

# Contents

<b>1 Variational Quantum Algorithms</b>	<b>2</b>
<i>Bavithra Govintha Rajah (under the supervision of Prof. Dr. Fabian Hassler)</i>	
1.1 Introduction . . . . .	2
1.2 Variational quantum algorithms . . . . .	3
1.2.1 Outline of the algorithm . . . . .	4
1.2.2 The variational paradigm and general quantum systems . . . . .	4
1.3 Variational ansatz construction . . . . .	6
1.3.1 Problem-inspired ansatz construction . . . . .	7
1.3.2 Hardware-specific ansatz construction . . . . .	7
1.4 Cost function . . . . .	7
1.5 Measurement scheme . . . . .	8
1.5.1 Measurement of Pauli strings . . . . .	9
1.5.2 Precision and Chebyshev's inequality . . . . .	9
1.5.3 Measurement scheme optimization . . . . .	9
1.6 Optimization routine . . . . .	10
1.6.1 Gradient based methods . . . . .	10
1.6.2 Gradient free approaches . . . . .	10
1.7 Applications . . . . .	10
1.8 Hands-on tutorial: Simulating $H_2$ molecule using VQE . . . . .	10
1.9 Outlook and conclusions . . . . .	11
App. 1.A The variation principle . . . . .	12
App. 1.B Pauli basis . . . . .	13
App. 1.C Minimizing the VQE cost function . . . . .	14
1.C.1 Gradient estimation . . . . .	14

# Chapter 1

# Variational Quantum Algorithms

**Bavithra Govintharajah**

Supervisor: Prof. Dr. Fabian Hassler

*don't forget the abstract*

## 1.1 Introduction

As quantum devices became available, it was evident that progress in both the algorithmic and the hardware side was going to require the realization of proof-of-principle implementations of quantum algorithms. The modest resource requirements and flexibility of VQE compared to other algorithms, positioned this approach as one of the first milestones in the pipeline of quantum computing experiments.

Experimental advancements in quantum computation can be measured with different figures of merit. One main requirement is that the number of available physical qubits must exceed a specific threshold to explore classically intractable problems. At the moment of writing this document, the largest circuit model quantum processors reported comprise of just 10-100 qubits with error rates of  $10^{-3} - 10^{-2}$ . This is called the Noisy Intermediate Scale Quantum (NISQ) era, a term first coined by John Preskill in 2018 [1]. NISQ devices are characterized by the limited coherence times, lower gate fidelities and fewer qubits. Therefore, NISQ devices will not have adequate resources to implement quantum error correction and as a consequence only shallow circuits can be executed. However, these machines will be large enough to implement quantum dynamics that cannot be efficiently simulated with existing classical computers because classical computers are notoriously bad at simulating dynamics of highly entangled many-particle quantum systems <sup>1</sup>.

The introduction of VQE and QAOA in 2014 coincided with the transition of experimental quantum computing from mostly an academic and public endeavor to an industrial enterprise.

The inevitable question is: will these early imperfect quantum computers deliver such answers? We hope this is the case, and we believe that a viable path towards a definite affirmative answer lies in the further development and perfecting of quantum variational approaches. Variational methods emerged in the last few years as a promising way to utilize the the existing NISQ devices instead of waiting for a universal quantum computer.

---

<sup>1</sup>Even just storing arbitrary quantum states in classical computers is difficult, because the amount of classical resources required to represent a given quantum state is related to the degree of entanglement of the state. This problem just gets worse when studying the dynamics. Check Exercise 4.46 in Ref. [?].

The very first variational quantum algorithms (VQA); the Variational Quantum Eigen-solver (VQE) [?] and the Quantum Approximate Optimization Algorithm (QAOA) [?] paved the way for a new paradigm in quantum algorithm design, by taking the limitations of NISQ devices into consideration. Before these proposals, quantum algorithmic development were aimed at fault tolerant quantum computers with error corrections capabilities. In contrast, VQAs seek to remove the stringent requirements in coherent evolution by formulating the computation as an approximate optimization problem. Here, a functional<sup>2</sup> in the space of heuristic functions defined by a variational quantum circuit is optimized. From this perspective variational quantum information processing and VQAs concern a strategy for finding extrema of *quantum functionals*, which are defined as mappings from the state space of a quantum system onto a scalar [?]. The flexibility in the definition of these heuristics allows the expansion of operability of these approached within the NISQ regime, circumventing the need for quantum error correction.

which sections deal with what, which examples are given

## 1.2 Variational quantum algorithms

VQAs arise from three main concepts in variational quantum information processing as given below [?]:

1. *Approximate nature of the calculations:* Instead of guaranteeing an exact solution to the problem, VQAs offer approximate solutions. This helps us gain more flexibility in the number of resources required, in exchange for performance guarantees on the algorithm.
2. *The use of variational quantum circuits:* Variational quantum circuits (also called variational ansatz or variational forms or parametrized quantum circuits) are quantum circuits in which some of the unitary operations have tunable parameters obtained by parametrizing the corresponding quantum gates. These parameters are systematically updated when the algorithm is executed. To give an oversimplified analogy, the variational quantum circuit is like a stencil of the problem where the sizes of the holes can be varied. Computational problems which encode solutions as extreme values of a functional can be naturally described using the variational quantum circuits as heuristics. The flexibility in choosing variational circuits allows one to pick circuits that can operate within the capabilities of the NISQ devices. This operational robustness thus removes the requirement for error correction.
3. *Employing hybrid quantum-classical computing strategy:* The variational quantum circuits require an optimization routine to tune the parameters. This task is allocated to a classical computer. The optimization routine is carried out in a quantum-classical loop via classical communication. This feature motivates the idea of hybrid quantum-classical computing. The key idea of this approach is to 'divide and conquer' where the main computational task is divided into subtasks and allocated between the classical and quantum computers. By outsourcing pre-processing and post-processing involved to a classical computer, it is possible to create algorithms that exhibit advantage using fewer quantum resources than an algorithm that implements the entire task only on a quantum device.

These elements are combined to create a general strategy that can be applied to find approximate solutions to optimization problems. The main goal of a VQA is to find

---

<sup>2</sup>A functional is simply a mapping that takes a function as input and returns a scalar as output.

a function that extremizes the value of specific quantities of interest depending on the problem. For example, in the Variational Quantum Eigensolver (VQE) [1], the goal is to find a quantum circuit that prepares the ground state for a given Hamiltonian, usually the electronic structure problem. Another prominent example of a VQA is the Quantum Approximate Optimization Algorithm (QAOA) [2] where the goal is to find a quantum circuit that prepares the state which maximizes a problem specific cost function with encoded constraints.

### 1.2.1 Outline of the algorithm

In practice, VQAs are carried out following an iterative procedure on a Quantum Processing Unit (QPU) with continuous feedback to and from a classical central processing unit (CPU)<sup>3</sup>. A VQA consists of several modular components split between the QPU and CPU and can be readily connected, extended, and enhanced individually with developments in algorithms and quantum hardware. A full protocol consists of:

1. *An initial state preparation*: we start by preparing the initial input state  $|\varphi\rangle$ ;
2. *Application of a parametrized unitary*: a parameterized unitary  $\hat{U}(\vec{\theta})$  is applied to the input state on the QPU preparing the output state  $\hat{U}(\vec{\theta})|\varphi\rangle$ . This parametrized unitary is defined by the choice of variational ansatz. It should correspond to a quantum circuit that cannot be efficiently computed using classical resources<sup>4</sup>;
3. *A cost function and a method to estimate/evaluate it*: the value of the function is estimated using a quantum circuit. Ultimately, the estimation procedure reduces to performing measurements on the output state. In order to estimate the function to a given accuracy, Step 1 has to be repeated several times to prepare trial states to collect sufficient measurement statistics;
4. *A feedback loop with Classical optimization routine*: based on the function estimate from Step 3. A classical optimization technique proposes a new set of values for the optimal variational parameters  $\vec{\theta}'$ .

Steps 1-4 are repeated until some problem specific convergence criteria are satisfied. The following sections will deal with each of the component in depth and a summarized diagram of the full protocol is shown in Fig. ??.

### 1.2.2 The variational paradigm and general quantum systems

The variational principle in quantum mechanics states that, for a given quantum state  $|\Psi\rangle$  and an observable  $\hat{O}$ , the expectation value

$$\langle \hat{O} \rangle_{|\Psi\rangle} = \frac{\langle \Psi | \hat{O} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \geq \lambda_0, \quad (1.1)$$

where  $\lambda_0$  is the lowest eigenvalue of operator  $\hat{O}$ . The equality holds only when  $\langle \hat{O} \rangle_{|\Psi\rangle} = \lambda_0$  and  $|\Psi\rangle$  is said to be in the ground state. For a detailed discussion on the variational principle see Appendix 1.A.

---

<sup>3</sup>This is similar to the Graphical Processing Unit (GPU) and the CPU working together in a normal computer. In this case the QPU plays the role of a GPU.

