Variational Quantum Circuit Optimization via Interior Point Method

Jiaxin Huang

Dept. of Computer Science
University of Hong Kong
jiaxin.huang@connect.hku.hk

Abstract—**Variational quantum algorithms (VQAs) have emerged as a promising approach to leverage near-term quantum devices and demonstrate quantum supremacy. However, highly expressive variational quantum circuits, known as ansatzes, often encounter trainability issues, such as the barren plateau problem, where gradients of trainable parameters vanish exponentially with increasing qubit numbers. To address these challenges, this project introduces specially designed circuit templates, focusing on the State Efficient Ansatz (SEA) that utilizes Schmidt decomposition to reduce problem size and amplify gradients. In particular, a sub-case of SEA is studied, where a layer of R_y rotations serves as the Schmidt coefficient layer. The expressibility of this ansatz is investigated by casting it as an optimization problem. Then the problem is solved using the primal-dual interior point method, alongside a comparison with the conventional stochastic gradient descent method. The results demonstrate the superior convergence of the interior point method. Additionally, the observed solution structures offer potential refinements to the problem formulation, opening avenues for future generalization. This study contributes to the advancement of variational quantum algorithms, offering insights into optimizing circuit design and overcoming trainability limitations.**

I. Introduction

Variational quantum algorithms hold the promise of demonstrating quantum supremacy on noisy intermediate-scale quantum (NISQ) devices in the near future [1]. However, it has been observed that highly expressive variational quantum circuits, known as ansatzes, can suffer from trainability issues [2]. One such issue is the barren plateau phenomenon, which occurs when the gradients of the trainable parameters exponentially vanish as the number of qubits increases [3].

To address these challenges, researchers have proposed specially designed circuit templates that aim to strike a balance between expressibility and trainability [4]. Among these templates is the State Efficient Ansatz (SEA) [5], which leverages Schmidt decomposition [6] to reduce the problem size by a factor of two, thereby quadratically amplifying the gradients. The structure of SEA is illustrated in Figure [insert figure number]. Unlike general ansatzes that require unitary universality, SEA only assumes wavefunction universality for the Schmidt coefficient layer, denoted as U_1. However, an efficient method for constructing U_1 that satisfies these properties remains unclear.

In this project, we focus on studying a specific sub-case of SEA, where U_1 is instantiated as a layer of R_y rotations, given by:

$$U_1(\boldsymbol{\theta}) = \bigotimes_{i=1}^{N} R_y(\theta_i), \tag{1}$$

where θ_i represents the rotation angles. Our objectives are twofold. First, we investigate the expressibility of this ansatz, exploring whether it can approximate any arbitrary wavefunctions. Second, given the critical importance of trainability in VQA research, we employ the primal-dual interior-point method to optimize this ansatz, comparing its efficiency with the conventional gradient descent method commonly used in the literature.

II. Literature Review

The optimization of variational quantum circuits currently relies on classical optimizers, as there is no established theory for designing quantum optimizers. In particular, machine learning optimizers have gained significant popularity due to their success in the field and their similarity to variational quantum circuits. Various machine learning optimization techniques have been widely adopted, ranging from the vanilla stochastic gradient descent method to more advanced gradient-based methods such as RMSprop and Adam [7], [8].

However, it has recently been discovered that universal quantum ansatzes may encounter the barren plateau problem, which refers to the exponential vanishing of gradients for trainable parameters as the number of qubits, denoted as N, increases. This phenomenon poses a significant limitation on the scalability of gradient-based optimization methods [3]. Consequently, alternative approaches that do not rely on gradients, known as gradient-free methods, have gained attention and consideration in the optimization of variational quantum circuits.

In a recent study by Monroig et al. [9], four gradient-free optimization methods (SLSQP, COBYLA, CMA-ES, and SPSA) were compared across multiple systems of varying sizes. This work aimed to assess the effectiveness and performance of these methods in addressing the barren plateau problem. By evaluating their performance on different systems, the researchers sought to provide insights into the applicability and potential limitations of gradient-free optimization techniques in the context of variational quantum circuits.

