HONOURS PROJECT UPDATE
for Semester V

ALAPAN CHAUDHURI[1]

Dec 9, 2021

PREFACE

In this report, I talk about the projects the pursued as a part of my Honours studies (on Quantum Computation and Information) during the semester (Monsoon 2021). It contains an extensive series of notes on Quantum Information (which I have worked on throughout the semester). Finding concise, rigorous and yet complete notes on the topic is hard to come across – thus, hopefully my contribution will help in that regard.

Other than my notes, this report serves as an update on my research work throughout the semester. This includes paper summaries and original approaches on different problems (e.g., QPIR) or frameworks (e.g., Cirq) that I have looked into.

Lastly, I would like to thank Prof. Indranil Chakrabarty and Prof. Prasad Krishnan as well as my friends (peers and seniors) for their contributions and inputs which I deeply value.

[1] *Alapan Chaudhuri, CSTAR & CQST, IIITH*

CONTENTS

1 PRIVATE INFORMATION RETRIEVAL

1.1 Introduction

Private Information Retrieval (PIR) refers to the problem of "privately" retrieving a file out of M messages from N distributed databases (or servers) in such a way that no individual database can tell which file has been retrieved. Thus, the goal of such as a system is to protect the privacy of the query (the file we are interested in retrieving). Each database can only tell that a file has been requested and the response time, but they do not know exactly which one we were interested in retrieving.

There are two fields of study within PIR, namely - information theoretic PIR (IT-PIR) and computational PIR (CPIR). ITPIR is faster since it uses cheap cryptographic operations and is information theoretically secure. But, it assumes non-collusion and requires a minimum of 2 databases.

CPIR, on the other hand, is much slower but we can use it for colluding servers as well as just a single server (or database).

1.2 Major motivations for QPIR

- Protecting our online query privacy.
- PIR for replicated and MSD coded storage as well as for colluding servers has gained a lot of interest recently.
- Promise of better rates of transmission are possible with quantum communication.

Theorem 1. *If the data is stored on only one server (or equivalently, if all servers are colluding), then the only thing we can do to achieve perfect privacy is to download the entire database.*

- On the other hand, if the database is replicated or coded on multiple servers that are not all communicating, then we can do better.
- Upload cost ignored: the number of bits in the files ($\sim$ download cost) is assumed to be much bigger than the number of files on the servers ($\sim$ upload cost).
- Symmetric PIR (SPIR): the user is only able to decode the file that he has requested, and learns nothing about the other files, i.e., privacy is guaranteed also for the server.

1.3 Definitions within PIR

- Correctness: there exists a functional D such that $D(A^K, Q^K, K) = x^K, \forall K$.
- t-collusion: any t servers may collude, i.e., exchange their received queries.
- t-PIR scheme: a scheme that protects against t-collusion, i.e., any set of at most t colluding nodes learns no information about the index K of the desired file.
- PIR rate: $R = \frac{\text{number of bits in a file}}{\text{number of received bits}}$ and in case of QPIR we have number of received qubits as the denominator.
- Capacity: supremum of PIR rates of all possible PIR schemes, for a fixed parameter setting.

1.4 Classical PIR Model

Let us consider the classical PIR setup with respect to multiple replicated servers.

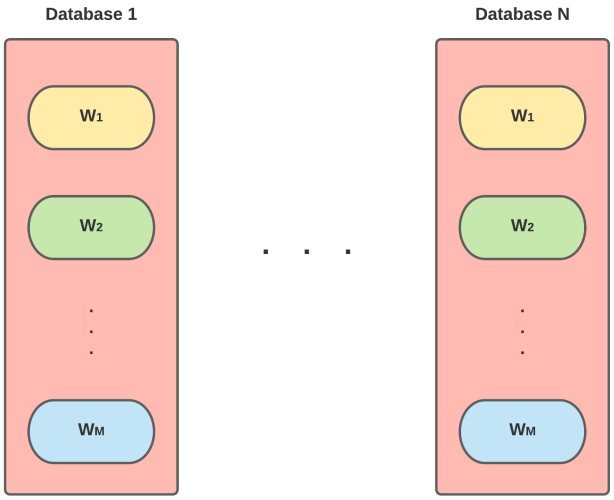


Figure 1: Replicated databases or servers

Now, if $N = 1$, i.e., we have only a single server then all that we can do is download all the files $W_1, W_2, \ldots, W_M$ and the related rate of this trivial PIR scheme is $1/M$.

The first interesting scheme that we come across is when $N = 2$.

- Given: two replicated databases, as shown in the figure, with M one-bit messages. And, we are trying to retrieve the message W_i.
- Consider a random vector $\vec{h} \in \mathcal{L}(\{0,1\}^M)$.

$$\vec{h} = \begin{bmatrix} h_1 & h_2 & \ldots & h_M \end{bmatrix}$$

- Our queries are based on $\vec{h}$ and $\vec{W}$.
- We have two queries in this scheme. The first one is made to server 1 ($Q_1^{[i]}$) and the last one is made to server 2 ($Q_2^{[i]}$).
- $Q_1^{[i]}$: retrieve dot product of $\vec{h}$ and $\vec{W}$.
- Thus, $A_1^{[i]} = \sum_{j=1}^{M} h_j W_j$.
- $Q_2^{[i]}$: retrieve dot product of $\vec{h'}$ and $\vec{W}$ where $\vec{h'} = \begin{bmatrix} h_1 & h_2 & \ldots & h_i + 1 & \ldots & h_M \end{bmatrix}$
- Thus, $A_2^{[i]} = \sum_{j=1}^{M} h_j W_j + W_i$.
- Therefore, we have $W_i = A_2^{[i]} - A_1^{[i]}$ and the retrieval rate is given by $1/2$.

1.5 N-server system with M-multibit messages

- Given: N replicated databases, as shown in the figure, with M multi-bit ($L = N - 1$) messages. And, we are trying to retrieve the message W_i.

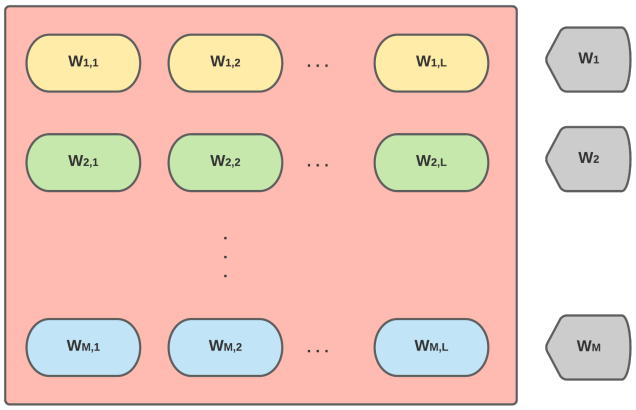


Figure 2: N replicated databases

- Consider a random vector $\vec{h} \in \mathcal{L}(\{0,1\}^{M\times L})$.

$$\vec{h} = \begin{bmatrix} \vec{h_1} \\ \vec{h_2} \\ \vdots \\ \vec{h_M} \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,L} \\ h_{2,1} & h_{2,2} & \dots & h_{2,L} \\ \vdots & \vdots & & \vdots \\ h_{M,1} & h_{M,2} & \dots & h_{M,L} \end{bmatrix}$$

- Our queries are based on $\vec{h}$ and $\vec{W}$.
- We have two queries in this scheme. The first one is made to server 1 ($Q_1^{[i]} \to A_1^{[i]}$) and the last one is made to server N ($Q_N^{[i]} \to A_N^{[i]}$).

$$A_1^{[i]} = \sum\sum h_{j,k} W_{j,k}$$

$$A_2^{[i]} = \sum\sum h_{j,k} W_{j,k} + W_{i,1}$$

$$A_N^{[i]} = \sum\sum h_{j,k} W_{j,k} + W_{i,N-1}$$

- Therefore, we can obtain $\vec{W_i}$ from $\vec{A^{[i]}}$ and the rate of retrieval is $R = \frac{N-1}{N}$.

1.6 Formulating PIR

- Queries and files or messages are independent and have zero mutual information.

$$I(Q_1^{[i]}, \dots, Q_N^{[i]}; W_1, \dots, W_M) = 0$$

- Answers are fully determined by messages and queries.

$$H(A_n^{[i]} | Q_n^{[i]}, W_1, \dots, W_M) = 0,\ n \in \{1, \dots, N\}$$

- Constraint of Reliability: $H(W_i | A_1^{[i]}, \dots, A_N^{[i]}, Q_1^{[i]}, \dots, Q_N^{[i]}) = o(L)$
- Constraint of Privacy: $I(Q_n^{[i]}; i) = 0,\ n \in \{1, \dots, N\}$
- Retrieval rate of PIR is given by the following.

$$R = \frac{H(W_i)}{\sum_{n=1}^{N} H(A_n^{[i]})}$$

- Capacity of PIR is given by the supremum of PIR rates of all possible PIR schemes, for a fixed parameter setting.

1.7 Capacity of Classical PIR

The capacity of classical PIR is given by C_{PIR}.

$$C_{PIR} = \frac{1 - 1/N}{1 - (1/N)^M}$$

We can construct the C_{PIR} scheme by making sure that we maximally utilize interference or interference alignment.

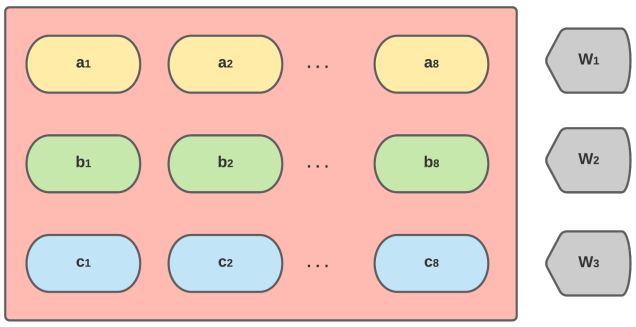


Figure 3: Database

Let us understand the scheme via the example shown above. Here, we have 2 servers and 3 multibit files ($N = 2,\ M = 3$).

Server			
DB_1	a_1	b_1	c_1
DB_2	a_2	b_2	c_2
DB_1	$a_3 + b_2$	$a_4 + c_2$	$b_3 + c_3$
DB_2	$a_5 + b_1$	$a_6 + c_1$	$b_4 + c_4$
DB_1		$a_7 + b_3 + c_3$	
DB_2		$a_8 + b_4 + c_4$	

Table 1: Query results

Rate of this above scheme is given as follows.

$$R = \frac{8}{14} = C_{PIR}(N = 2, M = 3)$$

2 PROBABILISTIC PIR

Now, let us look into some probabilistic schemes. This approach is used to construct queries with asymmetric lengths.

$$R = \frac{L}{\sum_{n=1}^{N} \mathbb{E}[l_n]}$$

The retrieval rate of such a scheme is constructed as given above. And, in this case too we have objectives such as:

- Obtain optimal retrieval rate for probabilistic PIR $= C_{PIR}$
- Minimize message length
- Minimize upload cost

2.1 $N = 2,\ M = 2$ probablistic PIR scheme

Probability	Requesting W_1	
	DB_1	DB_2
1/2	ϕ	W_1
1/2	$W_1 + W_2$	W_2

$$\text{Retrieval Rate } = \frac{1}{1/2 \times 1 + 1/2 \times 2} = \frac{2}{3} = C_{PIR}(N = 2, M = 2)$$

Probability	Requesting W_2	
	DB_1	DB_2
1/2	ϕ	W_2
1/2	$W_1 + W_2$	W_1

3 CAPACITY OF DIFFERENT PIR SCHEMES

- Classical PIR
$$C_{PIR} = \frac{1 - 1/N}{1 - (1/N)^M}$$
- T-colluding PIR
$$C_{COL} = \frac{1 - T/N}{1 - (T/N)^M}$$
- U-robust PIR
$$C_{ROB} = \frac{1 - \frac{T}{N-U}}{1 - (\frac{T}{N-U})^M}$$
- B-byzantine PIR
$$C_{BYZ} = \frac{1 - \frac{T}{N-2B}}{1 - (\frac{T}{N-2B})^M}$$
- Symmetric PIR
$$C_{SPIR} = 1 - \frac{1}{N}$$
- (N, K) MDS-coded PIR
$$C_{MDS} = \frac{1 - \frac{K}{N}}{1 - (\frac{K}{N})^M}$$

3.1 Tabulated results for important cases

Capacities	**PIR**	**SPIR**	**QPIR**
Replicated (non-colluding)	$1 - 1/N$	$1 - 1/N$	1
Replicated (T-collusion)	$1 - T/N$	$1 - T/N$	$\geqslant 2/(T+2)$
(N, K) MDS-coded (T-collusion)	$1 - (K+T-1)/N$ *	$1 - (K+T-1)/N$	$\geqslant 2/(K+T-1)$

* *conjectured to be true by Freij-Hollanti et. al.*

4 QUANTUM PRIVATE INFORMATION RETRIEVAL

In the QPIR problem with multiple servers, the objective is for a user to retrieve a classical file by downloading (entangled) quantum systems from multiple replicated servers, while maintaining the privacy constraint that identity of the downloaded file remains unknown to each server.

In the quantum counterpart for the classical PIR problem, we shall be dealing with non-colluded replicated servers.

