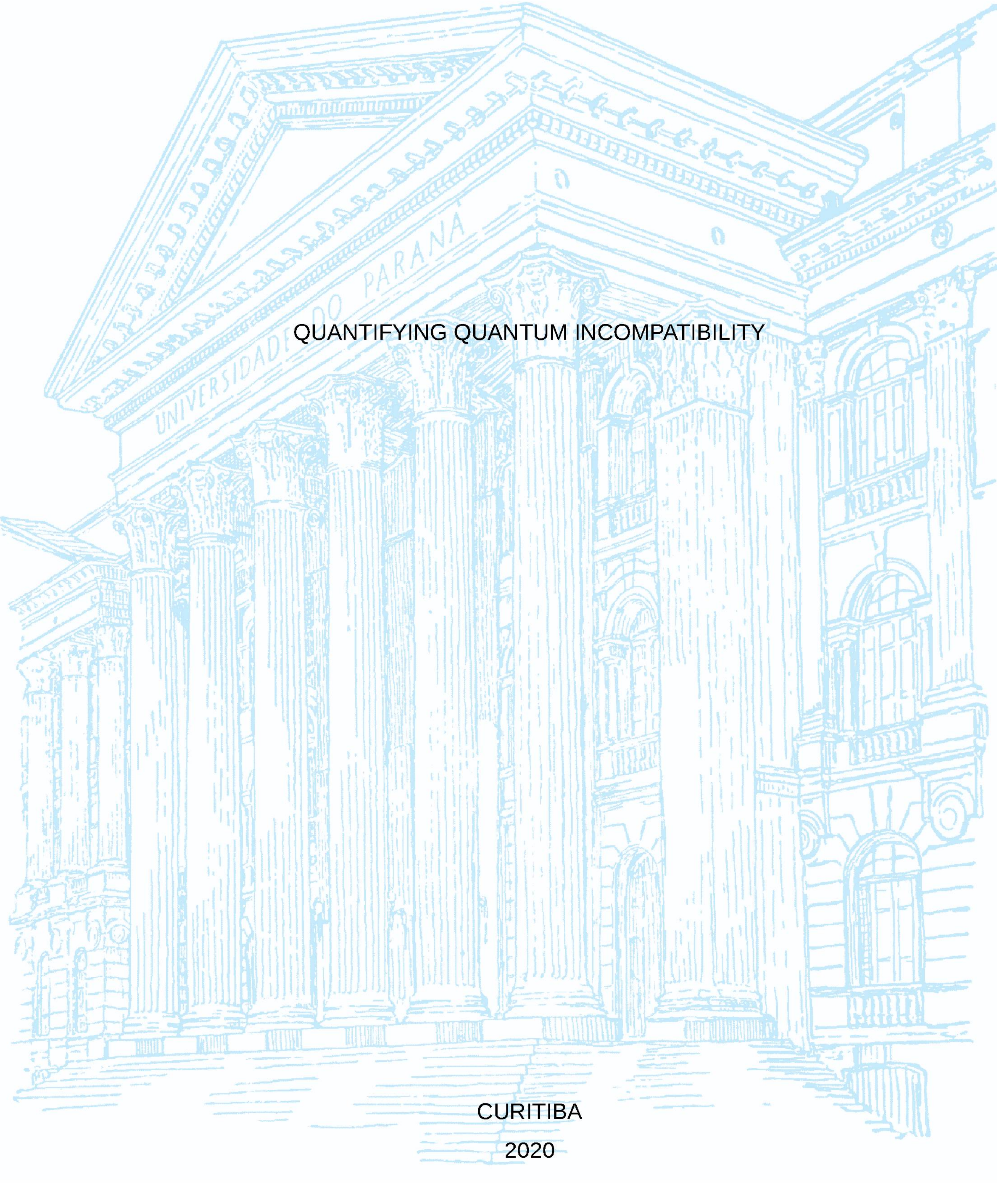
UNIVERSIDADE FEDERAL DO PARANÁ

EVERLYN MARTINS



QUANTIFYING QUANTUM INCOMPATIBILITY

CURITIBA

2020

EVERLYN MARTINS

QUANTIFYING QUANTUM INCOMPATIBILITY

Dissertação apresentada ao Curso de Pós-Graduação em Física, Setor de Ciências Exatas, da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Física.

Orientador: Prof. Dr. Renato M. Angelo

CURITIBA

2020

Catalogação na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

M386q Martins, Everlyn
Quantifying quantum incompatibility [recurso eletrônico] Everlyn Martins. – Curitiba, 2020.

Dissertação - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Física, 2020.

Orientador: Renato Moreira Angelo.

1. Teoria quântica. 2. Física. I. Universidade Federal do Paraná. II. Angelo, Renato Moreira. III. Título.

CDD: 530.12

Bibliotecária: Vanusa Maciel CRB- 9/1928

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em FÍSICA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **EVERLYN MARTINS** intitulada: **"Quantifying Quantum Incompatibility"**, sob orientação do Prof. Dr. RENATO MOREIRA ANGELO, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.
A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 29 de Julho de 2020.

Assinatura Eletrônica
29/07/2020 17:29:53.0
RENATO MOREIRA ANGELO
Presidente da Banca Examinadora (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
29/07/2020 17:38:09.0
JONAS MAZIERO
Avaliador Externo (UNIVERSIDADE FEDERAL DE SANTA MARIA)

Assinatura Eletrônica
29/07/2020 19:16:33.0
MILTON MASSUMI FUJIMOTO
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

"If you are receptive and humble, mathematics will lead you by the hand."
–Paul A. M. Dirac

Agradecimentos

Primeiramente, agradeço à Jesus Cristo por tudo.

Meu sincero agradecimento ao Professor Renato por toda a paciência, dedicação, comprometimento, disponibilidade e pela oportunidade em trabalharmos juntos durante esses dois anos.

Agradeço, também, à minha família por todo o apoio e por estamos juntos nessa jornada.

Agradeço à banca de defesa e pré-defesa por avaliar este trabalho, e pelas contribuições significativas ao texto final. Todas as observações, críticas e correções foram de suma importância.

Ao Matheus Savi, meu colega, agradeço por inúmeras discussões e pela contribuição inestimável aos trabalhos, e também pela paciência e o tempo que ele teve em explicar muitos conceitos físicos.

Ao Grupo de Fundamentos em Mecânica Quântica, organizado pelos Professores Renato e Alexandre, agradeço pelas diversas discussões e debates, além do rico conhecimento adquirido durante às sessões.

Meus agradecimentos, também, aos meus amigos(as) Erick M. Veronese, Carlos Toro, Raphael Gatto, Vitor Skrzypczak, Emilia Grzyb, Márcia R. Martins, Alyona Stepnova, Priscila Cavassin e Valeria Gashkova pelas conversas, discussões e contribuições à minha aprendizagem nesta vida.

Ao Programa de Pós-graduação em Física da UFPR, que consiste em todos os professores, funcionários e infraestrutura, em especial ao Leandro C. Florentino, secretário, por toda a paciência em sanar minhas dúvidas e questões burocráticas, e bate-papos.

E, agradeço ao CNPq pelo apoio financeiro.

RESUMO

Um dos aspectos mais intrigantes da mecânica quântica remonta aos trabalhos dos fundadores da teoria, Heisenberg e Bohr, que revelaram aspectos fundamentais não clássicos relacionados às medições dentro do domínio quântico. A incompatibilidade é um elemento essencial que faz com que os sistemas microscópicos se comportem de maneira "inesperada". Em livros acadêmicos, geralmente associado com relações de comutação, esse conceito é interpretado como uma construção puramente algébrica no espaço de Hilbert, em vez de uma noção operacionalmente mensurável em laboratórios reais. Em face à esses dilemas, é proposto dois cenários diferentes para a medições projetivas sequenciais que definem a incompatibilidade dentro de um contexto físico definido não apenas pelos observáveis, mas também pelo estado quântico, embora não seja uma abordagem geral, é um procedimento com resultados muito intuitivos. A primeira noção relaciona incompatibilidade com violações da regra de Bayes em configurações envolvendo medições de dois observáveis usando dois ordenamentos opostos. Nesse caso, um quantificador é introduzido com base na divergência Kullback-Leibler, com uma regularização apropriada. A segunda abordagem vincula a incompatibilidade quântica à quantidade de informação que pode ser extraída de um sistema em um cenário que envolve um canal de comunicação. Aqui, a incompatibilidade é reconhecida como um recurso para testar a segurança de canais de comunicação contra vazamento de informações (espionagem). Além disso, é definida uma medida para quantificar o grau de incompatibilidade de contexto que é facilmente computável, admite uma interpretação geométrica e é máxima somente se as auto-bases dos observáveis envolvidos forem mutuamente não-enviesadas.

Palavras-chave: incompatibilidade, recursos quânticos, violação da regra de Bayes, informação e medições quânticas.

ABSTRACT

One of the most intriguing aspects of quantum mechanics tracks back to the works by the founders of the theory, Heisenberg and Bohr, who unveiled fundamental non-classical aspects related to measurements in the quantum realm. Incompatibility is a special element that makes microscopic systems behave in an "unexpected" way. Generally associated in academic books with commutation relations, this concept has been interpreted as a pure algebraic construction in Hilbert space rather than an operationally measurable notion in actual laboratories. In this thesis, two different frameworks concerning sequential projective measurements are proposed which define incompatibility within a physical context defined not only by observables but also by the quantum state, although not being a general approach it is an insightful procedure. The first one relates incompatibility with violations of the Bayes rule in setups involving measurements of two observables using two opposite orderings. In this case, a quantifier is introduced with basis on the Kullback-Leibler divergence, with an appropriate re-scaling. The second approach links quantum incompatibility with the amount of information that can be extracted from a system within a scenario involving a communication channel. Here, incompatibility is recognized as a resource to test the safety of communication channels against information leakage (espionage). Moreover, a measure to quantify the degree of informational incompatibility is defined that is easily computable, admits a geometrical interpretation, and is maximum only if the eigenbases of the involved observables are mutually unbiased.

Keywords: incompatibility, quantum resources, Bayes' rule violation, information and quantum measurements.

CONTENTS

CHAPTER 1

INTRODUCTION

Likewise superposition, entanglement, coherence, non-commutativity, etc., the concept of incompatibility is only relevant in the Quantum Mechanics realm since classical physics has not measurement-related concerns. The relevance of incompatible observables can be tracked back to Heisenberg's Uncertainty Principle [Heisenberg (1928)] and Bohr's Complementarity Principle [Bohr (1928)]. The first states that the product of the standard deviations of position and momentum must have a lower bound. In Heisenberg's words "the more precisely the position is determined, the less precisely the momentum is known, and conversely". Position and momentum are not only complementary quantities but also incompatible observables; they are complementary in the sense that they cannot be measured simultaneously. However, the complementarity principle goes further stating that every measurement is performed by a macroscopic apparatus over a microscopic object, and it influences the microscopic object state. Copenhagen interpretation [Gomatam (2007)] expresses that the microscopic object together with the apparatus determines the result of a measurement. In this sense, incompatibility should not be thought of as property solely of the observables. The incompatible nature of some observables is one of the defining properties of quantum mechanics; one that implies an important departure from classical principles. On the other hand, the uncertainty principle and complementarity have been revisited recently [Saha (2020)].

It is possible to establish a connection between the uncertainty principle and the commutation relations (commutators) of two complementary observables. Thus, in the scientific literature, it is not uncommon to find books relating to commutators with compatibility or incompatibility. The authors in [Dirac (1930)–Banks (2018)] state that, if observables commute they are compatible and if not they are incompatible. Authors in [Dicke (1960)–Moretti (2019)] define compatibility and incompatibility as to whether observables can be measured simultaneously or not, respectively. References [Mandl (1992)–Kirsten (2006)] relate incompatibility directly to the uncertainty principle. Authors in [Schiff (1968), Weinberg (2015)] do not even make a mention to compatibility or incompatibility, at least directly. From these samples, it is fair to say that authors are not cohesive when treating incompatibility. Surprisingly, the state of affairs turns out to be such that one of the most fundamental concepts in quantum theory is not a solid subject in academic books.

An important obstacle to consider commutators for quantifying incompatibility is that they are algebraic structures that are not directly measurable, once with the measurement results is possible to obtain only probabilities and expectation values. It should be expected, therefore, a suitable quantifier of incompatibility is a function of probabilities and the outputs of a measurement. Another obstacle is that commutators do not take into account the physical state of the system. The whole structure of the quantum theory is formulated in a way that the physical predictions are obtained not only with operators but also with the quantum state. Thus, one may wonder whether the ultimate notion of incompatibility should depend only on the observables. After all, in every actual test of the concept, one always needs to prepare and measure the system, which amounts to initially setting a quantum state for the system. In other words, it is desirable from physics to have a definition of incompatibility in the physical space-time rather than in the Hilbert space.

Another point regarding the importance of quantum states in incompatibility is that when approaching classicality, there should not exist any measurement of incompatibility. Therefore, intuition requires that quantum incompatibility should vanish as the system approaches the classical domain—instance that is usually accomplished through the quantum state. Accordingly, measurement incompatibility has been shown to disappear under noise [Schultz (2015)]. In such approach, however, one can use the duality relation $\mathrm{Tr}[\Gamma(\rho)\,X] = \mathrm{Tr}[\rho\,\Gamma_*(X)]$ to maintain a state-independent notion of measurement incompatibility. Indeed, one can always interpret any local noisy channel Γ, leading ρ to a classical state, as implying some degree of fuzziness in the X measurement. Nevertheless, this concept seems to be related more to experimental imperfections [Designolle (2019)] than to fundamental classicalization processes involving the discard of correlated systems [Zurek (2007), Dieguez (2018)]. A subtler classical scenario can be conceived as follows. As far as heavy bodies are concerned, measurements are expected to be nearly nondisturbing, so that the resulting physical state should be independent of the ordering with which two noncommuting observables are measured. We then have a clear dependence of the notion of measurement incompatibility with an intrinsic property (the mass) of the probed body. In this case, it is less obvious how to effectively rephrase classicality in the formal structure of the measurement operators. In contrast, the mass is naturally encoded in the quantum state which is a solution of the Schödinger equation.

The authors in reference [Cohen (1991)] propose an interesting perspective regarding incompatibility. Their argument is that "when two observables are compatible, the physical predictions are the same, whatever the order of performing the two measurements (provided that the time interval which separates them is sufficiently small). The probabilities of obtaining either one observable output then the other observable output or the reverse other are identical. On the other hand, if the observables do not commute, the preceding arguments are no longer valid.". In the face of that, a quantifier first is proposed. When performing measurements of observables for a given state, it is convenient to identify a set called "context" whose elements are the state in question and the observables. The context generates joint probabilities that, according to quantum mechanics, do not obey Bayes' rule if all elements of a context do not commute, which means that the joint probability distributions differ from each other. If all elements of a context do commute, then the joint probability distributions are equal, not violating Bayes' rule. This is astonishing, since Bayes' rule in one of the most fundamental concepts of classical probability theory. In this perspective, it is reasonable to elect a measure that quantifies the distance between these probability distributions. A possible candidate is a Kullback-Leibler divergence (KL divergence), which is a distance of two probability distributions. At first sight, this quantifier may manifest two operational drawbacks. First, it runs from 0 to ∞. However, it is not difficult to introduce a monotonic function confining this large interval to $[0, 1]$. With that, the same information provided by the KL divergence is entirely rescaled to a convenient image. Second, the computational effort involved grows with the Hilbert space dimension. Although, as we show later, the calculations are mapped on vectors of dimension $d^2 - 1$, which, therefore, do no involve exponential increase with d.

Some other paths in the literature are taken, instead of quantifying incompatibility, some works contextualize it in a variety of other different physical concepts. For instance, a faithful symptom of incompatibility is violations of *joint measurability* [Busch (1986)–Ziman (2016)]—the hypothesis that a set of measurements can be decomposed in terms of a single "parent" measurement. Incompatibility also unveils interconnections between the so-called *measurement incompatibility* and *nonlocal resources*, as for instance *Bell nonlocality* [Fine (1982)–Bene (2018)] and Einstein-Podolsky-Rosen *steering* [Quintino (2014)–Uola (2015)], in the sense that incompatibility is a required ingredient to accomplish them. As for a quantitative assessment of the concept, *incompatibility robustness* measures have been introduced [Busch (2013), Heinosaari (2015)] with basis on the amount of noise needed to render the measurements (or devices [Haapsalo (2015)]) compatible. From that, further developments were accomplished within the contexts of *device-independent characterizations* [Cavalcanti (2016)–Chen (2018)], *state-discrimination tasks* [Toigo (2018)–Cavalcanti (2019)], and *quantum programmability* [Buscemi (2020)], through which operational interpretations were conceived to measurement incompatibility. Recently, however, unexpected features have been noted for some widely used *robustness-based measures* of incompatibility [Designolle (2019)], some measures does not satisfy some natural properties of incompatibility, for instance, some measures are not maximum when considering mutually unbiased bases.

The author in reference [Cohen (1991)] also argues that: "*when two observables A and B are compatible, the measurement of B does not cause any loss of information previously obtained from a measurement of A (and vice versa) but, on the contrary, adds to it. Moreover, the order of measuring the two observables A and B is of no importance. This last point, furthermore, enables us to envisage the simultaneous measurement of A and B.*" and "*two incompatible observables cannot be measured simultaneously. It can be seen... that the second measurement causes the information supplied by the first one to be lost. If, for example, after the sequence: 'we measure A and find, for example, a_1; the system goes into the state $|u_1\rangle$. We then measure B and find, for example, b_2; the state of the system becomes $|v_2\rangle$: we measure A again, we can no longer be sure of the result, since $|2\rangle$ is not an eigenvector of A. All that was gained by the first measurement of A is thus lost*". In contrast to the previous argument, this is a different perspective to look at incompatibility; it involves a different physical concept. This requires "new" tools to deal with incompatibility in terms of information. To this end, the usual object to compare two quantum states is the relative entropy, a non-metric 'distance' between two quantum states. An important step toward the introduction of a physical concept within quantum information field is its statement in terms of a communication task. This move is interesting because it attaches an operational interpretation of the concept and automatically indicates that such a concept can be viewed as a quantum resource. The task considered here is the certification of the safety of a communication channel, which is of primal relevance for quantum cryptography [Gisin (2002)]. Where the idea of a context is still in place, since the relative entropy depends on the input state and also what observables are considered. Although several quantifiers can, in principle, be introduced, it is shown that not all are capable of reproducing the expected results.

This work is structured as follows. In the second chapter, it is introduced the mathematical tools used throughout this manuscript. In it, a brief review about quantum mechanics is presented regarding density operators, reduced states and conditional states. A brief review of classical probability theory is also presented, with conditional and joint probabilities. A somewhat detailed presentation of "quantum" probability theory is introduced, which is nothing but the adaptation of conditional and joint probabilities for quantum measurements. In addition, it is shown how to write states in a general Bloch representation and also in discrete space and momentum representations.

In the third chapter, an incompatibility framework is reviewed intuitively. An argument is designed to contextualize Bayes' rule violation in sequential measurements relating it to incompatibility. Using a symmetric KL divergence of two joint probabilities, a measure of Bayes' divergence is proposed, with an appropriate normalization, which quantifies the incompatibility of a context in the interval $[0, 1]$. As an application, a state of the qubit is considered, together with its generalization, and a state in discrete space and momentum representation.

In the fourth chapter, a framework is introduced to test the safety of a communication channel, thus linking quantum incompatibility with information—the most fundamental resource for quantum information and quantum thermodynamics tasks [Horodecki (2003)–Costa (2020)]. Our approach employs a key principle powering quantum cryptography [Gisin (2002)], namely, that no information can be extracted from a system without disturbing it [Fuchs (1996), Busch (2009)]. As an application, a state in a d-dimensional space is considered in a generalized Bloch representation and also a system described in a discrete space and momentum representation.

In the fifth and last chapter, the conclusions are presented. Some comparisons are shown between the models, and their differences are discussed. In particular, we show how our information-based framework for contextual incompatibility can be used to quantify measurement incompatibility, thus being related to non-commutativity. We then close this work with perspectives for future research.

CHAPTER 2

THEORETICAL BACKGROUND

In this chapter, we will introduce some important mathematical tools that will be used in this work. To ensure that the reader does not become overwhelmed with all the mathematical rigour, the concepts presented maximally simplified to be possible for one grasp only what is necessary without forgetting the underlying physical motivation. Because the main subject of this work is incompatibility in the quantum realm, it is natural to start with a review of important definitions in quantum mechanics. Then, we discuss how to use quantum theory to describe probabilities distributions, which are quantities that can be obtained experimentally. Although these probabilities describe quantum systems, *i.e.*, they are 'quantum probabilities', it is still possible to apply the KL divergence to compute the "distance" between two probability distributions. Next, we introduce a somewhat detailed introduction to Bloch's representation in a d-dimensional Hilbert space, applying it to study general probability distributions. The following subject is a brief introduction to classical and quantum information theory. Finally, we present a discretization method allowing for a simplified treatment of continuous variables, such as position and momentum.

2.1 Review of Quantum Mechanics

The most important framework in this manuscript is the quantum theory, whose algebraic tools are used throughout. For this reason, a brief review of quantum mechanics is given next which aims at gathering the main concepts necessary for the comprehension of our findings, setting the notation, and introducing some advanced notions.

2.1.1 Quantum States

In the previous section, we provide some examples of quantum systems, yet it was not discussed how to describe in a general mathematical and physical structure those kinds of quantum systems. The quantum theory is based on observations through experiments, which give us a pretty good view of the quantum world. Therefore, the theory gives us postulates that are sufficient to describe all quantum phenomena. However, this physical theory requires a mathematical framework. In this section, we will introduce a physical and mathematical prescription. This will give us all the needed tools to describe quantum states of systems like those given previously.

The first postulate is related to the quantum states of physical systems and sets up the mathematical structure required.

Postulate 1. (Quantum state) *The 'state' of an isolated physical system in time t is associated to a complex vector $|\psi(t)\rangle \in \mathcal{H}$ also so-called state vector, where $\mathcal{H}$ is a vector space so-called Hilbert space or state space with an inner product and $|\psi(t)\rangle$ must be normalized and describes a pure ensemble.*

Some systems can be decomposed in a linear combination of orthonormal measurement basis $\{|\psi_i\rangle\} \in \mathcal{H}$, referred as 'superposition' states which give us the interference phenomena, and it can be written as

$$|\psi(t)\rangle = \sum_i c_i(t)\,|\psi_i\rangle\,. \tag{2.1.1}$$

In the space $\mathcal{H}$ inner products are denoted by $\langle\phi|\psi\rangle$, where $\langle\phi|$ is called the 'dual' of the vector $|\psi\rangle$ and it is not a vector, but a functional, a mathematical object (a map) that takes vectors into a (complex) number. This means that $\langle\phi|$ is not an element of $\mathcal{H}$, but instead $\langle\phi| \in \mathcal{H}^*$, where $\mathcal{H}^*$ is called the 'dual space' of $\mathcal{H}$, *i.e.*, the space of all linear functionals of $\mathcal{H}$.[1] For orthonormal basis we have that $\langle\psi_i|\psi_j\rangle = \delta_{ij}$, where δ_{ij} is the 'Kronecker delta'.

The simplest quantum system is a qubit, which was already mentioned in the previous section as an example of a quantum system. In computation basis, a general state vector can be written as

$$|\psi\rangle = c_0\,|0\rangle + c_1 e^{\imath\theta}\,|1\rangle\,,^2 \tag{2.1.2}$$

where $c_0, c_1 \in \mathbb{C}$ and $\theta \in \mathbb{R}$ with $|c_0|^2 + |c_1|^2 = 1$.

2.1.2 Quantum Systems

In the present work, we shall deal only with (finite) d-dimensional Hilbert spaces $\mathcal{H} \simeq \mathbb{C}^d$, *i.e.*, $\mathcal{H}$ is isomorphic to the d-dimensional complex number field $\mathbb{C}^d$, where $d > 1 \in \mathbb{N}$. To any physical degree of freedom, a space with inner product $\mathcal{H}$ can be defined, whose dimension is determined by the characteristics of the system. Two examples of physical systems are given below:

- **Photon polarization:** Photons with left-handed and right-handed circular polarization have two degrees of freedom, so $\mathcal{H} \simeq \mathbb{C}^2$ and its eigenbasis is denoted by $\{|l\rangle, |r\rangle\}$, respectively. Physical systems with two degrees of freedom are named 'qubit', therefore, they have a two-dimensional state space.

- **Spin-1/2 Particles:** Fundamental particles have discrete internal degrees of freedom—the spin. The electron, for instance, is a spin-1/2 particle and as such, admits a description $\mathcal{H} \simeq \mathbb{C}^2$ via the so-called 'computational basis' $\{|0\rangle, |1\rangle\}$ (eigenbasis of σ_z).

These are the most common and pedagogical examples of physical systems whose internal degree of freedoms are described by d-dimensional Hilbert spaces. They are pedagogical examples; however, one can think of other kinds of systems.

[1] Both spaces, Hilbert space and its dual, are isomorphic $\mathcal{H} \simeq \mathcal{H}^*$, thus for all practical purposes $\langle\psi|$ is the transpose conjugate of $|\psi\rangle$.*Riesz representation theorem* affords a convenient description of the dual space stablishing a relationship between $\mathcal{H}$ and $\mathcal{H}^*$

[2] To avoid confusion, the symbol i is only used for indices and $\imath$ is the imaginary number of the set $\mathbb{C}$.

We can also have *composite systems* which are composed of multiple particles. In this work, the maximum number of particles that we are dealing with is two, *i.e.*, systems with two particles (or subsystems) with d degrees of freedom each are associated with a *composite Hilbert space* $\mathcal{H}_{\mathcal{XY}} \equiv \mathcal{H}_{\mathcal{X}} \otimes \mathcal{H}_{\mathcal{Y}}$. However, we will not give much emphasis on the composite ones, because when treating incompatibility, we restrain to the case of $\mathcal{H}$ (a single system) and not $\mathcal{H}_{\mathcal{XY}}$ (two systems), since it is a simpler scenario and fully captures the notion of incompatibility.

2.1.3 Operators

Definition 1. (Operator) *A 'linear operator' or simply 'operator'* $X : \mathcal{H} \to \mathcal{H}$ *acting on a Hilbert space* $\mathcal{H}$ *is a rule that assigns to each* $|\psi\rangle \in \mathcal{H}$ *another vector* $X|\psi\rangle = |\phi\rangle \in \mathcal{H}$.[3]

An operator X is said bounded if its norm satisfies $\| X|\psi\rangle \| < \infty, \forall |\psi\rangle$. The set of all bounded operators of $\mathcal{H}$ is denoted $\mathscr{B}(\mathcal{H})$. Operators also have duality, operators called 'ad-joint operators' or 'dual operators' denoted $X^\dagger$ and acting accordingly to the rule $\langle\psi| X^\dagger = \langle\phi|$, where $\langle\psi|, \langle\phi| \in \mathcal{H}^*$.

Commutator and Anti-commutator

Definition 2. (Commutator) *Given two operators* X *and* Y *the 'commutator' is given by*

$$[X, Y] = XY - YX \tag{2.1.3}$$

and the 'anti-commutator' is given by

$$\{X, Y\}_+ = XY + YX. \tag{2.1.4}$$

Notice that $X, Y \in \mathscr{B}(\mathcal{H})$, if $X \in \mathscr{B}(\mathcal{H}_{\mathcal{X}})$ and $Y \in \mathscr{B}(\mathcal{H}_{\mathcal{Y}})$ their commutator is null, one acts in one space and the other acts on a complete different space. Two operators X and Y are said to be 'commutative' if $[X, Y] = XY - YX = 0$ and 'non-commutative' if $[Y, X] \neq 0$.

Degree of Non-commutativity

Definition 3. (Non-commutativity degree) *The 'non-commutativity degree' of two operators* X *and* Y *is given by*

$$\mathscr{N}_{\{X,Y\}} := \frac{1}{4} \| [X, Y] \|, \tag{2.1.5}$$

where $\|X\| := \sqrt{\mathrm{Tr}(X^\dagger X)}$ *is the 'Hilbert-Schmidt norm' of* X *and* $1/4$ *is only a proper normalization factor.*

The last definition is a state-independent object whose multiplication constant $1/4$ is merely a normalization factor. Notice when $\mathscr{N}_{\{X,Y\}} = 0$, X and Y commute.

[3] An useful way to write an operator is in terms of vectors, $X = |\phi\rangle \langle\psi|$.

Trace Operation

Given an operator X and two complete orthonormal bases $\{|\psi_i\rangle, |\phi_j\rangle\}$, the 'trace' of X is written as

$$\operatorname{Tr} X = \sum_{i=1}^{\dim \mathcal{H}} \langle \psi_i | X | \psi_i \rangle \, . \tag{2.1.6}$$

The trace operation has the following properties:

i) **Base invariance**: $$\sum_i \langle \psi_i | X | \psi_i \rangle = \sum_j \langle \phi_j | X | \phi_j \rangle \, . \tag{2.1.7}$$

ii) **Cyclic permutation**: $$\operatorname{Tr}(XYZ) = \operatorname{Tr}(YZX) = \operatorname{Tr}(ZXY). \tag{2.1.8}$$

iii) **Partial trace**: $$\operatorname{Tr}_{\mathcal{X}}(X \otimes Y) = \operatorname{Tr}(X) Y \quad \textit{and} \quad \operatorname{Tr}_{\mathcal{Y}}(X \otimes Y) = \operatorname{Tr}(Y) X. \tag{2.1.9}$$

Here, $X \otimes Y$ denotes the 'tensor product' between X and Y which describes systems consisting of multiple subsystems, it acts on the composite space $\mathcal{H}_{\mathcal{X}} \otimes \mathcal{H}_{\mathcal{Y}}$. And $\operatorname{Tr}_{\mathcal{X}}$ denotes the trace only in the space where X acts on, *i.e.*, $\mathcal{H}_{\mathcal{X}}$. And $\operatorname{Tr}_{\mathcal{Y}}$ denotes the trace only in the space where Y acts on, *i.e.*, $\mathcal{H}_{\mathcal{Y}}$.

Projection Operator

An example of Hermitian operator is the so-called 'projection operator' or 'projector'. Given an orthonormal basis $\{|\mathrm{x}_n^i\rangle\} \in \mathcal{H}$ of an operator X with eigenvalue x_n, the projector in the subspace corresponding to the eigenvalue x_n is given by

$$X_n = \sum_{i=1}^{g_n} |\mathrm{x}_n^i\rangle \langle \mathrm{x}_n^i| \, . \tag{2.1.10}$$

Henceforth, the convention is that X, Y, ... denote generic operators whereas X_n, Y_m, ... denote projectors. X_n is called a projector because it projects any vector $|\psi\rangle$ onto the subspace defined by the X-eigenbasis $\{|\mathrm{x}_n^i\rangle\} \in \mathcal{H}$.

Any projector has the following properties:

i) **Idempotency**: $$X_n^2 = X_n. \tag{2.1.11}$$

ii) **Completeness**: $$\sum_n X_n = \mathbb{1}. \tag{2.1.12}$$

iii) **Orthogonality**: $$X_n X_m = X_m X_n = X_n \delta_{nm}. \tag{2.1.13}$$

Pauli Matrices

An important example of Hermitian operators are the Pauli matrices; they are 2×2 complex matrices and also *generators* of the $SU(2)$ group, written in the form

$$\sigma_x := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.1.14}$$

They have the following properties:

i) $$\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = -i\sigma_x\sigma_y\sigma_z = \mathbb{1};\tag{2.1.15}$$

ii) $$\det(\sigma_i) = -1, \qquad \mathrm{Tr}(\sigma_i) = 0;\tag{2.1.16}$$

iii) $$[\sigma_i, \sigma_j] = 2i\varepsilon_{ijk}\sigma_k, \qquad \{\sigma_i, \sigma_j\}_+ = 2\delta_{ij}\mathbb{1};\tag{2.1.17}$$

where ε_{ijk} is the Levi-Civita symbol.

Other useful properties, which can be derived from the previous ones are:

$$\sigma_i\sigma_j = \delta_{ij}\mathbb{1} + i\varepsilon_{ijk}\sigma_k,\tag{2.1.18}$$

and

$$(\boldsymbol{a}\cdot\boldsymbol{\sigma})(\boldsymbol{b}\cdot\boldsymbol{\sigma}) = (\boldsymbol{a}\cdot\boldsymbol{b})\mathbb{1} + i(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{\sigma},\tag{2.1.19}$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ and $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3$. Notice that the Pauli matrices in equation (2.1.14) together with the identity $\mathbb{1}$ form a complete set of operators in $SU(2)$, such that any 2×2 matrix representing the operator X can be written in terms of them, that is

$$X \doteq c_0\mathbb{1} + c_1\sigma_x + c_2\sigma_y + c_3\sigma_z = \begin{pmatrix} c_0 + c_3 & c_1 - ic_2 \\ c_1 + ic_2 & c_0 + c_3 \end{pmatrix}\tag{2.1.20}$$

where $c_i \in \mathbb{C}$.