In [10], Díez-Valle et al. also formulated the placement of quantum gates in a circuit template as a quadratic unconstrained binary optimization (QUBO) problem. Traditionally, the placement of quantum gates within a circuit has been approached using heuristic algorithms and expert knowledge.

Traditionally, the discrete placement of quantum gates within a circuit is approached using heuristic algorithms and expert knowledge. Only continuous parameters within the gates are optimized. Inspired by classical automated machine learning (AutoML, [11]), Díez-Valle et al. proposed a novel perspective by casting the gate placement problem as a quadratic unconstrained binary optimization (QUBO) problem, leveraging the power of optimization techniques commonly used in combinatorial optimization [10].

III. Problem formulation

Let $|\varphi\rangle$ be a $2N$ qubit quantum state to be synthesized. It admits Schmidt decomposition

$$
|\varphi\rangle = \sum_{j=1}^{2^N} \omega_j |a_j\rangle |b_j\rangle
$$
$$
= (U \otimes V) \sum_{j=1}^{2^N} \omega_j |j\rangle |j\rangle, \tag{2}
$$

for some unitaries U, V and $\omega_j \geq 0$.

Let $|\psi\rangle$ be the output of the SEA circuit, with SC layer $U_1 = \bigotimes_{i=1}^N R_y$ and some LBC layer U_2, U_3. Below we analyze the effect of SEA on $2N$ basis states $|0\rangle^{\otimes 2N}$

We start with the SC layer. Let $\theta \in \mathbb{R}$ be a rotation angle. The $R_y(\theta)$ rotation gate applied to the basis state $|0\rangle$ produces the effect

$$
|0\rangle \to R_y(\theta)|0\rangle = \sin\theta|0\rangle + \cos\theta|1\rangle \tag{3}
$$

Therefore, given a vector $\theta \in \mathbb{R}^N$ of N rotation angles, the SC layer $U_1(\theta)$ applied to N basis states $|0\rangle^{\otimes N}$ gives

$$
|0\rangle^{\otimes N} \to |\theta\rangle := U_1(\theta)|0\rangle^{\otimes N}
$$
$$
= \left(\bigotimes_{i=1}^N R_y(\theta_i) \right) |0\rangle^{\otimes N}
$$
$$
= \bigotimes_{i=1}^N \left(R_y(\theta_i)|0\rangle \right) \tag{4}
$$
$$
= \bigotimes_{i=1}^N \left(\sin\theta_i|0\rangle + \cos\theta_i|1\rangle \right)
$$
$$
= \sum_{i=1}^{2^N} \mu_i |i\rangle,
$$

where

$$
\mu := \bigotimes_{i=1}^N \begin{bmatrix} \sin\theta_i \\ \cos\theta_i \end{bmatrix} \in \mathbb{R}^{2^N}. \tag{5}
$$

As an example, when $N = 2$, $\mu = \begin{bmatrix} \sin\theta_1 \sin\theta_2 \\ \sin\theta_1 \cos\theta_2 \\ \cos\theta_1 \sin\theta_2 \\ \cos\theta_1 \cos\theta_2 \end{bmatrix}$.

Then, the remaining layers give

$$
|0\rangle^{\otimes 2N} \to |\psi\rangle := \mathrm{SEA}\, |0\rangle^{\otimes 2N}
$$
$$
= (U_2 \otimes U_3)\, \mathrm{CNOT}\left(|\theta\rangle, |0\rangle^{\otimes N} \right)
$$
$$
= (U_2 \otimes U_3) \sum_{i=1}^{2^N} \mu_i\, \mathrm{CNOT}\left(|i\rangle, |0\rangle^{\otimes N} \right) \tag{6}
$$
$$
= (U_2 \otimes U_3) \sum_{i=1}^{2^N} \mu_i |i\rangle |i\rangle.
$$

