THE UNIVERSITY OF HONG KONG

FYP18026 FINAL YEAR PROJECT

FINAL REPORT

Quantum Machine Learning with Superposition of Causal Order

Author
Gao Ning 3035234521

Supervisor
Prof. G. CHIRIBELLA

13 April. 2019

Abstract

Research has shown that quantum computation is capable of improving significantly the performance of classical machine learning algorithms. On the other hand, classical machine learning algorithms have been proven useful for understanding quantum resources. In this report, starting from defining a new measure of incompatibility for projective measurements, we proceed to tackle a specific quantum learning task: clustering (projective) measurement devices using a new computational primitive known as quantum switch.

Acknowledgement

I would like to express my sincere gratitude to my supervisor Prof. Giulio Chiribella, without whom I would never have the chance to explore the wonderful world of quantum information and computation. Prof. Giulio Chiribella has always been the strongest help and the greatest inspiration to me throughout this project. His helpfulness, his erudition, and his enthusiasms in research will always be what I shall learn from. I would also like to thank Prof. Giulio Chiribella's group members: Dr.Some Sankar Bhattacharya, Mr.Xiaobin Zhao, Mr.Mo Yin, Mr.Ge Bai and Mr.Fei Meng for helping me to understand many important concepts and ideas which are crucial in this project. I would also like to say thank you to Mr.Mao Wenxu, who is also a final year student working on his final year project under Prof.Giulio Chiribella's supervision, many times you gave me the power to move on. Finally, I want to say thank you to all that have been with me and supported me, you guys are awesome.

Contents

List of Figures

List of Tables

1 Introduction

Since first introduced by Richard P. Feynman in 1981 in his epochal-making speech [1], tremendous progress on the study of quantum computation has been made in both theory and physical realization. Along the line of research, many algorithms were proposed and gained public attention, although they were only of pure theoretical interests in 90s when created. Examples include the well-known Shor's algorithm, a quantum algorithm which factorizes a large integer into its prime factors [2], and the Grovers search algorithm [2] which was later extended to a general scheme called amplitude amplification [3]. The Shor's algorithm surprised the public for it not only demonstrated the potential of quantum computer to provide exponential speedup against their classical counterpart, but it overthrew some most widely used public-key cryptography schemes such as RSA scheme [2]. The latter one, though offers only quadratic speedup, has been proven extremely useful, since searching problems are often important building blocks of other much more complicated algorithms [2].

In addition to its astonishing application in cryptography, quantum computation is also considered advantageous in solving optimization problems. In particular, much efforts have been devoted into a quantum computation model called adiabatic quantum computation [4]. This model was first introduced to attack a long-standing problem called 3SAT problem which was the first known to be NP-complete [5]. Later on. It was then found useful in other tasks such as clustering algorithms, where researchers artfully transformed the clustering objectives into cost functions to be optimized [6].

The introduction of Adiabatic quantum computation into unsupervised machine learning is in fact only a small portion of the wide application of quantum computation techniques in tasks arising from the study of machine learning [7] [8]. Many classical machine learning algorithms witnessed the construction of their quantum counterparts. In most cases, the quantum versions are asymptotically much faster, although the quantum algorithms usually come with restraints that classical versions do not suffer. A typical example is the HHL algorithm which can solve a linear system of equations exponentially faster than their classical counterparts, under the constraint that the matrix can be simulated efficiently [9]. Other techniques are also popular in quantizing classical machine learning algorithms. For instance, combining the power of density operator exponentiation and phase estimation, one can perform principle component analysis efficiently [10]. When a machine learning task can be formulated as a search problem, Grovers search algorithm (and its variants) can usually offer speedup (mostly quadratically). Concrete examples are quantum clustering algorithms in which tasks such as searching for minimal are repeated [11].

On the other hand, one can also consider the possibility of utilizing classical machine learning techniques to accomplish quantum learning tasks. Existing machine learning algorithms provide general frameworks to understand, to extract and to define certain information of data [12]. This suggests we may consider making use of machine learning techniques to understand quantum resources, such as unitary gates and projective measurements. However, a typical clustering algorithm such as K means algorithm usually assumes the accessibility of the difference between individual data points. This leads to the question: how should we understand and quantify the difference between quantum resources such as projective measurements? We approach this problem from the perspective that marks the drastic distinction between classical and quantum world: incompatibility between measurements. Before diving into the details, we would like to emphasize that the usefulness of understanding and estimating the incompatibility between projective measurements goes way beyond solving a quantum learning task and we will discuss some of these in section 2.3.

A worth noting point is that there already exists several well accepted definitions of incompatibility for quantum measurements/devices in general operation theory, such as in [13] [14] [15]. On one hand, we would like to argue that our own definition fits well into these existing notions, on the other hand, we would also like to point out that our definition enjoys a very desirable feature as compared to its counterparts: it is easy to compute both in theory and in practice. That is to say, with constant time accesses to 2 measurement devices, we can estimate the incompatibility between such 2 devices (in our definition) to high precision. To achieve so, we assume the existence of a new computational primitive called quantum switch [16]. As we shall explain in the following sections, quantum switch is strongly related to the concept of superposition of causal order which appears mysteriously in the title of this report.

As an overview, this report is organized in the following fashion: section 2 is devoted to review the background knowledge used in this report. This includes basic quantum mechanics used in quantum information theory, the definition and basic properties of a quantum switch, and concepts in theory of incompatibility that are relevant to our project. In section 3 we present our original definition of incompatibility for projective measurements and the measure of incompatibility induced from our definition. We also discuss several important properties of our definition in this section. In section 4 we will discuss how to estimate the degree of incompatibility between projective measurements using quantum switch. From there, we will generalize our definition to a broader class of measurements. In section 5, we deal with an interesting quantum learning task: clustering a set of projective measurements. Then in section 6 we summarize the findings and discuss possible directions of future research, this serves as the conclusion of the whole report.

2 Background

2.1 Qubit, Density Operator, Measurement and Quantum Channel

[1] In classical computer science, the unit of data is bit, taking values 0 or 1. In quantum information, bit is replaced by quantum bit (qubit), with two perfectly distinguishable states $|0\rangle, |1\rangle \in \mathbb{C}^2$. In fact, a qubit can take all unit vectors $\alpha|0\rangle + \beta|1\rangle \in \mathbb{C}^2$ as its value. However, different vectors may correspond to the same state, in particular, $|\phi_1\rangle, |\phi_2\rangle \in \mathbb{C}^2$ correspond to the same state if they differ only by a global phase, i.e $|\phi_1\rangle = e^{i\gamma}|\phi_2\rangle$, for some $\gamma \in \mathbb{C}$. Hence, one may reparametrize a unit vector $|\phi\rangle = e^{i\gamma}[\cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}e^{i\psi}|1\rangle], \gamma \in [0, 2\pi], \theta \in [0, \pi], \psi \in [0, 2\pi]$ and consider only the part within []. Note that this allows one to identify a qubit with a point on the unit sphere $\mathbb{S}^2$ (known as Bloch Sphere), by the polar coordinate $(1, \theta, \psi)$. Note that by such identification, an orthornormal basis would correspond to a pair of antipodal points on BLoch Sphere. In general, we are allowed to work on qudits in a space $\mathbb{C}^d$, with d perfectly distinguishable states $|0\rangle, \cdots, |d-1\rangle$ (the standard basis). These states represented by unit vectors are said to **pure**. In general situation, a state can be **mixed**, in which situation, they are represented by density operator on the state space.

We can also consider a joint system, just like in classical case where we operate on $101100\ldots$. For two systems A, B with their corresponding state spaces $\mathbb{C}^{d_A}, \mathbb{C}^{d_B}$, the composite system AB is associated with the state space $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B} \cong \mathbb{C}^{d_A d_B}$. Hence, the composite of states and actions (to be defined) are all naturally defined under the tensor product. We summarize these in definition 1. A remark is from now on the space $\mathbb{C}^d$ equipped with the standard inner product will be referred to as "the Hilbert" space $\mathcal{H}^d$. In this report, we are not interested in the topological properties (the space is complete, etc) nor infinite dimensional Hilbert spaces.

Definition 1. *Associated to any isolated physical system is a complex vector space with inner product (a Hilbert space $\mathcal{H}^d$ for finite dimensional case or for infinite dimensional case $L^2[\mathbb{R}]$ a Lebesgue space) known as **state space** of the system. The system is completely described by its **density operator**, which is a positive operator ρ with trace one, acting on the state space of the system. If a quantum system is in the state ρ_i with probability p_i, then the density operator for the system is $\sum_i p_i \rho_i$. A pure state is any $\rho = |a\rangle\langle a|$, where $\big\||a\rangle\big\| = 1$, sometimes written as a state vector $|a\rangle$. In case $|a\rangle \in \mathcal{H}^2$, a pure state $|a\rangle = \alpha|0\rangle + \beta|1\rangle$ is called a qubit.*

Now we deal with what actions on a quantum state are allowed. For a reversible process, it is characterized by a unitary operator on $\mathcal{H}^d$. The rationale is that it must send a unit pure quantum state to unit pure quantum state (thus its an isometry), and it has to be reversible, so it must be a unitary operator (if we assume it to be linear). Formally:

Definition 2. *The (reversible) evolution of a closed quantum system $\mathcal{H}^d$ is described by a unitary*

[1] unless specified, all definitions are taken from [2] and Prof. Giulio Chiribella's lecture note

transformation U up to global phases. U acts on the density operator ρ (or the state vector $|a\rangle$) of the system by $U\rho U^\dagger$ (or $U|a\rangle$).

