$

for any $x \in X$, then g is called a **symmetry** of F . Such symmetries form a subgroup $G \subseteq \mathcal{S}$.

From Classical to Quantum Setup

Let X be a finite set with $|X|$ elements, and let W be a vector space of dimension $|X|$ with *standard basis* $\{|x\rangle\}_{x \in X}$.

The Hamiltonian H_P represents a function $F : X \rightarrow \mathbb{R}$ if

$$H_P |x\rangle = F(x) |x\rangle \quad \forall x \in X.$$

Classical \rightsquigarrow Quantum Dictionary:

- $X \rightsquigarrow W$ (Hilbert space);
- $F \rightsquigarrow H_P$ (problem Hamiltonian);
- Minima of F on $X \rightsquigarrow$ ground states of H_P .

Remark

The action of G on X extends to an action on W .

Actions of G and H_P on W commute:

$$H_P(g(w)) = g(H_P(w)) \quad \forall w \in W, \quad \forall g \in G.$$

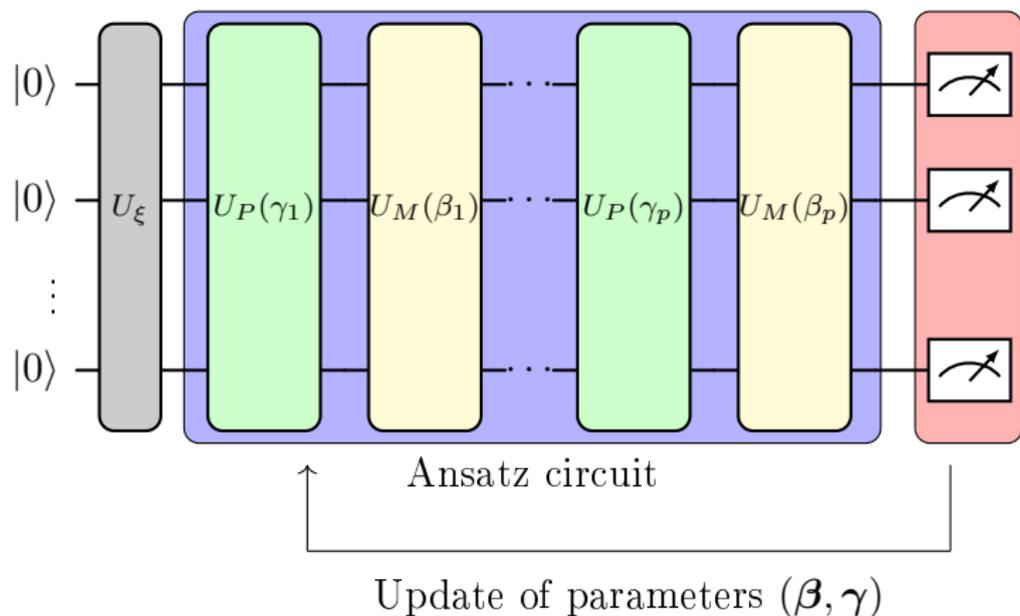
QAOA: Quantum Approximate Optimization Algorithm

- A central component of QAOA is the *mixer Hamiltonian* H_M .
- The negative of this operator, $-H_M$, should have a **unique ground state** $|\xi\rangle \in W$.
- **Circuit construction:**
 - Prepare initial state: $U_\xi |0\rangle^{\otimes n} = |\xi\rangle$;
 - Apply p alternating layers:

$$U_P(\gamma_j) := e^{-i\gamma_j H_P}, \quad U_M(\beta_j) := e^{-i\beta_j H_M};$$

- Measure in the *computational basis* to obtain outcomes $x \in X$.
- For each outcome x record the objective function value, $F(x)$;
- Using these values, compute the *empirical mean*
 $\langle H_P \rangle_{\beta, \gamma} = \langle \psi(\beta, \gamma) | H_P | \psi(\beta, \gamma) \rangle$;
- Use *classical optimization* to update parameters (β, γ) and minimize the empirical mean.

QAOA: Quantum Approximate Optimization Algorithm



Dynamical Lie Algebra and Symmetries

Let W be the Hilbert space of the system, H_P the problem Hamiltonian and H_M the mixer Hamiltonian. The *dynamical Lie algebra* \mathfrak{g} , is the real Lie algebra, generated by iH_P and iH_M :

$$\mathfrak{g} = \langle iH_M, iH_P \rangle_{\text{Lie}}.$$

Let G be the full group of symmetries of the objective function F . Then

- G commutes with H_P (the problem Hamiltonian)
- G may *not* commute with H_M (the mixer Hamiltonian), hence, not commute with \mathfrak{g} .

The remainder of this talk will be mostly devoted to the actions of \mathfrak{g} and (subgroups of) G on W , as well as interaction thereof:

$$\mathfrak{g} \curvearrowright W \curvearrowleft G.$$

Symmetry, Dynamics, and Mixer Design in QAOA

The **problem Hamiltonian H_P** encodes the classical objective function and is therefore **uniquely determined** by the optimization problem itself. However, there exists a substantial degree of **freedom in the choice of the mixer Hamiltonian H_M** , leading to distinct **dynamical Lie algebras \mathfrak{g}** .

This viewpoint suggests that \mathfrak{g} , being *parameterized* by the choice of mixer Hamiltonian (since H_P is fixed), naturally serves as a tool for analyzing and comparing mixer optimality.