Therefore, the synthesis loss can be characterized by the fidelity

$$
F = \langle \psi | \varphi \rangle
$$
$$
= \left(\sum_{i=1}^{2^N} \mu_i \langle i| \langle i| \right) \left(U_2^\dagger \otimes U_3^\dagger \right)
$$
$$
(U \otimes V) \left(\sum_{j=1}^{2^N} \omega_j |j\rangle |j\rangle \right) \tag{7}
$$
$$
= \sum_{i=1}^{2^N} \sum_{j=1}^{2^N} \mu_i \omega_j \left\langle i \middle| U_2^\dagger U \middle| j \right\rangle \left\langle i \middle| U_3^\dagger V \middle| j \right\rangle.
$$

This objective function involes optimization over unitaries U_1 and U_2, which is computationally difficult. We reduce the problem by assuming $U_2^\dagger U = I$ and $U_3^\dagger V = P$ for some permutation matrix P.

The rearrangement inequality states that, for any fixed pair of μ and ω, F is maximized when P is a permutation that rearranges μ to the same order as ω, i.e.,

$$
\omega_k < \omega_l \Leftrightarrow (P\mu)_k < (P\mu)_l. \tag{8}
$$

Therefore, we could write

$$
F = (P\mu)^T \omega. \tag{9}
$$

We further assume that $\boldsymbol{\mu}$ is homogenious across all of its 2^N dimensions. That is, if $\boldsymbol{\mu}^*$ is attainable, so is $P\boldsymbol{\mu}^*$ for all permutation P. (In fact this is not true due to the tensor product structure of $\boldsymbol{\mu}$.) Under this assumption, we can drop the matrix P by assuming $\boldsymbol{\omega}$ is given in ascending order, i.e.,

$$F = \boldsymbol{\mu}^T \boldsymbol{\omega}. \quad (10)$$

Moreover, we note that Schmidt decomposition guarantees $\boldsymbol{\omega} \succcurlyeq 0$. Therefore, F is maximized when $\boldsymbol{u} \succcurlyeq 0$, which is equivalent to $0 \preccurlyeq \boldsymbol{\theta} \preccurlyeq \frac{\pi}{2}\boldsymbol{I}$.

In summary, synthesizing a given state $|\varphi\rangle$ is equivalent to the following optimization problem:

$$\underset{\boldsymbol{\theta}}{\text{minimize}} - \boldsymbol{\mu}^T \boldsymbol{\omega}$$

$$\text{subject to } 0 \preccurlyeq \boldsymbol{\theta} \preccurlyeq \frac{\pi}{2}\boldsymbol{I}$$

$$\text{where } \boldsymbol{\mu} = \bigotimes_{i=1}^{N} \begin{bmatrix} \sin\theta_i \\ \cos\theta_i \end{bmatrix} \quad (11)$$

$$\omega_i \leq \omega_{i+1}, \forall i.$$

IV. ALGORITHM

The interior point method was chosen to solve the optimization problem. There is no need to implement phase 1 method because a feasible starting point can be easily found: currently, it's chosen as $\boldsymbol{\theta}^{(0)} = \frac{\pi}{4}\boldsymbol{1}$, which is the center of the feasible set. To accelerate convergence, the primal-dual form of the algorithm is used.

Several important functions are explained below. The objective function is an inner product of two vectors with length 2^N:

$$f_0(\boldsymbol{\theta}) := -\left(\bigotimes_{i=1}^{N} \begin{bmatrix} \sin\theta_i \\ \cos\theta_i \end{bmatrix}\right)^T \boldsymbol{\omega}. \quad (12)$$

The constraints are

$$\boldsymbol{g}(\boldsymbol{\theta}) := \text{concat}\left(-\boldsymbol{\theta}, \boldsymbol{\theta} - \frac{\pi}{2}\boldsymbol{1}\right) \preccurlyeq 0, \quad (13)$$

which corresponds to $0 \leq \theta_i \leq \frac{\pi}{2}$ exactly.