However, we are not restricted to reversible actions only, in fact, we can define axiomatically what are the quantum operations we are interested in and derive suitable representations for them. In most cases, we are working on "quantum channels", which captures all the quantum operations we are studying in this report.

Definition 3. *A **quantum** channel $\mathcal{E}$ from the set of density operators of the input space Q_1 to the set of density operators of the output space Q_2, satisfies the three axioms:*
***A1**: $\operatorname{Tr}\mathcal{E}(\rho)$ is the probability that the process represented by $\mathcal{E}$ occurs, when the initial state is ρ:*

$$0 \leq \operatorname{Tr}[\mathcal{E}(\rho) \leq 1, \tag{1}$$

***A2**: $\mathcal{E}$ is convex-linear on the set of density operators, that is, for probabilities p_i:*

$$\mathcal{E}(\sum_i p_i\rho_i) = \sum_i p_i\mathcal{E}(\rho_i) \tag{2}$$

***A3**: $\mathcal{E}$ is a **completely positive** map. That is $\mathcal{E}(A)$ must be positive for all positive A. Further more, if we introduce an extra system R of arbitrary dimensionality, it must be true that $\mathcal{I}\otimes\mathcal{E}(A)$ is positive for any positive operator A on the combined system RQ_1, where $\mathcal{I}$ is the identity map on R.*
*The map $\mathcal{E}$ satisfies axioms **A1**, **A2** and **A3** if and only if:*

$$\mathcal{E}(\rho) = \sum_i E_i\rho E_i^\dagger \tag{3}$$

for some set of operators E_i which map the input Hilbert space to the output space, and $\sum_i E_i^\dagger E_i \leq I$. $\{E_i\}$ are called Kraus operators.

Kraus representation of a channel is unique up to unitary transformation, i.e. the maps $\mathcal{E}_1(\rho) = \sum_i E_i\rho E_i^\dagger$ and $\mathcal{E}_2(\rho) = \sum_j F_j\rho F_j^\dagger$ represent the same channel if and only if there exists a unitary u_{ij} such that $E_i = \sum_j u_{ij}F_j$. This can be understood as different physical implementation of the same channel, one example is the completely depolarizing channel (over $\mathcal{H}^d$): $\mathscr{D}(\rho) = \operatorname{Tr}\rho\frac{I}{d} = \frac{1}{d^2}\sum_i^{d^2} U_i\rho U_i^\dagger = \frac{1}{d^2}\sum_{j,k}^{d} |j\rangle\langle k|\rho|k\rangle\langle j|$ where $\{U_i\}$ is any set of orthogonal unitaries (e.g. Heisenberg Weyl operators) and $\{\,|j\rangle\,\}$ is any orthonormal basis.

Although a quantum state can take uncountably infinitely many different value, we cannot extract an infinite amount of information from it. We can extract information of a state by performing a measurement. In quantum theory, measurements are described as positive operator valued measure (POVM).

Definition 4. *A POVM (over $\mathcal{H}^d$) is a set of positive operators $\{E_i\}$ such that $\sum_i E_i = I$. The index i refers to the measurement outcome that may occur. If the state of the quantum system is ρ, then the probability of that result i occurs is $p(i) = \operatorname{Tr}E_i\rho$ and the state of the system after measurement is up to the underlying channel (of this measurement).*

However, in this report, we are mainly interested in the case where the positive operator E_i are (orthogonal) projectors, i.e. $E_i^* = E_i^2 = E_i$. In such special case, the measurement is called **projective measurement**. When we are considering projective measurements, in convention, the state of the system after the measurement (with outcome i occurs) is $\frac{E_i\rho E_i}{trE_i\rho}$. Hence by definition 3, this projective measurement can be described by a quantum channel $\mathscr{N}(\rho) = \sum_i E_i\rho E_i$. When E_i are rank one projectors, the measurement is associated with an orthonormal basis $\{\,|\alpha_i\rangle\,\}$ such that $E_i = |\alpha_i\rangle\langle\alpha_i|$. In this case, with outcome being i, the system is "collapsed" into $|\alpha_i\rangle$.

2.2 Quantum Switch

Definition 5. *(quantum SWITCH)[17]:*

Denote the Kraus operators of the channel $\mathscr{N}_1$ as K_i^1 and the channel $\mathscr{N}_2$ as K_i^2. The overall Kraus operators of the resulting channel from the switching is:

$$W_{i,j} = K_j^1 K_i^2 \otimes |0\rangle\langle 0|_C + K_i^2 K_j^1 \otimes |1\rangle\langle 1|_C \tag{4}$$

Then the action of the quantum SWITCH is given by:

$$\mathcal{S}(\mathscr{N}_1, \mathscr{N}_2)(\rho \otimes \rho_c) = \sum_{i,j} W_{ij} \rho \otimes \rho_c W_{ij}^\dagger \tag{5}$$

In words, a switch map $\mathcal{S}$ is a super map that transforms two quantum channels $\mathscr{N}_1, \mathscr{N}_2$ into a channel which is classically controlled that performs either $\mathscr{N}_1\mathscr{N}_2$ or $\mathscr{N}_2\mathscr{N}_1$ conditionally on the outcome of a measurement of system C [16].

An important property is that such definition is invariant of Kraus representations. This means a supermap that controls between the orders of the channels to be applied is in fact uniquely defined as in definition 5. In addition to that, this means the switching of two channels is invariant of different physical realizations of the channels.

Perhaps the most interesting property of a quantum switch is that when the control system is in a superposition state $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$, the two channels are applied in a superposition of orders. This means one cannot say either one of the channels is applied first: quantum switch exhibit an indefinite causal order (sometimes called causally nonseparable)[16] [17].

2.2.1 Properties and Applications of Quantum Switch

Starting from the groundbreaking paper [16] which first presented the notion of the quantum switch theoretically, many efforts have been devoted into understanding the computational and informational advantages that the quantum switch can offer. The major property that signals the deviation of a communication scheme/computation model with indefinite causal structure from the classical models is summarized as a no go theorem in [16]:

Theorem 1. *The quantum switch cannot be computed deterministically by a circuit in which the two unknown oracles are called a single time in a fixed casual order.*

This no go theorem denies the possibility of a perfect implementation of quantum switch in a quantum computation model with definite causal structure. Hence confirms that quantum switch is indeed a new computational primitive. One concrete example of quantum switch providing computational advantage was then discovered in the task of quantum operation discrimination. Quoting the [18]: "two no-signalling channels that are not perfectly distinguishable in any ordinary quantum circuit can become perfectly distinguishable through the quantum superposition of circuits with different causal structure".Other works then identified the advantage indefinite causal structure can offer in Bell-inequality type games [19]. However, the power of quantum switch also applies when only unitary transformations are involved. For example, it can be shown that a generalized n-switch can be used to solve algorithmic problems that involve permutations of orders of multiplication of unitaries, for instance, applying n unitary gates in an order chosen from S_n as in [20], and testing certain "commutativity" properties for a set of unitaries as in [21]. In both cases, "quantum switch" offers significant speedup against "standard" model of quantum computation and classical computers.

In the area of quantum information theory, it was first observed in [17] that self-switching a completely depolarising channel allows the transmission of classical information, which is very astonishing and counter intuitive since two completely depolarising channels can never transmit any classical (and quantum) information however used in definite causal orders. More research followed this line to study the activation of classical channel capacity, [22] for example has started a discussion on the true cause of the activation.

Although it remains an open question of the true power of indefinite causal orders in classical communication tasks. Results in [23], [24] have confirmed that a quantum switch can offer unique advantages in quantum communication tasks against independent superposition of channels. In particular the recent article [24] proved that indefinite causal order enables perfect error correction for zero (quantum) capacity channels, where in contrast superposition of independent channels can never achieve so.

Our results are partially motivated by these findings. But to fully understand our work, we give a brief review of several existing definitions of "incompatibility" for measurements and devices in general.

2.3 Incompatibility

The idea of incompatibility dates back to the age of Heisenberg and Bohr, manifested through the famous Heisenberg's uncertainty principle [25] and the notion of complementarity [26]. These notions were the early attempts to mark the distinct feature of quantum physic: there exists quantum measurements that cannot be implemented simultaneously. For example the position and momentum observables are incompatible.

Similar to entanglement, incompatibility is a valuable resource allowed in quantum world but not classical world. [27] shows that only incompatible measurements enable violations of the famous Bell Inequality. Quantum public key distribution protocols such as BB84 involves incompatible qubit projective measurements. Hence the incompatibility is intriguing to study even by itself.

In recent years, many different ways to compare quantum observables were introduced, in a much more general framework in many cases. We briefly review two of such measure, and list the essential features for further comparison. Since we are only interested in quantum measurements, we will not study the following concepts in their most general forms.

