Pretraining of PQC with VGON and g-sim

Taihei Takahashi (Keio Univ・Yamamoto Lab)

Abstract

The Variational Quantum Algorithm (VQA) is a key application of Noisy Intermediate-Scale Quantum (NISQ) computing. However, a notorious phenomenon known as the Barren Plateau (BP) poses a significant obstacle for VQA, hindering its application to optimization problems. Recent research suggests that pretraining is a promising approach to mitigate BP in the VQA process. Pretraining requires efficient classical simulation of quantum circuits. A method known as g-sim, which utilizes the properties of Dynamical Lie Algebra (DLA) - deeply associated with BP, exists for this classical simulation. Pretraining using g-sim is particularly promising due to its relationship with DLA. Meanwhile, various methods have been proposed to circumvent the Barren Plateau, apart from pretraining. One promising method among these is VGON, an optimization algorithm that employs generative AI. While various methods have been proposed, the performance of their combinations has not been investigated until now. This presentation demonstrates the performance of multiple BP alleviation methods, including pretraining, VGON, and g-sim, in finding minimum energy states.

Introduction | BP, DLA, g-sim, Pretraining, VGON

Background①

Larocca's conjecture

$$\mathrm{Var}_\theta(\partial_\mu C(\theta)) \in \frac{1}{\mathrm{poly}(\dim(\mathfrak{g}))}.$$

The hypothesis represented by the above formula, proposed by Larocca et al.[1], is a relationship between the variance of the gradient during training $\mathrm{Var}_\theta(\partial \mu C(\theta))$ and the dimension of the system's Dynamical Lie Algebra (DLA) $\dim(\mathfrak{g})$. This hypothesis explains the well-known phenomenon "*Barren plateau*" (BP), where **the gradient disappears during training and training does not progress when the dimension of the system's DLA increases exponentially**.

Background②

g-sim algorithm

The g-sim algorithm, revived by Goh et al [2], is a **classical algorithm for simulating quantum many-body systems**. A characteristic of this algorithm is that the computational cost for simulation essentially depends not on the dimension of the system's Hilbert space, but **only on the dimension of the system's DLA**. The dependence of the spatial computational cost C of g-sim on the dimension of DLA, when the system's DLA is denoted as $\mathfrak{g}$, is expressed as follows

$$C \in \mathrm{poly}(\dim(\mathfrak{g})).$$

Background③

Classical simulability and avoidance of BP

According to Background1, for any quantum circuit that can avoid the BP, the dimension of its DLA $\mathfrak{g}$ should be $\mathrm{poly}(n)$ in relation to the number of quantum bits n. Meanwhile, based on Background 2, a quantum circuit whose dimension of DLA $\mathfrak{g}$ is $\mathrm{poly}(n)$ in relation to the number of quantum bits can be simulated classically by g-sim.

This implies that any quantum circuit that can avoid BP can be simulated classically. This was pointed out by Cerezo et al. [3]. Given this fact, in the study of quantum utility, it is more crucial to research methods to advance learning within the Barren Plateau, rather than aiming to avoid BP by reducing the dimension of the system's DLA.

Background④

Pretraining, Warm starting

"Warm starting" is a category of methods in quantum machine learning. The aim of these methods is to enable learning even in regions where the BP exists, by discovering good initial parameters or adopting biased Optimizers. Among these, *"Pretraining"* aims to avoid the BP by optimizing the parameters of the quantum circuit in advance using a classical computer. Goh et al. [1] demonstrated the effectiveness of this method for certain types of problems by **using g-sim for pretraining.**

Background⑤

VGON algorithm

The *VGON* algorithm, proposed by Zhang et al [4], is a **classical machine learning algorithm** that serves as a generative AI for good parameters of quantum circuits. It uses the measurement values of the observables in quantum circuits as the objective function and generates sets of parameters for the quantum circuits that minimize this function. A distinctive feature of VGON is its ability to generate multiple candidate parameters for minimization in a single learning process. This feature is advantageous in many-body problems where multiple minimum value candidates exist, making it resistant to getting trapped in local minima, a characteristic as being strong against BP.