2.1.4 Quantum State Generalization

Pure Density Matrix

Representing quantum states as state vector $|\psi(t)\rangle$ is not the most general formulation. An alternative mathematical description of the quantum state is the so-called 'density matrix' or 'density operator' $\rho(t)$. Instead of representing a quantum state as a vector one can represent a quantum state as a matrix written in terms of a state vector. Although mathematically different, state vectors and density operators are equivalent to pure ensembles. The definition of this density operator is

$$\rho := |\psi\rangle\langle\psi|.\tag{2.1.21}$$

A quantum state is called 'pure' if all elements of an ensemble of a given system are prepared in the same state $|\psi_i\rangle$, and the ensemble is represented by (2.1.21). However, we do not always have full information about the state. A property that only pure states satisfy is $\mathrm{Tr}(\rho^2) = 1$, because a pure density operator is a projector, which is idempotent.

Consider the state vector of a qubit written in equation (2.1.2), as a simple example. Its pure ensemble density operator which describes a qubit in a superposition of the normalized orthonormal vectors $|0\rangle$ *and* $|1\rangle$ can be written as

$$\rho = |\psi\rangle\langle\psi| = |c_0|^2|0\rangle\langle 0| + c_0c_1^*|0\rangle\langle 1| + c_1c_0^*|1\rangle\langle 0| + |c_1|^2|1\rangle\langle 1|,\tag{2.1.22}$$

where $|0\rangle\langle 1|$ and $|1\rangle\langle 0|$ are off-diagonal terms responsible for interference effects, $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$ are diagonal terms responsible for the populations, which can be clearly identified in matrix representation,

$$\rho \doteq \begin{pmatrix} |c_0|^2 & c_0c_1^* \\ c_0^*c_1 & |c_1|^2 \end{pmatrix}.\tag{2.1.23}$$

The probabilities are encoded in the coefficients $|c_0|^2$ and $|c_1|^2$, they are related to "populations" (diagonal terms). The off-diagonal terms $c_0c_1^*$ and $c_0^*c_1$ are called "quantum coherences", terms that carries the quantumness of the state differing it from the classical ones.

Mixed Density Matrix

Not always quantum systems are described by pure states. One can have 'statistical mixtures' as $\{\wp_i, |\psi_i\rangle\}$, the so-called 'mixed' states. Consider an ensemble consisting of N copies of a system where n_i systems are in the state $|\psi_i\rangle$, with $N = \sum_i n_i$. The probability of finding a given system in the state $|\psi\rangle$ is given by $\wp_i = n_i/N$ (also called statistical weight), where $\sum_i \wp_i = 1$. So, we can write the mixed ensemble density operator as a convex sum (classical superposition) of pure ensemble density operators ρ_i as

$$\rho_{mix} = \sum_i \wp_i \rho_i = \sum_i \wp_i |\psi_i\rangle \langle\psi_i| , \tag{2.1.24}$$

where the density operators ρ_{mix} stands for mixed ensembles and ρ_i for pure ensembles.[4]

Consider an ensemble with probability $\wp$ of having a qubit in the pure state $|0\rangle$ and $(1-\wp)$ for the pure state $|1\rangle$. The density operator for such a mixed state reads

$$\rho = \sum_i \wp_i |\psi_i\rangle \langle\psi_i| = \wp |0\rangle \langle 0| + (1-\wp) |1\rangle \langle 1| . \tag{2.1.25}$$

Using matrix representation, we clearly see that this state has no quantum coherence in the computational basis:

$$\rho \doteq \begin{pmatrix} \wp & 0 \\ 0 & 1-\wp \end{pmatrix} . \tag{2.1.26}$$

Example: One can ask what the difference is between pure and mixed states. Consider, as an illustrative example, the following three scenarios involving an ensemble with N copies of the same system.

- *First case* – Pure ensemble for N copies in state $|0\rangle$:

$$\rho = |0\rangle \langle 0| . \tag{2.1.27}$$

- *Second case* – Pure ensemble for N copies in state $|\psi\rangle = \sqrt{\frac{n_0}{N}} |0\rangle + e^{i\phi} \sqrt{\frac{n_1}{N}} |1\rangle$:

$$\rho = \frac{n_0}{N} |0\rangle \langle 0| + \frac{n_1}{N} |1\rangle \langle 1| + \frac{\sqrt{n_1 n_0}}{N} \left(e^{i\phi} |1\rangle \langle 0| + e^{-i\phi} |0\rangle \langle 1| \right) . \tag{2.1.28}$$

- *Third case* – Mixed ensemble for N copies, n_0 in state $|0\rangle$ and n_1 in state $|1\rangle$:

$$\rho = \frac{n_0}{N} |0\rangle \langle 0| + \frac{n_1}{N} |1\rangle \langle 1| \neq |\psi\rangle \langle\psi| . \tag{2.1.29}$$

The term in parenthesis in equation (2.1.28) is responsible for the quantum coherence of ρ in the computational basis. Also, it is noteworthy that only pure superpositions can encode relative phases (ϕ), which are measurable in interference experiments. For this reason, we cannot always represent quantum states by vector states.

In conclusion, the vector-based description of a physical state only applies when one has full knowledge of the preparation scheme (pure ensemble). In general, though, the density operator approach is the only way of considering statistical ignorance (further than the fundamental quantum fluctuations). In addition, mixed states also emerge when part of a multipartite system (the environment) is discarded, which is always the case in experiments.

[4] Henceforth both density operators will be called a density operator, physical state or simply 'state', as one can obtain pure ensemble density operators through mixed ensemble density operators.

Properties of Physical States

Every density operator must obey the following properties to represent a physical state:

i) **Hermiticity**: $\rho^\dagger = \rho$.

ii) **Normalization**: $\mathrm{Tr}\,\rho = 1$.

iii) **Positivity**: $\langle\psi|\rho|\psi\rangle \geq 0, \quad (\forall\ |\psi\rangle \neq 0)$ [5].

For every pure ensemble, we have that $\mathrm{Tr}\,\rho^2 = 1$. In fact, $\mathrm{Tr}(\rho^2)$ is a measure of the purity of the state ρ. And the mean value of an operator X is given by

$$\langle X\rangle = \mathrm{Tr}(X\rho) = \sum_i \wp_i \langle\psi_i|X|\psi_i\rangle\,. \tag{2.1.30}$$

Composite States

For single Hilbert space $\mathcal{H}$, the quantum state is given by equation (2.1.24). However one can consider a system consisting of two parts, each one with their own Hilbert spaces, *i.e.*, the part $\mathcal{X}$ is associated with the Hilbert space $\mathcal{H}_\mathcal{X}$ and the part $\mathcal{Y}$ associated to $\mathcal{H}_\mathcal{Y}$. So, the system Hilbert space, or the *composite space*, is the tensor product $\mathcal{H}_{\mathcal{X}\mathcal{Y}} = \mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{Y}$. Then, the density operator of the system is denoted $\rho \equiv \rho_{\mathcal{X}\mathcal{Y}} \in \mathscr{B}(\mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{Y})$. If the joint system is prepared in a pure state, one can write

$$|\psi\rangle = \sum_{i,j} c_{i,j}\, |\psi_i\rangle_\mathcal{X} \otimes |\phi_j\rangle_\mathcal{Y}\,, \tag{2.1.31}$$

where $\{|\psi_i\rangle_\mathcal{X}\}$ is a basis in $\mathcal{H}_\mathcal{X}$ and $\{|\phi_j\rangle_\mathcal{Y}\}$ is a basis in $\mathcal{H}_\mathcal{Y}$. The pure state written in terms of equation (2.1.31) is

$$\rho_{\mathcal{X}\mathcal{Y}} = |\psi\rangle\langle\psi|_{\mathcal{X}\mathcal{Y}} = \sum_{i,j,k,l} c_{i,j}c^*_{k,l}\, |\psi_i\rangle\langle\psi_k|_\mathcal{X} \otimes |\phi_j\rangle\langle\phi_l|_\mathcal{Y}\,, \tag{2.1.32}$$

which acts on $\mathcal{H}_{\mathcal{X}\mathcal{Y}}$. The mixed state, we have

$$\rho_{\mathcal{X}\mathcal{Y}} = \sum_{i,j} \wp_{i,j}\, |\psi_i\rangle\langle\psi_i|_\mathcal{X} \otimes |\phi_j\rangle\langle\phi_j|_\mathcal{Y}\,, \tag{2.1.33}$$

where $\wp_{i,j}$ is a probability distribution, $\rho^i_\mathcal{X} = |\psi_i\rangle\langle\psi_i|_\mathcal{X} \in \mathscr{B}(\mathcal{H}_\mathcal{X})$ and $\rho^j_\mathcal{Y} = |\phi_j\rangle\langle\phi_j|_\mathcal{Y} \in \mathscr{B}(\mathcal{H}_\mathcal{Y})$.

Reduced Density Operator

An important concept when only part of the joint system is accessible is the ‘reduced density operator’ or ‘reduce state’. As previously, suppose one has a system of two parts with space $\mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{Y}$ and state $\rho_{\mathcal{X}\mathcal{Y}}$. When one is interested in part $\mathcal{X}$ only or have no access to the part $\mathcal{Y}$, a density operator $\rho_\mathcal{X}$ can be obtained via partial trace over the space $\mathcal{H}_\mathcal{X}$. In this way, one ensures that the remaining part is only an element of $\mathscr{B}(\mathcal{H}_\mathcal{X})$. Thereby, a reduced density operator belonging to $\mathscr{B}(\mathcal{H}_\mathcal{X})$ is written as

$$\rho_\mathcal{X} = \mathrm{Tr}_\mathcal{Y}(\rho_{\mathcal{X}\mathcal{Y}}). \tag{2.1.34}$$

[5] The inequality is only satisfied when ρ is written in a basis that is orthogonal to the basis of $|\psi\rangle \in \mathcal{H}$.

If one is interested in the converse, *i.e.*, in a reduced density operator belonging to $\mathscr{B}(\mathcal{H}_{\mathcal{Y}})$, one must trace the $\mathcal{H}_{\mathcal{X}}$ part, which gives

$$\rho_{\mathcal{Y}} = \mathrm{Tr}_{\mathcal{X}}(\rho_{\mathcal{X}\mathcal{Y}}). \tag{2.1.35}$$

As an example, let us consider the singlet state

$$|s\rangle = \frac{|0\rangle_{\mathcal{X}} |1\rangle_{\mathcal{Y}} - |1\rangle_{\mathcal{X}} |0\rangle_{\mathcal{Y}}}{\sqrt{2}}. \tag{2.1.36}$$

The density operator is

$$\rho_{\mathcal{X}\mathcal{Y}} = |s\rangle\langle s| = \frac{1}{2}\Big(|0\rangle\langle 0|_{\mathcal{X}} \otimes |1\rangle\langle 1|_{\mathcal{Y}} + |1\rangle\langle 1|_{\mathcal{X}} \otimes |0\rangle\langle 0|_{\mathcal{Y}} - |0\rangle\langle 1|_{\mathcal{X}} \otimes |1\rangle\langle 0|_{\mathcal{Y}} - |1\rangle\langle 0|_{\mathcal{X}} \otimes |0\rangle\langle 1|_{\mathcal{Y}} \Big). \tag{2.1.37}$$

Tracing $\mathcal{H}_{\mathcal{Y}}$ we obtain the reduced density operator

$$\rho_{\mathcal{X}} = \mathrm{Tr}_{\mathcal{Y}}(\rho_{\mathcal{X}\mathcal{Y}}) = \frac{|0\rangle\langle 0|_{\mathcal{X}} + |1\rangle\langle 1|_{\mathcal{X}}}{2} = \frac{\mathbb{1}}{2}, \tag{2.1.38}$$

and tracing $\mathcal{H}_{\mathcal{X}}$ we obtain

$$\rho_{\mathcal{Y}} = \mathrm{Tr}_{\mathcal{X}}(\rho_{\mathcal{X}\mathcal{Y}}) = \frac{|0\rangle\langle 0|_{\mathcal{Y}} + |1\rangle\langle 1|_{\mathcal{Y}}}{2} = \frac{\mathbb{1}}{2}. \tag{2.1.39}$$

2.1.5 Quantum Measurements

In quantum theory, possible values obtained from a measurement of a physical quantity are the eigenvalues of the observable representing that quantity. However, when operators were introduced in section 2.1.3, it was stated that operators act on Hilbert spaces, *i.e.*, they act on complex vectors that can produce complex eigenvalues. But in laboratories, we do not obtain complex numbers as measurement results. So, the following postulates are restrictions to the wide range of operators to consider only those which describe real physical measurements.

Observables

The only possible way to investigate quantum states is through measurements. So, the next definition introduces observables

Definition 4. (Observable) *Observable is a Hermitian operator X whose eigenvectors form an orthonormal basis that span the state space in which that observable exists.*

The following postulate is about what describes physical measurements

Postulate 2. (Measurements) – *Every measurable physical quantity χ is described by an Hermitian operator whose basis span $\mathcal{H}$ which X acts on and is called 'observable'.*

We have seen that projectors take a vector state and project it into another vector. If we have an observable X whose eigenstates are $\{|x_i\rangle\}$ with eigenvalues x_i, using the properties of projectors then

$$X = X\mathbb{1} = X\sum_{i=1}^{d} X_i = \sum_{i=1}^{d} XX_i = \sum_{i=1}^{d} x_i X_i. \tag{2.1.40}$$

This procedure to write observables in terms of its eigenvalues and projectors is called 'discrete' *spectral decomposition* of the observable X.

Now we prove an important theorem concerning commuting observables.

Theorem 1. *Two observables X and Y can be decomposed on the same eigenstates basis $\{|x_i, y_j\rangle\}$ if and only if they commute* ($[X, Y] = 0$).

Proof. The eigenvalues of X and Y are denoted by x_i and y_j, respectively. Consider a state $|\psi\rangle$ whose basis is the eigenstates $\{|x_i, y_j\rangle\}$ similarly to equation (2.1.1). Thus, we can write

$$\begin{aligned} XY|\psi\rangle &= X\sum_i c_i Y|x_i, y_j\rangle = \sum_i c_i y_j X|x_i, y_j\rangle \\ &= \sum_i c_i x_i y_j |x_i, y_j\rangle = Y\sum_i c_i x_i |x_i, y_j\rangle \\ &= YX\sum_i c_i |x_i, y_j\rangle = YX|\psi\rangle . \end{aligned} \tag{2.1.41}$$

We proved that if the X and Y have the same basis decomposition, they commute. Now to prove that if X and Y commute, they must have the same basis decomposition, we consider a generic vector $|\psi\rangle$. It follows that

$$YX|\psi\rangle = XY|\psi\rangle = X(Y|\psi\rangle) = x_i(Y|\psi\rangle). \tag{2.1.42}$$

So, $|\psi\rangle$ is an eigenvector of X and also $Y|\psi\rangle$, such that the action of Y preserves the eigenspace of X. □

Results of Measurements

Not every mathematical solution is physically possible. Thus, the following postulate is about possible results of measurements,

Postulate 3. (Possible results) – *The measurement 'possible results' of χ are the eigenvalues of the observable X.*

Probability of Measurement Outputs

When doing measurements one can obtain probabilities of getting eigenvalues of an observable. For this reason, the postulate bellow is about measurement probabilities for finite-dimensional Hilbert space,

Postulate 4. (Probability) – *If $X|\mathrm{x}_n^i\rangle = \mathrm{x}_n|\mathrm{x}_n^i\rangle$, where $i = 1, 2, \ldots, g_n$, the 'probability' of obtaining x_n is given by $\mathbb{p}_{x_n} \equiv \mathbb{p}(\mathrm{x}_n) = \mathrm{Tr}(X_n\rho)$, where X_n is the 'projector' and ρ is the density operator.*

Although $\mathbb{p}_{x_n}$ is a '*quantum*' probability, in the sense, that it comes from a quantum state, it can also result in a 'classical' probability, which is the populations of the quantum state $\wp_n$. In the next section, more details are provided concerning the differences between "quantum" and classical probabilities.

States After Measurements

It has been postulated the possible results of a measurement, but the next postulate provides a picture to what happens to the state afterward, the famous collapse postulate.

Postulate 5. (State collapse) – *After obtained x_n in a measurement of χ, the state $|\psi\rangle$ collapses to a normalized state* $\frac{X_n|\psi\rangle}{\sqrt{\mathbb{P}(x_n)}}$.

After a measurement, the quantum state collapses to a projection of the eigenbasis $\{|\mathrm{x}_n^i\rangle\} \in \mathcal{H}$. The postulate 5 can also be written in terms of density operators as

$$\rho_n' = \frac{X_n \rho X_n}{\sqrt{\mathrm{Tr}(X_n \rho X_n)}} \tag{2.1.43}$$

where ρ_n' is the state after a measurement of X.

Dynamics of Quantum States

The dynamics of quantum states is postulated next, the also famous Schrödinger equation, which describes the quantum state time evolution.

Postulate 6. (Time evolution of states) – *A time evolution of state vector $|\psi(t)\rangle$ is governed by the 'Schrödinger equation' $i\hbar\frac{d}{dt}|\psi(t)\rangle = H(t)|\psi(t)\rangle$, where $H(t)$ is the 'Hamiltonian operator' of the system.*

The solution of the Schrödinger equation predicts what happens to the initial state after some time t and also the experimentally verifiable objects, such as probability distributions and expectation values. The Schrödinger equation is the time evolution of the vector state $|\psi\rangle$, but one can consider the time evolution of the density operator ρ, in this case, the equation that describes the density operator time evolution is the *von Neumann equation*, written as

$$i\hbar\frac{\partial \rho}{\partial t} = [H, \rho]. \tag{2.1.44}$$

2.1.6 Conditional State

Consider the following two types of systems.

Single-Partite Systems

Consider a quantum system described by a density operator ρ written in terms of vectors in $\mathcal{H}$. After a measurement of an observable X, the system is in an eigenstate of X, namely $\{|x_i\rangle\}$. Therefore, the system was restricted to a particular basis, but it does not prevent one to make other measurements, e.g., of an observable Y. The post-collapse state $\rho = |x_i\rangle\langle x_i|$ is called $\rho_{S|x_i}$, it is often interesting to call the state 'conditioned state to x_i' to emphasize what observable was measured with which output and is written as

$$\rho_{S|x_i} := \frac{X_i \rho X_i}{\mathrm{Tr}(X_i \rho X_i)}, \tag{2.1.45}$$

here, the 'x_i' subscript means that the system was measured by the observable X whose projector is X_i, *i.e.*, reinforcing the idea that the system is collapsed in one of the eigenstates of X. And the subscript 'S' stands for the rest of the system that has not collapsed; otherwise, it would not be a state.

We can show that ρ after the collapse is the state $\rho_{S|x_i}$ is, in fact, an eigenstate of X

$$\begin{aligned}\rho_{S|x_i} &= \frac{X_i \rho X_i}{\mathrm{Tr}(X_i \rho X_i)} \\ &= \frac{|x_i\rangle\langle x_i|\psi\rangle\langle\psi|x_i\rangle\langle x_i|}{\mathrm{Tr}(|x_i\rangle\langle x_i\psi\rangle\langle\psi|x_i\rangle\langle x_i|)} \\ &= \frac{|\langle x_i|\psi\rangle|^2 |x_i\rangle\langle x_i|}{|\langle x_i|\psi\rangle|^2} \\ &= |x_i\rangle\langle x_i| \\ &= X_i.\end{aligned} \tag{2.1.46}$$

Thus, $\rho_{S|x_i} = X_i$ is actually a projector in the subspace corresponding to the eigenvalue $x_i]$.

Composite Systems

Given a bipartite state $\rho \in \mathscr{B}(\mathcal{H}_{\mathcal{X}} \otimes \mathcal{H}_{\mathcal{Y}})$, a measurement of observable X yields the collapsed state (2.1.45), which upon partial trace, gives

$$\rho_{Y|x_i} := \mathrm{Tr}_{\mathcal{X}}\left[\frac{X_i\rho}{\mathrm{Tr}(X_i\rho)}\right] = \frac{\mathrm{Tr}_{\mathcal{X}}(X_i\rho)}{\mathbb{p}_{x_i}}, \tag{2.1.47}$$

where $\mathbb{p}_{x_i} = \mathrm{Tr}(X_i\rho)$, accordingly to postulate 4.

On the other hand in the opposite situation, *i.e.*, if one measures Y first and then X, then the conditional state is now given by

$$\rho_{X|y_j} := \frac{\mathrm{Tr}_{\mathcal{Y}}(Y_j\rho)}{\mathrm{Tr}(Y_j\rho)} = \frac{\mathrm{Tr}_{\mathcal{Y}}(Y_j\rho)}{\mathbb{p}_{y_j}}. \tag{2.1.48}$$

Examples

We provide two simple but intuitive examples of how to calculate conditional states.

1. Single systems: Suppose Alice has a Stern-Gerlach apparatus in the z-axis direction in her laboratory and also Bob in his laboratory. They agree that Alice prepares a system with spin 1/2 in a particular state and makes the first measurement, then send the resulting system to Bob, so that he can make his measurements. Afterwards, she prepares a particle in the pure state

$$|\psi\rangle = \alpha|+\rangle + \beta|-\rangle, \tag{2.1.49}$$

whose density operator is written as

$$\rho = |\alpha|^2 |+\rangle\langle+| + \alpha\beta^* |+\rangle\langle-| + \beta\alpha^* |-\rangle\langle+| + |\beta|^2 |-\rangle\langle-|. \tag{2.1.50}$$

After a measurement of the observable σ, Alice obtains +1/2 (spin up) or −1/2 (spin down). Thus, she obtains a pure state conditioned to her measurement outcome. For instance, suppose that Alice obtains +1/2 in her laboratory and then, send the state to Bob‘s laboratory, the state that she sent is the conditional state written as

$$\rho_{S|+} = \frac{\sigma_+\rho\sigma_+}{\mathrm{Tr}(\sigma_+\rho\sigma_+)} = \frac{|+\rangle\langle+|\rho|+\rangle\langle+|}{\langle+|\rho|+\rangle} = |+\rangle\langle+|. \tag{2.1.51}$$

where we simplified it using the properties of the trace and replaced equation (2.1.50). Therefore, the resulting state is an eigenstate of σ_z (the Pauli matrix for the z-direction). Something expected, since $|\psi\rangle$ collapses to one of the eigenstates of σ, which is $|+\rangle$, in this case. So, if Alice obtains +1/2, Bob receives a state of a particle with spin up with 100% of resulting +1/2, since his state is conditioned to Alice's measurements.

2. Composite systems: Suppose again the same scenario, that Alice and Bob have Stern-Gerlach apparatuses in z-axis in their laboratories and they agree that she makes the first measurement. The system to be measured is two particles with spin 1/2 described by a singlet state given by equation (2.1.37) that describes two entangled particles emitted from a source. Alice receives one particle and make the first measurement σ^A. Bob receives the other particle after Alice makes her measurement. The task is to write the state that Bob receives, since his state is again conditioned to Alice's measurements. We can write the state of the two particles emitted from the source as

$$\begin{aligned}\rho = &\frac{1}{2}\Big(|+\rangle\langle+|_{\mathcal{A}} \otimes |-\rangle\langle-|_{\mathcal{B}} + |-\rangle\langle-|_{\mathcal{A}} \otimes |+\rangle\langle+|_{\mathcal{B}}\Big)\\ &-\frac{1}{2}\Big(|+\rangle\langle-|_{\mathcal{A}} \otimes |-\rangle\langle+|_{\mathcal{B}} - |-\rangle\langle+|_{\mathcal{A}} \otimes |+\rangle\langle-|_{\mathcal{B}}\Big),\end{aligned} \tag{2.1.52}$$

where, $\mathcal{A}$ and $\mathcal{B}$ stands for the spaces where Alice's and Bob's observable operate, respectively. If Alice's result is +1/2, Bob receives the conditional state

$$\rho_{\mathcal{B}|+} = \frac{\mathrm{Tr}_{\mathcal{A}}\left(\sigma_+^{\mathcal{A}}\rho\right)}{\mathrm{Tr}\left(\sigma_+^{\mathcal{A}}\rho\right)} = \frac{\mathrm{Tr}_{\mathcal{A}}\left(|+\rangle\langle+|_{\mathcal{A}}\,\rho\right)}{\langle+|\rho|+\rangle_{\mathcal{A}}}, \tag{2.1.53}$$

or, simplifying it

$$\rho_{\mathcal{B}|+} = \frac{\frac{1}{2}\mathrm{Tr}_{\mathcal{A}}\left(|+\rangle\langle+|_{\mathcal{A}} \otimes |-\rangle\langle-|_{\mathcal{B}} - |+\rangle\langle-|_{\mathcal{A}} \otimes |-\rangle\langle+|_{\mathcal{B}}\right)}{1/2} = |-\rangle\langle-|_{\mathcal{B}}, \tag{2.1.54}$$

That is, if Alice's measurement gives spin up, Bob will receive an electron with spin down with 100% of chance.

2.2 Generalized Bloch Representation

We have seen that two-level systems can be represented by qubits which are associated with a 2-dimensional Hilbert space. However, one can have systems with N levels represented by d-dimensional Hilbert space. We will introduce the Bloch sphere first, which is a representation for qubits and then its generalization for d-dimensional spaces.

2.2.1 Bloch Sphere

Consider a system described by a 2-dimensional Hilbert space prepared in the state

$$|\psi\rangle = c_0|0\rangle + c_1 e^{i\theta}|1\rangle. \tag{2.2.1}$$

States

Since $\{\mathbb{1}, \boldsymbol{\sigma}\}$ form a complete set of operators that acts on $\mathcal{H} \simeq \mathbb{C}^2$, we can write the two-level system state as

$$\rho = \frac{\mathbb{1} + \boldsymbol{r} \cdot \boldsymbol{\sigma}}{2}, \tag{2.2.2}$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the Pauli matrices in vector notation, $\boldsymbol{r}$ is the Bloch vector, a vector with any direction in $\mathbb{R}^3$ with $|r| \in [0, 1]$. So, $\boldsymbol{r} \cdot \boldsymbol{\sigma} = \sum_{m=1}^{3} r_m \sigma_m$, where $\sigma_1 = \sigma_x$, $\sigma_2 = \sigma_y$ and $\sigma_3 = \sigma_z$, the (2.2.2) maps every ρ into one vector $\boldsymbol{r}$, therefore, the set of all physical states are mapped in a solid sphere, the *Bloch sphere*. When $r = 1$ we have a pure state, *i.e.*, pure states live in the surface of the Bloch sphere, since

$$\mathrm{Tr}(\rho^2) = \frac{1 + r}{2} \tag{2.2.3}$$

therefore, in this case, $\mathrm{Tr}(\rho^2) = 1$. When $r < 1$, ρ is mixed state, *i.e.*, mixed states live inside the Bloch sphere, since $\mathrm{Tr}(\rho^2) < 1$.

Equation (2.2.2) can be written in the matrix form as

$$\rho \doteq \frac{1}{2} \begin{pmatrix} 1 + r_z & r_x - i r_y \\ r_x + i r_y & 1 - r_z \end{pmatrix}. \tag{2.2.4}$$

Equations (2.2.2) and (2.2.4) are similar to (2.1.20).

In spherical coordinates, we can write $\boldsymbol{r} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$, where $|\boldsymbol{r}| = 1$, $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$, thus

$$\rho \doteq \begin{pmatrix} \cos^2(\theta/2) & e^{-i\phi}\sin(\theta/2)\cos(\theta/2) \\ e^{i\phi}\sin(\theta/2)\cos(\theta/2) & \sin^2(\theta/2) \end{pmatrix}. \tag{2.2.5}$$

Observables

In the particular case where one has a 2-dimensional Hilbert space associated with a two-level system, we can write the spectral decomposition of the observable X by equation (2.1.20), considering a simple case where $\mathrm{Tr}\, X = 0$, as well as for spin observables, then we can write

$$X = x_1\sigma_1 + x_2\sigma_2 + x_3\sigma_3 = \sum_{k=1}^{3} x_k \sigma_k, \tag{2.2.6}$$

with the condition that $x_1^2 + x_2^2 + x_3^2 = 1$, points on unit sphere. We can also write it as

$$X = \hat{\boldsymbol{x}} \cdot \boldsymbol{\sigma}, \tag{2.2.7}$$

such that X has the eigenvalue equation $X\,|x_k\rangle = k\,|x_k\rangle$, where $k = \pm 1$. So, the observable admits the spectral decomposition[6]

$$X = \sum_{k=-1,+1} k\, X_k, \tag{2.2.8}$$

and projectors

$$X_k = \frac{\mathbb{1} + k\,\hat{\boldsymbol{x}} \cdot \boldsymbol{\sigma}}{2}. \tag{2.2.9}$$

It is trivial to verify that (2.2.9) is a projector. Also, putting it in equation (2.2.8) results in (2.2.7).

[6]Where we denote the eigenvalues in the set $\{-1, +1\}$ rather than $\{1, 2\}$.

2.2.2 Generalized Pauli Matrices

To introduce the next topic, we have to introduce a generalization of the Pauli matrices, which are only valid to systems with two degrees of freedom. The main reference to this subject is [Aerts (2014)].

Consider a $d \times d$ unitary[7] matrix U with determinant +1. Such matrix belongs to the $SU(d)$ group. Consider the matrix U that has $d^2 - 1$ generators Λ_i and $d^2 - 1$ real parameters $u_i = (u_1, u_2, \ldots, u_{d^2-1})$, thus we can write in vector notation,

$$U = e^{i\boldsymbol{u}\cdot\boldsymbol{\Lambda}}. \tag{2.2.10}$$

As the Pauli matrices, Λ_i has similar properties:

i) Null trace: $$\mathrm{Tr}(\Lambda_i) = 0 \quad (\forall i); \tag{2.2.11}$$

ii) Product trace: $$\mathrm{Tr}(\Lambda_i \Lambda_j) = 2\delta_{ij}; \tag{2.2.12}$$

iii) Commutator: $$[\Lambda_i, \Lambda_j] = 2i \sum_{k=1}^{d^2-1} f_{ijk}\Lambda_k; \tag{2.2.13}$$

iv) Anti-commutator: $$\{\Lambda_i, \Lambda_j\}_+ = \frac{4}{d}\delta_{ij}\mathbb{1} + 2\sum_{k=1}^{d^2-1} d_{ijk}\Lambda_k. \tag{2.2.14}$$

where f_{ijk} is an anti-symmetric tensor and d_{ijk} is a symmetric tensor. These tensors are also called *structure constants* of the algebra and are given by

$$f_{ijk} = \frac{1}{4i}\,\mathrm{Tr}\left([\Lambda_i, \Lambda_j]\Lambda_k\right), \tag{2.2.15}$$

and

$$d_{ijk} = \frac{1}{4}\,\mathrm{Tr}\left(\{\Lambda_i, \Lambda_j\}_+\Lambda_k\right), \tag{2.2.16}$$

Therefore, for a given dimension d there are different f_{ijk} and d_{ijk} in kind and in number.

Examples

Two examples follow, one for the algebra of the group $SU(2)$ whose generators are the Pauli matrices which describe qubits, thus $\Lambda_1 = \sigma_x$, $\Lambda_2 = \sigma_y$ and $\Lambda_3 = \sigma_z$. The other for $SU(3)$ whose generators are the Gell-Mann matrices which describe qutrits (system with three degrees of freedom), thus $\Lambda_1 = \lambda_1$, $\Lambda_2 = \lambda_2$, $\Lambda_3 = \lambda_3$, $\Lambda_4 = \lambda_4$, $\Lambda_5 = \lambda_5$, $\Lambda_6 = \lambda_6$, $\Lambda_7 = \lambda_7$ and $\Lambda_8 = \lambda_8$.

1. Qubits: In this case, $d = 2$, we have 3 generators Λ_i's which are called Pauli matrices and 3 u_i's which form a 3-dimensional Bloch vector. Their properties are given by equations (2.1.15)–(2.1.19).

[7] Unitary operators must satisfy the condition: $UU^\dagger = U^\dagger U = \mathbb{1}$.

2. Qutrits: In this case $d = 3$, we have 8 generators Λ_i's which are the *Gell-Mann matrices* and 3 u_i's which consist of a 8-dimensional Bloch vector. They are given by:

$$\begin{aligned} &\lambda_1 := \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 := \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \\ &\lambda_4 := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \lambda_5 := \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \lambda_6 := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \\ &\lambda_7 := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 := \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \end{aligned} \tag{2.2.17}$$

2.2.3 Bloch Body

We introduce the *Bloch body*, which is a generalization of Bloch sphere, as named by [Appleby (2007)], for a d-dimensional object. The authors in reference [Aerts (2014)] introduce a generalized Bloch representation. In this section, we will give an introduction to some highlights which are convenient for us.

