Quantum Circuit Learning

BY K.MITARAI

I. Introduction.

1. In this paper, we present a new hybrid framework, which we call quantum circuit learning (QCL), for machine learning with a low-depth quantum circuit.

2. In QCL, we provide input data to a quantum circuit, and iteratively tunes the circuit parameters so that it gives the desired output.

3. Gradient-Based systematic optimization of parameters is introduced for the tuning just like back-propagation method utilized in feedforward neural networks.

4. We theoretically show that a quantum circuit driven by the QCL framework can approximate any analytical function if the circuit has a sufficient number of qubits.

5. The ability of the QCL framework to learn nonlinear functions and perform a simple classification task is demonstrated by numerical simulations.

II. Quantum Circuit Learning

A. Algorithms

We summarize the QCL algorithm on N qubit circuit:

1. Encode input data $\{x_i\}$ into some quantum state $|\psi_{\text{in}}(x_i)\rangle$ by applying a unitary input gate $U(x_i)$[1] to initialized qubits $|0\rangle$.

2. Apply a θ-parameterized unitary $U(\theta)$ to the input state and generate an output state $|\psi_{\text{out}}(x_i, \theta) = U(\theta)|\psi_{\text{in}}(x_i)\rangle$.[2]

3. Measure the expectation values of some chosen observables. Specifically, we use a subset of Pauli operators[3] $\{B_j\} \subset \{I, X, Y, Z\}^{\otimes N}$.[4] Using some output function F,[5] output $y_i = y(x_i, \theta)$ is defined to be $y(x_i, \theta) \equiv F(\{\langle B_j(x_i, \theta)\rangle\})$.

4. Minimize the cost function $L(f(x_i), y(x_i, \theta))$[6] of the teacher $f(x_i)$[7] and the output y_i, by tuning the circuit parameters θ iteratively.

5. Evaluate the performance by checking the cost function with respect to a data set that is taken independently from the training one.

B. Relation with existing algorithms

1. How can we construct the unitary input gate $U(x_i)$?
2. $U(\theta)$ is the ansatz; $|\psi_{\text{in}}(x_i)\rangle$ is the input state.
3. Pauli matrices or their tensor products are called Pauli operators.
4. ? Consider the case $N = 2$, $\{I, X, Y, Z\}^{\otimes 2} = \{I \otimes I, X \otimes X, Y \otimes Y, Z \otimes Z, I \otimes X, X \otimes I, I \otimes Y, Y \otimes I, I \otimes Z,, Z \otimes I, X \otimes Y, Y \otimes X, X \otimes Z, Z \otimes X, Y \otimes Z, Z \otimes Y\}$ and choose its subset $\{I \otimes I, X \otimes X, Y \otimes Y\}$ as the observables. Now we have the ouput state $|\psi_{\text{out}}(x_i, \theta)\rangle$ and its expectation of each observable is

$$\langle\psi_{\text{out}}(x_i, \theta)|I \otimes I|\psi_{\text{out}}(x_i, \theta)\rangle, \langle\psi_{\text{out}}(x_i, \theta)|X \otimes X|\psi_{\text{out}}(x_i, \theta)\rangle, \langle\psi_{\text{out}}(x_i, \theta)|Y \otimes Y|\psi_{\text{out}}(x_i, \theta)\rangle$$

We choose some proper function $F(\cdot, \cdot, \cdot)$ to define the output as

$$y(x_i, \theta) = F(\psi_{\text{out}}(x_i, \theta)|I \otimes I|\psi_{\text{out}}(x_i, \theta)\rangle, \langle\psi_{\text{out}}(x_i, \theta)|X \otimes X|\psi_{\text{out}}(x_i, \theta)\rangle, \langle\psi_{\text{out}}(x_i, \theta)|Y \otimes Y|\psi_{\text{out}}(x_i, \theta)\rangle).$$

5. Like the quadratic function? No, general functions are needed.
6. $L(\cdot, \cdot)$ is the cost function, like the quadratic cost.
7. True label in the supervised learning setting up.

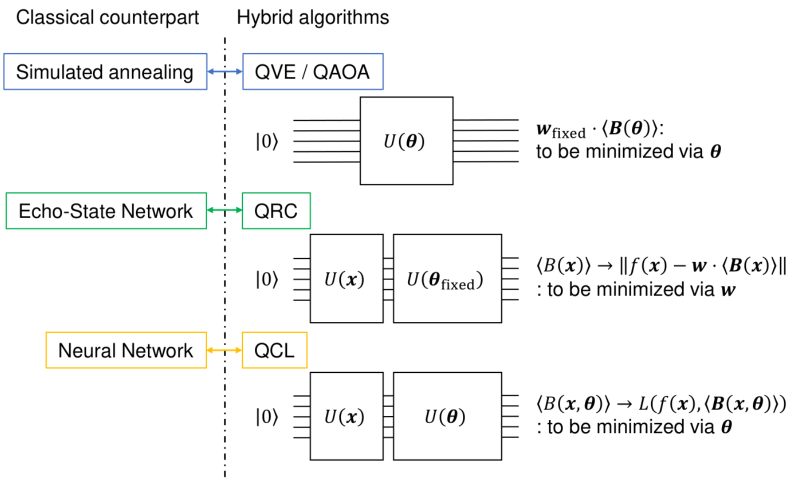

FIG. 1. Comparison of QVE[8]/QAOA[9], QRC[10], and presented QCL[11] framework. In QVE, the output of the quantum circuit is directly minimized. QRC and QCL both optimize the output to the teacher $f(x)$. QRC optimization is done via tuning the linear weight w, as opposed to QCL approach which tunes the circuit parameter θ.

C. Ability to approximate a function

First, we consider the case where input data are one dimension for simplicity. It is straightforward to generalize the following argument for higher dimensional inputs.