<sup>4</sup>One should remember that certain sets of quantum gates such as the Clifford group gates, compose quantum circuits that can be efficiently simulated (i.e in polynomial time) with classical resources according to the Gottesman-Knill theorem

In many instances, the eigenstates corresponding to the lowest few eigenvalues and their properties are of primary interest. In physical systems this is due to the fact that low energy states play a dominant role in determining the properties of a system at moderate temperatures, and in optimization problems they often encode the optimal solution. The variational principle leads to a simple method to find the ground state of a quantum system with  $\hat{O} = \hat{H}$ : prepare several parametrized 'trial states'  $|\Psi(\vec{\theta})\rangle$ , measure the expectation value of the Hamiltonian, choose the state where the expectation value is minimized. This brings us to the topic below.

### How to implement a variational method in a quantum computer?

Let us consider a quantum system  $\mathcal{S}$  composed of  $N$  qubits which will act as our quantum computer. Take a Hamiltonian  $\hat{H}$  of a different system  $\mathcal{Q}$  which need not have any relations to  $\mathcal{S}$  other than acting on a Hilbert space of less than  $N$  qubits. The goal is to find and study the system  $\mathcal{Q}$  using  $\mathcal{S}$ . The Hamiltonian  $\hat{H}$  naturally arises in the description of all physical systems. For example, this Hamiltonian could be derived from a physical system such as a collection of interacting spins or an interacting electronic system.

Hamiltonians are not the only operators that can be measured on quantum devices; in general any expectation value of an operator can be obtained by extending this method. Our attention is restricted to the class of operators whose expectation value can be measured efficiently on  $\mathcal{S}$  and mapped to  $\mathcal{Q}$  [?]. A sufficient condition for this property is that the operator  $\hat{O}$  has a decomposition into a polynomial sum of simple operators<sup>5</sup> as

$$\hat{O} = \sum_{\alpha} c_{\alpha} \hat{O}_{\alpha} \quad (1.2)$$

where  $\hat{O}$  acts on  $\mathcal{Q}$ ,  $\alpha$  runs over a number of terms polynomial in the size of the system,  $c_{\alpha}$  is a constant coefficient and each  $\hat{O}_{\alpha}$  has a simple measurement prescription on  $\mathcal{S}$ . This will enable straightforward measurement of the expectation values of  $\hat{O}$  on  $\mathcal{Q}$  by weighted summation of projective measurements of  $\mathcal{S}$ . A simple and relevant example is the decomposition of a Hermitian operator into a weighted sum of Pauli strings (See Appendix 1.B).

After deciding the operator and the corresponding measurement scheme through polynomial expansion as given by Eq. (1.2), the system  $\mathcal{S}$  needs to be prepared into the parametrized trial states. How they are prepared is the key element of VQAs and is discussed in detail under ???. To give a more general description, influence of the environment on the trial state preparation must be taken into account. This is better described by an ensemble given by a parametrized density matrix  $\rho(\vec{\theta})$ . In an ideal case where the state preparation is error free and a pure state is maintained,  $\rho(\vec{\theta}) = |\Psi(\vec{\theta})\rangle\langle\Psi(\vec{\theta})|$ . In density matrix formalism, the expectation value of an operator is given by

$$\langle \hat{O} \rangle_{\rho} = \text{Tr} [\rho \hat{O}] \quad (1.3)$$

and the variational principle in Eq. (1.1) still holds. It can be rewritten as

$$\langle \hat{O} \rangle_{\rho} (\vec{\theta}) = \langle \hat{O} \rangle (\vec{\theta}) = \text{Tr} [\rho(\vec{\theta}) \hat{O}] \geq \lambda_0. \quad (1.4)$$

The fact that this principle still holds for mixed states is the main reason for the robustness of VQAs to errors and environmental influence. By finding the set of parameters  $\vec{\theta}$  that

---

<sup>5</sup>Almost all the Hamiltonians describing physical systems such as the Ising Model, Heisenberg Model, and electronic structure problem can be written as a polynomial number of terms, with respect to the system size [?].

minimizes  $\langle \hat{O} \rangle$ , one is in effect, finding a set of *experimental parameters* that are most likely to produce the ground state on average, potentially affecting a blind purification of the state of the system being produced [?].

### 1.3 Variational ansatz construction

Variational quantum circuit or the ansatz construction is a crucial aspect of VQA that dictates the success of the algorithm. It is a very active field of research and in most cases finding an optimal ansatz for a problem is difficult than finding a solution to the computational problem itself. The variational ansatz is like a stencil of a given VQA where the parameters are like the prints and holes in it. They are continuously varied to prepare a lot of different trial states. One can make this parametrization classical - i.e, given a list of parameters  $\vec{\theta} \in \mathbb{R}^{N_{params}}$ , we can prepare a different trial state  $|\Psi(\vec{\theta})\rangle$  for each  $\vec{\theta}$ . One might ask whether the classical parametrization makes the calculation of  $\langle \hat{O} \rangle$  classically tractable, but this is avoided by using gate angles as parameters to control the quantum circuit. With this key idea in mind, now we are ready to define a variational ansatz formally according to [?].