States

The complete set that acts in $\mathcal{H} \simeq \mathbb{C}^d$ is $\{\mathbb{1}, \Lambda_1, \Lambda_2, \ldots, \Lambda_{d^2-1}\}$. Therefore, we can write a generic state as

$$\rho = \frac{\mathbb{1} + C_d \, \boldsymbol{r} \cdot \boldsymbol{\Lambda}}{d}, \tag{2.2.18}$$

with $\boldsymbol{r} = \sum_{i=1}^{d^2-1} r_i \hat{\boldsymbol{e}}_j \in \mathbb{R}^{d^2-1}$, $\{\hat{\boldsymbol{e}}_j\}$ is an orthonormal basis and C_d is a quantity that in *Grassmann algebra*[8] determines the dimension of the vector space of the *exterior product* of $\hat{\boldsymbol{e}}_j$, it is given by

$$C_d = \sqrt{\frac{d(d-1)}{2}}. \tag{2.2.19}$$

It is noteworthy that not all states are represented by (2.2.18) for $d \geq 3$.

By taking the sum $(\rho\Lambda_i + \Lambda_i\rho)/2$, then, its trace, one obtains the components of $\boldsymbol{r}$ which are

$$r_i = \sqrt{\frac{d}{2(d-1)}} \operatorname{Tr}(\rho\Lambda_i). \tag{2.2.20}$$

Generalized Vector Products

In $\mathbb{R}^d$ the inner product between two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is given by

$$\boldsymbol{a} \cdot \boldsymbol{b} = \sum_{i,j=1}^{d^2-1} \delta_{ij} a_i b_j, \tag{2.2.21}$$

[8]Also known as 'exterior algebra', it is the algebraic system whose product is the exterior product, the product between vectors in higher dimensional spaces which forms volumes and areas, cross product is a particular case of exterior product.

and the k-th component of the cross product between $\boldsymbol{a}$ and $\boldsymbol{b}$ is given by

$$(\boldsymbol{a} \times \boldsymbol{b})_k = \sum_{i,j=1}^{d^2-1} \varepsilon_{ijk} a_i b_j. \tag{2.2.22}$$

However for higher-dimensional vector space, for instance, in $\mathbb{R}^{d^2-1}$, it is required a generalization of those products. In $\mathbb{R}^{d^2-1}$, one has the so-called *wedge product*,

$$(\boldsymbol{a} \wedge \boldsymbol{b})_k := \sum_{i,j=1}^{d^2-1} f_{ijk} a_i b_j. \tag{2.2.23}$$

which is an anti-symmetric product.

Consider now a basis with two unit vectors

$$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{2.2.24}$$

If two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^2$ are spanned by this basis

$$\boldsymbol{u} = u_1 \boldsymbol{e}_1 + u_2 \boldsymbol{e}_1 \tag{2.2.25}$$

and

$$\boldsymbol{v} = v_1 \boldsymbol{e}_1 + v_2 \boldsymbol{e}_1. \tag{2.2.26}$$

The area of the parallelogram formed by $\boldsymbol{u}$ and $\boldsymbol{v}$ is

$$A := |\det(\boldsymbol{u}\boldsymbol{v})| = \left| \det \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix} \right| = |u_1 v_2 - v_1 u_2|, \tag{2.2.27}$$

and the wedge product of $\boldsymbol{u}$ and $\boldsymbol{v}$ is

$$\begin{aligned} \boldsymbol{u} \wedge \boldsymbol{v} &= (u_1 \boldsymbol{e}_1 + u_2 \boldsymbol{e}_1) \wedge (v_1 \boldsymbol{e}_1 + v_2 \boldsymbol{e}_1) \\ &= u_1 v_1 \, \boldsymbol{e}_1 \wedge \boldsymbol{e}_1 + u_1 v_2 \, \boldsymbol{e}_1 \wedge \boldsymbol{e}_2 + u_2 v_1 \, \boldsymbol{e}_2 \wedge \boldsymbol{e}_1 + u_2 v_2 \, \boldsymbol{e}_2 \wedge \boldsymbol{e}_2. \end{aligned} \tag{2.2.28}$$

But for any *bivector* which is defined as $\boldsymbol{e}_i \wedge \boldsymbol{e}_j$, we have the properties:

i)
$$\boldsymbol{e}_i \wedge \boldsymbol{e}_j = -\left(\boldsymbol{e}_j \wedge \boldsymbol{e}_i\right); \tag{2.2.29}$$

ii)
$$\boldsymbol{e}_i \wedge \boldsymbol{e}_i = \boldsymbol{e}_j \wedge \boldsymbol{e}_j = 0. \tag{2.2.30}$$

Therefore, we can simplify (2.2.28),

$$\boldsymbol{u} \wedge \boldsymbol{v} = (u_1 v_1 - v_1 u_2) \; \boldsymbol{e}_1 \wedge \boldsymbol{e}_2. \tag{2.2.31}$$

Thus, we interpret products like (2.2.23) as an oriented area. In the case of (2.2.31) it is a parallelogram oriented and depending on how its vertices are defined we have a clockwise or counterclockwise orientation, this being defined if (2.2.31) is negative or positive, respectively.

Another relevant operation is the so-called *star product*, which is a symmetric product. Its k-th component is defined as

$$(\boldsymbol{a} \star \boldsymbol{b})_k := \frac{C_d}{d-2} \sum_{i,j=1}^{d^2-1} d_{ijk} a_i b_j, \tag{2.2.32}$$

such that $\boldsymbol{a} \star \boldsymbol{b} = \boldsymbol{b} \star \boldsymbol{a}$.

With the products (2.2.23) and (2.2.32) in hand with (2.2.14) we can write a generalized version of (2.1.19) as

$$\begin{aligned}(\boldsymbol{a}\cdot\boldsymbol{\Lambda})(\boldsymbol{b}\cdot\boldsymbol{\Lambda}) &= \sum_{i,j=1}^{d^2-1} a_i b_j \Lambda_i \Lambda_j \\ &= \sum_{i,j=1}^{d^2-1} a_i b_j \left(i \sum_{k=1}^{d^2-1} f_{ijk}\Lambda_k + \sum_{k=1}^{d^2-1} d_{ijk}\Lambda_k + \frac{2}{d}\delta_{ij}\mathbb{1} \right),\end{aligned} \tag{2.2.33}$$

which, after simplification, reads

$$(\boldsymbol{a}\cdot\boldsymbol{\Lambda})(\boldsymbol{b}\cdot\boldsymbol{\Lambda}) = \frac{2}{d}(\boldsymbol{a}\cdot\boldsymbol{b})\,\mathbb{1} + i\,(\boldsymbol{a}\wedge\boldsymbol{b})\cdot\boldsymbol{\Lambda} + \frac{d-2}{C_d}(\boldsymbol{a}\star\boldsymbol{b})\cdot\boldsymbol{\Lambda}. \tag{2.2.34}$$

If $d = 2$ we recover all results (2.2.4) and (2.1.18) for $SU(2)$.

Following the same idea as for $SU(2)$, we can find the conditions that the vector $\boldsymbol{r}$ must satisfy. We will find that, pure states not just live in the surface of the Bloch body, but also satisfy $\boldsymbol{r} \star \boldsymbol{r} = \boldsymbol{r}$.

Observables

In $SU(2)$, we have seen that observables can be written as in (2.2.9). Now in $SU(d)$ for a given eigenbasis with d eigenvectors $\{|x_i\rangle\}$, an observable X admits a spectral decomposition in terms of the projectors

$$X_i = \frac{\mathbb{1} + C_d\,\hat{\boldsymbol{x}}_i\cdot\boldsymbol{\Lambda}}{d}, \tag{2.2.35}$$

where $\hat{\boldsymbol{x}}_i$ is an unit vector in $\mathbb{R}^{d^2-1}$. Applying the properties (2.2.35), we find

$$\begin{aligned}\sum_{i=1}^{d} X_i = \mathbb{1} &= \sum_{i=1}^{d}\left(\frac{\mathbb{1} + C_d\,\hat{\boldsymbol{x}}_i\cdot\boldsymbol{\Lambda}}{d}\right) \\ &= \mathbb{1} + \frac{C_d}{d}\sum_{i=1}^{d}\hat{\boldsymbol{x}}_i\cdot\boldsymbol{\Lambda}.\end{aligned} \tag{2.2.36}$$

It follows that in the second step we have to impose

$$\frac{C_d}{d}\sum_{i=1}^{d}\hat{\boldsymbol{x}}_i\cdot\boldsymbol{\Lambda} = 0 \tag{2.2.37}$$

implying that

$$\sum_{i=1}^{d}\hat{\boldsymbol{x}}_i = 0. \tag{2.2.38}$$

Note that the d vectors x_i do not form a basis in $\mathbb{R}^{d^2-1}$. They are not independent unit vectors since one can write any vector in terms of others. For instance, for qubits we have $\hat{\boldsymbol{x}}_1 + \hat{\boldsymbol{x}}_2 = 0$, hence $\hat{\boldsymbol{x}}_1 = -\hat{\boldsymbol{x}}_2$.

The spectral decomposition of a traceless observable X then reads

$$\begin{aligned} X &= \sum_{i=1}^{d} x_i X_i \\ &= \frac{1}{d}\sum_{i=1}^{d} x_i \left(\mathbb{1} + C_d\,\hat{\boldsymbol{x}}_i \cdot \boldsymbol{\Lambda}\right) \\ &= \frac{\mathbb{1}}{d}\sum_{i=1}^{d} x_i + \frac{C_d}{d}\sum_{i=1}^{d} x_i \left(\hat{\boldsymbol{x}}_i \cdot \boldsymbol{\Lambda}\right) \end{aligned} \tag{2.2.39}$$

which ensures that

$$\operatorname{Tr} X = \sum_{i=1}^{d} x_i \operatorname{Tr}(X_i) = \sum_{i=1}^{d} x_i = 0. \tag{2.2.40}$$

So replacing this result in equation (2.2.39) we have

$$X = \frac{C_d}{d}\sum_{i=1}^{d} x_i \left(\hat{\boldsymbol{x}}_i \cdot \boldsymbol{\Lambda}\right) \equiv \hat{\boldsymbol{x}} \cdot \boldsymbol{\Lambda}, \tag{2.2.41}$$

so that we obtain the following definition

$$\hat{\boldsymbol{x}} := \frac{C_d}{d}\sum_{i=1}^{d} x_i \hat{\boldsymbol{x}}_i. \tag{2.2.42}$$

2.3 Classical Probability Theory

The classical probability theory was put on a axiomatic system by Andrey Nikolaevich Kolmogorov in 1933 (see [Kolmogorov (1950), Ross (2012)]) which is based on 'probability spaces' (S, $\mathcal{F}$, μ). Where S is the set of all possible results in a experiment, called 'sample space'. $\mathcal{F}$ is a commutative algebra called 'σ-algebra', the space of events whose elements are subsets of S. And μ, a 'probability measure', is a function $\sigma : \mathcal{F} \longrightarrow [0,1]$. In this theory, every subset $A \in \mathcal{F}$ is a possible event and $\mathbb{p}(A)$, a function from $\mathcal{F}$ to the interval $[0,1] \in \mathbb{R}$, is the probability that event A happens. In what follows, we will give an intuitive definition on probability distributions.

2.3.1 Axioms

Any probability has the following properties or axioms, where A, B and C are subsets of the same spaces S.

i) If two events, A and B, are said impossible events, where $\varnothing$ is the empty set common to both events, then $\mathbb{p}(\varnothing) = 0$, and the probability of the intersection $\mathbb{p}(A \cap B) \equiv \mathbb{p}(A,B)$, the probability of "$A$ and B occur" in an experiment is

$$\mathbb{p}(A,B) = 0. \tag{2.3.1}$$

ii) If A^c is the complement event of A, then,

$$\mathbb{p}(A^c) = 1 - \mathbb{p}(A). \tag{2.3.2}$$

iii) If A and B are any two events, then, the probability of the union $\mathbb{p}(A \cup B)$ is the probability of "A or B occur" in an experiment, which is given by

$$\mathbb{p}(A \cup B) = \mathbb{p}(A) + \mathbb{p}(B) - \mathbb{p}(A, B). \tag{2.3.3}$$

iv) If A and B are mutually exclusive events,

$$\mathbb{p}(A \cup B) = \mathbb{p}(A) + \mathbb{p}(B). \tag{2.3.4}$$

v) If A, B and C are any three events,

$$\mathbb{p}(A \cup B \cup C) = \mathbb{p}(A) + \mathbb{p}(B) + \mathbb{p}(C) - \mathbb{p}(A, B) - \mathbb{p}(A, C) - \mathbb{p}(B, C) + \mathbb{p}(A, B, C). \tag{2.3.5}$$

2.3.2 Conditional Probability

Suppose that B is an event in S with $\mathbb{p}(B) \geq 0$. The probability of an event A occur once B has happened, the conditional probability of "A given B" is denoted $\mathbb{p}(A|B)$ and defined by

$$\mathbb{p}(A|B) := \frac{\mathbb{p}(A, B)}{\mathbb{p}(B)}. \tag{2.3.6}$$

2.3.3 Joint Probability

Suppose now A and B are any two events. Then, the probability of both events A and B occur is the so-called *joint probability* of "A and B". It is defined as

$$\mathbb{p}(A, B) := \mathbb{p}(A)\, \mathbb{p}(B|A). \tag{2.3.7}$$

Similarly, for the opposite situation, *i.e.*, the probability of both events B and A occur is given by

$$\mathbb{p}(B, A) := \mathbb{p}(B)\, \mathbb{p}(A|B), \tag{2.3.8}$$

where $\mathbb{p}(A, B) \equiv \mathbb{p}(B, A)$, this relation will play a central role in pur posterior discussions.

If A and B are independent events, i.e., the occurrences of A or B do not depend on each other, then equation (2.3.7) turns out to be

$$\mathbb{p}(A, B) = \mathbb{p}(A)\, \mathbb{p}(B). \tag{2.3.9}$$

We see clearly that if A and B are independent events $\mathbb{p}(B|A) = \mathbb{p}(B)$, since B does not depend on A and $\mathbb{p}(A|B) = \mathbb{p}(A)$, because does not depend on B.

2.3.4 Marginal Probability

Suppose that one sums equation (2.3.7) over all possible outputs of event A,

$$\sum_i \mathbb{p}(A_i, B) = \sum_i \mathbb{p}(A_i)\, \mathbb{p}(B|A_i). \tag{2.3.10}$$

The result is the *marginal probability* of the event B:

$$\mathbb{p}(B) := \sum_i \mathbb{p}(A_i, B) = \sum_i \mathbb{p}(A_i)\, \mathbb{p}(B|A_i), \tag{2.3.11}$$

The same can be done with the possible outputs of B in equation (2.3.8), *i.e.*,

$$\mathbb{p}(A) := \sum_j \mathbb{p}(A, B_j) = \sum_j \mathbb{p}(B_j)\, \mathbb{p}(A|B_j), \tag{2.3.12}$$

where here $\mathbb{p}(A)$ is called marginal probability of event A.

When one has a *binary variable* (events whose elements consist of A and its complement A^c) we can write the marginal probability of B, for instance, as

$$\mathbb{p}(B) = \mathbb{p}(B|A)\, \mathbb{p}(A) + \mathbb{p}(B|A^c)\, \mathbb{p}(A^c). \tag{2.3.13}$$

Consider, for instance, a simple case to illustrate conditional, joint and marginal probabilities. Suppose a box contains 24 toasters, 3 of which are defective. If two toasters are selected and tested, what is the probability that both are defective? The probability of the first toaster be defective is the marginal probability $\mathbb{p}(d_1) = 3/24 = 1/8$. Suppose that the first is indeed defective, now remains 23 toasters and there are 2 defective toasters left in the box, the probability that the second is defective is the conditional probability $\mathbb{p}(d_2|d_1) = 2/23$. The probability that both toasters are defective is the joint probability $\mathbb{p}(d_1, d_2) = p(d_1)\mathbb{p}(d_2|d_1) = 3/24 \times 2/23 = 1/92$. To obtain the marginal probability $\mathbb{p}(d_2)$, the inference over the result of the second toaster be defective not conditioned to the first, we can use equation (2.3.13), $\mathbb{p}(d_2) = p(d_1)\mathbb{p}(d_2|d_1) + p(d_1^c)\mathbb{p}(d_2|d_1^c)$, where $\mathbb{p}(d_1^c)$ is the probability of the first toaster not be defective, since there are 21 not defective out of 24, $\mathbb{p}(d_1^c) = 21/24$. The term $\mathbb{p}(d_2|d_1^c)$ is the conditional probability of the second be defective once the first is not, as there are 23 toasters left and 3 defective, then $\mathbb{p}(d_2|d_1^c) = 3/23$. Computing the marginal probability, we have, $\mathbb{p}(d_2) = 1/92 + 21/24 \times 3/23 = 1/8$, same result for $\mathbb{p}(d_1)$.

2.3.5 Bayes' Rule

In classical theory, the order of occurrence of two events does not matter, in the sense that events occurring in the order A and B or in the order B and A have the same probabilities, that is

$$\mathbb{p}(A, B) = \mathbb{p}(B, A). \tag{2.3.14}$$

Using (2.3.7) and (2.3.8), we obtain

$$\mathbb{p}(A)\, \mathbb{p}(B|A) = \mathbb{p}(B)\, \mathbb{p}(A|B). \tag{2.3.15}$$

Dividing by $\mathbb{p}(A)$ we get the conditional probability (2.3.6) in a different manner

$$\mathbb{p}(B|A) = \frac{\mathbb{p}(B)\, \mathbb{p}(A|B)}{\mathbb{p}(A)}. \tag{2.3.16}$$

The last equation is called *Bayes' rule*, it gives the probability of an event occurs with basis on previous knowledge. Bayes' rule connects two conditional probabilities $\mathbb{p}(B|A)$ and $\mathbb{p}(A|B)$, *i.e.*, one is the inverse of the other. Thus, Bayes' rule is a formula to "turn things around", being called by its creator Thomas Bayes as inverse probability.

Alternatively, we can write (2.3.16) by identifying the marginal probability $\mathbb{p}(A)$, so that

$$\mathbb{p}(B|A) = \frac{\mathbb{p}(B)\, \mathbb{p}(A|B)}{\sum_i \mathbb{p}(A_i)\, \mathbb{p}(B|A_i)}. \tag{2.3.17}$$

If one has a binary variable, then

$$\mathbb{p}(B|A) = \frac{\mathbb{p}(B)\, \mathbb{p}(A|B)}{\mathbb{p}(A)\, \mathbb{p}(B|A) + \mathbb{p}(A^c)\, \mathbb{p}(B|A^c)}. \tag{2.3.18}$$

2.4 Quantum Probability

In quantum theory, we have already introduced a postulate which gives the probability for the measurement of an observable. Given a state ρ and an observable X, the probability of obtaining the X eigenvalues x_n, is written as

$$\mathbb{p}_n = \mathrm{Tr}(X_n \rho). \tag{2.4.1}$$

Its classical correspondent is the probability $\mathbb{p}(A)$ of an event A. Rewriting equation (2.4.1) in a basis $\{|x_n\rangle\}$ using the properties of the trace, we get

$$\mathbb{p}_n = \langle x_n|\rho|x_n\rangle \, . \tag{2.4.2}$$

For pure states $\rho = |\psi\rangle\langle\psi|$, equation (2.4.2) reduces to

$$\mathbb{p}_n = |\langle x_n|\psi\rangle|^2 \, . \tag{2.4.3}$$

Equation (2.4.1) is the most general way to write probabilities within the quantum formalism, since it is valid both for pure and mixed ensembles.

Example: Qubit

Consider a qubit state given by (2.2.2) submitted to a measurement of the observable $X = \hat{\boldsymbol{x}} \cdot \boldsymbol{\sigma}$ whose projectors X_i are (2.2.9). The probability of getting the outcome x_i reads

$$\begin{aligned} \mathbb{p}_{x_i} &= \mathrm{Tr}(X_i \rho) \\ &= \frac{\mathrm{Tr}\left[(\mathbb{1} + i\hat{\boldsymbol{x}} \cdot \boldsymbol{\sigma})(\mathbb{1} + \boldsymbol{r} \cdot \boldsymbol{\sigma})\right]}{4} \\ &= \frac{\mathrm{Tr}\left(\mathbb{1} + \boldsymbol{r} \cdot \boldsymbol{\sigma} + i\hat{\boldsymbol{x}} \cdot \boldsymbol{\sigma} + i(\hat{\boldsymbol{x}} \cdot \boldsymbol{\sigma})(\boldsymbol{r} \cdot \boldsymbol{\sigma})\right)}{4}. \end{aligned} \tag{2.4.4}$$

Since we can write the product $(\hat{\boldsymbol{x}} \cdot \boldsymbol{\sigma})(\boldsymbol{r} \cdot \boldsymbol{\sigma})$ as (2.1.18) and $\mathrm{Tr}(\sigma_k) = 0$, then

$$\begin{aligned} \mathbb{p}_{x_i} &= \frac{\mathrm{Tr}\left[\mathbb{1} + \boldsymbol{r} \cdot \boldsymbol{\sigma} + i\hat{\boldsymbol{x}} \cdot \boldsymbol{\sigma} + i(\hat{\boldsymbol{x}} \cdot \boldsymbol{r})\mathbb{1} + \imath i(\hat{\boldsymbol{x}} \times \boldsymbol{r}) \cdot \sigma\right]}{4} \\ &= \frac{\left[1 + i(\hat{\boldsymbol{x}} \cdot \boldsymbol{r})\right] \mathrm{Tr}\,\mathbb{1}}{4} \\ &= \frac{1 + i(\hat{\boldsymbol{x}} \cdot \boldsymbol{r})}{2}. \end{aligned} \tag{2.4.5}$$

Notice that $\mathbb{p}_{x_i}$ depends both on $|r|$ and the angle between $\boldsymbol{r}$ and $\hat{\boldsymbol{x}}$.

Example: Qudits

For qudits, a system with d degrees of freedom, when $d = 2$ it is named qubit, we can compute probabilities using the generalized Bloch representation. First the probability of one obtaining the outputs x_i in a measurement of X over a state ρ reads

$$\mathbb{p}_{x_i} = \mathrm{Tr}(X_i \rho) = \frac{1 + (d-1)\hat{\boldsymbol{x}}_i \cdot \boldsymbol{r}}{d}. \tag{2.4.6}$$

If the input state is an eigenstate of X, *i.e.*, $\rho = X_{i'}$, then the probability will be $\mathbb{p}_{x_i} = \mathrm{Tr}(X_i X_{i'}) = \mathrm{Tr}(|x_i\rangle\langle x_i|x_{i'}\rangle\langle x_{i'}|) = \delta_{ii'}$, we can write the internal product between the unit vectors of an observable as

$$\hat{\boldsymbol{x}}_i \cdot \hat{\boldsymbol{x}}_{i'} = \frac{d\delta_{ii'} - 1}{d-1}. \tag{2.4.7}$$

satisfying equation (2.2.38), since $\sum_i \hat{\boldsymbol{x}}_i \cdot \hat{\boldsymbol{x}}_{i'} = (d-d)/(d-1) = 0$.

2.4.1 Conditional Probabilities

Single Systems

Quantum mechanics is adapted to describe not only single measurements. One can have multiple measurements, simultaneous or sequentially. The idea is the same as in classical theory. Suppose Alice receives a quantum state ρ and makes a measurement of the observable X. According to postulate 4, the probability of Alice to get x_i is $\mathbb{p}_{x_i}$. The post-measurement state is then sent to Bob, who will perform a measurement of observable Y. The probability of Bob getting y_j is given by $\mathbb{p}_{y_j|x_i}$, which is a probability distribution conditioned to every outcome x_i obtained in Alice's measurement. In other words, the probabilities obtained by Bob can be affected by Alice's previous measurements.

According to the quantum mechanical formalism, the referred conditional probability is written as

$$\mathbb{p}_{y_j|x_i} \equiv \mathbb{p}(y_j|x_i) := \mathrm{Tr}\left(Y_j \rho_{S|x_i}\right). \tag{2.4.8}$$

Given that $\rho_{S|x_i} = X_i$, we can rewrite (2.4.8) as

$$\mathbb{p}_{y_j|x_i} = \mathrm{Tr}\left(Y_j X_i\right). \tag{2.4.9}$$

If the opposite happens, that is, Bob measure Y and then Alice measure X, then the probability that Alice obtains x_i is conditioned to Bob's action according to

$$\mathbb{p}_{x_i|y_j} \equiv \mathbb{p}(x_i|y_j) := \mathrm{Tr}\left(X_i \rho_{S|y_j}\right), \tag{2.4.10}$$

which can be re-written as

$$\mathbb{p}_{x_i|y_j} = \mathrm{Tr}\left(X_i Y_j\right). \tag{2.4.11}$$

Notice that $\mathbb{p}_{x_i|y_j} = \mathbb{p}_{y_j|x_i}$, by the cyclic property of the trace.

Composite Systems

Consider now a system composed of two subsystems associated with a Hilbert space $\mathcal{H}_{\mathcal{XY}}$. Alice receives one particle, makes a measurement of X with probability $\mathbb{p}_{x_i}$. Bob receives the other particle and measures an observable Y. The conditional probability reads

$$\mathbb{p}_{y_j|x_i} \equiv \mathbb{p}(y_j|x_i) := \mathrm{Tr}_{\mathcal{Y}}\left(Y_j \rho_{Y|x_i}\right), \tag{2.4.12}$$

where $\rho_{Y|x_i}$ is the conditional state given by (2.1.47). The opposite situation can also occur, *i.e.*, Bob measures Y and then Alice measures X. The conditional probability now reads

$$\mathbb{p}_{x_i|y_j} \equiv \mathbb{p}(x_i|y_j) := \mathrm{Tr}_{\mathcal{X}}\left(X_i \rho_{X|y_j}\right). \tag{2.4.13}$$

Notice that the difference between single to composite systems lies in the conditional state and also the trace in the probability formulas.

Example: Qubit

Consider a qubit prepared in the state (2.2.2). Regardless of the ordering one chooses for the measurement of X and Y, we obtain

$$\begin{aligned}
\mathbb{p}_{y_j|x_i} = \mathbb{p}_{x_i|y_j} &= \operatorname{Tr}\left(Y_j X_i\right) \\
&= \frac{\operatorname{Tr}\left[(\mathbb{1} + i\hat{\boldsymbol{x}} \cdot \boldsymbol{\sigma})(\mathbb{1} + j\hat{\boldsymbol{y}} \cdot \boldsymbol{\sigma})\right]}{4} \\
&= \frac{\operatorname{Tr}\left(\mathbb{1} + j\hat{\boldsymbol{y}} \cdot \boldsymbol{\sigma} + i\hat{\boldsymbol{x}} \cdot \boldsymbol{\sigma} + ij(\hat{\boldsymbol{x}} \cdot \boldsymbol{\sigma})(\hat{\boldsymbol{y}} \cdot \boldsymbol{\sigma})\right)}{4}.
\end{aligned} \tag{2.4.14}$$

Using (2.1.18) and $\operatorname{Tr}(\sigma_k) = 0$, we find

$$\begin{aligned}
\mathbb{p}_{y_j|x_i} = \mathbb{p}_{x_i|y_j} &= \frac{\operatorname{Tr}\left[\mathbb{1} + j\hat{\boldsymbol{y}} \cdot \boldsymbol{\sigma} + i\hat{\boldsymbol{x}} \cdot \boldsymbol{\sigma} + i(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}})\mathbb{1} + \imath ij(\hat{\boldsymbol{x}} \times \hat{\boldsymbol{y}}) \cdot \sigma\right]}{4} \\
&= \frac{\left[1 + ij(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}})\right] \operatorname{Tr} \mathbb{1}}{4}
\end{aligned} \tag{2.4.15}$$

where $\operatorname{Tr} \mathbb{1} = 2$. Finally, we have

$$\mathbb{p}_{y_j|x_i} = \mathbb{p}_{x_i|y_j} = \frac{1 + ij(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}})}{2}. \tag{2.4.16}$$

Example: Qudits

Using general Bloch representation 2.2.3 we have

$$\mathbb{p}_{x_i|y_j} = \mathbb{p}_{y_j|x_i} = \operatorname{Tr}(X_i Y_j) = \frac{1 + (d-1)\hat{\boldsymbol{x}}_i \cdot \hat{\boldsymbol{y}}_j}{d}. \tag{2.4.17}$$

For qubits ($d = 2$) we have, e.g., $\hat{\boldsymbol{x}}_i = i\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}_j = j\hat{\boldsymbol{y}}$ and formula (2.4.16) is readily recovered.

2.4.2 Joint Probabilities

Single Systems

Suppose one is given the task to obtain the probability of measuring two observables X and Y. In classical physics, measurements just reveal given information about the systems without significantly disturbing the physical state. Then, when two observables, X and Y, are measured, the joint probability of finding x_i and y_j is $\mathbb{p}(x_i, y_j)$ regardless of the time ordering with which these observables are actually measured. In quantum mechanics, however, measurements fundamentally disturb the state of the system, in a way such that the ordering "X then Y" may lead to rather different results from "Y then X". In this case, it is fair for one to wonder whether some ambiguity will arise concerning the description of the joint probability $\mathbb{p}(x_i, y_j)$. This rationale reveals that, in fact, conditional probabilities may be more fundamental objects, for they allow us to write

$$\mathbb{p}_{x_i,y_j} := \mathbb{p}_{x_i}\mathbb{p}_{y_j|x_i} = \operatorname{Tr}(X_i\rho) \operatorname{Tr}\left(Y_j \rho_{S|x_i}\right). \tag{2.4.18}$$

In the case, it is clear that the joint probability is constructed under the premise that X is measured before Y. Using (2.4.9), we further obtain

$$\mathbb{p}_{x_i,y_j} := \operatorname{Tr}(X_i\rho) \operatorname{Tr}(Y_j X_i). \tag{2.4.19}$$

Now, when the inverse ordering is taken, quantum mechanics predicts that

$$\mathbb{p}_{y_j,x_i} := \mathbb{p}_{y_j}\mathbb{p}_{x_i|y_j} = \mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}\left(X_i\rho_{S|y_j}\right) \tag{2.4.20}$$

which, along with (2.4.11), reduces to

$$\mathbb{p}_{y_j,x_i} := \mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}(X_iY_j). \tag{2.4.21}$$

Since equation (2.4.19) is not generally equal to (2.4.21), the distinct notations $\mathbb{p}_{x_i,y_j}$ and $\mathbb{p}_{y_j,x_i}$ for the joint probability are justified. Recall that such a distinction is immaterial in classical physics.

If one sums over all possible values of j in equation (2.4.19) the result is

$$\begin{aligned}\sum_j \mathbb{p}_{x_i,y_j} &= \sum_j \mathrm{Tr}(X_i\rho)\,\mathrm{Tr}(Y_jX_i)\\ &= \mathrm{Tr}(X_i\rho)\sum_j \mathrm{Tr}(Y_jX_i)\\ &= \mathrm{Tr}(X_i\rho)\,\mathrm{Tr}\left(\sum_j Y_jX_i\right)\\ &= \mathrm{Tr}(X_i\rho)\,\mathrm{Tr}(X_i)\\ &= \mathrm{Tr}(X_i\rho),\end{aligned} \tag{2.4.22}$$

where we identify $\mathrm{Tr}(X_i\rho) = \mathbb{p}_{x_i}$. Then,

$$\mathbb{p}_{x_i} := \sum_j \mathbb{p}_{x_i,y_j}, \tag{2.4.23}$$

which, as in classical theory, is called the *marginal probability* of X. However,

$$\mathbb{p}_{x_i} \neq \sum_j \mathbb{p}_{y_j,x_i}. \tag{2.4.24}$$

is also a marginal probability but different from $\mathbb{p}_{x_i}$. This result is interesting since it shows that the ordering of the measurements really matters in quantum mechanics.

On the other hand, summing overall x_i in equation (2.4.21) we find

$$\begin{aligned}\sum_i \mathbb{p}_{y_j,x_i} &= \sum_i \mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}(X_iY_j)\\ &= \mathrm{Tr}(Y_j\rho)\sum_i \mathrm{Tr}(Y_jX_i)\\ &= \mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}\left(Y_j\sum_i X_i\right)\\ &= \mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}(Y_j)\\ &= \mathrm{Tr}(Y_j\rho),\end{aligned} \tag{2.4.25}$$

where we identify $\mathrm{Tr}(Y_j\rho) = \mathbb{p}_{y_j}$. Then

$$\mathbb{p}_{y_j} := \sum_i \mathbb{p}_{y_j,x_i}, \tag{2.4.26}$$

which is the *marginal probability* of Y. Once again, though, we have

$$\mathbb{p}_{y_j} \neq \sum_i \mathbb{p}_{x_i,y_j}. \tag{2.4.27}$$

Therefore, equations (2.4.24) and (2.4.27) violate the principles of classical theory. As we have seen in classical theory $\mathbb{p}_{x_i} = \sum_j \mathbb{p}_{x_i,y_j} = \sum_j \mathbb{p}_{y_j,x_i}$, but this fails in quantum theory. The reason lies in the order that one measures observables X and Y, since in general $\mathbb{p}_{x_i,y_j} \neq \mathbb{p}_{y_j,x_i}$. We are finding clues that somehow quantum probability theory differs from the classical one. Later on, we will show that it is connected to incompatibility, something inexistent in classical theory.

