# Quantum Circuit, Permutations, and Quantum Machine Learning

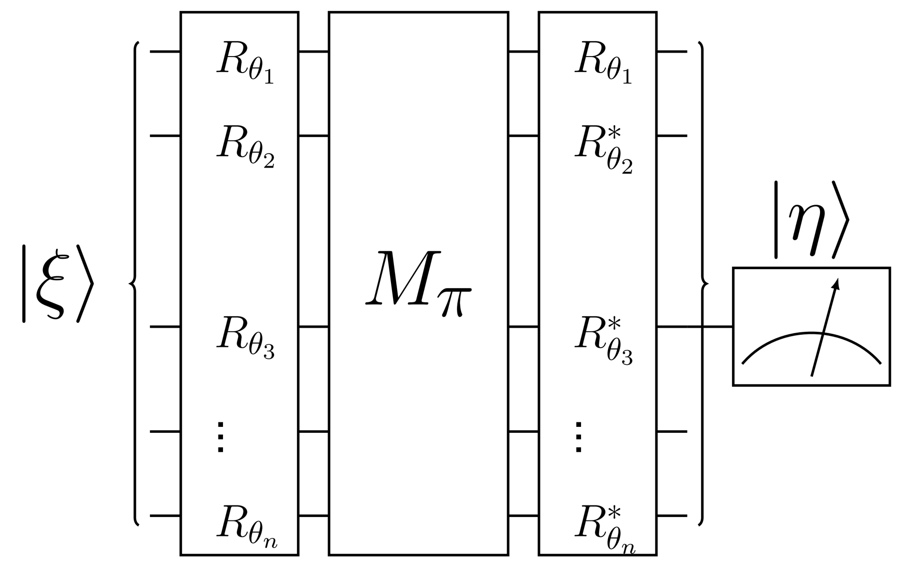
**Bozhang Chen, Linzhe Teng, Yun Wang, Shuzhen Zhang, Ivor Yidong Chen, Marius Junge**

## Introduction

**Goal**

Identify an unknown permutation gate implemented in a potentially complex quantum circuit.

**Problem state:**




By providing n input qubits after a rotation matrix: $(R(\theta_1) \otimes R(\theta_2) \otimes \cdots \otimes R(\theta_n))$ passing through a permutation Matrix $M_\pi$ and then into the conjugate transpose of the rotation matrix $\left(R(\theta_1^*) \otimes R(\theta_2^*) \otimes \cdots \otimes R(\theta_n)^*\right)$. Finally, we measure the result of our quantum circuit. **Which $\pi$ was actually used?**

## 1. Problem Analysis

**Definition.** *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 $\theta$ is randomly generated under a Uniform distribution $(0, 2\pi)$.
- the calculation will be:

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

**Lemma 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 2.** *If $\pi', \pi$ are from the same conjugacy class of the permutation symmetry group $S_n$:*

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

**Lemma 3.** *According to Lemma 1, the expected value of $\pi$: $\phi(\pi)$ 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 value of the function $f(\pi)$,

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

allows us to distinguish the distinctive classes of $M_\pi$

**Query:**

Rotation matrixs, $R(\theta)$, are considered query that from oracle.

## 2. Machine Learning Set up

This problem is mainly solved by a Machine Learning Theory approach:

1. Choose our input vectors $(\xi_\ell, \eta_\ell)_{\ell=1}^k$ cleverly for our training data.
2. Then, we could obtain the training data by our oracles $R_\theta$ and the implemented permutation matrix: $\pi'$: $\cup_{i=1}^m \{(\boldsymbol{\theta}_i, \left(y_{\xi,\eta}(\boldsymbol{\theta}_i, \pi')\right)_\ell |_{\ell=1}^k\}$.
3. The value $(y_{\xi,\eta}(\boldsymbol{\theta}_i, \pi'))_\ell$ have to be $\epsilon_1$ close to the true outcome of the oracle $\boldsymbol{\theta}_i$ combined with the implemented gate $\pi'$ after our measurement.
4. Our prediction value of the estimated permutation $\pi$ is $\hat{y}_{\xi,\eta} = f_{\xi,\eta}(\pi)$.

   **Definition.** *(Agnostic **PAC** 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 $\mathcal{H}$.*

- In our case, the loss function is defined as the following:

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

Minimizing the loss function $L$ gives us the optimal permutation gate $\pi$.
Also, we want to guarantee that with high probability $\mathbb{P} \geq 1 - \delta$ when sample size is large enough:

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

Where $\epsilon_1$ is the machine epsilon to the true outcome of our oracle circuit . $\epsilon_2$ is the error in our prediction.

## 3. Our procedures

### 3.1 Dummy Test

Without choosing the input vectors$\xi, \eta$, try to directly use the trace of our expected value as the output by our lemmas. Here's the results for $S_3$ and $S_5$:

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

**Table 3.1.1:** *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 3.1.1:** *Trace for classes in $S_5$*

We see that different conjugacy classes in a larger dimension may share the same trace value of $\phi(\pi)$, therefore we fail in our dummy test. Choosing appropriate input vectors $\xi, \eta$ is necessary to better find the optimal permutation.

### 3.2 Select appropriate input vector $\xi, \eta$ for $S_3$

- In $S_3$, suppose $|\xi\rangle = |\eta\rangle$ and $|\xi\rangle = |\xi_1\rangle \otimes |\xi_2\rangle \otimes |\xi_3\rangle$, and we have to make sure each qubit's $l_2$ norm is 1.
- 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 let $|\xi\rangle$ be represented as $[a, b, c, d, e, f, g, h] \in \mathbb{C}^8$.
- According to the calculation of $f(\pi)$ and Lemma 3, the output for different permutation operators would be:

$$f(1) : a^2 + b^2 + c^2 + d^2 + e^2 + f^2 + g^2 + h^2$$
$$f(12) : a^2 + b^2 + g^2 + h^2$$
$$f(13) : a^2 + c^2 + f^2 + h^2$$
$$f(23) : a^2 + d^2 + e^2 + h^2$$
$$f(123) = f(132) : a^2 + h^2$$

1. Since we are trying to make each output of $f(\pi)$ distinct, we then need to make sure each value in $|\xi\rangle$ is distinct.
2. Since $|\eta\rangle$ = $|\xi\rangle$ = $|\xi_1\rangle \otimes |\xi_2\rangle \otimes |\xi_3\rangle$ where each qubit's $l_2$ norm should be 1.
3. Therefore, we let $|\xi_i\rangle = \cos(\delta_i)|0\rangle + \sin(\delta_i)|1\rangle$ where $\delta$ randomly selected from $(0, \pi/4)$ to get 3 distinct $\delta_i$ to make sure the value of each $f(\pi)$ as distinct as possible.

### 3.3 Select appropriate input vector $\xi, \eta$ for $S_n$

- 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 $\pi$ will increase dramatically.
- Thanks to the sparsity of diagonal matrix for long cycle permutation, we can use Compressed Sensing technique.

## 4. Future Directions

Quantum circuits are expensive to produce but fast to test. Efficiency has to be improved in evaluating the classical data by

- Modifying loss function (for example, cross-entropy)
- work out compressed sensing technique for long cycle permutations.
- Use theoretical tools to estimate the required number $M$ of the training phase.

**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.*

*Support for this project was provided by the Illinois Geometry Lab and the Department of Mathematics at the University of Illinois at Urbana-Champaign.*