Let x and $\rho_{\text{in}}(x) = |\psi_{\text{in}}(x)\rangle\langle\psi_{\text{in}}(x)|$ be an input data and a corresponding density operator of input state. $\rho_{\text{in}}(x)$ can be expanded by a set of Pauli operators $\{P_k\} = \{I, X, Y, Z\}^{\otimes N}$ with $a_k(x)$ as coefficients, $\rho_{\text{in}}(x) = \sum_k a_k(x) P_k$.[12] A parameterized unitary transformation $U(\theta)$ acting on $\rho_{\text{in}}(x)$ creates the output state, which can also be expanded by $\{P_k\}$ with $\{b_k(x, \theta)\}$.[13]

Now let $u_{ij}(\theta)$ be such that $b_m(x, \theta) = \sum_k u_{mk}(\theta) a_k(x)$.[14] b_m is an expectation value of a Pauli observable itself[15], therefore, the output is linear combination of input coefficient functions a_k under unitarity constraints imposed on $\{u_{ij}\}$. ???

When the teacher $f(x)$ is an analytical function, we can show, at least in principle, QCL is able to approximate it by considering a simple case with an input state created by single-qubit rotations.

The tensor product structure of quantum system plays an important role in this analysis.

Let us consider a state of N qubits:

$$\rho_{\text{in}}(x) = \frac{1}{2^N} \bigotimes_{i=1}^{N} \left[I + x X_i + \sqrt{1 - x^2} Z_i \right]. \quad (1)$$

This state can be generated for any $x \in [-1, 1]$ with single-qubit rotations, namely, $\Pi_{i=1}^{N} R_i^Y(\sin^{-1} x)$, where $R_i^Y(\phi)$[16] is the rotation of ith qubit around y axis with angle ϕ.

8. Quantum variational eigensolver. More details: $\langle B(\theta)\rangle = (\langle B_1(\theta)\rangle, \langle B_2(\theta)\rangle, \ldots)$ is a vector, where $\langle B_1(\theta)\rangle$ is the expectation of the observable B_1, i.e., $\langle B_1(\theta)\rangle = \langle \psi_{\text{in}}(\theta)|B_1|\psi_{\text{in}}(\theta)\rangle$, where $|\psi_{\text{in}}(\theta)\rangle = U(\theta)|0\rangle$; $w_{\text{fixed}} \cdot \langle B(\theta)\rangle$ is the inner product between w_{fixed} and $\langle B(\theta)\rangle$.

9. Quantum approximation optimization algorithm. More details: $\|f(x) - w \cdot \langle B(x)\rangle\|$ is minimized by updating w.

10. Quantum reservior computing

11. Quantum circuit learning. More details: $L(f(x), \langle B(x, \theta)\rangle$ should be $L(f(x), F(\langle B(x, \theta)\rangle))$ where $\langle B(x, \theta)\rangle = (\langle B_1(x, \theta)\rangle, \langle B_2(x, \theta)\rangle, \ldots)$.

12. Decompose the density function of the input data into the linear combination of Pauli operators.

13. The output state is $|\psi_{\text{out}}(x, \theta)\rangle = U(\theta)|\psi_{\text{in}}(x)\rangle$ and it density function $\rho_{\text{out}}(x, \theta) = |\psi_{\text{out}}(x, \theta)\rangle\langle\psi_{\text{out}}(x, \theta)|$ can be expanded as $\rho_{\text{out}}(x, \theta) = \sum_k b_k(x, \theta) P_k$.

14. How do we know such $u_{ij}(\theta)$ exist? The linear system $a_1(x) u_{m1}(\theta) + a_2(x) u_{m2}(\theta) + \cdots + a_{\text{finite}}(x) u_{m,\text{finite}}(\theta) = b_m(x, \theta)$ has one equation but more unknowns so it always has a solution.

15. b_m is $b_m(x, \theta)$. How to understand this sentence?

16. $R_i^Y(\phi) = \begin{bmatrix} \cos\frac{\phi}{2} & -\sin\frac{\phi}{2} \\ \sin\frac{\phi}{2} & \cos\frac{\phi}{2} \end{bmatrix}$. Also, $\prod_{i=1}^{N} R_i^Y(\sin^{-1} x)$ should be $\bigotimes_{i=1}^{N} R_i^Y(\sin^{-1} x)$.

Remark. For a single qubit, the input state is $R^Y(\phi)|0\rangle = \begin{bmatrix} \cos\frac{\phi}{2} \\ \sin\frac{\phi}{2} \end{bmatrix}$,where $R^Y(\phi) = \begin{bmatrix} \cos\frac{\phi}{2} & -\sin\frac{\phi}{2} \\ \sin\frac{\phi}{2} & \cos\frac{\phi}{2} \end{bmatrix}$. Then the density matrix of this qubit is

$$\begin{aligned}
R^Y(\phi)|0\rangle(R^Y(\phi)|0\rangle)^\dagger &= \begin{bmatrix} \cos\frac{\phi}{2} \\ \sin\frac{\phi}{2} \end{bmatrix}\begin{bmatrix} \cos\frac{\phi}{2} & \sin\frac{\phi}{2} \end{bmatrix} \\
&= \begin{bmatrix} \cos^2\frac{\phi}{2} & \cos\frac{\phi}{2}\sin\frac{\phi}{2} \\ \sin\frac{\phi}{2}\cos\frac{\phi}{2} & \sin^2\frac{\phi}{2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1+\cos\phi}{2} & \frac{1}{2}\sin\phi \\ \frac{1}{2}\sin\phi & \frac{1-\cos\phi}{2} \end{bmatrix} \\
&= \frac{1}{2}\begin{bmatrix} 1+\cos\phi & \sin\phi \\ \sin\phi & 1-\cos\phi \end{bmatrix} \\
&= \frac{1}{2}I + \frac{1}{2}\begin{bmatrix} \cos\phi & \sin\phi \\ \sin\phi & -\cos\phi \end{bmatrix} \\
&= \frac{1}{2}(I + \sin\phi X + \cos\phi Z)
\end{aligned}$$

In our case $\phi = \sin^{-1}x$, so $\sin\phi = \sin(\sin^{-1}x) = x$. Let $y = \sin^{-1}x \in [-\pi/2, \pi/2]$. Then $\sin y = x$ and $\cos(\sin^{-1}x) = \cos y = \sqrt{1-\sin^2 y} = \sqrt{1-x^2}$. Therefore,

$$R^Y(\sin^{-1}x)|0\rangle(R^Y(\sin^{-1}x)|0\rangle)^\dagger = \frac{1}{2}\Big(I + xX + \sqrt{1-x^2}Z\Big)$$

and consequently (N qubits),

$$\rho_{\text{in}}(x) = \frac{1}{2^N}\bigotimes_{i=1}^N \left[I + xX_i + \sqrt{1-x^2}Z_i\right]$$

The state given by Eq. (1) has higher order terms up to the Nth with respect to x. Thus an arbitrary unitary transformation on this state can provide us with an arbitrary Nth order polynomial as expectation values of an observable. Terms like $x\sqrt{1-x^2}$ in Eq. (1) can enhance its ability to approximate a function.

