Application of Variational Quantum Eigensolver to Basic Transverse Ising Model

(Dated: February 15, 2023)

Hi, everyone

Usage:

Structure:

I. INTRODUCTION

A. Variation Quantum Algorithm (VQA)

From the early stage of studying quantum computer, that by its very quantum nature are promising for a new brightly age of computation coming along with three crucial keys of quantum theory: quantum probabilistic, superposition and entanglement. The distinct properties of quantum computer (QC) from classical computer make it be more unique in application, not only be a upgrade computational speed version of traditional computer. Currently, when the quantum hardware is rather limited on operational stats and noise resilience, despite of the ambiguity of realization of such advantages, Variational Quantum Algorithms (VQAs) are thought to be the best at outperforming conventional computers, which are able to implement well on near-term quantum devices known as the so-called Noisy Intermediate-Scale Quantum (NISQ) computers.

Nowadays, when the quantum revolution is being on the track, big tech companies like IBM, Google, D-Wave, etc. compete to build their own quantum computers to yield the quantum supremacy where the first-ever experimental demonstration achieved by the Google AI Quantum team [1], but that is still far from what we expect the quantum computer could do, while coherence time, connectivity of qubit and qubit number limitations keep us from being able to successfully run a long depth circuit and produce a significant result on the present noisy device, in addition to quantum gate implement issues. Those restrictions would yet be challenging to hardware scientists in the near future, posing obstacles that computer scientists would need to weigh against trainability, precision, and efficiency, nevertheless, these NISQ devices are capable of exploitation. Standing out among quantum algorithms envisioned to beat classical computer which are mostly designed for the fault-tolerant quantum computer, VQAs turn up regarding as a appropriate candidate compatible with the current defective quantum device to address these constraints. The fist two prominent applications of VQAs come up are Variation Quantum Eigensolver (VQE) and Quantum Approximate Optimization Algorithm (QAOA), these inherit the core scheme of a VQA is to utilize the hybrid routine, leverage the versatility and computational power of classical computer to handle the computation processing, and use quantum device to execute the quantum circuit. Hence, at this paper scope, we will investigate the standard procedure and properties of different methods implemented in a VQE application.

B. Variational Quantum Eigensolver (VQE)

VQE was initially published by Peruzzo et al. [2] and get into more details by Ref. [3], which is proposed as an efficiently alternative way to compute the ground state energy of a quantum chemistry problem i.e. ground-state He-H^+ molecular energy rather than Quantum Phase Estimation (QPE) demanding impractically huge numbers of quantum gates, consequently, VQE becomes a viable strategy highly well flexible with currently potential quantum resources, that can outpace conventional computers. Following the success of running on the photonic quantum processor combining with traditional device, many research works are favored to develop various types of VQA reviewed in Ref. [4, 5]. In essence, what makes VQE (generally VQAs) engrossing that is the trial state analysis step to choosing an ansatz adaptive to particular quantum device architecture and/or problems, subsequently, the suitable ansatz is executed congruently on the NISQ device, the remainders are then handled by the classical computer to generate remarkable results with the aid of error mitigation techniques. Instead of diagonalizing a matrix representation of Hamiltonian $\hat{H}$ to find out eigenvalues, that exponentially scale up the matrix size and calculation overhead in traditional computing as the problem grows up larger, VQE just focus on catching up the ground state energy E_g basically based on the variational principle

$$E_g \leq E\left[\Psi(\boldsymbol{\theta})\right] = \frac{\langle\Psi(\boldsymbol{\theta})|\hat{H}|\Psi(\boldsymbol{\theta})\rangle}{\langle\Psi(\boldsymbol{\theta})|\Psi(\boldsymbol{\theta})\rangle}. \tag{1}$$

The arbitrary quantum state $|\Psi(\boldsymbol{\theta})\rangle$ is a trial solution, the so-called *ansatz* parametrized by a unitary operator $\hat{U}(\boldsymbol{\theta})$, so that when the parameter $\boldsymbol{\theta}$ varies, the ansatz $|\Psi(\boldsymbol{\theta})\rangle$ is readily capable to span on the relevant quantum space where the ground state located. By modifying $\boldsymbol{\theta}$ value systematically, the VQE task is to travel around spanned space to extract the ground state minimizing the energy function $E\left[\Psi(\boldsymbol{\theta})\right]$ or in general VQA, the objective function described by Lagrangian $\mathcal{L}\left(\boldsymbol{\theta}, \langle\hat{H}\rangle_{\hat{U}(\boldsymbol{\theta})}\right)$

$$\min_{\boldsymbol{\theta}} \mathcal{L}\left(\boldsymbol{\theta}, \langle\hat{\mathcal{H}}\rangle_{\hat{\mathcal{U}}(\boldsymbol{\theta})}\right). \tag{2}$$

In fact, when the accuracy is not strictly rigorous, the minimum of the energy function is not necessarily re-