Practical Relevance

The **dimension of \mathfrak{g}** correlates with:

- Convergence behavior of QAOA, and
- Adaptive schemes such as *ADAPT-QAOA* and *QAOA-GPT*, where mixer operators are chosen iteratively from an operator pool.

Many combinatorial optimization problems are defined over

$$X = \mathbb{D}^n = \{0, 1, \dots, d-1\}^n,$$

the set of n -tuples with entries in the local alphabet

$\mathbb{D} = \{0, 1, \dots, d-1\}$. In many cases, the objective function $F : X \rightarrow \mathbb{R}$ is invariant under the action of the symmetric group $S_d = W(U_d)$, which acts collectively on the Hilbert space of n qudits ($W = (\mathbb{C}^{\otimes d})^{\otimes n}$) by permuting their local labels:

$$\sigma(x_1, x_2, \dots, x_n) := (\sigma(x_1), \sigma(x_2), \dots, \sigma(x_n)).$$

A Symmetry-Preserving Mixer Hamiltonian

In the case where X consists of **binary strings** of length n , the most common choice of mixer Hamiltonian is

$$B = \sum_{j=0}^{n-1} X_j,$$

where each X_j acts as a Pauli- X operator on the j th qubit. The ground state of this mixer is the **uniform superposition**

$$|\xi\rangle = |+\rangle^{\otimes n} = |++\dots+\rangle.$$

Henceforth, we will assume that the alphabet size satisfies $d = 2^\ell$, so that the local register associated with each site can be viewed as consisting of ℓ binary variables.

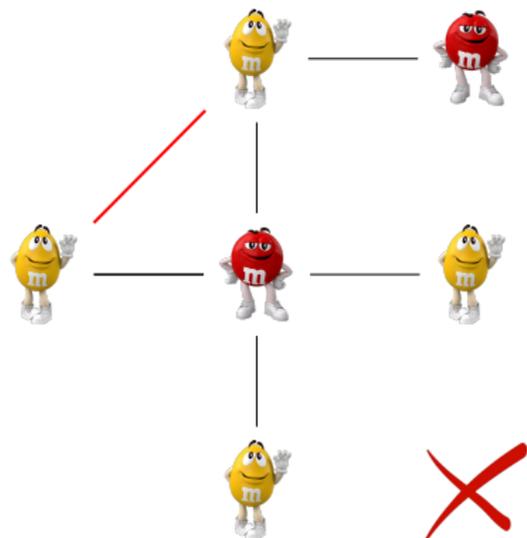
The **standard mixer Hamiltonian** does not commute with the action full symmetric group S_d , but only with a noticeably smaller subgroup. This motivates the introduction of an **alternative mixer Hamiltonian \mathbf{H}_M** , which preserves **commutativity with the entire group action**.

A Concrete Application

Consider the problem of coloring the vertices (edges) of a graph. A coloring is considered *proper* if no adjacent vertices (edges sharing a vertex) have the same color.

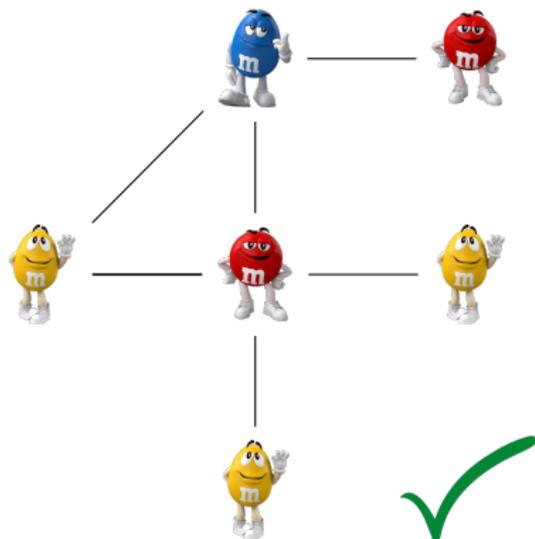
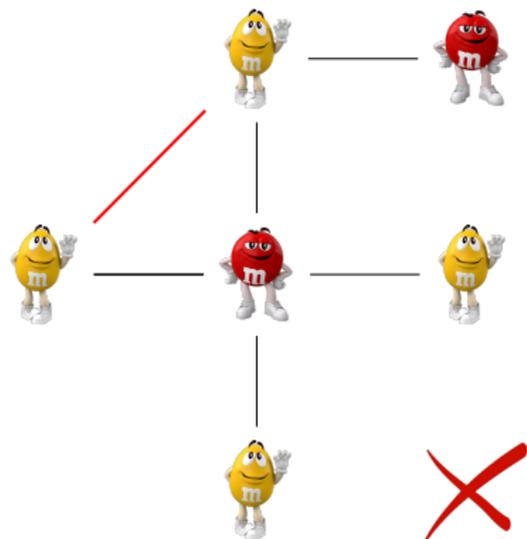
A Concrete Application

Consider the problem of coloring the vertices (edges) of a graph. A coloring is considered *proper* if no adjacent vertices (edges sharing a vertex) have the same color.



A Concrete Application

Consider the problem of coloring the vertices (edges) of a graph. A coloring is considered *proper* if no adjacent vertices (edges sharing a vertex) have the same color.