Composite Systems

Given a bipartite state $\rho \in \mathscr{B}(H_{\mathcal{X}} \otimes H_{\mathcal{Y}})$, upon measurements of $X \in \mathscr{B}(H_{\mathcal{X}})$ and $Y \in \mathscr{B}(H_{\mathcal{Y}})$ in this sequence, the joint probability distribution is given by

$$\mathbb{p}_{x_i,y_j} = \mathbb{p}_{x_i}\mathbb{p}_{y_j|x_i} = \mathrm{Tr}(X_i\rho)\,\mathrm{Tr}_{\mathcal{Y}}\left(Y_j\rho_{\mathcal{Y}|x_i}\right), \tag{2.4.28}$$

whereas for the opposite sequence, we find

$$\mathbb{p}_{y_j,x_i} = \mathbb{p}_{y_j}\mathbb{p}_{x_i|y_j} = \mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}_{\mathcal{X}}\left(X_i\rho_{\mathcal{X}|y_j}\right). \tag{2.4.29}$$

Example: Qubits

With probabilities (2.4.5) and conditional probabilities (2.4.16) for qubits we can write the joint probabilities

$$\mathbb{p}_{x_i,y_j} = \frac{1 + i(\hat{\boldsymbol{x}}\cdot\boldsymbol{r}) + ij\big[(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}}) + j(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}})(\hat{\boldsymbol{x}}\cdot\boldsymbol{r})\big]}{4}, \tag{2.4.30}$$

and

$$\mathbb{p}_{y_j,x_i} = \frac{1 + j(\hat{\boldsymbol{y}}\cdot\boldsymbol{r}) + ij\big[(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}}) + i(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}})(\hat{\boldsymbol{y}}\cdot\boldsymbol{r})\big]}{4}. \tag{2.4.31}$$

Example: Bloch Body

With probabilities (2.4.6) and the conditional probabilities (2.4.19) we can write the joint probabilities in generalized Bloch representation for outcomes x_i and y_j in respective measurements of X and Y:

$$\mathbb{p}_{x_i,y_j} = \frac{1 + (d-1)\big[\hat{\boldsymbol{x}}_i\cdot\hat{\boldsymbol{y}}_j + \hat{\boldsymbol{x}}_i\cdot\boldsymbol{r} + (d-1)(\hat{\boldsymbol{x}}_i\cdot\hat{\boldsymbol{y}}_j)(\hat{\boldsymbol{x}}_i\cdot\boldsymbol{r})\big]}{d^2}, \tag{2.4.32}$$

whereas for the converse order one has

$$\mathbb{p}_{y_j,x_i} = \frac{1 + (d-1)\big[\hat{\boldsymbol{x}}_i\cdot\hat{\boldsymbol{y}}_j + \hat{\boldsymbol{y}}_j\cdot\boldsymbol{r} + (d-1)(\hat{\boldsymbol{x}}_i\cdot\hat{\boldsymbol{y}}_j)(\hat{\boldsymbol{y}}_j\cdot\boldsymbol{r})\big]}{d^2}. \tag{2.4.33}$$

For one qubit ($d = 2$) we have, e.g., $\hat{\boldsymbol{x}}_i = i\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}_j = j\hat{\boldsymbol{y}}$, thus

$$\mathbb{p}_{x_i,y_j} = \frac{1 + i(\hat{\boldsymbol{x}}\cdot\boldsymbol{r}) + ij\big[(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}}) + j(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}})(\hat{\boldsymbol{x}}\cdot\boldsymbol{r})\big]}{4} \tag{2.4.34}$$

and

$$\mathbb{p}_{y_j,x_i} = \frac{1 + j(\hat{\boldsymbol{y}}\cdot\boldsymbol{r}) + ij\big[(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}}) + i(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}})(\hat{\boldsymbol{y}}\cdot\boldsymbol{r})\big]}{4}, \tag{2.4.35}$$

already obtained in equations (2.4.30) and (2.4.31).

2.5 Information Theory

In this section, a brief review is made about information theory. This is necessary because all quantifiers that will be defined in the main topics of this work are based on the quantum notion of information.

Our main interest is the physical meaning of information; however, it is a general subject, in the sense that it can be applied in a variety of contexts. We treat here, therefore, information as a physical quantity. The references which we follow are [Witten (2020), Preskill (2019-20)]. As usual, we start with the classical theory and then move to the quantum theory.

2.5.1 Classical Theory

The classical information theory does not take into account superposition phenomena, this means that quantum mechanics is not applied to the theory. The smallest amount of information is called *bit*, being denoted by 0 or 1. The bit can be used to quantify dichotomous variables only: the parity of a number, the face of a coin, the state (on or off) of a lamp, etc.

Shannon Entropy

To illustrate the idea behind Shannon entropy, suppose that Alice has a weather radio station[9]. She is hired by Bob to inform him every day if it the next day will be sunny or rainy with equal probability, *i.e.*, a probability of 50% of each happening. If she tells Bob that it will be raining the next day, then, Bob has gained '1 bit' of useful information about the weather, or in other words, his uncertainty has dropped by a factor of 2. Since his uncertainty was about 2 possibilities, now there is only 1 possibility. Thus his uncertainty has been divided by a factor of 2. It does not matter how many bytes she used to encode the message sent to Bob's receptor; every single exceeding bit of information is useless. The useful amount of information about raining or not raining is 1 bit since we have $2^1 = 2$. The number of bits is just log of the numerical factor by which his uncertainty is divided. In this case, the number of bits is given by $\log_2 2 = 1$.

Alice is also hired by Charlie, but he has a bigger demand. He wants to know 8 equally possible states, depending on how many clouds and rain there will be the next day. When she informs Charlie about the weather, his uncertainty will drop by a factor of 8, *i.e.*, he will gain 3 bits of useful information, since $2^3 = 8$. In this case, the number of bits is $\log_2 8 = 3$.

However, suppose that the weather states are not equally possible. In Bob's case, assume that there is $25\% = 0.25$ of chance of the next day to be sunny and $75\% = 0.75$ to be rainy. If Alice informs Bob that it will be sunny the next day, Bob's uncertainty has been divided by a factor 4. So, Bob has gained 2 bits of useful information, since $2^2 = 4$. In this case, the number of bits is $\log_2 4 = \log_2 1/0.25 = 2$ or $-\log_2 0.25 = 2$. If Alice informs that it will be rainy the next day, the number of bits of useful information is $-\log_2 0.75 = 0.415$. Therefore, when the next day is rainy, Bob gains less information than when it is sunny. But one can ask how much information Bob gains every day on average. The answer is

$$0.25 \times 2 + 0.75 \times 0.415 = 0.81125 \,\text{bits}. \tag{2.5.1}$$

This is the *Shannon entropy*, introduced by Claude Shannon in his 1948 paper [Shannon (1948)]. The Shannon entropy can be summarized by the following statement

[9]Weather radio station, telephone cable, etc...are examples of a classical channel of communication.

The Shannon entropy $H(X)$ quantifies the uncertainty degree about the random variable X before the message reception. Or, it quantifies the information degree about the random variable X after the message reception.

Considering a more general case, for instance, a message encoded as a string of two symbols a with probability $\mathbb{p}$ and b with probability $1 - \mathbb{p}$ like

$$abaabbabbaaa\ldots, \tag{2.5.2}$$

one can ask how many bits one can extract on average from this message with N bits. The answer is

$$H(\mathbb{p}) = -\mathbb{p}\log\mathbb{p} - (1-\mathbb{p})\log(1-\mathbb{p}), \tag{2.5.3}$$

where H is called *binary* Shannon entropy[10]. The number of bits in the message is NH.

In a even more general scenario, we can imagine a collection $X = \{x_1, x_2, \ldots, x_k\}$, e.g., an alphabet, with probabilities $\mathbb{p}_1, \mathbb{p}_2, \ldots, \mathbb{p}_k$. The entropy per letter is given by

$$H(\mathbb{p}_i) = -\sum_{i=1}^{k} \mathbb{p}_i \log \mathbb{p}_i. \tag{2.5.4}$$

Conditional Entropy and Joint Entropy

Consider now that Alice sends a report to Bob about the next day weather state with many letters $X = \{x_1, x_2, \ldots, x_k\}$. However, her signal is too weak since she employs a noisy channel of communication. Bob receives a slightly different message, but with letters $Y = \{y_1, y_2, \ldots, y_k\}$. One can ask how many bits Bob gains after Alice transmitted her report.

We have first to analyze what are the probability distributions. Suppose that in these rough conditions the probability that Alice sends x_i and Bob receives y_j is $\mathbb{p}(x_i, y_j)$. The probability that Bob receives y_j, summing over all weather possibilities, *i.e.*, summing over all Alice's intended symbols, is

$$\mathbb{p}(y_j) = \sum_{i=1}^{k} \mathbb{p}(x_i, y_j). \tag{2.5.5}$$

If Bob receives y_j, he can estimate the probability that Alice has sent x_i by writing the conditional probability

$$\mathbb{p}(x_i|y_j) = \frac{\mathbb{p}(x_i, y_j)}{\mathbb{p}(y_j)}. \tag{2.5.6}$$

The ignorance that Bob attributes to Alice signal after he receives y_j is the Shannon entropy conditioned to the acquisition of knowledge, that is

$$H(X|y_j) = -\sum_{i=1}^{k} \mathbb{p}(x_i|y_j)\log\Big(\mathbb{p}(x_i|y_j)\Big). \tag{2.5.7}$$

If one averages over all symbols that Bob receives, one obtains the average entropy that remains in Alice's signal once Bob receives y_j or the so-called *conditional entropy*, that is

$$H(X|Y) = \sum_{j=1}^{k} \mathbb{p}(y_j) H(X|y_j), \tag{2.5.8}$$

[10]If one uses $\log_2$, H has units of 'bits'. If one uses ln, H has units of 'nats'. It is conventioned that $0\log 0 \equiv 0$.

or,

$$H(X|Y) = -\sum_{j=1}^{k} \mathbb{p}(y_j) \sum_{i=1}^{k} \frac{\mathbb{p}(x_i, y_j)}{\mathbb{p}(y_j)} \log\left(\frac{\mathbb{p}(x_i, y_j)}{\mathbb{p}(y_j)}\right), \tag{2.5.9}$$

or, simply,

$$H(X|Y) := -\sum_{i,j=1}^{k} \mathbb{p}(x_i, y_j) \log\left(\frac{\mathbb{p}(x_i, y_j)}{\mathbb{p}(y_j)}\right). \tag{2.5.10}$$

It can be interpreted as information content that Bob still does not have about X after receiving y_j. After further manipulations, we find

$$\begin{aligned} H(X|Y) &= -\sum_{i,j=1}^{k} \mathbb{p}(x_i, y_j) \log \mathbb{p}(x_i, y_j) + \sum_{i,j=1}^{k} \mathbb{p}(x_i, y_j) \log p(y_j) \\ &= H(X,Y) - H(Y), \end{aligned} \tag{2.5.11}$$

where we identify the *joint entropy*

$$H(X,Y) := -\sum_{i,j=1}^{k} \mathbb{p}(x_i, y_j) \log \mathbb{p}(x_i, y_j). \tag{2.5.12}$$

Notice that it is similar to Shannon entropy, the difference being that it has two sums and joint probability distributions.

Properties of Entropy

The following properties take from [Chuang (2000)] are properties of the entropy:

i) $H(X,Y) = H(Y,X)$;

ii) $H(X) \leq H(X,Y)$, equality if and only if (iff) $Y = f(x)$;

iii) If there are k possible results for X, then $H(X) \leq \log k$, equality iff $\mathbb{p}(x) = 1/k$;

iv) **Subadditivity**: $H(X,Y) \leq H(X) + H(Y)$, equality iff X and Y are independents;

v) $H(X|Y,Z) \leq H(X|Y)$.

Mutual Information

However, noisy the signal may be, Bob can still gain information about Alice's original message by reading Y. This is possible whenever the messages X and Y share some correlation, which is quantified via the mutual information:

$$I(X:Y) := H(X) + H(Y) - H(X,Y), \tag{2.5.13}$$

or,

$$I(X:Y) = \sum_{i,j=1}^{k} \mathbb{p}(x_i, y_j) \log\left(\frac{\mathbb{p}(x_i, y_j)}{\mathbb{p}(x_i)\mathbb{p}(y_j)}\right). \tag{2.5.14}$$

This quantity gives us how much information X has about Y and vice-versa. Accordingly, it can be shown that $I(X:Y) = 0$ if and only if X and Y are uncorrelated.

Kullback-Leibler Divergence

Suppose Alice wonders if Bob got the right information that she sent every time, despite the noisy communication channel they have. The information associated with the messages is quantified by means of the probability distributions $\mathbb{p}$ and $\mathbb{q}$. To verify how different the distributions are, one can compute the quantity

$$D\Big(\mathbb{p}(X)\big\|\mathbb{q}(X)\Big) := \sum_{i=1}^{k} \mathbb{p}(x_i) \log\left(\frac{\mathbb{p}(x_i)}{\mathbb{q}(x_i)}\right), \tag{2.5.15}$$

which is called *Kullback-Leibler divergence* (KL divergence) or *relative entropy.* If the probabilities are equal, *i.e.*, $\mathbb{p}(x) = \mathbb{q}(x)$ we have $D = 0$, meaning that the broadcast was perfect, with no information lost. If $\mathbb{p}(x) \neq \mathbb{q}(x)$, Alice can establish a interval for D that she considers that it was a good broadcast. One may wonder whether D is a proper metric to quantify how "distant" the distributions are from each other. For this reason, we introduce the following theorem regarding KL divergences.

Theorem 2. ***(Non-negativity of KL Divergence)*** *For any two probability distributions* $\mathbb{p}(x)$ *and* $\mathbb{q}(x)$*, the KL divergence is always non-negative,* i.e., $D\Big(\mathbb{p}(x)\big\|\mathbb{q}(x)\Big) \geq 0$*, where the equality is satisfied if and only if* $\mathbb{p}(x) = \mathbb{q}(x)$*.*

Proof. To proof the previous theorem we have to introduce the useful inequality:

$$\log(x)\ln(2) = \ln(x) \leq x - 1 \tag{2.5.16}$$

or

$$-\log(x) \geq \frac{1-x}{\ln(2)}. \tag{2.5.17}$$

Therefore, we can rewrite (2.5.15) as

$$\begin{aligned} D\Big(\mathbb{p}(x)\big\|\mathbb{q}(x)\Big) &= \sum_{i=1}^{k} \mathbb{p}(x_i)\log\left(\frac{\mathbb{p}(x_i)}{\mathbb{q}(x_i)}\right) \\ &= -\sum_{i=1}^{k} \mathbb{p}(x_i)\log\left(\frac{\mathbb{q}(x_i)}{\mathbb{p}(x_i)}\right) \\ &\geq \frac{1}{\ln 2}\sum_{i=1}^{k} \mathbb{p}(x_i)\left(1 - \frac{\mathbb{q}(x_i)}{\mathbb{p}(x_i)}\right) \\ &= \frac{1}{\ln 2}\sum_{i=1}^{k} \left(\mathbb{p}(x_i) - \mathbb{q}(x_i)\right) \\ &= \frac{1-1}{\ln 2} \\ &= 0. \end{aligned} \tag{2.5.18}$$

□

Based on the previous theorem we can say that KL divergence is a good *distance*, however it does not satisfy triangle inequality, therefore it is not a metric.

Other important theorems presented in [Cover (2006)] follow.

Theorem 3. ***(Log-sum inequality)*** *If $x_1, x_2, \dots, x_n$ and $y_1, y_2, \dots, y_n$ are non-negative numbers then*

$$\sum_{i=1}^{n} x_i \log\left(\frac{x_i}{y_i}\right) \geq \left(\sum_{i=1}^{n} x_i\right) \log\left(\frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i}\right), \tag{2.5.19}$$

with equality if and only if $x_i/y_i = const.$

Theorem 4. ***(Convexity)*** *$D(\mathbb{p}||\mathbb{q})$ is convex in the pair $(\mathbb{p}, \mathbb{q})$, so that for the probability distributions $(\mathbb{p}_1, \mathbb{q}_1)$ and $(\mathbb{p}_2, \mathbb{q}_2)$, we have for all $\lambda \in [0, 1]$*

$$D\Big(\lambda \mathbb{p}_1 + (1-\lambda)\mathbb{p}_2 \Big\| \lambda \mathbb{q}_1 + (1-\lambda)\mathbb{q}_2\Big) \leq \lambda\, D\big(\mathbb{p}_1 \| \mathbb{q}_1\big) + (1-\lambda)\, D\big(\mathbb{p}_2 \| \mathbb{q}_2\big), \tag{2.5.20}$$

Theorem 5. ***(KL Divergence is not symmetric)*** *For any two probability distributions $\mathbb{p}(x)$ and $\mathbb{q}(x)$, the following equality is not always valid*

$$D\Big(\mathbb{p}(x) \Big\| \mathbb{q}(x)\Big) \neq D\Big(\mathbb{q}(x) \Big\| \mathbb{p}(x)\Big), \tag{2.5.21}$$

in general, for a given $x \in X$.

Example: Suppose that Alice selects two different weeks in the year to analyze the broadcast quality.

1. Week 1: During this week there were few clouds in the sky and sunny most days, therefore the broadcast should be good. Alice has the following data of what she sent and what Bob received for the probability of raining in that day:

	Day 1	**Day 2**	**Day 3**	**Day 4**
Alice	1/4	1/2	1/4	1/4
Bob	1/2	1/4	1/8	1/8

   With this data, she computes the KL divergence:

$$\begin{aligned} D &= \frac{1}{4}\log\left(\frac{\frac{1}{4}}{\frac{1}{2}}\right) + \frac{1}{2}\log\left(\frac{\frac{1}{2}}{\frac{1}{4}}\right) + \frac{1}{4}\log\left(\frac{\frac{1}{4}}{\frac{1}{8}}\right) + \frac{1}{4}\log\left(\frac{\frac{1}{4}}{\frac{1}{8}}\right) \\ &\approx 0.52 \end{aligned} \tag{2.5.22}$$

   so, D is a small number, indicating that the broadcast in that week was good.

2. Week 2: During this week there were a lot of clouds in the sky and with storms most days; therefore the broadcast should be bad, which caused a misunderstand in one of the messages lines by Bob.

   Then, the KL divergence with this data will be:

$$\begin{aligned} D &= \frac{1}{4}\log\left(\frac{\frac{1}{4}}{0}\right) + \frac{1}{2}\log\left(\frac{\frac{1}{2}}{\frac{1}{4}}\right) + \frac{1}{4}\log\left(\frac{\frac{1}{4}}{\frac{1}{8}}\right) + \frac{1}{4}\log\left(\frac{\frac{1}{4}}{\frac{1}{8}}\right) \\ &= \infty \end{aligned} \tag{2.5.23}$$

   where the convention used is $\mathbb{p}\log 1/0 = \mathbb{p}\log\infty \equiv \infty$, for any $\mathbb{p} > 0$. Thus, Alice concludes that in week 2 the broadcast was much significantly worse than week 1. In other words, Bob lost more information during week 2 than during week 1.

	Day 1	Day 2	Day 3	Day 4
Alice	1/4	1/2	1/4	1/4
Bob	0	1/4	1/8	1/8

Joint Kullback-Leibler Divergence

Suppose now that Alice wants to include Charlie in her broadcast analysis to verify how bad or how good are their communication channels. In this case, instead of (2.5.15) with probability distributions $\mathbb{p}(X)$ and $\mathbb{q}(X)$, she needs to adjust her measure, once now her probability distributions are the joint probability distributions $\mathbb{p}(x, y)$ and $\mathbb{q}(x, y)$. This is so because, using different channels to communicate, Alice actually delivers symbols $x_1, x_2, \ldots, x_k$ to Bob and $y_1, y_2, \ldots, y_l$ to Charlie. So, the KL divergence that she computes is now given by

$$D\Big(\mathbb{p}(X,Y)\Big\|\mathbb{q}(X,Y)\Big) = \sum_{i=1}^{k}\sum_{j=1}^{l} \mathbb{p}(x_i, y_j) \log\left(\frac{\mathbb{p}(x_i, y_j)}{\mathbb{q}(x_i, y_j)}\right), \tag{2.5.24}$$

where $\mathbb{p}$ and $\mathbb{q}$ are the probabilities received by Bob and Charlie, respectively, concerning the next day weather.

If Bob and Charlie also decide to compare the quality of their communication channels with Alice, then, they compute the following KL divergence

$$D\Big(\mathbb{q}(X,Y)\Big\|\mathbb{p}(X,Y)\Big) = \sum_{i=1}^{k}\sum_{j=1}^{l} \mathbb{q}(x_i, y_j) \log\left(\frac{\mathbb{q}(x_i, y_j)}{\mathbb{p}(x_i, y_j)}\right). \tag{2.5.25}$$

The symmetric KL divergence is written as

$$D_S = \frac{1}{2}\Big[D\Big(\mathbb{p}(x,y)\Big\|\mathbb{q}(x,y)\Big) + D\Big(\mathbb{q}(x,y)\Big\|\mathbb{p}(x,y)\Big)\Big], \tag{2.5.26}$$

or,

$$D_S = \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{l}\left[\mathbb{p}(x_i, y_j)\log\left(\frac{\mathbb{p}(x_i, y_j)}{\mathbb{q}(x_i, y_j)}\right) + \mathbb{q}(x_i, y_j)\log\left(\frac{\mathbb{q}(x_i, y_j)}{\mathbb{p}(x_i, y_j)}\right)\right]. \tag{2.5.27}$$

2.5.2 Quantum Theory

In classical theory, we saw that information is represented by bits $\{0, 1\}$, which can describe, for example, states of current and no-current in a wire. However, in quantum systems, we have seen physical quantities are not always well defined, so that two-level systems can be prepared not only in the orthogonal states $|0\rangle$ and $|1\rangle$, but also in a coherent superposition $\alpha|0\rangle + \beta|1\rangle$. In this capacity, one has a quantum bit (qubit).

For a given quantum state ρ Shannon entropy does not apply since it provides us with information associated with a probability distribution $\mathbb{p}$ of a random variable X. In other words, it was not conceived to deal with quantum states. For this reason, we introduce the following subject.

Von Neumann Entropy

A reasonable measure for quantum information must include quantum states, in some limit, the Shannon entropy. Thus, John von Neumann introduced in 1927 in his paper [von Neumann (1927)] the *von Neumann entropy*

$$S(\rho) = -\operatorname{Tr}(\rho \log \rho). \tag{2.5.28}$$

with properties.

i) $S(\rho) \geq 0$, the equality if and only if ρ is a pure state;

ii) $S(\rho) \leq \log d$, the equality if and only if ρ is a maximally mixed state;

iii) $S(\rho)$, is invariant under unitary transformations, $S(\rho) = S(U\rho U^{-1})$;

iv) **Concavity**: For $\mathbb{p}_i \geq 0$, $i = 1, 2, \ldots, n$, with $\sum_{i=1}^{n} \mathbb{p}_i = 1$, we have;

$$S\left(\sum_{i=1}^{n} \mathbb{p}_i \rho_i\right) \geq \sum_{i=1}^{n} \mathbb{p}_i S(\rho_i); \tag{2.5.29}$$

v) **Subadditivity**: $S(\rho) \leq S(\rho_{\mathcal{X}}) + S(\rho_{\mathcal{Y}})$, the equality if and only if $\rho = \rho_{\mathcal{X}} \otimes \rho_{\mathcal{Y}}$;

vi) For $\mathbb{p}_i \geq 0$, $i = 1, 2, \ldots, n$, with $\sum_{i=1}^{n} \mathbb{p}_i = 1$, and an orthonormal basis we have that

$$S(\rho) \leq \sum_{i=1}^{n} \mathbb{p}_i S(\rho_i) + H(\mathbb{p}_i); \tag{2.5.30}$$

such that

$$\sum_{i=1}^{n} \mathbb{p}_i S(\rho_i) \leq S\left(\sum_{i=1}^{n} \mathbb{p}_i \rho_i\right) \leq \sum_{i=1}^{n} \mathbb{p}_i S(\rho_i) + H(\mathbb{p}_i). \tag{2.5.31}$$

As Shannon entropy we can summarize the von Neumann in the following short statement:

The von Neumann entropy $S(\rho)$ is an uncertainty measure associated with ρ.

Being a Hermitian operator, ρ can be written in diagonalized form (spectral decomposition):

$$\rho = \sum_i \wp_i |\psi_i\rangle \langle\psi| = \sum_i \wp_i \rho_i, \tag{2.5.32}$$

where $\{|\psi_i\rangle\}$ is an orthonormal basis and $\wp_i$ are the ρ eigenvalues (which have interpretation of probabilities). Let the eigenvalue relation be: $A|a\rangle = a|a\rangle$. Since $f(A)|a\rangle = f(a)|a\rangle$, where f is any function. Then simply using the trace in a convenient base we have

$$\operatorname{Tr} f(A) = \sum_a \langle a|f(A)|a\rangle = \sum_a f(a). \tag{2.5.33}$$

Using $f(x) = -x\log(x)$, we obtain

$$S(\rho) = -\sum_i \wp_i \log \wp_i, \tag{2.5.34}$$

which is the Shannon entropy with probability distribution $\wp_i$. Because (2.5.34) can always be written, the non-negativity of the Shannon entropy implies that

$$S(\rho) \geq 0, \tag{2.5.35}$$

the equality only if we have pure states, *i.e.*, one of the $\wp_i$'s is 1 and the others are 0. The upper bound is the same as for Shannon entropy, that is, a state associated with a d-dimensional Hilbert space satisfies

$$S(\rho) \leq \log d, \tag{2.5.36}$$

with equality only if the state is maximally mixed

$$\rho = \frac{\mathbb{1}}{d}. \tag{2.5.37}$$

For a qubit, e.g., the state with maximum entropy $S(\rho) = \log 2$ is therefore,

$$\rho = \frac{\mathbb{1}}{2}. \tag{2.5.38}$$

Relative Entropy

We have introduced the KL divergence or relative entropy in (2.5.15). Given two density operators, the quantum counterpart of the KL divergence is the so-called relative entropy, which is defined as

$$S(\rho||\sigma) := \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\rho \log \sigma). \tag{2.5.39}$$

As its classical analogue, $S(\rho||\sigma) \geq 0$, with equality if and only if $\rho = \sigma$.

Relative entropy has the following properties:

i) **Monotonicity**: $S(\mathcal{N}(\rho)||\mathcal{N}(\sigma)) \leq S(\rho||\sigma)$, where $\mathcal{N}$ are quantum operations on ρ.

ii) **Convexity**: $S\left(\lambda\rho_1 + (1-\lambda)\rho_2||\lambda\sigma_1 + (1-\lambda)\sigma_2\right) \leq \lambda S\left(\rho_1||\sigma_1\right) + (1-\lambda) S\left(\rho_2||\sigma_2\right)$, where $0 \leq \lambda \leq 1$.

Mutual Information

For the quantum analogue of (2.5.13), we simply replace the probabilities $\mathbb{p}$'s by the quantum state ρ and the entropies, such that

$$\mathcal{I}_{\mathcal{X}:\mathcal{Y}} := S(\rho_{\mathcal{X}}) + S(\rho_{\mathcal{Y}}) - S(\rho). \tag{2.5.40}$$

Information Quantifier

We have talked about uncertainty about a quantum system so far. However, we did not introduce what exactly is the information of a quantum system. Naturally, it can be introduced as a complementary quantity to entropy; the latter is an uncertainty measure about the state ρ. The maximally mixed state, here denoted by $\mu = \mathbb{1}/d$, encodes no useful information, as its entropy is maximum $S(\rho) = \log d$. So we can define the *amount of information* in a state ρ as an entropic distance between ρ and μ as

$$I(\rho) = S(\rho||\mu). \tag{2.5.41}$$

Calculating the relative entropy between ρ and μ we have

$$\begin{aligned} S(\rho||\mu) &= \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\rho \log \mathbb{1}/d) \\ &= \mathrm{Tr}(\rho \log \rho) + \log d \, \mathrm{Tr}(\rho) \\ &= \log d - S(\rho). \end{aligned} \tag{2.5.42}$$

Therefore, we can define the amount of information encoded in ρ as

$$I(\rho) := \log d - S(\rho). \tag{2.5.43}$$

Pure state is the state with maximum information $\mathcal{I}(\rho) = \log d$, since $S(\rho) = 0$. On the other hand, maximally mixed state is the state with minimum information $\mathcal{I}(\rho) = 0$, once $S(\rho) = \mathbb{1}/d$. In the case of qubits, the pure state has information $\mathcal{I}(\rho) = \log 2$.

2.5.3 Coherence and Decoherence

In this section, we only will scratch the surface of the subject of coherence which is not the main topic of this work by far, following the material in [Landi (2018)].

In section 2.1.4, we discussed the differences between pure and mixed states. To briefly recapitulate, mixed states are those states that consist of an ensemble of particles in a statistical mixture, *i.e.*, not every particle are in the same state. In this case, the off-diagonal terms of the density matrix tend to vanish in some basis, and the interference effects are suppressed in that basis. However, pure states represent ensembles in which each copy of the system are in the same state. In this case, the quantum superposition present in each state, and its interference properties, are not chased out when we combine the statistics. I follow that interference effects can be observed and the density matrix has significant off-diagonal terms at least in one basis.

These ideas of 'quantumness' and 'classicality' (absence of interference effects) can be defined in a more operational way through the concept of *coherence*. Given an orthonormal basis $\{|\psi_i\rangle\}$ in a d-dimensional Hilbert space, we say that a state is *incoherent* (in this particular basis), states which only have diagonal terms, if it admits a diagonal decomposition as

$$\rho = \sum_{i=0}^{d-1} \mathbb{p}_i \left|\psi_i\right\rangle \left\langle\psi_i\right| . \tag{2.5.44}$$

As suggested, coherence is a concept that depends on the basis that one uses. Thus, depending on the changing of the basis, one can no longer have coherence.

We can track the origin of coherence with the process of *decoherence* (see for instance [Zurek (2003)]). Decoherence is the process of loss of coherence to the environment. In the sense that it starts with a state like (2.1.23) and end up like (2.1.26). Generally speaking, the 'environment' is any system (or set of systems) that interacts with the system but is discarded (via partial trace) from the quantum state description, either for operational or foundational reasons. The very quantum vacuum can play the role of the environment since it actually is an infinite collection of electromagnetic modes which couples with physical systems. This explains, for instance, why an excited hydrogen atom always decays to its ground state after some lifetime. In this process, the atom state loses coherence in the energy basis. In practice, it is not an easy task to completely isolate a quantum system from its surroundings.

On the other hand, the *maximally coherent states* in the same basis is

$$\left|\psi_d\right\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d-1} \left|\psi_i\right\rangle . \tag{2.5.45}$$

Coherence Quantifier

In reference [Baumgratz (2014)], the authors introduce a measure of coherence based on the relative entropy (see section 2.5). First, we define the set $\mathbb{I}$ of the incoherent states in a given basis $\{|\psi_i\rangle\}$ as the set of all diagonal states in the form

$$\delta = \sum_{i=1}^{d} \delta_i |\psi_i\rangle \langle\psi_i| . \tag{2.5.46}$$

The coherence quantifier is then defined as

$$C(\rho) = \min_{\delta\in\mathbb{I}} S(\rho||\delta). \tag{2.5.47}$$

To measure the "distance" between ρ and its closest incoherent state delta norms could be used, but for our purposes, it is sufficient the last definition. For a given general state ρ, let us consider its 'incoherent counterpart' in the basis $\{|\psi_i\rangle\}$, disregarding the terms outside the diagonal:

$$\rho_{\text{diag}} = \sum_{i=1}^{d-1} \rho_{ii} |\psi_i\rangle \langle\psi_i| . \tag{2.5.48}$$

One can write then the relative entropy in equation (2.5.47) as

$$S(\rho||\delta) = -S(\rho) - \mathrm{Tr}(\rho \log \delta). \tag{2.5.49}$$

For all $\delta \in \mathbb{I}$ once shows that

$$\mathrm{Tr}(\rho \log \delta) = \mathrm{Tr}(\rho_{\text{diag}} \log \delta), \tag{2.5.50}$$

and

$$S(\rho_{\text{diag}}||\delta) = -S(\rho_{\text{diag}}) - \mathrm{Tr}(\rho_{\text{diag}} \log \delta), \tag{2.5.51}$$

or, rewriting the last equation,

$$\mathrm{Tr}(\rho_{\text{diag}} \log \delta) = -S(\rho_{\text{diag}}||\delta) - S(\rho_{\text{diag}}). \tag{2.5.52}$$

Thus, we can replace it in the original relative entropy (2.5.49) to obtain

$$S(\rho||\delta) = S(\rho_{\text{diag}}) - S(\rho) + S(\rho_{\text{diag}}||\delta) \geq S(\rho_{\text{diag}}) - S(\rho). \tag{2.5.53}$$

The minimum is when $S(\rho_{\text{diag}}||\delta) = 0$ which occurs only when $\rho_{\text{diag}} = \delta$. Then, we can define in a better way equation (2.5.47) as the minimum of (2.5.53), then the amount of coherence is

$$C(\rho) \coloneqq S(\rho_{\text{diag}}) - S(\rho). \tag{2.5.54}$$

Notice that if ρ is written as equation (2.1.26), *i.e.*, in diagonal form, such that $\rho = \rho_{\text{diag}}$ we have $C(\rho) = 0$, which means that ρ does not have any coherence, as expected.