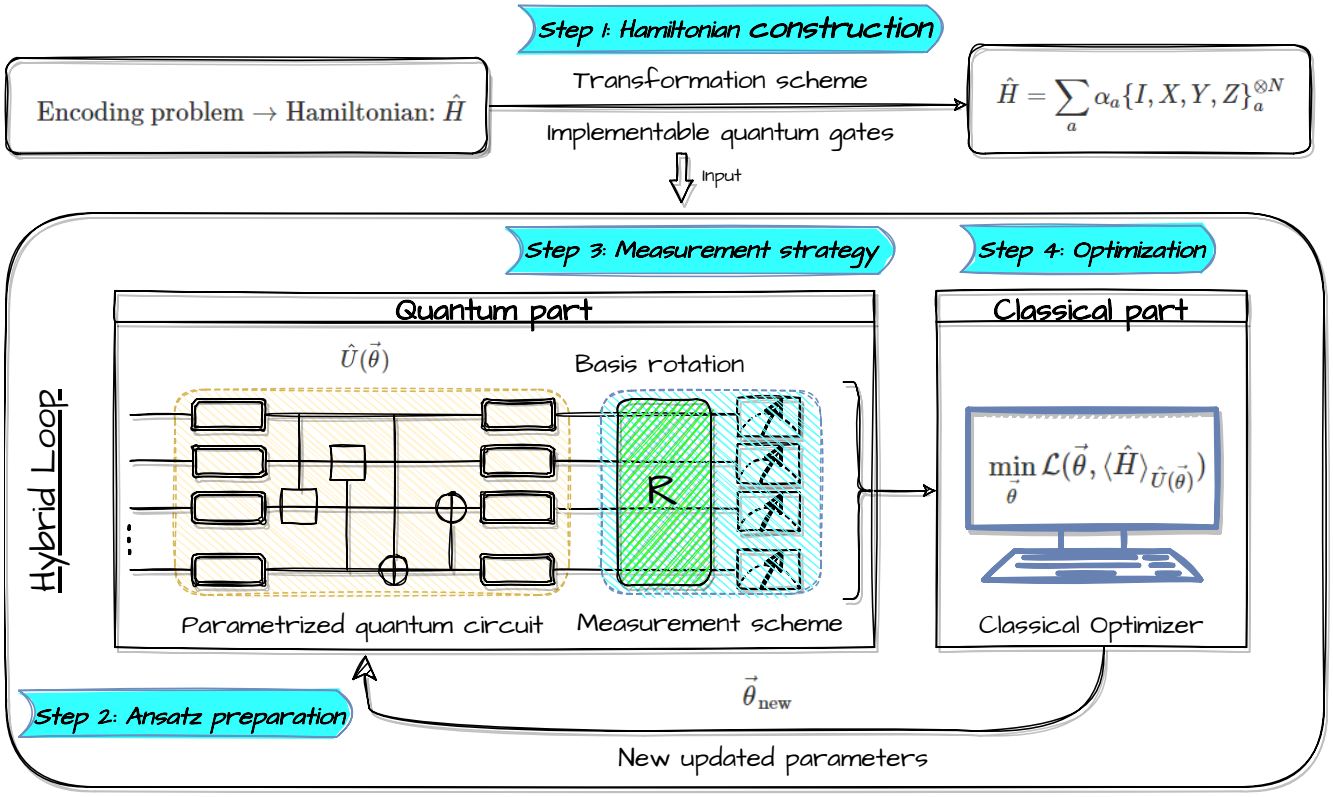


FIG. 1. VQE Architecture

quired a exactly ground state solution, due to the error in energy being in second order of the quantum state error. Formally, the classical approach for VQE insists an analytical expression of the lost function $\mathcal{L}\left(\boldsymbol{\theta}, \langle \hat{\mathcal{H}} \rangle_{\hat{\mathcal{U}}(\boldsymbol{\theta})}\right)$, in particularly, wave function expression $|\Psi(\theta)\rangle$ that becomes eventually unachievable going along with more and more problem complexities. Nevertheless, this weakness is literally the quantum advantage we have in VQE quantum computing approach, quantum computers are adopted to simulate quantum states, state expectation outcomes are subsequently measured to pipeline to calculate the lost function classically. And partly because of lost function precision greater quadratically than the state function, indeed, we need two states to sandwich operators, VQEs are hereby friendly with noisy devices, which can work out meaning results even though the states are disturbed by noises. These motivations are robust to let us dive deeper into VQE operation, ideally noise free assumption in this paper scope though, the noisy one will be address in future study.

The structure of a VQE algorithm is depicted schematically as Fig. 1 within four-basic steps: *Hamiltonian construction, Ansatz preparation, Measurement strategy* and *Optimization*.

Hamiltonian construction: Starting from a given abstract problem, eg. molecule ground state energy, shortest path, transshipment problems, an initially mathematical form called Hamiltonian $\hat{H}$ are modeled. The mapping step is needed to convert $\hat{H}$ into new form made out of operators that are able to execute on quantum computer.

