Doubly stochastic gradient descent

EEE 606: Adaptive Signal Processing

Rajesh Sathya Kumar, M.S. Computer Science, On-Campus

Overview

Problem Statement

- ✓ Understand and implement stochastic gradient descent in quantum computation
- ✓ Compare that will classical (vanilla) gradient descent

Objectives

1. Variational Quantum Eigensolver (VQE)

- ✓ Implement Quantum doubly stochastic gradient descent (QSGD) using pennylane for VQE
- ✓ Compare the convergence curve with vanilla gradient descent

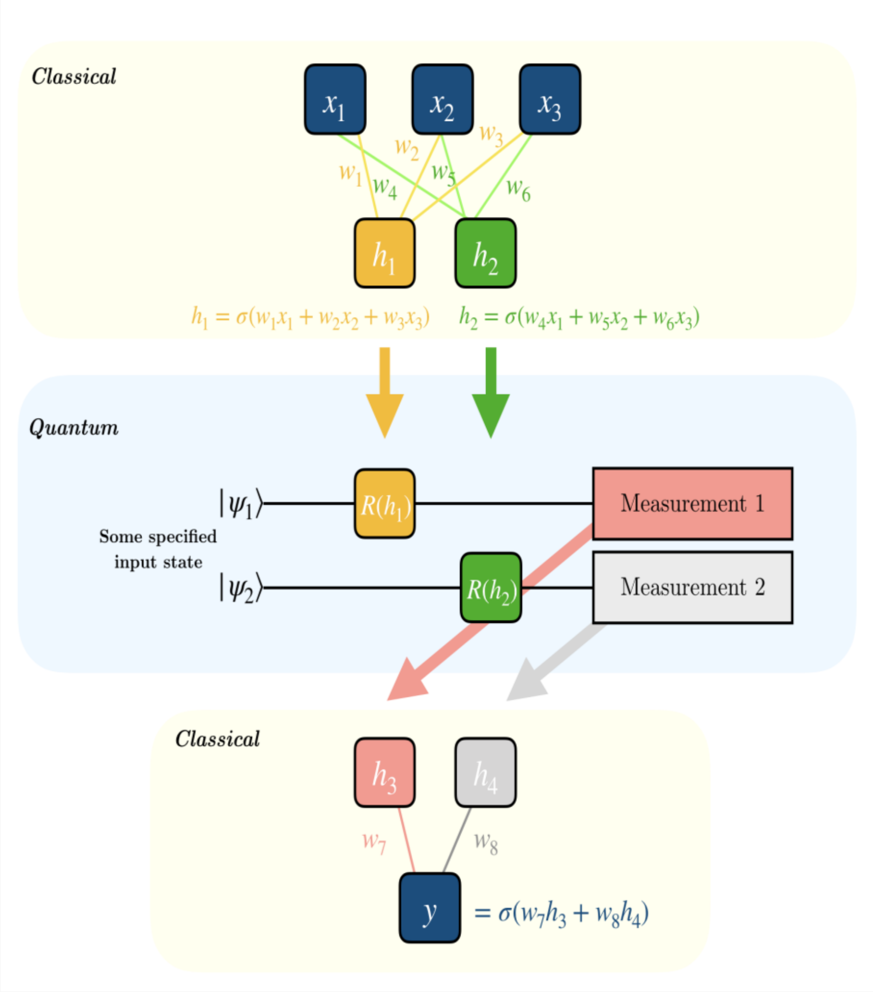
2. Hybrid Neural Network + QSGD

- ✓ Implement a Neural network with a Quantum Convolutional Layer using QSGD optimizer using pennylane
- ✓ Compare its convergence curve with Vanilla GD

Note: The recorded video will discuss about the theory and results of QSGD for the above problems

Main Reference Paper:

1. Ryan Sweke, Frederik Wilde, Johannes Jakob Meyer, Maria Schuld, Paul K. Fährmann, Barthélémy Meynard-Piganeau, Jens Eisert. **"Stochastic gradient descent for hybrid quantum-classical optimization."** arXiv:1910.01155, 2019.



General structure of a Hybrid quantum neural network

Preliminary Results

Theory: VQE:

We are given a Hamiltonian (Energy operator of a system) and we are trying to minimize the energy of the system using QSGD.