Coherence as Resource

In this section will be briefly discussed a way to study information processing under a limited set of allowed operations which are named *free operations*, since they do not generate resource, in this case, coherence (2.5.54) is a resource. *Incoherent operations* defines a set of all states which are diagonal in a particular basis which must be explicitly specified.

States with no resource are said to be *free states* and in this particular case of coherence, free states are called *incoherent states.* What quantifier (2.5.54) does is measure the distance between ρ and the closest state without coherence within the set of all states which do not have coherence, which is the set $\mathbb{I}$. The main question in resource theory is which set of operations one can perform and do not increase resource (see references [Chitambar (2019), Baumgratz (2014)]).

2.6 Discrete Position and Momentum Representation

In quantum mechanics, the conjugate variables position Q and momentum P are represented by Hermitian operators in a Hilbert space $\mathbb{H}$. However this Hilbert space $\mathbb{H}$ is not exactly that one we are considering in this work; it is an extended version of $\mathcal{H}$ presented in the section 2.1.3. That is, $\mathcal{H} \subseteq \mathbb{H}$, since it deals with infinite continuous basis, *i.e.*, 'unbounded operators' can act on it, which means that $\|R\,|\alpha\rangle\| \to \infty$. Here, $|\alpha\rangle \in \mathbb{H}$ and $R \in \mathscr{U}(\mathbb{H})$, where $\mathscr{U}(\mathbb{H})$ is the set of all unbounded operators that act on $\mathbb{H}$. Hence, $Q, P \in \mathscr{U}(\mathbb{H})$ and $Q, P \notin \mathscr{B}(\mathcal{H})$.

In the reference [Freire (2019)] the authors introduce the concept of 'discrete position and momentum representation' and their formulation restricts the position Q and momentum P to be bounded operators acting only in $\mathcal{H}$, *i.e.*, in their description $Q, P \in \mathscr{B}(\mathcal{H})$, that requires certain mathematical care. This discrete space will be very useful for us to study the incompatibility of position and momentum operators. For this reason, this subject is now briefly reviewed.

2.6.1 Basic Definitions and States

The main idea relies on the fact that real measurement apparatuses have a finite resolution. That is, one-dimensional detectors, for example, can be divided into other small one-dimensional detectors with resolution $\delta q \in \mathbb{R}^+$, in position measurements, and $\delta p \in \mathbb{R}^+$, in momentum measurements. An incident particle will be detected by the k-th detector if it is in such detector's resolution range. In the limit $\delta q, \delta p \to 0$ one should recover continuous position and momentum spaces.

The discrete position and momentum bases are denoted $\{|q_k\rangle\}$ and $\{|p_l\rangle\}$, respectively, where the index $k(l)$ will be used only for position (momentum) quantities. The eigenvalue equation for the position operator Q reads

$$Q\,|q_k\rangle = q_k\,|q_k\rangle\,, \tag{2.6.1}$$

where the eigenvalue $q_k \equiv k\,\delta q$ is associated with the eigenvector $|q_k\rangle$, where δq has a dimensional unit of length and $k \in \mathbb{Z}$. As we are dealing with Hilbert spaces, we must define an inner product between position eigenvectors, which is

$$\langle q_k | q_{k'} \rangle = \frac{\delta_{kk'}}{\delta q}, \tag{2.6.2}$$

where it is divided by δq to include the dimensional unit of length. In the limit $\delta q \to 0$ we obtain the Dirac delta function. The authors in reference [Freire (2019)] propose the projection operator associated with the operator Q as

$$Q_k = \delta q \left|q_k\right\rangle \left\langle q_k\right| . \tag{2.6.3}$$

The quantity in equation (2.6.3) has the desired properties:

$$Q_k Q_{k'} = \delta_{kk'} Q_{k'} \qquad \text{and} \qquad \sum_{k=-L_q}^{L_q} Q_k = \mathbb{1}, \tag{2.6.4}$$

where L_q is a parameter related to the position space dimension, whose definition will be provided later.

Given the above discrete position basis, any state vector $|\psi\rangle$ can be expanded as

$$|\psi\rangle = \sum_{k=-L_q}^{L_q} \delta q\, \psi(q_k) \left|q_k\right\rangle , \tag{2.6.5}$$

where $\psi(q_k) = \langle q_k|\psi\rangle$. Thus, the probability of one obtains q_k in a measurement of Q is given by

$$\mathbb{p}_k \equiv \mathbb{p}(q_k) = \delta q \; |\psi(q_k)|^2 , \tag{2.6.6}$$

which satisfies the normalization condition

$$\sum_k \mathbb{p}_k = 1. \tag{2.6.7}$$

Analogous definitions apply to the momentum space, the eigenvalue equation for the momentum operator P reads

$$P \left|p_l\right\rangle = p_l \left|p_l\right\rangle , \tag{2.6.8}$$

where the eigenvalue $p_l \equiv l\, \delta p$ is associated with the eigenvector $|p_l\rangle$, where δp has a dimensional unit of length and $l \in \mathbb{Z}$. The inner product between momentum eigenvectors is then

$$\langle p_l|p_{l'}\rangle = \frac{\delta_{ll'}}{\delta p}, \tag{2.6.9}$$

which also contains information about the spatial resolution. The projector associated with the momentum operator P is

$$P_l = \delta p \left|p_l\right\rangle \left\langle p_l\right| , \tag{2.6.10}$$

which also has the following properties

$$P_l P_{l'} = \delta_{ll'} P_{l'} \qquad \text{and} \qquad \sum_{l=-L_P}^{L_P} P_l = \mathbb{1}. \tag{2.6.11}$$

In momentum basis expansion the state vector $|\psi\rangle$ can be written as

$$|\psi\rangle = \sum_{l=-L_P}^{L_P} \delta p\, \psi(p_l) \left|p_l\right\rangle , \tag{2.6.12}$$

where $\psi(p_l) = \langle p_l|\psi\rangle$. And the probability of one obtains p_l in a measurement is

$$\mathbb{p}_l = \mathbb{p}(p_l) = \delta p \ |\psi(p_l)|^2, \tag{2.6.13}$$

with

$$\sum_l \mathbb{p}_l = 1. \tag{2.6.14}$$

The inner product between eigenvectors of the different bases is given by

$$\langle q_k|p_l\rangle = \frac{e^{2\pi i kl/\xi}}{\sqrt{2\pi\hbar}}. \tag{2.6.15}$$

As shown in reference [Freire (2019)], the dimensions of the spaces are constrained as $\xi = 2L + 1$, with $L = L_q = L_p$, where

$$\xi = \frac{2\pi\hbar}{\delta q\, \delta p}. \tag{2.6.16}$$

Moreover, it is shown that the resolutions δq and δp cannot be arbitrarily large, if one (naturally) requires that $L > 0$, then it is needed that $\xi > 1$, which implies $\delta q \delta p < 2\pi\hbar$.

Another important prescription of the information is the trace of an operator $\mathcal{O}$ in the momentum and position basis:

$$\mathrm{Tr}\, \mathcal{O} = \sum_k \delta q \, \langle q_k|\mathcal{O}|q_k\rangle = \sum_l \delta p \, \langle p_l|\mathcal{O}|p_l\rangle \,. \tag{2.6.17}$$

Notice that the trace is invariant under basis changes, a property that also holds for continuous variables like position and momentum.

2.6.2 Entropy in Position and Momentum Representation

Using the discretization procedure, we can write the von Neumann entropy for any quantum state σ as

$$\begin{aligned} S(\sigma) &= -\,\mathrm{Tr}\,(\sigma \ln \sigma) \\ &= -\sum_k \delta q \, \langle q_k|\sigma \ln \sigma|q_k\rangle \\ &= -\sum_l \delta p \, \langle p_l|\sigma \ln \sigma|p_l\rangle \,. \end{aligned} \tag{2.6.18}$$

One can write the general mixed state (2.1.24), e.g., in position basis as

$$\sigma = \sum_k w_k \, \delta q \, |q_k\rangle \, \langle q_k| \,, \tag{2.6.19}$$

where w_k is a probability distribution. Hence, using equation (2.6.2) and the trace properties given by equations (2.1.7)–(2.1.9), one computes

$$\begin{aligned}
\mathrm{Tr}\,(Q_k\sigma) &= \sum_{k'} \mathrm{Tr}\left((\delta q)^2 w_{k'} \left|q_k\right\rangle \left\langle q_k | q_{k'}\right\rangle \left\langle q_{k'}\right|\right) \\
&= \sum_{k'} w_{k'} (\delta q)^2 \frac{\delta_{kk'}}{\delta q} \mathrm{Tr}\left(\left|q_k\right\rangle \left\langle q_{k'}\right|\right) \\
&= w_k \delta q \left\langle q_k | q_k\right\rangle \\
&= w_k \delta q \frac{\delta_{kk}}{\delta q} \\
&= w_k,
\end{aligned} \tag{2.6.20}$$

resulting in the probability w_k. This means that we obtain the usual result of discrete variable quantum mechanics used so far in the text, thereby validating this procedure.

Computing the eigenvalue equation using equation (2.6.2),

$$\sigma \left|q_k\right\rangle = \sum_{k'} w_{k'}\, \delta q \left|q_{k'}\right\rangle \left\langle q_{k'} | q_k\right\rangle = w_k \left|q_k\right\rangle, \tag{2.6.21}$$

we see that the σ eigenvalues are w_k. We then obtain the von Neumann entropy in terms of Shannon entropy,

$$S(\sigma) = -\sum_{k=-L}^{L} w_k \ln w_k = H(w_k). \tag{2.6.22}$$

Notice that this series diverges as $L \longrightarrow \infty$.

CHAPTER 3

CONTEXT INCOMPATIBILITY VIA BAYES' RULE VIOLATION

3.1 Introduction

Previously we proved theorem 1, which states that for any two commuting observables X and Y, with eigenvalues x_i and y_j, respectively, there is a common eigenbasis $\{|x_i, y_j\rangle\}$ belonging to them, such that

$$\begin{aligned} X\,|x_i, y_j\rangle &= x_i\,|x_i, y_j\rangle\,, \\ Y\,|x_i, y_j\rangle &= y_j\,|x_i, y_j\rangle\,. \end{aligned} \tag{3.1.1}$$

This implies that measurements of X in the input state $|x_i, y_j\rangle$ will result in x_i, leaving the state unchanged, and a sequential measurement of Y will result in y_j, also leaving the system in the eigenstate $|x_i, y_j\rangle$. For a preparation like $|x_i, y_j\rangle$ the observables are said to be *compatible*. But if X and Y do not commute, one cannot either find or prepare a simultaneous eigenstate for them. Accordingly, it is widely accepted that these observables cannot be simultaneously measured, which renders X and Y to be termed incompatible.

3.2 Incompatibility Via Sequential Measurements

As discussed in many textbooks, the notion of incompatibility can be appreciated through a scheme involving sequential measurements. Let ρ, X, and Y be operators acting on a single-partite Hilbert space $\mathcal{H}$. Now, consider the experiments illustrated in 3.1, where it is assumed that $[X, Y] = 0$. We see that the Y measurement does not change the state prepared by the first X measurement, which justify the validity of Bayes's rule $\mathbb{p}_{x_i,y_j} = \mathbb{p}_{y_j,x_i}$.

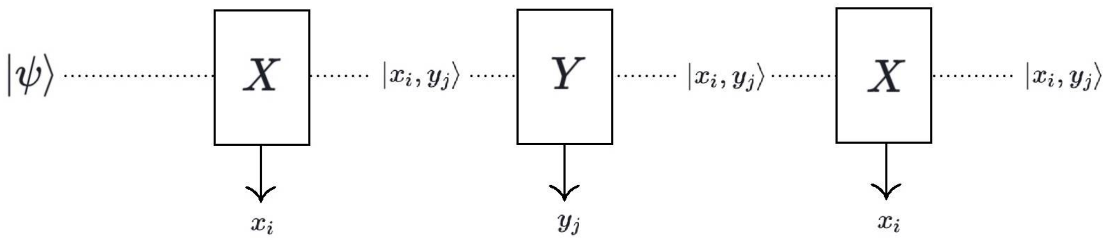


Figure 3.1: Schematic picture of compatible observables being measured in the input state as common eigenstates of both X and Y. In the whole process, the state is unchanged (or not spoiled) by the measurements. Therefore, they can be simultaneously measured.

On the other hand, when $[X, Y] \neq 0$, the Y measurement destroys the information, the observer gets with the first X measurement (see 3.2). In fact, the second X measurement will not generally give the same outcome as the first. In addition, it is clearly seen that the final state depends on the ordering one chooses for the measurements.

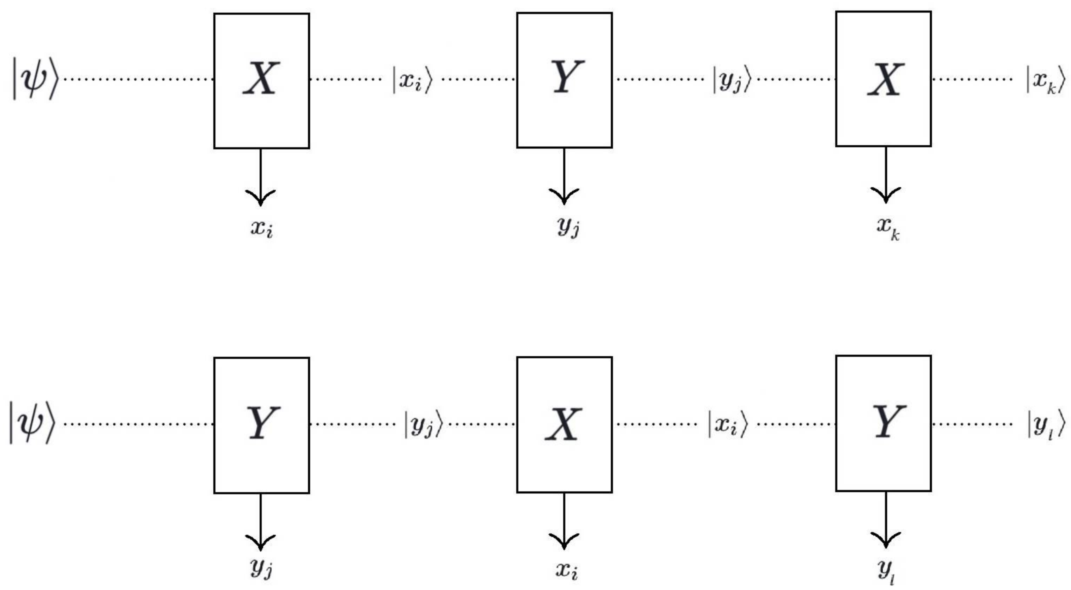


Figure 3.2: Schematic picture of incompatible observables being measured in the input state as a general state $|\psi\rangle$. The system is never left in the same state if different observables are measured sequentially, the state is spoiled by new measurements. Therefore, the state cannot be simultaneously measured.

In the present discussion, the role of the input state $|\psi\rangle$ seems to be innocuous. However, this cannot be the ultimate narrative for the incompatibility concept, because we know that such a notion does not exist in the classical limit. Moreover, the classical limit is usually implemented through decoherence process over the quantum state. We then expect that a more refined notion of incompatibility has to take the quantum state into account necessarily.

3.3 Bayes' Rule Violation

We have seen in section 2.3.5 that the so-called Bayes' rule is a formula in classical probability theory, allowing one to obtain the inverse conditional probability. However, this formula does not always apply to quantum systems, since measurements are invasive and hence, the ordering matters. As we show in this section, this is direct related to incompatibility.

3.3.1 Composite States

Consider a system associated with a Hilbert space $\mathcal{H}_{\mathcal{XY}}$ whose bipartite input state is $\rho_{\mathcal{XY}}$. Suppose one measures observables $X \in \mathscr{B}(\mathcal{H}_{\mathcal{X}})$ and $Y \in \mathscr{B}(\mathcal{H}_{\mathcal{Y}})$, also in the reverse order, with probability distributions $\mathbb{p}_{x_i,y_j}$ and $\mathbb{p}_{y_j,x_i}$. It is trivial to see that in this case $[X,Y]=0$ and then, according to the last section, X and Y are compatible observables. Writing the Bayes' rule for these probabilities

$$\mathbb{p}_{x_i,y_j} = \mathbb{p}_{y_j,x_i}, \tag{3.3.1}$$

using equations (2.4.28) and (2.4.29), both sides of last equation can be rewritten as

$$\mathbb{p}_{x_i} \operatorname{Tr}_{\mathcal{Y}} \left(Y_j \rho_{Y|x_i}\right) = \mathbb{p}_{y_j} \operatorname{Tr}_{\mathcal{X}} \left(X_i \rho_{X|y_j}\right), \tag{3.3.2}$$

and from equations (2.1.47) and (2.1.48), last equations turns out to be

$$\frac{\mathbb{p}_{x_i} \operatorname{Tr}_{\mathcal{Y}} \left(Y_j \operatorname{Tr}_{\mathcal{X}}(X_i \rho_{\mathcal{XY}})\right)}{\mathbb{p}_{x_i}} = \frac{\mathbb{p}_{y_j} \operatorname{Tr}_{\mathcal{X}} \left(X_i \operatorname{Tr}_{\mathcal{Y}}(X_j \rho_{\mathcal{XY}})\right)}{\mathbb{p}_{y_j}}, \tag{3.3.3}$$

using the properties of partial trace, it leads to

$$\operatorname{Tr}\left(X_i \otimes Y_j\, \rho_{\mathcal{XY}}\right) = \operatorname{Tr}\left(X_i \otimes Y_j\, \rho_{\mathcal{XY}}\right), \tag{3.3.4}$$

therefore when comparing two probability distributions of compatible observables with composite systems as input states, the Bayes' rule is valid. *Observables $X \otimes \mathbb{1}$ and $\mathbb{1} \otimes Y$ that acts on $\mathcal{H}_{\mathcal{XY}}$ are compatible, there is no Bayes' rule violation what so ever in this space.*

3.3.2 Single-Partite States

Now consider a system associated to a Hilbert space $\mathcal{H}$ with input state ρ and observables $X, Y \in \mathscr{B}(\mathcal{H})$, to be measured in both orders, with probability distributions $\mathbb{p}_{x_i,y_j}$ and $\mathbb{p}_{y_j,x_i}$. There are two possibilities regarding the observables in this space, $[X,Y]=0$ and $[X,Y]\neq 0$.

Commuting Observables

Assuming the validity of Bayes' rule for the given joint probabilities,

$$\mathbb{p}_{x_i,y_j} = \mathbb{p}_{y_j,x_i}, \tag{3.3.5}$$

and using equations (2.4.19) and (2.4.21), we have

$$\operatorname{Tr}(X_i\rho)\operatorname{Tr}(Y_jX_i) = \operatorname{Tr}(Y_j\rho)\operatorname{Tr}(X_iY_j). \tag{3.3.6}$$

When $[X,Y]=0$, both observables have common eigenbasis. This means that $X_i = Y_j$, resulting in

$$\operatorname{Tr}(X_i\rho)\operatorname{Tr}(X_jX_i) = \operatorname{Tr}(Y_j\rho)\operatorname{Tr}(X_iX_j), \tag{3.3.7}$$

which is indeed an equality. Therefore, if X and Y are compatible, $\mathbb{p}_{x_i,y_j}$ and $\mathbb{p}_{y_j,x_i}$ obey Bayes' rule.

Non-commuting Observables

If X and Y do not commute, joint probability of obtaining outcomes x_i and y_j can be written as

$$\begin{aligned}
\mathbb{p}_{x_i,y_j} &= \operatorname{Tr}(X_i\rho)\operatorname{Tr}(Y_j\rho_{S|x_i}) \\
&= \operatorname{Tr}(X_i\rho)\operatorname{Tr}\left(Y_j\frac{X_i\rho X_i}{\operatorname{Tr}(X_i\rho)}\right) \\
&= \operatorname{Tr}(Y_jX_i\rho X_i) \\
&= \operatorname{Tr}\Big(\left|y_j\right\rangle\left\langle y_j|x_i\right\rangle\left|x_i\right\rangle\left\langle x_i|\rho|x_i\right\rangle\Big) \\
&= \left|\left\langle x_i|y_j\right\rangle\right|^2\left\langle x_i|\rho|x_i\right\rangle .
\end{aligned} \tag{3.3.8}$$

The opposite ordering has joint probability of obtaining outcomes y_j and x_i reads

$$\begin{aligned}
\mathbb{p}_{y_j,x_i} &= \operatorname{Tr}(Y_j\rho)\operatorname{Tr}(X_i\rho_{S|y_j}) \\
&= \operatorname{Tr}(Y_j\rho)\operatorname{Tr}\left(X_i\frac{Y_j\rho Y_j}{\operatorname{Tr}(Y_j\rho)}\right) \\
&= \operatorname{Tr}(X_iY_j\rho Y_j) \\
&= \operatorname{Tr}\Big(\left|x_i\right\rangle\left\langle x_i|y_j\right\rangle\left|y_j\right\rangle\left\langle y_j|\rho|y_j\right\rangle\Big) \\
&= \left|\left\langle x_i|y_j\right\rangle\right|^2\left\langle y_j|\rho|y_j\right\rangle
\end{aligned} \tag{3.3.9}$$

therefore, $\mathbb{p}_{x_i,y_j} \neq \mathbb{p}_{y_j,x_i}$, that is, quantum mechanics admits Bayes' rule violation. Therefore, when ρ is not MU with X and Y, X and Y not commuting, the Bayes' rule is not valid.[1] This has to do with the ordering of the measurements: since the order matters, in this case, joint probability distributions differ. Which is a surprising result.

3.4 Bayes' Divergence

In this section is proposed a quantifier for incompatibility of two observables based on Bayes' rule violation and a possible '*normalization*'[2] for it. Here the quantifier compares two joint probability distributions, referring to opposite measurement orderings. If $\mathbb{p}_{x_i,y_j} = \mathbb{p}_{y_j,x_i}$, then we should have compatibility in place, but whenever $\mathbb{p}_{x_i,y_j} \neq \mathbb{p}_{y_j,x_i}$, then, incompatibility arises. See figures 3.3–3.4 for a depicted example for compatibility and figures 3.5–3.6 for a depicted example for incompatibility.

As we show below, KL divergence turns out to be the best candidate for a reasonable measure of incompatibility. Other measures, such as quantum relative entropy and trace norms of post-measurement states $\Phi_{XY}(\rho)$ and $\Phi_{YX}(\rho)$, as discarded since they fail to provide the expected result in some limits.

Definition 5. (Bayes' divergence) *For a given system described by a quantum state ρ associated with a d-dimensional non-degenerate discrete Hilbert space $\mathcal{H}$, a pair of observables X and Y with*

[1]Since, incompatibility is dependent on the state, when ρ is MU with the observables one does not have complete knowledge about the measuring system. In this case, although satisfying Bayes' rule, it is not possible to say whether the observables are incompatible or not.

[2]Normalization here is meant to be a re-scaling of the image of quantifier. For instance, instead of lying in the interval $[0, +\infty]$ a normalization puts the quantifier $[0, +1]$.

eigenvalues x_i and y_j, respectively, the 'Bayes' divergence' of the context $\mathbb{C} \equiv \{\rho, X, Y\} \subset \mathscr{B}(\mathcal{H})$ is defined as the symmetric Kullback-Leibler divergence between the joint probability distributions $\mathbb{p}_{x_i,y_j}$ and $\mathbb{p}_{y_j,x_i}$,

$$\mathcal{D}_{\mathbb{C}} := \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\left\{\mathbb{p}_{x_i,y_j}\ln\left[\frac{\mathbb{p}_{x_i,y_j}}{\mathbb{p}_{y_j,x_i}}\right] + \mathbb{p}_{y_j,x_i}\ln\left[\frac{\mathbb{p}_{y_j,x_i}}{\mathbb{p}_{x_i,y_j}}\right]\right\}, \tag{3.4.1}$$

with mathematical conditions $-0\ln 0 \equiv 0$ and $-\mathbb{p}_{x_i,y_j}\ln 0 = -\mathbb{p}_{y_j,x_i}\ln 0 \equiv \infty$ if $\mathbb{p}_{x_i,y_j} \geq 0$ and $\mathbb{p}_{y_j,x_i} \geq 0$.

3.4.1 Properties

It is worth to recall how probabilities, conditional probabilities and joint probabilities were defined in quantum mechanics. First, from postulate 4 we see that

$$\mathbb{p}_{x_i} := \mathrm{Tr}(X_i\rho), \tag{3.4.2}$$

and

$$\mathbb{p}_{y_j} := \mathrm{Tr}(Y_j\rho). \tag{3.4.3}$$

For conditional probability, we have

$$\mathbb{p}_{x_i|y_j} = \mathbb{p}_{y_j|x_i} := \mathrm{Tr}(X_iY_j). \tag{3.4.4}$$

Joint probabilities, consequently, are written as products of equations (3.4.2) and (3.4.3) with equation (3.4.4), that is

$$\mathbb{p}_{x_i,y_j} = \mathbb{p}_{x_i}\mathbb{p}_{y_j|x_i} := \mathrm{Tr}(X_i\rho)\,\mathrm{Tr}(X_iY_j), \tag{3.4.5}$$

and

$$\mathbb{p}_{y_j,x_i} = \mathbb{p}_{y_j}\mathbb{p}_{x_i|y_j} := \mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}(X_iY_j). \tag{3.4.6}$$

By replacing equation (3.4.5) and (3.4.6) in the definition (3.4.1), it results in

$$\begin{aligned}
\mathcal{D}_{\mathbb{C}} &= \frac{1}{2}\sum_{i,j=1}^{d}\left\{\mathrm{Tr}(X_i\rho)\,\mathrm{Tr}(X_iY_j)\ln\left[\frac{\mathrm{Tr}(X_i\rho)\,\mathrm{Tr}(X_iY_j)}{\mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}(X_iY_j)}\right] + \mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}(X_iY_j)\ln\left[\frac{\mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}(X_iY_j)}{\mathrm{Tr}(X_i\rho)\,\mathrm{Tr}(X_iY_j)}\right]\right\}\\
&= \frac{1}{2}\sum_{i,j=1}^{d}\left\{\mathrm{Tr}(X_i\rho)\,\mathrm{Tr}(X_iY_j)\ln\left[\frac{\mathrm{Tr}(X_i\rho)}{\mathrm{Tr}(Y_j\rho)}\right] + \mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}(X_iY_j)\ln\left[\frac{\mathrm{Tr}(Y_j\rho)}{\mathrm{Tr}(X_i\rho)}\right]\right\}\\
&= \frac{1}{2}\sum_{i,j=1}^{d}\left\{\mathrm{Tr}(X_i\rho)\,\mathrm{Tr}(X_iY_j)\ln\left[\mathrm{Tr}(X_i\rho)\right] - \mathrm{Tr}(X_i\rho)\,\mathrm{Tr}(X_iY_j)\ln\left[\mathrm{Tr}(Y_j\rho)\right]\right\}\\
&+ \frac{1}{2}\sum_{i,j=1}^{d}\left\{\mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}(X_iY_j)\ln\left[\mathrm{Tr}(Y_j\rho)\right] - \mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}(X_iY_j)\ln\left[\mathrm{Tr}(X_i\rho)\right]\right\}\\
&= \frac{1}{2}\left\{\sum_{i,j=1}^{d}\mathrm{Tr}(X_i\rho)\,\mathrm{Tr}(X_iY_j)\ln\left[\mathrm{Tr}(X_i\rho)\right] - \sum_{i,j=1}^{d}\mathrm{Tr}(X_i\rho)\,\mathrm{Tr}(X_iY_j)\ln\left[\mathrm{Tr}(Y_j\rho)\right]\right\}\\
&+ \frac{1}{2}\left\{\sum_{i,j=1}^{d}\mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}(X_iY_j)\ln\left[\mathrm{Tr}(Y_j\rho)\right] - \sum_{i,j=1}^{d}\mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}(X_iY_j)\ln\left[\mathrm{Tr}(X_i\rho)\right]\right\}.
\end{aligned} \tag{3.4.7}$$

Since, the trace is a linear mapping and using the completeness property of the projectors, *i.e.*, $\sum_i X_i = \mathbb{1}$ and $\sum_j Y_j = \mathbb{1}$, last equation can be simplified to

$$\begin{aligned}
\mathcal{D}_{\mathbb{C}} &= \frac{1}{2}\left\{\sum_{i=1}^{d}\mathrm{Tr}(X_i\rho)\,\mathrm{Tr}\left(X_i\left(\sum_{j=1}^{d}Y_j\right)\right)\ln\Big[\mathrm{Tr}(X_i\rho)\Big] - \sum_{i,j=1}^{d}\mathrm{Tr}(X_i\rho)\,\mathrm{Tr}(X_iY_j)\ln\big[\mathrm{Tr}(Y_j\rho)\big]\right\} \\
&+ \frac{1}{2}\left\{\sum_{j=1}^{d}\mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}\left(\left(\sum_{i=1}^{d}X_i\right)Y_j\right)\ln\big[\mathrm{Tr}(Y_j\rho)\big] - \sum_{i,j=1}^{d}\mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}(X_i\rho)\ln\Big[\mathrm{Tr}(X_i\rho)\Big]\right\} \\
&= \frac{1}{2}\left\{\sum_{i=1}^{d}\mathrm{Tr}(X_i\rho)\,\mathrm{Tr}\left(X_i\right)\ln\Big[\mathrm{Tr}(X_i\rho)\Big] - \sum_{i,j=1}^{d}\mathrm{Tr}(X_i\rho)\,\mathrm{Tr}(X_iY_j)\ln\big[\mathrm{Tr}(Y_j\rho)\big]\right\} \\
&+ \frac{1}{2}\left\{\sum_{j=1}^{d}\mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}\left(Y_j\right)\ln\big[\mathrm{Tr}(Y_j\rho)\big] - \sum_{i,j=1}^{d}\mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}(X_iY_j)\ln\Big[\mathrm{Tr}(X_i\rho)\Big]\right\},
\end{aligned} \tag{3.4.8}$$

where we used $\mathrm{Tr}(X_i) = \langle x_i|x_i\rangle = 1$ and $\mathrm{Tr}(Y_j) = \langle y_j|y_j\rangle = 1$, combining terms with same logarithms, we find

$$\begin{aligned}
\mathcal{D}_{\mathbb{C}} = &-\frac{1}{2}\sum_{i=1}^{d}\left\{\sum_{j=1}^{d}\mathrm{Tr}(Y_j\rho)\,\mathrm{Tr}(X_iY_j) - \mathrm{Tr}(X_i\rho)\right\}\ln\Big[\mathrm{Tr}(X_i\rho)\Big] \\
&-\frac{1}{2}\sum_{j=1}^{d}\left\{\sum_{i=1}^{d}\mathrm{Tr}(X_i\rho)\,\mathrm{Tr}(X_iY_j) - \mathrm{Tr}(Y_j\rho)\right\}\ln\big[\mathrm{Tr}(Y_j\rho)\big].
\end{aligned} \tag{3.4.9}$$

Identifying each term with the probabilities given in equations (3.4.2)–(3.4.6), the obtained result is

$$\mathcal{D}_{\mathbb{C}} = -\frac{1}{2}\left\{\sum_{i=1}^{d}\left[\sum_{j=1}^{d}\mathbb{p}_{y_j,x_i} - \mathbb{p}_{x_i}\right]\ln\mathbb{p}_{x_i} + \sum_{j=1}^{d}\left[\sum_{i=1}^{d}\mathbb{p}_{x_i,y_j} - \mathbb{p}_{y_j}\right]\ln\mathbb{p}_{y_j}\right\}. \tag{3.4.10}$$

Notice that if the pair of observables are maximally unbiased[3], then we have violation of Bayes' rule in place, *i.e.*, $\mathbb{p}_{x_i,y_j} \neq \mathbb{p}_{y_j,x_i}$, meaning that

$$\sum_{j=1}^{d}\mathbb{p}_{y_j,x_i} \neq \mathbb{p}_{x_i}, \tag{3.4.11}$$

and

$$\sum_{i=1}^{d}\mathbb{p}_{x_i,y_j} \neq \mathbb{p}_{y_j}. \tag{3.4.12}$$

As we have seen, this result is at odd with the theory of classical probability. In the case of non-commuting observables in any degree, $\mathcal{D}_{\mathbb{C}} > 0$. On the other hand if the pair of observables

[3]Mutually unbiased bases (MUB) are two orthonormal bases $\{|a_n\rangle\}$ and $\{|b_m\rangle\}$, that their square of the magnitude of the inner product $|\langle a_n|b_m\rangle|^2 = 1/d, \quad \forall n, m$. Thus, a system prepared in a state ρ belonging to $\{|a_n\rangle\}$ or $\{|b_m\rangle\}$, all its measurement outcomes can be predicted to respect to the other bases with equal probability.

commute, then relations (3.4.12) is an equality and either (3.4.11), implying that $\mathcal{D}_{\mathbb{C}}$. Therefore the Bayes' divergence is limited to the interval

$$\mathcal{D}_{\mathbb{C}} \in [0, +\infty], \tag{3.4.13}$$

meaning that Bayes' divergence 3.4.10 is a good measure for incompatibility, since it is null if $[X, Y] = 0$. Besides, we can state the following theorem.