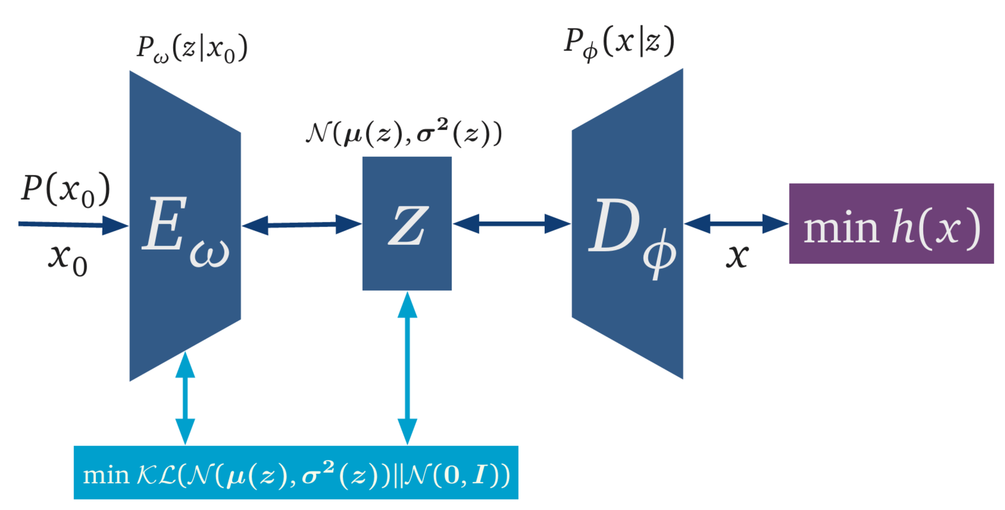

Diagram of VGON | VGON has a structure similar to VAE and consists of an encoder, a decoder, and a latent. Green parts are classical and a magenta part is quantum. x is PQC parameters generated by VGON (blue parts). x_0 is classically generated random parameters for inputs of VGON.

Weakness of VGON : A drawback of the VGON algorithm is that it depends on the gradient of each parameter of the Parametric Quantum Circuit (PQC) outputs. This poses a challenging issue for quantum computers in the NISQ era, as the accuracy of the gradient directly affects the accuracy of the output.

The reasons of BP alleviation methods combinations

We are going to verify the effects of the multiple BP alleviation methods combinations. In this presentation, we combine pretraining using *g-sim* and *VGON*. That is, we used g-sim simulator to simulate an objective function of VGON. The reasons for this combination are twofold:

Firstly, as mentioned earlier, VGON has a flaw that requires accuracy in the gradient. Algorithms that demand accuracy in the gradient like this are not noise-aware and are not suitable for execution on noisy quantum computers. On the other hand, in simulations using g-sim, exact expected values can be obtained, making efficient learning of VGON possible. Thus, powerful algorithms that are not noise-aware can also be used in pretraining.

Secondly, g-sim can potentially simulate efficiently on classical computers up to the limit where quantum computers do not have BP, i.e., where is the border of quantum supremacy. Also, because it is not affected by spatial constraints, it has the advantage of being able to flexibly assemble classical imitation circuits compared to other efficient classical simulation algorithms such as tensornetwork.

Method

To evaluate the effectiveness of the combination of BP alleviation methods, we first tested them against the *Noise-Induced Barren Plateau*. The problem used was a very simple weighted MAXCUT problem, and we compared learning **without pretraining and with pretraining using stochastic gradient descent (SGD)**.

MAXCUT Problem

The MAXCUT problem involves dividing nodes into two groups to maximize the sum of the weights of the edges between nodes belonging to different groups. To solve this problem using a quantum computer, we search for the minimum energy state of the Hamiltonian H given below, where w_{ij} represents the weight between nodes i and j:

$$H = \sum_{i<j} w_{ij} Z_i Z_j.$$

The problem involves 5 nodes, which are fully connected as shown in Fig 1. The weights between nodes are randomly selected, with specific values given to two decimal places in Fig 1.

Noisy Quantum Circuit and Noise-Induced Barren Plateau

To solve this problem, we executed the circuit shown in Figure 2 as the objective function for SGD, searching for the parameter θ that produces the minimum energy state. This circuit is noisy, with T1 time of 131.51 microseconds, T2 time of 95.75 microseconds, one qubit gate application time of 68 nanoseconds, Pauli gate error rate of 3.326e-4, and readout error rate of 1.720e-2. These values are sampled from ibm_fez device and it was calibrated 30 minutes ago. It is known that noisy quantum circuits have barren plateau even if its circuit is very simple, i.e. the time required for SGD to converge increases due to the uncertainty in gradients caused by noise. This phenomenon is called as *Noise-Induced Barren Plateau*.