Remark.

First, we assume $x \in [0, 1]$ and consider only $R^Y(\phi) = \begin{bmatrix} \cos\frac{\phi}{2} & -\sin\frac{\phi}{2} \\ \sin\frac{\phi}{2} & \cos\frac{\phi}{2} \end{bmatrix}$ embedding data x where $\phi = \sin^{-1}x$.

The case of single qubit

Step 1 embedding data

$$|\psi_{\text{in}}(x)\rangle = R^Y(\sin^{-1}x)|0\rangle$$

Step 2 pass through quantum circuit $U(\theta)$

$$|\psi_{\text{out}}(x)\rangle = U(\theta)|\psi_{\text{in}}(x)\rangle = UR^Y(\sin^{-1}x)|0\rangle$$

Step 3 Measure the qubit by an observale O

$$\langle\psi_{\text{out}}(x)|U|\psi_{\text{out}}(x)\rangle = \langle\psi_{\text{in}}(x)|U(\theta)^\dagger OU(\theta)|\psi_{\text{in}}(x)\rangle$$

Let $U = U(\theta)^\dagger OU(\theta)$. Then

$$\begin{aligned}
\langle U\rangle &= \langle\psi_{\text{in}}(x)|U|\psi_{\text{in}}(x)\rangle \\
&= (R^Y(\sin^{-1}x)|0\rangle)^\dagger UR^Y(\sin^{-1}x)|0\rangle
\end{aligned}$$

$$
\begin{aligned}
&= \begin{bmatrix} \cos\left(\frac{\sin^{-1}x}{2}\right) & \sin\left(\frac{\sin^{-1}x}{2}\right) \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \begin{bmatrix} \cos\left(\frac{\sin^{-1}x}{2}\right) \\ \sin\left(\frac{\sin^{-1}x}{2}\right) \end{bmatrix} \\
&= u_{11}\cos^2\left(\frac{\sin^{-1}x}{2}\right) + u_{22}\sin^2\left(\frac{\sin^{-1}x}{2}\right) + (u_{12}+u_{21})\cos\left(\frac{\sin^{-1}x}{2}\right)\sin\left(\frac{\sin^{-1}x}{2}\right) \\
&= u_{11}\frac{1+\sqrt{1-x^2}}{2} + u_{22}\frac{1-\sqrt{1-x^2}}{2} + (u_{12}+u_{21})\frac{x}{2} \\
&= \frac{u_{12}+u_{21}}{2}x + \frac{u_{11}-u_{22}}{2}\sqrt{1-x^2} + \frac{u_{11}+u_{22}}{2}
\end{aligned}
$$

Step 4 Apply the classicial transformation $a\langle U\rangle + b$

$$\frac{a(u_{12}+u_{21})}{2}x + \frac{a(u_{11}-u_{22})}{2}\sqrt{1-x^2} + \frac{a(u_{11}+u_{22})}{2} + b$$

which can only express first order polynomials[17].

The case of two qubits, for simplicity of calculations, we assume $U = U(\theta)^{\dagger} O U(\theta)$ can be decomposed $U = U_1 \otimes U_2$. Then

$$
\begin{aligned}
& \langle U\rangle \\
=\; & \langle \psi_{\text{in}}(x)|U_1 \otimes U_2|\psi_{\text{in}}(x)\rangle \\
=\; & (R_1^Y \otimes R_2^Y|00\rangle)^{\dagger}(U_1 \otimes U_2)R_1^Y \otimes R_2^Y|00\rangle \\
=\; & (R_1^Y|0\rangle)^{\dagger}U_1R_1^Y|0\rangle \otimes (R_2^Y|0\rangle)^{\dagger}U_2R_2^Y|0\rangle \\
& (R_1^Y|0\rangle)^{\dagger}U_1R_1^Y|0\rangle, (R_2^Y|0\rangle)^{\dagger}U_2R_2^Y|0\rangle \in \mathbb{C} \\
=\; & (R_1^Y|0\rangle)^{\dagger}U_1R_1^Y|0\rangle(R_2^Y|0\rangle)^{\dagger}U_2R_2^Y|0\rangle \\
=\; & (R_1^Y(\sin^{-1}x)|0\rangle)^{\dagger}U_1R_1^Y(\sin^{-1}x)|0\rangle(R_2^Y(\sin^{-1}x)|0\rangle)^{\dagger}U_2R_2^Y(\sin^{-1}x)|0\rangle \\
=\; & \left(u_{11}^1\frac{1+\sqrt{1-x^2}}{2} + u_{22}^1\frac{1-\sqrt{1-x^2}}{2} + (u_{12}^1+u_{21}^1)\frac{x}{2} \right)\left(u_{11}^2\frac{1+\sqrt{1-x^2}}{2} + u_{22}^2\frac{1-\sqrt{1-x^2}}{2} + (u_{12}^2+u_{21}^2)\frac{x}{2} \right)
\end{aligned}
$$

where $U_1 = (\, u_{ij}^1 \,), U_2 = (\, u_{ij}^2 \,)$ are the parameterized quantum circuits.

The quantum circuits can only approximate second order polynomials.