Theorem 6. *Bayes's divergence $\mathcal{D}_{\mathbb{C}}$ is maximum if and only if X and Y are mutually unbiased (MU) observables.*

Proof. If X and Y are MU, we have

$$\mathbb{p}_{x_i|y_j} = \mathbb{p}_{y_j|x_i} = \left|\langle x_i|y_j\rangle\right|^2 = \frac{1}{d} \tag{3.4.14}$$

we find

$$\begin{aligned} \mathbb{p}_{x_i,y_j} &= \frac{\mathbb{p}_{x_i}}{d}, \\ \mathbb{p}_{y_j,x_i} &= \frac{\mathbb{p}_{y_j}}{d}. \end{aligned} \tag{3.4.15}$$

Replacing in (3.4.10)

$$\begin{aligned} \mathcal{D}_{\mathbb{C}} &= -\frac{1}{2}\left\{\sum_{i=1}^{d}\left[\sum_{j=1}^{d}\frac{\mathbb{p}_{y_j}}{d} - \mathbb{p}_{x_i}\right]\ln \mathbb{p}_{x_i} + \sum_{j=1}^{d}\left[\sum_{i=1}^{d}\frac{\mathbb{p}_{x_i}}{d} - \mathbb{p}_{y_j}\right]\ln \mathbb{p}_{y_j}\right\} \\ &= -\frac{1}{2}\left\{\sum_{i=1}^{d}\left[\mathbb{p}_{y_j} - \mathbb{p}_{x_i}\right]\ln \mathbb{p}_{x_i} + \sum_{j=1}^{d}\left[\mathbb{p}_{x_i} - \mathbb{p}_{y_j}\right]\ln \mathbb{p}_{y_j}\right\} \end{aligned} \tag{3.4.16}$$

suppose that $\mathbb{p}_{x_i} = 1/d$, since the observables are MU, we have that $\mathbb{p}_{y_j} = 0$, leading us to

$$\mathcal{D}_{\mathbb{C}} \to +\infty. \tag{3.4.17}$$

□

X Eigenstates as Input State

Let the input state be an eigenstate of X, that is $\rho = X_i$. In this case, it is clear that $[\rho, X] = 0$. From the second line in equation (3.4.7) it follows that

$$\begin{aligned} \mathcal{D}_{\{X_i,X,Y\}} &= \frac{1}{2}\sum_{i,j=1}^{d}\left\{\mathrm{Tr}(X_iX_i)\,\mathrm{Tr}(X_iY_j)\ln\left[\frac{\mathrm{Tr}(X_iX_i)}{\mathrm{Tr}(Y_jX_i)}\right] + \mathrm{Tr}(Y_jX_i)\,\mathrm{Tr}(X_iY_j)\ln\left[\frac{\mathrm{Tr}(Y_jX_i)}{\mathrm{Tr}(X_iX_i)}\right]\right\} \\ &= \frac{1}{2}\sum_{i,j=1}^{d}\left\{\mathrm{Tr}(X_iY_j)\ln\left[\frac{1}{\mathrm{Tr}(X_iY_j)}\right] + \mathrm{Tr}(X_iY_j)\,\mathrm{Tr}(X_iY_j)\ln\left[\mathrm{Tr}(X_iY_j)\right]\right\} \\ &= \frac{1}{2}\sum_{i,j=1}^{d}\left[\mathrm{Tr}(X_iY_j) - 1\right]\mathrm{Tr}(X_iY_j)\ln\left[\mathrm{Tr}(X_iY_j)\right]. \end{aligned} \tag{3.4.18}$$

Interestingly, the same result can be shown to emerge for $\rho = Y_j$, that is

$$\mathcal{D}_{\{X_i,X,Y\}} = \mathcal{D}_{\{Y_j,X,Y\}} = \frac{1}{2}\sum_{i,j=1}^{d}\Big[\mathbb{p}_{x_i|y_j} - 1\Big]\mathbb{p}_{x_i|y_j}\ln \mathbb{p}_{x_i|y_j}, \tag{3.4.19}$$

notice that if one wants the cases where $\mathcal{D}_{\{X_i,X,Y\}} = \mathcal{D}_{\{Y_j,X,Y\}} = 0$, the result that one finds is that the only possible case is when $\mathbb{p}_{x_i|y_j} = \mathrm{Tr}(X_iY_j) = |\langle x_i|y_j\rangle|^2 = 1$, this case occurs if and only if $[X,Y] = 0$. However, if one wants the cases where $\mathcal{D}_{\{X_i,X,Y\}} = \mathcal{D}_{\{Y_j,X,Y\}} = +\infty$, then, the only possible case is when $\mathbb{p}_{x_i|y_j} = \mathrm{Tr}(X_iY_j) = |\langle x_i|y_j\rangle|^2 = 0$, this case occurs if and only if $[X,Y] \neq 0$ in its maximum value. In this cases, it is fair to say that when the input state is a particular state, the Bayes' divergence is a quantity only related to the commutation relation between X and Y.

3.4.2 Examples

In this section, we provide some examples which give an intuitive perspective about definition 5, where it was used equation (3.4.10) for simplicity.

Qubit

Let the input system be a qubit in the generic state

$$\rho = \frac{\mathbb{1} + \boldsymbol{r}\cdot\boldsymbol{\sigma}}{2}, \tag{3.4.20}$$

where $|\boldsymbol{r}| \in [0,1]$, with ρ being pure (mixed) for $r = 1$ ($r \leq 1$). Consider the observables $X = \hat{\boldsymbol{x}}\cdot\boldsymbol{\sigma}$ and $Y = \hat{\boldsymbol{y}}\cdot\boldsymbol{\sigma}$, with respective projectors

$$X_i = \frac{\mathbb{1} + i\,\hat{\boldsymbol{x}}\cdot\boldsymbol{\sigma}}{2}, \tag{3.4.21}$$

and

$$Y_j = \frac{\mathbb{1} + j\,\hat{\boldsymbol{y}}\cdot\boldsymbol{\sigma}}{2}, \tag{3.4.22}$$

where $i,j \in \{-1,1\}$.

Calculating the probabilities given by equations (3.4.2)–(3.4.6) using the above quantities, we find

$$\mathbb{p}_{x_i} = \frac{1 + i\,\hat{\boldsymbol{x}}\cdot\boldsymbol{r}}{2}, \tag{3.4.23}$$

and

$$\mathbb{p}_{y_j} = \frac{1 + j\,\hat{\boldsymbol{y}}\cdot\boldsymbol{r}}{2}. \tag{3.4.24}$$

The conditional probabilities are

$$\mathbb{p}_{x_i|y_j} = \mathbb{p}_{y_j|x_i} = \frac{1 + ij\,\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}}}{2}. \tag{3.4.25}$$

The joint probabilities, thus, are

$$\mathbb{p}_{x_i,y_j} = \frac{1 + i\,\hat{\boldsymbol{x}}\cdot\boldsymbol{r} + j\left(i + \hat{\boldsymbol{x}}\cdot\boldsymbol{r}\right)\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}}}{4}, \tag{3.4.26}$$

and

$$\mathbb{p}_{y_j,x_i} = \frac{1 + j\,\hat{\boldsymbol{y}} \cdot \boldsymbol{r} + i\left(j + \hat{\boldsymbol{y}} \cdot \boldsymbol{r}\right)\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}}}{4}. \tag{3.4.27}$$

Replacing equations (3.4.20)–(3.4.27) in equation (3.4.10) and simplifying it, we obtain the Bayes' divergence of a context written in terms of Bloch vectors of each element of the context $\mathbb{C} = \{\rho, X, Y\} \longrightarrow \mathbb{C} = \{\boldsymbol{r}, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}\}$, it follows that

$$\begin{aligned} \mathcal{D}_{\{\boldsymbol{r},\hat{\boldsymbol{x}},\hat{\boldsymbol{y}}\}} &= \frac{1}{2}\Big[\hat{\boldsymbol{y}} \cdot \boldsymbol{r} - (\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}})(\hat{\boldsymbol{x}} \cdot \boldsymbol{r})\Big] \ln\left(\frac{1 + \hat{\boldsymbol{y}} \cdot \boldsymbol{r}}{1 - \hat{\boldsymbol{y}} \cdot \boldsymbol{r}}\right) \\ &+ \frac{1}{2}\Big[\hat{\boldsymbol{x}} \cdot \boldsymbol{r} - (\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}})(\hat{\boldsymbol{y}} \cdot \boldsymbol{r})\Big] \ln\left(\frac{1 + \hat{\boldsymbol{x}} \cdot \boldsymbol{r}}{1 - \hat{\boldsymbol{x}} \cdot \boldsymbol{r}}\right). \end{aligned} \tag{3.4.28}$$

Notice that equation (3.4.28) depends on the input state through modulus $|\boldsymbol{r}|$ and the angles between the vector $\hat{\boldsymbol{x}}$ with $\hat{\boldsymbol{y}}$, $\hat{\boldsymbol{x}}$ with $\boldsymbol{r}$ and $\hat{\boldsymbol{y}}$ with $\boldsymbol{r}$. The inner products reveal an interesting features of our measure: $\mathcal{D}$ actually encompass information about the degree of non-commutativity between the whole context $\{\rho, X, Y\}$. For instance, $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}}$ is connected to the degree of non-commutativity of X with respect to Y by

$$\mathcal{N}_{\{X,Y\}} = \frac{1}{4}\left\|[X_i, Y_j]\right\| = |\hat{\boldsymbol{x}} \times \hat{\boldsymbol{y}}| = \sqrt{1 - (\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}})^2}.$$

If $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}} = 0$ then $\mathcal{N}_{\{X,Y\}} = 1$, its maximum value, implying that X and Y do not commute. On the other hand if $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}} = 1$ then $\mathcal{N}_{\{X,Y\}} = 0$, its minimum value, implying that X and Y commute.

The same is valid for

$$\mathcal{N}_{\{\rho,X\}} = \frac{1}{4}\left\|[\rho, X_i]\right\| = |\hat{\boldsymbol{x}} \times \boldsymbol{r}| = \sqrt{1 - (\hat{\boldsymbol{x}} \cdot \boldsymbol{r})^2},$$

and also for

$$\mathcal{N}_{\{\rho,Y\}} = \frac{1}{4}\left\|[\rho, Y_j]\right\| = |\hat{\boldsymbol{y}} \times \boldsymbol{r}| = \sqrt{1 - (\hat{\boldsymbol{y}} \cdot \boldsymbol{r})^2}.$$

It is possible write equation (3.4.28) instead of inner products but in terms of the degrees non-commutativity $\mathcal{N}_{\{X,Y\}}$, $\mathcal{N}_{\{\rho,X\}}$ and $\mathcal{N}_{\{\rho,Y\}}$ as

$$\begin{aligned} \mathcal{D}_{\{\rho,X,Y\}} &= \frac{1}{2}\left[\sqrt{1 - \mathcal{N}^2_{\{\rho,Y\}}} - \sqrt{\left(1 - \mathcal{N}^2_{\{X,Y\}}\right)\left(1 - \mathcal{N}^2_{\{\rho,X\}}\right)}\right] \ln\left(\frac{1 + \sqrt{1 - \mathcal{N}^2_{\{\rho,Y\}}}}{1 - \sqrt{1 - \mathcal{N}^2_{\{\rho,Y\}}}}\right) \\ &+ \frac{1}{2}\left[\sqrt{1 - \mathcal{N}^2_{\{\rho,X\}}} - \sqrt{\left(1 - \mathcal{N}^2_{\{X,Y\}}\right)\left(1 - \mathcal{N}^2_{\{\rho,Y\}}\right)}\right] \ln\left(\frac{1 + \sqrt{1 - \mathcal{N}^2_{\{\rho,X\}}}}{1 - \sqrt{1 - \mathcal{N}^2_{\{\rho,X\}}}}\right). \end{aligned} \tag{3.4.29}$$

In the particular $\mathcal{N}_{\{X,Y\}} = 1$ and $\hat{\boldsymbol{x}} = \hat{\boldsymbol{y}}$, $\mathcal{N}_{\{\rho,X\}} = \mathcal{N}_{\{\rho,Y\}} = 0$, we find

$$\mathcal{D}_{\{X,Y\}} = \lim_{\mathcal{N}_{\{X,Y\}} \to 1}\left[\sqrt{1 - \mathcal{N}^2_{\{X,Y\}}} - \sqrt{1 - \mathcal{N}^2_{\{X,Y\}}}\right] \ln\left(\frac{1 + \sqrt{1 - \mathcal{N}^2_{\{X,Y\}}}}{1 - \sqrt{1 - \mathcal{N}^2_{\{X,Y\}}}}\right) = 0, \tag{3.4.30}$$

which confirms that when X and Y commute, they are compatible.

Now if X and Y are maximally non-commuting, $\mathcal{N}_{\{X,Y\}} = 1$, in vector representation we get the inner product $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}} = 0$. Suppose for instance $\boldsymbol{r} = \hat{\boldsymbol{x}}$, so that, the input state is pure and eigenstate of X. Consequently, $\hat{\boldsymbol{x}} \cdot \boldsymbol{r} = 1$ and $\hat{\boldsymbol{y}} \cdot \boldsymbol{r} = 0$, and Bayes' divergence (3.4.28) is

$$\mathcal{D}_{\{r,\hat{x},\hat{y}\}} = \lim_{\hat{\boldsymbol{x}}\cdot\boldsymbol{r}\to 1} \frac{1}{2} \hat{\boldsymbol{x}} \cdot \boldsymbol{r} \ln \left(\frac{1 + \hat{\boldsymbol{x}} \cdot \boldsymbol{r}}{1 - \hat{\boldsymbol{x}} \cdot \boldsymbol{r}} \right) = +\infty. \tag{3.4.31}$$

therefore, when X and Y are maximally non-commuting $\mathcal{D} = +\infty$. For this reason, it is convenient introduce a normalization for $\mathcal{D}$.

Discrete Position and Momentum

Remember that in discrete position and momentum representation we can write probabilities as

$$\mathbb{p}_{q_k} = \delta q \, |\psi(q_k)|^2 , \tag{3.4.32}$$

and

$$\mathbb{p}_{p_l} = \delta p \, |\psi(p_l)|^2 . \tag{3.4.33}$$

Conditional probability distributions reduce to

$$\mathbb{p}_{q_k|p_l} = \mathbb{p}_{p_l|q_k} = \frac{\delta q \delta p}{2\pi\hbar} = \frac{1}{\xi}. \tag{3.4.34}$$

which is the result expected for MUB.

Joint probability distributions read

$$\mathbb{p}_{q_k,p_l} = \frac{\mathbb{p}_{q_k}}{\xi} = \frac{\delta q}{\xi} \, |\psi(q_k)|^2 \tag{3.4.35}$$

and

$$\mathbb{p}_{p_l,q_k} = \frac{\mathbb{p}_{p_l}}{\xi} = \frac{\delta p}{\xi} \, |\psi(p_l)|^2 . \tag{3.4.36}$$

Replacing equations (3.4.32)–(3.4.36) in equation (3.4.10) we get

$$\mathcal{D}_{\{\rho,Q,P\}} = -\frac{1}{2} \left\{ \sum_{k=1}^{\xi} \left[\sum_{l=1}^{\xi} \frac{\mathbb{p}_{p_l}}{\xi} - \mathbb{p}_{q_k} \right] \ln \mathbb{p}_{q_k} + \sum_{l=1}^{\xi} \left[\sum_{k=1}^{\xi} \frac{\mathbb{p}_{q_k}}{\xi} - \mathbb{p}_{p_l} \right] \ln \mathbb{p}_{p_l} \right\}. \tag{3.4.37}$$

Since, $\sum_k \mathbb{p}_{q_k} = 1$ and $\sum_l \mathbb{p}_{p_l} = 1$, we obtain

$$\mathcal{D}_{\{\rho,Q,P\}} = -\frac{1}{2} \left\{ \sum_{k=1}^{\xi} \left[\frac{1}{\xi} - \mathbb{p}_{q_k} \right] \ln \mathbb{p}_{q_k} + \sum_{l=1}^{\xi} \left[\frac{1}{\xi} - \mathbb{p}_{p_l} \right] \ln \mathbb{p}_{p_l} \right\}. \tag{3.4.38}$$

Since, $\mathbb{p}_{q_k} = |\langle q_k|\psi\rangle|^2 = 1/\xi$ and $\mathbb{p}_{p_l} = |\langle p_l|\psi\rangle|^2 = 1/\xi$ can never happen simultaneously, then $\{\rho, Q, P\}$ typically form an incompatible scenario, which is expected since Q and P are maximally non-commuting.

Consider the case where ρ is eigenstate of position, $\rho = Q_i$. Then $\mathbb{p}_{q_k} = 1$ and $\mathbb{p}_{p_l} = 0$, since now $\{\rho, P\}$ forms MUB. We arrive at

$$\mathcal{D}_{\{Q,P\}} = -\frac{1}{2} \left\{ \sum_{k=1}^{\xi} \left[\frac{1}{\xi} - 0 \right] \ln(0) \right\} = +\infty, \tag{3.4.39}$$

therefore, measuring Q and P lead to maximal incompatibility, which is an expected result.

General Bloch Representation

Using Bloch representation, we can write the quantum state as

$$\rho = \frac{1}{d}\left(\mathbb{1} + C_d \boldsymbol{r}\cdot\boldsymbol{\Lambda}\right) \tag{3.4.40}$$

and the projectors as

$$X_i = \frac{1}{d}\left(\mathbb{1} + C_d \hat{\boldsymbol{x}}_i\cdot\boldsymbol{\Lambda}\right) \tag{3.4.41}$$

and

$$Y_j = \frac{1}{d}\left(\mathbb{1} + C_d \hat{\boldsymbol{y}}_j\cdot\boldsymbol{\Lambda}\right), \tag{3.4.42}$$

where $C_d = \sqrt{\frac{d(d-1)}{d}}$.

Calculating the probabilities given by equations (3.4.2)–(3.4.6) with the above quantities in Bloch representation, we have the marginal probabilities

$$\mathbb{p}_{x_i} = \frac{1}{d}\left[1 + (d-1)\hat{\boldsymbol{x}}_i\cdot\boldsymbol{r}\right] \tag{3.4.43}$$

and

$$\mathbb{p}_{y_j} = \frac{1}{d}\left[1 + (d-1)\hat{\boldsymbol{y}}_j\cdot\boldsymbol{r}\right]. \tag{3.4.44}$$

The conditional probabilities read

$$\mathbb{p}_{x_i|y_j} = \mathbb{p}_{y_j|x_i} = \frac{1}{d}\left[1 + (d-1)\hat{\boldsymbol{x}}_i\cdot\hat{\boldsymbol{y}}_j\right]. \tag{3.4.45}$$

The joint probabilities, thus, are

$$\mathbb{p}_{x_i,y_j} = \frac{1}{d^2}\Big\{1 + (d-1)\left[\hat{\boldsymbol{x}}_i\cdot\hat{\boldsymbol{y}}_j + \hat{\boldsymbol{x}}_i\cdot\boldsymbol{r} + (d-1)(\hat{\boldsymbol{x}}_i\cdot\hat{\boldsymbol{y}}_j)(\hat{\boldsymbol{x}}_i\cdot\boldsymbol{r})\right]\Big\} \tag{3.4.46}$$

and

$$\mathbb{p}_{y_j,x_i} = \frac{1}{d^2}\Big\{1 + (d-1)\left[\hat{\boldsymbol{x}}_i\cdot\hat{\boldsymbol{y}}_j + \hat{\boldsymbol{y}}_j\cdot\boldsymbol{r} + (d-1)(\hat{\boldsymbol{x}}_i\cdot\hat{\boldsymbol{y}}_j)(\hat{\boldsymbol{y}}_j\cdot\boldsymbol{r})\right]\Big\}. \tag{3.4.47}$$

Replacing equations (3.4.43)–(3.4.47) in equation (3.4.10), we find

$$\mathcal{D}_{\{\boldsymbol{r},\hat{\boldsymbol{x}},\hat{\boldsymbol{y}}\}} = -\frac{1}{2}\frac{d-1}{d}\left\{\sum_{i=1}^{d}\left[v_i - \hat{\boldsymbol{x}}_i\cdot\boldsymbol{r}\right]\ln\frac{1+(d-1)\hat{\boldsymbol{x}}_i\cdot\boldsymbol{r}}{d} + \sum_{j=1}^{d}\left[u_j - \hat{\boldsymbol{y}}_j\cdot\boldsymbol{r}\right]\ln\frac{1+(d-1)\hat{\boldsymbol{y}}_j\cdot\boldsymbol{r}}{d}\right\}, \tag{3.4.48}$$

where

$$u_j := \frac{d-1}{d}\sum_{i=1}^{d}(\hat{\boldsymbol{x}}_i\cdot\hat{\boldsymbol{y}}_j)(\hat{\boldsymbol{x}}_i\cdot\boldsymbol{r}), \tag{3.4.49}$$

and

$$v_i := \frac{d-1}{d}\sum_{j=1}^{d}(\hat{\boldsymbol{x}}_i\cdot\hat{\boldsymbol{y}}_j)(\hat{\boldsymbol{y}}_j\cdot\boldsymbol{r}). \tag{3.4.50}$$

Notice that equation (2.4.7) is also a condition to be satisfied,

$$\hat{\boldsymbol{x}}_i\cdot\hat{\boldsymbol{y}}_j = \frac{d\delta_{ij}-1}{d-1}. \tag{3.4.51}$$

3.5 Bayes' Incompatibility

Having introduced Bayes' divergence, which can be viewed as a measure of Bayes' rule violation for two joint probability distributions of two sequential measurements, we can now introduce the Bayes' incompatibility which is a measure of incompatibility of the context $\mathbb{C} \equiv \{\rho, X, Y\}$ which is a set of a state ρ and two observables X and Y, if their joint probability distributions violate Bayes' rule through Bayes' divergence.

3.5.1 Definition

Considering the same kind of systems that is described by Bayes' divergence, we now introduce, the most important definition of this section.

Definition 6. (Bayes' incompatibility) *The 'Bayes' incompatibility' of the context* $\mathbb{C} \equiv \{\rho, X, Y\} \subset \mathscr{B}(\mathcal{H})$, *with joint probability distributions* $\mathbb{p}_{x,y}$ *and* $\mathbb{p}_{y,x}$ *is defined as*

$$\mathcal{B}_{\mathbb{C}} = 1 - e^{-\mathcal{D}_{\mathbb{C}}}, \tag{3.5.1}$$

where $\mathcal{D}_{\mathbb{C}}$ *is the Bayes' divergence of the two joint probability distributions.*

3.5.2 Example

Qubit

We can substitute equation (3.4.28) in (3.5.1), then obtaining a result for incompatibility of observables. Thus, we get

$$\mathcal{B}_{\{\boldsymbol{r},\hat{\boldsymbol{x}},\hat{\boldsymbol{y}}\}} = 1 - \left(\frac{1+\hat{\boldsymbol{y}}\cdot\boldsymbol{r}}{1-\hat{\boldsymbol{y}}\cdot\boldsymbol{r}}\right)^{-\frac{1}{2}\left[\hat{\boldsymbol{y}}\cdot\boldsymbol{r}-(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}})(\hat{\boldsymbol{x}}\cdot\boldsymbol{r})\right]}\left(\frac{1+\hat{\boldsymbol{x}}\cdot\boldsymbol{r}}{1-\hat{\boldsymbol{x}}\cdot\boldsymbol{r}}\right)^{-\frac{1}{2}\left[\hat{\boldsymbol{x}}\cdot\boldsymbol{r}-(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}})(\hat{\boldsymbol{y}}\cdot\boldsymbol{r})\right]}. \tag{3.5.2}$$

In what follows, we explore some limits.

i) $\underline{\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}} \to 1}$: In this case, the observable orientations are parallel to each other. This leads to commutativity, hence $\mathcal{N}_{\{x,y\}} = 0$ which is the same as $[X, Y] = 0$. However there are two interesting possibilities regarding the input state ρ, which are:

a) $\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{r}} \to 1$ and $\hat{\boldsymbol{y}}\cdot\hat{\boldsymbol{r}} \to 1$: In this case, the input state is oriented in the same direction as the observables. Another way to write this limit is by $\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{r}} = \hat{\boldsymbol{y}}\cdot\hat{\boldsymbol{r}} \equiv a$, so the Bayes' incompatibility is

$$\mathcal{B}_{\{\boldsymbol{r},\hat{\boldsymbol{x}},\hat{\boldsymbol{y}}\}} = 1 - \left(\frac{1+ra}{1-ra}\right)^{-r[a-a]} = 1 - 1 = 0. \tag{3.5.3}$$

Therefore, this is an instance of full compatibility. This situation is schematically illustrated by figure 3.3.

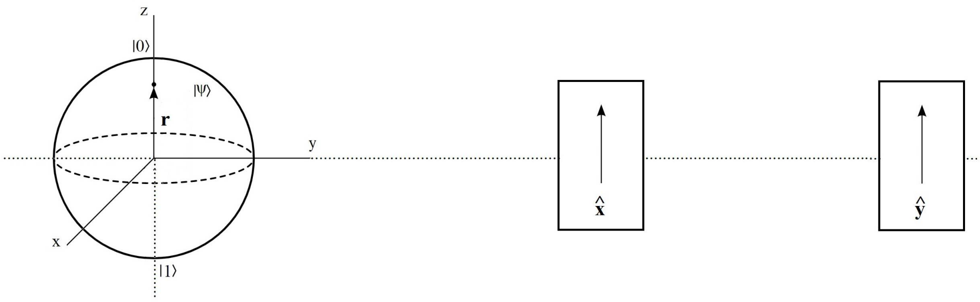


Figure 3.3: The input state is in z-axis direction and also the observables X and Y, therefore all them commute with each other.

b) $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{r}} \to 0$ and $\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{r}} \to 0$: In this case, $\hat{\boldsymbol{x}} = \hat{\boldsymbol{y}} = \hat{\boldsymbol{z}}$ and the input state is not oriented in the same direction as the observables; it is orthogonal to X and Y. So Bayes' incompatibility is

$$\mathscr{B}_{\{\boldsymbol{r},\hat{\boldsymbol{x}},\hat{\boldsymbol{y}}\}} = 1 - 1 = 0, \tag{3.5.4}$$

hence this is also a situation of compatibility. See the illustration in figure 3.4.

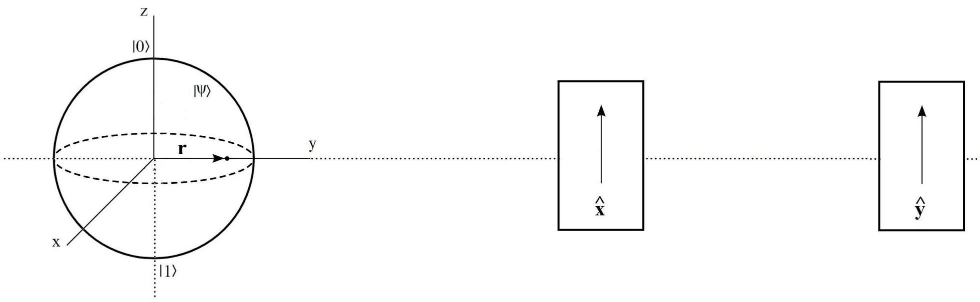


Figure 3.4: The input state is y-axis direction and the observables X and Y in z-axis direction, therefore all commute with each other.

ii) $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}} \to 0$: In this case, the observables orientation is perpendicular to each other. This leads to non-commutativity, hence $\mathscr{N}_{\{X,Y\}} = 1$. However there are two possibilities regarding the input state ρ, namely:

a) $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{r}} \to 1$, $\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{r}} \to 0$ and $r \to 1$: In this case, the input state is a pure state and is oriented in the same direction as the observable X, *i.e.*, an eigenstate of X, but perpendicular to Y. So the Bayes' incompatibility is

$$\mathscr{B}_{\{\boldsymbol{r},\hat{\boldsymbol{x}},\hat{\boldsymbol{y}}\}} = 1 - \left(\frac{1 + \hat{\boldsymbol{x}} \cdot \boldsymbol{r}}{1 - \hat{\boldsymbol{x}} \cdot \boldsymbol{r}}\right)^{-\frac{1}{2}\hat{\boldsymbol{x}}\cdot\boldsymbol{r}}\Bigg|_{\hat{\boldsymbol{x}}\cdot\boldsymbol{r}\to 1} = 1 - (1 - 1) = 1, \tag{3.5.5}$$

rendering a scenario of maximal incompatibility. This situation is depicted in figure 3.5.

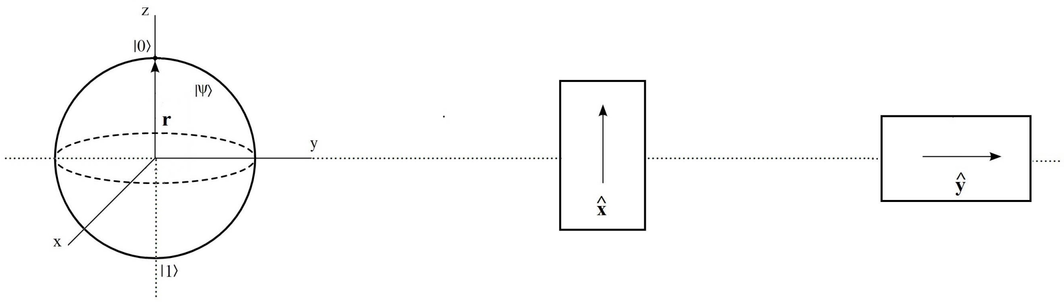


Figure 3.5: The input state and the observables X are in z-axis direction, but Y is in y-axis direction, therefore the input state commute with the observable X but not with observable Y.

b) $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{r}} \to 0$, $\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{r}} \to 1$ and $r \to 1$: In this case, the input state is a pure state and is not oriented in the same direction as the observables, it is orthogonal to X and Y. So the Bayes' incompatibility is

$$\mathcal{B}_{\{\boldsymbol{r},\hat{\boldsymbol{x}},\hat{\boldsymbol{y}}\}} = 1 - \left(\frac{1+\hat{\boldsymbol{y}}\cdot\boldsymbol{r}}{1-\hat{\boldsymbol{y}}\cdot\boldsymbol{r}}\right)^{-\frac{1}{2}\hat{\boldsymbol{y}}\cdot\boldsymbol{r}}\Bigg|_{\hat{\boldsymbol{y}}\cdot\boldsymbol{r}\to 1} = 1 - (1-1) = 1, \tag{3.5.6}$$

rendering a scenario of maximal incompatibility. This situation is depicted in figure 3.6.

