QUANTUM OPTIMIZATION ALGORITHMS.

In this lecture we discuss optimization methods using quantum computation approaches.

I. We are first concerned with combinatorial optimization problems which can be formulated as the minimization of a classical spin Hamiltonian. Taking inspiration from the quantum adiabatic theorem (seen last time) we can formulate the QAOA algorithm (resembling the quantum alternating method seen last time).

II. We then will discuss the variational quantum eigensolver (VQE) algorithm of great importance (in quantum chemistry for example)

to find the ground state energy and first excited states of a quantum system.

III. These algorithms all require to optimize over some set of parameters (be it QAOA or VQE). This can be done by formulating the eventual algorithm as a hybrid classical-quantum algorithm. In this context we will also discuss the parameter-shift rule which allows to easily perform gradient steps.

I. Combinatorial optimization problems.

Let $C_a : \{0,1\}^n \to \mathbb{R}_+$ be functions with Boolean arguments $z_1 \dots z_n$ and value in $\mathbb{R}_+$ (or some finite interval, or boolean in $\{0,1\}$).

We want to <u>minimize</u> (or maximize) the objective:

$$C(\underline{z}) = \sum_{a=1}^{m} C_a(\underline{z})$$

Often for each $a = 1 \dots m$ the function depends on $O(1)$ variables (and is said to be local) (with respect to n which is supposed to be large).

Using the mapping $s_i = (-1)^{z_i}$ or $z_i = \frac{s_i + 1}{2}$ we can rewrite $C(\underline{z})$ as some sort of classical spin system

with classical Hamiltonian :

$$H_C(\underline{s}) = \sum_{a=1}^{M} C_a(\ldots, \frac{1+s_i}{2}, \ldots)$$

One can use the idea outlined last time and introduce a time dependent family :

$$\hat{H}(t) = \left(1 - \frac{t}{T}\right) \sum_{i=1}^{N} X_i + \frac{t}{T} H_C(\underline{z})$$

where $0 \leq t \leq T$ and

$$H_C(\underline{z}) = \sum_{a=1}^{M} C_a(\ldots, \frac{1+z_i}{2}, \ldots)$$

Here we use info theoretical notation for Pauli matrices $X = \sigma^x$ & $Z = \sigma^z$.

For $t=0$ the ground state is

$$|\psi_0(0)\rangle = \bigotimes_{i=1}^{m} \frac{|0\rangle + |1\rangle}{\sqrt{2}} .$$

For $t=T$ the ground state is given by the minimum of $H_c(\underline{s})$: call it $\underline{s}^*$

$$|s_1^* \; s_2^* \; \dots \; s_m^*\rangle = |\psi_0(T)\rangle$$

(assuming its unique for simplicity)

Recall the main idea behind the adiabatic theorem: if the energy $E_0(t)$ of the GS $|\psi_0(t)\rangle$ is separated from the first excited state $E_1(t)$ by a gap $\geq \Delta$ then for T large enough $|\psi_0(T)\rangle \approx U(T)|\psi_0(0)\rangle$.

The QAOA is inspired by this theorem and evolves the state $|\psi_0(0)\rangle$ by a sequence of "small" unitary steps. The unitary steps are generated successively by the:

- "cost Hamiltonian" $H_C(Z_1 \dots Z_N)$ defined above
- the "mixer Hamiltonian" H_M. Here