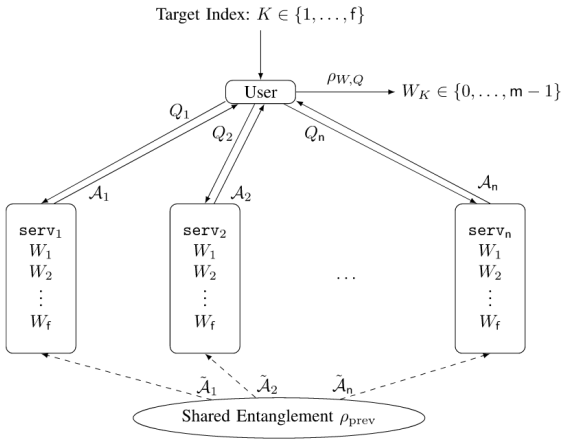


Figure 4: QPIR protocol showing the composite system of the servers initialized to an entangled state ρ_{perv}.

4.1 Description of the QPIR scheme

QPIR protocol, for n-servers and f-files with m-bits each, is given by the 4-tuple $\Phi(m)_{QPIR} = (\rho_{prev}, Enc_{user}, Enc_{serv}, Dec)$ where Enc_{user} denotes the user encoder, $Enc_{serv} := (Enc_{serv_1}, ..., Enc_{serv_n})$ denotes the collection of the server encoders, and Dec denotes the decoder.

To retrieve the W_K , the user encodes queries by user encoder given below. $Enc_{user}(K, R_{user}) = (Q_1, ..., Q_n)$ where Q_t is the set of query symbols to the t-th server for any $t \in \{1, \dots, n\}$. The n queries $Q1, \dots, Q_n$ are sent to the servers $serv_1, \dots, serv_n$, respectively.

Finally, each server $serv_t$ applies a CPTP map Λ_t depending on $Q_t, W_1, \dots, W_f$. The final received state of the user is given above.

$$\rho_{W,Q} := \Lambda_1, \dots, \Lambda_n(\rho_{prev})$$

4.2 Tools for QPIR

Consider multiple 2-qubit systems $\mathcal{H}_i \otimes \mathcal{H}_j$ in maximally entangled state ϕ.

$$\phi = \frac{00 + 11}{\sqrt{2}}$$

We then define an unitary operator $W(a, b)$, where $a, b \in \mathbb{F}_2$, over the system.

$$W(a, b) = (-1)^{ab} \sum_{i=0}^{l-1} (-1)^{ai} i + bi$$

Theorem 2. *The set* $\mathcal{B}_{\mathbb{F}_2^2} := \{B_{(a,b)} := W(a,b)\phi\phi W(a,b)^T \mid a,b \in \mathcal{F}_2\}$ *is a projection valued measure. And, the measurement defined by the POVM* $\mathcal{B}_{\mathbb{F}_2^2}$ *is called the Bell measurement.*

4.2.1 Two sum transmission protocol

- Alice and Bob prepare qubits $\mathcal{H}_A$ and $\mathcal{H}_B$ in the maximally entangled state ϕ.
- Alice and Bob apply the unitaries $W_A(a_1, a_2)$ on $\mathcal{H}_A$ and $W_B(b_1, b_2)$ on $\mathcal{H}_B$, respectively.
- Alice and Bob send their qubits (respectively $\mathcal{H}_A, \mathcal{H}_B$) to Carol through two quantum channels.
- Carol performs a Bell measurement on the system $\mathcal{H}_A \otimes \mathcal{H}_B$ and obtains $(a_1 + b_1, a_2 + b_2)$ as the protocol output.

Rate-one QPIR protocol In this section, we propose a $R_{QPIR} = 1$ 2-server QPIR protocol with the perfect security and negligible upload cost. This protocol is constructed from the idea of the classical two-server PIR protocol.

- To retrieve: W_K
- In each server: $W_1, \ldots, W_f \in \{0, \ldots, l^2 - 1 =: m_l - 1\}$
- We assume that $serv_1$ and $serv_2$ possess the l-dimensional quantum systems $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively.
- The maximally entangled state ϕ in $\mathcal{A}_1 \otimes \mathcal{A}_2$ is shared at the beginning of the protocol.

The QPIR protocol for retrieving W_K is described as follows.

1. Depending on the target file index K, the user chooses a subset R_{user} of $\{1, , f\}$ uniformly. Let $Q_1 := R_{user}$ and we have:
$$Q_2 = \begin{cases} Q_1 - K, & \text{if } \{K\} \in Q_1 \\ Q_1 \cup K, & \text{if } \{K\} \notin Q_1 \end{cases}$$
2. The user sends the queries Q_1 and Q_2 to $serv_1$ and $serv_2$, respectively.
3. $serv_1$ calculates $H_1 := \sum_{i \in Q_1} W_i \in \mathbb{Z}_l^2$ and applies $W(H_1)$ on the quantum system $\mathcal{A}_1$.
4. Similarly, $serv_2$ calculates $H_2 := \sum_{i \in Q_2} W_i \in \mathbb{Z}_l^2$ and applies $W(H_2)$ on the quantum system $\mathcal{A}_2$.
5. The state on $\mathcal{A}_1 \otimes \mathcal{A}_2$ is $(W(H_1) \otimes W(H_2))\phi$.
6. $serv_1$ and $serv_2$ send the quantum systems $\mathcal{A}_1$ and $\mathcal{A}2$ to the user, respectively.
7. The user performs a POVM $\mathcal{B}_{\mathbb{F}_2^2}$ where each POVM element is given by the following.
$$B_{(a,b)} := (W(a,b) \otimes I)\phi\phi(W(a,b)^\dagger \otimes I), \quad \text{if } K \in Q_1$$
$$B_{(a,b)} := (W(-a,-b) \otimes I)\phi\phi(W(-a,-b)^\dagger \otimes I), \quad \text{otherwise}$$
The user obtains the measurement outcome (a,b) as the retrieval result.

4.3 Conclusion

As described above, we have mostly explored Classical PIR schemes and topics in Quantum Shannon Theory during the extent of my study. Future work involves looking into different QPIR protocols, so as to come up with new novel schemes or results on capacity under different scenarios and settings.

5 GOOGLE CIRQ

During the semester, I have also had some experience in working with Google Cirq. Amongst my contributions, I have the following:

- Currently working on OpenQASM 3 implementation
- Added $r(\theta, \phi)$ gate (with Zeeshan Ahmed)

6 CREATING ARBITRARY SUPERPOSITIONS

Creating any arbitrary superposition is an important problem to solve and implement in every quantum programming language. Our (Zeeshan Ahmed and myself) approach to the problem is described below.

6.1 Problem Statement

Design a general circuit that accepts vectors with random positive integral values of size n with m bits in length for each element and finds the superposition of indices such that the elements of the vector at those indices have different values for every two adjacent bits in their binary representation.

6.1.1 Example

So, if we were to provide an example — consider the vector $[1, 5, 7, 10]$. Now, I shall describe the computation that would lead us to our solution.

Classical Part:

- First, we convert the elements of the array into their respective binary representation (as required).

$$[1, 5, 7, 10] \rightarrow [0001, 0101, 0111, 1010]$$

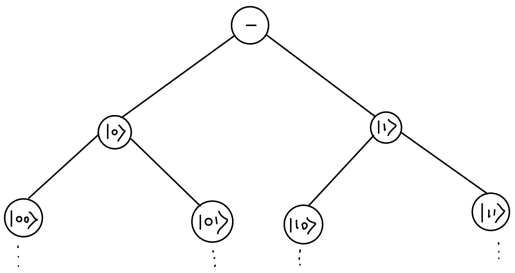
- Now, in the list of binary representations we search for those numbers with different values for every adjacent bit pairs. In this case, they are $[0101, 1010]$ that is 5 and 10 respectively.
- Now, we need to obtain their indices in accordance with the original array.

$$\text{indices} = [1, 3]$$

- Finally, we return the binary representation of these indices.

$$[01, 11]$$

Quantum Part:



- Given, the list of indices in their binary representation consider each element of the list to denote a quantum state. Now we have a list of quantum states. Thus, in this case we have $[|01\rangle, |11\rangle]$.
- Create a state that is equal superposition of all such states in the above list. The final superposition obtained in this case is shown below.

$$\frac{1}{\sqrt{2}}(|01\rangle + |11\rangle)$$

6.2 Solving for the Classical Part

For the classical part of the problem, given an arbitrary array of elements we have to return a list of suitable indices. This is easy enough.

Algorithm: Suppose we have an array arr, then we iterate over each of its elements and convert them into their required binary representation and for every such binary representation, we do the following:

- Let s be the string denoting the binary representation of $arr[i]$.
- Then, if $s[j] \neq s[j-1]\ \forall\ j \in [1, ..., \text{len}(s)-1]$ then we know that we have found a binary representation where every two adjacent bits are different.
- We append the index i to a list, say L.
- Finally after all iterations, we obtain L as the required list of suitable indices.

6.3 Solving for the Quantum Part

Now, upon obtaining the list of indices, we form an array (called params) of length $2^{\text{binary len}(n-1)}$ where n is the number of elements of arr (the original vector or array of random positive integers).

Now, params will be an array representation of the required superposed state we expect to get. For example, in the example we had shown above with $arr = [1, 5, 7, 10]$, our params array will be as follows.

$$\text{params} = [0, \frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}]$$

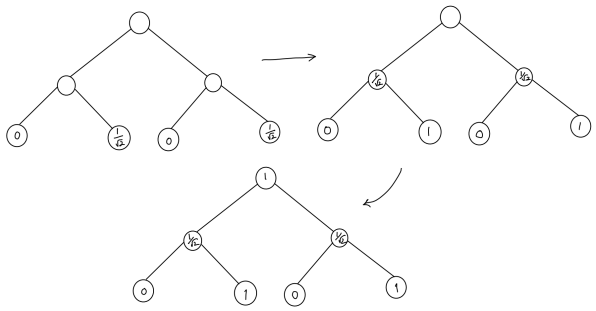
Now using this array we create a binary tree (which I later learned is similar to the bifurcation graph of a QRAM) of the following structure.

So basically, as we go lower in the tree we can represent larger states. For every parent node $|v\rangle$ we have $|v\rangle \otimes |0\rangle$ and $|v\rangle \otimes |1\rangle$ as its two children nodes.

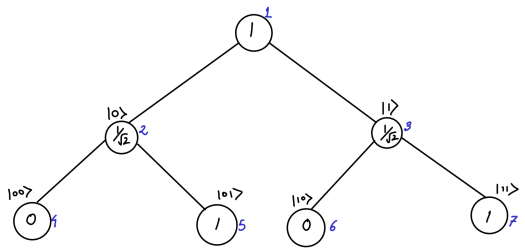
Now, each node (which represents a state) stores a value (amplitude) k such that there is a probability of k^2 to reach it from its parent node. Thus, we can use this

tree structure to represent any arbitrary superposition of its leaves.

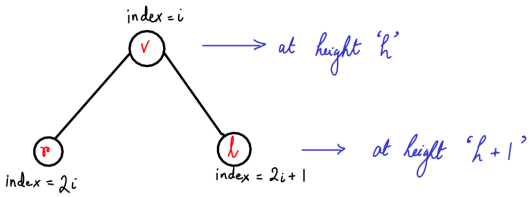
We build this tree bottom up and the amplitudes are shifted from children nodes to parent node, similar to as described below.



So finally we have the following complete tree.



This tree represents the rolled back probability amplitude distributions and the structure is used to apply controlled rotations on the qubits to obtain the desired state represented by the overall amplitude distribution in leaf nodes (namely $[0, \frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}]$).



6.3.1 Computation at every node

Suppose, we are at index i which falls at height h. Based on the value l, the angle of rotation will be decided. When we are at node i, the rotational gate applied is $R_y(\theta)$ on the $q[h+1]$ (the qubit at height $h+1$).

$$R_y(\theta) = \begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}$$

Here, $\theta = 2\cos^{-1}(l)$. Whether or not $R_y(\theta)$ is applied, depends on the control qubits (all qubits till the height h represented by $q[:h+1]$) and control values (binary representation of i without the most significant bit).

The $R_y(\theta)$ gate is applied only when the values of the list $q[:h+1]$ are same as that of the list of control values.

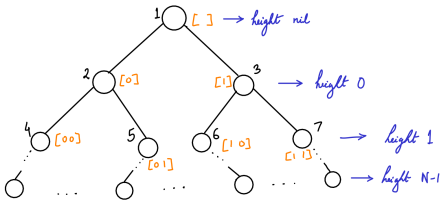
For example, if $q[:h+1] = [|+\rangle, |0\rangle]$ and control values $= [1, 0]$ then only 50% of times the associated rotation would be applied, if we were to simulate the circuit ex-

ecution. The below code shows the implementation of the above idea of controlled rotation in `cirq`.

```
cirq.YPowGate(exponent=angle)
    .on(qubits[h])
    .controlled_by(*qubits[h],control_values=bits)
```

6.3.2 Conclusion

Controlled rotations like as explained above are applied at every node other than the leaf nodes. Thus, the last controlled rotations are applied on the qubit at the height of the leaves.



Here, we have a total number of qubits = N and total nodes = $2^0 + 2^1 + ... + 2^N$. Respective list of control values for computation at every node is shown in orange.