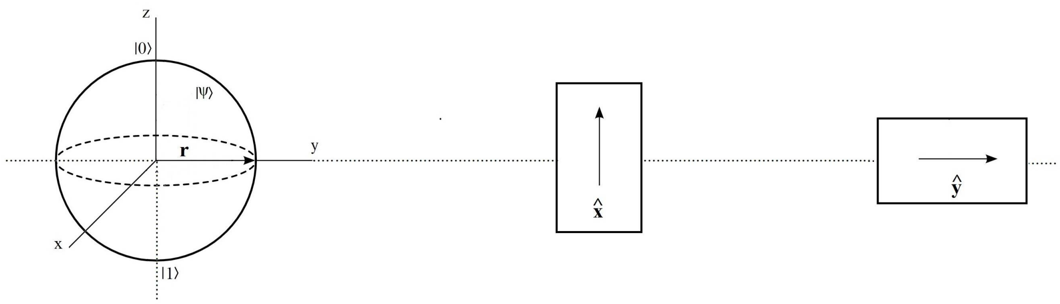


Figure 3.6: The input state and the observables Y are in y-axis direction, but X is in z-axis direction, therefore the input state commute with the observable Y but not with observable X.

3.6 Comparison with other Quantifiers

The authors in reference [Shukla (2019)] introduces a measure also using KL divergence (or relative entropy), which they call *relative incompatibility*. The quantifier is denoted $Q_{X,Y}(\rho)$, and is defined as

$$Q_{X,Y}(\rho) = D(\mathbb{p}_X || \mathbb{p}'_X). \tag{3.6.1}$$

The key idea of this work is to link incompatibility with differences between opposite measurement orderings. Although the authors do not make any explicit statement concerning the dependence of their measure on the quantum state, it is clear that likewise ours, the resulting measure refers to a context. Their definition uses the pairs of marginal probabilities $\{\mathbb{p}_X, \mathbb{p}'_X\}$ and $\{\mathbb{p}_Y, \mathbb{p}'_Y\}$ obtained from the joint probabilities $\mathbb{p}_{x_i,y_j}$ and $\mathbb{p}_{x_i,y_j}$, respectively. The probability $\mathbb{p}_X$

is obtained when X is measured first and $\mathbb{p}'_X$ when X is measured after Y. The latter is the marginal probability distribution defined as

$$\mathbb{p}'_{x_i} = \sum_j \mathbb{p}_{x_i,y_j}. \tag{3.6.2}$$

The same notation can be used for the reverse order, but now using the variable Y, that is, the probability $\mathbb{p}_Y$ is obtained when Y is measured first and $\mathbb{p}'_Y$ when Y is measured after X, the latter is the marginal probability distribution defined as

$$\mathbb{p}'_{y_j} = \sum_i \mathbb{p}_{y_j,x_i}. \tag{3.6.3}$$

It is possible to symmetrize the relative entropy by doing,

$$Q^{(s)}_{X,Y}(\rho) = \frac{D(\mathbb{p}_X||\mathbb{p}'_X) + D(\mathbb{p}_Y||\mathbb{p}'_Y)}{2}. \tag{3.6.4}$$

For a generic qubit state, we have

$$D(\mathbb{p}_X||\mathbb{p}'_X) = \frac{1}{2}(1+\hat{\boldsymbol{x}}\cdot\boldsymbol{r})\log\left[\frac{1+\hat{\boldsymbol{x}}\cdot\boldsymbol{r}}{1+(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}})(\hat{\boldsymbol{y}}\cdot\boldsymbol{r})}\right] + \frac{1}{2}(1-\hat{\boldsymbol{x}}\cdot\boldsymbol{r})\log\left[\frac{1-\hat{\boldsymbol{x}}\cdot\boldsymbol{r}}{1-(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}})(\hat{\boldsymbol{y}}\cdot\boldsymbol{r})}\right] \tag{3.6.5}$$

and

$$D(\mathbb{p}_Y||\mathbb{p}'_Y) = \frac{1}{2}(1+\hat{\boldsymbol{y}}\cdot\boldsymbol{r})\log\left[\frac{1+\hat{\boldsymbol{y}}\cdot\boldsymbol{r}}{1+(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}})(\hat{\boldsymbol{x}}\cdot\boldsymbol{r})}\right] + \frac{1}{2}(1-\hat{\boldsymbol{y}}\cdot\boldsymbol{r})\log\left[\frac{1-\hat{\boldsymbol{y}}\cdot\boldsymbol{r}}{1-(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}})(\hat{\boldsymbol{x}}\cdot\boldsymbol{r})}\right]. \tag{3.6.6}$$

The difference from relative incompatibility and Bayes' incompatibility is that the latter has in its core two direct fundamental principles, one is Bayes' rule from probability theory, and the other is incompatibility from quantum theory. In relative incompatibility, Bayes' rule is not a fundamental principle, in the sense that relative incompatibility is born only from quantum principles regarding measurements.

In figure 3.7, a comparison is shown between these measures for a "conical" context defined by $\hat{\boldsymbol{x}}\cdot\boldsymbol{r} = \hat{\boldsymbol{y}}\cdot\boldsymbol{r}$. In this case, the vectors $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ lie on the inclined surface of a cone whose symmetry axis is given by $\boldsymbol{r}$. By virtue of such angular equivalence of $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ relative to $\boldsymbol{r}$, we expect the measures to provide incompatibility information solely about the couple $\{X,Y\}$. To corroborate this reasoning, we bring the non-commutativity $\mathscr{N}_{\{X,Y\}}$ to the discussion. The first panel gives a parametric plot between $\mathscr{B}_{\mathbb{C}}$ and $\mathscr{N}_{\{X,Y\}}$ as a function of $\hat{\boldsymbol{x}}\cdot\boldsymbol{y}$ for several values of $\hat{\boldsymbol{x}}\cdot\boldsymbol{r}$ (which is equal to $\hat{\boldsymbol{y}}\cdot\boldsymbol{r}$). The results show that the measures $\mathscr{B}_{\mathbb{C}}$ and $\mathscr{N}_{\{X,Y\}}$ turn out to be monotonic functions of each other, that is, they are equivalent in essence. The second and third panels provide similar information, but between different quantifiers. In particular, the third panel shows that, for generic qubit states, our measure $\mathscr{B}_{\mathbb{C}}$ is equivalent to the symmetric relative incompatibility $Q^{(s)}_{X,Y}$. We have no study so far proving such equivalence for higher dimensions.

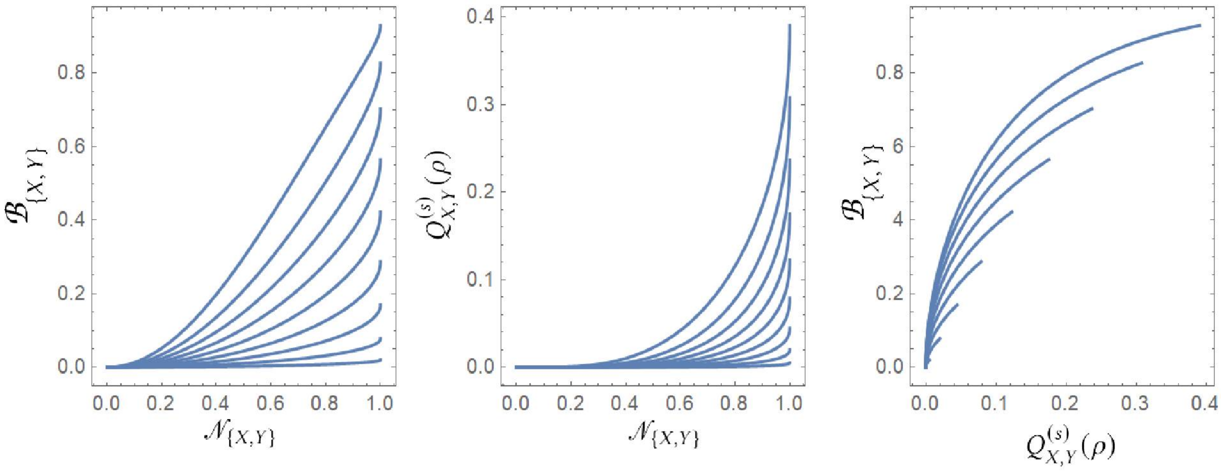


Figure 3.7: In the graphs, each blue curve refers to a specific value of $\hat{\boldsymbol{x}} \cdot \boldsymbol{r}$ (which is equal to $\hat{\boldsymbol{y}} \cdot \boldsymbol{r}$) in the set $\{0.1, ..., 0.9\}$ with increments of 0.1, so that the lower curve corresponds to 0.1 and the upper one to 0.9. Each graph furnishes parametric plots between the indicates measures [incompatibility via Bayes' rule violation ($\mathcal{B}_{\mathbb{C}}$)); symmetric relative incompatibility ($Q^{(s)}_{X,Y}$); non-commutativity ($\mathcal{N}_{\{X,Y\}}$)] as a function of $\hat{\boldsymbol{x}} \cdot \boldsymbol{y} \in [0, 1]$. The results reveal monotonic relations between the measures, thus indicating that they are all equivalent in essence. The third panel shows that our measure of incompatibility for the couple $\{X, Y\}$ is equivalent to the symmetric relative incompatibility.

CHAPTER 4

CONTEXT INCOMPATIBILITY VIA INFORMATION LEAKAGE

In the previous chapter, an incompatibility quantifier was introduced based on the Bayes' rule violation of two joint probability distributions referring to two distinct ordering for the measurement. In this section, a route is taken that has remained unexplored so far. By considering a scenario designed to test the safety of a communication channel, we link quantum incompatibility with information. To start with, we have to provide the reader with some supplementary tools.

4.1 Quantum Operations

The state evolution of closed systems is described by

$$\rho' = U\rho U^{\dagger}, \tag{4.1.1}$$

where U are unitary operators, that is

$$UU^{\dagger} = U^{\dagger}U = \mathbb{1}. \tag{4.1.2}$$

This operation being unitary preserves not only normalization of the density operator, but also Hermiticity, the trace, the positivity and it leads to the identity of itself. This is not the only operation that maps a density operator in another one. Generally, operation with such characteristics are called *quantum operations* or *quantum channels* and are formally described in terms of the *completely positive and trace preserving* (CPTP) map

$$\rho' \equiv \Phi_K(\rho) := \sum_n K_n \rho K_n^{\dagger}, \tag{4.1.3}$$

where $\{K_n\}$ is an arbitrary set of operators, called *Kraus operators*, satisfying the condition

$$\sum_n K_n^{\dagger} K_n = \mathbb{1}. \tag{4.1.4}$$

The evolution operator is a particular case of Kraus operators when there is only one operator in $\{K_n\}$.

In the particular case where K_n is a projector on the basis $\{|k_n\rangle\}$, it follows that

$$\begin{aligned}
\Phi_K(\rho) &= \sum_n K_n \rho K_n \\
&= \sum_n |k_n\rangle \langle k_n| \rho |k_n\rangle \langle k_n| \\
&= \sum_n \langle k_n|\rho|k_n\rangle |k_n\rangle \langle k_n| \\
&= \sum_n \mathbb{p}_n K_n,
\end{aligned} \tag{4.1.5}$$

where $\mathbb{p}_n = \mathrm{Tr}(K_n\rho) = \langle k_n|\rho|k_n\rangle$, in this case $\Phi_K(\rho)$ is called *non-selective von Neumann measurements.* A further operation in the basis $\{|j_m\rangle\}$ gives

$$\begin{aligned}
\Phi_{JK}(\rho) \equiv \Phi_J\left(\Phi_K(\rho)\right) &= \sum_m J_m \Phi_K(\rho) J_m \\
&= \sum_{m,n} \mathbb{p}_n J_m K_n J_m \\
&= \sum_{m,n} \mathbb{p}_n |j_m\rangle \langle j_m| |k_n\rangle \langle k_n| |j_m\rangle \langle j_m| \\
&= \sum_{m,n} \mathbb{p}_n |\langle j_m|k_n\rangle|^2 |j_m\rangle \langle j_m| \\
&= \sum_{m,n} \mathbb{p}_n \mathbb{p}_{m|n} J_m \\
&= \sum_{m,n} \mathbb{p}_{m,n} J_m,
\end{aligned} \tag{4.1.6}$$

where $\mathbb{p}_{m|n} = \mathrm{Tr}(J_m K_n) = |\langle j_m|k_n\rangle|^2$. Although we have used the notation $\mathbb{p}_{m,n} = \mathbb{p}_n \mathbb{p}_{m|n}$, the reader should recall from previous discussion that Bayes' rule does not apply in the quantum context, in general.

A quantum operation maps density operators to density operators with the following properties:

i) **Linearity**: $$\Phi_K(\alpha\rho_1 + \beta\rho_2) = \alpha\Phi_K(\rho_1) + \beta\Phi_K(\rho_2). \tag{4.1.7}$$

ii) **Hermiticity**: $$\rho = \rho^\dagger \quad \text{implies} \quad \Phi_K(\rho) = \Phi_K^\dagger(\rho). \tag{4.1.8}$$

iii) **Positivity**: $$\rho \geq 0 \quad \text{implies} \quad \Phi_K(\rho) \geq 0. \tag{4.1.9}$$

iv) **Preserves trace**: $$\mathrm{Tr}\left(\Phi_K(\rho)\right) = \mathrm{Tr}(\rho). \tag{4.1.10}$$

v) **Entropy increasing**: $$S\left(\Phi_K(\rho)\right) \geq S(\rho). \tag{4.1.11}$$

Although our analysis is restricted to single-partite systems, these properties also apply for multipartite states with K acting on one of the parts.

Trace Norm of CPTP Maps

Given the Hilbert-Schmidt norm of any density operator ρ

$$\|\rho\| = \sqrt{\operatorname{Tr}(\rho^2)}, \tag{4.1.12}$$

We have

$$\begin{aligned}
\|\Phi_{YX}(\rho) - \Phi_X(\rho)\|^2 &= \operatorname{Tr}\Big(\Phi_{YX}(\rho) - \Phi_X(\rho)\Big)^2. \\
&= \operatorname{Tr}\Phi_{YX}^2(\rho) + \operatorname{Tr}\Phi_X^2(\rho) - 2\operatorname{Tr}\Big(\Phi_{YX}(\rho)\Phi_X(\rho)\Big) \\
&= \|\Phi_{YX}(\rho)\|^2 + \|\Phi_X(\rho)\|^2 - 2\operatorname{Tr}\Big(\Phi_{YX}(\rho)\Phi_X(\rho)\Big).
\end{aligned} \tag{4.1.13}$$

To calculate the last term, let us consider non-selective measurements, then it is possible to write

$$\begin{aligned}
\operatorname{Tr}\left(\Phi_X^2(\rho)\right) &= \operatorname{Tr}\left(\sum_i \mathbb{p}_{x_i} X_i\right)^2 \\
&= \operatorname{Tr}\left(\sum_i \mathbb{p}_{x_i}^2 X_i\right) \\
&= \sum_i \mathbb{p}_{x_i}^2,
\end{aligned} \tag{4.1.14}$$

and

$$\begin{aligned}
\operatorname{Tr}\left(\Phi_{YX}^2(\rho)\right) &= \operatorname{Tr}\left(\sum_j \mathbb{w}_j Y_j\right)^2 \\
&= \operatorname{Tr}\left(\sum_i \mathbb{w}_j^2 Y_j\right) \\
&= \sum_j \mathbb{w}_j^2,
\end{aligned} \tag{4.1.15}$$

where $\mathbb{w}_j \equiv \sum_i \mathbb{p}_{x_i}\mathbb{p}_{y_j|x_i}$. Calculating also,

$$\begin{aligned}
\operatorname{Tr}\Big(\Phi_{YX}(\rho)\Phi_X(\rho)\Big) &= \operatorname{Tr}\left(\sum_i \mathbb{p}_{x_i} X_i \sum_j \mathbb{w}_j Y_j\right) \\
&= \sum_j \mathbb{w}_j^2,
\end{aligned} \tag{4.1.16}$$

hence

$$\operatorname{Tr}\Big(\Phi_{YX}(\rho)\Phi_X(\rho)\Big) = \|\Phi_{YX}(\rho)\|^2. \tag{4.1.17}$$

Returning to (4.1.13), we finally obtain

$$\|\Phi_{YX}(\rho) - \Phi_X(\rho)\|^2 = \|\Phi_X(\rho)\|^2 - \|\Phi_{YX}(\rho)\|^2. \tag{4.1.18}$$

This result will be useful later on.

4.2 Stinespring Theorem

Consider a system whose state is ρ_S coupled to an external system (the "environment") $\mathbb{E}$ prepared in a particular pure state $|\psi\rangle_{\mathbb{E}}$ with any dimension. The total system is associated with $\mathcal{H} \otimes \mathcal{H}_{\mathbb{E}}$ and the time evolution is described by a unitary operation U. Tracing out the entanglement with the external system, we find the reduced state

$$\rho_S(t) = \mathrm{Tr}_{\mathbb{E}}\left(U(t)\rho_{S\mathbb{E}}U^{\dagger}(t)\right), \tag{4.2.1}$$

where $\rho_{S\mathbb{E}} = \rho_S \otimes |\psi\rangle\langle\psi|_{\mathbb{E}}$. To take the trace over the environment, a basis $\{|\mathbb{E}_i\rangle\}$ can be introduce, such that,

$$\rho_S(t) = \sum_i \langle\mathbb{E}_i| U(t) |\psi\rangle_{\mathbb{E}} \rho_S \langle\psi|_{\mathbb{E}} U^{\dagger}(t) |\psi\rangle_{\mathbb{E}_i}. \tag{4.2.2}$$

Introducing $K_i \equiv \langle\mathbb{E}_i| U |\psi\rangle_{\mathbb{E}}$, then last equation turns out to be

$$\rho_S(t) = \sum_i K_i \rho_S K_i^{\dagger} = \Phi_K(\rho_S), \tag{4.2.3}$$

which is a CPTP map, since $\sum_i K_i K_i^{\dagger} = \mathbb{1}$. Here is the key content of the Stinespring theorem: getting the reduced state of the system by tracing out the environment in an entangling unitary dynamics is equivalent to the application of a CPTP map over the system.

4.3 Informational Incompatibility of a Physical Context

We are now ready to introduce the second result of this work (see reference [Martins (2020)] for the preprint submitted for publication). Let $\mathbb{C} \equiv \{\rho, X, Y\} \subset \mathscr{B}(\mathcal{H})$ be a context, a set formed by a state ρ, observables $X = \sum_j x_j X_j$ and $Y = \sum_k y_k Y_k$ all of them acting on $\mathcal{H}$.

4.3.1 Protocol

Consider the generic protocol depicted in the figure 4.1. Alice prepares a state ρ, with informational content given by

$$I(\rho) = \ln d - S(\rho), \tag{4.3.1}$$

where $S(\rho) = -\mathrm{Tr}(\rho \ln \rho)$ is the von Neumann entropy of ρ. After measuring X, without registering the outcome of any particular run of the experiment (unrevealed measurement), Alice transforms the initial preparation into

$$\Phi_X(\rho) = \sum_{j=1}^{d} X_j \, \rho \, X_j = \sum_{j=1}^{d} \mathbb{p}_{x_j} X_j, \tag{4.3.2}$$

where $\mathbb{p}_{x_j} = \mathrm{Tr}(\rho X_j)$ and Φ_X is a CPTP map given by (4.1.5), which removes quantum coherence from ρ in the basis $\{|x_j\rangle\}$. After she measures X, the informational resource I is reduced to the value (from equation (4.1.11))

$$I^i \equiv I(\Phi_X(\rho)) = \ln d - H(\mathbb{p}_{x_j}), \tag{4.3.3}$$

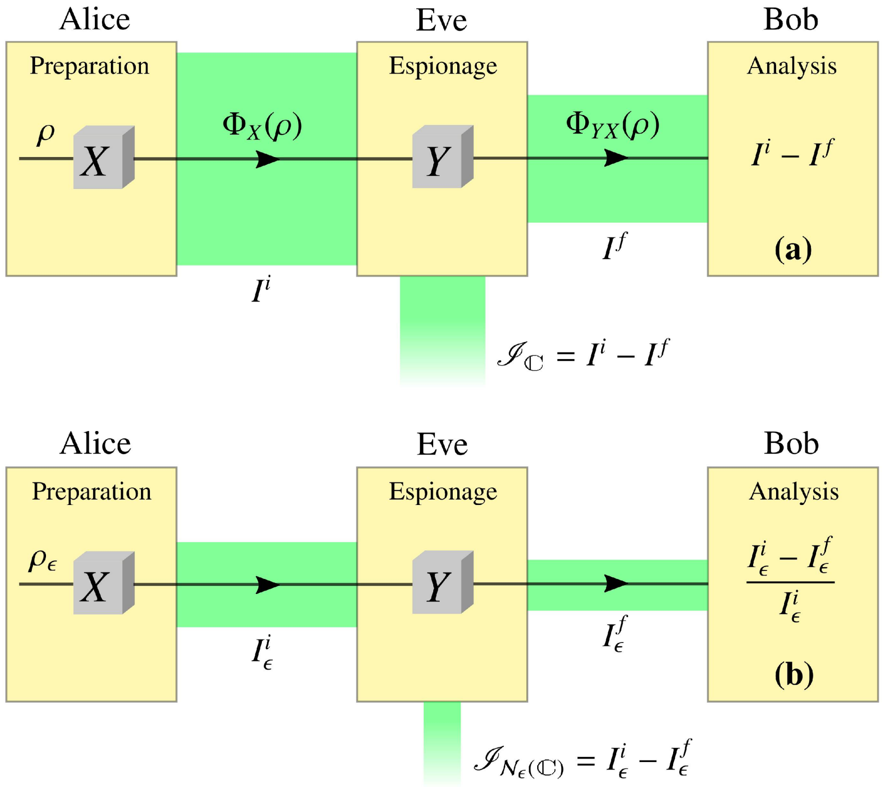


Figure 4.1: (a) For a preparation ρ, Alice measures an observable X and thus sets an amount $I^i = I(\Phi_X(\rho))$ of information (depicted by the first green thick stripe). Aware of the calibration procedure performed by Alice, the trusted partner Bob makes state tomography and then checks the information received. Upon the action of an eavesdropper, Eve, who measures Y, the received information actually is just $I^f = I(\Phi_{YX}(\rho))$. The incompatibility of a context $\mathbb{C} \equiv \{\rho, X, Y\}$ is a resource, quantified as $\mathscr{I}_{\mathbb{C}} \coloneqq I^i - I^f$, that allows for Alice and Bob to detect, via information leakage, Eve' espionage. (b) Aiming at discouraging potential eavesdroppers, Alice now prepares a highly noisy state $\rho_\epsilon = \mathcal{N}_\epsilon(\rho)$ [Equation (4.5.1)] and injects a very limited amount of information I^i_ϵ in the channel. Bob receives an even more restrictive amount I^f_ϵ of information. However, because the information leakage is proportional to the injected information, by looking at the ratio $(I^i_\epsilon - I^f_\epsilon)/I^i_\epsilon$ he still succeeds to detect Eve's intervention.

where $H(\mathbb{p}_{x_j})$ is the Shannon entropy of the probability distribution $\mathbb{p}_{x_j}$. The system is then delivered to Bob, who expects to receive an amount I^i of informational resource, as prearranged with Alice. If he indeed receives this amount of information, then it is ascertained that the channel is safe from information leakage (external espionage).

Consider, however, that there is an eavesdropper in the communication channel, Eve, who intercepts the system sent by Alice and probe it by measuring the observable Y. The procedure is conducted by means of a unitary transformation $U \in \mathscr{B}(\mathcal{H} \otimes \mathcal{H}_\mathcal{E})$ that entangles the system, which left Alice's laboratory in the state $\Phi_X(\rho) \in \mathcal{H}$, with Eve's apparatus $\mathcal{E}$ initially prepared in a state $\rho_0^\mathcal{E} \in \mathscr{B}(H_\mathcal{E})$. The composite state $\Omega_0 = \Phi_X(\rho) \otimes \rho_0^\mathcal{E}$ thus evolves into $\Omega_t = U\Omega_0 U^\dagger$. Using Eq. (4.3.1) we can rewrite the mutual information, defined by

$$\mathcal{I}(\Omega_t) = S(\rho_t) + S(\rho_t^\mathcal{E}) - S(\Omega_t), \tag{4.3.4}$$

in the form

$$I(\Omega_t) = I(\rho_t) + I(\rho_t^\mathcal{E}) + \mathcal{I}(\Omega_t), \tag{4.3.5}$$

where $\rho_t^{\mathcal{E}}$ is the reduced state of the apparatus and $\rho_t = \mathrm{Tr}_{\mathcal{E}}\,\Omega_t = \Phi_{YX}(\rho)$ (via the Stinespring theorem). Since the von Neumann entropy is invariant under unitary transformations, which guarantees that $I(\Omega_0) = I(\Omega_t)$, the last equation can be written as

$$I(\Omega_t) = I(\Omega_0) = I\left(\Phi_{YX}(\rho)\right) + I(\rho_t^{\mathcal{E}}) + \mathcal{I}(\Omega_t), \tag{4.3.6}$$

or

$$I\left(\Phi_X(\rho)\otimes\rho_0^{\mathcal{E}}\right) = I(\Phi_X(\rho)) + I(\rho_0^{\mathcal{E}}) = I\left(\Phi_{YX}(\rho)\right) + I(\rho_t^{\mathcal{E}}) + \mathcal{I}(\Omega_t). \tag{4.3.7}$$

Finally,

$$I(\Phi_X(\rho)) + I(\rho_0^{\mathcal{E}}) = I\left(\Phi_{YX}(\rho)\right) + I(\rho_t^{\mathcal{E}}) + \mathcal{I}(\Omega_t). \tag{4.3.8}$$

Introducing the notation $\Delta I_{\mathcal{E}} \equiv I(\rho_t^{\mathcal{E}}) - I(\rho_0^{\mathcal{E}})$ and $I^f \equiv I(\Phi_{YX}(\rho))$, it is possible to rearrange the above equation as

$$I^i - I^f = I(\Phi_X(\rho)) - I(\Phi_{YX}(\rho)) = \Delta I_{\mathcal{E}} + \mathcal{I}(\Omega_t), \tag{4.3.9}$$

where the map $\Phi_{YX}(\rho)$ is written as in equation (4.1.6), that is,

$$\Phi_{YX}(\rho) = \sum_{j,k} \mathbb{p}_{y_k|x_j}\mathbb{p}_{x_j}\, Y_k, \tag{4.3.10}$$

with $\mathbb{p}_{k|j} = |\langle x_j|y_k\rangle|^2 = \mathrm{Tr}(X_j Y_k)$, so that

$$I^f = \ln d - H\left(\sum_j \mathbb{p}_{y_k|x_j}\mathbb{p}_{x_j}\right). \tag{4.3.11}$$

In equation (4.3.9) we see that the resource consumed from Alice's system, $\mathscr{I}_{\mathbb{C}} \equiv I^i - I^f$, was used to change the local information of Eve's apparatus ($\Delta I_{\mathcal{E}}$) and to increase the correlations between the system and the apparatus. If $\mathscr{I}_{\mathbb{C}} > 0$, Alice and Bob then discover that the channel is being spied upon [Fig. 4.1(a)], since Eve is acquiring information in the channel. Now, using $\sum_k Y_k^2 = \mathbb{1}$ and $Y_k\Phi_Y(\sigma) = \Phi_Y(\sigma)Y_k$, one shows that

$$\mathrm{Tr}[\sigma g(\Phi_Y(\sigma))] = \mathrm{Tr}[\Phi_Y(\sigma) g(\Phi_Y(\sigma))] \tag{4.3.12}$$

for any state σ and function g. This allows us to write

$$S(\sigma||\Phi_Y(\sigma)) = S(\Phi_Y(\sigma)) - S(\sigma), \tag{4.3.13}$$

with

$$S(\sigma||\varrho) = \mathrm{Tr}[\sigma(\ln\sigma - \ln\varrho)] \geq 0 \tag{4.3.14}$$

being the relative entropy (equality holding if and only if $\sigma = \varrho$). We then arrive at the form

$$\mathscr{I}_{\mathbb{C}} = I(\Phi_X(\rho)) - I(\Phi_{YX}(\rho)) = S(\Phi_X(\rho)||\Phi_{YX}(\rho)), \tag{4.3.15}$$

through which we can check that there are only two instances in which $\mathscr{I}_{\mathbb{C}} = 0$, *i.e.*, the context $\mathbb{C}$ is compatible:

(i) $[X,Y] = 0$ $(\forall\rho)$ (the operators share the same set of eigenstates and $\Phi_{YX}(\rho) = \Phi_X(\rho)$);

(ii) $\Phi_X(\rho) = \mathbb{1}/d$ $(\forall X, Y)$ (implying that Bob can not check the channel safety, *i.e.*, $I^i = I^f = 0$).

On the other hand the consumed resource $\mathscr{I}_{\mathbb{C}}$ reaches its maximum value, $\ln d$, when $\rho = X_j$ (an eigenstate of X) and in addition, the X and Y eigenbases form mutually unbiased bases (MUB) [Durt (2010)], that is, $|\langle x_j|y_k\rangle|^2 = 1/d$. Therefore, from Bob's (Eve's) viewpoint, noncommutativity and $I^i > 0$ are necessary ingredients—*resources*—for a successful leakage detection (information acquisition). Thus, with respect to the protocol depicted by Fig. 4.1(a), the following concept is introduced.

Definition 7. (Informational incompatibility) *Informational incompatibility is the resource encoded in a context $\mathbb{C} \equiv \{\rho, X, Y\}$ that allows one to test the safety of a communication channel against information leakage. Quantified via $\mathscr{I}_{\mathbb{C}} = I^i - I^f$* [equation (4.3.15)], *it is operationally related to the amount of information subtracted from the system upon an external measurement.*

Connection with Coherence

It is interesting to note that a connection can be made with quantum coherence—as we have seen in section 2.5.3 a well-established quantum resource quantified by the $\{|y_k\rangle\}$-basis relative entropy of coherence,

$$C_Y(\rho) = S(\rho||\Phi_Y(\rho)) \tag{4.3.16}$$

(see references [Baumgratz (2014)] and [Streltsov (2017)]). It is trivial to verify that

$$\mathscr{I}_{\mathbb{C}} = C_Y(\Phi_X(\rho)), \tag{4.3.17}$$

which shows that the informational incompatibility can be viewed as the amount of Y-coherence that is encoded in an X-incoherent state $\Phi_X(\rho)$. This is the manner whereby the incompatibility of the set $\{X, Y\}$ is captured by $\mathscr{I}_{\mathbb{C}}$.

4.4 Resource Theory of Informational Incompatibility

As briefly discussed in section 2.5.3, a resource theory requires free operations, free states, states with maximum resource and a quantifier. In our framework, all these objects are switched to free context, context with the maximum resource, free operations and a quantifier or resource monotone ($\mathscr{I}_{\mathbb{C}}$). The resourceless (free) contexts, defined as $\mathbb{C}^{\text{free}}$ such that $\mathscr{I}_{\mathbb{C}^{\text{free}}} = 0$, are

$$\mathbb{C}_1^{\text{free}} = \{\rho, X, Y\} \text{ s.t. } [X, Y] = 0 \ (\forall\, \rho), \tag{4.4.1a}$$
$$\mathbb{C}_2^{\text{free}} = \{\tfrac{\mathbb{1}}{d}, X, Y\} \ (\forall\, X, Y), \tag{4.4.1b}$$
$$\mathbb{C}_3^{\text{free}} = \{Y_k, X, Y\} \text{ s.t. } |\langle x_j|y_k\rangle|^2 = 1/d. \tag{4.4.1c}$$

Regarding $\mathbb{C}_2^{\text{free}}$ and $\mathbb{C}_3^{\text{free}}$, which result in $I^i = 0$, these correspond to context for which, Alice does not send any information to Bob. The following proposition formally proves that (4.4.1) are the only existing free context for $I_{\mathbb{C}}$.