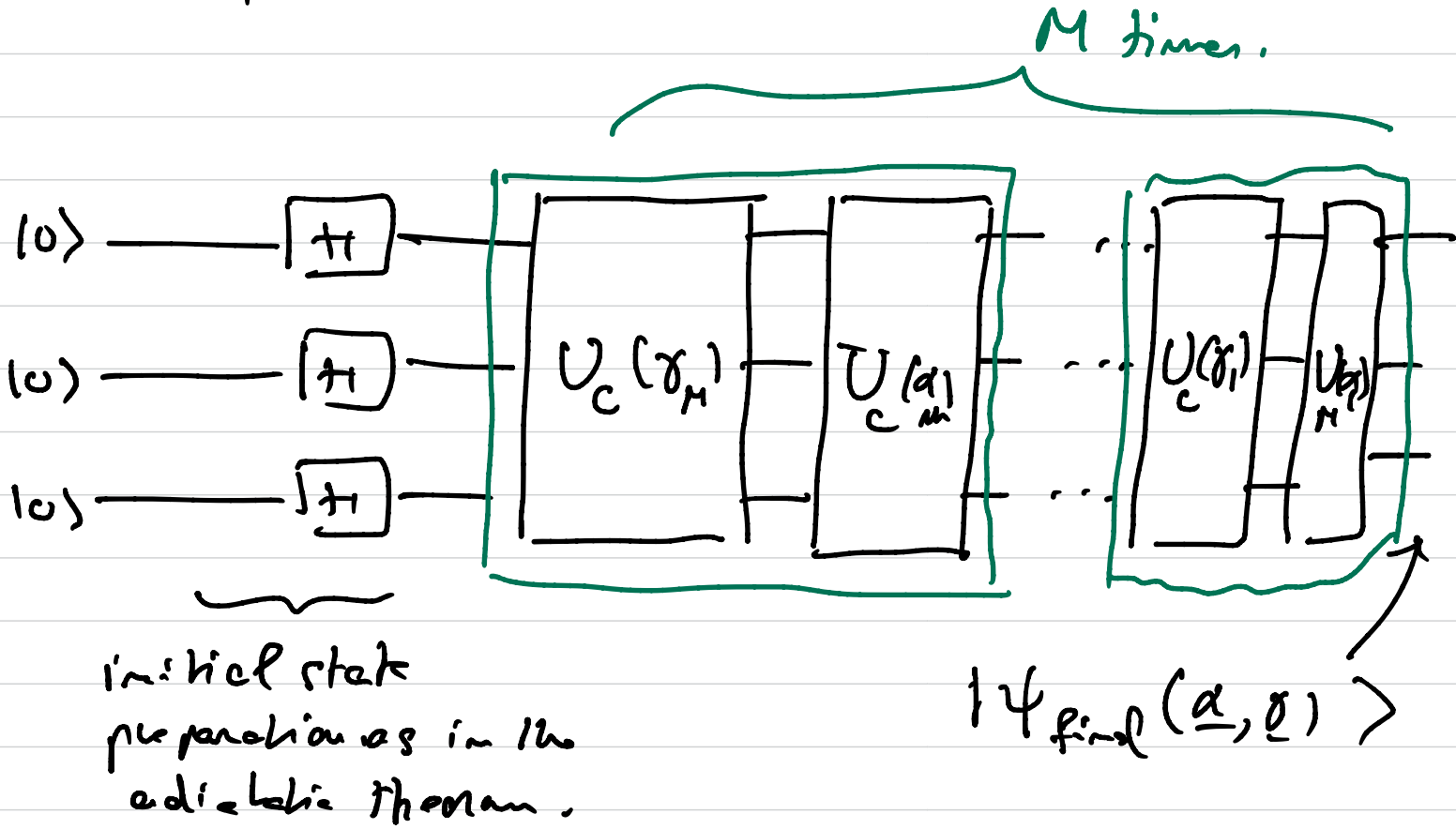
$$H_M = \sum_{i=1}^{N} X_i = \sum_{i=1}^{N} \mathbb{1} \otimes \dots \otimes \mathbb{1} \otimes X_i \otimes \mathbb{1} \dots \otimes \mathbb{1}$$

We take a sequence of parameters ("angles") to be eventually optimized upon $\gamma_1 \dots \gamma_M$, $\alpha_1 \dots \alpha_M$ (for M "large" enough) and successively apply for $i = 1 \dots M$:

$$U_M(\alpha_i)\, U_C(\gamma_i) = e^{-i\gamma_i H_C}\, e^{-i\alpha_i H_M}$$

The quantum circuit of QAOA corresponds to:

$$\prod_{i=1}^{M} \left(U_M(\alpha_i)\, U_C(\gamma_i) \right) H^{\otimes M} |0\rangle^{\otimes M}$$



We see that the basic difference with the Quantum Alternating method lies in the fact that $|\psi_{final}(\underline{\alpha}, \underline{\gamma})\rangle$ depends on "variational parameters" $\underline{\alpha}$, $\underline{\gamma}$.

