The Kirkwood-Dirac Quasi-probability: Characterising "Quantumness"

Arpan Akash Ray

A dissertation submitted for the partial fulfilment of BS-MS dual degree in Science

Indian Institute of Science Education and Research, Mohali

April 28, 2022

Certificate of Examination

This is to certify that the dissertation titled **The Kirkwood-Dirac Quasi-probability: Characterising "Quantumness"** submitted by **Arpan Akash Ray** (Reg. No. MS17046) for the partial fulfilment of BS - MS Dual Degree programme of the institute, has been examined by the thesis committee duly appointed by the institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.

Dr. Manabendra Nath Bera   Dr. Kinjalk Lochan   Dr. Sandeep K. Goyal

Dr. Manabendra Nath Bera
(Supervisor)

Dated: April 28, 2022

Declaration

The work presented in this dissertation has been carried out by me under the guidance of Dr. Manabendra Nath Bera at the Indian Institute of Science Education and Research, Mohali.

This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

Arpan Akash Ray
(Candidate)
Dated: April 28, 2022

In my capacity as the supervisor of the candidates project work, I certify that the above statements by the candidate are true to the best of my knowledge.

Dr. Manabendra Nath Bera
(Supervisor)
Dated: April 28, 2022

Acknowledgements

This project would not have been possible without the support of many people. Many thanks to my adviser, Dr. Manabendra Nath Bera, who provided me with constant guidance and helped make sense of some confusions that arose. Also thanks to my group members, who often asked instigating questions which led to invigorating discussions.

Special thanks to Sourav Das for his insights and valuable comments. And finally, thanks to my parents, and close friends who endured this long process with me, always offering support and love.

Abstract

In quantum mechanics, complex numbers appear naturally. This is, however, in contrast to classical mechanics. But for quantum theory, they are essential! In this thesis, we study a complex quasi-probability distribution which appears in the study of quantum information, known as the Kirkwood-Dirac distribution. We attempt to understand if this distribution stores meaningful information about a quantum state and propose a formalism to quantify it. Further, we also motivate how the Kirkwood-Dirac distribution may contain information about the coherence in a quantum state. We provide considerable evidence to support this claim, by deriving a coherence quantifier from the Kirkwood-Dirac distribution. We end by summarising our findings, and discussing future work on the subject.

List of Figures

Contents

Chapter 1

Introduction

"Negative energies and probabilities should not be considered as nonsense. They are well-defined concepts mathematically, like a negative of money."

P.A.M. Dirac

The Physical Interpretation of QM, 1942

1.1 "Quantumness"

Quantum mechanics has been distinguished from its classical counterparts due to many features. The most notable of them being superposition. A qubit can be a superposition of both $|0\rangle$ and $|1\rangle$. However, a classical bit can only be either 0 or 1. This fundamental difference sets the theory of quantum information apart, which is well known and acknowledged.

However, one more striking difference in quantum mechanics is the natural emergence of complex numbers. This would suggest that wherever these numbers appear, they should indicate true *quantumness*. The thesis is an in-depth investigation in this direction through an interesting quasi-probability distribution.

1.2 What is a Probability?

Probabilities appear in vastly different areas of study. However, the core meaning of a probability remains unchanged throughout these analyses; a probability function quantifies the likelihood of a favourable outcome. Thus a probability always encodes some information about the subject at hand. The question of how to then quantify this information was first addressed by Shannon (Shannon, 1948).

Given a random variable, $X \sim \{p_i\}$, the Shannon entropy is defined as

$$H(X) = -\sum_i p_i \log p_i. \tag{1.1}$$

The operational meaning of this formula is as follows. Consider a classical information source generating - $1,2,3,4$. No of bits required to encode this information is two, trivially. Now let's say that the source generates the same information, but probabilistically. Let us assume the Probability of occurrence of each digit is - $0.5, 0.25, 0.125, 0.125$ respectively. We can now use the bias between the source outputs to compress the information such that fewer bits are used to store commonly occurring symbols such as 1, and more bits are used to store rarely occurring symbols like 3 and 4. Now, the Shannon said that the *least* number of classical bits required to store is the Shannon entropy function defined above. A quick calculation reveals that $H(X) = 7/4$, which is less than two!

Hence the Shannon entropy provides a way to compress data losslessly and helps us, in a way, to quantify information.

1.3 But then, What is a *Quantum* Probability?

The von Neumann entropy is defined as (v. Neumann, 1933)

$$S(\rho) = -\mathrm{Tr}\{\rho \log \rho\}, \tag{1.2}$$

for a density matrix ρ. Notice how the probabilities are now replaced with a density matrix. But a density matrix is also a probability distribution.

Consider a basis $|i\rangle\langle i|$ in which ρ is diagonal. Then (1.2) becomes

$$S(\rho) = -\sum_i \lambda_i \log \lambda_i,$$

where $\{\lambda_i\}$ are the eigenvalues of ρ (remember, ρ is positive semi-definite and hermitian). It would seem that ρ is just like a classical probability distribution, up to a diagonalization.

However, a quantum state is much more than a classical probability! It is then, thus, safe to assume that the *quantumness* we are after must be hidden in the off-diagonal terms of a density matrix. This is not an absurd assumption. The l_1-norm of quantum coherence (which is often considered a quantifier of superposition (Aberg, 2006)) is defined as the sum of off-diagonal elements of a density matrix.

What kind of probability distribution can encode this information of *quantumness* in the off-diagonal terms? What constraints will such a distribution have? Moreover, how can one quantify the information held by it? Can we also concretely prove that such distributions do contain *quantumness*? And if so, how? In the following few pages, we will try to understand

and maybe answer these questions from the perspective of an interesting mathematical function: the Kirkwood-Dirac (KD) quasi-probability distribution.

1.4 The Roadmap

The thesis is divided into two parts. The first part covers questions about the informational content of the Kirkwood-Dirac distribution. We start with the definition and features of the KD distribution in Chapter 2. Chapter 3, we look at measure-theoretic definitions to mathematically make sense of our quasi-probability function. In Chapter 4, we derive an entropic measure from the KD distribution and explain its significance as a conditional entropy. Concluding our analysis of the information-theoretic aspects, we move on to part II, where we study the KD distribution in the light of quantum coherence. Chapter 5 serves as an introduction to the resource theory of coherence. Chapter 6 elaborates on the motivations behind considering that the KD distribution might hold information about the coherence of a quantum state. Finally, in Chapter 7, we define a coherence quantifier using the KD distribution. A summary of the entire text is provided in Chapter 8.

Part I

Quantifying the Informational Content

Chapter 2

Preliminaries

2.1 The Kirkwood-Dirac Distribution

Introduced by Dirac (Dirac, 1945) and further worked upon by Kirkwood, the Kirkwood-Dirac (KD) distribution is constructed for a given density matrix, using two sets of non-orthogonal projectors. Consider the complete set of observables A_m and B_n. Here,

$$A_m = |a_m\rangle\langle a_m|\,, B_n = |b_n\rangle\langle b_n|\,. \tag{2.1}$$

For simplicity we will only consider rank one projectors. Of course, $\sum_m A_m = 1$ and $\sum_n B_n = 1$. Then the distribution is given by

$$P(a_m, b_n) = \mathrm{Tr}\{\rho A_m B_n\}. \tag{2.2}$$

This definition satisfies all but one of Kolmogorov's axioms of probability. The KD is a non real function, hence it cannot be classified as a probability distribution in the usual sense. It was shown that the negativity and complexness of this function is due to the non-commutativity of the observables A, B (Johansen, 2007).

The marginals of the function, however, give rise to usual probabilities in quantum mechanics:

$$\begin{aligned}\sum_m P(a_m, b_n) = \sum_m \mathrm{Tr}\{\rho A_m B_n\} = \mathrm{Tr}\{\rho B_n\} \\ \sum_n P(a_m, b_n) = \sum_n \mathrm{Tr}\{\rho A_m B_n\} = \mathrm{Tr}\{\rho A_m\}.\end{aligned} \tag{2.3}$$

Hence, the normalised KD distribution satisfies

$$\sum_{m,n} P(a_m, b_n) = \sum_{m,n} \mathrm{Tr}\{\rho A_m B_n\} = 1. \tag{2.4}$$

At this juncture, the distribution is looking like a joint probability function. In later sections, we will expand upon that idea.

2.2 Informational Completeness

We can write (2.2) in the following manner,

$$P(a_m, b_n) = \mathrm{Tr}\{\rho A_m B_n\} = \langle b_n| \rho |a_m\rangle \langle a_m|b_n\rangle . \tag{2.5}$$

Consider given a density matrix ρ and two non compatible complete observable $A_m = |a_m\rangle \langle a_m|$ and $B_n = |b_n\rangle \langle b_n|$. Then, in b_n basis, one can write a matrix representation of ρ as,

$$\langle b_n| \rho |b_k\rangle = \sum_m \langle b_n| \rho |a_m\rangle \langle a_m|b_k\rangle .$$

Substituting the above equation in this expression, we get,

$$\langle b_n| \rho |b_k\rangle = \sum_m P(a_m, b_n) \frac{\langle a_m|b_k\rangle}{\langle a_m|b_n\rangle}. \tag{2.6}$$

Any state, then, may be written as

$$\rho = \sum_m P(a_m, b_n) \frac{|b_n\rangle \langle a_m|}{\langle a_m|b_n\rangle} = \sum_m P(a_m, b_n) \rho_{n|m}. \tag{2.7}$$

$\rho_{m|n}$ are basis vectors in the two-state vector (TSV) formalism (see Appendix A). Thus, given such set of projectors, the KD distribution is informationally complete.

2.3 In Weak Measurement Theory

2.3.1 Background

Interestingly, the KD distribution can be measured via a scheme called weak measurements. Consider the following scheme which is detailed in (Aharonov and Vaidman, 1990).

Starting with a pre-selection $|i\rangle$, we set up a weak interaction Hamiltonian for the measurement of A:

$$H = -g(t) q_n \otimes A.$$

Here, $g(t)$ is a normalized function with compact support near the time of measurement, and q is a canonical variable of the measuring device with conjugate momentum p. The index n refers to the nth system in the ensemble or nth measuring device.

The weak value, is defined as:

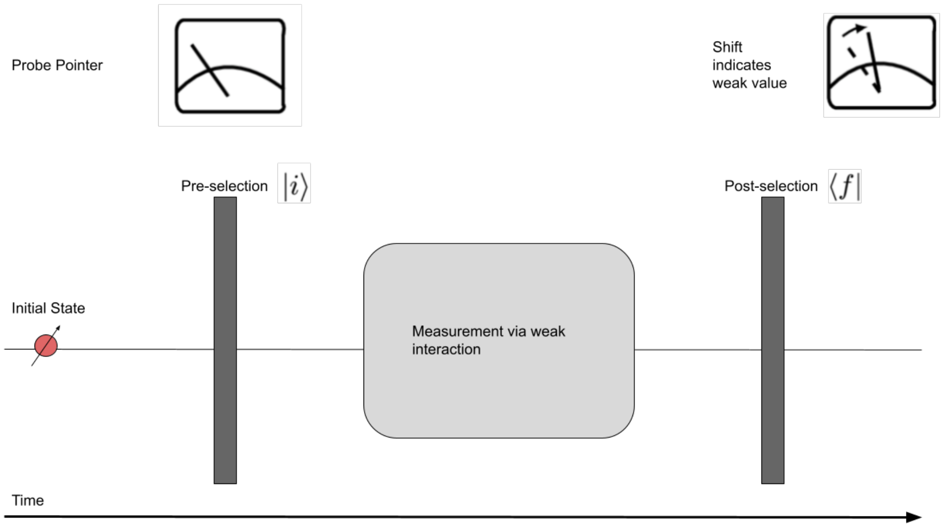

Figure 2.1: The weak measurement setup.

$$A_w = \frac{\langle f|A|i\rangle}{\langle f|i\rangle}. \tag{2.8}$$

It can be shown that the real and imaginary parts of the weak value can be estimated to a desired precision. We use the Gaussian wave function for the measurement device,

$$\frac{1}{\sqrt{\Delta q}(2\pi)^{1/4}} \exp\left\{-\frac{q_n^2}{4(\Delta q)^2}\right\}.$$

The final wave function of the measuring device in the p representation, to a good approximation, turns out to be proportional to:

$$\exp\left\{-(\Delta q)^2\left(p - \frac{\langle f|A|i\rangle}{\langle f|i\rangle}\right)^2\right\}.$$

Thus, the probability distribution of p is a Gaussian with spread $\Delta p = (2\Delta q)^{-1}$ centered at $p = \mathrm{Re}(A_w)$. The imaginary part is calulated by the distrubution of q. In the q representation, we have:

$$\exp\{iq\mathrm{Re}(A_w)\}\exp\left\{-\frac{\left(q + 2(\Delta q)^2\mathrm{Im}(A_w)\right)^2}{4(\Delta q)^2}\right\}.$$

The probability distribution of q is a Gaussian with the same spread Δq centered at $q = 2(\Delta q)^2\mathrm{Im}(A_w)$.

