Machine Learning for VQE

Son Nguyen

Optimizing parameters

We have a quantum state $|\psi(\vec{\theta})\rangle$, where $\vec{\theta} = (\theta_1, \theta_2, \theta_3, \ldots, \theta_n)$ are the parameters of the quantum circuit. The goal is to find the optimal θ that minimizes the expectation value of the Hamiltonian H.

The cost function to minimize the expectation value (energy) of the Hamiltonian is given by:

$$E(\vec{\theta}) = \langle\psi(\vec{\theta})|H|\psi(\vec{\theta})\rangle$$

The optimizer's goal is to find the set of parameters $\vec{\theta}$ that $E(\vec{\theta})$.

Gradient-based optimization

Gradient descent is a common type of optimizer. It uses the gradient of the cost function with respect to the parameters:

$$\theta_i^{(t+1)} = \theta_i^{(t)} - \eta \nabla E(\vec{\theta}^{(t)})$$

Where:

- η is the learning rate.
- $\theta_i^{(t)}$ is the parameter at iteration t.
- $\nabla E(\vec{\theta}^{(t)})$ is the gradient of the cost function with respect to the parameter θ_i.

The optimizer will stop when $|E(\vec{\theta}^{(t+1)}) - E(\vec{\theta}^{(t)})|$ is smaller than threshold ϵ.

The final parameters $\vec{\theta}^*$ will be those that minimzie the energy. The ansatz state $|\psi(\vec{\theta^*})\rangle$ approximates the ground state of the Hamiltonian.

Parameter-Shifting Rule for parameter θ_i:

$$\nabla E(\vec{\theta}^{(t)}) = \left(\frac{\partial E(\vec{\theta})}{\partial \theta_0}, \frac{\partial E(\vec{\theta})}{\partial \theta_1}, \frac{\partial E(\vec{\theta})}{\partial \theta_2}, \ldots, \frac{\partial E(\vec{\theta})}{\partial \theta_n} \right)$$

$$\frac{\partial E(\vec{\theta})}{\partial \theta_i} \approx \frac{E\left(\vec{\theta} + s_k\right) - E\left(\vec{\theta} - s_k\right)}{2 s_k}$$

where s_k is a unit shift in the k-th parameter, i.e., $s_k = (0, \ldots, 0, \text{sk}, 0, \ldots, 0)$ with $s_k \neq 0$. The parameter-shift rule allows us to estimate the gradient using the difference between the cost function evaluated at two nearby points in parameter space.

The gradient-based algorithm rely on direct measurement of the gradient of the loss function with respect to the parameters being optimized. These measurement typically yeild an estimate of the gradient because the underlying data usually include added noise which is inherent to the quantum system.

SPSA - simultaneous perturbation stochastic approximation

SPSA is a gradient-free optimization method that estimates the gradient using a finite difference method. It is robust to noise and does not require the gradient to be calculated analytically. While executing a variational algorithm using a Quantum Assembly Language (QASM) simulator or a real device, SPSA would be the most recommended choice among the optimizers.

1. Initialize the parameters $\vec{\theta}$.
2. For each iteration:
   (a) Randomly choose a perturbation vector $\vec{\Delta}$ with elements ± 1 (Bernoulli distribution with probability of $\frac{1}{2}$ for each ± 1 outcome).
   (b) Perturbation Step. Let $\vec{\theta}^{(t)}$ be the parameter vector at iteration t. The pertubed parameter vectors are:

$$\vec{\theta}_+^{(t)} = \vec{\theta}^{(t)} + c^{(t)}\Delta\vec{\theta}^{(t)}$$

$$\vec{\theta}_-^{(t)} = \vec{\theta}^{(t)} - c^{(t)}\Delta\vec{\theta}^{(t)}$$

   where $c^{(t)}$ is is a small perturbation factor that decreases over iterations.
   (c) The quantum circuit is executed twice for each iteration, once with $\vec{\theta}_+^{(t)}$ and once with $\vec{\theta}_-^{(t)}$. The expectation value of the Hamiltonian is measured for each circuit execution.

$$E(\vec{\theta}_+^{(t)}) = \langle\psi(\vec{\theta}_+^{(t)})|H|\psi(\vec{\theta}_+^{(t)})\rangle$$

$$E(\vec{\theta}_-^{(t)}) = \langle\psi(\vec{\theta}_-^{(t)})|H|\psi(\vec{\theta}_-^{(t)})\rangle$$

   (d) Gradient approximation. SPSA approximates the gradient for each parameter based on the difference in energy between the 2 perturbed states:

$$\hat{g}_i^{(t)} = \frac{E(\vec{\theta}_+^{(t)}) - E(\vec{\theta}_-^{(t)})}{2c^{(t)}\Delta\theta_i^{(t)}}$$

The entire gradient vector $\hat{g}^{(t)}$ is computed with just 2 evaluations of the cost function, regardless of the number of parameters.

(e) Update the parameters:

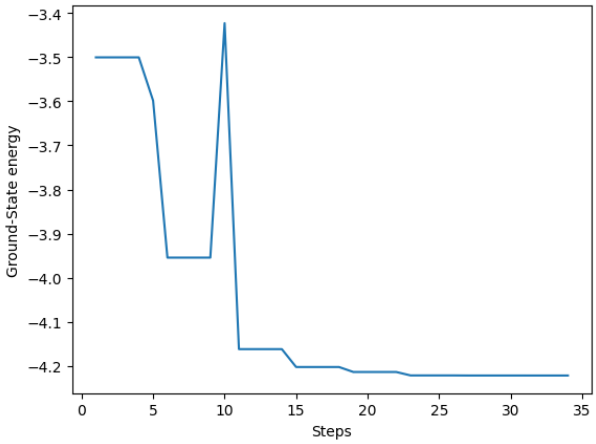
$$\vec{\theta}^{(t+1)} = \vec{\theta}^{(t)} - a^{(t)}\hat{g}^{(t)}$$

where $a^{(t)}$ is the learning rate that decreases over iterations.
Parmaters Schedule:

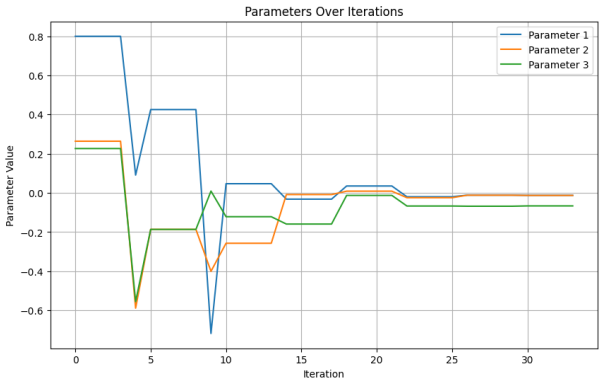
$$a^{(t)} = \frac{a}{(A+t+1)^{\alpha}}$$

$$c^{(t)} = \frac{c}{(t+1)^{\gamma}}$$

where A, α, γ control the decay of the learning rate and perturbation factor.



(a) Optimizing the Cost Function (Energy)



(b) Parmaters change over iterations

Figure 1: The optimization process using SPSA method (HeH+ molecule)

References

- Simultaneous Perturbation Stochastic Approximation
- Quantum Gradient Descent