A Concrete Application

We will focus on the edge coloring problem. To each edge $e \in E$, one associates ℓ bits $e_0, e_1, \dots, e_{\ell-1}$, the values of which uniquely determine its color.

The function χ_c is defined as follows:

$$\chi_c(c') := \begin{cases} 1, & \text{if } c'_i \equiv c_i \text{ for all } i \in \{1, \dots, \ell\} \\ 0, & \text{otherwise} \end{cases}$$

This function serves as the characteristic function of a color: it has value 1 on color c and 0 on all other colors.

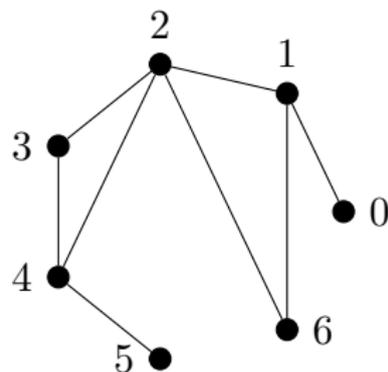
The objective function $F_\Gamma(C) := \sum_{e \bullet f} \sum_{c \in \mathcal{C}} \chi_c(C(e)) \chi_c(C(f))$ computes the number of adjacent edges of coinciding color.

Remark

A coloring C is proper if and only if $F_\Gamma(C) = 0$.

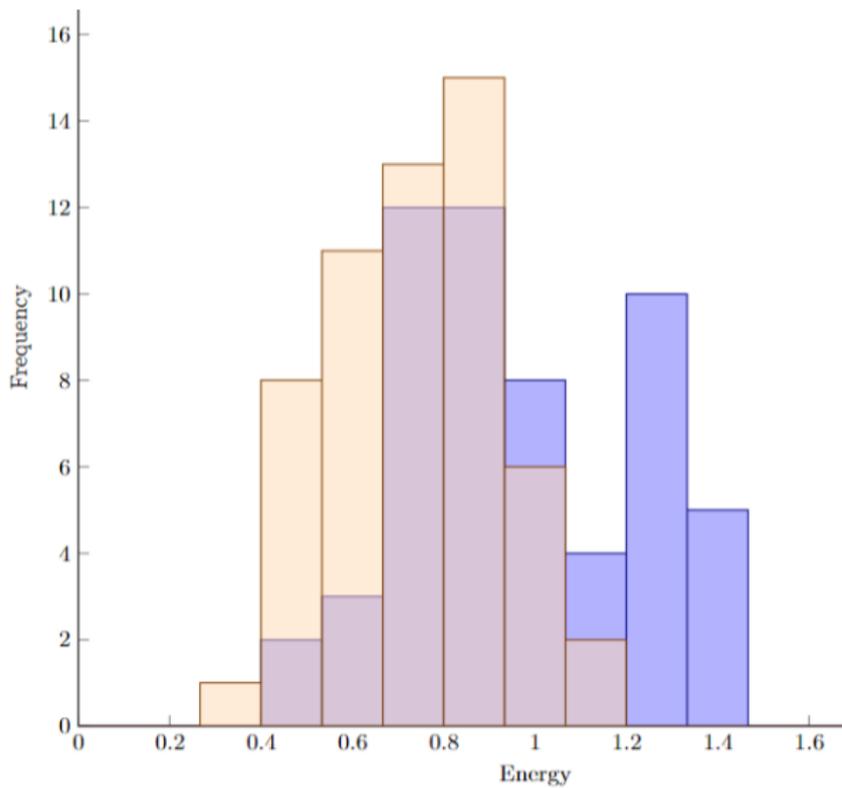
We would like to showcase a performance **comparison between two versions of QAOA** (standard and the newly proposed one) in determining appropriate edge colorings for the graph. Both algorithms are configured iteratively with a depth parameter of $p = 9$. Through over 50 independent trials for each scenario, we observe **statistically significant differences in mean values at the 1.5% significance level**, with the **new variant consistently demonstrating lower means**. Moreover, we note **considerably lower median and minimal values** in the experiments utilizing the **newly introduced mixer Hamiltonian** compared to the classical one.

Graph 1 (4 colors)

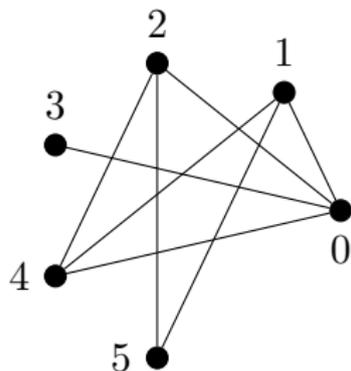


	Mean	Median	Min	Energy < 1
QAOA	0.9696	0.9316	0.4814	33/56
QAOA _{new}	0.7437	0.7388	0.3691	51/56