Pretraining and SGD

To evaluate the effect of pretraining, we compared SGD with random initial values and with initial parameters obtained by pretraining the circuit parameters using VGON and gsim. The results of this pretraining are shown in Fig 3. The results of SGD are shown in Fig 4. Their learning rate is 1.0e-7.

Results

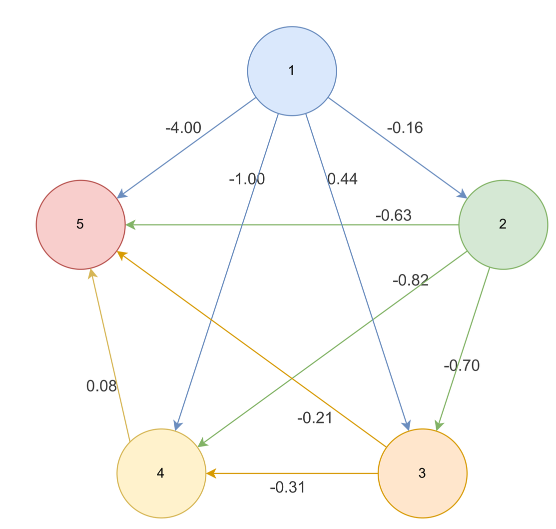

Figure1 | Weighted MAXCUT problem graph. 5 nodes 20 edges. Randomry generated weights are shown on the edge.

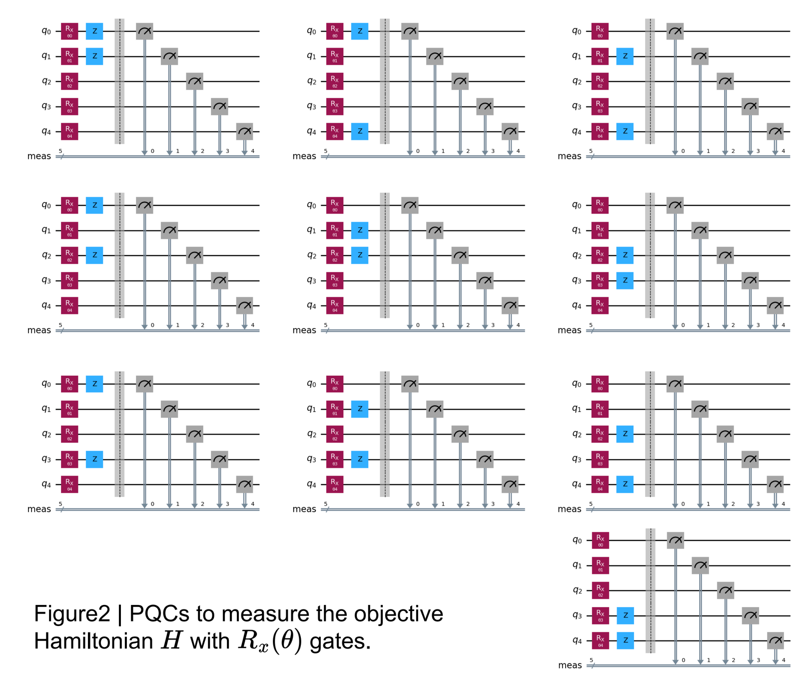

Figure2 | PQCs to measure the objective Hamiltonian H with $R_x(\theta)$ gates.

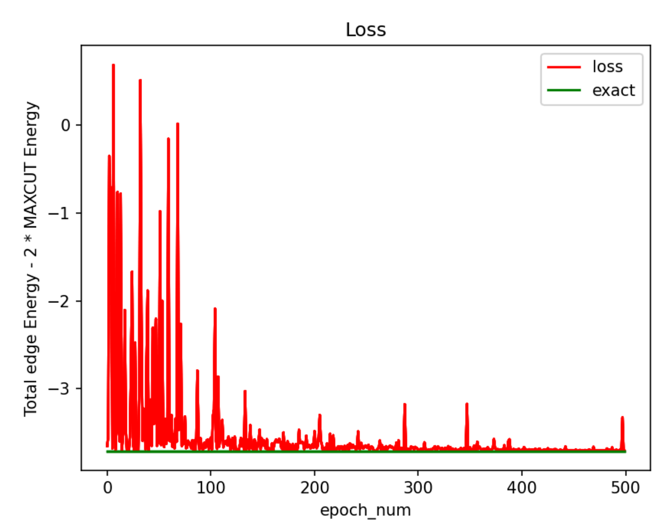

Figure3 | Pretraining loss transition. This pretraining is conducted by VGON with gsim simulator. VGON shows exponential convergence to the global minimum, -3.707....