**Definition 1.3.1.** A *variational ansatz* on  $n_p$  parameters corresponds to a pair  $(U, |\vec{0}\rangle)$ , where  $U$  is smooth map from the parameter space  $\vec{\theta} \in \mathbb{R}^{n_p}$  to the unitary operator  $U(\vec{\theta})$  on  $\mathbb{C}^{2^N}$ , and  $|\vec{0}\rangle \in \mathbb{C}^{2^N}$  is the initial state, which is acted on to generate the variational state  $|\psi(\vec{\theta})\rangle = U(\vec{\theta})|\vec{0}\rangle$ , with variational energy  $E(\vec{\theta}) = \langle \psi(\vec{\theta}) | H | \psi(\vec{\theta}) \rangle$

Assuming that the vectors  $\vec{\theta}$  are continuously defined, a variational ansatz can be imagined to sweep out a (possibly open) manifold in the Hilbert space of the states. In simpler terms, it explores some smooth space that lies within the set of allowed Hilbert space. With the exception of possible boundaries, the dimension of this manifold is precisely  $N_{params}$ . This points out a limit of the variational paradigm; to explore the entire  $2^N$ -dimensional Hilbert space and guarantee that the ground state is found,  $O(2^N)$  parameters [?] and a similar scaling of time are required <sup>6</sup>. So an exponential advantage is not expected. Instead, the 'art' of a VQA design is in choosing a manifold that contains points *sufficiently close* to the true ground state, and the 'art' of application design is in choosing a problem such that 'sufficiently close' is achievable.

With the limitations in mind, the variational circuit must be chosen carefully, because the choice of ansatz greatly affects the performance of a VQA. From the perspective of the problem, the ansatz influences both the closeness to the final solution and the convergence speed. Moreover, given the exponentially small fraction of the Hilbert space that will be explored, randomly generated ansatz cannot be expected to be 'good enough' due to the difficulties faced when optimizing such an arbitrary ansatz. Other than the computational concerns, NISQ hardware has to be taken into account as well. Due to the short coherence times of NISQ devices, deeper circuits are more susceptible to errors, and some gates can be costly to compose with hardware native gates. Accordingly, most of the ansätze developed to date are categorized either as more problem-inspired or more hardware efficient, depending on their structure and application. In the following subsections, common constructions of both the types are presented.

---

<sup>6</sup>This doesn't come off as a surprise, because performing such a task is known to be QMA-hard problem [?].

### 1.3.1 Problem-inspired ansatz construction

An arbitrary unitary evolution can be generated by an Hermitian operator  $\hat{g}$  which defines an evolution in terms of the parameter  $t$ ,

$$G(t) = e^{-i\hat{g}t}. \quad (1.5)$$

### 1.3.2 Hardware-specific ansatz construction

## 1.4 Cost function

After constructing a variational ansatz, the focus should be on finding a way to extract useful classical information from it. Therefore, the next step in a VQA is to encode the problem into an objective function, sometimes called the cost function. From Section 1.2.2, cost function is just the expectation value of the operator being measured. To reformulate it appropriately, the standard definition of a VQA cost function, denoted  $f_{\hat{O}}(\vec{\theta})$  is the expectation value of an observable  $\hat{O} \in \text{Herm}(\mathbb{C}^{2^N})$ , that is,

$$f_{\hat{O}}(\vec{\theta}) = \langle \psi(\vec{\theta}) | \hat{O} | \psi(\vec{\theta}) \rangle \quad (1.6)$$

where normalization of the wavefunction is assumed. Let us look at a couple of examples of cost functions.

#### Example 1 : Approximating the ground state energy of a physical system

Information about the dynamics of a system is encoded in its Hamiltonian  $\hat{H}$ , and it naturally arises in the description of all physical systems. In physics, one is often interested in computing the ground state energy  $E_0$  of a Hamiltonian  $H$  given by,

$$E_0 = \min_{|\psi\rangle} \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle}. \quad (1.7)$$

In this case it is apparent that the corresponding cost function is given by

$$f_{\hat{H}}(\vec{\theta}) = \langle \psi(\vec{\theta}) | \hat{H} | \psi(\vec{\theta}) \rangle \quad (1.8)$$