Error analysis

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + o(x^5)$$

$$\sqrt{1-x^2} = 1 - \frac{1}{2}x^2 - \frac{1}{4}x^4 + o(x^4)$$

$$ax + b\sqrt{1-x^2} + c = b + c + ax - \frac{b}{2}x^2 - \frac{b}{4}x^4 + o(x^4)$$

$$\Rightarrow b + c = 0, a = 1 \Rightarrow c = -b, a = 1$$

$$
\begin{aligned}
\sin x - x - b\sqrt{1-x^2} - b &= x - \frac{x^3}{3!} + \frac{x^5}{5!} + o(x^5) - \left[x - \frac{b}{2}x^2 - \frac{b}{4}x^4 + o(x^4) \right] \\
&= \frac{b}{2}x^2 + \frac{b}{4}x^4 + o(x^4) - \frac{x^3}{3!} + \frac{x^5}{5!} + o(x^5) \\
&= \left(\frac{b}{2} - \frac{x}{3!}\right)x^2 + \left(\frac{b}{4} - \frac{x}{5!}\right)x^4 + o(x^4) \\
&\approx O(x^2)
\end{aligned}
$$

Important notice in the example given above is that the highest order term x^N is hidden in an observable $X^{\otimes N}$. To extract x^N from Eq. (1), one needs to transfer the nonlocal observable $X^{\otimes N}$

17. Roughly, regard $\sqrt{1-x^2}$ as the first order polynomial.

to a single-qubit observable using entangling gate such as the controlled-NOT gate. ???

Entangling nonlocal operations are the key ingredients of the nonlinearity of an output. ???

The above argument can readily be generalized to multi-dimensional inputs. Assume that we are given with d-dimensional data $x=\{x_1, x_2, \ldots, x_d\}$ and want higher terms up to the n_kth $(k=1,\ldots,d)$ for each data, then encode this data into a $N=\sum_k n_k$-qubit quantum state as $\rho_{\text{in}}(x)=\frac{1}{2^N}\bigotimes_{k=1}^{d}\left(\bigotimes_{i=1}^{n_k}\left[I+x_kX_i+\sqrt{1-x_k^2}Z_i\right]\right)$. These input states automatically has an exponentially large number of independent functions as coefficient set to the number of qubits. The tensor product structure of quantum system readily "calculates" the product such as x_1x_2.

Remark. In our application, we only need 2 qubits, i.e., $N=2$, since the objective is of order 2 in basis functions $\sin x$ and $\cos x$, and the input data is 100 dimensional, i.e., $\delta=(\,\delta_1\ \ \delta_2\ \ \ldots\ \ \delta_{100}\,)$. Then $n_1+n_2+\cdots+n_{100}=N=2$. Since the basis functions we need are $\sin x$ and $\cos x$, we don't need the operation $\sin^{-1}\delta$.

(1) one qubit and one dimensional data: the qubit $R^Y(\delta)|0\rangle$ and an observable U, then

$$\begin{aligned}\rho_{\text{in}}(\delta) &= |\psi_{\text{in}}(\delta)\rangle\langle\psi_{\text{in}}(\delta)| \\ &= R^Y(\delta)|0\rangle(R^Y(\delta)|0\rangle)^\dagger \\ &= \begin{bmatrix}\cos\left(\frac{\delta}{2}\right)\\ \sin\left(\frac{\delta}{2}\right)\end{bmatrix}\begin{bmatrix}\cos\left(\frac{\delta}{2}\right) & \sin\left(\frac{\delta}{2}\right)\end{bmatrix} \\ &= \begin{bmatrix}\cos^2\left(\frac{\delta}{2}\right) & \cos\left(\frac{\delta}{2}\right)\sin\left(\frac{\delta}{2}\right)\\ \sin\left(\frac{\delta}{2}\right)\cos\left(\frac{\delta}{2}\right) & \sin^2\left(\frac{\delta}{2}\right)\end{bmatrix} \\ &= \begin{bmatrix}\frac{\cos\delta+1}{2} & \frac{1}{2}\sin\delta\\ \frac{1}{2}\sin\delta & \frac{1-\cos\delta}{2}\end{bmatrix} \\ &= \frac{1}{2}I+\frac{1}{2}\begin{bmatrix}\cos\delta & \sin\delta\\ \sin\delta & -\cos\delta\end{bmatrix} \\ &= \frac{1}{2}(I+\sin\delta X+\cos\delta Z)\end{aligned}$$

and

$$\begin{aligned}(R^Y(\delta)|0\rangle)^\dagger U R^Y(\delta)|0\rangle &= \begin{bmatrix}\cos\left(\frac{\delta}{2}\right) & \sin\left(\frac{\delta}{2}\right)\end{bmatrix}\begin{bmatrix}u_{11} & u_{12}\\ u_{21} & u_{22}\end{bmatrix}\begin{bmatrix}\cos\left(\frac{\delta}{2}\right)\\ \sin\left(\frac{\delta}{2}\right)\end{bmatrix} \\ &= u_{11}\cos^2\left(\frac{\delta}{2}\right)+u_{22}\sin^2\left(\frac{\delta}{2}\right)+(u_{12}+u_{21})\cos\left(\frac{\delta}{2}\right)\sin\left(\frac{\delta}{2}\right) \\ &= u_{11}\frac{\cos\delta+1}{2}+u_{22}\frac{1-\cos\delta}{2}+(u_{12}+u_{21})\frac{\sin\delta}{2} \\ &= \frac{1}{2}(u_{11}-u_{22})\cos\delta+\frac{1}{2}(u_{12}+u_{21})\sin\delta+\frac{1}{2}(u_{22}+u_{11})\end{aligned}$$