The residual is

$$\boldsymbol{r}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \begin{bmatrix} \boldsymbol{r}_{\text{dual}} \\ \boldsymbol{r}_{\text{cent}} \end{bmatrix} := \begin{bmatrix} \nabla f_0(\boldsymbol{\theta}) + D\boldsymbol{g}(\boldsymbol{\theta})^T \boldsymbol{\lambda} \\ -\text{diag}(\boldsymbol{\lambda})\boldsymbol{g}(\boldsymbol{\theta}) - \frac{1}{t}\boldsymbol{1} \end{bmatrix}. \quad (14)$$

Note that there is no primal residual as the optimization problem has no equality constraints.

In the primal-dual form, solving the optimization method is equivalent to find the root

$$\boldsymbol{r}(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*) = 0. \quad (15)$$

By Newton method, this can be iteratively solved by solving the linear approximation

$$\boldsymbol{r}(\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}) + [D\boldsymbol{r}(\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)})](\Delta\boldsymbol{\theta}, \Delta\boldsymbol{\lambda})^T = 0, \quad (16)$$

where $D\boldsymbol{r}$ is the Jacobian

$$Dr(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla^2 f_0(\boldsymbol{\theta}) + \sum_{i=1}^{2N} \lambda_i \nabla^2 g_{i(\boldsymbol{\theta})} & D\boldsymbol{g}(\boldsymbol{\theta})^T \\ -\text{diag}(\boldsymbol{\lambda})D\boldsymbol{g}(\boldsymbol{\theta}) & -\text{diag}(\boldsymbol{g}(\boldsymbol{\theta})) \end{bmatrix} \quad (17)$$

After finding the direction $(\Delta\boldsymbol{\theta}, \Delta\boldsymbol{\lambda})$, the backtracking line search is used to determine the step size. Backtracking line search finds the larest step size s which:

- $0 < s < 1$;
- maintains the primal constraint $\boldsymbol{g}(\boldsymbol{\theta}) \preccurlyeq 0$;
- maintains the dual constraint $\boldsymbol{\lambda} \succcurlyeq 0$;
- reduces $\|\boldsymbol{r}\|_2$ by at least a factor of αs (currently, α is set as 0.5).

Lastly, the algorithm terminates when dual residual is smaller than a predetermined threshold $\varepsilon_{\text{feas}}$, and the surrogate duality gap

$$\eta := -\boldsymbol{g}(\boldsymbol{\theta})^T \boldsymbol{\lambda} \quad (18)$$

is smaller than another threshold ε.

The complete pseudocode is shown in Algorithm 1.

To evaluate the performance of the interior point method, a stochastic gradient descend (SGC) method is also implemented. The gradient descend method simply computes the gradient of the objective function

$$\nabla f_0(\boldsymbol{\theta}) = \begin{bmatrix} -\left(\begin{bmatrix} \cos\theta_1 \\ -\sin\theta_1 \end{bmatrix} \otimes \left(\bigotimes_{i=2}^{N} \begin{bmatrix} \sin\theta_i \\ \cos\theta_i \end{bmatrix}\right)\right)^T \boldsymbol{\omega} \\ \vdots \\ -\left(\left(\bigotimes_{i=1}^{N-1} \begin{bmatrix} \sin\theta_i \\ \cos\theta_i \end{bmatrix}\right) \otimes \begin{bmatrix} \cos\theta_N \\ -\sin\theta_N \end{bmatrix}\right)^T \boldsymbol{\omega} \end{bmatrix} \quad (19)$$

and update parameters

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + s\nabla f_0(\boldsymbol{\theta}) \quad (20)$$

by some step size s.