2.3.1 Joint Measurability

Definition 6. [13] *Given n POVMs M^k with positive operators $\{M^k_{x_k}\}_{x_k\in\Omega_k}$ where Ω_k are the sets of measurement outcomes (for each POVM M^k respectively). POVMs M^k are said to be jointly measurable if there exists a "joint observable" M (POVM) that has the set of measurement outcome $\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$ and output a list of outcome $(x_1, x_2, \cdots, x_n)$ at each measurement rounds. In addition, M must satisfies $M^k_{x_k} = \sum_{l\neq k}\sum_{x_l} M_{x_1,x_2,\cdots,x_k,\cdots,x_n}$. As a convention, M^k is called a marginal of M, M is called the joint observable of $M^1, \cdots, M^n$.*

To test if two arbitrary observables (POVMs) are compatible is in general a difficult problem, but the task is much simpler in the case of projective measurements and there is an equivalent condition:

Proposition 1. *If measurements M^k as in definition 6 are projective measurements, then $M^1, \cdots, M^n$ are jointly measurable if and only if $M^k_{x_k}$ and $M^j_{x_j}$ commutes for all k, j, x_k, x_j, i.e. the observables commute.*

A proof is given in [13]. When given n commutative observables M^k (having the same outcome space Ω), we can define the joint observable M to be $M_{x_1,x_2,\cdots,x_n} = J_n(M^1_{x_1}, M^2_{x_2}, \cdots, M^n_{x_n}) = \frac{1}{n!}\sum_{\sigma\in S_n} M^{\sigma(1)}_{x_{\sigma(1)}} M^{\sigma(2)}_{x_{\sigma(2)}} \cdots M^{\sigma(n)}_{x_{\sigma(n)}}$, where S_n is the nth symmetric group, $J_n()$ is the generalized Jordan product, and $x_1, x_2, \cdots, x_n \in \Omega$. Note that this is a joint observable if and only if all $M_{x_1,x_2,\cdots,x_n}$ are positive. One can easily show that $M^{\sigma(1)}_{x_{\sigma(1)}} M^{\sigma(2)}_{x_{\sigma(2)}} \cdots M^{\sigma(n)}_{x_{\sigma(n)}}$ is positive because the individual observabals commute. Hence M is indeed the joint observable. Then using a necessary condition proven in [28], it is also possible to show that commutativity is necessary for joint measurability.

Indeed, such property matches the traditional understanding which equates compatibility with commutativity for projective measurements. Operationally this can be understood as commutative projective measurements do not disturb each other. Such intuition will be made clear after we review another defintion of the measure of incompatibility for projective measurements.

Definition 7. *The degree of compatibility between two measurements M^1, M^2 is defined to be the greatest number $0 \leq \lambda \leq 1$ such that the $\lambda M^1 + (1-\lambda)D^1$ and $\lambda M^2 + (1-\lambda)D^2$ are jointly compatible for some trivial devices D^1, D^2.*

In reality, the degree of compatibility between an arbitrary pair of quantum measurement is difficult to compute. There exists no known analytic formula that can be used to compute the degree. It is also not known, in such definition, what pairs of quantum measurements are maximally incompatible in a finite dimensional Hilbert space $\mathcal{H}^d$. There does exists several useful lower bounds (tight in infinite dimensional Hilbert space $L^2(\mathbb{R})$) as analysed in [29].

2.3.2 Measure of Incompatibility by Probability Distribution

Since joint measurability cannot give us much intuition on how should maximally incompatible projective measurements be like in finite dimensional Hilbert spaces, it makes sense to change a perspective and consider to study the incompatibility by comparing probability distributions. The follwoing definition is taken from [15].

Definition 8. *Given a measure of distance $D(\cdot,\cdot)$ on the space of discrete probability distributions, two projective measurements $\mathscr{N}_1, \mathscr{N}_2$ on d dimensional Hilbert space $\mathcal{H}^d$ with rank one projectors $\{P_i\}, \{Q_j\}$ respectively, and an arbitrary density operator ρ; we define the probability distributions: $Pr_\rho^{\mathscr{N}_1}(i) = \operatorname{Tr} P_i\rho$, $Pr_\rho^{\mathscr{N}_2\to\mathscr{N}_1}(i) = \operatorname{Tr} P_i(\sum_j Q_j\rho Q_j)$ (first do measurement $\mathscr{N}_2$ then do measurement $\mathscr{N}_1$). The **"measure of incompatibility of $\mathscr{N}_2$ with $\mathscr{N}_1$"** (relative to $D(\cdot,\cdot)$) is defined to be $\mathcal{Q}(\mathscr{N}_2 \to \mathscr{N}_1) := sup_\rho D(Pr_\rho^{\mathscr{N}_2\to\mathscr{N}_1}, Pr_\rho^{\mathscr{N}_1})$. The **degree of incompatibility between $\mathscr{N}_1$ and $\mathscr{N}_2$** is then defined as the average: $\mathcal{Q}(\mathscr{N}_1,\mathscr{N}_2) := \frac{\mathcal{Q}(\mathscr{N}_2\to\mathscr{N}_1)+\mathcal{Q}(\mathscr{N}_1\to\mathscr{N}_2)}{4}$*

Assume $D(\cdot,\cdot)$ to be the sup distance, it is shown in [15], $\mathcal{Q}(\mathscr{N}_1,\mathscr{N}_2)$ vanishes if and only if the two measurements commutes. In words this means if the two observable commutes, a measurement on $\mathscr{N}_2$ does not disturb the outcome of measurement $\mathscr{N}_1$ and vice versa. It can also be shown that if the measurements are on mutually unbiased bases then the their degree of incompatibility $\mathcal{Q}(\mathscr{N}_1,\mathscr{N}_2)$ is maximized. Mutually unbiased basis means for $\{P_i = |\alpha_i\rangle\langle\alpha_i|\}$ and $\{Q_j = |\beta_j\rangle\langle\beta_j|\}$, the orthonormal bases $\{|\alpha_i\rangle\}, \{|\beta_j\rangle\}$ satisfies $|\langle\alpha_i|\beta_j\rangle|^2 = \frac{1}{d}, \forall i,j$. In other words, measuring $|\beta_j\rangle$ on the basis $\{|\alpha_i\rangle\}$ generates completely random result (a uniform distribution). In fact, such projective measurements can be used in quantum key distribution protocols [30].

These known results form the basic intuitions in our definition of the measure of incompatibility for projective measurements. In particular, we expect our measure vanishes when the measurements commutes and go to maximal when the measurements are on mutually unbiased basis.

3 A New Measure of Incompatibility for Projective Measurement

In this section, we present the definition directly and prove several easy mathematical properties to illustrate why this is a good definition. Note in this section we mainly focus on rank one projective measurement.

Definition 9. *Consider two (rank one) projective measurements $\mathscr{N}_1$ and $\mathscr{N}_2$ on $\mathcal{H}^d$, with POVM elements $\{P_i = |\alpha_i\rangle\langle\alpha_j|\}$ and $\{Q_j = |\beta_j\rangle\langle\beta_j|\}$ respectively, where $\{|\alpha_i\rangle\}$ and $\{|\beta_j\rangle\}$ are orthonormal. The degree of incompatibility between $\mathscr{N}_1$ and $\mathscr{N}_2$ is defined to be $(d_{avg}(\mathscr{N}_1,\mathscr{N}_2))^2 = 2 - \frac{2}{d}\sum_{i,j}^d \operatorname{Tr} P_iQ_jP_iQ_j$. $\mathscr{N}_1$ and $\mathscr{N}_2$ are said to be compatible if $d_{avg}(\mathscr{N}_1,\mathscr{N}_2) = 0$*

3.1 Basic Properties

In this subsection, we prove several important properties of our definition. These properties show that our definition fits well into existing results. In this way, we hope to convince the reader that our definition is sensible. We also show in this subsection, our definition of incompatibility in fact defines a metric in strict mathematical sense on rank one projective measurements. A significant consequence is that we can use $d_{avg}(\cdot,\cdot)$ to do a quantum learning task, we discuss this in section 4.

Consider the change of coordinate matrix $A_{ij} = \langle\alpha_i|\beta\rangle$ between the two orthonormal basis. As a computational aid, observe that:

$$\begin{aligned}\operatorname{Tr} P_i Q_j P_i Q_j &= \operatorname{Tr} |\alpha_i\rangle\langle\alpha_i|\beta_j\rangle\langle\beta_j|\alpha_i\rangle\langle\alpha_i|\beta_j\rangle\langle\beta_j| \\ &= \langle\alpha_i|\beta_j\rangle\langle\beta_j|\alpha_i\rangle\langle\alpha_i|\beta_j\rangle\langle\beta_j|\alpha_i\rangle = |A_{ij}|^4\end{aligned} \tag{6}$$

This allows $d_{avg}(\mathscr{N}_1,\mathscr{N}_1)$ to be written in a more compact form:

$$(d_{avg}(\mathscr{N}_1,\mathscr{N}_1))^2 = 2 - \frac{2}{d}\|A\|_4^4 \tag{7}$$

Where $\|A\|_4$ is the 4-norm of A.
The following proposition shows that our definition reflects conventional ideas of incompatibility between projective measurements discussed in section 2.3.

Proposition 2. *1) Two (rank one) projective measurements $\mathscr{N}_1,\mathscr{N}_2$ are compatible with respect to definition 9 if and only if they are on the same basis, i.e, the two observables commute or in other words, they are compatible in the sense of joint measurability.*
2) Two (rank one) projective measurements $\mathscr{N}_1,\mathscr{N}_2$ are maximally incompatible with respect to definition 9 if and only if they are mutually unbiased, i.e. $|\langle\alpha_i,\beta_j\rangle|^2 = \frac{1}{d}, \forall i,j$.

