Mutual Unbiased Bases as a Countermeasure to Barren Plateaus

Ittay Alfassi*

September 28, 2022

Contents

*Email: ittay.al@cs.technion.ac.il

1 Introduction

Noisy, Intermediate-Scale Quantum (NISQ) computers are quantum computers that are limited certain ways. They are usually limited in the number of qubits, connectivity between qubits, noise, and maximal circuit depth. A computational paradigm that utilizes NISQ devices is the one of hybrid quantum-classical algorithms. Most hybrid algorithms translate a given problem to an optimization problem. The parameters of the optimization problem control some parametrized quantum state, the problem's cost function is evaluated on the quantum computer, and the optimzation steps are performed by the classical computer. These algorithms are called Variational Quantum Algorithms (VQAs) [1]. TODO: Make sure this is accurate!

While being a leading methodology for the use of quantum computers today, VQAs suffer from several problems. One such problem is the problem of Barren Plateaus [2]. Informally, barren plateaus imply that over the optimization parameter space, the gradient of the cost function is negligibly small and the optimization process does not how to converge.

This problem affects many VQAs that employ different types of quantum circuits and classical optimizers (even those that are gradient-free). As the number of qubits and the depth of the quantum circuits increases, this problem worsens exponentially. Effectively, this means that VQAs will not be able to give an advantage over their classical counterparts in problems that are not small in size.

To address this issue, in this project, I will try and utilize Mutually Unbiased Bases (MUBs), their states, and their properties. The basic intuition behind this method is that, for any number of qubits, the set of all MUB states spans the entire Hilbert space (with real coefficients).

Thus, inserting them (in some fashion) into the optimization process can allow for an "exhaustive search" over the Hilbert space of all states.

Unfortunately, this method did not succeed. However, I will detail the different aspects of this approach, the experiments implemented, the conclusions I reached, and possible methods to still utilize the idea of MUB states in VQAs.

2 Preliminaries

2.1 Variational Quantum Algorithms

A Variational Quantum Algorithm is an algorithm that takes some computational problem, and solved is by hybrid classical-quantum optimization. VQAs are comprised of several elements: The quantum circuit, the cost function, and the optimizer.

1. **The Quantum Circuit.** Every VQA problem uses a quantum circuit with some parametric values. Its structure can be generally written as the following:

$$U(\vec{\theta}) = \prod_{l=1}^{L} U_l(\theta_l) W_l$$

   Where $U_l(\theta_l) = \exp(-i\theta_l V_l)$, V_l is a Hermitian operator, and W_l is a generic unitary operator that does not depend on any angle θ_l.

In some VQAs, an **initial state** $|\psi_0\rangle$ is also a part of the algorithm, and it is related to the quantum circuit.

The term *ansatz* (German for "approach"/"attempt") is used in literature to either describe the initial state $|\psi_0\rangle$, the parametric circuit $U(\vec{\theta})$, or the application of the parametric circuit to the initial state $|\psi(\vec{\theta})\rangle = U(\vec{\theta})\,|\psi_0\rangle$. In this project, I will use the term ansatz to refer to the parametric quantum circuit.

Ansatzes[1] can either be problem-specific or problem-agnostic.

A problem-specific ansatz is tailored to the problem solved by the VQE using a-priori knowledge on the structure of the problem. Through this tailoring there is a better chance that the circuit can reach values that are optimal for the problem at hand. When using a problem-specific ansatz, the initial state is usually also tailored according to the same information.

A problem-agnostic ansatz is not tailored to the problem specifically, and can be used with no relevant information. It is usually comprised of repeated layers of quantum gates. Each layer is comprised of quantum gates that are usually easy to implement directly on quantum hardware, in contrast to problem-specific ansatzes. Problem-agnostic ansatzes are usually called "hardware-efficient" for this reason. In this project, I focused on using hardware-efficient ansatzes.

