Quantum Circuit &
Quantum Machine Learning II

IGL Scholars: Shuzhen Zhang, Bozhang Chen, Linzhe Teng, Yun Wang
Graduate leader: Ivor Yidong Chen
Faculty Advisor: Marius Junge

May 2023

1 Abstract

In quantum circuit, suppose we have random permutation gate π, we shows that we can classify the π by using a quantum algorithm we constructed. The form of our algorithm is based on a quantum query algorithm, and can be understood as the quantum analog to a classical machine learning algorithm, where the random permutation can be viewed as a quantum oracle operator. We feed sets of random values into our algorithm to apply to the oracle and fit the observations to a model to retrieve information about the oracle, and we can also increasing the accuracy and reduce computation by using compressed sensing for a large but finite number of qubits.

Keywords: Quantum Circuit, Permutation, Machine Learning, Quantum Query, Compressed Sensing

2 Background

2.1 Linear Algebra

Definition 2.1. **Hilbert Space** is a finite-dimensional real or complex vector space $\mathcal{H}$ equipped with an **inner product** $\langle v, w \rangle$ such that the norm turns $\mathcal{H}$ a complete metric space.

- A matrix U is **Unitary** if $U^\dagger = U^{-1}$ where $U^\dagger$ is the conjugate transpose.
- A matrix A is **self-adjoint** if $A^\dagger = A$
- A matrix is **symmetric positive semi-definite** if it is self-adjoint and all its eigenvalues are nonnegative.
- The trace of a matrix A is defined as the sum of all its diagonal elements: $\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$
- The **tensor product** of two vector spaces V and W is denoted as $V \otimes W$ which is a bilinear map: $V \times W \to V \otimes W$ that maps a pair of basis $(v, w), v \in V, w \in W$ to $v \otimes w$.

2.2 Permutation Group

Definition 2.2. For all $n \in \mathbb{N}$, a **Permutation** π is a bijection function such that

$$\pi : \{1, 2, ..., n\} \to \{1, 2, ..., n\}$$

for n element, we have $n!$ different permutations.

Example: let n=3, then we say S_3 is the all permutations of {1,2,3}, and $\pi = (123) \in S_3$, such π do the action that: $1 \to 2 \to 3 \to 1$

Definition 2.3. for any $n \in \mathbb{N}$, Every $\pi \in S_n$ has a cycle decomposition.

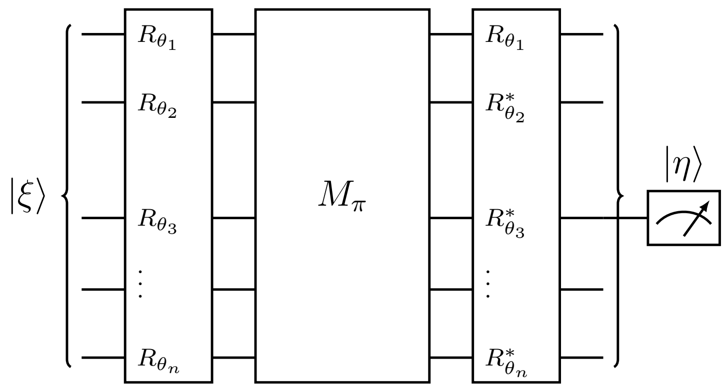
Example: $(123) \in S_3$ can be write as cycle decomposition that: $(12)(23)$

2.3 Quantum Computing

- The unit vector in quantum computing is called a **qubit**, and the state of a qubit which is a superposition of the states 0 & 1 can be represented mathematically as:
$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$
where α, β are complex numbers which satisfy $|\alpha|^2 + |\beta|^2 = 1$.
- The **density** operator of the system is defined as $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ where ρ is positive semi-definite and $\mathrm{tr}(\rho) = 1$.
- Quantum gates (unitary, invertible matrix) could operate on qubits and alter their states. A quantum circuit is composed by basic quantum gates and the measurement which unfortunately could alter the state of qubit (collapse of a quantum state).

3 Problem Statement

Quantum Circuits



- $|\xi\rangle$ and $|\eta\rangle$ are the input vectors that we choose to help distinguish the distinctive classes of M_π.
- R_{θ_i} and $R^*_{\theta_i}$, in basis matrix Z=$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, are random rotational matrices in the form of $\otimes_{i=1}^n \exp(\mathrm{i}\theta_i Z)$ which are the conjugate transpose of each other.
- M_π is an unknown permutation matrix.

Problem: How do we find the permutation matrix, M_π, using as fewer measurement as possible? What is the optimal strategy of choosing input vectors, $|\xi\rangle$ and $|\eta\rangle$ to distinguish the permutation?