An optimal solution will be given $\underline{\alpha}^*, \underline{\gamma}^*$ that minimize

$$E_{final}(\underline{\alpha}, \underline{\gamma}) = \langle \psi_{final}(\underline{\alpha}, \underline{\gamma}) | H_C | \psi_{final}(\underline{\alpha}, \underline{\gamma}) \rangle$$

One hopes that for M large enough $|\psi_{final}(\underline{\alpha}^*, \underline{\gamma}^*)\rangle$ has a large overlap with $|s_1^*, s_2^*, \ldots, s_N^*\rangle$ the minimizers of the classical Hamiltonian.

For the optimization steps we can proceed as follows:

- $\langle \psi_{final} | H_C | \psi_{final} \rangle$ can be obtained by repeated measurements in the computational basis since H_C depends only on $Z_1 \ldots Z_N$ and is diagonal in that basis.

- $(\underline{\alpha}, \underline{\gamma})$ are updated according to a gradient step (say):

$$\underline{\alpha}' = \underline{\alpha} - \eta \nabla_{\underline{\alpha}} E_{final}(\underline{\alpha}, \underline{\gamma})$$

$$\underline{\gamma}' = \underline{\gamma} - \eta \nabla_{\underline{\gamma}} E_{final}(\underline{\alpha}, \underline{\gamma}).$$

- We iterate.

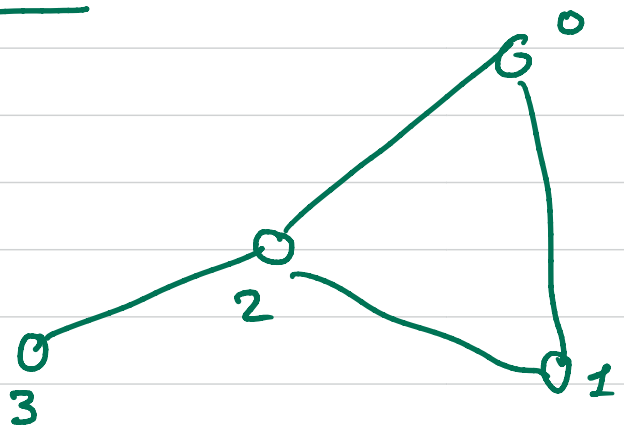
Computing the gradients above may to always be easy. In section III we discuss special cases where this can somehow be done efficiently.

An example: Minimum vertex cover of a graph.

We wish to show here a typical application of these ideas to a combinatorial optimization problem. In the exercises you will implement QAOA for this problem.

Given a graph $G = (V, E)$ we want to find the minimum size of a subset of the vertices such that every edge touches at least once this subset. This subset is called the minimal cover.

Example:



$V = \{0, 1, 2, 3\}$ $E = \{(01), (12), (02), (23)\}$

Minimal covers are $\{0, 2\}$ and $\{1, 2\}$. Note that $\{2, 3\}$ would not qualify as a cover because edge $(0, 1)$ does not touch the set. Also $\{0, 1, 2\}$ is a cover but is not minimal. Ect...

To implement QAOA we first have to map this problem onto the minimization problem of a certain classical spin system. This can be done through a general sort of method which we outline here for the example. Hopefully it is more or less clear how to proceed for other cases. For more information we refer to arXiv: 1302.5843 v3 (Andrew Lucas: Ising formulations of many NP problems).

For the graph $G = (V, E)$ and the minimum vertex cover problem we should minimize:

$$\sum_{v \in V} x_v$$

where $x_v = 1$ if the vertex is covered (colored black) and $x_v = 0$ if the vertex is not covered (colored white).

We must implement the constraint that for each edge $(u, v) \in E$ at least $x_v = 1$ or $x_u = 1$.