2. **The Cost Function.** The cost function $C(\vec{\theta})$ defines the goal of the VQA. It is defined as a function from the parameters of the ansatz to the real numbers. The goal of the VQA is to find
$$\vec{\theta}^* = \arg\min_{\vec{\theta}} C(\vec{\theta})$$
The cost function always depends on the ansatz, some Hermitian operators, and some initial states. Generally, it can be written as
$$C(\vec{\theta}) = \sum_k f_k(\mathrm{Tr}[O_k U(\vec{\theta})\rho_k U^\dagger(\vec{\theta})])$$
Where $\{f_k\}$ is a set of functions, $\{O_k\}$ is a set of operators, and $\{\rho_k\}$ is a set of input states.

   The trademark of VQAs is that they use a quantum computer to estimate the cost function $C(\vec{\theta})$ (or its derivatives) while leveraging the power of classical optimizers to train the parameters $\vec{\theta}$.

3. **The Optimizer.** While not necessarily dependent on the problem the VQA is trying to solve, the classical optimizer is an important part of the algorithm. The optimizer is run on a classical computer, and performs an iterative process: it uses data on the current parameters of the cost function to calculate their value in the next iteration. More details can be found in subsection 2.2.

2.1.1 Example: Variational Quantum Eigensolver

One of the earliest examples for a VQA is the VQE algorithm, proposed by Peruzzo et al. [3]. In this algorithm, The goal is to find the lowest eigenvalue of some Hamiltonian H.

[1] The correct plural form is Ansaetze.

The cost function is straightforwardly defined as the expectation value of the parametric state over the Hamiltonian:

$$C(\vec{\theta}) = \langle\psi(\vec{\theta})| H |\psi(\vec{\theta})\rangle = \langle\psi_0| U^\dagger(\vec{\theta}) H U(\vec{\theta}) |\psi_0\rangle$$

In case H is a molecular Hamiltonian (converted to a qubit Hamiltonian using an appropriate transformation), a problem-specific ansatz can be used. One such option is the Unitary Coupled-Cluster (UCC) ansatz, together with the Hartree-Fock state as the initial state.

Of course, a hardware-efficient ansatz can be used for any structure of H.

2.1.2 Example: Variational Quantum Compiling

A different example for a VQA is the problem of Variational Quantum Compiling (VQC), initially defined by Khatri et al. [4] as Quantum-Assisted Quantum Compiling (QAQC).

In VQC, the input of the problem is a general n-qubit unitary operation V[2]. The goal of the algorithm is to control the parameters of a parametric quantum circuit $U(\vec{\theta})$ so its behavior will be similar to that of V.

Formally, there are two possible goals: either that $U(\vec{\theta})$ and V will perform the same operation on all states, or that $U(\vec{\theta})$ and V will perform the same operation on a specific input state. In this project, I will use the latter as the problem definition. Thus, we wish that $V |0\rangle = U(\vec{\theta}) |0\rangle$.

There are two different cost functions presented in [4]. The first is straightforward, and is defined by

$$C^{\text{global}} = 1 - \left|\langle 0| UV^\dagger |0\rangle\right|^2$$

Another cost function, called the local cost function, is defined by

$$C^{\text{local}} = 1 - \frac{1}{n}\sum_{j=1}^{n} p_0^{(j)}$$

where

$$p_0^{(j)} = \text{Tr}[(|0\rangle\langle 0|_j \otimes I) U^\dagger V |0\rangle\langle 0| V^\dagger U]$$

and $|0\rangle\langle 0|_j \otimes I$ is the tensor product of the matrix $|0\rangle\langle 0|$ in the jth term and identity matrices in all other terms. The operational meaning of $p_0^{(j)}$ is the probability to obtain the zero measurement outcome on qubit qubit j for the state $U^\dagger V |0\rangle$.

it is proven in [5] that

$$C^{\text{local}} \leq C^{\text{global}} \leq nC^{\text{local}}$$

However, C^{global} suffers from a provably vanishing gradient as n increases, while C^{local} does not. Thus, we will use C^{local} or a linear combination of both costs in our experiments.

[2] In the Khatri paper, the target unitary is denoted as U, while the ansatz is denoted V. I switched the notations to stay consistent with the rest of the examples and papers.