Of course, the uncertainty principle would not allow us to determine both these spreads from a single measurement. But, by using a large number of measurements, one can measure the weak value with arbitrary precision (Aharonov and Vaidman, 1990).

2.3.2 Link to the Kirkwood-Dirac distribution

Consider the pre-selected and post-selected states to be pure and be given by

$$\rho_i = |i\rangle\langle i|\,, \quad \rho_f = |f\rangle\langle f|\,.$$

Then the weak value can be written as:

$$A_w = \frac{\langle f|A\,|i\rangle}{\langle f|i\rangle} = \frac{\mathrm{Tr}\{\rho_i\rho_f A\}}{\mathrm{Tr}\{\rho_i\rho_f\}}. \tag{2.9}$$

The numerator is exactly the KD distribution. This implies that computing the weak value is equivalent to computing the distribution. In general, we define the following function:

$$\mathscr{K}_{mn} = \frac{\mathrm{Tr}\{\rho B_n A_m\}}{\mathrm{Tr}\{\rho B_n\}}. \tag{2.10}$$

Here, the measurement is given by A_m, and the post-selection is given by B_n. A weak measurement allows us to precisely determine the KD distribution.

To quantify the informational content of the distribution, we need to understand the quantity $\mathscr{K}_{mn}$ mathematically. We, therefore, turn towards measure theory in an attempt to make our understanding mathematically rigorous.

Chapter 3

Measure Theory

3.1 Basics

A set in mathematics is a collection of objects. Sets are also the simplest mathematical structures, and more complex theories (for example, topology and manifolds) are developed by adding complications to sets. For our use, we define the following.

Definition 1. *Let Ω be a set. A class $\mathscr{F}$ of subsets of Ω is said to be an algebra if*

1. $\Omega \in \mathscr{F}$,
2. $\Omega - A = A^c \in \mathscr{F}$ *whenever* $A \in \mathscr{F}$,
3. $A \cup B \in \mathscr{F}$ *whenever* $A, B \in \mathscr{F}$.

It follows that $\varnothing \in \mathscr{F}$, and by de Morgan's law $A \cap B \in \mathscr{F}$ whenever $A, B \in \mathscr{F}$. Thus an algebra is closed under the process of taking finite unions and intersections. If in addition, $\mathscr{F}$ is closed under taking countable unions, then it is called a σ-algebra.

In the simplest of terms, measure is a way to assign a notion of volume to sets. Such a function, then, should follow some simple common sense axioms. It should evaluate to zero for empty sets; should be positive for all inputs and for a union of disjoint inputs, it should be equal to the sum of each input. We thus have the following definition.

Definition 2. *Let Ω be a space and $\mathscr{F}$ be an algebra. Then a function $\mu : \mathscr{F} \to \mathbb{R}$ is called a measure if*

1. $\mu(\varnothing) = 0$,
2. $\mu(A) \geq 0$ *for all* $A \in \mathscr{F}$,

3. *If $A_1, A_2, \ldots$ is a sequence of disjoint sets in $\mathscr{F}$ such that $\cup_{i=1}^{\infty} A_i \in \mathscr{F}$ then,*

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

The last condition is called countable additivity. Finite additivity follows from the countable additivity. A measure thus defines a way to compare arbitrary sets. Probability, the likelihood of occurrence, is just a normalised measure. Hence for a probability measure,

$$\mu(\Omega) = 1.$$

Now Ω is the largest volume in the σ- algebra. This ensures that a probability measure takes values between 0 and 1. We formalise this in the following definition.

Definition 3. *A function P is called a probability when, given a real valued random variable $X : \mathscr{F} \to \mathbb{R}$:*

$$P(X = x) = \frac{\mu(X^{-1}(x))}{\mu(\Omega)}.$$

Where $X^{-1} : \mathbb{R} \to \mathscr{F}$ denotes the inverse of X.

These measures we defined were all real valued. The definition of a measure can be naturally extended to include complex valued measures.

Definition 4. *Let Ω be a space and $\mathscr{F}$ be an algebra. Then a function $\mu : \mathscr{F} \to \mathbb{C}$ is called a complex measure if*

$A_1, A_2, \ldots$ is a sequence of disjoint sets in $\mathscr{F}$ such that $\cup_{i=1}^{\infty} A_i \in \mathscr{F}$ then,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

3.2 In Quantum Theory

Quantum mechanics is a theory based entirely upon probabilities. It is, therefore, natural that measure theory is an important part of the mathematical backbone of quantum mechanics. However, the abstract notions of measures must be connected to the realm of testable physical theories. The definitions below are an attempt in this direction.

First we would define a quantum mechanical experiment (Feintzeig, 2015).

Definition 5. *A simple quantum mechanical* experiment *is a triple $(\mathscr{H}, \psi, \mathscr{O}_n)$, where $\mathscr{H}$ is a Hilbert space, $\psi \in \mathscr{H}$ is a unit vector, and $\mathscr{O}_n = P_1, \ldots, P_n$ is a set of n projection operators on $\mathscr{H}$.*

ψ represents the quantum state in which the system is prepared. A probability p_i is assigned to the outcome of each projection operator in the following fashion:

$$p_i = \langle \psi | P_i | \psi \rangle .$$

This definition encompass our expectations from a quantum mechanical experiment. It can be easily generalised for density operators ρ. Then, for $\rho \in \mathscr{H}$,

$$p_i = \mathrm{Tr}\{P_i \rho\}.$$

We now attempt to connect quantum mechanics with measure theory with the definition of a *Classical Probability Space Representation* (Feintzeig, 2015).

Definition 6. *A quantum mechanical experiment* $(\mathscr{H}, \psi, \mathscr{O}_n)$ *is said to have a* Classical Probability Space Representation *(CPSR) iff there is a classical probability space* $(\Omega, \mathscr{F}, \mu)$ *and a map* $\zeta : \mathscr{O}_n \rightarrow \mathscr{F}$ *such that*

1. *For each* $P_i \in \mathscr{O}_n$,
$$\mu(\zeta(P_i)) = \langle \psi | P_i | \psi \rangle = p_i.$$

2. *For pairwise commuting* $\{P_i\} \in \mathscr{O}_n$,

$$\mu(\cap_i \zeta(P_i)) = \langle \psi | P_i P_j P_k ...... | \psi \rangle = p_{ijk...}.$$

3.3 Kirkwood-Dirac Distribution as a Complex Measure

In equation (2.10), A_m, B_n are PVMs (projection valued measure). Therefore to under $\mathscr{K}_{\Updownarrow\backslash}$, we employ a rigorous definition for a PVM. (See Chapter 9 of (Baggett, 1991) and Chapter 1 of (Folland, 1994))

Definition 7. *Let there be a measurable space* $(\Omega, \mathscr{F})$ *where* $\mathscr{F}$ *is the* σ*-algebra generated by* Ω*. Let there also be a Hilbert space* $\mathscr{H}$ *over* $\mathbb{C}$*. Then a PVM (projection valued measure) is a function*

$$\Pi : \mathscr{F} \rightarrow \mathscr{L}(\mathscr{H})$$

which satisfy the following:

1. *For all* $E_i \in \mathscr{F}$, $\Pi(E_i) = P_i$ *where* P_i *is an orthogonal projection operator.*

2. $\Pi(\Omega) = \mathbb{I}$ *and* $\Pi(\varnothing) = 0$.

3. *For all* $E_i, E_j \in \mathscr{F}$, $\Pi(E_i \cap E_j) = \Pi(E_i)\Pi(E_j)$.

4. *If* $\{E_1, E_2, ...\}$ *is a countable collection of pairwise disjoint elements in* $\mathscr{F}$ *then* $\{\Pi(E_1), \Pi(E_2), ...\}$ *is a pairwise orthogonal collection of projections and,*

$$\Pi\left(\bigcup_i E_i\right) = \sum_i \Pi(E_i).$$

Where $\mathscr{L}(\mathscr{H})$ is the space of all bounded linear operators on the Hilbert space $\mathscr{H}$.

Theorem 1. *A PVM naturally admits a CPSR.*

Proof. Consider the set of PVMs $\{\Pi_i\}$. From the definition above, for all $E_i \in \mathscr{F}$, $\Pi(E_i) = P_i$. Define a set $\mathscr{O}_n$ only containing all such $P_i \in \mathscr{O}_n$

Define the natural measure associated with $(\Omega, \mathscr{F})$, for $|\psi\rangle \in \mathscr{H}$:

$$\mu : E \mapsto \langle \Pi(E)\psi|\psi\rangle\,.$$

Our setup is now complete with both a quantum mechanical experiment $(\mathscr{H}, \psi, \mathscr{O}_n)$ and classical probability space $(\Omega, \mathscr{F}, \mu)$. We now look for the function $\zeta : \mathscr{O}_n \to \mathscr{F}$.

By definition of $\mathscr{O}_n$ and the definition of a PVM, the inverse functionΠ^{-1} exists. Identify $\zeta = \Pi^{-1}$.

Then, we can trivially show from the properties of a PVM, ζ satisfies:

1. For each $P_i \in \mathscr{O}_n$,

$$\mu(\zeta(P_i)) = \mu(\Pi^{-1}(P_i)) = \mu(E_i) = \langle\psi| P_i |\psi\rangle = p_i.$$

2. For pairwise commuting $\{P_i\} \in \mathscr{O}_n$,

$$\mu(\cap_i \zeta(P_i)) = \mu(\cap_i \Pi^{-1}(P_i)) = \mu(\cap_i E_i) = \langle\psi| P_i P_j P_k ......|\psi\rangle = p_{ijk....}$$

□

Now we have all the ingredients for our formulation. Let the quantum mechanical experiment be the measurement of the observable A. A particular outcome m is given by the the projection operator $A_m = |a_m\rangle\langle a_m|$. This operator is a PVM, and hence a map from the σ-algebra of the classical probability space representation.

Theorem 2. *Given a pair of states (normalised)* $|\phi\rangle, |\psi\rangle \in \mathscr{H}$, *a projection-valued measure always admits a complex measure,* S_Π, *which can be restricted to a probability measure,* $S_\Pi|\psi$.

Proof. Define a map: $S_\Pi : \mathscr{F} \to \mathbb{C}$.

$$S_\Pi : E \mapsto \langle \Pi(E)\psi|\phi\rangle .$$

It can be easily verified that that that it follows σ-additivity for all countable unions of events. Thus S_Π is a complex measure.