This means for $(u, v) \in E$: $(1 - x_v)(1 - x_u) = 0$,

Thus we want covers that satisfy:

$$\sum_{(u,v) \in E} (1 - x_u)(1 - x_v) = 0 \quad .$$

This suggest the following Hamiltonian:

$$H_c = \sum_{v \in V} x_v + \lambda \sum_{(u,v) \in E} (1-x_u)(1-x_v)$$

where λ is a "Lagrange multiplier". It is easy to see that we should take $\lambda > 1$.

Proof: Indeed we should show that for $\lambda > 1$ we have $\min_{\underline{x} \in \{0,1\}^V} H_c(\underline{x})$ is attained on covers. In this case

$$\min_{\underline{x} \in \{0,1\}^V} H_c(\underline{x}) = \min_{\underline{x} \text{ covers}} \left(\sum_{v \in V} x_v \right)$$

and we are minimizing the size of the cover.

Now take a cover $\underline{\tilde{x}}$, so that is satisfies for all $(u,v) \in E$ $(1-\tilde{x}_u)(1-\tilde{x}_v) = 0$. Flip a vertex j from $\tilde{x}_j = 1$ (covered) to $\tilde{x}_j = 0$ (uncovered) such that also for at least one edge the new $\underline{x}'$ is not a cover. Then

$$H_c(\underline{x}') = H_c(\tilde{\underline{x}}) - 1 + \underbrace{\lambda}$$

at least λ is payed if only one edge is not covered.

We want $H_c(\tilde{\underline{x}}) < H_c(\underline{x}')$ which implies $\lambda > 1$. ∎

Of course this is not the only possibility. One can use other coefficients in front of the monomials in $H_c(\underline{x})$. The last step is to make a mapping to spins: $x_v = \frac{S_v + 1}{2}$. This yields eventually an antiferromagnetic Ising model in an external magnetic field. (for vertex cover).

II. Variational Quantum Eigensolver alg. (VQE).

Recall the variational principle of quantum mechanics. Let H be a (hermitian) Hamiltonian. The GS energy E_0 satisfies:

$$E_0 = \min_{\|\psi\|=1} \langle \psi | H | \psi \rangle.$$

The variational method use a class of "trial" states $|\psi(\lambda)\rangle$ with λ some set of parameters and attempts to estimate E_0 by minimizing in the restricted set $\{|\psi(\lambda)\rangle\}$. In general this yields an upper bound

$$E_0 \leq \min_{\lambda} \langle \psi(\lambda) | H | \psi(\lambda) \rangle.$$

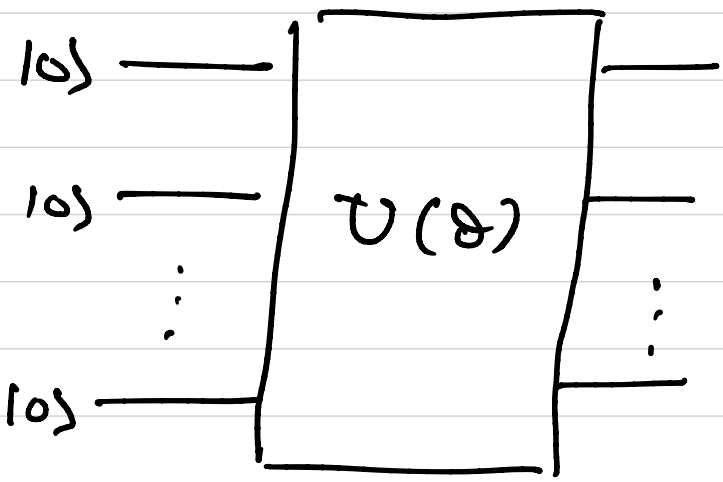
The trial class may be a class of functions motivated by the physics of the problem. Historically the principle was applied to such classes depending on a small number (or one) parameters. In recent years people explored with some success the idea of taking Neural Network representations for $|\psi(\vec{\theta})\rangle$ with many parameters ("weights") for $\vec{\theta}$. Another approach is to take $|\psi(\vec{\theta})\rangle$ as the output of a parametrized quantum circuit and optimize over the parameters entering in the gates of the circuit. This is the main idea of VQE.