#### Example 2 : Approximating optimization problems

For problems unrelated to physical systems, encoding into an equivalent qubit Hamiltonian is still possible, thereby opening a path to solving them on a quantum computer. To solve a combinatorial optimization problem using a quantum algorithms, the problem is first encoded onto a quantum system> one of the most widely used model in physics, that is also used to represent optimization problems, is the Ising model. A quantum version of it can be obtained easily by replacing the spin variables with Pauli  $Z$ -operators

$$\hat{H}_C \equiv \hat{H}(\sigma_1^z, \dots, \sigma_n^z) = - \sum_{1 \leq i < j \leq n} J_{ij} \sigma_i^z \sigma_j^z - \sum_{i=1}^n h_i \sigma_i^z \quad (1.9)$$

where  $J_{ij}$  and  $h_i$  are real numbers, and  $\sigma_i^z$  refers to the Pauli  $Z$ -operator acting on the  $i^{\text{th}}$  qubit. In fact, several NP-complete optimization problems and even many NP-hard problems can naturally be expressed as a problem of finding the ground state, or minimum energy configuration, of a quantum Ising Hamiltonian by choosing appropriate values for  $J_{ij}$  and  $h_i$  [?].

To sum up the idea, a cost function defines a hyper-surface called the cost landscape that maps the trainable parameters to real numbers. The task of the optimizer is to traverse the cost landscape and find the global minima. The choice of optimizer depends on how complicated the landscape is, as discussed in detail under Section 1.6. Therefore, the choice of a good objective function is also crucial to ensure convergence to the solution in addition to the variational ansatz choice. A good objective function of a problem should meet the following criteria.

1. *Faithfulness*: The optimum of the cost function must correspond to the solution of the problem
2. *Efficient estimation*: It should be possible to estimate the cost function efficiently by performing a finite number of measurements with classical post-processing.
3. *Operational meaningfulness*: It's useful when the value of the cost function reflects the quality of a solution (e.g. lower cost means better solution).
4. *Trainability*: It must be possible to efficiently optimize the parameters in order to meet a convergence criteria

Even though the properties given above sound trivial, they are difficult to satisfy when running an actual experiment on NISQ devices with constraints on circuit depth and ancilla requirements.

## 1.5 Measurement scheme

Let us now address the question of how to evaluate the VQA cost function defined in Eq. 1.6. Using Eq. (1.2) and Theorem 1.B.1, most interesting observables can be written as a linear combination of a polynomial number of Pauli strings with real coefficients, i.e.,

$$\hat{O} = \sum_{i=1}^{\text{poly}(N)} c_i \sigma_i^\alpha \quad (1.10)$$

where  $\alpha = x, y, z$  identify the Pauli spin operator. For example, most physically motivates Hamiltonians are  $k$ -local. That is, each of the Pauli strings in the sum acts nontrivially on at most  $k$  qubits. Examples of  $k$ -local systems include Ising and Heisenberg spin systems, van der Waals gases, strong and weak interactions, and lattice gauge theories. In fact, any system that is consistent with special and general relativity evolves according to local interactions [?]. In a  $k$ -local system there are  $\binom{n}{k}$  possible subsets of  $k$  qubits. Therefore the number of Pauli strings that appear in the decomposition of a  $k$ -local Hamiltonian is bounded by a polynomial. When Eq. (1.10) is encoded into a const function as in Eq. (1.6) we find that

$$\begin{aligned} f_{\hat{O}}(\vec{\theta}) &= \langle \psi(\vec{\theta}) | \hat{O} | \psi(\vec{\theta}) \rangle \\ &= \sum_{i=1}^{\text{poly}(N)} \langle \psi(\vec{\theta}) | c_i \sigma_i^\alpha | \psi(\vec{\theta}) \rangle \\ &= \sum_{i=1}^{\text{poly}(N)} c_i \langle \psi(\vec{\theta}) | \sigma_i^\alpha | \psi(\vec{\theta}) \rangle. \end{aligned} \quad (1.11)$$

From this we see that the evaluation of the cost function boils down to evaluating the expectation value of a weighted sum of Pauli strings assuming that the values of  $c_i$  are known beforehand. Expectation value of a tensor product of an arbitrary number of Pauli operators can be estimated by measuring each of the qubit locally [?]. This is an operation that requires a coherence time of  $O(1)$  (i.e. incurring a constant cost in time) under the assumption that parallel qubit rotations and readouts are possible [?]. Subsection 1.5.1 demonstrates how to find the expectation value of a Pauli string.