Proof.
To facilitate the proof, we define a new matrix on A, by letting $B \in \mathsf{Lin}\mathbb{R}^d$, a real linear operator, to be such that $B_{jk} = |A_{jk}|^2$. Observe that $\|A\|_4^4 = \|B\|_2^2$ where $\|B\|_2$ is the 2-norm of B.
Note that the change of coordinate matrix A between two orthonormal bases is unitary, making B double stochastic and positive in each entry. In particular this makes the maximum of $\|A\|_4^4$ easy to compute:
$\sum_j B_{ij}^2 \le (\sum_j B_{ij})^2 = 1, \forall i = 1,...,d$, then $\|B\|_2^2 \le d$ with equality holds if and only if for each row of the matrix B there is only one none-zero entry, which is equal to 1. Equivalently, A is almost a permutation matrix except that each none zero entry in A can have a global phase. This proves (1).

To see (2), we first note that $\|A\|_4^4 = \|B\|_2^2$ has an obvious lower bound: The 2-norm of B has a lower bound which is its spectral radius, and noting that B is double stochastic and symmetric, we conclude $\|B\|_2^2 \le 1$. This lower bound is reached, only when $B_{ij} = \frac{1}{d}$, which can be shown by applying the Lagrange multiplier method on the minimization problem. However, $B_{ij} = \frac{1}{d}$ exactly means $|\langle\alpha_i|\beta_j\rangle|^2 = \frac{1}{d}$, i.e the two bases are mutually unbiased. One example is when $|\alpha_j\rangle$ is Fourier transformed of $|\beta_k\rangle$:

$$|\alpha_j\rangle = \frac{1}{\sqrt{d}}\sum_{k=0}^{d} e^{\frac{2\pi ijk}{d}}|\beta_k\rangle, i^2 = -1, \forall j \in \{1,...,d\} \tag{8}$$

This makes $|\langle\alpha_j|\beta_k\rangle|^2 = (\frac{1}{\sqrt{d}}e^{\frac{2\pi ijk}{d}})^2 = \frac{1}{d}$.

Proposition 2 shows that d_{avg} we defined has nice mathematical properties to serve as a measure of incompatibility between measurements. But what it actually means deserves some more explanations. Now note that the measurements $\mathscr{N}_1,\mathscr{N}_2$ have a conventional implementation, their underlying channels can be described as $\mathscr{N}_1(\rho) = \sum_i P_i\rho P_i$ and $\mathscr{N}_2(\rho) = \sum_j P_j\rho P_j$ respectively, so we may rewrite:

$$\sum_{i,j} \operatorname{Tr} P_i Q_j P_i Q_j = \sum_i \operatorname{Tr} P_i \mathscr{N}_2(P_i)$$

where $\operatorname{Tr} P_i\mathscr{N}_2(P_i)$ is exactly the probability of measurement $\mathscr{N}_1$ produces outcome i when its measuring the output state of a measurement $\mathscr{N}_2$ on P_i. In this sense, each individual terms

$\operatorname{Tr} P_i \mathscr{N}_2(P_i)$ quantifies how much the measurement $\mathscr{N}_1$'s element P_i is disturbed by a measurement on $\mathscr{N}_2$. Summing over all the $\operatorname{Tr} P_i \mathscr{N}_2(P_i)$, our definition d_{avg} can be seen as quantifying the averaged mutual disturbance between two measurement bases. So instead of measuring the probability distribution's mutual disturbance such as in definition 8, our definition reveals directly the complementarity between each measurement elements of the two measurements.

In the following subsection, we prove another interesting property of $d_{avg}(\cdot,\cdot)$, that is, it satisfies triangle inequality.

3.2 A Metric

If we assume that $d_{avg}(\cdot,\cdot)$ is indeed a metric for now, we can see there's an interesting connection between $d_{avg}(\cdot,\cdot)$ and the Bures distance between two densitoy operators ρ_1, ρ_2, defined as [31]:

$$D_B(\rho_1,\rho_2)^2 = 2(1-\sqrt{F(\rho_1,\rho_2)}), F(\rho_1,\rho_2) = [\operatorname{Tr}(\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}})]^2 \tag{9}$$

Here $F(\rho_1,\rho_2)$ is the fidelity between ρ_1,ρ_2. Note that the equation (9) and definition 9 admit a very similar form, it suggests us to consider defining a fidelity for (rank one) projective measurements:

Definition 10.

$$\hat{F}(\mathscr{N}_1,\mathscr{N}_2) = (\frac{1}{d}(\sum_{i,j} \operatorname{Tr} P_i Q_j P_i Q_j))^2 \tag{10}$$

We may now in the report refer to $\hat{F}(\mathscr{N}_1,\mathscr{N}_2)$ as a fidelity.

Theorem 2. *$d_{avg}(\mathscr{N}_1,\mathscr{N}_2)$ is a metric between $\mathscr{N}_1,\mathscr{N}_2$, on the space of all rank one projective measurements over the Hilbert space $\mathcal{H}^d$.*

Proof.
Note that in proposition 2, we've already shown that d_{avg} vanishes if and only the projective measurements are on the same basis up to global phases and it's obviously non-negetive. The only non-trivial part to prove is actually the triangle inequality. To prove this, we associate to each projective measurements a density operator (Choi operator):

$$\tau_{\mathscr{N}} = (\mathscr{N} \otimes Id)(|\psi\rangle\langle\psi| \otimes |\psi\rangle\langle\psi|)$$

Where $|\psi\rangle$ is a maximally entangled state chosen according to the measurement basis concerned. In particular, we define for $\mathscr{N}_1$, $\tau_{\mathscr{N}_1} = (\mathscr{N}_1 \otimes Id)(\sum_{i,j} |\alpha_i\rangle\langle\alpha_j| \otimes |\alpha_i\rangle\langle\alpha_j|) = \sum_i |\alpha_i\rangle\langle\alpha_i| \otimes |\alpha_i\rangle\langle\alpha_i|$, and for $\mathscr{N}_2$, $\tau_{\mathscr{N}_2} = (\mathscr{N}_2 \otimes Id)(\sum_{i,j} |\beta_i\rangle\langle\beta_j| \otimes |\beta_i\rangle\langle\beta_j|)$
$= \sum_j |\beta_j\rangle\langle\beta_j| \otimes |\beta_j\rangle\langle\beta_j|$.
The key observation is that $d_{avg}(\mathscr{N}_1,\mathscr{N}_2)$ corresponds to the induced metric of $\langle\cdot,\cdot\rangle$ (the Hilbert Schmidt inner product) on $\boldsymbol{B}(\mathbb{C}^d \otimes \mathbb{C}^d)$ up to a constant factor $\sqrt{\frac{1}{d}}$:

$$\begin{aligned}
\langle\tau_{\mathscr{N}_1},\tau_{\mathscr{N}_2}\rangle_F = \operatorname{Tr} \tau^\dagger_{\mathscr{N}_1}\tau_{\mathscr{N}_2} &= \operatorname{Tr} \frac{1}{d}\sum_i P_i \otimes P_i \frac{1}{d}\sum_j Q_j \otimes Q_j \\
&= \frac{1}{d^2}\sum_{i,j} \operatorname{Tr} |\alpha_i\rangle\langle\alpha_i|\beta_j\rangle\langle\beta_j| \otimes |\alpha_i\rangle\langle\alpha_i|\beta_j\rangle\langle\beta_j| \\
&= \frac{1}{d^2}\sum_{i,j} \langle\alpha_i|\beta_j\rangle\langle\beta_j|\alpha_i\rangle\langle\alpha_i|\beta_j\rangle\langle\beta_j|\alpha_i\rangle \\
&= \frac{1}{d^2}\sum_{i,j} \operatorname{Tr} P_i Q_j P_i Q_j \\
&= \frac{1}{d}\sqrt{\hat{F}(\mathscr{N}_1,\mathscr{N}_2)}
\end{aligned} \tag{11}$$

$$
\begin{aligned}
\|\tau_{\mathcal{N}_1} - \tau_{\mathcal{N}_2}\|_F^2 &= \langle \tau_{\mathcal{N}_1} - \tau_{\mathcal{N}_2}, \tau_{\mathcal{N}_1} - \tau_{\mathcal{N}_2} \rangle_F \\
&= \langle \tau_{\mathcal{N}_1}, \tau_{\mathcal{N}_1} \rangle_F + \langle \tau_{\mathcal{N}_2}, \tau_{\mathcal{N}_2} \rangle_F - 2\langle \tau_{\mathcal{N}_1}, \tau_{\mathcal{N}_2} \rangle \\
= \frac{2}{d} - \frac{2}{d^2} \sum_{i,j} \operatorname{Tr} P_i Q_j P_i Q_j &= 2(1 - \sqrt{\hat{F}(\mathcal{N}_1, \mathcal{N}_2)}) \\
&= \frac{1}{d} d_{avg}(\mathcal{N}_1, \mathcal{N}_2)^2
\end{aligned} \tag{12}
$$

Equation (21) shows that $d_{avg}(\mathcal{N}_1, \mathcal{N}_2)$ is precisely $\|\tau_{\mathcal{N}_1} - \tau_{\mathcal{N}_2}\|_F$ up to a constant factor. Then the correctness of $d_{avg}(\cdot, \cdot)$ as a metric is an immediate consequence of $\|\cdot, \cdot\|_F$ being a metric. For instance, triangle inequality follows from $d_{avg}(\mathcal{N}_1, \mathcal{N}_3) = \|\tau_{\mathcal{N}_1} - \tau_3\| \leq \|\tau_{\mathcal{N}_1} - \tau_2\| +]\|\tau_{\mathcal{N}_2} - \tau_3\| = d_{avg}(\mathcal{N}_1, \mathcal{N}_2) + d_{avg}(\mathcal{N}_2, \mathcal{N}_3)$

Note that $d_{avg}(\mathcal{N}_1, \mathcal{N}_2)$ is defined so by dropping a constant factor $\sqrt{\frac{1}{d}}$ compared to $\|\tau_{\mathcal{N}_1}, \tau_{\mathcal{N}_2}\|$, due to the fact that the dimension d can be exponentially large in terms of the number of subsystems involved. When d increases exponentially as the number of systems increases, we do not want the value of the incompatibility also drops exponentially, which would possibly be a big obstacle for estimating teh degree of incompatibility between measurement devices.