Consider the special case:

$$S_\Pi|_\psi : E \mapsto \langle \Pi(E)\psi|\psi\rangle .$$

Then

1. $S_\Pi|_\psi(\Omega) = \langle \Pi(\Omega)\psi|\psi\rangle = \langle \mathbb{I}\psi|\psi\rangle = 1,$
2. $S_\Pi|_\psi(\varnothing) = \langle \Pi(\varnothing)\psi|\psi\rangle = 0.$

Hence, it restricts to a probability measure.

□

Using this, We can, then define the KD distribution as a complex measure.

Theorem 3. *Consider the PVMs.*

$$\Pi_A(E_m) = A_m = |a_m\rangle\langle a_m| \text{ and } \Pi_B(E_n) = B_n = |b_n\rangle\langle b_n|$$

Then the function

$$K_\Pi(E_m, E_n) = \mathrm{Tr}\{\rho\Pi_A(E_m)\Pi_B(E_n)\} \tag{3.1}$$

is a normalised complex measure in both arguments.

Proof. By definition, K_Π is a complex valued function. Hence the first condition is satisfied.

As Π_A, Π_B are PVMs, we can prove that for any countable union of disjoint events,

$$K_\Pi\left(\bigsqcup_k E_k, \bigsqcup_l E_l\right) = \sum_k\sum_l K_\Pi(E_k, E_l)$$

This proves the second condition.

As for the normalisation, we already saw in (2.4), that the distribution is normalised.

□

Thus defined, the KD distribution is realised as a complex measure *in each argument*. Therefore, we see that the function is a normalised complex measure producing Kolmogorovian probability functions as marginals. This allows us to regard it as a joint probability distribution.

3.4 A Conditional Probability

The weak value, as shown in section (2.3.2), is given by the KD distribution divided by the probability of post-selection. However, this leads to an interesting observation. We recall Baye's rule from probability theory,

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}.$$

X,Y are random variables acting on the σ algebra $\mathscr{F}$. Then, we can draw a direct comparison to the previously defined quantity:

$$\mathscr{K}_{mn} = \frac{\mathrm{Tr}\{\rho A_m B_n\}}{\mathrm{Tr}\{\rho B_n\}}.$$

$\mathscr{K}_{mn}$ is therefore a conditional probability. However, its a function of the KD distribution, and hence is a complex distribution! Therefore, we call it a conditional quasi probability distribution. This analysis now brings us the the final chapter of this part, where we quantify the informational content of $\mathscr{K}_{mn}$.

Chapter 4

A Conditional Entropy

4.1 Motivation from Information Theory

Given a conditional probability function, one can follow classical information theory to define a conditional entropy (Cover, 2006). For a random variable X conditioned on selection of a particular instance y of random variable Y, the Shannon entropy is defined as:

$$H(X|Y=y) = -\sum_x Pr(x|y)\log Pr(x|y). \tag{4.1}$$

where $Pr(x|y)$ is the conditional probability of x given the occurrence of y.

However, y occurs probabilistically, hence the total entropy should be the expectation of the above quantity, and is hence defined as:

$$H(X|Y) = \sum_y Pr(y)H(X|Y=y). \tag{4.2}$$

We wish to proceed in a similar manner for the quantum quasi-probability defined above. However, first, we discuss a previous work on the subject.

4.2 Salek's Approach

In a recent work (Salek et al., 2014), the following was defined as a conditional state:

$$\rho_{\psi|\phi} = \frac{|\phi\rangle\langle\psi|}{\langle\psi|\phi\rangle}.$$

The state can be retrieved by multiplying $\rho_{\psi|\phi}$ by probability $Pr(\phi)$ of the system ending up in the state $|\phi\rangle$ and summing over a complete basis $\{\phi\}$.

$$\rho_s = \sum_\phi P(\phi)\frac{|\phi\rangle\langle\psi|}{\langle\psi|\phi\rangle} \tag{4.3}$$

They then, propose the following definition for a conditional entropy:

$$S_c(\rho_{\psi|\phi}) = -\frac{1}{2}\mathrm{Tr}\Big\{\rho_{\psi|\phi}\log\left(\rho_{\psi|\phi}\rho^{\dagger}_{\psi|\phi}\right)\Big\}. \tag{4.4}$$

They provide the following justification. The von Neumann entropy is an expectation of a logarithm,

$$S = \langle -\log\rho\rangle\,;$$

as is the proposed definition,

$$S_c(\rho_{\psi|\phi}) = \left\langle -\log\left(\rho_{\psi|\phi}\rho^{\dagger}_{\psi|\phi}\right)^{1/2}\right\rangle.$$

Where $P(\phi) = \mathrm{Tr}\{\rho\,|\phi\rangle\langle\phi|\}$. The TSV basis like quantities are substituted in the formula for entropy; then the associated probability is multiplied ($P(\phi)$) to calculate the final entropy of the state:

$$S_C(\rho_s) = \sum_n P(\phi)S_c(\rho_{\psi|\phi}). \tag{4.5}$$

Further computation reveals,

$$S_C(\rho_s) = \sum_n |\langle b_n|\psi\rangle|^2\log|\langle b_n|\psi\rangle|. \tag{4.6}$$

when $\rho_s = |\psi\rangle\langle\psi|$.

This formulation has a couple of glaring limitations. Firstly, as the authors mention, this is a special case of pre selection in a pure state. In general, we expect an ideal process to work for any state ρ. Therefore a generalised formula is needed. Secondly, it lacks sufficient justification as to why $\rho_{n|m}$ must be considered as a conditional state. In the next few pages, we attempt to provide a better formulation.

4.3 A Complex Conditional Quasi-state

We come back to the natural question we asked in 4.1. Is a similar analysis is possible for quantum states conditioned on particular post-selections? We wish to proceed the same way as Shannon, to define conditional entropy. As we are working in the quantum realm, to define quantum conditional entropy, we must develop a notion of a quantum conditional state.

A quantum state can be written down with a single set of basis vectors in $\mathscr{H}$. We can choose a basis in which the state is diagonal. Then, the coefficients of the basis are a probability distribution. In case of the conditional probability distribution $\mathscr{K}_{mn}$, we require *two* sets of basis vectors for a complete description. Fortunately, the well developed two-state formalism comes to our rescue. (Appendix A)

We thus assume the TSV basis with coefficients given by $\mathscr{K}_{mn}$. This allows us to define our conditional quantum quasi-state:

$$\rho_{A|n} = \sum_m \mathscr{K}_{mn} \frac{|b_n\rangle \langle a_m|}{\langle a_m|b_n\rangle}. \tag{4.7}$$

Then the original state ρ can be retrieved as

$$\rho = \sum_n \mathrm{Tr}\{\rho B_n\}\rho_{A|n}.$$

Foremost, we require a logarithmic expression for calculation of entropy. The standard von Neumann formula will lead to complex numbers in the logarithm; which would in turn lead to complex values of entropy itself. This is not ideal as it is not immediately clear how to compare complex numbers.

The most trivial modification to the formula is using a singular value decomposition to get rid of complex numbers inside the logarithm. This modification leads us the definition of entropy proposed by Salek.

$$S_c(\rho_{A|n}) = -\frac{1}{2}\mathrm{Tr}\Big\{\rho_{A|n}\log\Big(\rho_{A|n}\rho_{A|n}^{\dagger}\Big)\Big\}. \tag{4.8}$$

And the total entropy as

$$S_C(\rho) = \sum_n P(b_n) S_c(\rho_{A|n}).$$

$$\rho_{A|n} = \sum_m \mathscr{K}_{mn} \frac{|b_n\rangle \langle a_m|}{\langle a_m|b_n\rangle} = \sum_m \frac{\langle b_n|\rho|a_m\rangle}{\langle b_n|\rho|b_n\rangle} \frac{|b_n\rangle \langle a_m|}{\langle a_m|b_n\rangle} \tag{4.9}$$

Then the original state ρ can be retrieved as

$$\rho = \sum_n \mathrm{Tr}\{\rho B_n\}\rho_{A|n}.$$

In the above equation, all the complex parts are contained in $\rho_{A|n}$.

4.3.1 Conditioning a quantum entropy

Given the such a conditioned KD distribution, we would like to determine the entropy, or the information content, of the distribution. Intuitively, we would expect more information in a conditional scenario; in analogy with conditional probability where the odds of making the correct prediction with the increase in available information.

From (4.8), entropy is given by,

$$S_c(\rho_{A|n}) = -\frac{1}{2}\operatorname{Tr}\left\{\rho_{A|n}\log\left(\rho_{A|n}\rho^\dagger_{A|n}\right)\right\}.$$

After a some calculations (Appendix B.1), we end up with the expression:

$$S_c(\rho_{A|n}) = \log(\operatorname{Tr}[\sigma_n]) - \frac{1}{2\operatorname{Tr}[\sigma_n]}\operatorname{Tr}\left[\sigma_n\log\left(\sigma^\dagger_n\sigma_n\right)\right]; \tag{4.10}$$

where $\sigma_n = |b_n\rangle\langle b_n|\rho$.

Finally, substituting $S_c(\rho_{A|n})$ in (4.5),

$$S_C(\rho) = \sum_n\left[\operatorname{Tr}[\sigma_n]\log(\operatorname{Tr}[\sigma_n]) - \frac{1}{2}\operatorname{Tr}\left[\sigma_n\log\left(\sigma^\dagger_n\sigma_n\right)\right]\right]. \tag{4.11}$$

4.3.2 Interesting features and discussion

If ρ is pure, $\rho = |\psi\rangle\langle\psi|$ and $b_n = |b_n\rangle\langle b_n|$. We see that the expression reduces to the formula (4.6) found in the previous section:

$$S_C(\rho_s) = S_C(\psi|B) = \sum_n |\langle b_n|\psi\rangle|^2\log|\langle b_n|\psi\rangle|.$$

If ρ is diagonal in B basis, $\rho = \sum_n\lambda_n|b_n\rangle\langle b_n|$. Then $\operatorname{Tr}[\sigma_n] = \lambda_n$.

Then, (4.11) gives,

$$S_C(\rho) = \sum_n[\lambda_n\log(\lambda_n) - \lambda_n\log(\lambda_n)] = 0. \tag{4.12}$$

This actually is a remarkable observation. We know that the von Neumann entropy is basis independent. This function is crucially not. Moreover, S_C is maximum $(= 0)$ when the state is diagonal in the reference basis, implying that the information quantified by S_C is lost upon diagonalization. Thus, S_C quantifies the information about the *superposition* in a quantum state.

Consider an initial state in the post-selection basis,

$$\rho = \sum_{i,j}\rho_{ij}\,|b_i\rangle\langle b_j|\,.$$

Then, $\operatorname{Tr}\{\sigma_n\} := \rho_{nn}$, which gives,

$$S_C(\rho) = \sum_n\rho_{nn}\log\left[\frac{\rho_{nn}}{\sqrt{(\rho^2)_{nn}}}\right] = \frac{1}{2}\sum_n\rho_{nn}\log\left[\frac{\rho^2_{nn}}{(\rho^2)_{nn}}\right]. \tag{4.13}$$

We will now prove an important theorem.

Theorem 4. *For all quantum states $\rho \in \mathscr{H}$, $S_C(\rho) \leq 0$; and $S_C = 0$ iff the state is diagonal*

Proof. $\rho^2 = \sum_{i,j,k} \rho_{ij}\rho^*_{jk} |b_i\rangle\langle b_k|$. This allows one to show,

$$\log\left[\frac{\rho_{nn}^2}{(\rho^2)_{nn}}\right] = \log\left[\frac{\rho_{nn}^2}{\sum_i |\rho_{ni}|^2}\right]$$

Of course, $\sum_i |\rho_{ni}|^2 \geq \rho_{nn}^2$ and hence, $S_C(\rho) \leq 0$.
For a quantum state, ρ_{nn} cannot be zero for all n. Hence, $S_C(\rho) = 0$ iff $\log\left[\frac{\rho_{nn}^2}{(\rho^2)_{nn}}\right] = 0$ which is true iff $\sum_i |\rho_{ni}|^2 = \rho_{nn}^2$ implying the state is diagonal.