2.2 Optimizers for Variational Quantum Algorithms

Classical optimizers perform the task of finding a (hopefully global) minimum of some scalar function. An optimizer for a function f needs to find the vector $\vec{x} \in S$ with $S \subseteq R^n$ that has a minimal or maximal value. In this project, we will only use optimizers for minimization.

In each step, an optimizer is given a point in the parameter space of the cost function $\vec{\theta_i}$, and is tasked with finding a new guess for parameter values $\vec{\theta}_{i+1}$, which hopefully has a lower cost value.

The optimizer can either use the value of the cost function, its partial derivative (its gradient vector), or its second partial derivatives (its Hessian matrix).

Methods that use the second derivative are called Quasi-Newton methods (as Newton's method uses the second derivative). Methods that use the first derivative are called Gradient-Based methods. Methods that do not use the derivative at all are called Gradient-Free methods.

In this project, I focused on two gradient-free optimizers: the COBYLA optimizer and the Powell optimizer.

2.2.1 COBYLA

Constrained Optimization BY Linear Approximation (COBYLA) is a numerical optimization method for constrained problems that does not use the derivative. It was invented by Powell, and is based on linear approximations to the objective function and each constraint. More details can be found in [6].

2.2.2 Powell

Powell's method is also called the "conjugate direction" method. The optimization uses a set of directions in each step. It performs 1-dimensional bi-directional minimization along each vector in the direction set. Using the results of these minimizations, the direction set is updated (a new direction is added, and the most dominant vector is deleted). More details can be found in [7].

TODO: Check with Tal how much detail is required on the optimizers.

References

[1] M. Cerezo et al. "Variational quantum algorithms". In: *Nature Reviews Physics* 3.9 (Aug. 2021), pp. 625–644. DOI: 10.1038/s42254-021-00348-9. URL: https://doi.org/10.1038/s42254-021-00348-9.

[2] Jarrod R. McClean et al. "Barren plateaus in quantum neural network training landscapes". en. In: *Nature Communications* 9.1 (Nov. 2018). Number: 1 Publisher: Nature Publishing Group, p. 4812. ISSN: 2041-1723. DOI: 10.1038/s41467-018-07090-4. URL: https://www.nature.com/articles/s41467-018-07090-4 (visited on 07/14/2022).

[3] Alberto Peruzzo et al. "A variational eigenvalue solver on a photonic quantum processor". en. In: *Nature Communications* 5.1 (July 2014). Number: 1 Publisher: Nature Publishing Group, p. 4213. ISSN: 2041-1723. DOI: 10.1038/ncomms5213. URL: https://www.nature.com/articles/ncomms5213 (visited on 05/29/2022).

[4] Sumeet Khatri et al. "Quantum-assisted quantum compiling". en. In: *Quantum* 3 (May 2019). arXiv:1807.00800 [quant-ph], p. 140. ISSN: 2521-327X. DOI: 10.22331/q-2019-05-13-140. URL: http://arxiv.org/abs/1807.00800 (visited on 09/25/2022).

[5] Kunal Sharma et al. “Noise resilience of variational quantum compiling”. en. In: *New Journal of Physics* 22.4 (Apr. 2020). Publisher: IOP Publishing, p. 043006. ISSN: 1367-2630. DOI: 10.1088/1367-2630/ab784c. URL: https://doi.org/10.1088/1367-2630/ab784c (visited on 07/14/2022).

[6] M. J. D. Powell. “A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation”. In: *Advances in Optimization and Numerical Analysis*. Springer Netherlands, 1994, pp. 51–67. DOI: 10.1007/978-94-015-8330-5_4. URL: https://doi.org/10.1007/978-94-015-8330-5_4.

[7] M. J. D. Powell. “An efficient method for finding the minimum of a function of several variables without calculating derivatives”. In: *The Computer Journal* 7.2 (Feb. 1964), pp. 155–162. DOI: 10.1093/comjnl/7.2.155. URL: https://doi.org/10.1093/comjnl/7.2.155.