Problem Analysis

Definition 3.1. Let $\pi \in S_n$ be a permutation. Then we donate:

$$\phi(\pi) = \mathbf{E}_\theta(R(\theta) M_\pi R^*(\theta)).$$

Here we have:

- $R(\theta) := \otimes_{i=1}^n \exp\{i\theta_i Z\}$ with $Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, and θ is randomly generated under a Uniform distribution $(0, 2\pi)$.
- The calculation will be:

$$\phi(\pi) = \frac{1}{(2\pi)} \int_0^{2\pi} \cdots \int_0^{2\pi} \left(R(\theta_1) \otimes \cdots \otimes R(\theta_n)\right) M_\pi \left(R^*(\theta_1) \otimes \cdots \otimes R^*(\theta_n)\right) d\theta_i \cdots d\theta_n$$

An example: (12) in S3:

$$\begin{aligned} \mathbf{E}_\theta(\phi(\pi)) &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{2\pi} \int_0^{2\pi} (R_{\theta_1} R^*_{\theta_2}) \otimes (R_{\theta_2} R^*_{\theta_1}) \otimes (R_{\theta_3} R^*_{\theta_3})\, d\theta_1 d\theta_2 d\theta_3 \\ &= \frac{1}{2\pi} \left(\int_0^{2\pi} \left(\begin{bmatrix} e^{i\psi} & 0 \\ 0 & e^{-i\psi} \end{bmatrix} \otimes \begin{bmatrix} e^{-i\psi} & 0 \\ 0 & e^{i\psi} \end{bmatrix} \right) d\psi \otimes \int_0^{2\pi} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} d\theta_3 \right), \ \psi = \theta_1 - \theta_2 \\ &= \mathrm{Diag}(1,1,0,0,0,0,1,1) \end{aligned}$$

Hence, we have the following lemmas:
Lemma 3.1. We denote the set $\{1, 2, \cdots, n\}$ as $[n]$, let $S \subset [n]$, and define set A as the following:

$$a_i = 1 \text{ if } i \in S, a_i = 0 \text{ otherwise}$$

Where $|A| = n$ and a_i is the i-th element in the set A. Moreover, let $|e_A\rangle \in (l^2)^{\otimes n}$ be the following basis vector:

$$|e_A\rangle := |0\rangle^{A^c} \otimes |1\rangle^A.$$

Then we have $\phi(\pi)|e_A\rangle = |e_A\rangle$ if and only if $\pi(A) = A$, and $\phi(\pi)|e_A\rangle = 0$ otherwise.
Lemma 3.2. If π', π are from the same conjugacy class of the permutation symmetry group S_n:

$$\mathrm{Tr}(\phi(\pi')) = \mathrm{Tr}(\phi(\pi))$$

Proof: If π' and π are in the same conjugacy class, then $|\pi'| = |\pi|$. By lemma 1, we would know the number of 1 of $\phi(\pi)$ will be same as $\phi(\pi')$. Therefore, $\mathrm{Tr}(\phi(\pi)) = \mathrm{Tr}(\phi(\pi'))$.

Lemma 3.3. According to Lemma 1, the expected value of π: $\phi(\pi)$ will be a diagonal matrix containing only 0s and 1s. And longer permutation cycle means fewer 1s in the diagonal matrix.

Let $\xi, \eta \in (\ell_2^2)^{\otimes n}$, more specifically $\xi := \otimes_{i=1}^n \xi_i$ and $\eta := \otimes_{i=1}^n \eta_i$ are input vectors such that the function $f(\pi)_{\xi,\eta}$ is defined as the expected value of $\langle \xi|\phi(\pi)|\eta\rangle$:

$$f_{\xi,\eta}(\pi) : \pi \to \mathbb{R}, f_{\xi,\eta}(\pi) = \mathbb{E}_\theta \left(\langle \xi|R(\theta)M_\pi R^*(\theta)|\eta\rangle\right)$$

4 Machine Learning

Definition 4.1. Agnostic **PAC** (Probably Approximately Correct) Learnability for General Loss Functions.
A hypothesis class $\mathcal{H}$ is Agnostic **PAC** learnable if for every $\epsilon, \delta \in (0, 1)$ and for every distribution $\mathcal{D}$ over a set $\mathcal{Z}$, there exists $h \in \mathcal{H}$ when sample size $m \geq m_{\mathcal{H}(\epsilon,\delta)}$ such that with high probability at least $1 - \delta$:

$$L_{\mathcal{D}}(h) \leq \min_{h' \in H} L_{\mathcal{D}}(h') + \epsilon$$

where $L_{\mathcal{D}}(h)$ is the expected value of the loss function and h' is the best predictor in $\mathcal{H}$.