□

This function can thus naturally detect the presence of off-diagonal terms, which is a property of a coherence measure. Moreover, if S_C does quantify the basis dependant information about superposition, it should be able to quantify *coherence*. These facts thus calls for an analysis through the lens of quantum coherence which we explore in detail throughout the next part.

Part II

Towards a Measure of Coherence

Chapter 5

Quantifying Superposition - Quantum Coherence

5.1 Introduction

Simply put, coherence is the amount of superposition in a quantum state. Superposition is the heart of quantum mechanics. Quantum states are much more than a classical probability distribution. As discussed earlier, the diagonal elements of a density matrix form a valid probability distribution, but a general quantum state is not diagonal. The off-diagonal elements, therefore, contain information of superposition or *quantumness* (Aberg, 2006). Coherence is exactly the quantifier of this quantumness. One of the simplest and well-known coherence measures is the l_1-norm,

$$C_{l_1}(\rho) = \sum_{j \neq k} \rho_{jk}. \tag{5.1}$$

Where ρ_{jk} are the elements of ρ in the chosen basis.

Coherence is, of course, a basis dependent concept, which is why we first need to fix a reference basis. Generally, the physics of the problem dictates the reference basis under investigation. Alternatively, it is given by a task for which coherence is required (Streltsov et al., 2017).

5.2 Incoherent Operations

Given a d-dimensional Hilbert space $\mathscr{H}$ with an orthonormal reference basis set $\{|i\rangle\}$, incoherent states are defined as diagonal matrices,

$$\sigma = \sum_i p_i |i\rangle\langle i|. \tag{5.2}$$

A CPTP map is said to be an incoherent operation if it can be written as,

$$\Lambda[\rho] = \sum_m K_m \rho K_m^\dagger \quad ; \sum_m K_m K_m^\dagger = \mathbb{I}, \tag{5.3}$$

and for the set I of all incoherent states for any dimension d,

$$K_m \rho K_m^\dagger \subseteq I \ \forall\, m, \quad \rho \in I.$$

Of course, $I \subset \mathscr{B}(\mathscr{H})$, where $\mathscr{B}(\mathscr{H})$ denotes the set of all bounded trace class operators on $\mathscr{H}$. We need to ensure that above is true for all m, because it is important that coherence is not generated for observers who have access to individual outcomes m.

Incoherent operations can be broadly classified as follows (Baumgratz et al., 2014).

1. Non-selective - These operations are given by

   $$\Lambda[\rho] = \sum_m K_m \rho K^\dagger m \quad ; \sum_m K_m K_m^\dagger = \mathbb{I}$$

   where all K_m have the same dimension ($d_{OUT} \times d_{IN}$). These are called Maximally Incoherent operations (MIO).

2. Selective - These are given by

   $$\sigma_n = \frac{K_n \rho K_n^\dagger}{\mathrm{Tr}\left[K_n \rho K_n^\dagger\right]} \in I$$

   Here each K_n can have different dimensions ($d_n \times d_{IN}$). These are called Incoherent operations (IO)

3. Extra conditions - Many conditions can be imposed to restrict the class of incoherence operations. We study a particular class called Strictly Incoherent Operations (SIO).

We define the dephasing operation

$$\Delta[\rho] = \sum_{i=0}^{d-1} |i\rangle\langle i| \, \rho \, |i\rangle\langle i| \tag{5.4}$$

An SIO will not use any information of coherence from the input state.
So, if such an operation is represented by Kraus operators $\{K_m\}$ then,

$$\mathrm{Tr}\left[P_i K_m \rho K_m^\dagger\right] = \mathrm{Tr}\left[P_i K_m \Delta[\rho] K_m^\dagger\right]$$

where $P_i = |i\rangle\langle i|$. This ensures that the outcomes of a measurement in the reference basis P_i applied to the output σ_m are independent of the coherence of ρ.
SIOs have the following form

$$K_m = \sum_i c_i^m \, |\pi_m(i)\rangle\langle i| \, . \tag{5.5}$$

Here $\pi_\lambda \in S_n$ is a permutation. S_n is the permutation group of n dimensions. Formally,

$$\pi_\lambda : \{|i\rangle\}_{d_\pi} \rightarrow \{|i\rangle\}_{d_\pi}.$$

SIOs have the property that $K_m^\dagger$ are also IOs. This means that an incoherent operation is also strictly incoherent iff π is injective (Appendix C.1). Define:

$$|\Psi_j\rangle = \frac{1}{\sqrt{j}} \sum_{i=1}^{d} |i\rangle$$

This is a d-dimensional maximally coherent state. This state allows for the deterministic generation of all $d \times d$ dimensional density matrices using free operations of incoherence (Baumgratz et al., 2014).

5.3 Resource Theory

A resource theory is determined by some particular constraints. One defines a set of free states I; the classes of operation F are considered s.t. $f \in F \;\; ; f : I \rightarrow I$. Any state outside I are considered a resource and has a cost. Another direction is to use free operations on a maximally mixed state (devised of any useful resource) to generate a set I.

Resource theories are extremely important for various reason. If coherence is a resource, then there must be ways to measure coherence. We discuss this in the following section.

5.4 Measuring Coherence

A coherence measure C is expected to follow the axioms listed below (Streltsov et al., 2017).

A1 Non-negativity: It is expected of a measure to only output non-negative number, so that 0 (the smallest output) is reserved for states without any coherence.

$$C(\rho) \geq 0 \text{ and } C(\rho) = 0 \text{ iff } \rho \in I \tag{5.6}$$

A2 Monotonicity: It is expected that C does not increase under the action of incoherent operations, as such operations cannot create coherence.

$$C(\rho) \geq C(\Lambda[\rho]) \text{ with } \Lambda[I] \subseteq I \tag{5.7}$$

A3 Strong monotonicity: Making the last property stronger, it is also expected that C does not increase on average under selective incoherent operations.

$$C(\rho) \geq \sum_{\lambda} q_{\lambda} C(\rho_{\lambda}); \quad \rho_{\lambda} = \frac{1}{q_{\lambda}} K_{\lambda} \rho K_{\lambda}^{\dagger} \tag{5.8}$$

for incoherent Kraus operators K_{λ}, and $q_{\lambda} = \mathrm{Tr}\left[K_n \rho K_n^{\dagger}\right]$

A4 Convexity: Finally, it is expected that coherence should decrease upon mixing.

$$C(\sum_{i} p_i \rho_j) \leq \sum_{i} p_i C(\rho_j) \tag{5.9}$$

Conditions (A1) and (A2) are minimal requirements for C to be a meaningful resource quantifier for any coherence-based task. Condition (A3) says coherence should not increase under incoherent measurements even if one has access to the individual measurement outcomes. If a measure follows (A3) and (A4), it also follows (A2). For the purpose of this thesis, we will call a quantity C that fulfills conditions (A1), (A2), and (A4) a coherence monotone. A quantity C will be further called a coherence measure if it fulfills (A1 - A4).

Alternative or additional desirable requirements for coherence measures may be proposed, but for the present work, this definition is sufficient. With the knowledge of the resource theory of coherence, we return to the matter at hand. Is the theory of coherence linked to the KD distribution? In the following chapter, we explore some facts which try to further motivate this belief.

Chapter 6

From Quasi-probability to Quantum Coherence

In this chapter, we review two different works on the subject of visibility of interference. We will show how these serve as further motivation to explore the link between the Kirkwood-Dirac distribution and quantum coherence.

6.1 An Experiment of Indefinite Causal Order

In a recent work (Ban, 2021), the following scheme of sequential measurements performed in indefinite causal order was proposed.

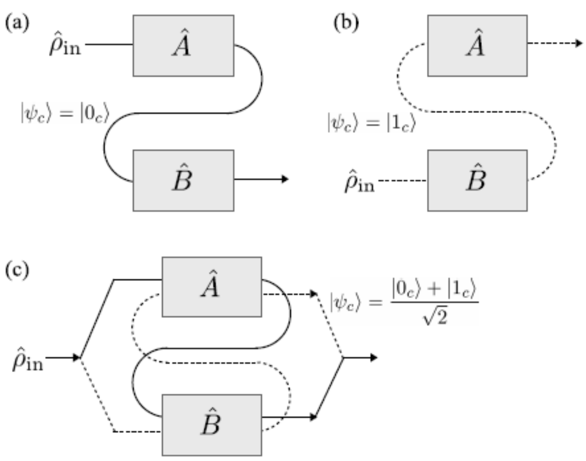

Figure 6.1: Schematic of measurements in different causal orders.

Two physical quantities, $\hat{A}, \hat{B}$ of a quantum system S are measured sequentially. (a) When the control system is in a computational state $|0\rangle_c$, $\hat{A}$ is measured first. (b) When C is in the other computational state $|1\rangle_c$, they are measured in the reverse order. (c) When S is in the superposition state $|\Psi_c\rangle = (|0\rangle_c + |1\rangle_c)/\sqrt{2}$, the two different orders of the sequential measurements are superposed. In (a) and (b), the two measurements are performed in the definite causal order while in (c), they are carried out in the indefinite causal order.

The interferometric experimental setup is shown below. Here ϕ is a measurement on the control qubit and the detectors and beam splitters are labelled by D and BS respectively.

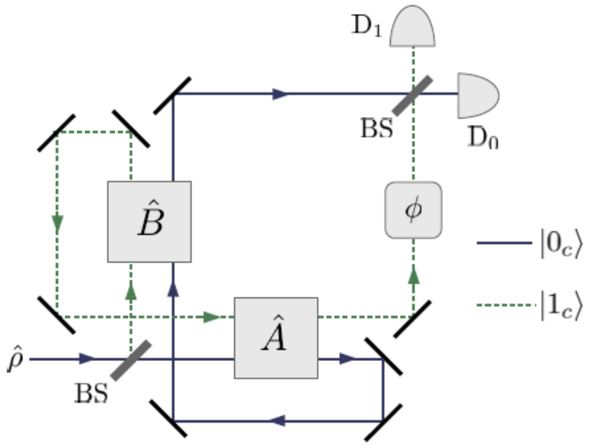

Figure 6.2: The interferometric setup.

The final interference pattern was analysed. To calculate the pattern, the following were used.

Input state:

$$\rho_{in} = \rho \otimes |\phi_A\rangle \langle\phi_A| \otimes |\phi_B\rangle \langle\phi_B| \otimes |0\rangle \langle 0|_C$$

Control states:

$$|\psi_\pm(\phi)\rangle = \frac{1}{\sqrt{2}}|0_C\rangle \pm e^{i\phi}\frac{1}{\sqrt{2}}|1_C\rangle$$

Beam splitters:

$$\mathscr{B}_\psi = |\psi_+(0)\rangle\langle 0|_C + |\psi_-(0)\rangle\langle 1|_C$$

The measurement on C:

$$\Pi_\pm(\phi) = |\psi_\pm(\phi)\rangle\langle\psi_\pm(\phi)|$$

The evolution of the system is given by:

$$\mathscr{U}_{SABC} = \mathscr{U}_{SB}\mathscr{U}_{SA} \otimes |0\rangle\langle 0|_C + \mathscr{U}_\phi \mathscr{U}_{SA}\mathscr{U}_{SB} \otimes |1\rangle\langle 1|_C$$

Where $\mathscr{U}_\phi$ is a measurement by on the control system by the device ϕ and $\mathscr{U}_{SA}$ is the unitary which denotes measurement A being performed on the system. Also, $\langle\phi_A(a_m)|\,\mathscr{U}_{SA}\,|\phi_A(a_m)\rangle = |a_m\rangle\langle a_m|$, as we will be considering projective measurements. Similarly, one may define

$\mathscr{U}_{SB}$.
Finally, the output state is:

$$\rho_{out} = \mathscr{B}_\psi \mathscr{U}_{SABC} \mathscr{B}_\psi \rho_{in} \mathscr{B}^\dagger{}_\psi \mathscr{U}^\dagger_{SABC} \mathscr{B}^\dagger_\psi$$

Now, there are two cases.