### 1.5.1 Measurement of Pauli strings

### 1.5.2 Precision and Chebyshev's inequality

Suppose we wish to estimate  $E = \langle \psi(\vec{\theta}) | \mathcal{P} | \psi(\vec{\theta}) \rangle$ , for some Pauli string  $\mathcal{P}$ , up to a precision  $\varepsilon > 0$ . In order to achieve that, let

$$\mathcal{P} = \sum_{k=1}^{2^N} \lambda_k |\phi_k\rangle\langle\phi_k| \quad (1.12)$$

denote the spectral decomposition of  $\mathcal{P}$ . Here  $E$  can be taken as the expected value of a random variable  $X$  with possible outcomes  $\{\lambda_i\}_{i=1}^{2^N}$  with

$$\mathbb{P}(X = \lambda_k) = \left| \langle \phi_k | \psi(\vec{\theta}) \rangle \right|^2. \quad (1.13)$$

That is,  $X$  is the random variable representing the outcome of measuring operator  $\hat{O}$  on the state  $|\psi(\vec{\theta})\rangle$ . *Chebyshev's inequality* is a well known result in probability theory which states that if we take  $M$  copies of  $X$ , denoted by  $X_1, \dots, X_M$ , then

$$\mathbb{P}\left(\left| \frac{\sum_{i=1}^M X_i}{M} - E \right| \geq \varepsilon\right) \leq \frac{\sigma^2}{M\varepsilon^2}, \quad (1.14)$$

where  $\sigma$  denotes the variance of  $X$ . Since the variance of  $X$  can be assumed to be bounded above by a constant, from Eq. (1.14) it can be deduced that

$$M \sim \frac{1}{\varepsilon^2} \quad (1.15)$$

measurements must be performed, to obtain, with high probability, an estimate of  $E$  to within additive precision  $\varepsilon$ . The limitations in current NISQ hardware has severe implications on how well the above strategy will allow one to estimate the energies  $E$ , as the measurement outcomes will become noisy. This inherent statistical nature of the outcomes will influence the optimization strategy. This is discussed in Section 1.6.

### 1.5.3 Measurement scheme optimization

The strategy presented here is a very naive way of estimating the cost function by sampling every Pauli string and using the outcome statistics to estimate the corresponding expectation value. As mentioned before, the key advantage of this approach is that the coherence time to make a single measurement after preparing the state is  $O(1)$ . Conversely, the disadvantage of this approach is the scaling of the total number of operations, as a function of the desired precision  $\varepsilon$  as  $O(\varepsilon^{-2})$  [?]. Moreover, this scaling also reflects the number of state preparation repetitions required. This has led to an active area of research in measurement optimization strategies. One example of such a technique would be to use

the commutation properties of Pauli strings and make simultaneous measurements. The problem is then to divide the Pauli strings into a small number of commuting subsets. It turns out that this problem is equivalent to the clique cover problem, which is known to be NP-complete [?]. Researchers have proposed new methods using approximation algorithms for the clique cover problem as a grouping, with various degrees of success. By speeding up the measurement scheme with new techniques, it is possible to realize a faster algorithm with efficient use of quantum resources.

## 1.6 Optimization routine

Optimization is the next critical aspect of VQAs. Numerical optimization is a vast field, and therefore identifying and adapting existing optimization techniques that are suited to quantum variational approaches requires considerable effort because success of a VQA depends on the reliability and efficiency of the classical optimization method used. The cost function is statistical by design. Unlike a typical optimization routine in classical computing, the inherently stochastic environment due to the limited number of measurements, hardware noise, and the presence of barren plateaus makes the choice of an optimizer difficult. This has led to active research in the development of quantum-aware optimizers.

The following should be kept in mind when choosing an optimizer.

1. Due to short coherence times in the NISQ era, complicated analytical gradient circuits cannot be implemented
2. The optimizer should be resilient to noisy data and precision on objective function evaluation that is limited by the number of shots in the measurement.
3. Objective function evaluations take a long time, which means shots frugal optimizers are favored.

Given below is a small selection of optimizers used commonly or promoted for VQAs.

### 1.6.1 Gradient based methods

### 1.6.2 Gradient free approaches

#### Barren plateaus

#### Effect of noise

## 1.7 Applications

## 1.8 Hands-on tutorial: Simulating $H_2$ molecule using VQE

- short outline of VQE
- flow chart of the algorithm
- mapping and second quantization
- calculating the cost function example at bond length
- comparing two different variational forms
- different optimizers and effect of noise

- qubi reduction and symmetries

VQE consists of three different nested iterations.

1. outer iterations : to update parameters  $\vec{\theta}$  of quantum state  $|\Psi(\vec{\theta})\rangle$ ,
2. middle iteration: to evaluate the expectations value by calculating weighted sum of pauli strings, and
3. inner iterations : to evaluate the expectation value of a pauli string through sampling.