(2) two qubits and one dimensional data[18],

$$\begin{aligned}\rho_{\text{in}}(\delta) &= R_1^Y(\delta)\otimes R_2^Y(\delta)|00\rangle(R_1^Y(\delta)\otimes R_2^Y(\delta)|00\rangle)^\dagger \\ &= [R_1^Y(\delta)|0\rangle\otimes R_2^Y(\delta)|0\rangle][R_1^Y(\delta)^\dagger|0\rangle\otimes R_2^Y(\delta)^\dagger|0\rangle] \\ &= [R_1^Y(\delta)|0\rangle R_1^Y(\delta)^\dagger|0\rangle]\otimes[R_2^Y(\delta)|0\rangle R_2^Y(\delta)^\dagger|0\rangle] \\ &= \frac{1}{2^2}(I+\sin\delta X_1+\cos\delta Z_1)\otimes(I+\sin\delta X_2+\cos\delta Z_2)\end{aligned}$$

18. $(A\otimes B)^\dagger=A^\dagger\otimes B^\dagger$ and $(R|\phi\rangle)^\dagger=(|\phi\rangle)^\dagger R^\dagger=\langle\phi|R^\dagger$.

and

$$
\begin{aligned}
& (R_1^Y(\delta)\otimes R_2^Y(\delta)|00\rangle)^\dagger U_1\otimes U_2 R_1^Y(\delta)\otimes R_2^Y(\delta)|00\rangle \\
= \; & [(R_1^Y(\delta)|0\rangle)^\dagger \otimes (R_2^Y(\delta)|0\rangle)^\dagger] U_1\otimes U_2 [R_1^Y(\delta)|0\rangle \otimes R_2^Y(\delta)|0\rangle] \\
= \; & [(R_1^Y(\delta)|0\rangle)^\dagger U_1 R_1^Y(\delta)|0\rangle] \otimes [(R_2^Y(\delta)|0\rangle)^\dagger U_2 R_2^Y(\delta)|0\rangle] \\
& (R_1^Y(\delta)|0\rangle)^\dagger U_1 R_1^Y(\delta)|0\rangle, (R_2^Y(\delta)|0\rangle)^\dagger U_2 R_2^Y(\delta)|0\rangle \in \mathbb{C} \\
= \; & [(R_1^Y(\delta)|0\rangle)^\dagger U_1 R_1^Y(\delta)|0\rangle][(R_2^Y(\delta)|0\rangle)^\dagger U_2 R_2^Y(\delta)|0\rangle] \\
= \; & \left[\frac{1}{2}(u_{11}^1 - u_{22})\cos\delta + \frac{1}{2}(u_{12}^1 + u_{21}^1)\sin\delta + \frac{1}{2}(u_{22}^1 + u_{11}^1)\right]\cdot \\
& \left[\frac{1}{2}(u_{11}^1 - u_{22}^1)\cos\delta + \frac{1}{2}(u_{12}^1 + u_{21}^1)\sin\delta + \frac{1}{2}(u_{22}^1 + u_{11}^1)\right]
\end{aligned}
$$

(3) $d=100, \delta = (\begin{array}{cccc} \delta_1 & \delta_2 & \ldots & \delta_{100} \end{array})^T$. For each dimension, we need order 1 polynomial in terms of $\sin\delta$ and $\cos\delta$, so we set $n_k = 1$ and $n_1 + n_2 + \cdots + n_{100} = N = 100$. Then

$$
\begin{aligned}
\rho_{\mathrm{in}}(\delta) &= \frac{1}{2^{100}} \bigotimes_{k=1}^{100} (I + \sin\delta_k X + \cos\delta_k Z) \\
&= \frac{1}{2^{100}} (I + \sin\delta_1 X + \cos\delta_1 Z) \otimes \ldots \otimes (I + \sin\delta_{100} X + \cos\delta_{100} Z)
\end{aligned}
$$

D. Possible quantum advantages

We have shown by above discussions that approximation of any analytical functions is possible with the use of nonlinearity created by the tensor product. In fact, nonlinear basis functions are crucial for many methods utilized in classical machine learning. They require a large number of basis functions to create a complex model that predicts with high precision.

However, the computational cost of learning increases with respect to the increasing number of basis functions. To avoid this problem, the so-called kernel trick method, which circumvents the direct use of a large number of them, is utilized.

In contrast, QCL directly utilizes the exponential number of functions with respect to the number of qubits to model the teacher, which is basically intractable on classical computers. This is a possible quantum advantage of our framework, which was not obvious from the previous approaches like QVE or QAOA.

Moreover, let us now argue about the potential power of QCL representing complex functions.

Suppose we want to learn the output of QCL that is allowed to use an unlimited resource in the learning process, via classical neural networks. Then it has to learn the relation between inputs and outputs of a quantum circuit, which, in general, includes universal quantum cellular automata[19]. This certainly could not be achieved using a polynomial-size classical computational resource to the size (qubits and gates) of QCL. This implies that QCL has a potential power to represent more complex functions than the classical counterpart.

Further investigations are needed including the learning costs and which actual learning problem enjoys such an advantage.

E. Optimization procedure

In QVE, it has been suggested to use gradient-free methods like Nelder-Mead. However, gradient-based methods are generally more preferred when the parameter space becomes large. In neural networks, backpropagation method, which is basically gradient descent, is utilized in the learning procedure.

To calculate a gradient of an expectation value of an observable with respect to a circuit parameter θ, suppose the unitary $U(\theta)$ consists of a chain of unitary transformations $\prod_{j=1}^{l} U_j(\theta_j)$ on a state ρ_{in} and we measure an observable B. For convenience, we use notation $U_{j:k} = U_1 \cdots U_k$. Then $\langle B(\theta)\rangle$ is given as $\langle B(\theta)\rangle = \mathrm{Tr}(B U_{l:1} \rho_{\mathrm{in}} U_{l:1}^\dagger)$.

19. 通用量子元胞自动机its mechanism could be super compicated.

III. Numerical Simulations

We demonstrate the performance of QCL framework for several prototypical machine learning tasks by numerically simulating a quantum circuit in the form of Fig. 2 with $N=6$ and $D=6$. $U\big(\theta_j^{(i)}\big)$ in Fig. 2 is an arbitrary rotation of a single qubit.