In additional to the correctness of the metric, we can gain some more insights from the preceding proof.

Remark 1. *Equation (11) further confirms the intuition that $\hat{F}$ shares similar properties with the definition of fidelity between density operators. Note that by Uhlmann's theorem [2]* [2]*:*

Theorem 3. *Suppose ρ and σ are states of a quantum system Q. Introduce a second quantum system R which is a copy of Q. Then*

$$\sqrt{F(\rho, \sigma)} = max_{|\psi\rangle, |\phi\rangle} |\langle \psi | \phi \rangle|, \tag{13}$$

where the maximization is over all purifications $|\psi\rangle$ of ρ and $|\phi\rangle$ of σ into RQ.

This suggests that a fidelity between density operators can be defined from inner product between purifications. Very similarly, the square root of $\hat{F}$ also admits a form (up to a constant factor $\frac{1}{d}$) of inner product between Choi-Jamiolkowski representations of projective measurements.

Now we conclude the discussion of basic mathematical properties of our measure of incompatibility between projective measurement by doing a comparison with the existing definitions.

The most obvious distinction is perhaps the degree of incompatibility is easy to compute in terms of our definition and automatically assumes an analytic form. While in both definition 7 and definition 8, no known analytic form exists and one need to solve a usually difficult maximization problem to obtain the degree of incompatibility (in fact, in the case of joint measurability, to solve the degree of incompatibility between a known pair of observables might itself be an open problem in the community). In terms of how should projective measurements be considered compatible, our definition, definition 6,7, and definition 8 give the same requirement: commutativity of observable. In terms of how projective measurements should be considered maximally incompatible (in a finite dimensional Hilbert Space), our definition and definition 8 both requires measurements to be on mutually unbiased bases (which has clear physical meanings), but such question still cannot be answered under definition 6, 7. The most significant advantage of our definition we have shown is that $d_{avg}(\cdot, \cdot)$ defines a metric. This opens up a lot of possibilities, for instance, we might consider using this degree of incompatibility to do learning tasks, we will address this issue in section 5.

But before that, in section 4, we show an amazing result which makes our definition also useful in practice: given 2 unknown (projective) measurement devices, we can estimate their degree of incompatibility in constant time using a quantum switch.

[2]Note that in Nielsen & Chuang's book, fidelity is defined as the square root of the fidelity that is adapted in this note

4 Efficient Estimation of d_{avg} Using a Quantum Switch and Generalizations

Suppose that we are given two measurement devices A and B that does measurements $\mathscr{N}_1$ and $\mathscr{N}_2$ (as in definition 9), is it possible to test how incompatible the two measurements are? Of course one may follow the definition 9, to do so you need to know $\{P_i\}, \{Q_j\}$ first and you can, by doing a quantum operation tomography whose time complexity (number of calls to $\mathscr{N}_1$, and $\mathscr{N}_2$) scales as $O(d^4)$, where d is the dimension of $\mathcal{H}^d$ (the Hilbert space the measurements are acting on) and we know that d increases exponentially as the number of subsystems [2]. To compute the degree of incompatibility in terms of definition 7 or definition 8 is even more difficult, because having information of $\{P_i\}, \{Q_j\}$ (the full descriptions of the measurements) does not necessarily means the degree can be calculated as we have discussed at the end of section 3.

However, it turns out that we do not need to know anything about the implementations of the measurement devices. Using a quantum switch, we can estimate $d_{avg}(\mathscr{N}_1, \mathscr{N}_2)$ in a constant time.

4.1 Protocol to Estimate $d_{avg}(\mathscr{N}_1, \mathscr{N}_2)$

1. Input the state $\frac{I}{d} \otimes |+\rangle\langle+|_c$ to the quantum switch $\mathcal{S}(\mathscr{N}_1, \mathscr{N}_2)$ ($\mathscr{N}_1, \mathscr{N}_2$ are the two projective measurements as in definition 9), the state after the evolution is then (discarding all measurement results):

$$\begin{aligned}\mathcal{S}(\mathscr{N}_1,\mathscr{N}_2)(\frac{I}{d}\otimes|+\rangle\langle+|_c) &= \sum_{i,j}(P_iQ_j\otimes|0\rangle\langle 0|_c + Q_jP_i\otimes|1\rangle\langle 1|_c)\frac{I}{d}\otimes|+\rangle\langle+|_c(Q_jP_i\otimes|0\rangle\langle 0|_c + P_iQ_j\otimes|1\rangle\langle 1|_c)\\ &= \frac{1}{2d}(\sum_{i,j}P_iQ_jQ_jP_i\otimes|0\rangle\langle 0|_c + Q_jP_iP_iQ_j\otimes|1\rangle\langle 1|_c + P_iQ_jP_iQ_j\otimes|0\rangle\langle 1|_c + Q_jP_iQ_jP_i\otimes|1\rangle\langle 0|_c)\\ &= \frac{1}{2d}(I\otimes I_c + \sum_{i,j}P_iQ_jP_iQ_j\otimes|0\rangle\langle 1|_c + Q_jP_iQ_jP_i\otimes|1\rangle\langle 0|_c)\end{aligned} \tag{14}$$

2. Then we measure the control system C on Fourier basis $\{\ |+\rangle, |-\rangle\ \}$. The probability of getting an outcome "-" is:

$$\begin{aligned}p(-) &= \frac{1}{2d}\operatorname{Tr}\langle-|_c(I\otimes I_c + \sum_{i,j}P_iQ_jP_iQ_j\otimes|0\rangle\langle 1|_c + Q_jP_iQ_jP_i\otimes|1\rangle\langle 0|_c)|-\rangle_c\\ &= \frac{1}{2} - \frac{1}{4d}(\sum_{i,j}\operatorname{Tr}P_iQ_jP_iQ_j + \operatorname{Tr}Q_jP_iQ_jP_i)\\ &= \frac{1}{2} - \frac{1}{2d}(\sum_{i,j}\operatorname{Tr}P_iQ_jP_iQ_j)\end{aligned} \tag{15}$$

Observe that:

$$4p(-) = d_{avg}(\mathscr{N}_1, \mathscr{N}_2)^2 \tag{16}$$

(16) means to estimate $d_{avg}(\mathscr{N}_1, \mathscr{N}_2)$, it suffices to estimate $p(-)$ instead. However, $p(-)$ is just the probability of getting an outcome "-" for a qubit measurement (which is of dimension 2). So each time we repeat the process, we will obtain an outcome either $+$ or $-$ in a constant time (only relevant to the time needed to do measurements $\mathscr{N}_1, \mathscr{N}_1$). Then by repeating the process, we can obtain arbitrarily accurate estimate of $p(-)$ which will lead to arbitrarily accurate estimate of $d_{avg}(\mathscr{N}_1, \mathscr{N}_2)$. This means to estimate $d_{avg}(\mathscr{N}_1, \mathscr{N}_2)$, the number of accesses to the measurement devices A, B is constant of the dimension d, inverse polynomial of the accuracy we need.

Such result shows that our definition can be very useful in practice. For instance, imagine you want to set up a circuit to do quantum key distribution. To do so, whether using a BB84 protocol or the protocol proposed in [32] (where eavesdropper cannot alter the data because Bell inequality is used to detect any such disturbance), you will need do measurements that are incompatible. Hence if you are equipped with a quantum switch, you are then able to pick out a desirable pair

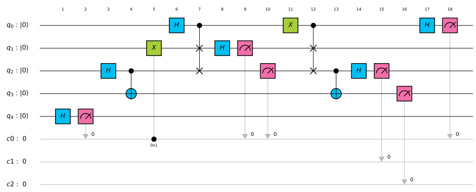

Figure 1: Circuit for qubit measurements

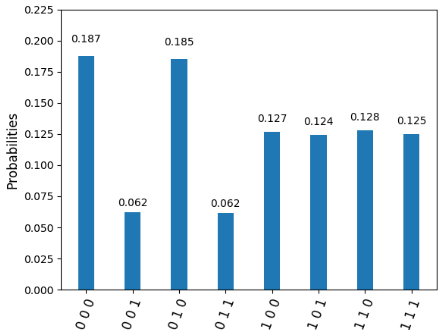

Figure 2: Simulation: 40000 shots

of measurement devices very quickly even if you do not know the really implementations of these devices.