Case 1: The control system is untouched. This means that no measurement is performed by the device ψ and $\mathscr{U}_\phi = \mathbb{I}$

$$P(m,n) = \langle \phi_A(a_m)| \langle \phi_B(b_n)| \mathrm{Tr}\{\rho_{out}\} |\phi_B(b_n)\rangle |\phi_A(a_m)\rangle = \frac{1}{2} |\langle a_m|b_n\rangle|^2 \left(\langle \rho \rangle_{a_m} + \langle \rho \rangle_{b_n}\right) \tag{6.1}$$

This is similar to $P(a|b) = 1/2[P(b|a)P(a) + P(a|b)P(b)]$; which is just a statistical mixture of measurement orders with equal probabilities.

Case 2: A measurement is performed on the control system, and the probability is calculated like above. $\mathscr{U}_\phi = |0\rangle\langle 0|_C + e^{i\phi} |1\rangle\langle 1|_C$.

$$P_\phi(i,m,n) = \frac{1}{4} |\langle a_m|b_n\rangle|^2 \left(\langle \rho \rangle_{a_m} + \langle \rho \rangle_{b_n} + 2(-1)^i Re[e^{i\phi} P(a_m,b_n)]\right) \tag{6.2}$$

In this case, the situation is more complicated. $P_\phi(i,m,n)$ is the joint probability that the outcomes (a_m,b_n) are obtained and the system is found on the path $|i\rangle$ by the detector D_i. The interference pattern, however, provides us the probability that given the outcomes (a_m,b_n) are measured, the system is found on the path $|i\rangle$ by the detector D_i.

This is of course a conditional probability which can be calculated simply:

$$P_\phi(i|m,n) = \frac{1}{2}\left[1 + (-1)^i V_{mn} cos(\phi + \Phi_{mn})\right] \tag{6.3}$$

Where

$$V_{mn} = \frac{2|\mathscr{K}(a_m,b_n)|}{\langle b_n|\rho|b_n\rangle + \langle a_m|\rho|a_m\rangle}, \quad \Phi_{mn} = \arg(\mathscr{K}(a_m,b_n))$$

For any interference experiment, the visibility is given by:

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

Clearly, V_{mn} is a visibility functional, which in this case is conditional. Therefore, it is shows that the KD distribution appears while measuring the conditional visibility in indefinite causal order experiments.

6.2 Conditional Visibility and Coherence

In another article (Biswas et al., 2017), the authors considered the following simple Mach-Zehnder interferometer.

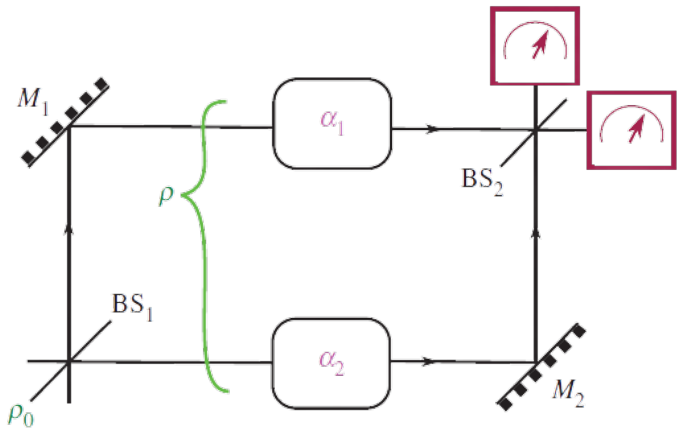

Figure 6.3: The Mach-Zehnder interferometric setup.

Consider a multi-path interferometer spanning a d-dimensional Hilbert space $\mathscr{H}$ (ignoring all other internal or spatial degrees of freedom). A pure state inside the interferometer is given by

$$|\psi\rangle = \sum_j c_j |j\rangle ,$$

whereas a mixed state is given by

$$\rho = \sum_{j,k=1}^{d} \rho_{jk} |j\rangle\langle k| .$$

There are two components of the interferometric experiment - the first is the local phase shifts α_j introduced in the paths of the particle which is given by the following unitary

$$U(\alpha) = \sum_j e^{i\alpha_j} |j\rangle\langle j| ,$$

and evolves the state in the following manner

$$\rho(\alpha) = U(\alpha)\rho U(\alpha)^\dagger = \sum_{j,k=1} d e^{i(\alpha_j - \alpha_k)} \rho_{jk} |j\rangle\langle k| .$$

The second is the detector at the output which mathematically is a POVM $M = (M_\omega)$ with outcomes ω from a suitable space Ω.

After having chosen $\alpha = (\alpha_1,..\alpha_d)$ the experimenter will observe outcomes $\omega \in \Omega$ sampled from the Born distribution:

$$P_{M|\rho}(\omega|\alpha) = \mathrm{Tr}\, U(\alpha)\rho U(\alpha)^{\dagger} M_{\omega}.$$

In such an experiment, given a fixed ρ, the distribution $P = P_{M|\rho}$ can vary as a function of the local phases. This degree of variability is the visibility of the interference pattern, thus we define a visibility functional $V = V[P]$ on the conditional distributionsn $P(\omega|\alpha)$.

For the simplest case of the known Mach-Zehnder interferometer in two dimensions, we observe the interference with respect to a relative phase shift. Consider,

$$\rho = \rho_{11} |1\rangle\langle 1| + \rho_{22} |2\rangle\langle 2| + \rho_{12} |1\rangle\langle 2| + \rho_{21} |2\rangle\langle 1|,$$

with the diagonal phase unitary,

$$U(\alpha) = e^{i\alpha_1} |1\rangle\langle 1| + e^{i\alpha_2} |2\rangle\langle 2|$$

and a measurement with POVM elements,

$$|\mu\rangle\langle\mu|, |\mu\rangle = \mu_1 |1\rangle + \mu_2 |2\rangle.$$

Now, with $\alpha = \alpha_1 - \alpha_2$ and writing $\rho_{12} = |\rho_{12}|e^{i\beta}, \bar{\mu}_1\mu_2 = |\mu_1\mu_2|e^{i\gamma}$, the output probability is,

$$P_{M|\rho}(\mu|\alpha) = \rho_{11}|\mu_1|^2 + \rho_{22}|\mu_2|^2 + 2|\rho_{12}\mu_1\mu_2|\cos(\alpha+\beta+\gamma). \tag{6.4}$$

The fluctuation in the above quantity is governed by the coefficient $|\rho_{12}\mu_1\mu_2|$ which we can thus conclude to be the visibility.

We wish, of course, to treat the state ρ as a resource, i.e. as a given, of which we are supposed to make the best. It then, makes sense to optimize the visibility $V[P_{M|\rho}]$ over all possible measurements. However, it is clear that the most beautiful coherent superposition in the state may be lost by an unsuitable choice of measurements.

The authors discuss the following example. Consider the state:

$$\rho = 1/3(|0\rangle + |1\rangle)(\langle 0| + \langle 1|) + 1/3\, |2\rangle\langle 2|.$$

With measurement $M^{(0)}$ in the basis,

$$\left\{ |0\rangle, \frac{1}{\sqrt{2}}(|1\rangle \pm |2\rangle) \right\}.$$

Then, probabilities come out to be $1/3$ for all outcomes irrespective of the phases.
This leads to all coherence in

$$\rho(\alpha) = 1/3(|0\rangle + e^{i\alpha}|1\rangle)(\langle 0| + e^{-i\alpha}\langle 1|) + 1/3|2\rangle\langle 2|,$$

to be lost.

Intuitively, we expect the best choice to bring out the coherence in ρ to be the projective measurement $M^{(1)}$ in the basis ,

$$\left\{ \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle), |2\rangle \right\},$$

for which the detection probabilities are $(\frac{2}{3}cos^2(\alpha/2), \frac{2}{3}, sin^2(\alpha/2), \frac{1}{3})$.

Thus motivated, given a visibility functional $V[P]$, they argue to optimize the visibility over all measurements, to get the best out of ρ. Define:

$$C_V(\rho) = \sup_M V[P] \tag{6.5}$$

This leads to the theorem for optimisation of visibility functionals.

Definition 8. *For a probability distribution $\{q_i\}$, $\bar{P} = \sum_i q_i P_i(\mu|\alpha)$ implies $V[\bar{P}] = \sum_i q_i V[P_i(\mu|\alpha)]$ for each $\mu \in M_i$. Then, such a visibility functional is called* weakly affine.

Theorem 5. *For any regular and weakly affine visibility functional $V[P]$, C_V is a coherence measure that is strongly monotonic under strictly incoherent operations (SIOs). If V is convex in P, then C_V is convex in ρ.*

Proof. See (Biswas et al., 2017)

□

6.3 The Link

This analysis then boils down to a simple fact, which is captured by the following schematic.

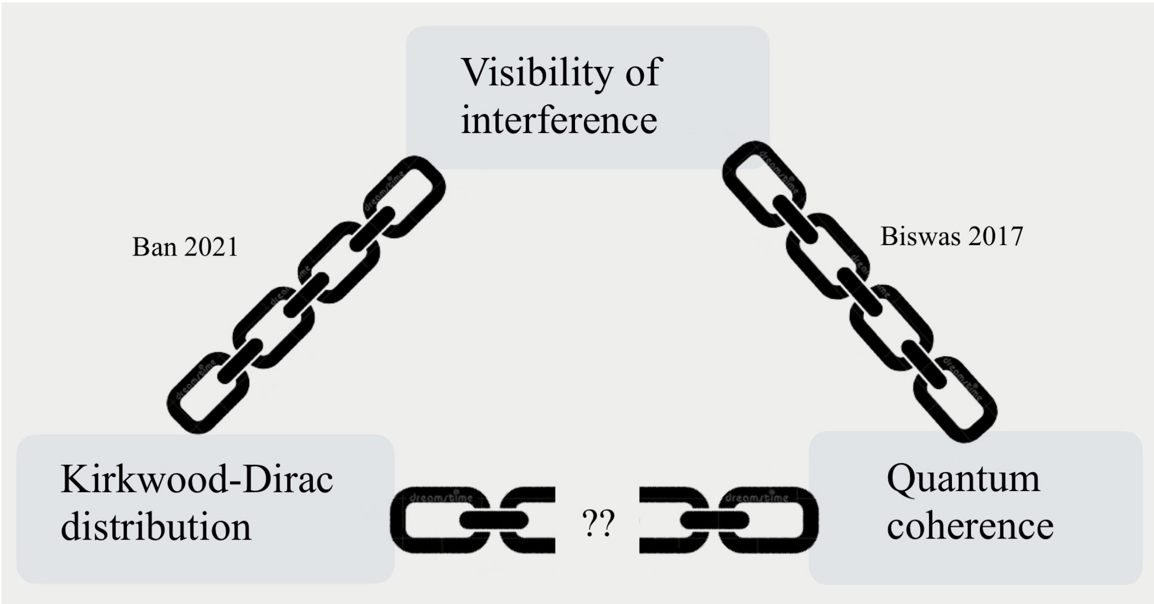

Figure 6.4: The missing link.

We see that the visibility of interference is linked to the Kirkwood-Dirac distribution, and the visibility is in turn linked to quantum coherence. Is it then possible to find a direct link between the distribution and coherence? We explore this question in the next chapter.