t-test p-value is $3.053993311768478 \cdot 10^{-7}$

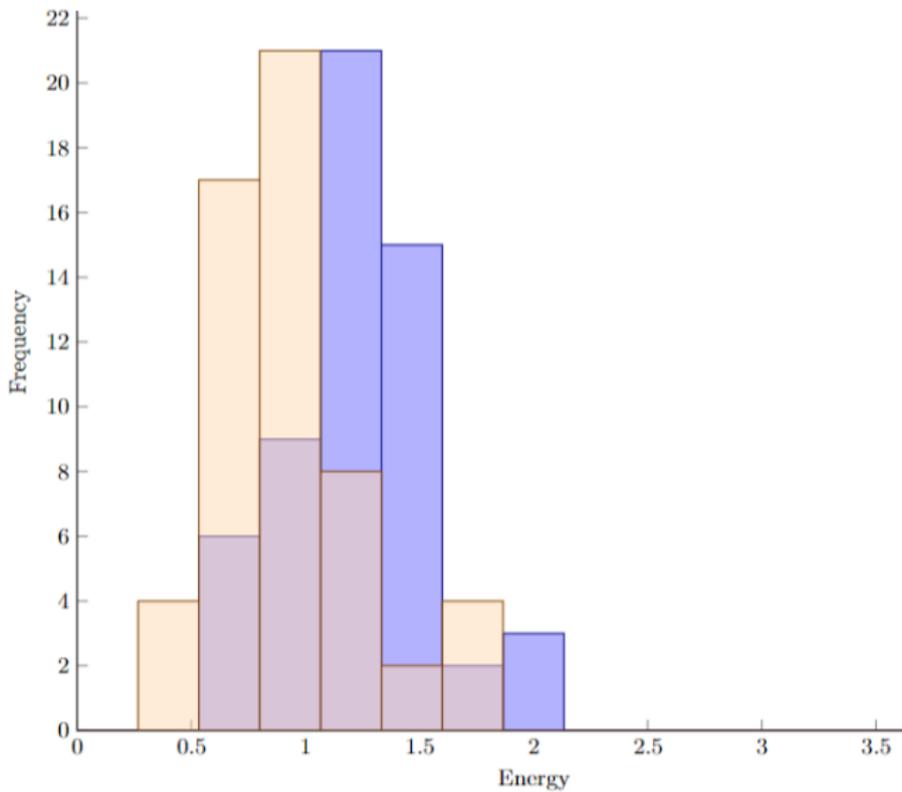


Graph 2 (4 colors)

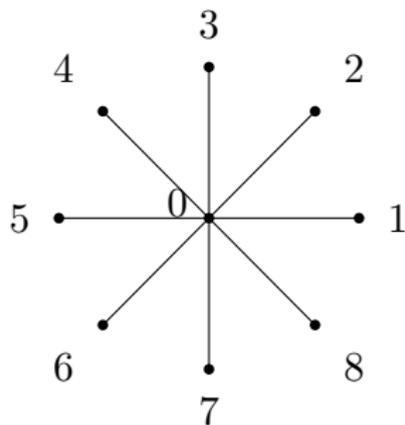


	Mean	Median	Min	Energy < 1
QAOA	1.2495	1.2417	0.6533	11/56
QAOA _{new}	0.9344	0.8857	0.3691	35/56

t-test p-value is $1.9806919304846427 \cdot 10^{-6}$

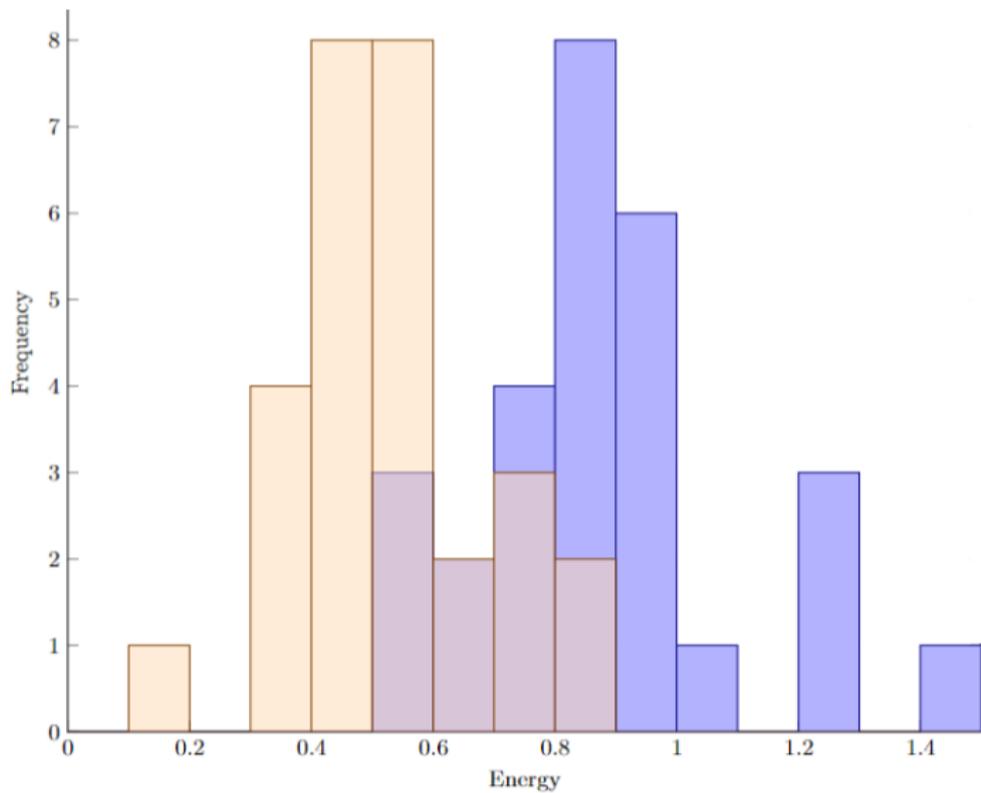


Graph 3 (8 colors)