We therefore set $|\psi(\vec{\theta})\rangle = U(\vec{\theta})\,|0\rangle^{\otimes m}$ where $U^\dagger(\vec{\theta})\,U(\vec{\theta}) = U(\vec{\theta})\,U^\dagger(\vec{\theta}) = \mathbb{1}$.

The variational energy is

$$E(\vec{\vartheta}) = \langle 0 | U^\dagger(\vec{\vartheta}) H U(\vec{\vartheta}) | 0 \rangle$$

and we know that $E(\vec{\vartheta}) \geq E_0$. The variational state is obtained as:



$U(\vec{\vartheta})|0\rangle = |\psi(\vec{\vartheta})\rangle$
variational state

The art is to choose $U(\vec{\vartheta})$ a sufficiently simple one & two qubit gate circuit which is at the same time sufficiently expressive.

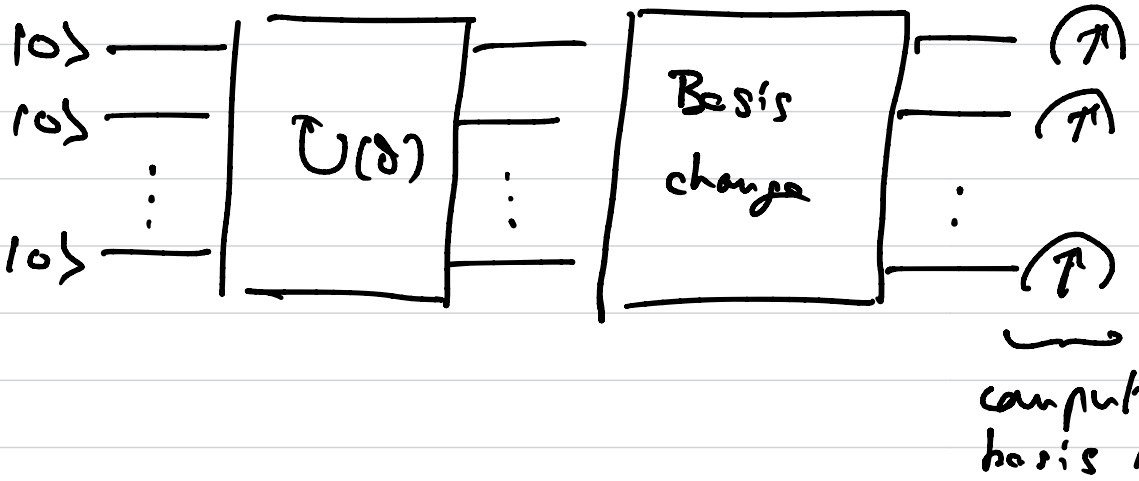
In practice we must compute the GS energy of a quantum spin Hamiltonian (e.g TFIM)

a some fermionic hamiltonian that has been qubitized (by a Jordan-Wigner transform or a Bravyi-Kitaev transform). So to compute

$$\langle 0|U^\dagger(\theta)\, H\, U(\theta)|0\rangle = \sum_{k=1}^{M} \langle 0|U^\dagger(\theta) A_k U(\theta)|0\rangle$$

where A_k is a monomial in $\{I, X, Y, Z\}$ matrices we collect outputs of the circuits:

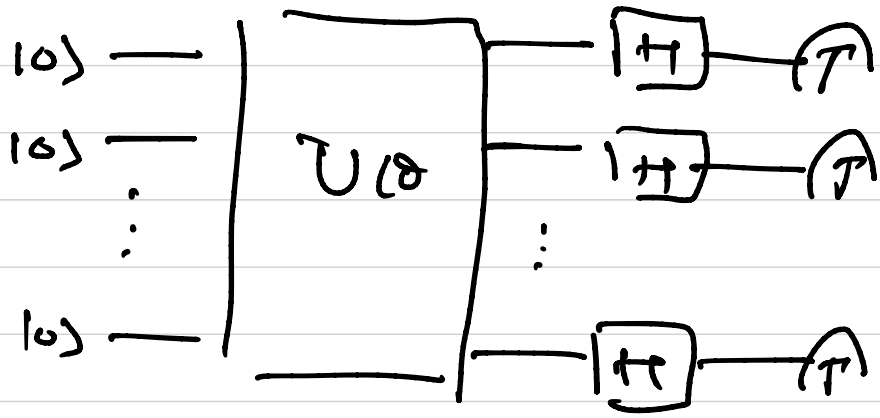
$k = 1 \dots M$,



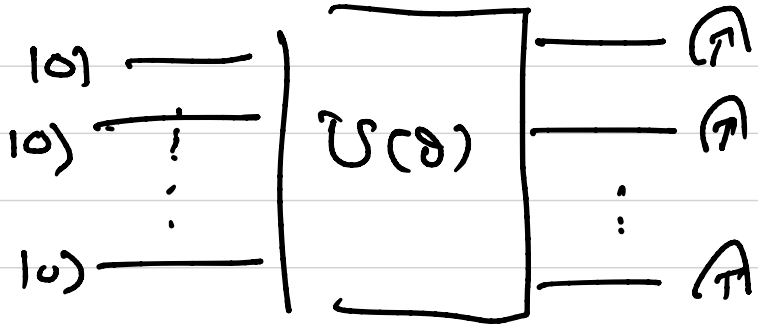
For example for the TFIM:

$$H = -\Gamma \sum_{i=1}^{N} X_i + \sum_{\langle i,j \rangle} J_{ij} Z_i Z_j$$

we use the circuits:



for the transverse field term and



for the Ising term.

In the exercises you will discuss constructions for $U(\theta)$. Once $\langle 0 | U^\dagger(\theta) H U(\theta) | 0 \rangle$ is estimated by circuits outputs the parameters $\underline{\theta}$ are updated typically through a gradient step:

$$\underline{\theta}' = \underline{\theta} - \eta \nabla_{\underline{\theta}} E(\underline{\theta})$$

Then the process is iterated. Thus the VQE is a hybrid classical-quantum algorithm.

The gradient step requires to compute $\nabla_\theta E(\theta)$ which is not a priori easy as we have access to a discrete set of values of the function $E(\theta)$. This is discussed in section III.

VQE for excited states:

We discuss excited states. The variational principle can be generalized to get excited states. For the first excited state discussed here for simplicity we have:

$$E_1 = \min_{\substack{|\psi\rangle \perp |\psi_0\rangle \\ \|\psi\| = 1}} \langle \psi | H | \psi \rangle$$

Thus we get an upper bound on E_1 by minimizing over a variational class:

$$\langle \psi(\theta) | H | \psi(\theta) \rangle - \lambda |\langle \psi_0 | \psi(\theta) \rangle|^2$$

where λ is a Lagrange multiplier.