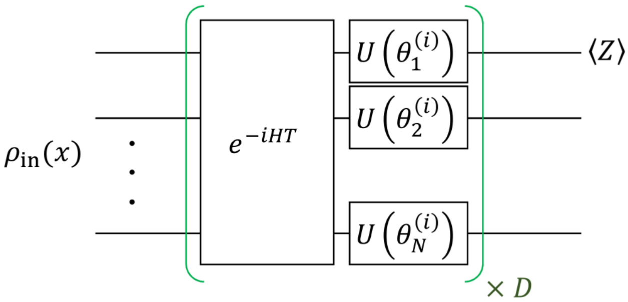

Fig.2. Quantum circuit used in numerical simulations. The parameter θ of single qubit arbitrary unitaries $U\big(\theta_j^{(i)}\big)$ are optimized to minimize the cost function. D denotes the depth of the cicuit.

Remark. The notation $\times D$ means repreating the same circuit structure D times; N is the number of total qubits; $U\big(\theta_j^{(i)}\big)$ is the unitary matrix for the jth qubit at the ith layer.

We use the decomposition $U\big(\theta_j^{(i)}\big) = R_j^X\big(\theta_{j1}^{(i)}\big) R_j^Z\big(\theta_{j2}^{(i)}\big) R^X\big(\theta_{j3}^{(i)}\big)$. H is Hamiltonian[20] of a fully connected transverse Ising model:

$$H = \sum_{j=1}^{N} a_j X_j + \sum_{j=1}^{N} \sum_{k=1}^{j-1} J_{jk} Z_j Z_k. \quad (4)$$

The coefficients a_j and J_{jk} are taken randomly from uniform distribution on $[-1,1]$. Evolution time T is fixed to 10.

The results shown throughout this section are generated by the Hamiltonian with the same coefficients. Here we note that we have checked that similar results can be achieved with different Hamiltonians.

The dynamics under this form of Hamiltonian can generate a highly entangled state and is, in general for a large number of qubits, not efficiently simulatable on a classical computer. Equation (4) is the basic form of interaction in trapped ions or superconducting qubits, which makes the time evolution easily implementable experimentally.

θ is initialized with random numbers uniformly distributed on $[0,2\pi]$. In all numerical simulations, outputs are taken from Z expectation values. To emulate a sampling, we added small Gaussian noise with standard deviation σ determined by $\sigma = \sqrt{2/N_s}(\langle Z\rangle^2 - 1)/4$, where N_s and $\langle Z\rangle$ are the number of samples and a calculated expectation value, to $\langle Z\rangle$.

First, we perform the fitting of $f(x) = x^2, e^x, \sin x, |x|, |x|$ as a demonstration of the representability of nonlinear functions [18]. We use the normal quadratic loss for the cost function. The number of teacher samples is 100. The output is taken from the Z expectation value of the **first qubit** as shown in Fig. 2. In this simulation, we allow the output to be multiplied by a constant a which is initialized to unity. This constant a and θ are simultaneously optimized. The input state $\rho_{\text{in}}(x)$ is prepared by applying $U_{\text{in}}(x) = \prod_j R_j^Z(\cos^{-1} x^2) R_j^Y(\sin^{-1} x)$ to initialized qubits $|0\rangle$. This unitary creates a state similar to Eq. (1).

20. Suppose that the $2n$-by-$2n$ matrix A is written as the block matrix $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ where a, b, c, and d are n-by-n matrices. Then the condition that A be Hamiltonian is equivalent to requiring that the matrices b and c are symmetric and $a + d^T = 0$. Another equivalent condition is that A is of the form JA is symmetric where $J = \left[\begin{smallmatrix} 0 & I_n \\ -I_n & 0 \end{smallmatrix}\right]$.

Remark. Consider the two qubit case.

$$U_{\text{in}}(x) = R_1^Z(\cos^{-1}x^2)R_1^Y(\sin^{-1}x) \otimes R_2^Z(\cos^{-1}x^2)R_2^Y(\sin^{-1}x)$$

Quantum preparation (one qubit)

$$\begin{aligned}
|\psi_{\text{in}}(x)\rangle &= R_1^Z(\cos^{-1}x^2)R_1^Y(\sin^{-1}x)|0\rangle \\
&= \left[\cos\left(\frac{\cos^{-1}x^2}{2}\right)I - i\sin\left(\frac{\cos^{-1}x^2}{2}\right)Z\right]\begin{bmatrix} \cos\left(\frac{\sin^{-1}x}{2}\right) & -\sin\left(\frac{\sin^{-1}x}{2}\right) \\ \sin\left(\frac{\sin^{-1}x}{2}\right) & \cos\left(\frac{\sin^{-1}x}{2}\right) \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} \cos\left(\frac{\cos^{-1}x^2}{2}\right) - i\sin\left(\frac{\cos^{-1}x^2}{2}\right) & 0 \\ 0 & \cos\left(\frac{\cos^{-1}x^2}{2}\right) + i\sin\left(\frac{\cos^{-1}x^2}{2}\right) \end{bmatrix}\begin{bmatrix} \cos\left(\frac{\sin^{-1}x}{2}\right) \\ \sin\left(\frac{\sin^{-1}x}{2}\right) \end{bmatrix} \\
&= \begin{bmatrix} \left(\cos\left(\frac{\cos^{-1}x^2}{2}\right) - i\sin\left(\frac{\cos^{-1}x^2}{2}\right)\right)\cos\left(\frac{\sin^{-1}x}{2}\right) \\ \left(\cos\left(\frac{\cos^{-1}x^2}{2}\right) + i\sin\left(\frac{\cos^{-1}x^2}{2}\right)\right)\sin\left(\frac{\sin^{-1}x}{2}\right) \end{bmatrix} \\
&= \begin{bmatrix} \left(\sqrt{\frac{x^2+1}{2}} - i\sqrt{\frac{1-x^2}{2}}\right)\cos\left(\frac{\sin^{-1}x}{2}\right) \\ \left(\sqrt{\frac{x^2+1}{2}} + i\sqrt{\frac{1-x^2}{2}}\right)\sin\left(\frac{\sin^{-1}x}{2}\right) \end{bmatrix}
\end{aligned}$$