	Mean	Median	Min
QAOA	0.8726	0.8569	0.502
QAOA _{new}	0.5227	0.5073	0.17

t-test p-value is $1.2230598272375008 \cdot 10^{-8}$



Decomposition of W According to G -action

Under the action of G , the space W decomposes as a direct sum of **isotypic components** (irreducible representations with multiplicities):

$$W = \bigoplus_i W_i.$$

Because the actions of G and H_P commute, this decomposition is preserved by H_P :

$$H_P(W_i) \subseteq W_i.$$

In particular, one of these subspaces is the *subspace of G -invariants*:

$$W^G = \{w \in W \mid g(w) = w, \forall g \in G\}.$$

Decomposition of W According to G -action

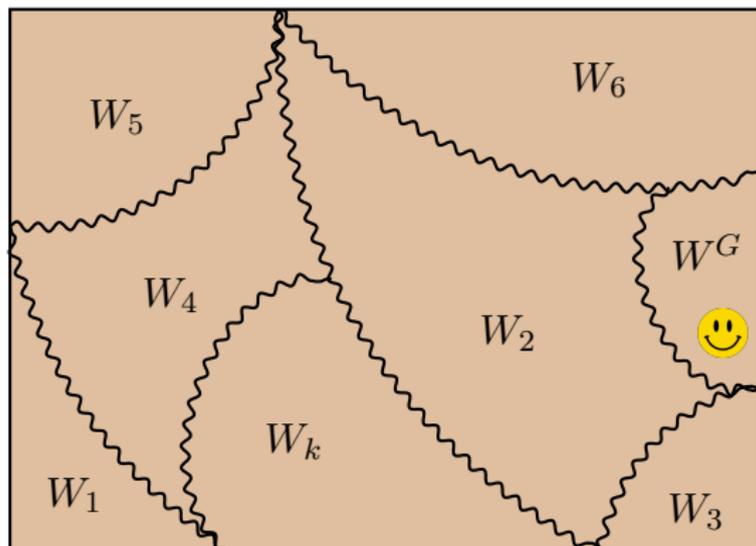


Figure: Decomposition of W

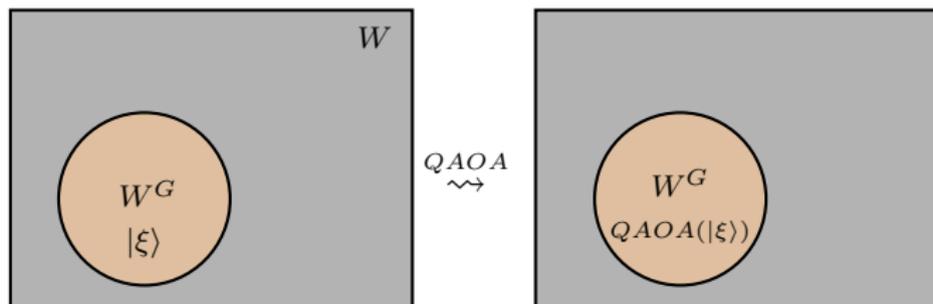
Where Does QAOA 'Live'?



Remark

Notice that the uniform superposition $|\xi\rangle = |++\dots+\rangle$ is inside W^G for any $G \subseteq \mathcal{S}$.

Suppose we employ a QAOA with the initial state $|\xi\rangle$, and the objective function has a group of symmetries G . **If the mixer Hamiltonian commutes with G , then the algorithm will operate within the subspace W^G prior to the final measurement.**



Where Does QAOA 'Live'?

Question

Is it possible to pick an initial state and mixer Hamiltonian so that QAOA 'runs' in a different $W_i \neq W^G$?

If there exists a mixer Hamiltonian $H_{M,i}$ that meets the following criteria:

- the lowest energy eigenspace of $H_{M,i}$ is one-dimensional and is spanned by $|\xi_i\rangle \in W_i$,
- $H_{M,i}$ preserves the direct sum decomposition of W (suffices to commute with G -action),

then one can establish a *reduced* QAOA with the same problem Hamiltonian H_P , mixer Hamiltonian $H_{M,i}$ and initial state $|\xi_i\rangle$. Thus defined QAOA operates within the subspace W_i prior to the final measurement.

We would like to summarize the arguments in favor of choosing the *reduced* QAOA on the subspace of invariants $W^G \subseteq W$ over any other W_i , when considering COPs with classical symmetry groups that include S_d .

- 1 The subspace W^G has the smallest dimension for a sufficiently large number of qudits, n .
- 2 It is impossible to ensure the convergence of reduced QAOA _{i} (operating on W_i) as the number of iterations p approaches infinity with the help of Perron-Frobenius Theorem for any other W_i .
- 3 The uniform superposition $\xi = |++\dots+\rangle$, which is used as the initial vector in the standard QAOA resides in W^G .