The idea of VQE is to set

$$|\psi(\theta)\rangle = U(\underline{\theta}) |0\rangle^{\otimes n}$$

$$|\psi_0\rangle = U(\underline{\theta}_0) |0\rangle^{\otimes n}$$

$\uparrow$

found previously by minimizing

$$\langle 0 | U^\dagger(\theta_0) H U(\theta_0) | 0 \rangle.$$

Thus we now minimize: $\mathcal{L}(\theta, \lambda) \equiv$

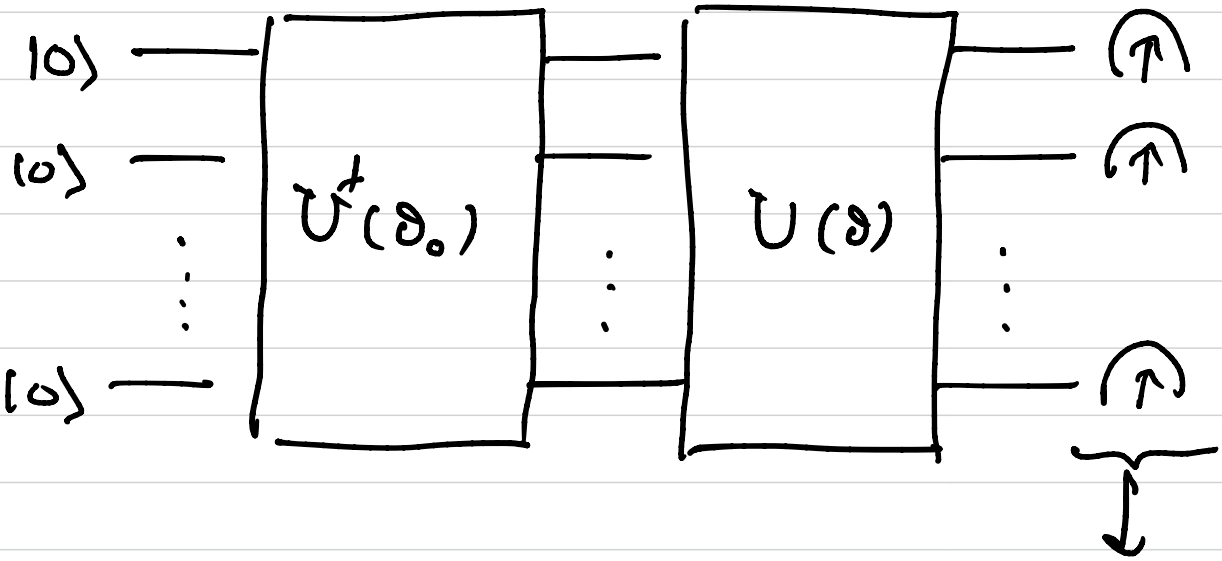
$$\langle 0 | U^\dagger(\theta) H U(\theta) | 0 \rangle - \lambda \left| \langle 0 | U^\dagger(\theta_0) U(\theta) | 0 \rangle \right|^2$$

What are the circuits to compute this objective fct $\mathcal{L}(\theta, d)$?

For the first term of $\mathcal{L}(\theta, d)$ we proceed as before. For the second term we note it is given by $\underbrace{|\langle 0 | \Phi \rangle|^2}$ where $|\Phi\rangle = U^\dagger(\theta_0) U(\theta)|0\rangle$

probability to measure $|0\rangle$ if the state is $|\Phi\rangle$.

So we can use the circuit:



postselect $|0\rangle$ and get fraction $|\langle 0|\Phi\rangle|^2$.

III. Optimization with gradient descent steps & parameter shift rule.

As discussed in previous paragraphs our focus is a state of the form

$$|\psi(\theta)\rangle = U_\ell(\theta_\ell) \cdots U_2(\theta_2) U_1(\theta_1) |0\rangle$$

and loss fcts of the form

$$\mathcal{L}(\theta) = \langle \psi(\theta)| H |\psi(\theta)\rangle$$

where H is a hermitian operator (typically a polynomial or macromial of Pauli matrices), Here $\theta = (\theta_1, \theta_2 \ldots, \theta_\ell)$ and each θ_i may be a vector. Typically we are trying to minimize a landscape (or objective) with many parameters which is not convex. This is a difficult optimization problem. The usual approach is to use a GD or Stochastic GD

(even though this doesn't guarantee that we find the true global minimum).

Typical VQE or QAOA or Alt Method Algorithm

1) At iteration i, use the quantum computer to estimate $\mathcal{L}(\vartheta^i)$ with the current value of variational parameters ϑ^i and estimate also gradients $G_k(\vartheta^i) = \nabla_{\vartheta_k} \mathcal{L}(\vartheta^i)$.

2) Update the parameter values according to a GD step: $\vartheta_k^{i+1} = \vartheta_k^i - \eta\, G_k(\vartheta^i)$

where η is a "small" learning rate.

3) Iterate until convergence according to a suitable criterion.

We have seen how to compute the loss function by using measurement with suitable circuits. One can proceed similarly for gradients but in specific applications there are more efficient strategies to do so.

Parameter shift rule.