Ansatz preparation: To pick out a good ansatz, an essential condition must be satisfied is that the space spanned by the ansatz contains the desired state extremize objective function. An obvious one is the generic ansatz spread all Hilbert space, a quantum circuit for this kind of ansatz is usually generated by control-$U_3(\boldsymbol{\theta})$ gates, so that all problems could be ultimately settled without ansatz concern. But not the end of the story,

Measurement

Optimization: Gradient descent, derivative free

C. Transverse Ising model (Ising quantum version)

We consider the 1D Ising model that describes the nearest neighbour interactions of the spin projection along z axis and uniform external magnetic field along x axis (in priciple, perpendicular to the z axis).

$$H = J \sum_{n=1}^{N-1} \sigma_{n-1}^z \sigma_n^z + h \sum_{n=0}^{N-1} \sigma_n^x \tag{3}$$

defined by the two spin exchange interaction factor J and representative strength of the external field h.

- Quantum effect: Cannot observe simultaneously all terms
- Physical phase: Ordered, Disordered, Gapless
- Construct cost function from Ising model

II. ANSATZ CONSTRUCTION

As mentioned above, we need the $U(\boldsymbol{\theta})$ operator can span all quantum Hilbert space so that we can use to solve all issues. Indeed, we do not, in practice though, an general $U(\theta)$ operator is hard to implement experimentally, because of noisy and weighty composite $C-U_3$ gate, and actually does not acquire overall good operation even in ideal simulator. Hence, the ansatz implementation for a particular problem's purpose is an serious studying process of VQE working roadmap. Additional to the compatibility with the current quantum device, basing on symmetries and properties of our problem that could be leveraged to shrink the parameter space, which are the important criteria to adopt an ansatz.

A. Symmetry of Transverse Ising model

At the scope of this article, we consider three properties of TIM so that we take its suggesting information into account to decrease the size of ansatz

- *Real representation.* Using the eigenstates of the σ_Z (Pauli-z) operator as elementary binary computational basis, in terms of which, σ^z, σ^x are real matrices, due to that, we are able to represent the TIM Hamiltonian in the real form, the eigenstates of TIM Hamiltonian can thus choose to be real for conventional. Namely, considering $|\Psi\rangle = \sum_n C_n|n\rangle$ is a eigenstate of H_{TIM} expanding in the computational basis $|n\rangle$. The real form of H_{TIM} means the Hermitian real element matrix, that induces coefficients satisfied $C_n^* C_m = C_n C_m^* \ \forall m, n \in [0, 2^N - 1]$, then, in generality

$$C_n = r_n e^{i(c+n_n\pi)}, \quad c \text{ is a constant}, \tag{4}$$

  or the angles in the complex plane of coefficients C_n differ from every other by an factor $k\pi$, where k and n_n are integer. Then, the complex angle can be shifted to the real coefficient $C_n^* = C_n$ by according to the quantum global phase principle.

- *Local interaction.* The first term $H_{\text{TIM}}^{\text{LI}}$ in TIM Hamiltonian describing the kind of neighbouring spin interaction along the z-axis

$$H_{\text{TIM}}^{\text{LI}} = \sum_{n=1}^{N-1} \sigma_{n-1}^z \sigma_{n+1}^z. \tag{5}$$

  The interaction of two local spin formulating by this term causes a sort of entanglement structure between every two local spin of the ground state energy. nevertheless, in the case of the order phase, when the external magnetic field is dominant, this interaction can be broken down, each spin will be free interacting and align in the same direction of the magnetic field.

- *Total spin-flip symmetry* . The most featured symmetry of Ising model, can be referred to the classical counterpart, Time-reversal symmetry (generally, $\mathcal{Z}_2$ symmetry). Under the total spin-flip transformation in the z direction $(\sigma^x)^{\otimes N}$

$$\left[(\sigma^x)^{\otimes N}, H_{\text{TIM}}\right] = 0, \tag{6}$$

  which implies the TIM Hamiltonian remains unchanged. This one leads to the eigenstate $|\Psi\rangle$ and $(\sigma^x)^{\otimes N}|\Psi\rangle = |\tilde{\Psi}\rangle$ have the same energy, the relation between them thus put in two cases

$$\langle\Psi|\tilde{\Psi}\rangle = \begin{cases} 0 & \text{Degeneracy} \\ \pm 1 & \text{Non-degeneracy} \end{cases}, \tag{7}$$

  where, degenerate case $g = 0$ is trivial, at non-degenerate $g > 0$ case, the expanded coefficients in the z-direction basis representation having a structure

$$C_n = \pm C_{2^N-1-n}. \tag{8}$$

B. Ansatz selection

For the chosen ansatz, the conventional real coefficients of eigenstates tell us that is enough to span in a real quantum parameter space for finding the ground state energy, and the linear entanglement mapping coming from information of the local interaction term in Hamiltonian. Moreover, to accommodate the device constraints, in speculative way, basing on these conditions, the available common use RealAmplitude ansatz from Qiskit open source is our good candidate for the implementation.

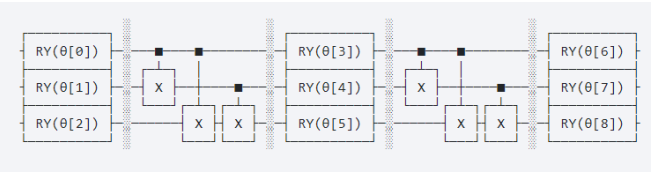


FIG. 2. A RealAmplitudes circuit with 2 repetitions on 3 qubits with 'full' entanglement

The additional reason for choosing such a Hardware-efficient ansatz architecture [6] comes from the efficient implementation of quantum optimization to attain our

full quantum algorithm purpose, which will be addressed in the next subsection III B.