However, there might be hope for other W_i 's 

Graph	Mean energy	Median energy	Min energy	Share of outcomes with $E < 1$
Γ_1, W	0.726	0.7056	0.3584	41/50
Γ_1, W^G	0.5692	0.4673	0.1923	47/50
Γ_1, W_1	0.5726	0.5142	0.1621	47/50
Γ_2, W	0.9696	0.9316	0.4814	33/56
Γ_2, W^G	0.7437	0.7388	0.3691	51/56
Γ_2, W_1	0.8688	0.7148	0.3964	47/56
Γ_3, W	1.2495	1.2417	0.6533	11/56
Γ_3, W^G	0.9344	0.8857	0.3691	35/56
Γ_3, W_1	0.7334	0.6763	0.2598	50/56
Γ_4, W	1.4857	1.5313	0.7382	6/56
Γ_4, W^G	1.1959	1.1074	0.5117	21/56
Γ_4, W_1	1.2415	1.1489	0.4395	20/56
Γ_5, W	1.3469	1.3066	0.6162	14/50
Γ_5, W^G	0.9149	0.9507	0.3516	30/50
Γ_5, W_1	0.94123	0.9375	0.2939	27/50

Reduced QAOA energy for edge coloring with $p = 9$ (on W, W^G and W_1)

Recall that the classical Grover search algorithm addresses the task of identifying a marked element among 2^n possibilities, where the set \mathcal{M} of marked elements may contain one or several items. The algorithm achieves query complexity $\mathcal{O}(\sqrt{2^n/|\mathcal{M}|})$, where $|\mathcal{M}|$ denotes the number of marked elements.

The QAOA formulation of Grover search employs the *Grover mixer*, defined as the projector onto the initial state:

$$G_M := |\xi\rangle \langle \xi| = U_\xi |0 \cdots 0\rangle \langle 0 \cdots 0| U_\xi^\dagger$$

and, is commonly referred to as the *Grover-mixer QAOA* (GM-QAOA).

An Important Observation and Its Consequences

Let \mathfrak{g}_ξ be the DLA of GM-QAOA corresponding to the initial state $|\xi\rangle$. For a DLA $\mathfrak{g} = \langle iH_M, iH_P \rangle_{\text{Lie}}$, generated by Hamiltonians H_M and H_P , the **commutant** $\mathcal{C}(\mathfrak{g}) = \mathcal{C}(H_M, H_P)$ is the set of all linear operators on W that commute with both H_M and H_P .

In the physical setting, elements of the commutant correspond to *conserved quantities*—operators that remain invariant under the dynamics generated by the DLA.

Observation

The Grover mixer commutes with all elements in the centralizer of H_P that preserve the one-dimensional subspace $\mathbb{C}|\xi\rangle$.

Consequences:

- among the DLAs associated with a QAOA circuit for a fixed problem Hamiltonian and initial state, the commutant $\mathcal{C}(\mathfrak{g}_\xi)$ is **maximal** with respect to inclusion;
- the associative algebra $\mathcal{A}(\mathfrak{g}_\xi)$ generated by \mathfrak{g}_ξ is **minimal** with respect to inclusion.

One More Important Consequence

Let $|\xi\rangle = \frac{1}{\sqrt{|X|}} \sum_{x \in X} |x\rangle$ be the *uniform superposition state*, and let

$W_0 \subseteq W$ denote the *cyclic representation* of the DLA \mathfrak{g}_ξ , generated by $|\xi\rangle$.

For each value λ_j of the objective function F , define the normalized uniform vector over the corresponding level set:

$$|\xi_j\rangle := \frac{1}{\sqrt{n_j}} \sum_{x:F(x)=\lambda_j} |x\rangle, \quad j = 1, \dots, r.$$

Key properties:

- The set $\{|\xi_1\rangle, \dots, |\xi_r\rangle\}$ forms an **orthonormal basis** for W_0 .
- The subspace W_0 is an **irreducible representation** of \mathfrak{g}_ξ .

Remark

One of the basis vectors above is the **uniform superposition of all optimal solutions**.

One More Important Consequence

Consequence. The inclusion of associative algebras $\mathcal{A}(\mathfrak{g}_\xi) \subseteq \mathcal{A}(\mathfrak{g})$ implies that

$W_0 \subseteq$ irreducible representation of any DLA \mathfrak{g}

associated with a QAOA circuit for the same problem Hamiltonian and initial state $|\xi\rangle$.

Therefore: any QAOA with the same initial state and problem Hamiltonian will reach an optimal solution for sufficiently large circuit depth.

The *loss function* in QAOA is defined as the expectation value of the problem Hamiltonian H_P with respect to the parameterized quantum state:

$$\ell_{\boldsymbol{\beta}, \boldsymbol{\gamma}}(\rho, H_P) := \langle \psi(\boldsymbol{\beta}, \boldsymbol{\gamma}) | H_P | \psi(\boldsymbol{\beta}, \boldsymbol{\gamma}) \rangle = \text{Tr} \left[U(\boldsymbol{\beta}, \boldsymbol{\gamma}) \rho U^\dagger(\boldsymbol{\beta}, \boldsymbol{\gamma}) H_P \right],$$

where $\rho = |\xi\rangle \langle \xi|$ denotes the pure-state density operator associated with the initial state $|\xi\rangle$.

Barren Plateaus: an Obstruction to Convergence

Definition

A *barren plateau* is a region in the parameter landscape of a variational quantum algorithm where the gradients of the cost function are exponentially suppressed in the system size n .

Consequence

When gradients vanish exponentially, classical optimization methods fail to efficiently navigate the landscape, leading to extremely slow or stalled convergence.

A key diagnostic of this behavior is the variance of the loss function over the parameter space: when this variance is exponentially small in the number of qubits n , the corresponding gradient magnitudes also vanish exponentially.