Our problem could be mainly solved by a **Machine Learning** approach: the goal is to find the optimal **permutation** π which minimize our **loss function** through learning phase and to show that our model follows the framework of PAC learning. The set-up of our problem is the following:

1. Choose k different input vectors $(\xi_\ell, \eta_\ell)_{\ell=1}^k$ cleverly to help us better distinguish the black box permutation gate, in another word, different permutation would correspond to different output values of our quantum circuit by our choice of ξ, η.

2. Our training data set: $\{(\boldsymbol{\theta}_i, (y_{\xi,\eta_1}(\boldsymbol{\theta}_i), \cdots, y_{\xi_k,\eta_k}(\boldsymbol{\theta}_i))) \mid_{i=1}^{m}\}$ could be obtrain through the oracles R_{θ_i} and the black box permutation π'. The output value $y_{\xi_\ell,\eta_\ell}(\boldsymbol{\theta}_i)$ will be: $\mathrm{Re}\langle\xi_\ell|R(\theta_i)M_{\pi'}R(\theta_i^*)|\eta_\ell\rangle$ after measurement.
   - Due to the collapse of the quantum state by measurement, our measured values would be different from the true output.
   - The difference is denoted by ϵ_1, and ϵ_1 could be considered as the machine epsilon of our quantum machine.
   - ϵ_1 error should be bounded, and the proof will be done by our professor.
3. The prediction value of the estimated permutation π: $\hat{y}_{\xi,\eta} = f_{\xi,\eta}(\pi)$.
4. The loss function is defined as:

$$L(\pi) = \frac{1}{k \cdot n} \sum_{i=1}^{n} \sum_{\ell=1}^{k} |(\hat{y}_{\xi,\eta}(\pi))_\ell - (y_{\xi,\eta}(\theta_i, \pi'))_\ell|^2$$

   Which is similar to the mean square error in linear regression, and we should minimize this function by finding out the most suitable permutation gate $\hat{\pi}$.
5. According the framework of PAC learning, it should be guaranteed that with high probability $\mathbb{P} \geq 1 - \delta$ when sample size m is large enough such that:

$$L(\pi) \leq \epsilon_1 + \epsilon_2$$

   Where ϵ_1 is the machine epsilon to the true outcome of our oracle circuit . ϵ_2 is the minimal error by the best operator in our set up.

5 Input vector $(\xi_\ell, \eta_\ell)_{\ell=1}^{k}$ for Low Complexity Permutations

5.1 Goal:

Given S_n, find a pair of the input vector $|\xi\rangle$ and $|\eta\rangle$ to distinguish the expected output for each permutation in S_n, where the input vectors are the tensor products of n qubits.

5.2 Consider permutations in S_3:

5.2.1 Dummy method

- **Removing input vectors**: After removing input vectors, the expected output for each permutation is the trace of each permutation matrix according to the lemma 5. Then permutations in the same conjugacy class would have the same trace. And for the higher order permutation group, two distinct conjugacy classes might have the same trace. For example, in S_5, conjugacy class (123) and conjugacy class (12)(34) have the same trace 8.

Class	Trace
(1)	8
(12)	4
(123)	2

Table 5.2.1: Trace for classes in S_3

Class	Trace
(1)	32
(12)	16
(123)	8
(12)(34)	8
(1234)	4
(12)(345)	4
(12345)	2

Table 5.2.2: Trace for classes in S_5

5.2.2 Enhanced method

1. **Intuition:** Prediction for π:

$$\begin{aligned}\hat{y}_{\xi,\eta} = f_{\xi,\eta}(\pi) &= \mathbf{E}_\theta\big(\langle\xi|R(\theta)M_\pi R^*(\theta)|\eta\rangle\big)\\ &= \langle\xi|\mathbf{E}_\theta\big(R(\theta)M_\pi R^*(\theta)\big)|\eta\rangle\end{aligned}$$