Chapter 7

Measuring Coherence

7.1 The Non-classicality in the Kirkwood-Dirac Distribution

As seen in the previous chapter, we have sufficient reasons to believe that the Kirwood-Dirac distribution contains information about the *quantumness* or the coherence in a quantum state. However, to prove this assertion, we should be able to derive a coherence measure from the distribution. It is logical to expect, then, that this information about coherence is hidden in the non-classical part of the distribution. But which is the non-classical part? In the following sections, we attempt to construct an answer to this question.

7.1.1 Measurements and the Wigner formula

In quantum mechanics, projective measurements may be selective or non-selective. Again consider observables A and B. For a non selective measurement on ρ, we write,

$$\rho' = A_m \rho A_m + (1 - A_m)\rho(1 - A_m). \tag{7.1}$$

Where $1 - A_m$ is the orthogonal complement to A_m, simply shown by

$$A_m(1 - A_m) = (1 - A_m)A_m = 0.$$

Similarly, a selective measurement gives,

$$\rho_s' = \frac{A_m \rho A_m}{\mathrm{Tr}\{\rho A_m\}}.$$

We now assume that a projective measurement of B is made on the system after a selective measurement of A_m. The probability of the measurement outcome B_n on the selective state ρ_s' is then,

$$P(b_n|a_m) = \mathrm{Tr}\,\rho_s' B_n = \frac{\mathrm{Tr}\,\rho A_m B_n A_m}{\mathrm{Tr}\,\rho A_m}.$$

Of course, this is not a joint probability in the classical sense, as reversing the order of A, B does not lead to the same probability. The numerator is known as the Wigner formula. The marginal probabilities obtained from this joint probability are asymmetric, as shown below:

$$\sum_n \mathrm{Tr}\,\rho A_m B_n A_m = \mathrm{Tr}\,\rho A_m,$$

$$\sum_m \mathrm{Tr}\,\rho A_m B_n A_m = \mathrm{Tr}\,\rho' B_n.$$

7.1.2 Breaking up the Kirkwood-Dirac distribution

The complex conjugate of the distribution is given by,

$$(\mathrm{Tr}\,\rho A_m B_n)^* = \mathrm{Tr}\,\rho B_n A_m. \tag{7.2}$$

This lead us to define:

$$Re(\mathrm{Tr}\,\rho A_m B_n) = \mathrm{Tr}\,\rho\{A_m, B_n\}.$$

Where $\{,\}$ denotes the anti-commutator. This however leads to an interesting observation. Consider the expression,

$$\mathrm{Tr}\,\rho A_m B_n A_m + \frac{1}{2}\mathrm{Tr}\big(\rho - \rho'\big)B_n = \mathrm{Tr}\,\rho A_m B_n A_m + \frac{1}{2}\mathrm{Tr}[\rho B_n - A_m \rho A_m B_n + (1 - A_m)\rho(1 - A_m)B_n]$$

$$= \mathrm{Tr}\,\rho B_n A_m + \mathrm{Tr}\,\rho A_m B_n = \mathrm{Tr}\,\rho\{A_m, B_n\}.$$

Here, ρ' is from (7.1). Hence the real part of the KD distribution can be expressed as (Johansen, 2007):

$$Re(\mathrm{Tr}\,\rho A_m B_n) = \mathrm{Tr}\,\rho A_m B_n A_m + \frac{1}{2}\mathrm{Tr}\big(\rho - \rho'\big)B_n. \tag{7.3}$$

The real part is therefore given by the Wigner formula and a modification term. The real part of the distribution gives the following marginals:

$$\sum_n Re(\mathrm{Tr}\,\rho A_m B_n A_m) = \mathrm{Tr}\,\rho A_m,$$

$$\sum_m Re(\mathrm{Tr}\,\rho A_m B_n A_m) = \mathrm{Tr}\,\rho B_n.$$

It appears that the the marginals and hence the property of the function being a joint probability is only contained in the real part. However, the real part does not determine the density matrix uniquely. Therefore, the imaginary part of the KD distribution contains crucial information about a quantum state. We now examine the imaginary part.
It is shown in Appendix B.3 that the imaginary part evaluates to:

$$Im(\mathrm{Tr}\,\rho A_m B_n) = \frac{1}{2}\mathrm{Tr}\big(\rho - \rho'\big)B_n^{\pi/2}. \tag{7.4}$$

Here $B_n^{\pi/2}$ is the selectively phase rotated projector (Appendix B.3). The marginals of the imaginary part, of course, do not contribute,

$$\sum_m \mathrm{Im}(\mathrm{Tr}\,\rho A_m B_n) = 0,$$
$$\sum_n \mathrm{Im}(\mathrm{Tr}\,\rho A_m B_n) = 0.$$

7.1.3 The non-classicality

Where should we expect the non-classicality to be hidden in the KD distribution? The Wigner formula is a non negative probability distribution. It describes the probability of successive measurement and is thus a classical probability. Further, it can be trivially seen that for the classical scenario of commuting observables A_m, B_n, the KD distribution is completely given by the Wigner term.

We suspect, then, that removing the Wigner formula from the KD distribution should give us a function which holds the information about the non-classicality in ρ:

$$\mathrm{Tr}\,\rho A_m B_n - \mathrm{Tr}\,\rho A_m B_n A_m = \frac{1}{2}\mathrm{Tr}\big(\rho - \rho'\big)B_n + i\frac{1}{2}\mathrm{Tr}\big(\rho - \rho'\big)B_n^{\pi/2}.$$

It can be shown, again trivially, that this expression takes a simpler form:

$$\mathrm{Tr}\,\rho A_m B_n - \mathrm{Tr}\,\rho A_m B_n A_m = \mathrm{Tr}\Big\{A_m^{\perp}\rho A_m B_n\Big\}. \tag{7.5}$$

Here $A_m^{\perp} = 1 - A_m$.

7.2 Search for Coherence in the Non-classical Part

Consider, independently, the non-classical part of the KD distribution, as a function of given destiny matrix ρ and a reference basis (POVM) B-

$$\mathbb{P}_Q(\rho, B) = \sum_{m,n}\Big|\mathrm{Tr}\Big\{A_m^{\perp}\rho A_m B_n\Big\}\Big|. \tag{7.6}$$

Note that in this case we have considered POVMs instead of PVM. However, this does not change our proposed measure.

Definition 9. *Let $C(\rho)$ be a function on the space of all density matrices $\mathscr{H}$ such that:*

$$C(\rho) : \mathscr{H} \to \{\mathbb{R}^+, 0\}.$$

Then C is defined as:

$$C(\rho) = \sup_B \mathbb{P}_Q = \sup_B \sum_{m,n} \left| \mathrm{Tr}\left\{ A_m^\perp \rho A_m B_n \right\} \right|. \tag{7.7}$$

Analogous to the case of the visibility functional, we have the following theorem.

Theorem 6. *For strictly incoherent operations (SIO), the $C(\rho)$ is a non increasing function.*

Proof. Consider a strictly incoherent operation,

$$\Lambda[\rho] = \sum_\lambda K_\lambda \rho K_\lambda^\dagger.$$

Now consider the quantity $\mathbb{P}_Q$ for such a state,

$$\mathbb{P}_Q(\Lambda[\rho], B) = \sum_{m,n} \left| \mathrm{Tr}\left\{ A_m^\perp \sum_\lambda K_\lambda \rho K_\lambda^\dagger A_m B_n \right\} \right|.$$

It can be shown that the Kraus operator commute with A (Appendix C.2). Then we have,

$$\mathbb{P}_Q(\Lambda[\rho], B) = \sum_{m,n} \left| \mathrm{Tr}\left\{ A_m^\perp \rho A_m \sum_\lambda K_\lambda^\dagger B_n K_\lambda \right\} \right|.$$

B is a POVM, hence define another POVM, $\tilde{B} = \sum_\lambda K_\lambda^\dagger B_n K_\lambda$. Then,

$$\mathbb{P}_Q(\Lambda[\rho], B) = \sum_{m,n} \left| \mathrm{Tr}\left\{ A_m^\perp \rho A_m \tilde{B}_n \right\} \right| = \mathbb{P}_Q(\rho, \tilde{B}) \leq C(\rho).$$

The last inequality follows from the definition of C, as the measurement $\tilde{B}$ is eligible for ρ but may be suboptimal. The measurements B_n can be chosen to optimise the LHS. Thus we get the required condition,

$$C(\Lambda[\rho]) \leq C(\rho). \tag{7.8}$$

This concludes the proof. □

From the construction of the function, C follows axiom (A1) for a coherence measure. The theorem above shows it also satisfies (A2). Thus C defined here is a coherence monotone. This is considerable evidence towards to support our intuition that this function holds information of non-classicality or quantumness in ρ.

7.3 Conditional Entropy as a Natural Coherence Measure

Let us come back to (4.13).

$$S_C(\rho) = \frac{1}{2}\sum_n \rho_{nn} \log\left[\frac{\rho_{nn}^2}{(\rho^2)_{nn}}\right]$$

From theorem 4, we know that $S_C(\rho) \leq 0$ and the bound is reached iff the state is diagonal.

Therefore, the function $C_S(\rho) = -S_C(\rho)$ corresponds perfectly to the first property (A1) of a coherence measure (Section 5.4). This begs the question that $C_S(\rho)$ might be a coherence measure, similar to the relative entropy of coherence. However this requires further investigation of the properties of C_S. For axiom (A4), one can also show that S_C, the entropic measure is concave, making the proposed measure C_S convex. The monotonicity conditions, axioms (A2) and (A3), need to be explicitly shown. These conditions often are difficult to prove. We have obtained preliminary results in this direction, which we discuss below.

Consider a strictly incoherent operation,

$$\Lambda[\rho] = \sum_\lambda K_\lambda \rho K_\lambda^\dagger = \sum_\lambda p_\lambda \rho_\lambda.$$

Here, $p_\lambda = \mathrm{Tr}\left[K_n \rho K_n^\dagger\right]$ and $\rho_\lambda = \frac{1}{p_\lambda} K_\lambda \rho K_\lambda^\dagger$. These Kraus operators are defined as in (5.5),

$$K_\lambda = \sum_i c_i^\lambda \left|\pi_\lambda(i)\right\rangle\langle i| .$$

The permutations π are bijective. From the property of Kraus operators: $\sum_\lambda K_\lambda K_\lambda^\dagger = 1$, we get

$$\sum_\lambda \left|c_i^\lambda\right|^2 = 1, \quad \forall i. \tag{7.9}$$

Now, we will prove the strong monotonicity criteria (A3) for a restricted class of SIOs. Known as TIO (Transnationally invariant operations) (Baumgratz et al., 2014), these operations are a subset of SIO with fixed permutation π. This class constitutes the free operations for $U(1)$ asymmetry resource theory (Chitambar and Gour, 2016; Korzekwa, 2013).

Theorem 7. *For TIOs, the measure C_S is strongly monotonic.*

Proof. See Appendix C.3. □

As illustrated, this result may provide some indications about the nature of C_S and S_C. However, this is a topic of current investigation. This concludes the work currently done. We will now provide a short summary and discuss future directions.

Chapter 8

Conclusions

This thesis attempts to study the properties of the Kirkwood-Dirac quasi-probability that arise in various quantum mechanical analyses. In Chapter 2, we learned what the Kirkwood-Dirac distribution is and how it is defined as a complete, normalised joint function. We also saw how weak measurements could be used to measure the distribution in an experimental setting.

In Chapter 3, we analysed the distribution through the lens of measure theory and established that it is normalised joint complex measure. This allowed us to understand that the quantities we can experimentally measure are conditional probabilities, requiring a suitable conditional entropic function to quantify its informational content.

In Chapter 4, we examined classical information theory to motivate a similar analysis in the quantum case. We looked at a previous work on the subject and discussed its limitations. We then proposed a more general formula to conquer these limitations. Finally, we looked at a few very interesting properties of our quantum conditional entropic formula.