6.4 Epilogue

After all operations are performed, the final circuit is obtained. The value of the state vector can be easily tested upon simulating the circuit. In the example with original vector $[1, 5, 7, 10]$ the circuit obtained is as follows.

```
0: ──Y^0.5──(0)──@──
             |   |
1: ──────────Y───Y──
```

Upon simulating the circuit, we obtain the representation of the required state vector (quantum state with required superposition defined according to the problem).

$$[0j,\ (-0.5+0.5j),\ 0j,\ (-0.5+0.5j)]$$

On removing global phase, we get the above quantum state as $[0, \frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}]$ which is basically what we wanted.

$$[0, \frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}] \equiv \frac{1}{\sqrt{2}}|01\rangle + |11\rangle$$

Moreover, the documentation provided in the code(s) is quite extensive and can be referred to for any further doubts or query regarding the implementation of the above idea.

6.5 Additional Study

Creating arbitrary superpositions is a subproblem. It comes under the bigger problem of creating any desired quantum state. It would be fun to provide some light on known practices to solve this problem as well.

`Qiskit` (by IBMQ) uses a method proposed by Shende et. al. which in short assumes to have started with the desired quantum state and reduces that state to

$|00\ldots0\rangle$ using a circuit. As a result, the reverse of this circuit would provide us with the necessary initialization circuit.

6.5.1 TLDR of the Idea

Given any state $|a\rangle$, we can construct $R_y(-\theta)R_Z(-\phi)$ such that we obtain the following.

$$R_y(-\theta)R_Z(-\phi)|a\rangle = re^{i\gamma}|0\rangle$$

Now, when there are more than one qubit, say n qubits, we factorize the state vector

$$\begin{aligned}|\psi\rangle &= k_1|0...00\rangle + k_2|0...01\rangle + k_3|0...10\rangle + \cdots + k_{2^n}|1...11\rangle \\ &= |0...0\rangle(k_1|0\rangle + k_2|1\rangle) + \cdots + |1...11\rangle(k_i|0\rangle + k_{2^n}|1\rangle) \\ &= |0...0\rangle(|a\rangle_1) + \cdots + |1...11\rangle(|a_{2^{n-1}}\rangle)\end{aligned}$$

such that $R_y(-\theta_j)R_Z(-\phi_j)|a\rangle_j = r_j e^{i\gamma_j}|0\rangle$ for a certain index j.

$$U = \begin{bmatrix} R_y(-\theta_1)R_Z(-\phi_1) & & & \\ & R_y(-\theta_2)R_Z(-\phi_2) & & \\ & & \ldots & \\ & & & R_y(-\theta_{2^{n-1}})R_Z(-\phi_{2^{n-1}}) \end{bmatrix}$$

Therefore, the above unitary U can be implemented as a "quantum multiplexor" gate (since it is a block diagonal matrix) such that upon applying it to $|\psi\rangle$ we get the following result.

$$U|\psi\rangle = \begin{bmatrix} r_1 e^{i\gamma_1} \\ \ldots \\ r_{2^{n-1}} e^{i\gamma_{2^{n-1}}} \end{bmatrix} \otimes |0\rangle$$

In short, this is how $\langle$`circuit`$\rangle$`.initialize(`$\langle$`desired_vector`$\rangle$`,`$\langle$`list of qubits`$\rangle$`)` is accurately implemented in `Qiskit`.

7 QUANTUM ML: TRAINING A VARIATIONAL CIRCUIT

This work was done in collaboration with **Zeeshan Ahmed** and **Shreyas Pradhan**.

7.1 Problem Statement

To train a quantum variational circuit that will transform according to the following map:

$$\begin{aligned}f(0000) &= 0011 \\ f(0001) &= 0101 \\ f(0010) &= 1010 \\ f(0011) &= 1100\end{aligned}$$

The input states were chosen arbitrarily, and the output states were fixed. The transformation for other states is irrelevant.

Framework used: Pennylane.

7.2 QVCs

QVCs are the prime example of the intersection between quantum computing and classical machine learning.

A quantum circuit can be taken as a model with parameters of different kinds that can be trained for a transformation.

There are two kinds of parameters:

- Adaptable $\hat{\theta} \rightarrow (\theta_1, \cdots \theta_n)$
- Non-adaptable $\hat{x} \rightarrow (x_1, \cdots x_n)$

The whole circuit can be modeled as one single unitary transformation that is parameterized by these parameters, and the transformation can be written as: $U(\hat{\theta}, \hat{x})$.

To use machine learning, we need to convert the qubit states into classical information. For this, we shall use observables. To maximize the distance between the possible outputs, we use 'PauliZ', which has $\{-1, 1\}$ as the observable values.

Note: We also need to convert the target values into the above format to draw reasonable accuracy and cost.

The circuits' gates can be thought of as hyperparameters that do not change while they are being trained. To capture the function's entire essence, we need to have both single qubit and multi-qubit gates.

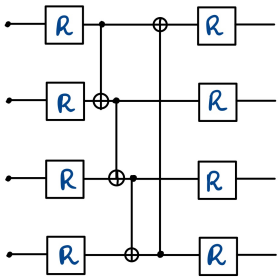


Figure 5: Circuit

7.3 The Citcuit

- The CNOT gates were required to count for the relation between different qubits.
- A saturation at 1.4 cost was observed with one layer of R gates and CNOT gates.
- Adding another layer of R gates dropped the cost to 0.6.
- Each R gate takes in 3 parameters:

$$R(\theta, \phi, \omega)$$

- **Embedding**: Before the input can be fed into the circuit, it must be converted into a state. We will use the basis embedding:

$$0000 \rightarrow 0000$$

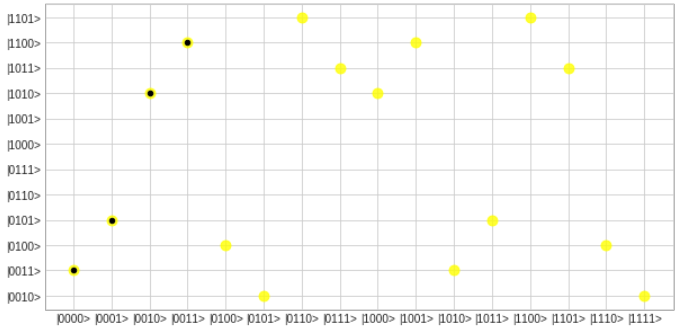
The kind of embedding determines the non-adaptable parameters of the circuit.

- **Cost Function**: Since the distance between the predicted state and the required state directly can't be made, we instead apply our loss function over the expected value of the observable `PauliZ`.
- **Training**: The QVC algorithm is now used to train the circuit, i.e. find the correct parameters for each of the gates. The parameters can be modeled into a weight matrix W.
- **Initialization**: The initial values of the matrix are set to small random values.
- **Optimization**: The weight matrix will now have to be optimized to reduce the cost, for which we will use gradient descent paired with Nesterov Momentum optimization.

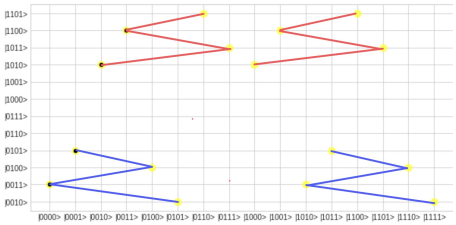
7.4 Results

The learning rate was initially set to 0.5 which did not give low cost due to a great amount of fluctuation in the parameters. The rate was then gradually reduced and decent results were obtained at 0.01 learning rate.

Final Cost: After 200 iterations the final cost was 0.6856744 with full accuracy.



- After the network was trained, it was tested for all the 16 possible classical inputs to the circuit. For visualization, it is easier to scatter plot both the prediction and the expected value. The plot will also help us draw insights about the states' transformations that aren't considered on the map.
- The diagram (next slide) shows a recurring pattern since the training samples had only the first 4 points. The rest of the values are adjusted accordingly, and the pattern is repeated.



The graph doesn't change when run multiple times, that means the function generated is not random for the other points but is deterministic based on the training samples.

8 NOTES ON QUANTUM INFORMATION

8.0.1 Models of Classical Computation

The following serve as universal models of computation:

- Turing Machines
- λ-calculus
- Circuits

8.0.2 Super-Universal

Circuits can describe computations which are beyond what a Turing machine can do. We can indeed solve a constrained version of the halting problem under some circuit model.

8.0.3 Landauer's Principle

Landauer's Principle explains that when info is erased, it requires work. Each bit erased $\implies \Delta S = -K\ln(2)$. Hence, heat associated is given by $Q = -KT\ln(2)$.

Hence, to keep constant temp, that amount of work needs to be put in. This generalises to reversible processes requiring work.

8.0.4 Quantum Computation

The idea of a QC model is to have information in superposition. If we want Turing Machines to be part of the quantum computational model then we will somehow have to make the read/write head to be in a superposition. But that is not possible (doubt). Hence we use a circuit where we keep the bits in a superposition and operate on them with gates.

8.0.5 History of Quantum Mechanics and Computation

Major questions in QM started arising with the EPR paradox in 1936 when faster than light travel was questioned.

Here's the paradox: Consider 2 particles that are moving away. Their positions are x and $-x$ and momenta are p and $-p$. Now we can't measure the exact position and momentum of one particle due to the uncertainty principle. But what if we measure the position of one particle and momentum of the other at the same time (Like a split second difference)? Assuming there is no interaction between the particles, we would in principle know the position and momentum of one particle at a time (cause we can know position or momentum of particle 1 if we measure that quantity on the second particle cause its just the negative of it). This would violate the uncertainty principle. Hence, there has to be some sort of interaction between the two particles that makes the wave function of particle 2 also collapse instantaneously when I measure position or momentum of particle 1. This mean that the particles were somehow connected. This was the paradox. But Schrodinger said this is possible by entanglement.

In 1964 Bell gave his Theorem and then it was later verified too which proved that Quantum Mechanics is cool with EPR pair. But there was another problem that how is faster than light communication possible? It was later proved that faster than light communication is not possible indeed as you cannot actually send 'information' with the entanglement coordination. Basically you cannot manipulate the

particles to actually transmit any useful information.

Non-cloning theorem: In physics, the no-cloning theorem states that it is impossible to create an independent and identical copy of an arbitrary unknown quantum state, a statement which has profound implications in the field of quantum computing among others.

8.1 On Formulation of Quantum Mechanics

Quantum Mechanics had a really dysfunctional and troubled childhood. And more often than not, we are taught of how it grew rather than what it stands for. As a by-product, quantum mechanics comes across as some strange physical theory. However, that is no where near the whole picture.

Quantum Theory is just as much (if not more) an extension of Classical Probability – as it is a physical theory that describes the nature of the universe. Just as Classical Probability Theory, Quantum Mechanics can be fundamentally formulated by pure thought alone without any particular appeal to experiment.

It is what we get upon conserving the L_2 norm rather than the L_1 norm (as in Classical Probability) along with addition of the continuity axiom and the idea of measurement.

8.1.1 Main Axioms

The axiomatic formulation of Quantum Mechanics can be stated as follows.

- The state of a system encodes probability of outcomes in a vector, say $[\alpha_1, \alpha_2, \ldots \alpha_n]$, such that $\sum_{\forall i} |\alpha_i|^2 = 1$, or that the L_2 norm is preserved.
- There exists a continuous reversible transformation on a system between any two pure states of that system. This is called the axiom of continuity.
- Measurement in a standard basis results in a collapse of the state to whatever outcome is obtained. The outcome is governed by the probability distribution.

Let us not worry about measurement or continuity, for a while, and deal with the idea of preservation of norm.

Classical Probability involves preservation of L_1 norm and thus all associated transformations are required to be L_1 preserving and they are referred to as stochastic matrices.

In case of Quantum Mechanics, all transformations require to be L_2 norm preserving and the matrices which possess this property, upon introducing complex numbers as well, are UNITARY MATRICES!

But what about other norms? Or, is it that God is partial to L_1 and L_2? Turns out that God is indeed partial!

Theorem 3. *Let* T *be a linear transformation such that it preserves* L_p*. Then, if* T *is a non-trivial transformation such that it doesn't involve re-shuffling of the elements of the vector – then* $p = 1$ *or* $p = 2$*.*

Building upon our major axioms we also need a formulation for building composite systems.

- Associated with every system, exists degree of freedom and dimension.

- Degree of Freedom of a system (K) denotes the minimum number of probability measurements needed to determine the state.
- Dimension of a system (N) denotes the maximum number of states that can be reliably distinguished from one another in a single shot measurement.
- A composite system consisting of two subsystems A and B having dimension N_A and N_B respectively, and number of degrees of freedom K_A and K_B respectively, has dimension $N = N_A N_B$ and number of degrees of freedom $K = K_A K_B$.

Now, let's see if we can derive some of the idiosyncrasies of Quantum Mechanics from the above mentioned axioms. Moreover, we shall also discuss the two most basic constraints which serve as the backbone of no-go theorems.

- Linearity
- Unitarity

Any consequences of linearity and unitarity of quantum theory must be respected.

- **Why do we use complex numbers, instead of real numbers?** The axiom of continuity requires us to operate on a field which is algebraically closed. Thus, we need complex numbers. In simpler terms, if you want every unitary operation to have a square root, then you have to go to the complex numbers.
- **Why is quantum computing reversible?** Any quantum gate must thus be implemented as a unitary operator and is therefore reversible.
- **Why Linearity?** When, we stated out our axioms – linearity was implied in the idea of state transformation. However, let us move out of our formulation and think about why linearity might be necessary.
  The first compelling argument for linearity was made by Wigner and Bargmann who proved that quantum dynamics must be linear if it does not change absolute values of inner products of state vectors.
  However, the assumption – upon which this argument stands – is not strong enough. In fact, Steven Weinberg among others have indeed come up with non-linear formulations of Quantum Mechanics.
  So, can we come up with an argument that is stronger? Fortunately, we can! And, guess what? It is inherently based on Computational Complexity. If quantum mechanics were non-linear, then one could build a computer to solve **NP**-complete problems in polynomial time.