input: $\omega \succcurlyeq 0, \varepsilon_{\text{feas}} > 0, \varepsilon > 0, \nu > 1,$
$$\beta \in (0, 1), \alpha \in (0, 1)$$

output: $\theta, f_0(\theta)$

1 $\lambda \leftarrow \mathbf{1}_{2N}$
2 **repeat**
3    $t \leftarrow 2\nu N/\eta$
4   **solve** $A \begin{bmatrix} \Delta\theta \\ \Delta\lambda \end{bmatrix} + r = 0$ for $\begin{bmatrix} \Delta\theta \\ \Delta\lambda \end{bmatrix}$
5    $s \leftarrow \min(0.99, -\lambda_i/\Delta\lambda_i : \Delta\lambda_i < 0)$
6   **repeat**
7      $s \leftarrow \beta s$
8   **until** $g(\theta + s\Delta\theta) \preccurlyeq 0$, and $\|r(\theta + s\Delta\theta, \lambda + s\Delta\lambda)\|$
$$\leq (1 - \alpha s)\|r(\theta, \lambda)\|$$
9    $\theta \leftarrow \theta + s\Delta\theta$
10   $\lambda \leftarrow \lambda + s\Delta\lambda$
11   $r_{\text{dual}} \leftarrow r(\theta, \lambda)[: N]$
12   $\eta \leftarrow -g(\theta)^T\lambda$
13 **until** $\|r_{\text{dual}}\| \leq \varepsilon_{\text{feas}}, \eta \leq \varepsilon$
14 **return** $\theta, f_0(\theta)$

Algorithm 1: Primal-dual interior point method

V. Evaluation

We evaluate the optimization methods by using randomly sampled $2N$-qubit quantum state $|\varphi\rangle$. The coefficient ω can be obtained by singular value decomposition

$$M_{|\varphi\rangle} = U \operatorname{diag}(\omega)V, \tag{21}$$

where $M_{|\varphi\rangle}$ is the inverse vectorziation of $|\varphi\rangle$ by rewriting the 2^{2N} dimentional vector as a $2N \times 2N$ matrix.

We tested two scenarios where $N = 5$ and $N = 10$. For the interior point optimizer, the stopping criteria is when $r_{\text{dual}} \leq \varepsilon_{\text{feas}}$ and $\eta \leq \varepsilon$. We used $\varepsilon_{\text{feas}} = \varepsilon = 10^{-2}$ for both $N = 5$ and $N = 10$ cases. For the stochastic gradient descend optimizer, there is no explicit stopping criteria. So we terminat it after the same iteration number of the interior point optimizer on the same problem. In experiments, we observe that the typical iteration number is 51 for $N = 5$ and 97 for $N = 10$.

The optimization results are shown in Figure 2 and Figure 3. In Figure 2. we observe that the wavefunction generated by the interior point method (orange) exhibits a closer approximation to ω (green), which agrees with the fact that the objective function decreases faster for interior point method in Figure 3. Moreover, in both Figure 3a and Figure 3b, the curve for interior point method exhibits larger fluctuation than the curve for stochastic gradient descend method. This can be explained by the fact that the interior point method is not a descend method. It steadily descreases the residual r, but not necessarily the objective function $-F$.

Finally, we note that for large quantum systems, for example $N = 10$ in Figure 2b, the optimal value exhibits a very clear pattern: the ω is almost a straight line from 0 to $\sqrt{\frac{3}{2^N}}$, and μ is roughly four disjoint line sengments. Such clear pattern may allow us to drop some assumptions in Section III, hence further improving synthesis accuracy.
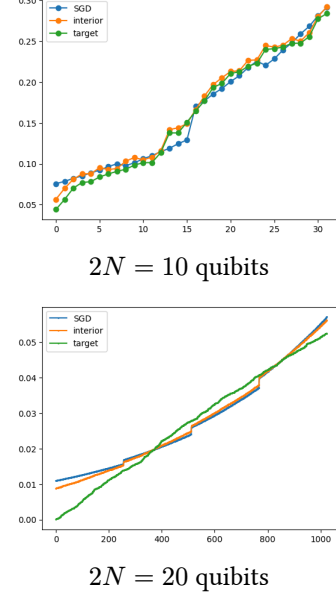

$2N = 10$ qubits

$2N = 20$ qubits

Figure 2: The wavefunction μ and ω after optimization.