## 1.9 Outlook and conclusions

Will NISQ devices running VQAs be able to outperform classical algorithms that find approximate solutions to the same computational problems? Nobody knows *yet*, but there is hope. Even if the early generation NISQ devices are incapable of competing with classical methods that were honed over decades of research, experimental results encourage that VQAs can soon surpass classical methods, and so spur further scientific advancements.

rewrite: Given that variational algorithms are heuristic and provide approximate solutions, one might wonder whether there will be value in these approaches once error-correction becomes available, especially compared to quantum algorithms with a proven advantage on the same tasks. The first thing to consider is that designing algorithms with proven advantage is challenging and so far only a relatively small number of such algorithms have been discovered. Therefore, for many applications, it is possible that variational algorithms will be the only quantum algorithms available in the error-correction era. A second consideration is that variational algorithms are designed to optimize the use of quantum resources, and therefore it is likely that their error-corrected implementations are more efficient than implementations of their counterparts with proven advantage. Consequently, quantum algorithms might be still competitive with traditional algorithms regarding computational cost. Furthermore, since formal demonstrations of advantage are based on asymptotic considerations, there is a chance for variational algorithms to be more efficient and sufficiently accurate for specific instances of a problem. Finally, the third consideration is that variational algorithms can be combined with traditional quantum algorithms to obtain even more powerful approaches. A concrete example of this can be VQE. Techniques employed for simulating quantum systems with provable speed-up, such as quantum phase estimation, require the preparation of states with sufficient overlap with the eigenstates of the Hamiltonian to measure eigenvalues with high probability. Using VQE, it will be possible to prepare such states, ultimately boosting the success probability of phase estimation.

Finally, we point out that the success of variational approaches ultimately depends on the quality of the quantum devices and the quantum operations they implement. Therefore, maximizing the utility of quantum devices, in particular, early machines, is another way to push quantum computing towards the practical frontier. These improvements can be introduced at the software level by finding more efficient protocols to implement quantum operations common to many algorithms.

The work presented here is only a part of the continuously growing body of research in variational algorithms [?, ?, ?, ?]. The history of this field has just started, and many exciting developments await for us.

## App. 1.A The variation principle

Consider an arbitrary physical system with Hamiltonian  $H$ . The time-independent Schrödinger equation then reads as,

$$H |\varphi_n\rangle = E_n |\varphi_n\rangle; \quad n = 0, 1, 2, \dots \quad (1.16)$$

Although the Hamiltonian  $H$  is known, this is not necessarily the case for its eigenvalues  $E_n$  and the corresponding eigenstates  $|\varphi_n\rangle$ . Hence, the variational methods are the most useful when an exact diagonalization of  $H$  is not feasible or not known prior. Such instances of  $H$  also come with a drawback as it is difficult, if not impossible, to evaluate the error in a calculation when the final exact solution is unknown. This method is most useful when an exact diagonalization of  $H$  is not feasible and not known prior.

Following Ref [?, ?, ?], the solution of Eq (1.16) with the approximate state  $|0\rangle$ , recast in the form of a variation principle for the energy written as an expectation value is given by,

$$E[\tilde{0}] = \frac{\langle \tilde{0} | H | \tilde{0} \rangle}{\langle \tilde{0} | \tilde{0} \rangle}. \quad (1.17)$$

Here,  $|\tilde{0}\rangle$  denotes some approximate eigenstate. The square brackets indicate that the energy functional depends on the form of the wavefunction rather than a set of parameters<sup>7</sup>. See Appendix ?? for more details on exact and approximate wavefunctions.

As the first step, a one-to-one relationship between the solutions to the Schrödinger equation and the stationary points of the energy functional  $E[\tilde{0}]$  is established. Let  $|0\rangle$  represent a solution to the Schrödinger equation (1.16) and  $|\delta\rangle$ , an allowed variation. The approximate state to the eigenstate is then given by,

$$|\tilde{0}\rangle = |0\rangle + |\delta\rangle. \quad (1.18)$$

Inserting (1.18) in (1.17) followed by an expansion in orders of  $|\delta\rangle$  around  $|0\rangle$  yields,

$$\begin{aligned} E[0 + \delta] &= \frac{\langle 0 | H | 0 \rangle + \langle 0 | H | \delta \rangle + \langle \delta | H | 0 \rangle + \langle \delta | H | \delta \rangle}{\langle 0 | 0 \rangle + \langle 0 | \delta \rangle + \langle \delta | 0 \rangle + \langle \delta | \delta \rangle} \\ &= E_0 + \langle 0 | H - E_0 | \delta \rangle + \langle \delta | H - E_0 | 0 \rangle + O(\delta^2) \\ &= E_0 + O(\delta^2) \end{aligned} \quad (1.19)$$

The first order variation in  $E[\tilde{0}]$  therefore vanishes whenever  $|\tilde{0}\rangle$  corresponds to one of the eigenstates  $|0\rangle$ . This shows that the eigenstates of the Schrödinger equation represent stationary points of the energy functional.