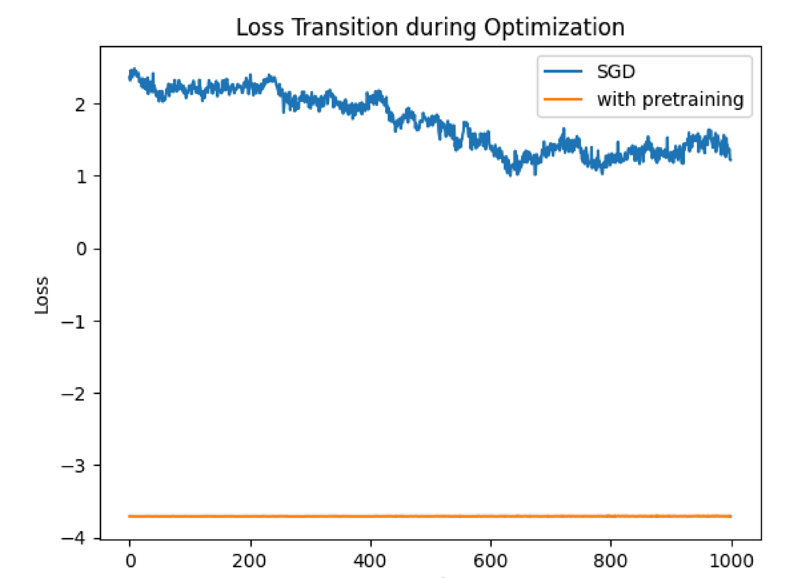

Figure4 | The progress of learning of SGD on Noisy Quantum Circuits. Blue line (SGD) is without pretraining and yellow line is with pretraining. The minimum values of (with pretraining) is placed at 803th iteration and the value is -3.709...

Pretraining improves convergence ability in noisy circuit

discussions and future work

- Regarding Fig. 3, it can be seen that the loss function during pretraining converges exponentially to its minimum. However, for problems with such simple landscapes, efficient exploration can be also achieved with SGD, so there is no advantage in using VGON for this problem. Problems where VGON is advantageous are those where the loss function of the problem during pretraining has many local minima, and it is unclear which one will become the true minimum after adding gates post-pretraining. Investigating such complex cases is a future work.
- Fig. 4 shows that the process of SGD with pretraining explores near the minimum value. This is in contrast to the case without pretraining, where the effect of the Noise-Induced Barren Plateau causes wandering at large values. Therefore, Based on Fig. 4, pretraining has the ability to avoid the Noise-Induced Barren Plateau. Future work includes investigating the effectiveness of pretraining in Barren Plateaus other than Noise-Induced ones.
- Theoretical future work includes establishing the theoretical advantages of using VGON. VGON, which can output diverse candidate points compared to regular SGD, is considered superior as a pretraining learning method, but currently lacks theoretical justification.
- Other future work includes exploring noise-aware optimization methods. When continuing learning with a quantum computer after pretraining, there are options for how to perform optimization, and algorithms such as Bayesian optimization or imaginary time evolution, which are not noise-aware SGD, will be considered.

Refferences

[1] Diagnosing Barren Plateaus with Tools from Quantum Optimal Control (Martin Larocca et.al.) [Quantum 6, 824 (2022)]
[2] Lie-algebraic classical simulations for variational quantum computing (Matthew L. Goh et.al.) [arXiv:2308.01432 (2023)]
[3] Does provable absence of barren plateaus imply classical simulability Or, why we need to rethink variational quantum computing (M. Cerezo et al.) [arXiv:2312.09121 (2023)]
[4] Variational Optimization for Quantum Problems using Deep Generative Networks (Lingxia Zhang et al.) [arXiv:2404.18041 (2024)]

24' 台湾