   Where $\mathbf{E}_\theta\big(R(\theta)M_\pi R^*(\theta)\big)$ is a diagonal matrix where only contains 0, if the corresponding basis $e_i \neq M_\pi(e_i)$, and 1 otherwise.
   Assume the basis for the permutation matrix is $e_{000}, e_{001}, e_{010}, e_{011}, e_{100}, e_{101}, e_{110}, e_{111}$, and since the input vectors in S_3 are the tensor products of 3 qubits: $|\xi\rangle = |\xi_1\rangle \otimes |\xi_2\rangle \otimes |\xi_3\rangle = [a_1, b_1, c_1, d_1, e_1, f_1, g_1, h_1] \in \mathbb{C}^8$ & $|\eta\rangle = |\eta_1\rangle \otimes |\eta_2\rangle \otimes |\eta_3\rangle = [a_2, b_2, c_2, d_2, e_2, f_2, g_2, h_2] \in \mathbb{C}^8$.
   Then:

$$\begin{aligned} f(1) &= a_1\cdot a_2 + b_1\cdot b_2 + c_1\cdot c_2 + d_1\cdot d_2 + e_1\cdot e_2 + f_1\cdot f_2 + g_1\cdot g_2 + h_1\cdot h_2\\ f(12) &= a_1\cdot a_2 + b_1\cdot b_2 + g_1\cdot g_2 + h_1\cdot h_2\\ f(13) &= a_1\cdot a_2 + c_1\cdot c_2 + f_1\cdot f_2 + h_1\cdot h_2\\ f(23) &= a_1\cdot a_2 + d_1\cdot d_2 + e_1\cdot e_2 + h_1\cdot h_2\\ f(123) &= f(132) : a_1\cdot a_2 + h_1\cdot h_2\end{aligned}$$

   Therefore, if there are input vectors $|\xi\rangle$ and $|\eta\rangle$ that have eight distinct elements, we are able to distinguish all permutations in S_3 except (123) and (132).

2. **Amplitudes:** Since the sum of the absolute squares of the amplitudes for each qubit should be 1, then we could let each qubit be:

$$\cos(\gamma)|0\rangle + \sin(\gamma)|1\rangle$$

3. **Method:** Based on point (1) and point (2), in S_3, we can construct $\langle\xi|^\dagger = |\eta\rangle = |\eta_1\rangle \otimes |\eta_2\rangle \otimes |\eta_3\rangle$, where $|\eta_i\rangle = \cos(\gamma_i)|0\rangle + \sin(\gamma_i)|1\rangle$ and $\gamma_i \in \left(\frac{\pi}{0}, \frac{\pi}{4}\right)$

5.3 General case for S_n

Lemma 5.1. For any choice of ξ, η, the expected value of $\hat{y}_{\xi,\eta}(\pi)$ is not distinguishable if π is the longest cycle in the permutation group, in another word: π a single cycle permutation which contains all elements in the group.

Proof. The only basis which would not be altered by the longest cycle π contains all 0s or all 1s, and thanks to our lemma 3.1, there are only two 1s in the diagonal matrices of $\phi(\pi)$ and these two 1s are always in the same position. Thus, $\hat{y}_{\xi,\eta}(\pi)$ will be the same value for this conjugacy class with the longest cycle. $\square$

Lemma 5.1 tells us that the permutation with the longest cycle could not be distinguished by our algorithm and the only thing we could tell is that the permutation is from this conjugacy class.
In the general case for S_n, we could utilize a similar trick as we do in S_3. However, as the number of qubits increases, the classical computational cost for determination of estimator π will increase dramatically. Thus, Compressed Sensing will be introduced here due to the sparsity of diagonal matrix for long cycle permutation.

6 Compressed Sensing

6.1 Trace and Length of Cycle

let $M_\pi : \ell_2^{2^n} \to \ell_2^{2^n}, \forall \pi \in S_n$ and π is the form of cycle decomposition such that

$$\pi = \Pi_{i=1}^r \pi_i, \forall \pi_i \in S_n$$

it has length r and $\ell_2^{2^n} = \otimes_{i=1}^r M_\pi(i)$.

Example:

- $(12345) = (12)(23)(34)(45)$ has length 4.
- $(12)(34)$ has length 2 same as $(123) = (12)(23)$
- e has length 0

Suppose we have permutation group S_n, and cycle π which has length r, then we have:

Lemma 6.1. The trace of M_π with cycle's length has relation that: $\mathrm{tr}(M_\pi) = 2^{n-r}$

Proof: Proof by induction. Suppose in S_n permutation group, for any $\pi \in S_n$ and π has length $r = 0, 1, ..., k$ and $k < n-1$, and it satisfy: $\mathrm{tr}(M_\pi) = 2^{n-r}$