As we have discussed in section 2.2.1, we cannot implement a quantum switch dterministically by a circuit in which two unknown oracles (in our case, measurements) are called a single time in a fixed causal order. But we are not denied from the possibility of implementing the procedure probabilistically. Figure 1 shows a quantum circuit that implements the protocol for qubit measurements on Fourier basis and computational basis (d = 2). In this circuit, the Quantum switch is successfully implemented when classical registers read out $c_1c_2 = 00$ [16][33]. Figure 2 shows the result of running this circuit for 40000 shots (in a simulator), note that only the first 2 columns are valid statistic. Simulation says that the degree of incompatibility is almost 1 which is exactly the theoretical value.

However, the implication of this procedure goes beyond what we have discussed. It suggests that for any quantum measurements $\mathcal{N}_1, \mathcal{N}_2$, given that we know what its underlying quantum channel is, we can apply the procedures (1) (2), then define the the degree of incompatibility between the two measurements as $2\sqrt{p(-)}$. The most obvious generalization is then to consider the case where the projectors are of rank higher than 1. Using the quantum channel we mentioned in the last paragraph of section 2.1, we have the following result:

4.2 Generalizations

Definition 11. *Consider two projective measurements $\mathcal{N}_1$ and $\mathcal{N}_2$ on $\mathcal{H}^d$, with POVM elements $\{ P_i \}_{i=1}^{k_1}$ and $\{ Q_j \}_{j=1}^{k_2}$ respectively, where P_i, Q_j are orthogonal projectors. The degree of incompatibility between $\mathcal{N}_1$ and $\mathcal{N}_2$ is defined to be $(d_{avg}(\mathcal{N}_1, \mathcal{N}_2))^2 = 2 - \frac{2}{d}\sum_{i=1}^{k_1}\sum_{j=1}^{k_2} \operatorname{Tr} P_iQ_jP_iQ_j$. $\mathcal{N}_1$ and $\mathcal{N}_2$ are said to be compatible if $d_{avg}(\mathcal{N}_1, \mathcal{N}_2) = 0$*

We are not assuming the projectors P_i, Q_j to be of rank one anymore, the consequence is then the proof of proposition 2 does not apply. However, it turns out that the proposition 2 still holds, in this way, our definition is now fully fitted into the conventional ideas. We summarize this also as a proposition.

Proposition 3. *1) Two projective measurements $\mathcal{N}_1, \mathcal{N}_2$ are compatible with respect to definition 11 if and only if they commute or in other words, they are compatible in the sense of joint measurability.*
2) Two projective measurements $\mathcal{N}_1, \mathcal{N}_2$ are maximally incompatible with respect to definition 11 if and only if all the projectors are of rank one and onto mutually unbiased basis, i.e. $P_i = |\alpha_i\rangle\langle\alpha_i|, Q_j = |\beta_j\rangle\langle\beta_j|$ and $|\langle\alpha_i, \beta_j\rangle|^2 = \frac{1}{d}, \forall i, j$.

Before diving into the proof, we first state a lemma here. A proof of this lemma can be found in [34] Corollary 7.62:

Lemma 1. *Let A, B be n by n Hermitian matrices: If A and B are positive semidefinite, then AB is diagonalizable and has non-negative eigenvalues.*

Proof.
For (1) We first prove that $\operatorname{Tr} P_iQ_jP_iQ_j \geq 0$:
Note that $P_i^2 = P_i$ we have $\operatorname{Tr} P_iQ_jP_iQ_j = \operatorname{Tr} P_iP_iQ_jP_iQ_j = \operatorname{Tr}(P_iQ_j)P_i(Q_jP_i)$. Note that $(P_iQ_j)P_i(Q_jP_i)$ is a postivie semi-definite operator since for any $|x\rangle \in \mathcal{H}^d$, $\langle x|(P_iQ_j)P_i(Q_jP_i)|x\rangle = (\langle x|P_iQ_j)P_i(Q_jP_i|x\rangle) \geq 0$ (by definition the orthogonal projector P_i is positive semi-definite).

Recall the standard inner product on the space of linear operators: $\langle A, B\rangle = \operatorname{Tr} B^\dagger A$. Note that $\operatorname{Tr} P_iQ_jP_iQ_j$ is essentially the inner prodcut between P_iQ_j and Q_jP_i, and noting the non-negativity of $\operatorname{Tr} P_iQ_jP_iQ_j$ we may then apply the Cauchy Schwartz Inequality:

$$\begin{aligned}\sum_{i=1}^{k_1}\sum_{j=1}^{k_2} \operatorname{Tr} P_iQ_jP_iQ_j &= \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} |\operatorname{Tr} P_iQ_jP_iQ_j| = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} |\langle P_iQ_j, Q_jP_i\rangle| \\ &\leq \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} \|P_iQ_j\|\|Q_jP_i\| = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} \sqrt{\operatorname{Tr} P_iQ_jQ_jP_i}\sqrt{Q_jP_iP_iQ_j} \\ &= \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} \sqrt{\operatorname{Tr} P_iQ_j}\sqrt{\operatorname{Tr} Q_jP_i} = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} \operatorname{Tr} P_iQ_j \\ &= \operatorname{Tr} \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} P_iQ_j = \operatorname{Tr} I = d\end{aligned} \tag{17}$$

Where the second inequality follows from the Cauchy Schwartz inequality.
Note that for the equality to hold, we must have $P_iQ_j = \alpha Q_jP_i$ for some $\alpha \in \mathbb{C}$. Note that in this case $0 \leq \operatorname{Tr} P_iQ_jP_iQ_j = \langle P_iQ_j, Q_jP_i\rangle = \langle P_iQ_j, \alpha P_iQ_j\rangle = \alpha\|P_iQ_j\|^2$. Thus $\alpha \geq 0$. On the other hand $\operatorname{Tr} P_iQ_j = \operatorname{Tr} Q_jP_i$, so we must have $\alpha = 1$ or $\operatorname{Tr} P_iQ_j = 0$. In the latter case, $\operatorname{Tr} P_iQ_j = \|P_iQ_j\|^2 = 0$ means $P_iQ_j = 0 = Q_jP_i$. So in either cases, P_i, Q_j must be commutative.

A remark here is that in the case of rank one projectors, commutativity is equivalent to say that the bases coincides and thus equivalent to say that the two measurements are the same. But in non rank one case, we obviously do not have such property. So we cannot expect for a metric.

For (2), we first consider only $\sum_j \operatorname{Tr} P_iQ_jP_iQ_j$, fixing i. Consider the matrix P_iQ_j, assume P_i is of rank r_i, then P_iQ_j is certainly of rank less then r_i. Since P_i, Q_j are orthogonal projectors, they are automatically Hermitian and positive semi-definite, so by lemma 1, we can diagonalize $P_iQ_j = X\Delta X^{-1}$, where Δ is a diagonal matrix with non-negative entries, of which at most r_i are non-zero. Let $\lambda_{j1}, \cdots \lambda_{jr_i}$ be such eigenvalues (not necessarily all non-zero). Considering all possible $j = 1, \cdots k_2$:

$$r_i = \operatorname{Tr} P_i = \sum_{j=1}^{k_2} \operatorname{Tr} P_iQ_j = \sum_{j}^{k_2}\sum_{s=1}^{r_i} \lambda_{js} \tag{18}$$

Now if we consider $\operatorname{Tr} P_iQ_jP_iQ_j = \operatorname{Tr}(P_iQ_j)^2$, it is immediately obvious that $\operatorname{Tr} P_iQ_jP_iQ_j = \sum_{s=1}^{r_i} \lambda_{js}^2$. Now from equation (18) and the constraint that all $\lambda_{js} \geq 0$, it is easy to show, using Lagrange multiplier method, we have:

$$\sum_{j}^{k_2} \operatorname{Tr} P_iQ_jP_iQ_j \geq \sum_{j}^{k_2}\sum_{s=1}^{r_i} (\frac{r_i}{k_2 r_i})^2 = \frac{r_i}{k_2} \tag{19}$$

So this means we have the lower bound:

$$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2} \operatorname{Tr} P_iQ_jP_iQ_j \geq \sum_{i=1}^{k_1} \frac{r_i}{k_2} = \frac{d}{k_2}. \tag{20}$$

Similarly, if we go from Q_j, then we would obtain another lower bound $\frac{d}{k_1}$ and thus we have:

$$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2} \operatorname{Tr} P_iQ_jP_iQ_j \geq max(\frac{d}{k_1}, \frac{d}{k_2}). \tag{21}$$

So from here, (2) can be easily seen.

In principle, we can also prove the triangle inequality, by associating to the measurement $\mathscr{N}_1$ for instance $\sum_i P_i \otimes P_i$. But we are in general not interested in this case since for non rank one projective measurements, we cannot obtain a metric anyway.