Conversely, to show that all stationary points of  $E[\tilde{0}]$  represent eigenstates of the Schrödinger equation, let  $|0\rangle$  be a stationary point of the energy functional. For a variation  $|\delta\rangle$ , expansion of the energy functional around the stationary point similar to (1.19) gives,

$$\langle 0 | H - E[0] | \delta \rangle + \langle \delta | H - E[0] | 0 \rangle = 0. \quad (1.20)$$

For the variation  $i|\delta\rangle$ ,

$$\langle 0 | H - E[0] | \delta \rangle - \langle \delta | H - E[0] | 0 \rangle = 0. \quad (1.21)$$

Combining Eq. (1.20) and Eq. (1.21) one obtains,

$$\langle \delta | H - E[0] | 0 \rangle = 0. \quad (1.22)$$

---

<sup>7</sup>In most applications like in VQAs, the wavefunction is described in terms of a finite set of parameters. During such cases, the usual notation for functions is used.

Since this relation holds for any arbitrary  $|\delta\rangle$ , the eigenvalue equation is then,

$$H|0\rangle = E[0]|0\rangle. \quad (1.23)$$

This shows that each stationary point  $E[0]$  of the energy functional  $E[\tilde{0}]$  also represents a solution to the Schrödinger equation with eigenvalue  $E[0]$ . To summarize in one line, the variational principle states that the solutions of the Schrödinger equation (1.16) are equivalent to a variational optimization of the energy functional (1.17).

Due to the flexible nature of variational methods, they can be adapted to very diverse problem scenarios. It also gives great scope to physical intuition in the choice of trial function. It's important to keep in mind that even though good approximations of eigenvalues are obtained rather easily, the approximate states may come with certain completely unpredictable erroneous features that cannot be checked.

## App. 1.B Pauli basis

The contents of this section are from [?].

**Definition 1.B.1.** *The state of an  $N$ -qubit quantum register is represented by a norm-1 vector in the Hilbert space  $\mathcal{H} = \mathbb{C}^{2^N}$ , under the association  $|\psi\rangle \in \mathcal{H} \equiv e^{i\phi}|\psi\rangle$  for  $\phi \in \mathbb{R}$ .*

**Definition 1.B.2.** *The **Pauli basis** on  $N$  qubits is defined as  $\mathbb{P}^N := \{\mathbb{I}, X, Y, Z\}^{\otimes N}$ , where  $\mathbb{I}, X, Y, Z$  are the  $2 \times 2$  matrices on  $\mathbb{C}^2$ :*

$$\mathbb{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

**Theorem 1.B.1.** *The Pauli basis is a basis for the set of  $2^N \times 2^N$  complex valued matrices. It is also a basis for the set of Hermitian matrices if one chooses real coefficients.*

**Remark.** *The Pauli basis is not a group under matrix multiplication, as the single-qubit Pauli matrices pick up a factor of  $i$  on multiplication. The closure of the Pauli basis is the Pauli group  $\Pi^N = \{\pm i\} \times \mathbb{P}^N$ ; this is four times as large, and no longer has the basis properties of  $\mathbb{P}^N$ .*

Some useful properties of  $\mathbb{P}^N$ :

- $P^2 = \mathbb{I}$  for all  $P \in \mathbb{P}^N$ .
- For  $P, Q \in \mathbb{P}^N$ , either  $[P, Q] = 0$ , or  $\{P, Q\} := 0$ , and  $P$  commutes with precisely half of  $\mathbb{P}^N$
- $P \in \mathbb{P}^N \neq \mathbb{I}$  has only two eigenvalues,  $\pm 1$  and the dimension of the corresponding eigenspaces is precisely  $2^{N-1}$  (i.e. each  $P$  divides  $\mathbb{C}^{2^N}$  in two)
- This division by two may be further continued, given  $P, Q \neq \mathbb{I}$  such that  $[P, Q] = 0$ ,  $P$  and  $Q$  divide the Hilbert space into 4 eigenspaces (labeled by combinations of their eigenvalues).
- To generalize, one can form a  $[N, k]$  **stabilizer group**  $\mathcal{S}$ , generated by  $k$  commuting, Hermitian, non-generating elements of  $\mathbb{P}^N$  (up to a complex phase); this diagonalizes  $\mathbb{C}^{2^k}$  into  $2^k$  unique eigensectors of dimension  $2^{N-k}$ . When  $N = k$ , these sectors contain single eigenstates, which are called **stabilizer states**.
- Given such a stabilizer state  $|\psi\rangle$  and Hermitian  $P \in \mathbb{P}^N$ , either  $P|\psi\rangle = \pm|\psi\rangle$  or  $\langle\psi|P|\psi\rangle = 0$ .

## App. 1.C Minimizing the VQE cost function

### 1.C.1 Gradient estimation

Parameter shift rule

Stochastic finite difference approximation