8.2 On Linear Algebra

The field of Quantum Computation and Information relies heavily on the understanding of Linear Algebra. Here, we shall provide a recap of the same.

8.2.1 Linear Operators and Matrices

$$A(\Sigma a_i |v_i\rangle) = \Sigma a_i A|v_i\rangle$$

Now, see a linear operator is just a matrix. Suppose $A : V \to W$ and $|v_1\rangle, |v_2\rangle, ..., |v_m\rangle$ are basis of V and $|w_1\rangle, |w_2\rangle, ..., |w_n\rangle$ is a basis of W then,

$$A|v_j\rangle = \Sigma_i A_{ij} |w_i\rangle$$

8.2.2 Inner Products

Ok so imagine an operation $(\_,\_) : V \times V \to \mathbb{C}$ such that the following shit holds ok?

1. $(|v\rangle, \Sigma_i \lambda_i |w_i\rangle) = \Sigma_i \lambda_i (|v\rangle, |w_i\rangle)$
2. $(|v\rangle, |w\rangle) = (|w\rangle, |v\rangle)^*$
3. $(|v\rangle, |v\rangle) \geqslant 0$ and $= 0$ iff $|v\rangle$

In finite dimensions, inner product space i.e., vector spaces equipped with inner prducts for all $|v\rangle \in$ vector space = Hilbert Space. Consider $|i\rangle$ & $|j\rangle$ to be orthonormal basis, we have the following.

$$\langle v|w\rangle = (\Sigma_i v_i |i\rangle, \Sigma_j w_j |j\rangle) = \Sigma_i \Sigma_j v_i^* w_j \delta_{ij} = \Sigma_i v_i^* w_i = |v\rangle^\dagger |w\rangle$$

8.2.3 Norm of a vector

$$\|v\| = \sqrt{\langle v|v\rangle}$$

We can say that $|v\rangle$ is normalized iff $\|v\| = 1$. A set of $|a_i\rangle$ vectors is orthonormal if $\langle a_i|a_j\rangle = \delta_{ij}$ i.e., $\forall\, i \neq j\ \langle a_i|a_j\rangle = 0$ and $\langle a_i|a_j\rangle = 1\ \forall\, i = j$.

8.2.4 Gram Schmidt: for orthonormal basis

$$|v_{k+1}\rangle = \frac{|w_{k+1}\rangle - \Sigma_{i=1}^{k} \langle v_i|w_{k+1}\rangle |v_i\rangle}{\| |w_{k+1}\rangle - \Sigma_{i=1}^{k} \langle v_i|w_{k+1}\rangle |v_i\rangle \|},\ |v_1\rangle = |w_1\rangle / \||w_1\rangle\|$$

8.2.5 Outer Product

$$|w\rangle\langle v|(|v'\rangle) = |w\rangle |v\rangle^\dagger |v'\rangle = |w\rangle\langle v|v'\rangle = \langle v|v'\rangle |w\rangle$$

- From this notion we obtain the completeness relation, $\Sigma_i |i\rangle\langle i| = I$.
- $A = I_w A I_v = \Sigma_{ij} |w_j\rangle\langle w_j|A|v_i\rangle\langle v_i| = \Sigma_{ij} \langle w_j|A|v_i\rangle |w_j\rangle\langle v_i|$
- Cauchy Schwarz: $\langle v|v\rangle\langle w|w\rangle \geqslant \langle v|w\rangle\langle w|v\rangle = |\langle v|w\rangle|^2$

8.2.6 Hilbert Space

A Hilbert Space $\mathcal{H}$ is complete which means that every Cauchy sequence of vectors admits in the space itself. Under this hypothesis there exist Hilbert bases also known as complete orthonormal systems of vectors in $\mathcal{H}$. For any orthonormal basis of $\mathcal{H}$, we have the following.

$$\text{Orthonormality} \equiv \langle \psi_i|\psi_j\rangle = \delta_{ij}$$
$$\text{Completeness} \equiv \sum_i |\psi_i\rangle\langle\psi_i| = I$$

8.2.7 Eigenvectors and Eigenvalues

Under a given linear transformation A, $A|v\rangle = \lambda|v\rangle$ where $\exists\, |v\rangle$ s.t. they do not get shifted off their span.

All such vectors are referred as eigenvectors and $(A - \lambda I)|v\rangle = 0 \implies \det|A - \lambda I| = 0$ gives all possible eigenvalue. If all $\lambda_i \geqslant 0$, it is positive semi-definite and if they are > 0, it is positive definite.

8.2.8 Eigenspace

It is the space of all vectors with a given eigenvalue λ. When an eigenspace is more than one dimensional, we call it degenerate.

8.2.9 Adjoints and Hermitian

Suppose $A : V \to V$ then $\exists\ A^\dagger : V \to V$ such that $\forall\ |v\rangle,\ |w\rangle \in V$ we have,

$$(|v\rangle, A|w\rangle) = (A^\dagger|v\rangle, |w\rangle)$$

This operator is called as the adjoint or Hermitian conjugate of the operator A.

$$(|v\rangle, A|w\rangle) = \langle v|A|w\rangle = \mathbf{v}^\dagger A\mathbf{w} = (A^\dagger\mathbf{v})^\dagger\mathbf{w} = (A^\dagger|v\rangle, |w\rangle)$$

- We have, $(AB)^\dagger = B^\dagger A^\dagger$
- $|v\rangle^\dagger = \langle v|$

8.2.10 Some defintions

- Normal matrices: $AA^\dagger = A^\dagger A$
- Hermitian matrices: $A^\dagger = A$
- Unitary matrices: $AA^\dagger = I$
- A normal matrix is Hermitian if and only if it has real eigenvalues.
- If $\langle x|A|x\rangle \geqslant 0, \forall\ |x\rangle$ then A is positive semi-definite and has positive eigenvalues.

8.2.11 Some properties

- If a Hermitian matrix has positive eigenvalues then it is positive semi-definite.
- If $M = AA^\dagger$ then it is both Hermitian and positive semi-definite.
- All positive semi-definite operators are Hermitian, by definition.

8.2.12 Spectral Decomposition

A linear operator is diagonalizable if and only if it is normal. Some notes and derivation regarding the above: $A\vec{v} = \lambda\vec{v} = \Sigma_i \lambda_{ij}\vec{q_i}$ where q_i's are linearly independent eigenvalues of A.

$$AQ = Q\Lambda \text{ where } Q = \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix} \implies A = Q\Lambda Q^{-1}$$

$$A = IAI = (P + Q)A(P + Q) = PAP + QAP + PAQ + QAQ$$

$$\implies A = \lambda P^2 + 0 + 0 + QAQ$$

$$\implies A = \lambda P^2 + QAQ$$

8.2.13 Matrices and Vectors

In the following statements we are dealing with $\{|i\rangle\}$ as a *orthonormal* basis set.

$$I = \Sigma |i\rangle\langle i|$$

$$|\psi\rangle = \Sigma \sigma_i |i\rangle, \text{ where } \sigma_i = \langle i|\psi\rangle$$

Now, to represent a operator or linear transformation as matrix in orthonormal basis.

$$A_{ij} = \langle i|A|j\rangle$$

$$A = \Sigma_{i,j} \langle i|A|j\rangle |i\rangle\langle j|$$

$$\text{tr}(A) = \Sigma_i \langle i|A|i\rangle$$

Now, diagonalization for any normal matrix.

$$M = \Sigma_i \lambda_i |i\rangle\langle i| = \Sigma_i \lambda_i P_i, \text{ where } P_i^\dagger = P_i$$

$$f(M) = \Sigma_i f(\lambda_i) |i\rangle\langle i|$$

where λ_i are eigenvalues of M under a given orthonormal basis set $\{|i\rangle\}$ for vector space V, each $|i\rangle$ is an eigenvector of M with eigenvalue λ_i.

If M is Hermitian, all eigenvalues (λ_i s) are non-negative.

8.2.14 Tensor Products

$$\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \otimes \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} = \begin{bmatrix} a_{1,1}\begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} & a_{1,2}\begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \\ a_{2,1}\begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} & a_{2,2}\begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} a_{1,1}b_{1,1} & a_{1,1}b_{1,2} & a_{1,2}b_{1,1} & a_{1,2}b_{1,2} \\ a_{1,1}b_{2,1} & a_{1,1}b_{2,2} & a_{1,2}b_{2,1} & a_{1,2}b_{2,2} \\ a_{2,1}b_{1,1} & a_{2,1}b_{1,2} & a_{2,2}b_{1,1} & a_{2,2}b_{1,2} \\ a_{2,1}b_{2,1} & a_{2,1}b_{2,2} & a_{2,2}b_{2,1} & a_{2,2}b_{2,2} \end{bmatrix}.$$

- $z|vw\rangle = (z|v\rangle) \otimes (|w\rangle) = (|v\rangle) \otimes (z|w\rangle)$
- $(|v_1\rangle + |v_2\rangle) \otimes |w\rangle = |v_1 w\rangle + |v_2 w\rangle$
- $|v\rangle \otimes (|w_1\rangle + |w_2\rangle) = |vw_1\rangle + |vw_2\rangle$
- $|\psi\rangle^{\otimes k} = |\psi\rangle \otimes \ldots \otimes |\psi\rangle$ k times
- $(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger$

8.2.15 Linear Product

$A \otimes B$ forms the linear operator that acts on $V \otimes W$ vector space givern that A acts on V and B acts on W.

$$(A \otimes B)(\Sigma_i a_i |v_i\rangle \otimes |w_i\rangle) = \Sigma_i a_i A|v_i\rangle \otimes B|w_i\rangle$$

8.2.16 Inner Product

$$(\Sigma_i a_i |v_i\rangle \otimes |w_i\rangle, \Sigma_i a_j |v'_j\rangle \otimes |w'_j\rangle) = \Sigma_{ij} a_i^* b_j \langle v_i|v'_j\rangle\langle w_i|w'_j\rangle$$

8.2.17 Trace

Properties of trace are given below as follows.

- $\text{tr}(A) = \Sigma_i A_{ii}$
- $\text{tr}(A) = \Sigma_i \langle i|A|i\rangle$ for orthonormal basis
- $\text{tr}(AB) = \text{tr}(BA)$
- $\text{tr}(zA + B) = z \cdot \text{tr}(A) + \text{tr}(B)$

The above properties yield certain implications as follows.

- $\text{tr}(UAU^\dagger) = \text{tr}(A)$
- $\text{tr}(A|\psi\rangle\langle\psi|) = \Sigma_i \langle i|A|\psi\rangle\langle\psi|i\rangle$
- $\text{tr}(A) = \sum_i \lambda_i$, $\det(A) = \prod_i \lambda_i$ with algebraic multiplicities

$$\|A\| = \sqrt{\text{tr}(A^T A)}$$

8.2.18 Partial Trace

Entanglement excludes the possibility of associating state vectors with individual subsystems. Therefore, we introduce density matrices and the corresponding idea of reduction preformed with partial trace.

$$\text{tr}(A \otimes B) = \text{tr}(A) \cdot \text{tr}(B)$$

$$\rho_{AB} : \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B} \xrightarrow{\text{tr}_B} \rho_A : \mathcal{H}_\mathcal{A}$$

$$\text{tr}_B(AB) = A\,\text{tr}(B)$$

8.2.19 Hilbert-Schimdt Inner Product

L_v forms the vector space of operators over the Hilbert space V. Then, we can show that L_v is also a Hilbert space with ****$\text{tr}(A^\dagger B)$ as the inner product operator on $L_v \times L_v$.

Also, we have $\text{div}(L_v) = \dim(V)^2$.

8.2.20 Commutator and Anti-commutator

$$[A, B] = AB - BA$$
$$\{A, B\} = AB + BA$$

8.2.21 Theorem of Simultaneous Diagonalization

Suppose A and B are both Hermitian matrices, then $[A, B] = 0$ iff $\exists$ orthonormal basis such that both A and B are diagonal with respect to that basis.

8.2.22 Polar Value Decomposition

If A is any linear operator and U is a unitary then J, K are positive operators, such that

$$A = UJ = KU, \text{ where } J = \sqrt{A^\dagger A} \text{ and } K = \sqrt{AA^\dagger}$$

Moreover, if A^{-1} exists, then U is unique.

8.2.23 Singular Value Decomposition

SVD in general is given as $U\Sigma V^T$

- It generalizes the eigen decomposition of a square normal matrix with an orthonormal eigen basis to any $m \times n$ matrix.
- Σ is an $m \times n$ rectangular diagonal matrix with non-negative real numbers on the diagonal (called singular values).
- Corollary: If A is a square matrix and $\exists\ U, V$ unitaries then D is a diagonal matrix, such that
$$A = UDV$$
where D has non-negative values.
- Corollary: If A has non-negative eigenvalues then, $A = U^\dagger DU$ is possible where D has non-negative values.
- If A is square both SVD and EVD exist but might not be same.
- If A is a square symmetric matrix both SVD and EVD exist and are equivalent.
- If A is non-square only SVD is possible.