This concluded the first part of the thesis. We were successfully able to propose a preliminary formula for quantifying the informational content of the Kirkwood-Dirac quasi probability distribution, which does satisfy some promising properties. The function seems to quantify information inaccessible to the von Neumann entropy; it can distinguish between the same states written in two different bases. However, it is still unclear how useful this information is? In other words, in what context would one be able to utilise the information quantified by S_C? Moreover, it is also immediately not clear how to compare the information quantified here to the von Neumann entropy. This is currently a topic of research for this project. We are also looking for other desirable properties that this function may satisfy, such as concavity.

The form of the function, and the fact that it quantifies the information hidden in the off-diagonal terms, motivates us to analyse the Kirkwood-Dirac distribution under the light of quantum coherence.

We give a general introduction to the resource theory of coherence in Chapter 5. In Chapter 6, we take a look at a more concrete motivation for linking coherence and quasi-probabilities. This finally helps us in Chapter 7 to define a coherence monotone from the non-classical part of the Kirkwood-Dirac distribution.

The existence of such a coherence quantifier is in itself a remarkable fact. It allows us to claim with considerable evidence that the Kirkwood-Dirac distribution does contain information about the off-diagonal elements of a density matrix, the amount of superposition in a quantum state. However, the defined function still cannot be elevated to the status of a measure as it does not follow the axiom (A3). Moreover, the function C_S also looks like a promising candidate for a coherence measure. We already saw that it is strongly monotonic for a restricted class of IOs, and if S_C is proven to be concave, C_S immediately follows (A4). Thus, we are also concentrating on exploring this direction to come up with more concrete evidence.

Appendix A

Brief Note on the Two-State Vector Formalism

Definition 10. *Given a Hilbert space $\mathscr{H}_I$ spanned by sets $\{|\alpha\rangle\}$ and $\{|\beta\rangle\}$, , a linear space of two-states $\mathscr{H}_{II}$ is defined by all the linear combination of the two-state $\{|\alpha\rangle\langle\beta|\}$ (Reznik and Aharonov, 1995).*

The most general $\rho_{\alpha|\beta} \in \mathscr{H}_{II}$ is given by

$$\rho_{\alpha|\beta} = \sum_{\alpha,\beta} c_{\alpha,\beta} \, |\alpha\rangle\langle\beta| \, .$$

Here c is a complex constant.
The inner product is defined as - $\langle\rho^1_{\alpha|\beta}, \rho^2_{\alpha|\beta}\rangle = \mathrm{Tr}\left(\rho^{1\dagger}_{\alpha,\beta}\rho^2_{\alpha,\beta}\right)$.

Define $\tilde{\mathscr{H}} \subset \mathscr{H}_{II}$ as the subset containing all physical states. Then, $\forall \rho \in \tilde{\mathscr{H}}_{II}$, $\mathrm{Tr}(\rho) \neq 0$ or $\langle \mathbb{I}, \rho\rangle \neq 0$. In other words, $\tilde{\mathscr{H}}_{II}$ contains all states not orthogonal to $\mathbb{I}$.
To form a normalised basis for this space, let their be two basis sets for $\mathscr{H}_I$ given by

$$S_1 = \{|\alpha\rangle_i\},$$

$$S_2 = \{|\beta\rangle_i\},$$

such that $\langle\beta|\alpha\rangle \neq 0$ for all $|\alpha\rangle_i, |\beta\rangle_i \in S_1, S_2$ respectively.

Now, we define the basis for $\tilde{\mathscr{H}}_{II}$ as

$$\tilde{\rho}_{\alpha|\beta} = \frac{|\alpha\rangle\langle\beta|}{\langle\beta|\alpha\rangle}.$$

Further,

$$\langle\tilde{\rho}_{\alpha|\beta}, \tilde{\rho}_{\alpha'|\beta'}\rangle = \frac{\delta_{\alpha\alpha'}\delta_{\beta\beta'}}{\langle\alpha|\beta\rangle\,\langle\beta'|\alpha'\rangle}.$$

If $\langle\alpha|\beta\rangle = \langle\alpha'|\beta'\rangle$, then the above inner product becomes

$$\frac{\delta_{\alpha\alpha'}\delta_{\beta\beta'}}{|\langle\alpha|\beta\rangle|^2}.$$

In this theory, we define the two-state amplitude (Reznik and Aharonov, 1995),

$$\rho(a,b) = \frac{\left\langle \rho_{\alpha|\beta}, \rho \right\rangle}{\left\langle \rho_{\alpha|\beta}, \rho_{\alpha|\beta} \right\rangle} = \langle a| \rho |b\rangle \langle b|a\rangle = P(a,b).$$

This is indeed the KD distribution. This suggests the operational meaning of all the analysis in this thesis may be found in the TSV formalism.

Appendix B

Details on Some Calculations

B.1 Detailed calculations for the derivation of quantum conditional entropy

As discussed in (4.8), entropy is given by,

$$S_c(\rho_{A|n}) = -\frac{1}{2}\mathrm{Tr}\Big\{\rho_{A|n}\log\Big(\rho_{A|n}\rho^{\dagger}_{A|n}\Big)\Big\}.$$

Simplifying, we find,

$$S_c(\rho_{A|n}) = -\frac{1}{2}\sum_m \langle b_n| \frac{\langle a_m|\rho\,|b_n\rangle}{\langle\rho\rangle}\log\Big(\rho_{A|n}\rho^{\dagger}_{A|n}\Big)\,|a_m\rangle\,,$$

where $\langle\rho\rangle = \langle b_n|\rho\,|b_n\rangle$.
Now,

$$\log\Big(\rho_{A|n}\rho^{\dagger}_{A|n}\Big) = \log\left[\sum_{m,m'}\frac{\langle a_m|\rho\,|b_n\rangle\langle b_n|\rho\,|a'_m\rangle}{\langle\rho\rangle^2}\,|a_m\rangle\langle a'_m|\right] = \log\left[\frac{\rho\,|b_n\rangle\langle b_n|\,\rho}{\langle\rho\rangle^2}\right].$$

Substituting this above, we get,

$$\log\langle\rho\rangle - \frac{1}{2\langle\rho\rangle}\langle\rho\log(\rho\,|b_n\rangle\langle b_n|\,\rho)\rangle.$$

We make a final substitution,

$$\sigma_n = |b_n\rangle\langle b_n|\,\rho\,.$$

Finally we have (4.11)

$$S_C(\rho) = \sum_n\left[\mathrm{Tr}[\sigma_n]\log(\mathrm{Tr}[\sigma_n]) - \frac{1}{2}\mathrm{Tr}\Big[\sigma_n\log\Big(\sigma_n^{\dagger}\sigma_n\Big)\Big]\right].$$

B.2 Proof of existence of the logarithm in (4.11)

Consider a general mixed state $\rho = \sum_i q_i |\psi_i\rangle \langle\psi_i|$ and $\sum_i q_i = 1$ for $q \in [0,1]$
Now, let $|\Psi_0\rangle_n = \sum_i q_i \langle\psi_i|b_n\rangle |\psi_i\rangle$
It can be shown that $\sigma_n \sigma_n^\dagger = |\Psi_0\rangle \langle\Psi_0|_n$
We require a normalisation, the definition for which is

$$\mathcal{N}_\phi = (\sum_i q_i^2 |\langle\psi_i|b_n\rangle|^2)^{1/2}$$

Then the logarithim is

$$\log \sigma_n \sigma_n^\dagger = \left[\log\left(|\mathcal{N}_\phi|\right) |\Psi\rangle \langle\Psi|_n\right]$$

where

$$|\Psi\rangle_n = \frac{|\Psi_0\rangle_n}{\mathcal{N}\phi}$$

As $|\mathcal{N}_\phi| > 0$, the existence is guaranteed.

B.3 Detailed calculations for the break up of the Kirkwood-Dirac distribution

Note that from (7.2), we can write,

$$Im(\mathrm{Tr}\,\rho A_m B_n) = \mathrm{Tr}\,\rho[A_m, B_n].$$

Where $[,]$ denotes the commutator.
Now, for an observable A_m, we have,

$$e^{i\phi A_m} = 1 + (e^{i\phi} - 1)A_m,$$

where ϕ is a real parameter. We thus have,

$$e^{i\phi Am} A_n = [1 + (e^{i\phi} - 1)\delta_{mn}]A_n.$$

Using this, we can write ρ' from (7.1) as:

$$\rho' = \frac{1}{2}[\rho + e^{i\pi A_m} \rho e^{-i\pi A_m}].$$

We now evaluate the following expression,

$$\frac{1}{2}\operatorname{Tr}\big(\rho-\rho'\big)e^{i(\pi/2)A_m}B_ne^{-i(\pi/2)A_m} =$$
$$\frac{1}{2}\operatorname{Tr}\left[\frac{1}{2}\rho e^{i(\pi/2)A_m}B_ne^{-i(\pi/2)A_m}-\frac{1}{2}e^{i\pi A_m}\rho e^{-i\pi A_m}e^{i(\pi/2)A_m}B_ne^{-i(\pi/2)A_m}\right].$$

Define $B_n^{\pi/2}=e^{i(\pi/2)A_m}B_ne^{-i(\pi/2)A_m}$ as a projector obtained by performing a selective phase rotation $e^{i(\pi/2)A_m}$ on the projector B_n.
Evaluating $B_n^{\pi/2}$,

$$e^{i(\pi/2)A_m}B_ne^{-i(\pi/2)A_m}=(1+(i-1)A_m)B_n(1-(i+1)A_m).$$

Substituting in the above equation and simplifying other parts, we get,

$$\frac{1}{2}\operatorname{Tr}\left[\frac{1}{2}\rho(1+i^-A_m)B_n(1-i^+A_m)-\frac{1}{2}(1-2A_m)\rho(1-2A_m)(1+i^-A_m)B_n(1-i^+A_m)\right].$$

Where $i^+=i+1$ and $i^-=i-1$.
Finally, this results in

$$\begin{aligned}&-\operatorname{Tr}\rho B_nA_m-\operatorname{Tr}\rho A_mB_nA_m-\operatorname{Tr}(\rho-A_m\rho A_m+(1-A_m)\rho(1-A_m))B_n\\ &=\operatorname{Tr}\rho A_mB_n-Re(\operatorname{Tr}\rho A_mB_n)-\frac{1}{2}\operatorname{Tr}\big(\rho-\rho'\big)B_n=Im(\operatorname{Tr}\rho A_mB_n).\end{aligned}$$

Therefore, we obtain the equation (7.4) (Johansen, 2007),

$$Im(\operatorname{Tr}\rho A_mB_n)=\frac{1}{2}\operatorname{Tr}\big(\rho-\rho'\big)B_n^{\pi/2}.$$

Appendix C

Proofs of Some Theorems

C.1 On strictly incoherent operations

Theorem 8. *An incoherent operations is also strictly incoherent iff π is injective.*

Proof. An IO K is said to be SIO when $K^\dagger$ is also incoherent. We first assume $\pi(i)$ be injective. Now consider the conjugate IO:

$$K_m^\dagger = \sum_i \bar{c}(i) \left|i\right\rangle\!\left\langle \pi_m(i)\right| .$$

As $\pi_m(i)$ is one-one, $\exists\, i' \in S_\pi$ s.t. $|\pi_m(i)\rangle = |i\rangle$. Also, for any $|i\rangle \in \{|i\rangle\}_{d_\pi}$, if $\exists$ an image of it under π, then the image is unique.
Thus, $K_m^\dagger = \sum_i \bar{c}(i) \left|\pi_m(i')\right\rangle\!\left\langle i'\right|$. This is exactly the form of K_m, which is incoherent. Hence, the conjugate is also incoherent.
Now, assume that $K_m^\dagger$ is incoherent. Then we have,

$$K_m'^\dagger = \sum_j \bar{c}(j) \left|\pi_m(j)\right\rangle\!\left\langle j\right| .$$