For traveling around full quantum Hilbert space we need to tune $2^{N+1}-2$ degrees of freedom, $+1$ inside of the exponent indicates the complex number condition, and -2 turns up from normalization and global phase conditions. After bringing TIM's properties we investigate from section II A to dissect, *Real representation* Eq. (4) helps us eliminate $+1$ within the exponent and *Total spin-flip symmetry* Eq. (8) reduces a half as much numbers of parameters that we have to search the ground state, finally

$$2^{N+1}-2 \to 2^{N-1}-1. \tag{9}$$

The factor *repetition (reps)* of our RealAmplitude ansatz is used to modify number of layer L, which changes directly ansatz number of parameter p according to $p = N(L+1)$. Following Eq. (9), we roughly evaluate a necessary number of layer $L \geq \frac{2^{N-1}-1}{N} - 1$, note that this is not a rigorous condition because ansatz's entanglement mapping also contributes, subdominant though, henceforth, it is a good estimation for surface analysis. Yet, we can realize that, not only TIM, but Eq. (9) also holds for all symmetrically real Hamiltonian being symmetric under $(\sigma^x)^{\otimes N}$ operator. Actually, TIM has some peculiarly hidden properties that can put down more value on L, and through conducting experimental survey, we are able to choose a reliable value for L, even being independent of the number of qubit N, see IV B for more detailed results.

III. OPTIMIZATION

The next subroutine in VQE work map, which is a purely technical perspective we investigate to study the performance of different optimization methods, provides us a larger view of the way to reaching the faster and more precise convergence of our target value. In TIM, we try to minimize the cost function defined in Eq. **??** using common methods of optimization categorized into two class operations.

A. Classical Operation

Classical operation means regardless of the analysis quantum structure of cost function, in particular ansatz structure, we are able to find the optimal parameters minimizing the objective function. Normally, in sense of operation, we can do the classical optimization process independently from each other VQE parts and even with your problem-specific concerns, and appears that you can develop your optimization algorithm for general variational issues.

Constrained Optimization BY Linear Approximation (COBYLA). One of the most powerful derivative free methods, which is favoured by many users in lately decades, COBYLA makes use of linear interpolation of the objective function at each iteration by a unique linear polynomial function at the vertices for finding a optimal vector parameter within the trust region, then feeding optimal point evaluated to the objective function to get the value improving the next iteration of approximation [7–10]. In VQE, COBYLA's linear interpolation at each step is a classical subroutine running on CPU using only one the objective value from QPU evaluation running. And so because of ignored derivative information, it would beneficially avoid several problems in analysis optimizations, especially Barren plateaus landscape, in return less accuracy for more parameters ($p > 9$) [10]. For a birdeye view, new update COBYLA versions like UOBYQA, NEWUOA, and BOBYQA count the objective function's curvature information to increase convergence [11–13].

Finite Different (FD). Let's dawn on one of the primary first-order derivative-based approaches, where the parameter $\boldsymbol{\theta} \in \mathbb{R}^p$ is updated at the iteration k-th by

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \eta_k \nabla f(\boldsymbol{\theta}_k), \tag{10}$$

which we usually name the Gradient Descent (GD) method. Finite Different is a numerical technique to calculate the gradient vector $\nabla f(\boldsymbol{\theta}_k)$ without analytical function's texture, which is the basic one used in the gradient descent method. We use the well-known central difference formula

$$\nabla f(\boldsymbol{\theta}_k)_i \simeq \frac{f(\boldsymbol{\theta}_k + \epsilon\vec{i}) - f(\boldsymbol{\theta}_k - \epsilon\vec{i})}{2\epsilon}, \tag{11}$$

where $\vec{i} \in \mathbb{R}^p$ is i-th the unit vector and ϵ is a infinitesimal change, with the error proportional to $\mathcal{O}(\epsilon^2)$. The smaller ϵ we use, the more exact result we get. However, due to the restricted accuracy of a classical computer, we are unable to achieve a value of *epsilon* that is too small, which even becomes greater when implemented on a quantum device, where at least the sample error enters the picture.

Simultaneous Perturbation Stochastic Approximation (SPSA). To surmount some of obstacles emerge from gradient descent operation on near-term device, the idea of a perturbing stochastic approximation method is a good choice. SPSA generates an unbiased estimator $\tilde{f}(\boldsymbol{\theta}_k)$ of gradient by simultaneously radomly perturbating gradient direction of parameters [14], we replace the ordinary gradient vector $\nabla f(\boldsymbol{\theta}_k)$ by

$$\nabla f(\boldsymbol{\theta}_k) \to \nabla \tilde{f}(\boldsymbol{\theta}_k) = \frac{f(\boldsymbol{\theta} + s_k\vec{\Delta}_k) - f(\boldsymbol{\theta}_k - s_k\vec{\Delta}_k)}{2s_k}\vec{\Delta}_k, \tag{12}$$

where $\vec{\Delta}_k$ is a random perturbed vector sampled from the zero-mean distribution, usually, the Bernoulli distribution is used. We can see that all parameters are simultaneously shifted by a random amount ($\pm s_k$), our

computation therefore only requires two objective function evaluations per iteration. Whereas standard gradient computation time is scaled along to number of parameters, the SPSA optimizer is, however, independent, that saves a lot of time as we work on higher parameter regime. Besides that, noise from quantum circuit executions computing the objective value be regarded as a stochastic perturbation part absorbed to the SPSA procedure. These advantages promote SPSA and its other version to be efficient techniques in the NISQ era.