8.2.24 Rank of a matrix

- Rank = number of dimensions in column space.
- The row rank is the largest number of rows of A that constitute a linearly independent set.
- The column rank is the largest number of columns of A that constitute a linearly independent set. Moreover, column-rank = row-rank for $A \in \mathbb{R}^{m\times n}$.
- $rank(A \in \mathbb{R}^{m\times n}) \leqslant min(m, n)$ and matrix is called full rank if equality holds.
- $rank(A^T) = rank(A)$
- $rank(AB) \leqslant min(rank(A), rank(B))$
- $rank(A + B) \leqslant rank(A) + rank(B)$

8.2.25 Projection and Spaces

$$Proj(y; A) = argmin_{v\in R(A)}\|v - y\|_2 = A(A^TA)^{-1}A^Ty$$

- $\mathcal{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$ denotes all vectors in $\mathbb{R}^n$ that land at the origin after transformation. It is also called kernel.
- $\mathcal{R}(A) = \{v \in \mathbb{R}^m : v = Ax,\ x \in \mathbb{R}^n\}$ denotes the space spanned by the transformed basis vectors in $\mathbb{R}^n$.
- $\mathcal{R}(A^T)$ and $\mathcal{N}(A)$ are orthogonal spaces which together span $\mathbb{R}^n$.
- Determinant $\neq 0$ implies that the matrix has an inverse.

8.2.26 Quadratic Forms

Reminder: we are in real $\mathbb{R}$ space.

$$x^TAx = (x^TAx)^T = x^T(\frac{1}{2}A + \frac{1}{2}A^T)x$$

8.2.27 Moore-Penrose Pseudoinverse

$$A^{\text{left inv}} = (A^\dagger A)^{-1} A^\dagger$$

$$A^{\text{right inv}} = A^\dagger (A A^\dagger)^{-1}$$

This is a pseudo inverse formalism with left and right inverses.

$$A^{\text{left inv}} A = I$$
$$A A^{\text{right inv}} = I$$

8.3 Postulates of Quantum Mechanics

Here, I shall state here the four major generic postulates of Quantum Mechanics stated in terms of both state-vector formalization and density-matrix formalization.

8.3.1 State is a vector

1. Isolated physical system is given by its state vector operating on a certain Hilbert space.
2. Evolution of a closed quantum system is given by a unitary transformation. In its physical interpretation we have this postulate governed by the Schrodinger Equation, as stated.
$$i\hbar \frac{d|\psi\rangle}{dt} = H|\psi\rangle$$
The Hamiltonian is a hermitian operator and has a spectral decomposition, $H = \sum E|E\rangle\langle E|$.
3. The state space of a composite physical system is the tensor product of the state spaces of the component systems.
$$|\psi\rangle = |\psi_1\rangle \otimes ... \otimes |\psi_n\rangle$$
4. Quantum measurements are described by a collection $\{M_m\}$ of measurement operators acting on the state space of the system. Probability that upon measurement the outcome is $m = p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle$ and the state of the system becomes as follows.
$$|\psi\rangle \xrightarrow{\text{on measuring}} \frac{M_m|\psi\rangle}{\sqrt{p(m)}} = \frac{M_m|\psi\rangle}{||M_m|\psi\rangle||}$$
Measurement operators also follow the completeness equation, described as follows.
$$\sum_m M_m^\dagger M_m = I$$

8.3.2 State is a density matrix

1. Isolated physical system is given by its density matrix operating on a certain Hilbert space.
2. Evolution of a closed quantum system is given by a unitary transformation as $\rho \xrightarrow{U} U\rho U^\dagger$.
3. The state space of a composite physical system is the tensor product of the state spaces of the component systems.
$$\rho = \rho_1 \otimes ... \otimes \rho_n$$

4. Quantum measurements are described by a collection $\{M_m\}$ of measurement operators acting on the state space of the system. Probability that upon measurement the outcome is $m = p(m) = tr(M_m^\dagger M_m \rho)$ and the state of the system becomes as follows.

$$\rho \xrightarrow{\text{on measuring}} \frac{M_m \rho M_m^\dagger}{p(m)}$$

Measurement operators follow the completeness equation, $\sum_m M_m^\dagger M_m = I$.

8.4 Measurement

Quantum Measurements are descrivbed by a collection of measurement operators $\{M_m\}$ with $p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle$, $\sum p(m) = 1$.

$$|\psi\rangle \xrightarrow{\text{on measurement}} \frac{M_m|\psi\rangle}{\sqrt{p(m)}}$$

If $M_i = |i\rangle\langle i|$, $\forall i$ then we are measuring in a computational basis.

8.4.1 Non-distinguishability of arbitrary states

We cannot distinguish any two arbitrary non-orthogonal quantum states.

Proof: Let $|\psi_1\rangle$ and $|\psi_2\rangle$ be two non-orthogonal states. Then $\exists\ E_1, E_2$ such that

$$E_1 = \sum_{j:f(j)=1} M_j^\dagger M_j$$

and similarly E_2 then $\langle\psi_1|E_1|\psi_1\rangle = 1$ and $\langle\psi_2|E_2|\psi_2\rangle = 1$.

Thus, $\sqrt{E_2}|\psi_1\rangle = 0$ and let $\psi_2 = a\psi_1 + b\phi$ where ψ and ϕ are orthogonal.

$$\langle\psi_2|E_2\psi_2\rangle = |\beta|^2\langle\psi_2|\phi\psi_2 \leqslant |\beta|^2 < 1$$

Hence, proved by contradiction.

8.4.2 Projective Measurements

We can use projective measurement formalism for any general measurement too. In case of projective measurements, $M = \sum mP_m$ and $p(m) = \langle\psi|P_m|\psi\rangle$.

$$|\psi\rangle \to \frac{P_m|\psi\rangle}{\sqrt{p(m)}}$$

$$E(M) = \sum mp(m) = \langle\psi|M|\psi\rangle = \langle M\rangle$$

$$\Delta(M) = \sqrt{\langle M^2\rangle - \langle M\rangle^2}$$

Here, $E(M)$ is expectation and $\Delta(M)$ is standard deviation or the square root of variance.

8.4.3 POVM Measurements

POVM Measurements are a formalism where only measurement statistics matters.

$$\{E_m\} \to \sum E_m = I,\ p(m) = \langle\psi|E_m|\psi\rangle$$

Here, each of E_m are hermitian.

8.4.4 Global Phase doesn't matter

We say $e^{i\theta}|\psi\rangle \equiv |\psi\rangle$ but why? Because, $\langle\psi|M_m^\dagger M_m|\psi\rangle = \langle\psi|e^{-i\theta}M_m^\dagger M_m e^{i\theta}|\psi\rangle$. However, be aware that the global phase is quite different from the relative phase.

8.5 Density Matrices

We can represent a system as an ensemble of pure states $\{p_i, \psi_i\}$. Now, if you have exact knowledge of the system then it is for sure in a pure state, i.e., $\rho = |\psi\rangle\langle\psi|$. However, if we have classical uncertainty amongst the possible states. The system can be represented as a mixed state $\rho = \sum_i p_i\psi_i$ where the probabilities p_i are classical in nature.

This formulation helps us a lot in dealing with quantum information, noisy systems and helps us represent measurements better, as well. Why? Because it provides a convenient means for describing quantum systems whose state is not completely known.

8.5.1 Properties

- **Theorem 1**: An operator ρ is a density operator if and only if it is both positive ($\rho = \rho^\dagger$) and $tr(\rho) = 1$. Converse is easy to prove. If an operator is both positive and has trace as one, then it shall have a spectral decomposition of the form $\sum_i \lambda_i|i\rangle\langle i|$. For the direct proof, let us consider $\rho = \sum_i p_i tr(|\psi_i\rangle\langle\psi_i|) = 1$.

- **Theorem 2**: The sets $|\tilde{\psi}\rangle$ and $|\tilde{\phi}\rangle$ generate the same density matrix if and only if,

$$|\tilde{\psi}_i\rangle = \sum_j u_{ij}|\tilde{\phi}_j\rangle$$

  where u_{ij} is a unitary matrix of complex numbers, with indices i and j, and we 'pad' whichever set of vectors $|\tilde{\psi}_i\rangle$ or $|\tilde{\phi}_j\rangle$ is smaller with additional vectors 0 so that the two sets have the same number of elements. As a consequence of the theorem, note that $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i| = \sum_j q_j|\phi_j\rangle\langle\phi_j|$ if and only if we have the following as true for some unitary matrix u_{ij}

$$\sqrt{p}_i|\psi_i\rangle = \sum_j u_{ij}\sqrt{q}_j|\phi_j\rangle$$

- **Theorem 3**: If ρ is a density operator, then ρ is a pure state if and only if $tr(\rho^2) = 1$ and mixed state if and only if $tr(\rho^2) < 1$.

- **Theorem 4**: Observable M has expectation $\sum_x \langle\psi_x|M|\psi_x\rangle = tr(M\rho)$.

8.5.2 Reduced Density Operator

This is the single-most important application of density operator formulation is the existence of reduced density operator. It is defined as follows.

$$\rho_A = tr_B(\rho_{AB})$$

This allows us to talk about sub-systems of a composite system.

$$tr_B(|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|) = |a_1\rangle\langle a_2|tr(|b_1\rangle\langle b_2|)$$

8.5.3 Schmidt Decomposition

Suppose $|\psi\rangle$ is a pure state of a composite system, AB. Then, there exists orthonormal states $|i_A\rangle$ for system A, and orthonormal states $|i_B\rangle$ of system B such that

$$|\psi\rangle = \sum_i \lambda_i |i_A\rangle |i_B\rangle$$

where λ_i are non-negative real numbers satisfying $\sum_i \lambda_i^2 = 1$ known as Schmidt co-efficients.

8.5.4 Purification

Suppose we have a mixed state ρ_A for a system A. Then, we can introduce another system R such that AR forms a pure state $|AR\rangle$ and $\rho_A = tr_R(|AR\rangle\langle AR|)$.

Given $\rho_A = \sum_i p_i |i_A\rangle\langle i_A|$, we shall have the following where $|i_R\rangle$ are orthonormal basis states.

$$|AR\rangle = \sum_i \sqrt{p_i} |i_A\rangle |i_R\rangle$$

8.6 Bell's Inequality

In a classical experiment, we have the following setup where Alice can choose to measure either Q or R and Bob chooses either S or T. The measurements are performed imultaneously and far off from each other.

$$E(QS) + E(RS) + E(RT) - E(QT) = E(QS + RS + RT - QT)$$

$$\implies E(QS) + E(RS) + E(RT) - E(QT) = \sum_{q,r,s,t} p(q,r,s,t)(qs + rs + rt - qt)$$

$$\implies E(QS) + E(RS) + E(RT) - E(QT) \leqslant \sum_{q,r,s,t} p \times 2 = 2$$

and thereby we obtain the inequality $E(QS) + E(RS) + E(RT) - E(QT) \leqslant 2$.

This is one of the set of Bell inequalities, the first of which was found by John Bell. This one in particular is named CHSH inequality.

8.6.1 Quantum Anomaly

In the quantum case, let us consider the measurements to be based on the following observables over the EPR pair $|\psi\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$.

$$Q = Z_1, R = X_1$$
$$S = \frac{-Z_2 - X_2}{\sqrt{2}}$$
$$T = \frac{Z_2 - X_2}{\sqrt{2}}$$

Then, we have the following result.

$$E(QS) + E(RS) + E(RT) - E(QT) = \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} = 2\sqrt{2} > 2$$

Thus, in other words, CHSH inequality doesn't hold.

8.6.2 Interpretation

The fact that CHSH doesn't hold in the quantum scenario implies that two of the major assumptions about nature is wrong in case of the classical experiment. The assumptions are:

- **Realism**: Q, R, S, T are physical quantities which have defininte values irrespective of observation.
- **Locality**: Alice's measurement doesn't influence that of Bob's.

Thus, the result of CHSH being false when accounted for the quantum mechanical properties of nature (we can perform the experiment in a lab with particles) suggests that nature cannot be locally real and neither can any true mathematical representation of it be locally real.

8.7 Bloch Sphere and Rotations

We can represent any ρ (density matrix) as $\frac{1}{2}(I + \vec{r} \cdot \vec{\sigma})$.

$$e^{iAx} = \cos(x) + i\sin(x)A$$

Thus, $R_x(\theta) = e^{i\theta X/2}$, $R_y(\theta) = e^{i\theta Y/2}$ and $R_z(\theta) = e^{i\theta Z/2}$.

$$R_{\hat{n}}(\theta) = e^{i\theta \hat{n} \cdot \vec{\sigma}/2}$$

In general, we have the above equation where $\vec{\sigma} = X\hat{i} + Y\hat{j} + Z\hat{k}$.

$$X^2 = Y^2 = Z^2 = -iXYZ = I$$

$$R_{\hat{n}}(\alpha) = R_z(\phi)R_y(\theta)R_z(\alpha)R_y(-\theta)R_z(-\phi) = R_z(\phi)R_y(\theta)R_z(\alpha)R_y(\theta)^\dagger R_z(\phi)^\dagger$$

8.7.1 Theorems

1. Any arbitrary single qubit unitary operator can be written in the form $U = e^{i\alpha}R_{\hat{n}}(\theta)$.
2. Suppose U is a unitary operation over a single qubit then $\exists\ \alpha, \beta, \gamma, \delta$ such that $U = e^{i\alpha}R_{\hat{n}}(\beta)R_{\hat{m}}(\gamma)R_{\hat{n}}(\delta)$.
3. There exists unitaries A, B, C for any given unitary U such that $ABC = I$ and $U = e^{i\alpha}AXBXC$.