Let $y = \frac{\cos^{-1}x^2}{2} \in \left[0, \frac{\pi}{2}\right]$. Then $\cos\left(\frac{\cos^{-1}x^2}{2}\right) = \cos y$ and $\cos^{-1}x^2 = 2y$ or $\cos(2y) = x^2$. Since $\cos(2y) = 2\cos^2 y - 1$ we have $2\cos^2 y - 1 = x^2$ or $\cos^2 y = \frac{x^2+1}{2}$ or $\cos y = \sqrt{\frac{x^2+1}{2}} \geq 0$. That is, $\cos\left(\frac{\cos^{-1}x^2}{2}\right) = \sqrt{\frac{x^2+1}{2}}$. Then it's easy to see $\sin\left(\frac{\cos^{-1}x^2}{2}\right) = \sqrt{\frac{1-x^2}{2}}$. Note we can not simplify $\sin\left(\frac{\sin^{-1}x}{2}\right)$ and $\cos\left(\frac{\sin^{-1}x}{2}\right)$.

Then we derive the formula for the two-qubit case

$$\begin{aligned}
|\psi_{\text{in}}\rangle &= U_{\text{in}}(x)|00\rangle \\
&= R_1^Z(\cos^{-1}x^2)R_1^Y(\sin^{-1}x) \otimes R_2^Z(\cos^{-1}x^2)R_2^Y(\sin^{-1}x)|00\rangle \\
&= R_1^Z(\cos^{-1}x^2)R_1^Y(\sin^{-1}x)|0\rangle \otimes R_2^Z(\cos^{-1}x^2)R_2^Y(\sin^{-1}x)|0\rangle \\
&= \begin{bmatrix} \left(\sqrt{\frac{x^2+1}{2}} - i\sqrt{\frac{1-x^2}{2}}\right)\cos\left(\frac{\sin^{-1}x}{2}\right) \\ \left(\sqrt{\frac{x^2+1}{2}} + i\sqrt{\frac{1-x^2}{2}}\right)\sin\left(\frac{\sin^{-1}x}{2}\right) \end{bmatrix} \otimes \begin{bmatrix} \left(\sqrt{\frac{x^2+1}{2}} - i\sqrt{\frac{1-x^2}{2}}\right)\cos\left(\frac{\sin^{-1}x}{2}\right) \\ \left(\sqrt{\frac{x^2+1}{2}} + i\sqrt{\frac{1-x^2}{2}}\right)\sin\left(\frac{\sin^{-1}x}{2}\right) \end{bmatrix} \\
&= \begin{bmatrix} \left(\sqrt{\frac{x^2+1}{2}} - i\sqrt{\frac{1-x^2}{2}}\right)\cos\left(\frac{\sin^{-1}x}{2}\right)\left(\sqrt{\frac{x^2+1}{2}} - i\sqrt{\frac{1-x^2}{2}}\right)\cos\left(\frac{\sin^{-1}x}{2}\right) \\ \left(\sqrt{\frac{x^2+1}{2}} - i\sqrt{\frac{1-x^2}{2}}\right)\cos\left(\frac{\sin^{-1}x}{2}\right)\left(\sqrt{\frac{x^2+1}{2}} + i\sqrt{\frac{1-x^2}{2}}\right)\sin\left(\frac{\sin^{-1}x}{2}\right) \\ \left(\sqrt{\frac{x^2+1}{2}} + i\sqrt{\frac{1-x^2}{2}}\right)\sin\left(\frac{\sin^{-1}x}{2}\right)\left(\sqrt{\frac{x^2+1}{2}} - i\sqrt{\frac{1-x^2}{2}}\right)\cos\left(\frac{\sin^{-1}x}{2}\right) \\ \left(\sqrt{\frac{x^2+1}{2}} + i\sqrt{\frac{1-x^2}{2}}\right)\sin\left(\frac{\sin^{-1}x}{2}\right)\left(\sqrt{\frac{x^2+1}{2}} + i\sqrt{\frac{1-x^2}{2}}\right)\sin\left(\frac{\sin^{-1}x}{2}\right) \\ \end{bmatrix}
\end{aligned}$$

$$= \begin{bmatrix} \left(\sqrt{\frac{x^2+1}{2}} - i\sqrt{\frac{1-x^2}{2}}\right)^2 \cos^2\left(\frac{\sin^{-1}x}{2}\right) \\ \left(\sqrt{\frac{x^2+1}{2}} - i\sqrt{\frac{1-x^2}{2}}\right)\left(\sqrt{\frac{x^2+1}{2}} + i\sqrt{\frac{1-x^2}{2}}\right)\cos\left(\frac{\sin^{-1}x}{2}\right)\sin\left(\frac{\sin^{-1}x}{2}\right) \\ \left(\sqrt{\frac{x^2+1}{2}} + i\sqrt{\frac{1-x^2}{2}}\right)\left(\sqrt{\frac{x^2+1}{2}} - i\sqrt{\frac{1-x^2}{2}}\right)\sin\left(\frac{\sin^{-1}x}{2}\right)\cos\left(\frac{\sin^{-1}x}{2}\right) \\ \left(\sqrt{\frac{x^2+1}{2}} + i\sqrt{\frac{1-x^2}{2}}\right)^2 \sin^2\left(\frac{\sin^{-1}x}{2}\right) \end{bmatrix}$$

$$= \begin{bmatrix} \left(x^2 - i\sqrt{1-x^4}\right)\frac{1+\cos(\sin^{-1}x)}{2} \\ 1\cdot\frac{1}{2}\sin(\sin^{-1}x) \\ 1\cdot\frac{1}{2}\sin(\sin^{-1}x) \\ \left(x^2 + i\sqrt{1-x^4}\right)\frac{1-\cos(\sin^{-1}x)}{2} \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} \left(x^2 - i\sqrt{1-x^4}\right)\left(1+\sqrt{1-x^2}\right) \\ x \\ x \\ \left(x^2 + i\sqrt{1-x^4}\right)\left(1-\sqrt{1-x^2}\right) \end{bmatrix}$$