B. Quantum Operation

As opposed to classical one, the quantum operation require the quantum information extracted from the structure of the trial wave function. This sort of behaviour inevitably entails the optimization process into to whole problem analysis, especially related close to Ansatz Construction.

Parameter-shift rules (PSR). Inspiring from the classical shift rules in the exact derivative computation of some special function, we are successfully able to figure out the analytical value of objective function's derivatives using quantum devices. For any kinds of quantum gates have the form

$$\hat{\mathcal{G}}(\theta) = e^{-i\theta\hat{G}} \tag{13}$$

generated by the Hermitian operator $\hat{G}$. Supposed that we have an expectation value of $\hat{H}$ is our objective function $f(\theta) = \langle\psi(\theta)|\hat{H}|\psi(\theta)\rangle$ parametrized by θ. The ansatz wave function $|\psi(\theta)\rangle = \hat{U}(\theta)|\psi\rangle_I$ is made up of $\hat{U}(\theta) = \hat{A}\hat{\mathcal{G}}(\theta)\hat{B}$, with $\hat{A}$, $\hat{B}$ are arbitrarily other operators. As a result, the partial derivative of the $f(\theta)$ with respect to θ is thus obtained via the parameter-shift rules [15]

$$\partial_\theta f(\theta) = s\left[f(\theta + \frac{\pi}{4s}) - f(\theta - \frac{\pi}{4s})\right]. \tag{14}$$

The partial derivative $\partial_\theta = \frac{\partial}{\partial\theta}$ implies we can generalize to set of multiple parameters $\{\theta_i\}$. The evaluation of $f(\theta \pm \frac{\pi}{4s})$ can be easy to run on quantum computer, then the exact amplitude of vector gradient is derived. Within the aim of this paper, the given parametric ansatz is made out of Pauli rotations (r_x, r_y, r_z), $s = \frac{1}{2}$ is selected accordingly. The state of the art formalism for more generic PQCs are also able to be derived [16, 17].

Quantum Natural Gradient Descent (QNG). How to improve the convergence? The global fixed learning rate η_k is apparently not a good choice, one does not be well sensitive to the model information with respect to parameter changes. Initially, many attempts try to tune the learning rate η_k such as learning rate schedulers that vary η_k after an iteration, laterly, adaptive methods that count previous iteration values, or capturing the curvature of the objective function using the diagonal approximation of the Hessian, where each η_k is the inverse Hessian's diagonal element instead of being equally fixed. Those methods educe that a we can embed further information rather than only just gradient vector, to make the optimal step size for our variational quantum algorithms. Likewise the classical counterpart using Fisher Information matrix (FIM), Quantum Natural Gradient Descent invokes the quantum geometry of the wave function, which can make us get the optimal spot faster. Namely, we transform to the Riemann parameter space using metric $g \in \mathbb{R}^{p\times p}$ of Projected Hilbert space $\mathcal{PH}$ instead of current Euclidean space $g = \mathbb{1}_{p\times p}$. The update Eq. (10) turns to

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \eta_k g^+(\boldsymbol{\theta}_k)\nabla f(\boldsymbol{\theta}_k), \tag{15}$$

$g^+(\boldsymbol{\theta}_k)$ means it is a pseudo-inverse local metric. Riemann parameter space metric g is general defined

$$ds^2 = g_{ij}d\theta_i d\theta_j. \tag{16}$$

The Hilbert space $\mathcal{H}$ of "bare" quantum states reduces to the $\mathcal{PH} = \mathcal{H}/U(1)$ space as we ignore the local phase $U(1)$ of quantum states [18], the quantum distance in $\mathcal{PH}$ is then based on to calculate the Riemann metric g

$$\begin{aligned} ds^2 &= 1 - |\langle\psi_\theta, \psi_{\theta+d\theta}\rangle|^2 \\ &= 1 - \left|1 + \frac{1}{2}\langle\psi_\theta|\partial_i\partial_j\psi_\theta\rangle d\theta_i d\theta_j + \langle\psi_\theta|\partial_i\psi_\theta\rangle d\theta_i\right|^2 \\ &= \mathrm{Re}[G_{ij}]d\theta_i d\theta_j. \end{aligned} \tag{17}$$