8.8 More on Bloch Sphere and Rotations

A single qubit operator can be represented as $U = a_0 I + a_1 X + a_2 Y + a_3 Z$.
Also, such a unitary can also be represented this way,

$$U = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

and thus, we obtain the following equivalences.

$$a_0 = (a+d)/2,\ a_1 = (b+c)/2$$
$$a_2 = (c-b)/2i,\ a_3 = (d-a)/2$$

Also from $UU^\dagger = I$ we get,

$$|a_0|^2 + |a_1|^2 + |a_2|^2 + |a_3|^2 = 1$$
$$a_0^* a_1 + a_1^* a_0 + i a_2^* a_3 - i a_3^* a_2 = 0$$
$$a_0^* a_2 - i a_1^* a_3 + a_2^* a_0 + i a_3^* a_1 = 0$$
$$a_0^* a_3 + i a_1^* a_2 - i a_2^* a_1 + a_3^* a_0 = 0$$

where we define $|a_0| = \cos(\theta/2)$ then $|a_1|^2 + |a_2|^2 + |a_3|^2 = |\sin(\theta/2)|$.
Then define,

$$n_x = |a_1|/|\sin(\theta/2)|$$
$$n_y = |a_2|/|\sin(\theta/2)|$$
$$n_z = |a_3|/|\sin(\theta/2)|$$

and further we get $n_x^2 + n_y^2 + n_z^2 = 1$.

Now, we define $exp(i\alpha) = a_0/\cos(\theta/2)$ and denote the phase of a_1, a_2, a_3 as $\alpha_1, \alpha_2, \alpha_3$ respectively. By putting these in the other constraints we get, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha - \pi/2$.

$$a_0 = e^{i\alpha}\cos(\theta/2),$$
$$a_1 = -ie^{i\alpha}\sin(\theta/2)n_x,$$
$$a_2 = -ie^{i\alpha}\sin(\theta/2)n_y,$$
$$a_3 = -ie^{i\alpha}\sin(\theta/2)n_z$$

$$U = e^{i\alpha}(\cos(\frac{\theta}{2})I - i\sin(\frac{\theta}{2})(n_x X + n_y Y + n_z Z)) = e^{i\alpha}R_{\hat{n}}(\theta)$$

8.9 Quantum Channels

8.9.1 Introduction

Quantum Channels are generalizations of Quantum Operations. Since, in general our operations maybe noisy (inherently) and the system we are dealing with maybe open so Quantum Channels us the study of noisy open Quantum Operations.

$$\rho \xrightarrow{U} \rho'$$

The above only holds for closed quantum systems. Thus, in case of noisy open quantum systems we have the following.

$$\rho \xrightarrow{\mathcal{E}} \rho'$$

Here, $\mathcal{E}$ is a combination of:

- unitaries (U)
- adding systems $(\rho \to \rho \otimes \sigma)$
- subtracting systems or partial tracing $(\rho_{AB} \to tr_B \rho_{AB} = \rho_A)$

8.9.2 Review of Density Matrices

Moreover, to deal with noisy open systems we also need to review density matrices which are the general way of representing noisy quantum states.

$$\rho \equiv \text{noisy quantum state}$$

$$\mathcal{E} \equiv \text{noisy open quantum channel}$$

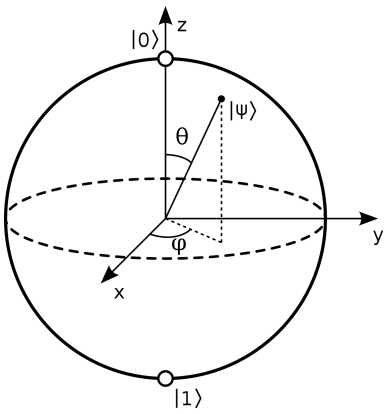


Figure 6: Bloch Sphere

Any general mixed or pure state can be given by the point $\bar{a} = (a_x, a_y, a_z)$ where $|\bar{a}|^2 \leqslant 1$. Now since, $\rho = \frac{1}{2}(I + \vec{a} \cdot \vec{\sigma})$ is positive semi-definite thus, $\text{eigs}(\rho) = \frac{1+||\vec{a}||}{2} \geqslant 0 \implies ||\vec{a}|| \leqslant 1$.

This results in the Bloch Ball representation for general mixed state qubits, which serves as a generalization of the Bloch Sphere that serves as representation for pure state qubits. Pure state qubits have eigenvalues of 0 or 1, $\lambda(\rho) = 0$ or 1.

The eigenvalue of the maximally mixed state is $I/2$ where the state is actually given by the following.

$$\frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) \equiv \frac{1}{2}(|+\rangle\langle +| + |-\rangle\langle -|)$$

8.9.3 Generalized Quantum Operations

We can refer to Generalized Quantum Operations as Quantum Channels (noisy and open) whose definitions we shall rediscover and formalize here.

Now, it should be noted that there are three major equivalent formalisms for Quantum Channels:

- Operational (Steinspring)
- Mathemtically Simplified (Kraus)
- Axiomatic (TPCP maps)

8.9.4 Introduction to Steinspring Representation

For closed systems we have a probabilistic combinations of the above (noisy closed channels).

$$|\psi\rangle \to U|\psi\rangle$$

$$|\psi\rangle\langle\psi| \to U|\psi\rangle\langle\psi|U^\dagger$$
$$\rho \to U\rho\, U^\dagger$$

This can be generalized for noisy open channels as combinations of unitaries (U), adding systems together ($\rho \to \rho \otimes \sigma$) and subtracting systems ($\rho_{AB} \to \mathrm{tr}_B(\rho_{AB}) = \rho_A$).

But why is partial tracing allowed? Doesn't this "delete" information? NO! The information has basically flown out of your lab to the heavens (but still exists in nature) and in most cases, you simply can't access it anymore.

$$\mathrm{tr}(\rho_{AB}) = \sum_b (I_A \otimes \langle b|)\ \rho\ (I_A \otimes |b\rangle) = (I_A \otimes \mathrm{tr}_B)(\rho_{AB})$$

8.9.5 Isometry

An operator $V \in \mathcal{L}(X),\ V : X \to Y$ is called an isometry if and only if $\|V\vec{v}\| = \|\vec{v}\|$ for all $\vec{v} \in X$. In other words, an isometry is a genralization of norm preserving operators.

Now, properties of linear isometries:

- $\langle v|v\rangle = \langle Vv|Vv\rangle = \langle v|V^\dagger V|v\rangle$
- Thus, $V^\dagger V = I_X$.

Linear isometries are not always unitary operators, though, as those require additionally that $X = Y$ and $VV^\dagger = I_X$.

By the Mazur–Ulam theorem, any isometry of normed vector spaces over $\mathbb{R}$ is affine. Affine transformations are those that preserve lines and parallelism (but not necessarily distances and angles).

8.9.6 Steinspring Representation

Isometries allow us to restate the requirements of a noisy open quantum channel. Now, it can be formed by combinations of isometries and partial trace. Moreover, the purpose of the partial trace is to trash the environment shit out of my lab.

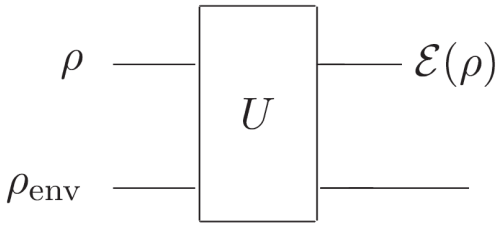
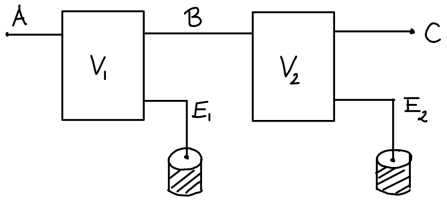


Figure 7: Channel

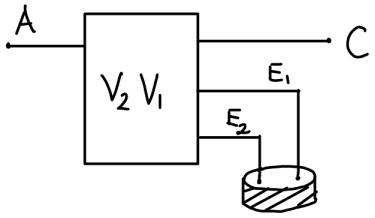
Thus, we have the following formalism.

$$V : A \to B \otimes E$$
$$\mathcal{E} : \mathcal{L}(A) \to \mathcal{L}(B)$$
$$\mathcal{E}(\rho) = \mathrm{tr}_E(V\rho\, V^\dagger)$$

Partial tracing can be delayed (deferred tracing), an idea similar to deferred measurements. Also, from a philosophical standpoint (Church of the Larger Hilbert Space) you can defer it forever.



Now, the above circuit is equivalent to the below circuit. As a result, inductively, combinations of isometries and partial trace is equivalent to the $\mathcal{E}$ operator.



8.9.7 Kraus Operator Picture

We can fix a orthonormal bases $\{|e\rangle\}$ for E, and have $\{V_e\}$ such that

$$V = \sum_e V_e \otimes |e\rangle$$

where $V_e \in \mathcal{L}(A), V_e : A \to B$ and furthermore

$$I = V^\dagger V = (\sum_{e_1} V_{e_1}^\dagger \otimes \langle e_1|)(\sum_{e_2} V_{e_2} \otimes |e_2\rangle)$$

such that $V_e \neq V_e^\dagger$in general and thus, we have the Kraus operator condition (stated below).

$$I = \sum_e V_e^\dagger V_e$$

Thus, we can have a equivalent and simplified formalism of the Steinspring Operator $(\mathcal{E})$, where $\{V_e\}$ are Kraus operators.

$$\mathcal{E}(\rho) = \mathrm{tr}_E(V\rho\, V^\dagger) = \mathrm{tr}_E(\sum_{e_1,e_2} V_{e_1}\rho\, V_{e_2}^\dagger \otimes |e_1\rangle\langle e_2|)$$
$$\implies \mathcal{E}(\rho) = \sum_e V_e\rho\, V_e^\dagger$$

Conversely, given $\{V_e\}$ satisfying Kraus operator conditions, $\mathcal{E}(\rho) = \sum_e V_e\rho\, V_e^\dagger$ is a quantum operator with $V = \sum_e V_e \otimes |e\rangle$ as the isometry involved in the transformation.

8.9.8 Examples of Kraus Operators for Quantum Channels

1. $\mathcal{E}(\rho) = U\rho\, U^\dagger \to$ single Kraus operator $\{U\}$
2. $\mathcal{E}(\rho) = \sum_e p_e U_e\ \rho\ U_e^\dagger \to$ Kraus operators $\{\sqrt{p_e}U_e\}$
3. $\mathrm{tr}(\rho_{AB}) = \sum_b (I_A \otimes \langle b|)\ \rho\ (I_A \otimes |b\rangle)$ where $V_b = (I_A \otimes \langle b|_B)$ thus we also have the following holding true, $\sum_b V_b^\dagger V_b = \sum_b I_A \otimes |b\rangle\langle b| = I_A \otimes I_B = I$.
4. Ultimate Refrigerator: $\mathcal{E}(\rho) = |0\rangle\langle 0|$, regardless of the initial state ρ.

$$V_0 = |0\rangle\langle 0|, V_1 = |0\rangle\langle 1|$$
$$V_0\rho\, V_0^\dagger + V_1\rho\, V_1^\dagger = |0\rangle\langle 0|$$

5. Depolarizing Channel: $\mathcal{E}(\rho) = (1-p)\rho + \rho(I/2) \rightarrow$ simple kind of white noise effect
6. Amplitude Damping Channel: $\mathcal{E}_\gamma(\rho) = K_0\rho\ K_0^\dagger + K_1\rho\ K_1^\dagger$ where we have the following

$$K_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix} \text{ and, } K_1 = \begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix}$$

8.9.9 Axiomatic Definitions

By God, properties of $\mathcal{E}$ should be as follows:

- Linear (assume this or die, basically) and Hermitian preserving
- Trace preserving (TP) such that $\text{tr}(\mathcal{E}(\rho)) = \text{tr}(\rho) = 1$
- Completely positive (CP) since $\rho \geqslant 0 \implies (\mathcal{E} \otimes I)(\rho) \geqslant 0$

But why completely positive? Well, because we are dealing with isometries of form $A \rightarrow B \otimes E$.

Here, completely positive means that if $\mathcal{E}$ is applied on a subsystem, then the complete system and the subsystems must remain positive.

8.9.10 Equivalence of TPCP maps and Kraus operators

- **Equivalence Theorem**: If $\mathcal{E}(\rho) = \sum_e V_e\rho\ V_e^\dagger$ then $\mathcal{E}$ is TPCP and converse holds true as well. The map $\rho \rightarrow \mathcal{E}(\rho)$is injective.
- **Equivalence of Operators**: Given 2 operators $\mathcal{E}$ and $\mathcal{F}$ with operator elements $\{\mathcal{E}_i\}$ and $\{\mathcal{F}_i\}$ respectively, if $\mathcal{E} = \mathcal{F}$ then we have the following.

$$\mathcal{E}_i = \sum_j u_{ij}\mathcal{F}_j$$

- **Choi's theorem**: Let $\mathcal{E} : \mathbb{C}^{n\times n} \rightarrow \mathbb{C}^{m\times m}$ be a linear map. Then, the following are equivalent:
  1. $\mathcal{E}$ is n-positive (i.e. $\mathcal{E}(A) \in \mathbb{C}^{m\times m}$ is positive whenever $A \in \mathbb{C}^{n\times n}$ is positive).
  2. The matrix $\Gamma_\mathcal{E}$, sometimes called the Choi matrix of $\mathcal{E}$, is positive. Here, ϕ is a maximally entangled state in the suited dimension.

$$\Gamma_\mathcal{E} = (\text{id}_n \otimes \mathcal{E})\left(|\phi\rangle\langle\phi|\right)$$

  3. $\mathcal{E}$ is completely positive.