Typically we have $U_k(\theta_k) = e^{-i S_k \frac{\theta_k}{2}}$ where S_k is a Hermitian operator which is an involution: $S_k^2 = \mathbb{1}$.

for example $S_k \in \{I, X, Y, Z\}$ or

$S_k \in \{ X_i \otimes X_j, Y_i \otimes Y_j, Z_i \otimes Z_j \}$ ecc...

For involutions we have a generalized form of Euler's formula:

$$e^{-i S_k \frac{\theta_k}{2}} = \left(\cos\frac{\theta_k}{2}\right)\mathbb{1} - i\left(\sin\frac{\theta_k}{2}\right) S_k$$

You can check this formula using $S_k^2 = \mathbb{1}$ and the Taylor expansion.

Thus for any operator K we have

$$K(\theta_k) \equiv U_k^\dagger(\theta_k)\, K\, U_k(\theta_k)$$
$$= \hat{A} + \hat{B}\cos\theta_k + \hat{C}\sin\theta_k .$$

(Exercise: compute operators $\hat{A}$, $\hat{B}$, $\hat{C}$ which are independent of θ_k).

$$\Rightarrow \nabla_{\theta_k} K(\theta_k) = -\hat{B}\sin\theta_k + \hat{C}\cos\theta_k$$
$$= \frac{1}{2}\left[K\left(\theta_k + \tfrac{\pi}{2}\right) - K\left(\theta_k - \tfrac{\pi}{2}\right)\right]$$
$$= \frac{1}{2}\Big[U^\dagger\left(\theta_k + \tfrac{\pi}{2}\right) K\, U\left(\theta_k + \tfrac{\pi}{2}\right)$$
$$- U^\dagger\left(\theta_k - \tfrac{\pi}{2}\right) K\, U\left(\theta_k - \tfrac{\pi}{2}\right)\Big]$$

Thus if we have a loss of the form :

$$\mathcal{L}(\theta) = \langle 0 | W_k^\dagger U^\dagger(\theta_k) V_k^\dagger H V_k U(\theta_k) W_k | 0 \rangle$$

where $W_k = U_{k-1}(\theta_{k-1}) \cdots U_1(\theta_1)$

$V_k = U_\ell(\theta_\ell) \cdots U_{k+1}(\theta_{k+1})$

<u>are independent of θ_k,</u>

we get something of the form :

$$\nabla_{\theta_k} \mathcal{L}(\theta) = \nabla_{\theta_k} \langle \psi_k | U^\dagger(\theta_k) K_k U(\theta_k) | \psi_k \rangle$$

$$= \langle \psi_k | \underbrace{\nabla_{\theta_k} \left(U^\dagger(\theta_k) K_k U(\theta_k) \right)}_{} | \psi_k \rangle$$

$$= \langle \psi_k | \frac{1}{2} \Big[U^\dagger(\theta_k + \tfrac{\pi}{2}) K_k U(\theta_k + \tfrac{\pi}{2}) - U^\dagger(\theta_k - \tfrac{\pi}{2}) K_k U(\theta_k - \tfrac{\pi}{2}) \Big] | \psi_k \rangle$$

$$= \frac{1}{2} \Big[\mathcal{L}(\theta_1 \cdots \theta_{k-1}, \theta_k + \tfrac{\pi}{2}, \theta_{k+1} \cdots \theta_\ell) - \mathcal{L}(\theta_1 \cdots \theta_{k-1}, \theta_k - \tfrac{\pi}{2}, \theta_{k+1} \cdots \theta_\ell) \Big]$$

This formula is very gratifying as it tells us that to compute a gradient instead for varying infinitesimally θ_k we can just compute the loss at two points $\theta_k + \frac{\pi}{2}$ & $\theta_k - \frac{\pi}{2}$.

Thus when we implement VQE in this context it is enough to run step 1 above for θ_k, $\theta_k + \frac{\pi}{2}$ and $\theta_k - \frac{\pi}{2}$. Step 2 is then immediate.

In conclusion the parameter shift rule and similar ideas can often simplify greatly the implementation of hybrid classical-quantum optimization algorithms.

■