We expand $|\psi_{\theta+d\theta}\rangle = |\psi_\theta\rangle + |\partial_i\psi_\theta\rangle d\theta_i + \frac{1}{2}|\partial_i\partial_j\psi_\theta\rangle d\theta_i d\theta_j$ upto second order in $d\theta$, where $\partial_i = \frac{\partial}{\partial\theta_i}$. The Quantum Geometry tensor (QGT) $G_{ij} \in \mathbb{R}^{p\times p}$ has two parts, the anti-symmetric imaginary part $\sigma_{ij} = -\sigma_{ji}$ related to gauge field of $U(1)$ is eventually vanished, the only contribution comes from the real symmetric part $g_{ij} \equiv g_{ij}(\boldsymbol{\theta})$ reflects the quantum distance in $\mathcal{PH}$ space, where its formula is

$$g_{ij}(\boldsymbol{\theta}) = \mathrm{Re}\left[\langle\partial_i\psi_\theta|\partial_j\psi_\theta\rangle - \langle\partial_i\psi_\theta|\psi_\theta\rangle\langle\psi_\theta|\partial_j\psi_\theta\rangle\right]. \tag{18}$$

The Fubini-Study metric tensor $g_{ij}(\boldsymbol{\theta})$ is a quantum analogue of the Fisher information matrix (QFIM). Given that it is proportional to p^2, the complicated computing and computational cost of $g(\boldsymbol{\theta})$ is high and incompatible with the short-term quantum device. To solve this problem, an approximation strategy is required.

- *QNG - Block Diagonal Approximation (QNG-BDA).* Moreover than Diagonal Approximation, where we just count p elements in diagonal line, the QNG-BDA employing the Parametric Family circuit is conveniently deploy entirely computation of approximated metric tensor on the quantum computer. The parametric unitary operator $\hat{U}(\boldsymbol{\theta})$ acts on the initial state $|\psi\rangle_I$ entailing L layers

  $$\hat{U}(\boldsymbol{\theta}) = S_L P_L(\theta^L)\ldots S_1 P_1(\theta^1), \tag{19}$$

  where:

- S_l are the static parts, which usually denote to the entanglement layer.
- $P_l(\theta^l)$ are the parametric parts, which can be decomposed to single qubit gates $P_l(\theta^l) = \bigotimes_{i=1}^{N} R_i(\theta_i^l)$.
- The single qubit gate $R_i(\theta_i^l) = \exp\{[i\theta_i^l K_i]\}$ is constructed from Hermitian generator K_i with the parameter $\theta_i^l \in \theta^l = \{\theta_1^l, \ldots, \theta_N^l\}$.

Such type of parametrized circuit whose nice properties we can prospect to yield a block diagonal form of QGT running completely on quantum processor, each block corresponding to each layered vector parameter $\theta^l \in \boldsymbol{\theta} = \theta^1 \oplus \ldots \oplus \theta^L$. We denote

$$\hat{U}_n^m = S_m P_m(\theta^m) \ldots S_n P_n(\theta^n), \tag{20}$$

then $\hat{U}(\boldsymbol{\theta}) \equiv \hat{U}_1^L = \hat{U}_{l+1}^L S_l P_l(\theta^l) \hat{U}_1^{l-1}$. The partial derivative state can be written in form

$$\begin{aligned} |\partial_i \psi_\theta\rangle &= \partial_i \hat{U}(\boldsymbol{\theta})|\psi\rangle_I && (21)\\ &= \hat{U}_{l+1}^L S_l \partial_i P_l(\theta^l) \hat{U}_1^{l-1}|\psi\rangle_I && (22)\\ &= \hat{U}_l^L (iK_i)\, \hat{U}_1^{l-1}|\psi\rangle_I && (23)\\ &= \hat{U}_l^L (iK_i)\, |\psi_\theta^{(l-1)}\rangle. && (24) \end{aligned}$$

Note that $[K_i, K_j] = 0$ in each layer, the unitarity of $\hat{U}_n^m$ and overlap of partial derivative states tell us the block matrix element

$$\begin{aligned} g_{ij}^{(l)}(\boldsymbol{\theta}) =& \langle \psi_\theta^{(l-1)}|K_i K_j|\psi_\theta^{(l-1)}\rangle \\ &- \langle \psi_\theta^{(l-1)}|K_i|\psi_\theta^{(l-1)}\rangle \langle \psi_\theta^{(l-1)}|K_j|\psi_\theta^{(l-1)}\rangle \end{aligned} \tag{25}$$

of block-diagonal form of Fubini-Study metric [19]

$$g_{ij}(\boldsymbol{\theta}) = \begin{array}{c} \\ \theta^1 \\ \theta^2 \\ \vdots \\ \theta^L \end{array} \begin{array}{c} \begin{array}{cccc} \theta^1 & \theta^2 & \ldots & \theta^L \end{array} \\ \begin{pmatrix} g^{(1)} & 0 & \ldots & 0 \\ 0 & g^{(2)} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & g^{(L)} \end{pmatrix} \end{array}. \tag{20}$$

QNG-BDA approximation involves L quantum evaluations and works well on models with weak correlation.