8.9.11 Measurements as Quantum Operations

Measurement can be thought of as a quantum operation where the input is any quantum state and the output is classical, $\mathcal{E}(\rho) = \sum_x p_x|x\rangle\langle x|$ (diagonal density matrix).

The probabilities, p_x, should depend on the state. Further, p_x should be a linear function of the density matrix $p_x = \text{tr}(M_x\rho)$. From this we can work out the properties that M_x should obey:

- Normalization: $\text{tr}(\mathcal{E}(\rho)) = \text{tr}(p_x) = \text{tr}(M_x\rho) = 1 \implies \sum M_x = I$

- Positive Semi-definite: $(\mathcal{E} \otimes I)(\rho) \geqslant 0 \implies p_x \geqslant 0$ thus $\text{tr}(M_x \rho) \geqslant 0, \forall\ M_x$ over ρ thus, we have $M_x \geqslant 0 \implies M_x's$ are positive semi-definite $\implies M_x = E_x E_x^\dagger$

Thus, we have the following (these conditions still leave room for noisy measurements, etc).

$$\sum_x M_x = I$$
$$M_x \geqslant 0$$

We can also talk about non-demolition measurements, which do not discard the quantum systems after measurements. Consider the following quantum channel.

$$\mathcal{E}(\rho) = \sum_e V_e \rho V_e^\dagger \otimes |e\rangle\langle e| = \sum_e \frac{V_e \rho V_e^\dagger}{\text{tr}(V_e \rho V_e^\dagger)} \otimes \text{tr}(V_e \rho V_e^\dagger) |e\rangle\langle e|$$

This channel has the following interpretation: with probability as

$$p_e = \text{tr}(V_e \rho V_e^\dagger) = \text{tr}(V_e V_e^\dagger \rho) = \text{tr}(M_e \rho)$$

the state of the system after the application of this channel is $\rho_e = (V_e \rho V_e^\dagger)/\text{tr}(V_e \rho V_e^\dagger)$.

This is similar to having a measurement that outputs the post measured state ρ_e with probability p_e.

8.9.12 Quantum Norms and Distance Metrics

Motivation: Now that we have defined channels and states of information, how do you differentiate between two items of information? What does it mean to say that information is preserved by some process?

Well for these questions it is necessary to develop distance measures. There is a certain arbitrariness in the way distance measures are defined, both classically and quantum mechanically, and the community of people studying quantum computation and quantum information has found it convenient to use a variety of distance measures over the years. Two of those measures, the trace distance and the fidelity, have particularly wide currency today.

8.9.13 Norms

Norm is a distance metric such that

1. $|cv| = |c|.|v|$
2. $|v + w| \leqslant |v| + |w|$
3. $|v| = 0$ iff $v = 0$

Examples of norms are as follows:

- Manhattan (L_1)
- Euclidean (L_2)
- L_p norm $\rightarrow |v|_{L_p} = (\sum v_i^p)^{1/p}$

8.9.14 Schatten p-norms

1. L_1 norm corresponds to probability distributions
2. L_2 norm corresponds to pure states
3. $L_\infty = \max_i |v_i|$

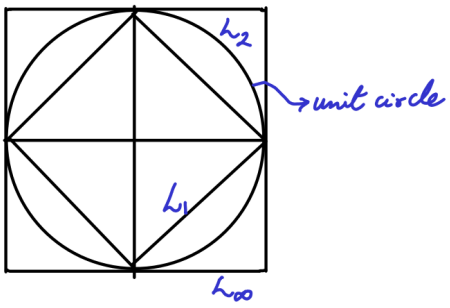
Similarly, schatten p-norm : $|M|_{S_p} = |\sigma(M)|_{S_p}$ where $\sigma(M)$ is a vector of singular values of a matrix, say M.

$$S_1 = \sum \text{singular values}$$
$$S_2 = \sqrt{\sum \text{singular values}^2}$$
$$S_\infty = \max_i |\text{singular values}|$$

In case of density matrices (in general to positive semi-definite Hermitian operators), we have singular values = eigen values thus, $\sigma(M) = \lambda(M)$.

$$|X|_{S_p} = |\sigma(X)|_{L_p} = (\sum_i \sigma_i^p)^{1/p}$$

If $X \geqslant 0$, then $\sigma(X) = \lambda(X)$ then, $|X|_{S_1} = \text{tr}X \implies |\rho|_{S_1} = \text{tr}(\rho)$. Thus, ρ is a density matrix $\iff \rho \geqslant 0$ and $|\rho|_{S_1} = 1$.



8.9.15 Measurements

Let us consider a simple system with $\{M, I - M\}$ as measurement operators. Then, we have

$$M \geqslant 0, I - M \geqslant 0 \iff M \leqslant I$$
$$\iff 0 \leqslant M \leqslant I$$
$$\iff |M|_{S_\infty} \leqslant 1$$

But why is it that two objects with different norms can co-exist in the same operator framework? The answer lies in the notion of duality.

8.9.16 Duality

Given a norm $|.|$ there exists a dual norm such that, $|x|_* = \max_{|y| \leqslant 1} |\langle x, y \rangle|$.

1. L_2 is dual to L_2
2. L_1 and L_∞ are dual to each other
3. S_2 is dual to S_2
4. S_1 and S_∞ are dual to each other

8.9.17 Introduction to Trace Norm

$$T(\rho,\sigma) = \frac{1}{2}|\rho-\sigma|_{S_1} = \frac{1}{2}\ \mathrm{tr}\,|\rho-\sigma|$$

Then, $\forall M$ where M denotes measurement,

$$|\mathrm{tr}(M\rho)-\mathrm{tr}(M\sigma)| = |\mathrm{tr}(M(\rho-\sigma))| \leqslant |M|_{S_\infty}|\rho-\sigma|_{S_1} \leqslant |\rho-\sigma|_{S_1} = 2T(\rho,\sigma)$$

In fact, we can tighten this inequality even further to obtain

$$|\mathrm{tr}(M\rho)-\mathrm{tr}(M\sigma)| \leqslant T(\rho,\sigma) \implies T(\rho,\sigma) = \max_M |\mathrm{tr}\ M(\rho-\sigma)|$$
$$\implies T(\mathcal{E}(\rho)-\mathcal{E}(\sigma)) \leqslant \max_M |\mathrm{tr}\ M(\rho-\sigma| = T(\rho,\sigma)$$

and finally we get the required inequality

$$T(\mathcal{E}(\rho),\mathcal{E}(\sigma)) \leqslant T(\rho,\sigma)$$

Thus, I can never increase the distance of two states, no matter what. So, basically when you wanna protect against noise all we are doing is slowing down the rate of noise (indistinguishable nonsense).

8.9.18 Solving the Lindblad equation

Any ideal open quantum system will undergo Markovian dynamics provided that its evolution satisfies a Master equation.

$$\frac{d\rho}{dt} = \mathcal{L}(\rho) = -i[H,\rho] + \sum_k \gamma_k(V_k\rho V_k^\dagger - \frac{1}{2}\{V_k^\dagger V_k,\rho\})$$

Let $\rho = \frac{1}{2}(I + \vec{r}\cdot\vec{\sigma})$ and differentiate $\rho(t)$ to obtain $\dot{\rho}(t)$ in terms of $\frac{d\vec{r}}{dt}$.

$$\frac{d\rho}{dt} = \frac{1}{2}(\dot{r}_x\sigma_x + \dot{r}_y\sigma_y + \dot{r}_z\sigma_z)$$

We equate $\dot{\rho} = \mathcal{L}(\rho)$ using the above representation for $\dot{\rho}$ and ρ and solve the equation by solving the individual differential equations obtained. After we solve for ρ, we obtain Λ_t.

$$\Lambda_t : \rho_0 \to \rho_t, \quad \rho(t) = \Lambda_t(\rho(0))$$

Now, with Choi's theorem we check if Λ_t is CP by checking whether Γ_Λ is positive semi-definite.

$$\Gamma_\Lambda = C|\psi\rangle\langle\psi|$$
$$C = (I\otimes\Lambda_t)$$

Now, we find the eigenvalues and eigenvectors of C and represent it as $C = \sum_i \lambda_i|v_i\rangle$. Then, we can obtain Kraus operators $\{K_i\}_{\forall i}$ such as follows (check below).

$$v_i = \begin{bmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix} \implies K_i = \sqrt{\lambda_i}\begin{bmatrix} -1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

8.9.19 Markovianity and Information flow

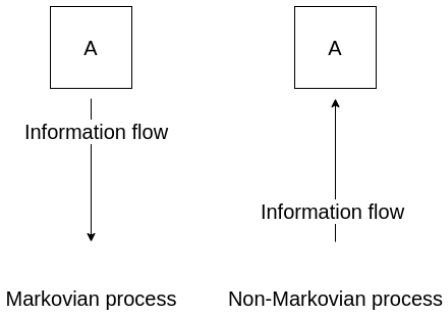


Figure 8: Information flow

Consider 2 states of a system, A and B present in the same environment. Since we are considering open quantum systems, there is interactions between the system and the environment. When treating the systems as probability distributions, the states are said to be collapsing to a common point if it becomes harder to distinguish the states as time passes. This process is classified as flow of information from the system to the environment.

In contrast, If the 2 states grow further apart from each other, there is a flow of information from the environment to the systems. Quantifying the ability to distinguish the states is done by defining a distance metric between the states. Lesser distance implies that it is harder for the states to be distinguished from each other.

A process is said to be Markovian if the evolution from state to another only depends on the present state. By this definition we can show that non-Markovian processes lead to information flow into the system as shown 8.

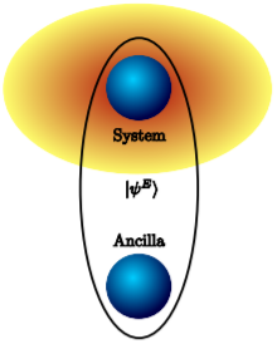
Consider a process A with the initial state $\rho(0)$, since the process is non-Markovian, there exists a pair of states$\rho(t), \rho(t+\tau)$ such that $\rho(t+\tau) = f(\rho(t))$. In other words, the state at time $t+\tau$ depends on the state t. This means that the process must gain information from outside the system, leading to information inflow into the system.

9 ENTANGLEMENT AND NON-MARKOVIANITY

9.1 Introduction

During our study of Quantum Systems, mostly we have encountered Markovian evolutions (which are supposed to be the ideal case). This arises from weak coupling with an environment that acts as a memoryless reservoir.

This works very well from quantum optics, but when dealing with interacting many-body systems. Why does it not work for multi-body systems? Because, subsystem's coupling strength may be comparable to the coupling to the bath in case of interacting many-body systems.



In this report, we discuss Markovianity and non-Markovianity for quantum systems and how we can quantify them.

9.2 Classical and Quantum Markov Processes

Any ideal open quantum system will undergo Markovian dynamics provided that its evolution satisfies a Master equation.

$$\frac{d\rho}{dt} = \mathcal{L}(\rho) = -i[H,\rho] + \sum_k \gamma_k (V_k \rho V_k^\dagger - \frac{1}{2}\{V_k^\dagger V_k, \rho\})$$

This Markovian quantum system given by TP maps $(\mathcal{E})$ that define one-parameter semigroup of CP maps such that $\mathcal{E}_{r_1+r_0} = \mathcal{E}_{r_1}\mathcal{E}_{r_0}$.

- $\mathcal{E}_{t_2,t_0} = \mathcal{E}_{t_2,t_1}\mathcal{E}_{t_1,t_0}$ where $\mathcal{E}_{t_j,t_i} = \mathcal{T}e^{\int_{t_i}^{t_j} d\mathcal{T}} \longrightarrow$
- This is similar to classical Chapman-Kolmogorov equation (given below in functional and matrix form).

$$p_{i_1,\ldots,i_{n-1}}(f_1,\ldots,f_{n-1}) = \int_{-\infty}^{\infty} p_{i_1,\ldots,i_n}(f_1,\ldots,f_n)\, df_n$$

$$P(t+s) = P(t)P(s)$$

At the level of one-point probabilities, divisible and Markovian processes are equivalent. It is required to known the complete hierarchy of time-conditional probabilities to make any distinction.

$$\mathbb{P}(x_1,t_1) = \sum_{x_0\in\mathcal{X}} T(x_1,t_1|x_0,t_0)\mathbb{P}(x_0,t_0)$$

9.2.1 *Relations between classical and quantum Markovian systems*

Consider a Markov process $\{X(t), t \in I\}$. Given any two time instants t_1 and t_2 we have $T(x_2,t_2|x_1,t_1) = P(x_2,t_2|x_1,t_1)$.

Let us consider a system in a quantum state given by some (non-degenerate) density matrix ρ, the spectral decomposition yields

$$\rho = \sum_x p(x)|\psi(x)\rangle\langle\psi(x)|$$

Here the eigenvalues $p(x)$ form a classical probability distribution, which may be interpreted as the probabilities for the system to be in the corresponding eigenstate

$|\psi(x)\rangle$, $\mathbb{P}(|\psi(x)\rangle) = p(x)$.

Consider now some time evolution of the quantum system such that the spectral decomposition of the initial state is preserved, $\rho(t_0) = \sum_x p(x, t_0)|\psi(x)\rangle\langle\psi(x)|$,

$$\rho(t) = \sum_x p(x, t)|\psi(x)\rangle\langle\psi(x)| \in \mathcal{S}$$

where $\mathcal{S}$ denotes the set of quantum states with the same eigenvectors as $\rho(t_0)$.