Base case: $\pi = e, r = 0$

$$\mathbf{E}_\theta[R_\theta M_e R_\theta^*] = \frac{1}{2\pi} \otimes_{i=1}^n \int_0^{2\pi} R_{\theta_i} R_{\theta_i}^* d\theta_i = \otimes_{i=1}^n \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_{2^n}$$

then $\mathrm{tr}(M_\pi) = 2^n = 2^{n-0}$
Let $\pi = \pi_1 \cdot \pi_2$, which π_2 is a element of 2-cycle decomposition of π with $r_{\pi_1} = k, r_{\pi_2} = 1$, then $r_\pi = k+1, k+1 < n$

$$\mathbf{E}[R_\theta M_\pi R_\theta^*] = \frac{1}{2\pi} \int_0^{2\pi} (R_{\theta_1} \otimes ...R_{\theta_n})(M_\pi)(R_{\theta_1}^* \otimes ...R_{\theta_n}^*) d(\theta_1, .., \theta_n)$$

$$= \frac{1}{2\pi} \int_0^{2\pi} (R_{\theta_1} \otimes ...R_{\theta_n})(M_{\pi_1} \otimes M_{\pi_2})(R^*_{\theta_1} \otimes ...R^*_{\theta_n}) d(\theta_1, .., \theta_n)$$

$$= \frac{1}{2\pi} \otimes_{i=1}^n \int_0^{2\pi} R_{\theta_i}(M_{\pi_1} \otimes M_{\pi_2}) R^*_{\theta_{(\pi_1 * \pi_2)(i)}} d\theta_i d\theta_{(\pi_1 \cdot \pi_2)(i)}$$

$$= \frac{1}{2\pi} (\otimes_{i=1}^n \int_0^{2\pi} R_{\theta_i}(M_{\pi_1}) R^*_{\theta_{\pi_1(i)}} d\theta_i d\theta_{\pi_1(i)}) (\otimes_{i=1}^n \int_0^{2\pi} R_{\theta_i}(M_{\pi_2}) R^*_{\theta_{\pi_2(i)}} d\theta_i d\theta_{\pi_2(i)})$$

Then we know $\text{Diag}(M_\pi) = \text{Diag}(M_{\pi_1}) \otimes \text{Diag}(M_{\pi_2})$**.**

- Since we know $\text{tr}(M_{\pi_1}) = 2^{n-k}$ and $\text{tr}(M_{\pi_2}) = 2^{n-1}$
- It shows that $\text{Diag}(M_{\pi_1})$ has dimension 2^{n-k} and $\text{Diag}(M_{\pi_2})$ has dimension 2^{n-1}
- Because of $\pi_2 \neq \pi_1$, $\pi_2 \notin \pi_1$, and Identity matrix has dimension 2^n
- Then $\text{Diag}(M_{\pi_1})\text{Diag}(M_{\pi_2})$ means $\text{Diag}(M_{\pi_1})$ will reduce dimension by half, which the dimension of $\text{Diag}(M_\pi)$ will be 2^{n-k-1}
- Therefore, $\text{tr}(M_\pi) = 2^{n-(k+1)}$, induction success.

6.2 Application

Lemma 6.2. Let the value of $\text{Diag}(M_\pi)$ be $\tau(M_\pi)$ and define: $\tau(M_\pi) = \frac{\text{tr}(M_\pi)}{2^n} = 2^{-r}$

Theorem: There is a function $G : \ell_1^{2^n} \to \ell_2^m$, with $m \approx 2^{n-r}$ such that for $\pi, \pi' \in S_n : \|G(\pi) - G(\pi')\|_2 \geq \epsilon$, then two permutations can be classify

Lemma 6.3. When τ is smaller, our prediction will be more precise

In this case, since we proved before that when cycles are longer, our $\text{tr}(M_\pi)$ will be smaller, which means our τ will be smaller. Therefore, in our measurement procedure, we could focus on longer cycles which will reduce our computation cost and increase the precision.

7 In the future

Due to time limitations, there is significant potential for improving our machine learning algorithm and proof for some content needs to be completed in this problem.

1. To demonstrate the PAC learnability of our algorithm, a formal proof should be given to show that with enough data size n, the learning error is bounded.
2. More experiments of loss function (e.g. cross-entropy) could be done to improve the stability and accuracy of our model.
3. Continue in understanding the compressed sensing technique, and its application for long cycle permutations with a formal proof.
4. In order to reduce the construction cost of quantum circuit and error introduced by measurement, optimizing our circuit to reduce the number of measurements is necessary as well.

References

[1] Shalev-Shwartz, S., & Ben-David, S. Understanding Machine Learning: From Theory to Algorithms,*Cambridge University Press,(pp. 22-28), 2014*
[2] S. Arunachalam, J. Briet, and C. Palazuelos. Quantum query algorithms are completely bounded forms. *SIAM J. Comput, 48(3):903 925, 2019.*