Quantum Natural-Simultaneous Perturbation Stochastic Approximation (QN-SPSA). Inheriting the SPSA idea to calculate the gradient, it is capable to put up to generate the Hessian matrix with fewer evaluations, the so-call 2-SPSA. Normally, the Hessian or FIM matrix consumes $\mathcal{O}(p^2)$ quantum expectation, QNG-BDA enables us turn to relying on number of layer $\mathcal{O}(L)$, saving us lots of computational sources. Even further, QN-SPSA just only execute four quantum running to obtain full of the second order derivative matrix. Let's consider the Hessian of the Fubini-Study metric

$$H_{ij}(\boldsymbol{\theta}) \equiv g_{ij}(\boldsymbol{\theta}) = -\frac{1}{2}\partial_i \partial_j |\langle \psi_\theta, \psi_{\tilde{\theta}}\rangle|^2 \Big|_{\tilde{\theta}=\theta}, \tag{26}$$

which is just Hessian form of Eq. (18) so that we can deploy the 2-SPSA method to generate QN-SPSA matrix. We can see the equivalence

$$\begin{aligned} &-\frac{1}{2}\partial_i \partial_j |\langle \psi_\theta, \psi_{\tilde{\theta}}\rangle|^2 \Big|_{\tilde{\theta}=\theta} = -\partial_i \mathrm{Re}\left\{\langle \psi_\theta, \psi_{\tilde{\theta}}\rangle \langle \psi_{\tilde{\theta}}, \partial_j \psi_\theta\rangle\right\}\big|_{\tilde{\theta}=\theta} \\ &= -\mathrm{Re}\left\{\langle \psi_\theta, \partial_i \partial_j \psi_\theta\rangle + \langle \partial_i \psi_\theta, \psi_\theta\rangle \langle \psi_\theta, \partial_j \psi_\theta\rangle\right\} \\ &= -\mathrm{Re}\left\{-\langle \partial_i \psi_\theta, \partial_j \psi_\theta\rangle + \langle \partial_i \psi_\theta, \psi_\theta\rangle \langle \psi_\theta, \partial_j \psi_\theta\rangle\right\}, \end{aligned} \tag{27}$$

which is exactly same as Eq. (18). Applying 2-SPSA approach to compute second order derivative of function $F(\boldsymbol{\theta}, \tilde{\theta}) = -\frac{1}{2}|\langle \psi_\theta, \psi_{\tilde{\theta}}\rangle|^2$ instead of our loss function $f(\boldsymbol{\theta})$ as being in Newton method. The core estimator in second order SPSA is perturbed by two random vector $\vec{\Delta}_k^1, \vec{\Delta}_k^2 \in \mathcal{U}^p\{-1, 1\}$ at step k-th

$$\begin{aligned} \triangle F = \frac{-1}{2} \Big[& F(\boldsymbol{\theta}_k + s_k \vec{\Delta}_k^1 + s_k \vec{\Delta}_k^2, \boldsymbol{\theta}_k) \\ & - F(\boldsymbol{\theta}_k + s_k \vec{\Delta}_k^1, \boldsymbol{\theta}_k) + F(\boldsymbol{\theta}_k - s_k \vec{\Delta}_k^1, \boldsymbol{\theta}_k) \\ & - F(\boldsymbol{\theta}_k - s_k \vec{\Delta}_k^1 + s_k \vec{\Delta}_k^2, \boldsymbol{\theta}_k) \Big], \end{aligned} \tag{28}$$

which is composed of four terms respective to four quantum expectations we run on quantum processor. Then, the Fubini-Study metric Eq. (26) is replaced by QN-SPSA metric at k-th iteration

$$g^k(\boldsymbol{\theta}) \to \bar{H}^k(\boldsymbol{\theta}) = \frac{\triangle F}{4 s_k^2} \left(\vec{\Delta}_k^1 \vec{\Delta}_k^{2\mathrm{T}} + \vec{\Delta}_k^2 \vec{\Delta}_k^{1\mathrm{T}} \right), \tag{29}$$

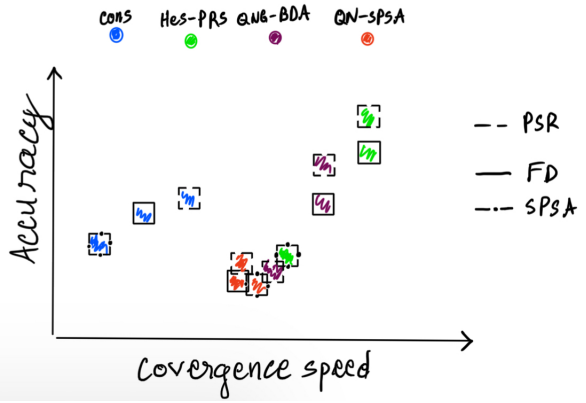
where $\bar{H}^k(\boldsymbol{\theta}) \in \mathbb{R}^{p \times p}$ and s_k is a small positive hyperparameter being able to be tuned. The metric is, however, still too much stochastic, some helpful techniques is invoked to resolve that hitch [20, 21], say, counting information from previous updates to smooth the estimator

$$\tilde{H}^k = \frac{k}{k+1} \tilde{H}^{k-1} + \frac{1}{k+1} \bar{H}^k, \tag{30}$$

eventually, to satisfy the positive semi-definite warranting the local convex analysis and invertibility condition of $\tilde{H}^k$, the below replacement is needed respectively

$$\tilde{H}^k \to \sqrt{\tilde{H}^k \tilde{H}^k} + \beta \mathbb{1}, \tag{31}$$

where the second term $\beta \mathbb{1} \in \mathbb{R}^{p \times p}$ corresponds to invertibility condition. The effect from geometry metric is suppressed as the positive regulator gets huge value $\beta \gg 0$, that is indeed standard gradient descent, and be more unstable, which means larger deviation in the average sample result, when $\beta \to 0$. That makes constant regulator β is thus a trade-off between Quantum Natural information and numerical instability.