Since this process can be seen as a classical stochastic process on the variable x, which labels the eigenstates $|\psi(x)\rangle$, we consider it to be divisible if the evolution of $p(x, t)$ satisfies the classical definition of divisibility. In such a case, there are transition matrices $T(x_1, t_1|x_0, t_0)$, such that

$$p(x_1, y_1) = \sum_{x_0 \in \mathcal{X}} T(x_1, t_1|x_0, t_0)p(x_0, t_0)$$

can be written in terms of density matrices as $\rho(t_1) = \mathcal{E}_{(t_1,t_0)}[\rho(t_0)]$.

Here, $\mathcal{E}_{(t_1,t_0)}$ is a dynamical map that preserves the spectral decomposition of $\rho(t_0)$ and satisfies the following equation.

$$\begin{aligned}\mathcal{E}(t_1, t_0)\rho(t_0) &= \sum_{x_0 \in X} p(x_0, t_0)\mathcal{E}(t_1, t_0)[|\psi(x)\rangle\langle\psi(x)|] \\ &= \sum_{x_0 \in X} T(x_1, t_1|x_0, t_0)p(x_0, t_0)[|\psi(x)\rangle\langle\psi(x)|]\end{aligned}$$

Furthermore, $\mathcal{E}(t_2, t_1)$ preserves positivity and trace of any state in $\mathcal{S}$ and obeys the composition law.

$$\mathcal{E}(t_3, t_1) = \mathcal{E}(t_3, t_2)\mathcal{E}(t_2, t_1)$$

	Classical	Quantum
Normalization	$\sum_{x_2 \in \mathcal{X}} T(x_2, t_2\|x_1, t_1) = 1$	$\mathcal{E}_{(t_2,t_1)}$ trace-preserving
Positivity	$T(x_2, t_2\|x_1, t_1) \geq 0$	$\mathcal{E}_{(t_2,t_1)}$ completely positive
Composition Law	$T(x_3, t_3\|x_1, t_1) = \sum_{x_2 \in \mathcal{X}} T(x_3, t_3\|x_2, t_2)T(x_2, t_2\|x_1, t_1)$	$\mathcal{E}_{(t_3,t_1)} = \mathcal{E}_{(t_3,t_2)}\mathcal{E}_{(t_2,t_1)}$

Figure 9: Comparision between Classical and Quantum Markovianity

9.3 Understanding Markovianity

We can understand Markovianity under serveral paradigms as follows.

1. Information flow: In case of a Markovian evolution, information goes out of the system, to the environment and does not come back. However, in case of non-Markovian evolution, there is influx of information from environment to system, at some point of time.

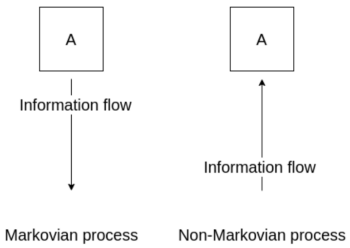


Figure 10: Information Flow

2. Trace distance: In general, for open quantum systems (under Markovian dynamics) you can never increase the distance of two states, no matter what. So, basically when you wanna protect against noise all we are doing is slowing down the rate of noise (indistinguishable nonsense).

$$T(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leqslant T(\rho, \sigma)$$

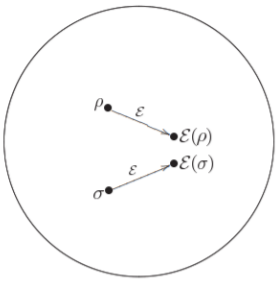


Figure 11: Trace distance

Definition of trace distance given below.

$$T(\rho, \sigma) = \frac{1}{2}|\rho - \sigma|_{S_1} = \frac{1}{2}\ \mathrm{tr}\ |\rho - \sigma|$$

However, if the distance between them increases then they become more distinguishable. This implies that information has entered the system from the environment, further implying non-Markovianity.

3. Coupling: Weak coupling of system to a memoryless reservoir results in Markovianity, whereas strong coupling mostly leads to non-Markovianity.
4. Entanglement: For Markovian evolutions, the decay of the entanglement with an ancillary system (in environment) will be monotonically decreasing.

9.4 Measures of Quantum non-Markovianity

9.4.1 Optimization problem

The measure of non-Markovianity can be formulated as an optimization problem. Let $\mathcal{E}$ be the physical dynamical map that is unknown to us, then we are interested in finding the map from the family of Markovian Maps $\mathcal{E}^{M}_{t_0+\epsilon,t_0}$ such that its S_1 norm with $\mathcal{E}$ is as low as possible. At the same time, we also want to find the maximum non-Markovianity over the entire duration, hence we treat ϵ as a variable and choose the one that gives us the largest non-Markovianity. This leads to a min-max problem. The major disadvantage of this formulation is that it requires the knowledge of $\mathcal{E}$ beforehand, which we definitely don't know about.

$$\max_{\epsilon>0} \min_{\mathcal{E}^M} \|\mathcal{E}_{(t_0+\epsilon,t_0)} - \mathcal{E}^{M}_{(t_0+\epsilon,t_0)}\|$$

9.4.2 Entanglement measure

We know that decay of entanglement with an ancillary system will be monotonically decreasing for Markovian evolutions.

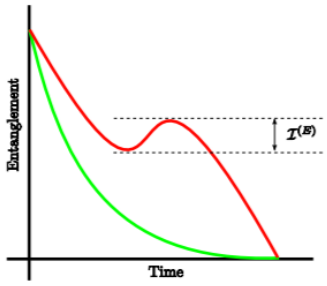


Figure 12: Entanglement vs Time

Thus, we can construct some measure for non-Markovianity of an unknown quantum evolution by computing the amount of entanglement between system and ancilla at different times within a selected interval $[t_0, t_{max}]$ and check for strict monotonic decrease of the quantum correlations.

So how do we compute entanglement? We use measures such as:

- Logarithmic negativity: $\mathbf{E}(\rho_{SE}) = \log_2 \|\rho_{SE}^{T_S}\|_1$
- Concurrence: $\mathbf{E}(\rho_{SE}) = \max(0, \sqrt{\lambda}_1\sqrt{\lambda}_2\sqrt{\lambda}_3\sqrt{\lambda}_4)$

where $\mathbf{E}(.)$ denotes measure of entanglement and S stands for system whereas E denotes environment and ρ_{SE} denotes a quantum state entangled with an ancillary system from the environment.

Now, we define $\mathcal{I}^{(E)}$ as:

$$\mathcal{I}^{(E)} = \int_{t_0}^{t_{max}} |\dot{\mathbf{E}}[\rho_{SA}(t)]|dt - \Delta\mathbf{E}$$

$$= \int_{t_0}^{t_{max}} |\dot{\mathbf{E}}[\rho_{SA}(t)]|dt - \int_{t_0}^{t_{max}} d\mathbf{E}[\rho_{SA}(t)]$$

where,

$$\Delta\mathbf{E} = \mathbf{E}[\rho_{SA}(t_0)] - \mathbf{E}[\rho_{SA}(t_{max})]$$

and,

$$\rho_{SA}(t_0) = |\phi\rangle\langle\phi|, \quad \phi = \frac{1}{d}\sum_{i=0}^{d-1} |i\rangle \otimes |i\rangle$$

Hence we obtain the sufficient condition to show that an evolution is non-Markovian as:

1. If $I(E) = 0$ then, the integral term is equal to $\Delta\mathbf{E}$. Which means there was no increase in entanglement at any point.
2. If $\mathcal{I}(E) > 0$, we conclude that the evolution is non-Markovian. However, there can be however non-Markovian quantum evolutions that remain undetected by the proposed measure.

9.5 Stronger condition for Markovianity

In the previous case (when using entanglement measure), it is worth noticing that that we never knew what the exact dynamical map was. We just kept applying it locally on our system and measuring the entanglement. Moreover, it wasn't a necessary condition for distinguishing Markovianity and non-Markovianity.

However, if we can reconstruct the dynamical map, then we can achieve a necessary and sufficient condition for non-Markovianity. This reconstruction can be done using tomography.

Now, this allows us to further construct a **necessary as well as sufficient condition** for Markovianity. Givn that we are dealing with dynamical map $\mathcal{E}_t$, we can say that if the map is CPTP, $\forall t$ then the evolution is Markovian and the following property shall hold.

$$\mathcal{E}_{t+\epsilon,0} = \mathcal{E}_{t+\epsilon,t}\mathcal{E}_{t,0}$$

We use the Choi-Jamiolkowski isomorphism to check whether $\mathcal{E}_{t+\epsilon,t}$ is CP, in other wordss we have to check if the following holds true.

$$(\mathbb{I} \otimes \mathcal{E}_{t+\epsilon,t})|\phi\rangle\langle\phi| \geqslant 0$$

Given the trace preserving property, we have the following as a measure of non-markovianity.

$$f_{NCP}(t+\epsilon,t) = \|(\mathbb{I} \otimes \mathcal{E}_{t+\epsilon,t})(|\phi\rangle\langle\phi|)\|_1$$

Therefore, $\mathcal{E}(t+\epsilon,t)$ is CP if and only if $f_{NCP}(t+\epsilon,t) = 1$, otherwise $f_{NCP}(t+\epsilon,t) > 1$ which we shall use to arrive at a measure of non-markovianity.

$$g(t) = \lim_{\epsilon\to 0^+} \frac{f_{NCP}(t+\epsilon,t)-1}{\epsilon} = \begin{cases} g(t) = 0, & \text{iff } \mathcal{E}_{t+\epsilon,t} \text{ is CP (markovian)} \\ g(t) > 0, & \text{otherwise (non-markovian)} \end{cases}$$

Further, we define $\mathcal{I}$ as follows.

$$\mathcal{I} = \int_0^\infty g(t)dt$$

Now, $\mathcal{I}$ can be taken as a measure of non-Markovianity, and as long as $g(t)$ decreases fast enough and still is finite.

We can further obtained a normalized version of the above measure given by $\mathcal{D}_{NM}$.

$$\mathcal{D}_{NM} = \frac{\mathcal{I}}{\mathcal{I}+1} = \begin{cases} \mathcal{D}_{NM} = 0 \text{ when } \mathcal{I} = 0 \text{ (markovian)} \\ \mathcal{D}_{NM} \to 1 \text{ when } \mathcal{I} \to \infty \text{ (non-markovian)} \end{cases}$$

9.6 Examples

9.6.1 Single Damped Harmonic Oscillator

Single Damped Harmonic Oscillator: In this case, in order to visulaize the sensitivity of the proposed measure $\mathcal{I}^{(E)}$, two different spectral densities of the bath have

been considered, along with several initial temperatures.

Now, we have

$$J(\omega) = \alpha\omega^k e^{-\omega/\omega_c}$$

where $k = 1$ for ohmic spectral density and $k = 3$ for super-ohmic spectral density.

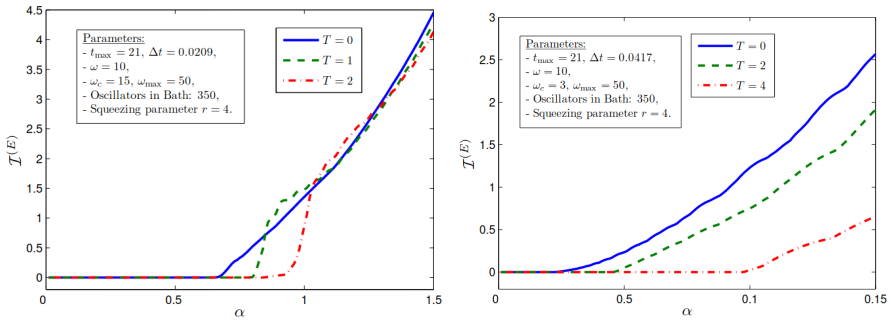


Figure 13: Ohmic and super-Ohmic spectral density

Non-markovian character increases as α increases. As $\alpha \to 0$, the evolution approaches markovian character.

9.6.2 *Pure Dephasing*

The dynamics of pure dephasing is given by the following Lindblad equation.

$$\frac{d\rho}{dt} = \gamma(t)(zz)$$

From the above we obtain,

$$g(t) = \begin{cases} 0, & \gamma(t) \geqslant 0 \\ -2\gamma(t), & \gamma(t) < 0 \end{cases}$$

$$\mathcal{I} = -2\int_{\gamma(t)<0} \gamma(t)dt$$

Thus, if $\gamma(t) \geqslant 0$ we have Markovian evolution, else if its negative (< 0) then we have non-Markovian evolution.

9.7 Conclusion

We explored some basic concepts related to what non-Markovian quantum evolution is and how it can be measured. We derived the necessary and sufficient conditions to indicate whether a dynamical map is Markovian or non-Markovian or not. We attempted to demonstrate our ideas by also using a classical analogue in thermodynamics as well.

REFERENCES

1. Song et. al. Capacity of Quantum Private Information Retrieval With Multiple Servers
2. Susana F. Huelga Angel Rivas and Martin B. Plenio. Entanglement and Non-Markovianity of Quantum Evolutions

3. Michael A. Nielsen and Isaac L. Chuang. Quantum Computation and Quantum Information

4. Google Quantum AI Cirq

5. B. Chor et. al. Private Information Retrieval

6. Alliax et. al. Quantum private information retrieval from mds-coded and colluding servers