Of course, K_m is an IO, thus $K_m = \sum_i c(i) \left|\pi_m(i)\right\rangle\!\left\langle i\right|$. Then $K_m^\dagger = \sum_i \bar{c}(i) \left|i\right\rangle\!\left\langle \pi_m(i)\right|$,

$$K_m^\dagger = K_m'^\dagger .$$

$$\therefore \left|\pi(j_0')\right\rangle = |j_0\rangle \implies j_0' = j_0 .$$

Thus π is injective. □

C.2 On strictly incoherent Kraus operators

Theorem 9. *Strictly incoherent Kraus operators commute with projections.*

Proof. We will prove this via explicit calculations. Consider rank one projector A_m and B_n. We evaluate the following expression, in context of section 7.2,

$$A_m^{\perp} K_\lambda^{\dagger} \rho K_\lambda A_m B_n$$

. Strictly incoherent Kraus operators have the form:

$$K_\lambda = \sum_i c_i^\lambda \left|\pi(i)\rangle\langle i\right|_\lambda .$$

Now,

$$A_m^{\perp} K_\lambda^{\dagger} \rho K_\lambda A_m B_n = \sum_{m' \neq m} \sum_{j,l} c_j^\lambda \bar{c}_l^\lambda \left|m'\right\rangle \left\langle m'\middle|j\right\rangle \left\langle \pi_\lambda(j)\right| \rho \left|\pi_\lambda(l)\right\rangle \left\langle l|m\right\rangle \left\langle m\right| B_n$$

$$= \sum_{m' \neq m} c_j^\lambda \bar{c}l^\lambda \delta_{m'j} \delta_{lm} \left\langle \pi_\lambda(j)\right| \rho \left|\pi_\lambda(l)\right\rangle \left|m'\right\rangle\!\left\langle m\right| B_n$$

$$= \sum_{m' \neq m} c_{m'}^\lambda \bar{c}_m^\lambda \left\langle \pi_\lambda(m')\right| \rho \left|\pi_\lambda(m)\right\rangle \left|m'\right\rangle\!\left\langle m\right| B_n.$$

Taking a trace,

$$\mathrm{Tr}\Big(A_m^{\perp} K_\lambda^{\dagger} \rho K_\lambda A_m B_n\Big) = \sum_{m' \neq m} c_{m'}^\lambda \bar{c}_m^\lambda \left\langle \pi_\lambda(m')\right| \rho \left|\pi_\lambda(m)\right\rangle \left\langle m\right| B_n \left|m'\right\rangle .$$

Now consider the opposing expression: $\mathrm{Tr}\Big(\rho A_m K_\lambda B_n K_\lambda^{\dagger} A_m^{\perp}\Big)$.

$$\mathrm{Tr}\Big(\rho A_m K_\lambda B_n K_\lambda^{\dagger} A_m^{\perp}\Big) = \mathrm{Tr}\left(\sum_{m' \neq m} \sum_{j,l} \bar{c}_j^\lambda c_l^\lambda \rho \left|m\right\rangle \left\langle m|\pi_\lambda(j)\right\rangle \left\langle j\right| B_n \left|l\right\rangle \left\langle \pi_\lambda(l)|m\right\rangle \left\langle m'\right|\right).$$

or,

$$\mathrm{Tr}\Big(\rho A_m K_\lambda B_n K_\lambda^{\dagger} A_m^{\perp}\Big) = \mathrm{Tr}\left(\sum_{m' \neq m} \sum_{j,l} \bar{c}_j^\lambda c_l^\lambda \delta_{\pi_\lambda(l),m} \delta_{m,\pi_\lambda(j)} \left\langle j\right| B_n \left|l\right\rangle \rho \left|m\right\rangle\!\left\langle m'\right|\right)$$

$$= \sum_{m' \neq m} \bar{c}_{k^{-1}(m)}^\lambda c_{k^{-1}(m')}^\lambda \left\langle k^{-1}(m)\right| B_n \left|k^{-1}(m')\right\rangle \left\langle m'\right| \rho \left|m\right\rangle .$$

Relabelling indices: $m' \to \pi_\lambda(m')$ and $m \to \pi_\lambda(m)$, the above expression reduces to

$$= \sum_{\pi_\lambda(m') \neq \pi_\lambda(m)} \bar{c}_m^\lambda c_{m'}^\lambda \left\langle m\right| B_n \left|m'\right\rangle \left\langle \pi_\lambda(m')\right| \rho \left|\pi_\lambda(m)\right\rangle .$$

Sum over $\pi_\lambda(m)$ or m is the same as $\pi_\lambda : \{|i\rangle\}_d \to \{|i\rangle\}_d$. Thus, such a permutation is permissible.

$\square$

C.3 On C_S being strongly monotonic (Theorem 7)

Theorem 7. *For TIOs, the measure C_S is strongly monotonic.*

Proof. We consider:

$$\sum_\lambda p_\lambda C_S(\rho) = -\sum_\lambda \sum_n \left\langle K_\lambda \rho K_\lambda^\dagger \right\rangle \log \frac{\left\langle K_\lambda \rho K_\lambda^\dagger \right\rangle^2}{\left\langle K_\lambda \rho K_\lambda^\dagger K_\lambda^\dagger \rho K_\lambda \right\rangle}. \tag{C.1}$$

Where $\langle A \rangle = \langle n| A |n\rangle$, for basis $|n\rangle_i$.
Now we show:

$$\left\langle K_\lambda \rho K_\lambda^\dagger \right\rangle = \sum_{j,j'} c_j^\lambda c_{j'}^\lambda \langle n|j\rangle \langle \pi(j)| \rho |\pi(j)\rangle \langle j'|n\rangle = \left|c_n^\lambda\right|^2 \langle \pi(n)| \rho |\pi(n)\rangle,$$

and similarly we can prove:

$$\left\langle K_\lambda \rho K_\lambda^\dagger K_\lambda^\dagger \rho K_\lambda \right\rangle = \left|c_n^\lambda\right|^2 \sum_j \left|c_j^\lambda\right|^2 |\langle \pi(j)| \rho |\pi(n)\rangle|^2$$

From $C.1$,

$$\sum_\lambda p_\lambda C_S(\rho) = -\sum_\lambda \sum_n \left|c_n^\lambda\right|^2 \left\langle K_\lambda \rho K_\lambda^\dagger \right\rangle \log \frac{\left|c_n^\lambda\right|^2 \langle \pi(n)| \rho |\pi(n)\rangle^2}{\sum_j \left|c_j^\lambda\right|^2 |\langle \pi(j)| \rho |\pi(n)\rangle|^2}.$$

WLOG, we change indices:

$$\pi(i) \to i; \quad i \to \pi^{-1}(i).$$

The existence of the inverse is guaranteed by bijectivity (SIO). Hence we finally have,

$$\sum_\lambda p_\lambda C_S(\rho) = -\sum_\lambda \sum_n \left|c_{\pi^{-1}(n)}^\lambda\right|^2 \langle n| \rho |n\rangle \log \frac{\left|c_{\pi^{-1}(n)}^\lambda\right|^2 \langle n| \rho |n\rangle^2}{\sum_j \left|c_{\pi^{-1}(j)}^\lambda\right|^2 |\langle j| \rho |n\rangle|^2}. \tag{C.2}$$

We will now calculate $C_S(\rho) - \sum_\lambda p_\lambda C_S(\rho_\lambda)$

$$= -\sum_n \langle n| \rho |n\rangle \log \frac{\langle n| \rho |n\rangle^2}{\sum_i |\langle n| \rho |i\rangle|^2} + \sum_\lambda \left|c_{\pi^{-1}(n)}^\lambda\right|^2 \log \frac{\left|c_{\pi^{-1}(n)}^\lambda\right|^2 \langle n| \rho |n\rangle^2}{\sum_j \left|c_{\pi^{-1}(j)}^\lambda\right|^2 |\langle j| \rho |n\rangle|^2}.$$

Now as $\sum_\lambda \left|c_i^\lambda\right|^2 = 1, \quad \Rightarrow \quad \sum_\lambda \left|c_{\pi^{-1}(i)}^\lambda\right|^2 = 1, \quad \forall i,$

$$= -\sum_\lambda \sum_n \left|c_{\pi^{-1}(n)}^\lambda\right|^2 \langle n| \rho |n\rangle \log \frac{\langle n| \rho |n\rangle^2}{\sum_i |\langle n| \rho |i\rangle|^2} \cdot \frac{\sum_j \left|c_{\pi^{-1}(j)}^\lambda\right|^2 |\langle \pi(j)| \rho |\pi(n)\rangle|^2}{\left|c_{\pi^{-1}(n)}^\lambda\right|^2 \langle \pi(n)| \rho |\pi(n)\rangle^2}.$$

Using concavity of logarithmic function, we obtain,

$$\leq -\sum_n \langle n|\,\rho\,|n\rangle \log \frac{\sum_j \sum_\lambda \left|c^{\lambda}_{\pi^{-1}(j)}\right|^2 |\langle n|\,\rho\,|j\rangle|^2}{\sum_i |\langle n|\,\rho\,|i\rangle|^2}$$

$$= -\sum_n \langle n|\,\rho\,|n\rangle \log 1 = 0.$$

Hence,

$$C_S(\rho) \geq \sum_\lambda p_\lambda C_S(\rho_\lambda). \tag{C.3}$$

Thus C_S is strongly monotonic under TIOs.

□

Bibliography

Aberg, J. (2006). Quantifying superposition.

Aharonov, Y. and Vaidman, L. (1990). Properties of a quantum system during the time interval between two measurements. *Phys. Rev. A*, 41:11–20.

Baggett, L. (1991). *Functional Analysis: A Primer*. Chapman & Hall Pure and Applied Mathematics. Taylor & Francis.

Ban, M. (2021). On sequential measurements with indefinite causal order. *Physics Letters A*, 403:127383.

Baumgratz, T., Cramer, M., and Plenio, M. B. (2014). Quantifying coherence. *Phys. Rev. Lett.*, 113:140401.

Biswas, T., García Díaz, M., and Winter, A. (2017). Interferometric visibility and coherence. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2203):20170170.

Chitambar, E. and Gour, G. (2016). Critical examination of incoherent operations and a physically consistent resource theory of quantum coherence. *Phys. Rev. Lett.*, 117:030401.

Cover, T. M. (2006). *Elements of information theory / Thomas M. Cover, Joy A. Thomas.* Wiley-Interscience, Hoboken, N.J, 2nd ed. edition.

Dirac, P. A. M. (1945). On the analogy between classical and quantum mechanics. *Rev. Mod. Phys.*, 17:195–199.

Feintzeig, B. (2015). Hidden variables and incompatible observables in quantum mechanics. *The British Journal for the Philosophy of Science*, 66(4):905–927.

Folland, G. (1994). *A Course in Abstract Harmonic Analysis*. Studies in Advanced Mathematics. Taylor & Francis.

Johansen, L. M. (2007). Quantum theory of successive projective measurements. *Phys. Rev. A*, 76:012119.

Korzekwa, K. (2013). PhD thesis.

Reznik, B. and Aharonov, Y. (1995). Time-symmetric formulation of quantum mechanics. *Phys. Rev. A*, 52:2538–2550.

Salek, S., Schubert, R., and Wiesner, K. (2014). Negative conditional entropy of postselected states. *Phys. Rev. A*, 90:022116.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.

Streltsov, A., Adesso, G., and Plenio, M. B. (2017). Colloquium: Quantum coherence as a resource. *Rev. Mod. Phys.*, 89:041003.

v. Neumann, J. (1933). Mathematische grundlagen der quantenmechanik. *Monatshefte für Mathematik und Physik*, 40(1):A31–A32.