Gradient Update rule:

$$\theta^{(t+1)} = \theta^{(t)} - \eta \nabla \mathcal{L}(\theta^{(t)}),$$

Expectation of an operator

$$\langle A_i \rangle = \langle 0 | U(\theta)^\dagger A_i U(\theta) | 0 \rangle$$

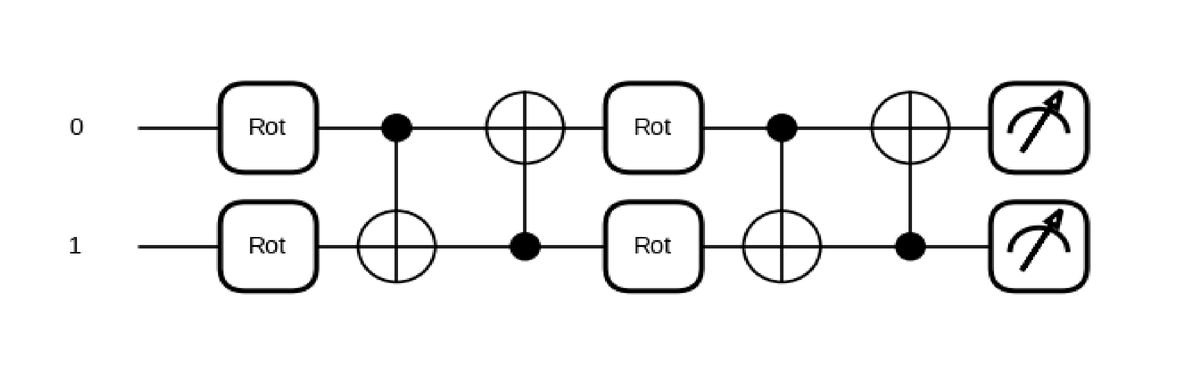
Loss function for a parameterized circuit

$$\mathcal{L}(\theta, \langle A_1 \rangle, \ldots, \langle A_M \rangle).$$

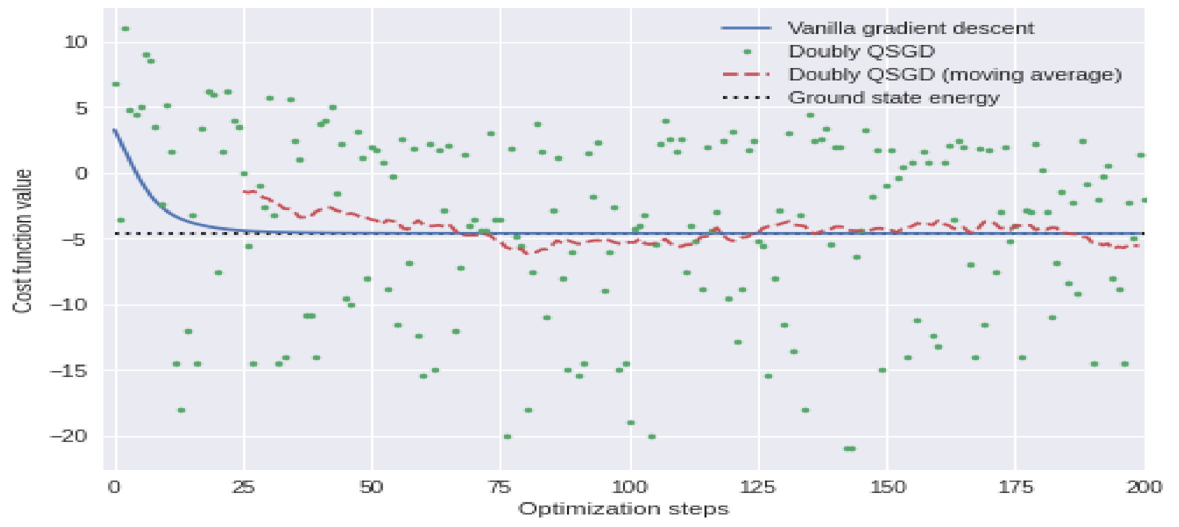
Hamiltonian as a sum of pauli matrices

$$H = \sum_{i,j=0,1,2,3} a_{i,j} (\sigma_i \otimes \sigma_j),$$

Variational Quantum Circuit (ansatz)



QGSD vs Vanilla Gradient Descent



Conclusion

Anticipated Results

VQE

- ✓ The convergence curve is not as smooth as the vanilla gradient descent, but converges to the correct optimum

MNIST

- ✓ Similarly, we should have the results converge to an optimum with a non-smooth curve
- ✓ Accuracy might be slightly lesser than vanilla gradient descent

Major Takeaway

- ✓ Stochastic gradient descent can be successfully implemented using quantum circuits
- ✓ The QSGD algorithm optimizes to a local minima, but the convergence curve is not as smooth as the vanilla gradient descent
- ✓ We have to use moving averages to properly minimize the loss function

My Learnings

- ✓ Understood the theory of stochastic gradient descent
- ✓ Learnt how to encode a stochastic optimizer for VQE problem