Then apply the gates $e^{-iHT} \otimes e^{-iHT}$ and the parameterized unitary matrix $U_1(\theta) \otimes U_2(\theta)$. For convenience, we use $U(\theta)$ to denote $(U_1(\theta) \otimes U_2(\theta)) \cdot (e^{-iHT} \otimes e^{-iHT})$. Then output quantum state is

$$\begin{aligned} |\psi_{\text{out}}(x)\rangle &= U(\theta)|\psi_{\text{in}}(x)\rangle \\ &= \begin{bmatrix} u_{11}(\theta) & u_{12}(\theta) & u_{13}(\theta) & u_{14}(\theta) \\ u_{21}(\theta) & u_{22}(\theta) & u_{23}(\theta) & u_{24}(\theta) \\ u_{31}(\theta) & u_{32}(\theta) & u_{33}(\theta) & u_{34}(\theta) \\ u_{41}(\theta) & u_{42}(\theta) & u_{43}(\theta) & u_{44}(\theta) \end{bmatrix} \frac{1}{2}\begin{bmatrix} \left(x^2 - i\sqrt{1-x^4}\right)\left(1+\sqrt{1-x^2}\right) \\ x \\ x \\ \left(x^2 + i\sqrt{1-x^4}\right)\left(1-\sqrt{1-x^2}\right) \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} u_{11}(\theta)\left(x^2 - i\sqrt{1-x^4}\right)\left(1+\sqrt{1-x^2}\right) + (u_{12}(\theta)+u_{13}(\theta))x \\ +u_{14}(\theta)\left(x^2 + i\sqrt{1-x^4}\right)\left(1-\sqrt{1-x^2}\right) \\ \\ u_{21}(\theta)\left(x^2 - i\sqrt{1-x^4}\right)\left(1+\sqrt{1-x^2}\right) + (u_{22}(\theta)+u_{23}(\theta))x \\ +u_{24}(\theta)\left(x^2 + i\sqrt{1-x^4}\right)\left(1-\sqrt{1-x^2}\right) \\ \\ u_{31}(\theta)\left(x^2 - i\sqrt{1-x^4}\right)\left(1+\sqrt{1-x^2}\right) + (u_{32}(\theta)+u_{33}(\theta))x \\ +u_{34}(\theta)\left(x^2 + i\sqrt{1-x^4}\right)\left(1-\sqrt{1-x^2}\right) \\ \\ u_{41}(\theta)\left(x^2 - i\sqrt{1-x^4}\right)\left(1+\sqrt{1-x^2}\right) + (u_{42}(\theta)+u_{43}(\theta))x \\ +u_{44}(\theta)\left(x^2 + i\sqrt{1-x^4}\right)\left(1-\sqrt{1-x^2}\right) \end{bmatrix} \end{aligned}$$

Although this unitary transformation is repeated several times, they are just linear maps with respected to the current state. So it will not change the form of the final quantum state.

Finally, we measure the first qubit using PauliZ matrix $Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Z has eigenvules $\lambda_1 = 1$ and $\lambda_2 = -1$, associated with eigenstate (or eigenvalue) $\alpha_1 = |0\rangle$ and $\alpha_2 = |1\rangle$. Then $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$. Let $|\psi\rangle = a_0|00\rangle + a_1|01\rangle + a_2|10\rangle + a_3|11\rangle$. We re-write it as $|\psi\rangle = |0\rangle(a_0|0\rangle + a_1|1\rangle) + |1\rangle(a_2|0\rangle + a_3|1\rangle)$ Then the Z expectation value of the first qubit is

$$\mathbb{E}[Z] = (|a_0|^2 + |a_1|^2) \cdot 1 + (|a_2|^2 + |a_3|^2) \cdot (-1).$$

In our case,

$$\begin{aligned}
\mathbb{E}[Z] \;=\; & \frac{1}{4}\Big[\Big(u_{11}x^2\sqrt{1-x^2} + u_{11}x^2 + u_{12}x + u_{13}x + u_{14}x^2\sqrt{1-x^2} + u_{14}x^2\Big)^2 + \\
& \Big(-u_{11}\sqrt{1-x^2}\sqrt{1-x^4} - u_{11}\sqrt{1-x^4} - u_{14}\sqrt{1-x^2}\sqrt{1-x^4} + u_{14}\sqrt{1-x^4}\Big)^2\Big] \\
& +\frac{1}{4}\Big[\Big(u_{21}x^2\sqrt{1-x^2} + u_{21}x^2 + u_{22}x + u_{23}x + u_{24}x^2\sqrt{1-x^2} + u_{24}x^2\Big)^2 + \\
& \Big(-u_{21}\sqrt{1-x^2}\sqrt{1-x^4} - u_{21}\sqrt{1-x^4} - u_{24}\sqrt{1-x^2}\sqrt{1-x^4} + u_{24}\sqrt{1-x^4}\Big)^2\Big] \\
& -\frac{1}{4}\Big[\Big(u_{31}x^2\sqrt{1-x^2} + u_{31}x^2 + u_{32}x + u_{33}x + u_{34}x^2\sqrt{1-x^2} + u_{34}x^2\Big)^2 + \\
& \Big(-u_{31}\sqrt{1-x^2}\sqrt{1-x^4} - u_{31}\sqrt{1-x^4} - u_{34}\sqrt{1-x^2}\sqrt{1-x^4} + u_{34}\sqrt{1-x^4}\Big)^2\Big] \\
& -\frac{1}{4}\Big[\Big(u_{41}x^2\sqrt{1-x^2} + u_{41}x^2 + u_{42}x + u_{43}x + u_{44}x^2\sqrt{1-x^2} + u_{44}x^2\Big)^2 + \\
& \Big(-u_{41}\sqrt{1-x^2}\sqrt{1-x^4} - u_{41}\sqrt{1-x^4} - u_{44}\sqrt{1-x^2}\sqrt{1-x^4} + u_{44}\sqrt{1-x^4}\Big)^2\Big]
\end{aligned}$$