One could also ask, how to define the degree of incompatibility between non-projective POVMs? In principle, we can also use the procedure in section 4.1 and the $p(-)$ to define a degree. However, the problem is in general, a POVM is not associated with a quantum channel, i.e, there could be infinitely many distinct implementations of the same POVM. Hence, when defining the degree of two POVMs, we must first declare what instruments that implement the POVMs are we assuming. However, if the POVM elements are restricted to be of rank one, then we can uniquely identify the underlying channels and explicitly define the degree of incompatibility. We give an example:

Consider a POVM M_1 with rank one POVM elements on $\mathcal{H}^2$: $\{\frac{2}{3} \mid \phi_i\rangle\langle\phi_i\mid\}_{i=1}^3$, where $|\phi_1\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}, |\phi_2\rangle = \frac{|0\rangle+e^{\frac{2\pi i}{3}}|1\rangle}{\sqrt{2}}, |\phi_3\rangle = \frac{|0\rangle+e^{\frac{-2\pi i}{3}}|1\rangle}{\sqrt{2}}$. Then what is the degree of incompatibility between this POVM and itself? To define the degree of incompatibility, we associate to the POVM a quantum channel which implements this POVM. Since all POVM elements are of rank 1, we can write down the Kraus representation of the unique quantum channel that implements this POVM. In fact, the Kraus operators are simply $\{\sqrt{\frac{2}{3}} \mid \phi_i\rangle\langle\phi_i\mid\}_{i=0}^3$. So following the procedure in section 4.1, after some calculations, we would observe that $p(-) = \frac{1}{2} - \frac{1}{4} \times \frac{3}{2} = \frac{1}{8}$ and thus $d_{avg}(M_1, M_1) = \frac{\sqrt{2}}{2}$. Note that this value is not 0! On one hand, it is not surprising, since indeed a POVM would disturb itself by repeated measurements, i.e the probability distribution generated by doing a POVM measurement twice can be different from the one generated by doing the POVM only once. On the other hand, if you look at it from a Joint device perspective, then clearly, a POVM should be compatible with itself, the joint measurement can simply be: do POVM once, copy the classical output, output the same 2 values. In such a way, to define and make sense of the definition of degree of incompatibility between general POVMs is trickier.

5 A Quantum Learning Task; Clustering Projective Measurement

What has been understood is that the degree of incompatibility between projective measurements defined by us is easy to compute/estimate in both theory and practice. It satisfies much of our intuitions for such a measure. It is also a metric (for rank one projective measurements). So it is natural to ask, cam we use this quantity to do quantum learning tasks? One obvious task is then to cluster a set of unknown (rank one) projective measurements using the degree of incompatibility as the metric, and classical clustering algorithms. To answer this question, we have done the following experiments.

5.1 Clustering Fourier Basis, Computational Basis and GBS Basis

Here we try to cluster a list of (rank one) projective measurements over $\mathcal{H}^d$ which are chosen uniformly at random from computational basis, Fourier basis and generalized bell states.
The Fourier basis of dimension $d = 2^t$ is defined as:

$$|j\rangle = \frac{1}{\sqrt{d}} = \sum_{k=0}^{d-1} e^{\frac{2\pi ijk}{d}} |k\rangle \tag{22}$$

A generalized bell state in a bipartite composite system $A \otimes B$ of dimension $d = \sqrt{d} * \sqrt{d}$ is obtained by applying the Heisenberg-Weyl operators on the maximally entangled bipartite state

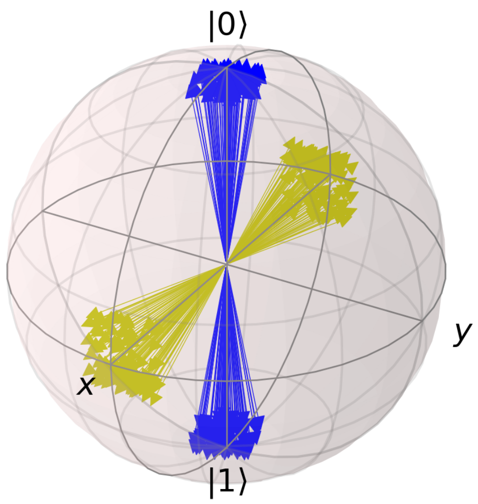

Figure 3: Fourier and Computational Measurements with Perturbation

$|\phi\rangle^{AB} = \frac{1}{\sqrt[4]{d}} \sum_{j=0}^{\sqrt{d}-1} |j\rangle^A |j\rangle^B$. Obtaining the basis $\{\ |\ \Phi_{m,n}\rangle^{AB}\ \}$ as:

$$|\Phi_{m,n}\rangle^{AB} = \frac{1}{\sqrt[4]{d}} \sum_{k=0}^{\sqrt{d}-1} e^{\frac{i2\pi mk}{\sqrt{d}}} |k\rangle^A |k \ominus n\rangle^B \tag{23}$$

The degree of incompatibility between projective measurements were pre-computed using definition 9 and presented to a clustering algorithm in the form of a distance matrix (of the randomly selected list of bases). For efficiency, the dimension of the systems were restricted to 4, 16 and 64 (corresponding to 2, 4 and 6 qubit systems) and the length of the list was restricted to 100. Summing up, the knowledge known to the clustering algorithm are the number of clusters (3 in this experiment) and the distance between projective measurements in the list.

The performance of the clustering algorithm was measured using "Adjusted Mutual Information" [35],where a score of 0 indicates a random clustering, 1 indicates a perfect match. In particular, we performed the list generation and clustering for 50 times and reported the time that a perfect match was obtained.
We chose to use the k-medoids algorithm, which is a variant of the standard k-means algorithm [36].

Experimental results showed that in all settings (dimension 4, 16, 64), the clustering algorithm always achieved a perfect labelling, i.e, the projective measurements were always correctly clustered. However, this was only true when we applied a cautious initial seeding process, where the initial medoids were selected in a k-means ++ style [37]. Note that a random initial seeding process could not achieve the same performance as the k-means ++ style initial seeding. This suggests that clustering projective measurements is highly sensitive to initial seeding.

5.2 Clustering with Perturbations

It is also interesting to know how the clustering algorithm reacts to perturbations. Suppose that we know the black boxed qubit projective measurements are either on Fourier basis $\{\ |+\rangle, |-\rangle\ \}$ or on Computational basis $\{\ |0\rangle, |1\rangle\ \}$. However, implementations of the measurement devices are not perfect and the bases are perturbed by a small factor. This can be visualized on the Bloch Sphere: note that qubit projective measurements are antipodal points on Bloch Sphere. By saying the projective measurements are perturbed, we mean that their representation on Bloch Sphere are perturbed by a small factor, see figure 3. Restricting that the perturbations are relatively small, for instance 22.5° off on Bloch Sphere as in figure 3, we did an experiment similar to the

previous one: generate a sequence of projective measurements on qubit, each one of them being either Fourier mesurements or Computational measurements, but randomly perturbed as in figure 3. Then construct the distance matrix, cluster the sequence of measurements into 2 clusters using k medoids algorithms with k means ++ style initial seeding.
We repeated the procedure for 50 times and in each time the clustering were always perfect. This result together with the previous one suggest that our definition of degree of incompatibility between projective measurements is a very useful metric in the task of clustering (rank one) projective measurements when the projective measurements are themselves well separated and the number of clusters is known, even in the presence of perturbation.

However, one would expect to extract some more general notions of similarity among the block boxed projective measurements. For instance: distinguish product measurements and entangled measurements. We deal with this in the following section.

5.3 Clustering Product Measurements and Entangled Measurements

Another interesting question to ask is if this clustering process is capable of distinguish product and entangled measurements? Consider the task where we are asked to cluster a set of projective measurements, each of which is either a product measurements (all basis elements in product states) or an entangled measurements (all basis elements in entangled states), into 2 clusters; product measurements being in one cluster and entangled measurements being in the other one.

It is clear that a product measurement and a entangled measurement always have non zero degree of incompatibility from proposition 2. This suggests that we might indeed cluster a set of product measurements and entangled measurements, although one should not expect a perfect clustering.

One can implement this idea in the following way:

Firstly, measurements are all assumed to be on 2 qubit systems $A \otimes B$ (of dimension 4) and generated at random: for product basis: first generate 2 bases at random for A and B, i.e. $\{\ |\alpha_0\rangle_A, |\alpha_1\rangle_A\ \}$ and $\{\ |\beta_0\rangle_B, |\beta_1\rangle_B\ \}$, then the basis for $A \otimes B$ is:

$\{\ |\alpha_0\rangle_A|\beta_0\rangle_B, |\alpha_1\rangle_A|\beta_0\rangle_B, |\alpha_0\rangle_A|\beta_1\rangle_B, |\alpha_1\rangle_A|\beta_1\rangle_B\ \}$;

Similarly for entangled basis, first generate at random $\{\ |\alpha_0\rangle'_A, |\alpha_1\rangle'_A\ \}$ and $\{\ |\beta_0\rangle'_B, |\beta_1\rangle'_B\ \}$, then the basis of $A \otimes B$ is put into:

$\{\ \frac{(|\alpha_0\rangle'_A|\beta_0\rangle'_B + |\alpha_1\rangle'_A|\beta 1\rangle'_B)}{\sqrt{2}}, \frac{(|\alpha_0\rangle'_A|\beta_0\rangle'_B - |\alpha_1\rangle'_A|\beta_1\rangle'_B)}{\sqrt{2}}, \frac{(|\alpha_0\rangle'_A|\beta_1\rangle'_B + |\alpha_1\rangle'_A|\beta_0\rangle'_B)}{\sqrt{2}}, \frac{(|\alpha_0\rangle'_A|\beta_1\rangle'_B - |\alpha_1\rangle'_A|\beta_0\rangle'_B)}{\sqrt{2}}\ \}$

Then following this scheme we generate at random a sequence of 100 product and entangled measurements. Calculate the distance matrix as before, use a k means ++ style initial seeding, set the number of clusters to be two and then apply the k medoids algorithm. Then measure the performance of each individual clustering process using "Adjusted Mutual Information". Finally, repeat the process for 50 time and report the average score.