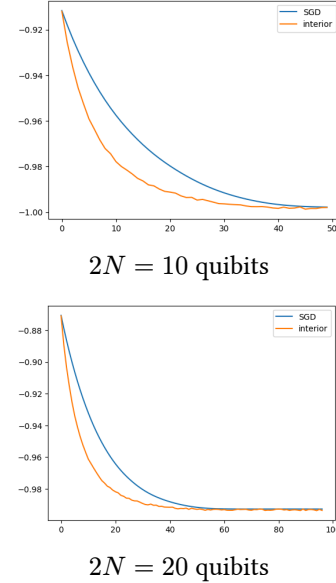$2N = 10$ qubits

$2N = 20$ qubits

Figure 3: Fidelity over optimization iterations.

VI. Discussion

In this project, the formulation of a variational quantum algorithm as an optimization problem has been investigated. The problem was tackled using two optimization methods, namely the primal-dual interior point method and the vanilla stochastic gradient descent method. Through comprehensive

experiments, it has been demonstrated that the interior point method exhibits faster convergence compared to the stochastic gradient descent method. This result sheds light on the potential for developing new optimizers that do not solely rely on the gradients of the objective function. By exploring alternative optimization methodologies, researchers may overcome the limitations imposed by the barren plateau problem in large variational circuits and pave the way for more efficient and effective variational quantum algorithms.

However, it is important to consider the computational efficiency implications associated with the interior point method. This approach necessitates the calculation of Hessian matrices for the objective and constraint functions, as well as solving a linear system with a dimension of $3N$ at each iteration. As a consequence, when optimizing large quantum circuits with a substantial number of tunable parameters, careful consideration must be given to strike a balance between computational efficiency and optimization effectiveness.

Moreover, the solution structures observed throughout this project offer valuable insights into potential refinements of the assumptions employed in the problem formulation. By refining these assumptions, it may be possible to improve computational efficiency or generalize the problem formulation, thereby expanding its applicability to a wider range of scenarios.

In summary, this project has successfully addressed the optimization of variational quantum algorithms by formulating them as optimization problems and exploring different solution methodologies. The findings highlight the superiority of the interior point method in terms of convergence speed and provide inspiration for the development of novel optimizers that can circumvent the barren plateau problem. Additionally, the computational efficiency considerations and the potential for refining the problem formulation assumptions underscore the need for continued research in this field to unlock the full potential of variational quantum algorithms.

References

[1] J. Preskill, "Quantum computing and the entanglement frontier", *arXiv: Quantum Physics*, 2012, [Online]. Available: https://api.semanticschola r.org/CorpusID:48828406

[2] K. Zhang, M.-H. Hsieh, L. Liu, and D. Tao, "Toward trainability of deep quantum neural networks", *arXiv preprint arXiv:2112.15002*, 2021.

[3] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes", *Nature communications*, vol. 9, no. 1, p. 4812, 2018.

[4] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, "Absence of Barren Plateaus in Quantum Convolutional Neural Networks", *Phys. Rev. X*, vol. 11, no. 4, p. 41011, Oct. 2021, doi: 10.1103/ PhysRevX.11.041011.

[5] X. Liu, G. Liu, J. Huang, H.-K. Zhang, and X. Wang, "Mitigating barren plateaus of variational quantum eigensolvers", *arXiv preprint arXiv: 2205.13539*, 2022.

[6] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*. Cambridge university press, 2010.

[7] A. Graves, "Generating sequences with recurrent neural networks", *arXiv preprint arXiv:1308.0850*, 2013.

[8] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", *arXiv preprint arXiv:1412.6980*, 2014.

[9] X. Bonet-Monroig *et al.*, "Performance comparison of optimization methods on variational quantum algorithms", *Physical Review A*, vol. 107, no. 3, p. 32407, 2023.

[10] P. D\iez-Valle, D. Porras, and J. J. Garc\ia-Ripoll, "Quantum variational optimization: The role of entanglement and problem hardness", *Physical Review A*, vol. 104, no. 6, p. 62426, 2021.

[11] X. He, K. Zhao, and X. Chu, "AutoML: A survey of the state-of-the-art", *Knowledge-Based Systems*, vol. 212, p. 106622, 2021.