IV. RESULTS AND DISCUSSION

A. Running results

Derivative-based method efficiency			
$\nabla f(\boldsymbol{\theta}_k)$ Evaluation / Adaptive η Evaluation	PRS ($2p$)	FD ($2p$)	SPSA (2)
Constant	3D	3D	$3D$
Hes-PRS ($3p^2$)	$1A$	$2A$	$4C$
QNG-BDA (L)	$2B$	$3B$	$4C$
QN-SPSA (4)	$4C$	$4C$	$4C$

TABLE I. Showing comparison among combinations of gradient and adaptive learning rate methods. Where p is the number of parameters, L is the number of PQC layer of the same rotation generator. Number ranking for accuracy, word ranking for convergence

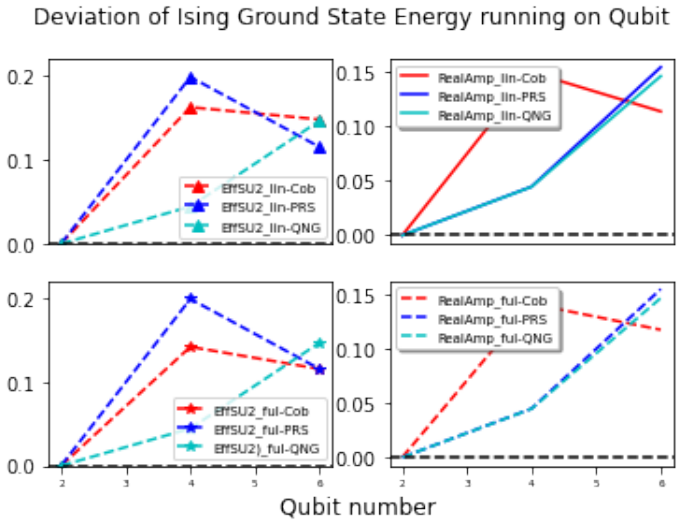


FIG. 3. Deviation comparison - Qubit running

B. Discussion

[1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, *et al.*, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).

[2] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'brien, A variational eigenvalue solver on a photonic quantum processor, Nature communications **5**, 1 (2014).

[3] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New Journal of Physics **18**, 023023 (2016).

[4] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, *et al.*, The variational quantum eigensolver: a review of methods and best practices, arXiv preprint arXiv:2111.05176 (2021).

[5] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, *et al.*, Noisy intermediate-scale quantum (nisq) algorithms, arXiv preprint arXiv:2101.08448 (2021).

[6] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature **549**, 242 (2017).

[7] M. J. D. Powell, Direct search algorithms for optimization calculations, Acta Numerica **7**, 287–336 (1998).

[8] M. J. Powell, A view of algorithms for optimization without derivatives, Mathematics Today-Bulletin of the Institute of Mathematics and its Applications **43**, 170 (2007).

[9] A. R. Conn, K. Scheinberg, and P. L. Toint, On the convergence of derivative-free methods for unconstrained optimization, Approximation theory and optimization: tributes to MJD Powell , 83 (1997).

[10] M. J. Powell, A direct search optimization method that models the objective and constraint functions by linear interpolation, in *Advances in optimization and numerical analysis* (Springer, 1994) pp. 51–67.

[11] M. J. Powell, Uobyqa: unconstrained optimization by quadratic approximation, Mathematical Programming

92, 555 (2002).
[12] M. J. Powell, The newuoa software for unconstrained optimization without derivatives, in *Large-scale nonlinear optimization* (Springer, 2006) pp. 255–297.
[13] M. J. Powell, The bobyqa algorithm for bound constrained optimization without derivatives, Cambridge NA Report NA2009/06, University of Cambridge, Cambridge **26** (2009).
[14] J. C. Spall, An overview of the simultaneous perturbation method for efficient optimization, Johns Hopkins apl technical digest **19**, 482 (1998).
[15] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, Physical Review A **99**, 10.1103/physreva.99.032331 (2019).
[16] L. Banchi and G. E. Crooks, Measuring Analytic Gradients of General Quantum Evolution with the Stochastic Parameter Shift Rule, Quantum **5**, 386 (2021).
[17] D. Wierichs, J. Izaac, C. Wang, and C. Y.-Y. Lin, General parameter-shift rules for quantum gradients, Quantum **6**, 677 (2022).
[18] R. Cheng, Quantum geometric tensor (fubini-study metric) in simple quantum system: A pedagogical introduction (2010).
[19] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum natural gradient, Quantum **4**, 269 (2020).
[20] J. Gacon, C. Zoufal, G. Carleo, and S. Woerner, Simultaneous perturbation stochastic approximation of the quantum fisher information, Quantum **5**, 567 (2021).
[21] A. Mari, T. R. Bromley, and N. Killoran, Estimating the gradient and higher-order derivatives on quantum hardware, Phys. Rev. A **103**, 012405 (2021).