Not surprisingly, the reported score was not as perfect as the first two experiments since product measurements themselves can still differ a lot (similarly for entangled measurements) in terms of degree of incompatibility. The average score was 0.33 and this result was stable upon repetition. This means we are able to distinguish random product measurement and entangled measurements to some extends and achieve a better performance than random assignments. But it is highly unlikely to achieve a perfect separation.

5.4 Summary of Experimental Results

Simulation results suggest that the metric defined in definition 9 can be a useful distance measure in tasks of clustering a set of projective measurements, given that a proper clustering algorithm is applied. An implicit assumption here is that the degree of incompatibility can be estimated up to a high precision in practice, since the only inaccuracy in our simulation is caused by the inaccuracy of floating point arithmetic in the classical computer.

We would also like to point out that although in our experiments, we only considered composite systems of no more than 6 qubits due to the limit of the power of our classical simulator, it is not to be assumed we cannot cluster projective measurements that act on large systems. In fact, in section 4.1 we have already argued that the degree of incompatibility between projective measurements can estimated in constant time however large the system is.

6 Conclusion

In conclusion, we have defined a new measure of how incompatible two (projective) quantum measurements can be. Such idea of incompatibility is a unique feature of quantum physic which distinguishes the quantum world and the classical world, and admits many applications in quantum information theory (see section 2.3 and section 4). In section 3 and section 4, we proved several mathematical properties to justify the meaningfulness of our definition from different perspective. In this report, we showed that our definition captures the essence of the other well accepted measure of incompatibility for quantum measurements. In particular, we showed our definition agrees with others on saying that projective measurements with commutative measurement elements are compatible. We also explicitly calculated the pair of measurements that are maximally incompatible (under our definition) and demonstrated that these measurements are indeed intuitively maximally incompatible, and we showed that they are very useful in quantum information theory.

What makes our definition stand out from the other existing measures is that using a new computational primitive: quantum switch, we can estimate the degree of incompatibility between two measurements in constant time of the dimension of the system. In this way, we have related our work with the idea of superposition of casual order. In terms of the research in quantum switch and indefinite causal order, our result (especially section 4) provides another example of quantum computation model with indefinite causal order allowing a better time complexity in computational tasks. We would also like to point out that the computability both in theory and in practice is another merit of our definition (see section 4).

In relation to machine learning, we argued in section 5 that the degree of incompatibility between (rank one) projective measurements being a metric enables us to perform a quantum learning task. We demonstrated that we can cluster a set of projective measurement devices using this metric (under minor assumptions) and suitable clustering algorithm. Here, we illustrated how quantum theory can actively interact with classical results in machine learning.

As a final comment, we would also like to point out some future directions of research. In section 4, we mentioned (and illustrated with an example) that the definition of degree of incompatibility can be extended general POVMs, give that we know the quantum channel that implements the measurement. However, the concrete analysis of this generalization still not exists, thus it would be interesting to study how such generalization would diverge from the known results. On the other hand, it would also be interesting to further explore the usefulness of our definition in quantum information and computation theory (one example was discussed in section 4).

References

[1] Simulating physics with computers. https://people.eecs.berkeley.edu/~christos/classics/Feynman.pdf. Accesses: 2018-10-20.

[2] Michael A. Nielsen and Isaac Chuang. Quantum computation and quantum information. *American Journal of Physics*, 70(5):558–559, 2002.

[3] G. Brassard, P. Hoyer, M. Mosca, and A. Tapp. Quantum Amplitude Amplification and Estimation. *eprint arXiv:quant-ph/0005055*, May 2000.

[4] B. Tamir and E. Cohen. Notes on Adiabatic Quantum Computers. *ArXiv e-prints*, December 2015.

[5] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser. Quantum Computation by Adiabatic Evolution. *eprint arXiv:quant-ph/0001106*, January 2000.

[6] S. Lloyd, M. Mohseni, and P. Rebentrost. Quantum algorithms for supervised and unsupervised machine learning. *ArXiv e-prints*, July 2013.

[7] Wikipedia contributors. Quantum — Wikipedia, the free encyclopedia, 2018. [Online; accessed 19-October-2018].

[8] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd. Quantum machine learning. , 549:195–202, September 2017.

[9] A. W. Harrow, A. Hassidim, and S. Lloyd. Quantum Algorithm for Linear Systems of Equations. *Physical Review Letters*, 103(15):150502, October 2009.

[10] S. Lloyd, M. Mohseni, and P. Rebentrost. Quantum principle component analysis. *nature physics*, July 2014.

[11] Esma Aïmeur, Gilles Brassard, and Sébastien Gambs. Quantum speed-up for unsupervised learning. *Machine Learning*, 90(2):261–287, Feb 2013.

[12] Kevin Patrick Murphy. *Machine Learning: a Probabilistic Perspective*. MIT Press, 2014.

[13] Teiko Heinosaari, Takayuki Miyadera, and Mario Ziman. An Invitation to Quantum Incompatibility. *arXiv e-prints*, page arXiv:1511.07548, November 2015.

[14] Teiko Heinosaari. Simultaneous measurement of two quantum observables: Compatibility, broadcasting, and in-between. *Physical Review A*, 93:042118, April 2016.

[15] Somshubhro Bandyopadhyay and Prabha Mandayam. Operational measure of incompatibility of noncommuting observables. *Physical Review A*, 87:042120, Apr 2013.

[16] G. Chiribella, G. M. D'Ariano, P. Perinotti, and B. Valiron. Quantum computations without definite causal structure. , 88(2):022318, August 2013.

[17] D. Ebler, S. Salek, and G. Chiribella. Enhanced Communication with the Assistance of Indefinite Causal Order. *Physical Review Letters*, 120(12):120502, March 2018.

[18] G. Chiribella. Perfect discrimination of no-signalling channels via quantum superposition of causal structures. , 86(4):040301, October 2012.

[19] O. Oreshkov, F. Costa, and Č. Brukner. Quantum correlations with no causal order. *Nature Communications*, 3:1092, October 2012.

[20] Stefano Facchini and Simon Perdrix. Quantum Circuits for the Unitary Permutation Problem. *arXiv e-prints*, page arXiv:1405.5205, May 2014.

[21] Mateus Araújo, Fabio Costa, and Časlav Brukner. Computational Advantage from Quantum-Controlled Ordering of Gates. , 113:250402, Dec 2014.

[22] A. A. Abbott, J. Wechs, D. Horsman, M. Mhalla, and C. Branciard. Communication through coherent control of quantum channels. *ArXiv e-prints*, October 2018.

[23] S. Salek, D. Ebler, and G. Chiribella. Quantum communication in a superposition of causal orders. *ArXiv e-prints*, September 2018.

[24] G. Chiribella, M. Banik, S. Sankar Bhattacharya, T. Guha, M. Alimuddin, A. Roy, S. Saha, S. Agrawal, and G. Kar. Indefinite causal order enables perfect quantum communication with zero capacity channel. *ArXiv e-prints*, October 2018.

[25] Paul Busch, Teiko Heinonen, and Pekka Lahti. Heisenberg's uncertainty principle. , 452:155–176, Nov 2007.

[26] P. Busch and P. J. Lahti. The complementarity of quantum observables: Theory and experiments. *Nuovo Cimento Rivista Serie*, 18:1–27, Apr 1995.

[27] Arthur Fine. Hidden variables, joint probability, and the bell inequalities. *Phys. Rev. Lett.*, 48:291–295, Feb 1982.

[28] Takayuki Miyadera and Hideki Imai. Heisenberg's uncertainty principle for simultaneous measurement of positive-operator-valued measures. *Physical Review A*, 78:052119, Nov 2008.

[29] Teiko Heinosaari, Jussi Schultz, Alessandro Toigo, and Mario Ziman. Maximally incompatible quantum observables. *arXiv e-prints*, page arXiv:1312.3499, Dec 2013.

[30] Nicolas J. Cerf, Mohamed Bourennane, Anders Karlsson, and Nicolas Gisin. Security of Quantum Key Distribution Using d-Level Systems. , 88:127902, Mar 2002.

[31] Wikipedia contributors. Bures metric — Wikipedia, the free encyclopedia, 2018. [Online; accessed 10-October-2018].

[32] Artur K. Ekert. Quantum cryptography based on bell's theorem. *Phys. Rev. Lett.*, 67:661–663, Aug 1991.

[33] Ibm q. https://www.research.ibm.com/ibm-q/. Accessed: 2018-10-20.

[34] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Matrix Analysis. Cambridge University Press, 2013.

[35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[36] Andrei Novikov. annoviko/pyclustering: pyclustering 0.8.1 release, May 2018.

[37] Wikipedia contributors. K-means++ — Wikipedia, the free encyclopedia, 2017. [Online; accessed 11-October-2018].