Lie Algebras Don't Lie

Let us decompose the dynamical Lie algebra \mathfrak{g} as

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_k \oplus \mathfrak{z},$$

into simple (compact) Lie algebras \mathfrak{g}_j together with a (possibly trivial) center \mathfrak{z} .

Then the variance of the loss function (for sufficiently large depth p) is approximately:

$$\text{Var}_{\beta, \gamma} [\ell_{\beta, \gamma}(\rho, H_P)] \approx \sum_{j=1}^k \frac{\mathcal{P}_{\mathfrak{g}_j}(\rho) \mathcal{P}_{\mathfrak{g}_j}(H_P)}{\dim(\mathfrak{g}_j)},$$

where $\mathcal{P}_{\mathfrak{g}_j}(H) := \text{Tr}[H_{\mathfrak{g}_j}^2]$ and $H_{\mathfrak{g}_j}$ denotes the orthogonal projection of an operator H onto the subalgebra $\mathfrak{g}_j \subseteq \text{End}(W)$.

An Important Class of Objective Functions

Let $\mathbb{B}^n := \{0, 1\}^n$ denote the set of all binary strings of length n . We say that an objective function $F: \mathbb{B}^n \rightarrow \mathbb{R}$ is *s-local* if it can be written as

$$F(x) = \sum_{j=1}^T f_j(x_{S_j}),$$

where each component function $f_j: \mathbb{B}^s \rightarrow \mathbb{R}$ depends only on the subset of bits indexed by $S_j \subseteq \{1, \dots, n\}$ with $|S_j| = s$. The total number of such local terms is $T \leq \binom{n}{s}$.

Main Result for Applications

Theorem

Suppose that the objective function $F(x)$ is s -local and takes only integer values, and the initial state $|\xi\rangle$ is not an eigenvector of the problem Hamiltonian H_P . Let

$$M := \max_{1 \leq j \leq T} \max_{y \in \mathbb{B}^s} |f_j(y)|$$

denote the maximum absolute value attained by the local component functions f_j .

Then, for sufficiently large circuit depth p , the variance of the loss function $\ell_{\beta, \gamma}(\rho, \hat{H}_P)$ over the parameters (β, γ) satisfies the lower bound

$$\text{Var}_{\beta, \gamma}[\ell_{\beta, \gamma}(\rho, \hat{H}_P)] \geq \frac{(s!)^2}{12 M^2 n^{2s}},$$

where \hat{H}_P is the normalized problem Hamiltonian with respect to the spectral norm.

Example 1: MaxCut

The *MaxCut* problem is a well-known combinatorial optimization problem defined on an undirected graph $G = (V, E)$. Given a binary assignment of vertices to two disjoint sets, the goal is to maximize the number of edges that have endpoints in different sets. Equivalently, the problem seeks a partition of the vertex set $V = \{1, \dots, n\}$ into two parts such that the number of *crossing edges*—those whose endpoints belong to different subsets—is maximized.

For a binary string $x = (x_1, \dots, x_n) \in \mathbb{B}^n$, where $x_i = 0$ if and only if vertex i is placed in the first subset of the cut, the objective function for MaxCut is given by:

$$F(x) = \sum_{(i,j) \in E} ((1 - x_i)x_j + x_i(1 - x_j)).$$

Lower bound on loss function variance for GM-QAOA:

$$\text{Var}_{\beta, \gamma} [\ell_{\beta, \gamma}(\rho, \hat{H}_P)] \geq \frac{2^2}{12 n^4} = \frac{1}{3n^4}.$$

Example 2: the m -SAT problem

The m -SAT problem asks whether there exists a bit string $x = (x_1, \dots, x_n) \in \mathbb{B}^n$ that satisfies a given collection of clauses, each of which is a disjunction of m literals (variables or their negations). We can encode this as an m -local objective function, where each clause function $f_j: \mathbb{B}^m \rightarrow \{0, 1\}$ evaluates to 1 if the clause is satisfied and 0 otherwise, and $S_j \subseteq \{1, \dots, n\}$ is the subset of variables appearing in the j -th clause.

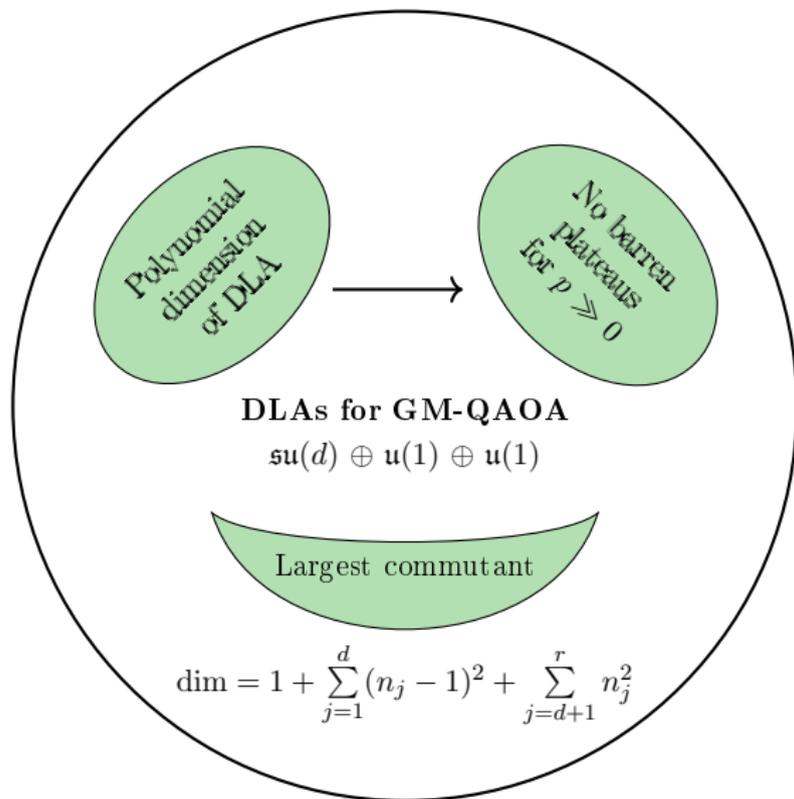
Applying the Theorem with $s = m$ and $M = 1$, we obtain a lower bound on the variance of the loss function:

$$\text{Var}_{\beta, \gamma} [\ell_{\beta, \gamma}(\rho, \hat{H}_P)] \geq \frac{(m!)^2}{12n^{2m}},$$

for sufficiently large circuit depth p .

Other examples with inverse polynomial bound on $\text{Var}_{\beta,\gamma}[\ell_{\beta,\gamma}(\rho, \hat{H}_P)]$ include weighted MaxCut, Max- k -VertexCover, Traveling Salesperson Problem

DLAs for GM-QAOA: Summary of Our Results



An Ongoing Project: Effect of Classical Reductions on QAOA Performance

In many combinatorial optimization problems defined on \mathbb{B}^n , the objective function F is invariant under the global bit-flip symmetry

$$x \mapsto \bar{x}, \quad \bar{x}_i = 1 - x_i.$$

This symmetry corresponds to an action of \mathbb{Z}_2 on \mathbb{B}^n .

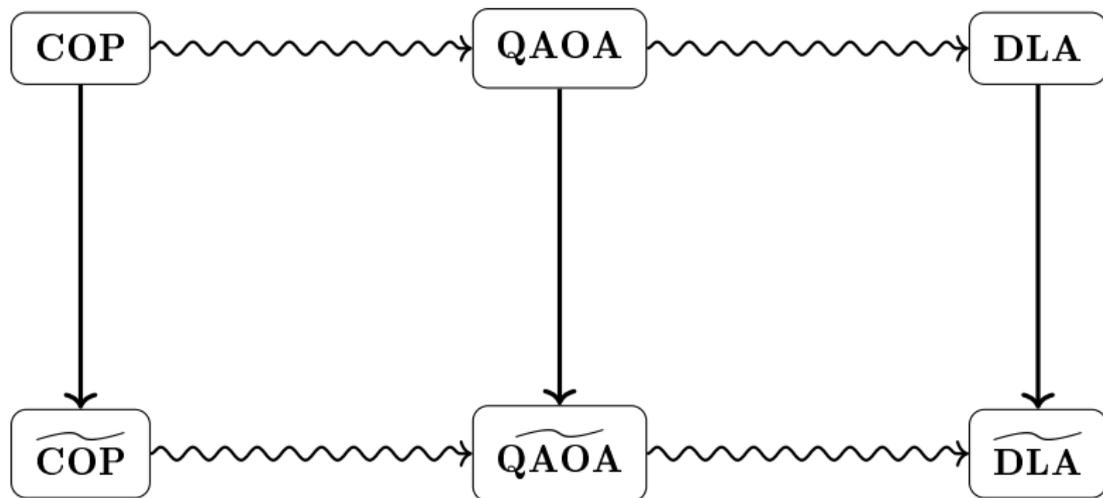
To construct a reduced problem, we fix one coordinate—say $x_s = 0$ for some $1 \leq s \leq n$ —as a representative of each \mathbb{Z}_2 -orbit. The reduced configuration space then becomes \mathbb{B}^{n-1} .

Remark

The solutions of the original problem are in a 2-to-1 correspondence with the solutions of the reduced problem. Equivalently, each solution of the original problem corresponds to a \mathbb{Z}_2 -orbit of solutions of the reduced problem. Thus, no solution information is lost under this reduction.

An Ongoing Project: Effect of Classical Reductions on QAOA Performance

We are investigating how classical symmetry reductions influence the expressive power of QAOA and the associated dynamical Lie algebras. We define the *reduced dynamical Lie algebras* as those generated by the two QAOA Hamiltonians acting on the reduced Hilbert space $(\mathbb{C}^2)^{\otimes n-1}$.



An Ongoing Project: Effect of Classical Reductions on QAOA Performance

Overview

We study how classical preprocessing (e.g., variable fixing or symmetry reduction) affects the expressive power of QAOA.

Observation

Although such reductions are classically straightforward, their influence on the **associated dynamical Lie algebras** can be highly nontrivial and **depends** sensitively on the choice of the **fixed coordinate** (representative bit).

Implication

Different classical reductions may lead to QAOA instances with distinct controllability properties and performance characteristics.

An Ongoing Project: Effect of Classical Reductions on QAOA Performance

This is a joint work with graduate students from the University of Delaware:

- Bao Bach
- José Falla
- Yuan-Chieh (Jimmy) Chen
- Cameron Ibrahim