Proposition 1. *With respect to contexts $\mathbb{C} = \{\rho, X, Y\}$, $\mathscr{I}_{\mathbb{C}} = 0$ if and only if (i) $[X, Y] = 0$ $(\forall \rho)$ or (ii) $I^i = I(\Phi_X(\rho)) = 0$.*

Proof. From $\mathscr{I}_{\mathbb{C}} = S(\Phi_X(\rho)||\Phi_{YX}(\rho))$ it follows that $\mathscr{I}_{\mathbb{C}} = 0$ if and only if $\Phi_{YX}(\rho) = \Phi_X(\rho)$. By its turn, this condition requires, via the definitions of Φ_X and Φ_{YX}, that

$$\sum_j p_{x_j} \left[\Phi_Y(X_j) - X_j\right] = 0, \tag{4.4.2}$$

with $\mathbb{p}_{x_j} = \mathrm{Tr}(X_j\rho) \geqslant 0$. There are only two ways of satisfying the above equation. First, if $\Phi_Y(X_j) = X_j$ for all ρ. This is equivalent to

$$\sum_k Y_k X_j Y_k = X_j. \tag{4.4.3}$$

Multiplying by Y_l either on the left-hand side or on the right-hand side, we obtain

$$Y_l X_j Y_l = Y_l X_j \tag{4.4.4}$$

or

$$Y_l X_j Y_l = X_j Y_l, \tag{4.4.5}$$

respectively, which imply $[X_j, Y_l] = 0$ and hence,

$$[X, Y] = \sum_{j,k} x_j y_k [X_j, Y_k] = 0. \tag{4.4.6}$$

On the other hand if $[X, Y] = 0$, then

$$[X^2, Y^2] = \sum_{j,k} x_j^2 y_k^2 [X_j, Y_k] = 0, \tag{4.4.7}$$

which demands that

$$Y_k X_j = X_j Y_k. \tag{4.4.8}$$

Multiplying by Y_k on the right-hand side and summing over k, we get

$$\sum_k Y_k X_j Y_k = X_j, \tag{4.4.9}$$

which satisfies equation (4.4.2) and completes the proof of item (i). The second way of satisfying equation (4.4.2) is by picking a uniform distribution, $\mathbb{p}_{x_j} = 1/d$, for in this case we find

$$\frac{1}{d} \sum_j [\Phi_Y(X_j) - X_j] = \frac{1}{d}[\Phi_Y(\mathbb{1}) - \mathbb{1}] = 0. \tag{4.4.10}$$

Since

$$\Phi_X(\rho) = \sum_j \mathbb{p}_{x_j} X_j, \tag{4.4.11}$$

in this case we have

$$\Phi_X(\rho) = \mathbb{1}/d \tag{4.4.12}$$

and

$$I^i = I(\Phi_X(\rho)) = \ln d - S(\Phi_X(\rho)) = 0. \tag{4.4.13}$$

Notice that $\Phi_X(\rho) = \mathbb{1}/d$ if either $\rho = \mathbb{1}/d$ or $\rho = Y_k$ with the X and Y eigenbases forming MUB. Conversely, $I^i = 0$ only if $\Phi_X(\rho) = \mathbb{1}/d$, implying that $\Phi_{YX}(\rho) = \Phi_X(\rho)$ and hence, the validity of equation (4.4.2). This proofs item (ii). □

Apart from them, any other context is either termed resourceful or maximally resourceful. With regard to free operations, which can be thought as CPTP maps, one property of the $S\left(\Phi_X(\rho)||\Phi_{YX}(\rho)\right)$ is that it does not increase under generic CPTP maps Γ, which implies that

$$\mathscr{I}_{\mathbb{C}} \geq \mathscr{I}_{\Gamma(\mathbb{C})}, \tag{4.4.14}$$

where

$$\Gamma(\mathbb{C}) \equiv \{\Gamma(\rho), X, Y\}, \tag{4.4.15}$$

provided that Γ commutes with the maps Φ_X and Φ_{YX}. In this case, it is clear that resource is never created upon the action of Γ. Also, to ensure that $\mathscr{I}_{\Gamma(\mathbb{C}^{\text{free}})} = 0$, we need to require Γ to be unital, so as not to make $\mathbb{C}_1^{\text{free}}$ resourceful upon Γ. Altogether, these aspects characterize the free operations Γ with respect to informational incompatibility. In our approach, we do not admit any operations on $\{X, Y\}$, as this would imply *unsharp observables* responsible for aspects of measurement fuzziness that have been disregarded from the outset.

4.5 Measurement Incompatibility

We now consider a noisy scenario [Fig. 4.1(b)]. To discourage any potential eavesdroppers, Alice introduces, in a controllable way, an amount $1 - \epsilon$ of noise in the input state, which then reads

$$\rho_\epsilon = \mathcal{N}_\epsilon(\rho) := (1-\epsilon)\frac{\mathbb{1}}{d} + \epsilon\,\rho, \tag{4.5.1}$$

where $\epsilon \in [0, 1]$ and $\mathcal{N}_\epsilon$ is a CPTP unital *noise map*. From the concavity of the entropy and the joint convexity of the relative entropy, the informational resource is written as

$$\begin{aligned} I_\epsilon^i = \ln d - S\left(\Phi_X(\rho_\epsilon)\right) &\leq \ln d - (1-\epsilon)\ln d - \epsilon S\left(\Phi_X(\rho)\right) \\ &= \epsilon \ln d - \epsilon S\left(\Phi_X(\rho)\right), \end{aligned} \tag{4.5.2}$$

or,

$$I_\epsilon^i \leq \epsilon\, I^i. \tag{4.5.3}$$

This shows that the input information is limited according to epsilon. Following the same procedure, one show that

$$\mathscr{I}_{\mathcal{N}_\epsilon(\mathbb{C})} \leq \epsilon\, \mathscr{I}_{\mathbb{C}}, \tag{4.5.4}$$

where $\mathcal{N}_\epsilon(\mathbb{C}) = \{\rho_\epsilon, X, Y\}$ and $\mathcal{N}_{\epsilon=0}(\mathbb{C}) = \mathbb{C}$.

Hence, the preparation ρ_ϵ implies, for $\epsilon \ll 1$, a very limited amount of information in the channel and an equally restrictive amount of consumable information. Aware of the amount of noise introduced, Bob can still check the security of the channel by looking at the amount of information that leaks per unit of injected information. To this end, Bob computes the ratio

$$\mathscr{R}_{\mathcal{N}_\epsilon(\mathbb{C})} := \frac{I_\epsilon^i - I_\epsilon^f}{I_\epsilon^i} = \frac{\mathscr{I}_{\mathcal{N}_\epsilon(\mathbb{C})}}{I_\epsilon^i}, \tag{4.5.5}$$

with $I_\epsilon^f \equiv I\left(\Phi_{YX}(\rho_\epsilon)\right),$. This is equivalent to

$$\mathscr{R}_{\mathcal{N}_\epsilon(\mathbb{C})} = \frac{I(\Phi_X(\rho_\epsilon)) - I(\Phi_{YX}(\rho_\epsilon))}{I(\Phi_X(\rho_\epsilon))}. \tag{4.5.6}$$

Denoting

$$
\begin{aligned}
\rho'_\epsilon \equiv \Phi_X(\rho_\epsilon) &= (1-\epsilon)\frac{\mathbb{1}}{d} + \epsilon\,\Phi_X(\rho) \\
&= (1-\epsilon)\frac{\mathbb{1}}{d} + \epsilon \sum_i \mathbb{p}_{x_i} X_i,
\end{aligned}
\tag{4.5.7}
$$

whose entropy reads

$$
\begin{aligned}
S(\rho'_\epsilon) &= -\operatorname{Tr}[\rho'_\epsilon \ln \rho'_\epsilon] \\
&= -\sum_i \langle x_i | \rho'_\epsilon \ln \rho'_\epsilon | x_i \rangle\,.
\end{aligned}
\tag{4.5.8}
$$

Noticing that

$$
\rho'_\epsilon \ln \rho'_\epsilon \,|x_i\rangle = \left(\frac{1-\epsilon}{d} + \epsilon \mathbb{p}_{x_i}\right)|x_i\rangle\,,
\tag{4.5.9}
$$

we find

$$
S(\rho'_\epsilon) = -\sum_i \left(\frac{1-\epsilon}{d}\right)\left[1 + \frac{\epsilon \mathbb{p}_{x_i} d}{1-\epsilon}\right]\left\{\ln\left(\frac{1-\epsilon}{d}\right) + \ln\left(1 + \frac{\epsilon \mathbb{p}_{x_i} d}{1-\epsilon}\right)\right\}.
\tag{4.5.10}
$$

In the large-noise limit $\epsilon d \ll 1$, it is possible to write, up to the second order in epsilon,

$$
\begin{aligned}
\ln\left(\frac{1-\epsilon}{d}\right) &\approx -\ln d - \epsilon - \frac{\epsilon^2}{2}, \\
\ln\left(1 + \frac{\epsilon \mathbb{p}_{x_i} d}{1-\epsilon}\right) &\approx \frac{\epsilon \mathbb{p}_{x_i} d}{1-\epsilon} - \frac{\epsilon^2 \mathbb{p}_{x_i}^2 d^2}{2(1-\epsilon)^2},
\end{aligned}
\tag{4.5.11}
$$

Then, we obtain

$$
S(\rho'_\epsilon) = \sum_i \left(\frac{1-\epsilon}{d} + \epsilon \mathbb{p}_{x_i}\right)\left\{\ln d + \epsilon + \frac{\epsilon^2}{2} - \frac{\epsilon \mathbb{p}_{x_i} d}{1-\epsilon} + \frac{\epsilon^2 \mathbb{p}_{x_i}^2 d^2}{2(1-\epsilon)^2}\right\}
\tag{4.5.12}
$$

and

$$
S(\rho'_\epsilon) = \ln d + \frac{\epsilon^2}{2}\left(1 - d\sum_i \mathbb{p}_{x_i}^2\right).
\tag{4.5.13}
$$

By use of (4.1.14), the last equation is simplified to

$$
S(\Phi_X(\rho_\epsilon)) = \ln d + \frac{\epsilon^2}{2}\left(1 - d\,\|\Phi_X(\rho)\|^2\right)
\tag{4.5.14}
$$

and by (4.1.15), one obtains

$$
S(\Phi_{YX}(\rho_\epsilon)) = \ln d + \frac{\epsilon^2}{2}\left(1 - d\,\|\Phi_{YX}(\rho)\|^2\right).
\tag{4.5.15}
$$

Returning to (4.5.6), with the last two equations, we find

$$
\mathscr{R}_{\mathcal{N}_\epsilon(\mathbb{C})} = \frac{\frac{\epsilon^2}{2} d\left(\|\Phi_X(\rho)\|^2 - \|\Phi_{YX}(\rho)\|^2\right)}{-\frac{\epsilon^2}{2}\left(1 - d\,\|\Phi_X(\rho)\|^2\right)}.
\tag{4.5.16}
$$

Using (4.1.18) and taking the limit of $\epsilon \to 0$, we finally obtain

$$\lim_{\epsilon\to 0} \mathscr{R}_{\mathcal{N}_\epsilon(\mathbb{C})} = \frac{||\Phi_{YX}(\rho) - \Phi_X(\rho)||^2}{||\Phi_X(\rho) - \mathbb{1}/d||^2} =: \mathscr{R}_{\mathbb{C}}, \tag{4.5.17}$$

The emerging expression for $\mathscr{I}_{\mathcal{N}_\epsilon(\mathbb{C})}/I^i_\epsilon$ results to be ϵ-independent and the $\epsilon \to 0$ limit trivially follows. Therefore, by use of this ratio, Bob can still check information leakage for arbitrarily large noise. An interesting feature of the formula (4.5.17) is that it is invariant upon noise maps of the form (4.5.1), that is, $\mathscr{R}_{\mathcal{N}_\epsilon(\mathbb{C})} = \mathscr{R}_{\mathbb{C}}$, for all ρ and $\epsilon \in [0, 1]$. This allows us to write, up to order ϵ^2,

$$\mathscr{I}_{\mathcal{N}_\epsilon(\mathbb{C})} \cong \mathscr{R}_{\mathbb{C}} I^i_\epsilon. \tag{4.5.18}$$

This result is interesting because it factorizes the role of the input information, thus showing that fundamental characteristics of incompatibility are encoded solely in the geometric measure $R_{\mathbb{C}}$, which is, moreover, epsilon-independent. Notice, in particular, that $R_{\mathbb{C}}$ is much easier to compute.

In search of a link between informational incompatibility and measurement incompatibility, the natural move is to restrict ourselves to the context $\mathbb{C}_j \equiv \{X_j, X, Y\}$. Then, setting $\rho = X_j$ in equation (4.5.1), we find

$$\mathscr{R}_{\mathbb{C}_j} = \frac{d}{d-1}\left(1 - ||\Phi_Y(X_j)||^2\right), \tag{4.5.19}$$

which is just the linear entropy of

$$\Phi_Y(X_j) = \sum_k \mathbb{p}_{y_k|x_j} Y_k = \sum_k \mathrm{Tr}(Y_k X_j) Y_k, \tag{4.5.20}$$

and

$$I^i_\epsilon \cong \epsilon^2 (d-1)/2. \tag{4.5.21}$$

It follows that

$$\mathscr{I}_{\mathcal{N}_\epsilon(\mathbb{C}_j)} \cong \mathscr{R}_{\mathbb{C}_j} I^i_\epsilon, \tag{4.5.22}$$

with I^i_ϵ keeping no dependence on the input state X_j. This result is relevant because it proves that the ratio $\mathscr{R}_{\mathbb{C}_j}$, suffices to capture the level of incompatibility in the context $\mathbb{C}_j$. However, it cannot be our definitive figure of merit for quantifying the measurement incompatibility of the set $\{X, Y\}$, since it considers only a single element of the X eigenbasis. We then examine the averaging

$$\mathscr{M}_{\{X,Y\}} := \frac{1}{d}\sum_{j=1}^{d} \mathscr{R}_{\mathbb{C}_j}. \tag{4.5.23}$$

By construction, $\mathscr{M}_{\{X,Y\}}$ tends to be a more appropriate quantifier of measurement incompatibility, for it (i) encompasses the contribution of all X eigenstates and (ii) is symmetrical upon the ordering permutation $X \leftrightarrow Y$, that is, $\mathscr{M}_{\{X,Y\}} = \mathscr{M}_{\{Y,X\}}$ (a desirable property for a measure meant to describe an algebraic relation between two observables). This point can be checked from the manipulated form

$$\mathscr{M}_{\{X,Y\}} = \sum_{j,k=1}^{d} \frac{\left\|[X_j, Y_k]\right\|^2}{2(d-1)} = \frac{1}{d-1}\left(d - \sum_{j,k=1}^{d} \left|\langle x_j | y_k\rangle\right|^4\right), \tag{4.5.24}$$

The first equality also makes it explicit a relation with the commutator

$$[X, Y] = \sum_{j,k} x_j y_k [X_j, Y_k], \tag{4.5.25}$$

which, however, also depends on the spectra of X and Y. This is an important reference to the well-known fact that, when projective measurements are concerned, joint measurability and commutativity turn out to be equivalent notions, although this is not true in general [Ziman (2016)]. Moreover, we have $0 \leq \mathscr{M}_{\{X,Y\}} \leq 1$, with the upper (lower) bound being reached for, and only for, MUB (commuting operators).

4.6 Geometrical Interpretation

Using the generalized Bloch representation in section 2.2, we can build a geometrical picture for the incompatibility measures introduced before. With this formalism, the map $\Phi_X(\rho)$ can be written as

$$\Phi_X(\rho) = \sum_{j=1}^{d} \mathbb{p}_{x_j} X_j = \sum_{j=1}^{d} \left(\frac{1 + (d-1)\hat{\boldsymbol{x}}_j \cdot \boldsymbol{r}}{d} \right) \left(\frac{\mathbb{1} + C_d \hat{\boldsymbol{x}}_j \cdot \boldsymbol{\Lambda}}{d} \right), \tag{4.6.1}$$

which can be simplified to

$$\Phi_X(\rho) = \frac{\mathbb{1} + C_d \boldsymbol{u} \cdot \boldsymbol{\Lambda}}{d} \tag{4.6.2}$$

with

$$\boldsymbol{u} := \frac{d-1}{d} \sum_{j=1}^{d} (\hat{\boldsymbol{x}}_j \cdot \boldsymbol{r})\, \hat{\boldsymbol{x}}_j. \tag{4.6.3}$$

The map $\Phi_{YX}(\rho)$ is written as

$$\Phi_{YX}(\rho) = \frac{\mathbb{1} + C_d \boldsymbol{v} \cdot \boldsymbol{\Lambda}}{d}, \tag{4.6.4}$$

with

$$\boldsymbol{v} = \frac{d-1}{d} \sum_{k=1}^{d} (\hat{\boldsymbol{y}}_k \cdot \boldsymbol{u})\, \hat{\boldsymbol{y}}_k, \tag{4.6.5}$$

where $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{d^2-1}$ and

$$\hat{\boldsymbol{y}}_i \cdot \hat{\boldsymbol{y}}_j = \frac{\delta_{ij}\, d - 1}{d-1}. \tag{4.6.6}$$

The considered observables assume the form $X = \hat{\boldsymbol{x}} \cdot \boldsymbol{\Lambda}$ and $Y = \hat{\boldsymbol{y}} \cdot \boldsymbol{\Lambda}$, where $\hat{\boldsymbol{x}} = (C_d/d) \sum_j x_j \hat{\boldsymbol{x}}_j$ and $\hat{\boldsymbol{y}} = (C_d/d) \sum_k y_k \hat{\boldsymbol{y}}_k$. The relations (4.6.2)–(4.6.4) allow us to speak of the incompatibility

$$\mathscr{I}_{\mathbb{C}} = H\left(\frac{1 + (d-1)\hat{\boldsymbol{y}}_k \cdot \boldsymbol{u}}{d} \right) - H\left(\frac{1 + (d-1)\hat{\boldsymbol{x}}_j \cdot \boldsymbol{r}}{d} \right) \tag{4.6.7}$$

of the "geometrical context" $\mathbb{C} = \{\boldsymbol{r}, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}\}$. In connection with the noisy state (4.5.1), we have $\mathcal{N}_\epsilon(\mathbb{C}) = \{\epsilon \boldsymbol{r}, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}\}$, for which we find a particularly insightful result for the proportionality ratio:

$$\mathscr{R}_{\mathbb{C}} = \frac{\|\boldsymbol{u} - \boldsymbol{v}\|^2}{\|\boldsymbol{u}\|^2} = 1 - \frac{\|\boldsymbol{v}\|^2}{\|\boldsymbol{u}\|^2}. \tag{4.6.8}$$

To compute the measurement incompatibility we set $\rho = X_j$, which implies that $\boldsymbol{r} = \hat{\boldsymbol{x}}_j = \boldsymbol{u}$ and hence,

$$\mathscr{R}_{\mathbb{C}_j} = 1 - \left\|\boldsymbol{v}_j\right\|^2, \tag{4.6.9}$$

with

$$\boldsymbol{v}_j = \frac{d-1}{d} \sum_{k=1}^{d} (\hat{\boldsymbol{y}}_k \cdot \hat{\boldsymbol{x}}_j)\, \hat{\boldsymbol{y}}_k. \tag{4.6.10}$$

It then follows that

$$\mathscr{M}_{\{X,Y\}} = 1 - \frac{1}{d} \sum_{j=1}^{d} \left\|\boldsymbol{v}_j\right\|^2 = 1 - \frac{d-1}{d^2} \sum_{j,k=1}^{d} (\hat{\boldsymbol{x}}_j \cdot \hat{\boldsymbol{y}}_k)^2. \tag{4.6.11}$$

Interestingly, the results (4.6.7)-(4.6.11) rephrase incompatibility in terms of the geometry defined by the vectors $\boldsymbol{r}, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}$. Here the free contexts (4.4.1) manifest themselves with respect to $\mathscr{I}_{\mathbb{C}}$ as $\mathbb{C}_1^{\text{free}} = \{\boldsymbol{r}, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}\}$ with

$$\hat{\boldsymbol{x}}_j \cdot \hat{\boldsymbol{y}}_k = \frac{d\delta_{jk} - 1}{d-1} \tag{4.6.12}$$

("parallel operators", since one has $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}} = \frac{1}{2}\sum_i x_i y_i$), $\mathbb{C}_2^{\text{free}} = \{\hat{0}, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}\}$ $(\forall\, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$, and $\mathbb{C}_3^{\text{free}} = \{\hat{\boldsymbol{y}}_k, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}\}$ with $\hat{\boldsymbol{x}}_j \cdot \hat{\boldsymbol{y}}_l = 0$ ("orthogonal vectors", since $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}} = 0$). Interestingly, as far as $\mathscr{M}_{X,Y}$ is concerned, we see that it vanishes for parallel (commuting) operators and reaches its maximum for orthogonal (MU) operators, this being the cerne of our geometrical interpretation.

The scenario becomes rather simple for generic qubit contexts. By setting $d = 2$, $C_d = 1$, $\boldsymbol{\Lambda} = (\sigma_1, \sigma_2, \sigma_3)$, where $\sigma_{1,2,3}$ are the Pauli matrices, $\hat{\boldsymbol{x}}_j = x_j \hat{\boldsymbol{x}}$, and $\hat{\boldsymbol{y}}_k = y_k \hat{\boldsymbol{y}}$ in the precedent formulas we readily obtain

$$\mathscr{I}_{\mathbb{C}} = h\left(\frac{1 + (\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}})(\hat{\boldsymbol{x}} \cdot \boldsymbol{r})}{2}\right) - h\left(\frac{1 + (\hat{\boldsymbol{x}} \cdot \boldsymbol{r})}{2}\right), \tag{4.6.13}$$

$$\mathscr{M}_{\{X,Y\}} = \mathscr{R}_{\mathbb{C}} = \mathscr{R}_{\mathbb{C}_j} = 1 - (\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}})^2, \tag{4.6.14}$$

where $h(\nu) = -\nu \ln \nu - (1-\nu)\ln(1-\nu)$ is the binary Shannon entropy and $\boldsymbol{r}, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}} \in \mathbb{R}^3$.

Some remarks are in order. First, $\mathscr{I}_{\mathbb{C}}$ is the only quantity that depends on the state ρ (via $\boldsymbol{r}$), this being the key aspect characterizing it as a informational incompatibility. In particular, this ensures that $\mathscr{I}_{\mathbb{C}} \to 0$ as $|\boldsymbol{r}| \to 0$ (decoherence-induced classical limit). The "large mass" classical limit, on the other hand, effectively comes via $\mathrm{x} \cdot \mathrm{y} \cong 1$, which implements the nondisturbance scenario and implies, via $\Phi_{YX}(\rho) \cong \Phi_X(\rho)$, that $\mathscr{I}_{\mathbb{C}} \cong 0$ (see the following sections for details). This regime is, of course, equivalent to the free context $\mathbb{C}_1^{\text{free}}$, where $\hat{\boldsymbol{x}} = \hat{\boldsymbol{y}}$ (implying parallel operators, that is, $[X, Y] = 0$). The informational incompatibility vanishes also when $\hat{\boldsymbol{x}} \cdot \boldsymbol{r} = 0$—case in which the first measurement X is incompatible with the input state—and monotonically increases with the quantifiers given by equation (4.6.14). Second, by taking

$$4\mathscr{N}_{\{X,Y\}}^2 = \tfrac{1}{4}\,\|[X,Y]\|^2 = |\hat{\boldsymbol{x}} \times \hat{\boldsymbol{y}}|^2 \tag{4.6.15}$$

as an estimate for the notion of noncommutativity, we see that the ratios $\mathscr{R}_{\mathbb{C}}$ and $\mathscr{R}_{\mathbb{C}_j}$, with $\mathbb{C}_j = \{X_j, X, Y\}$, and the measurement incompatibility $\mathscr{M}_{\{X,Y\}}$ are all indistinguishable concepts for qubit contexts. To a certain extent, this can be related to the bidirectional implication reported in reference [Heinosaari (2010)] between the notions of nondisturbance and commutativity.

4.7 Position and Momentum

Using the method introduced in 2.6 we can write the CPTP map given by equation (4.1.5), now for momentum and position observables in the previous formalism. Consider first a map for unrevealed position measurements, in $\{|q_k\rangle\}$ basis, which is given by

$$\Phi_Q(\rho) = \sum_k Q_k\, \rho\, Q_k, \tag{4.7.1}$$

where Q_k are position projectors given by equation (2.6.3) and $k \in \left[-L_q, L_q\right] = \left[-\frac{\xi-1}{2}, \frac{\xi-1}{2}\right]$, where ξ is function of the resolutions.

The unrevealed-measurement map for momentum is written as

$$\Phi_P(\rho) = \sum_l P_l\, \rho\, P_l. \tag{4.7.2}$$

The map of sequential measurements then reads

$$\Phi_{PQ}(\rho) = \sum_l P_l\, \Phi_Q(\rho)\, P_l. \tag{4.7.3}$$

One can also consider the opposite scenario, where the measurements are performed in reverse order, that is,

$$\Phi_{QP}(\rho) = \sum_k Q_k\, \Phi_P(\rho)\, Q_k. \tag{4.7.4}$$

Now consider a pure quantum state given by

$$\rho = |\psi\rangle\langle\psi| = \sum_{k,k'} (\delta q)^2\, \psi(q_k)\, \psi^*(q_{k'})\, |q_k\rangle\langle q_{k'}| \tag{4.7.5}$$

Calculating the unrevealed map for position measurements with equation (4.7.1) and using the state in equation (4.7.5), we have

$$\begin{aligned}
\Phi_Q(\rho) &= \sum_{k''} Q_{k''} \rho Q_{k''} \\
&= \sum_{k,k',k''} (\delta q)^4\, \psi(q_k)\, \psi^*(q_{k'})\, |q_{k''}\rangle\langle q_{k''}|q_k\rangle\langle q_{k'}|q_{k''}\rangle\langle q_{k''}| \\
&= \sum_{k,k',k''} (\delta q)^4\, \psi(q_k)\, \psi^*(q_{k'}) \frac{\delta_{k''k}}{\delta q}\frac{\delta_{k'k''}}{\delta q}\, |q_{k''}\rangle\langle q_{k''}| \\
&= \sum_k (\delta q)^2\, |\psi(q_k)|^2\, |q_k\rangle\langle q_k| \\
&= \sum_k \delta q\, |\psi(q_k)|^2\, Q_k.
\end{aligned}$$

From equation (3.4.32) we have

$$\Phi_Q(\rho) = \sum_k \mathbb{p}_k Q_k. \tag{4.7.6}$$

We can calculate the eigenvalue equation for $\Phi_Q(\rho)$, which is

$$\begin{aligned}\Phi_Q(\rho)\,|q_k\rangle &= \sum_{k'} \delta q \mathbb{p}_k \,|q_{k'}\rangle \langle q_{k'}|q_k\rangle \\ &= \mathbb{p}_k \sum_{k'} \delta q \frac{\delta_{k'k}}{\delta q}\,|q_{k'}\rangle \\ &= \mathbb{p}_k\,|q_k\rangle\,. \end{aligned} \tag{4.7.7}$$

so, $|q_k\rangle$ are also eigenstate of $\Phi_Q(\rho)$ with respective eigenvalues $\mathbb{p}_k$.

Calculating the unrevealed map for measurements of position and momentum, sequentially, as given by equation (4.7.3) we find

$$\begin{aligned}\Phi_{PQ}(\rho) &= \sum_l P_l\,\Phi_Q(\rho)\,P_l \\ &= \sum_{k,l} (\delta q)^2(\delta p)^2\,|\psi(q_k)|^2\,|p_l\rangle\langle p_l|q_k\rangle\langle q_k|p_l\rangle\langle p_l| \\ &= \sum_{k,l} (\delta q)^2(\delta p)^2\,|\psi(q_k)|^2\,\frac{e^{-2\pi ikl/\xi}}{\sqrt{2\pi\hbar}}\frac{e^{2\pi ikl/\xi}}{\sqrt{2\pi\hbar}}\,|p_l\rangle\langle p_l|\,. \end{aligned} \tag{4.7.8}$$

We have equation (2.6.16), which relates the resolutions product $\delta q\delta p$ to the space dimension ξ we can write the product as $\delta q\delta p = 2\pi\hbar/\xi$, therefrom we obtain

$$\begin{aligned}\Phi_{PQ}(\rho) &= \frac{1}{2\pi\hbar}\sum_{k,l}(\delta q)^2(\delta p)^2\,|\psi(q_k)|^2\,|p_l\rangle\langle p_l| \\ &= \frac{1}{\xi}\sum_{k,l}\delta q\,|\psi(q_k)|^2\,p_l \\ &= \frac{1}{\xi}\sum_k \mathbb{p}_k \sum_l p_l, \end{aligned}$$

with equations (2.6.7) and (2.6.11), we finally obtain

$$\Phi_{PQ}(\rho) = \frac{\mathbb{1}}{\xi}, \tag{4.7.9}$$

and it is trivial to compute its eigenvalue equation

$$\begin{aligned}\Phi_{PQ}(\rho)\,|p_l\rangle &= \frac{\mathbb{1}}{\xi}\,|p_l\rangle \\ &= \frac{1}{\xi}\,|p_l\rangle, \end{aligned} \tag{4.7.10}$$

where $\Phi_{PQ}(\rho)$ has eigenvalue $1/\xi$.

4.7.1 Informational Incompatibility for $\mathbb{C} = \{Q, P, \rho\}$

Combining equation (4.7.6) and (4.7.9), we can write $\mathscr{I}_\mathbb{C}$ for the context $\mathbb{C} = \{Q, P, \rho\}$, as

$$\mathscr{I}_\mathbb{C}(\rho) = S\left(\Phi_{PQ}(\rho)\right) - S\left(\Phi_Q(\rho)\right). \tag{4.7.11}$$

Replacing equation (4.7.6) and equation (4.7.9) in equation (4.7.11), we find

$$\mathscr{I}_{\mathbb{C}}(\rho) = S\left(\frac{\mathbb{1}}{\xi}\right) - S\left(\sum_k \mathbb{p}_k Q_k\right), \tag{4.7.12}$$

which can be expanded as

$$\mathscr{I}_{\mathbb{C}}(\rho) = -\sum_m \langle q_m| \frac{\mathbb{1}}{\xi} \ln \frac{\mathbb{1}}{\xi} |q_m\rangle + \sum_n \langle q_n| \left(\sum_k \mathbb{p}_k Q_k\right) \ln \left(\sum_k \mathbb{p}_k Q_k\right) |q_n\rangle . \tag{4.7.13}$$

Using equations (4.7.7) and (4.7.10), we obtain

$$\mathscr{I}_{\mathbb{C}}(\rho) = \ln \xi - H(\mathbb{p}_k). \tag{4.7.14}$$

Consider the case where the initial state ρ is a position eigenstate, *i.e.*, $\rho = Q_k$. Then equation (4.7.1) reduces to

$$\begin{aligned} \Phi_Q(Q_k) &= \sum_{k'} Q_{k'} Q_k Q_{k'} \\ &= \sum_{k'} Q_{k'} Q_k \delta_{k'k} \\ &= Q_k Q_k \\ &= Q_k, \end{aligned} \tag{4.7.15}$$

where the properties of equation (2.6.4) have been used. The corresponding eigenvalue equation is

$$\begin{aligned} \Phi_Q(Q_{k'}) |q_k\rangle &= \delta q \, |q_{k'}\rangle \langle q_{k'}|q_k\rangle \\ &= \delta q \, |q_{k'}\rangle \frac{\delta_{k'k}}{\delta q} \\ &= |q_k\rangle , \end{aligned} \tag{4.7.16}$$

so, the $\Phi_Q(Q_{k'})$ eigenvalue is 1.

The entropy of equation (4.7.15), using equations (2.6.18) and (4.7.16), is

$$\begin{aligned} S\left(\Phi_Q(Q_k)\right) &= \sum_m \langle q_m| \left(Q_k \ln Q_k\right) |q_m\rangle \\ &= -\ln 1 \sum_m \delta q \, \langle q_m|q_m\rangle \\ &= 0. \end{aligned} \tag{4.7.17}$$

Combining the results (4.7.14) and (4.7.17), we then have

$$\mathscr{I}_{\{Q,P\}} = \ln \xi. \tag{4.7.18}$$

This means that when one considers a Q eigenstate as input state, the informational incompatibility is maximum, indicating that Q and P are maximally incompatible, as expected.

4.8 Classical Limit

Suppose that, after being prepared in a generic state [see equation (2.2.18) with $d = 2$], an electron is submitted to a sequence of two Stern-Gerlach (SG) magnets, one with magnetic field $B_x\hat{\boldsymbol{x}}$ and the other with $B_y\hat{\boldsymbol{y}}$. Here the scenario is such that the physical interaction generates spin-position correlations that perfectly induce measurements of the observables $X = \hat{\boldsymbol{x}} \cdot \boldsymbol{\sigma}$ and $Y = \hat{\boldsymbol{y}} \cdot \boldsymbol{\sigma}$. That is, the electron spin couples with the magnetic fields in a way such that one can clearly distinguish between displacements of the electron trajectory along the axes x or y. In this case, the X and Y bases form MUB and $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}} = 0$, which maximizes both the context incompatibility and the resulting measurement incompatibility [see equations (4.6.13) and (4.6.14)]. Now, consider that the same SG apparatuses are used to measure the spin of an extremely massive particle. In this case, the coupling with the magnetic field is not enough to cause a significant deviation in the particle trajectory. In other words, the correlations established between position and spin are tiny, and the resulting "measurement" becomes effectively fuzzy. The larger the mass, the greater the difficulty to experimentally distinguish between the observables $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$. As a consequence, $\hat{\boldsymbol{x}} \cong \hat{\boldsymbol{y}}$ and $\mathscr{I}_{\mathbb{C}} \cong 0$.

But our approach admits an even deeper treatment of this problem. To make the point, let us imagine that the SG magnets are free to move upon interaction with the passing particle, that is, the apparatuses can receive kickbacks that guarantee total momentum conversation. Being free to move, the magnets earn the right to be treated quantum mechanically. In this case, via the Stinespring theorem, the unrevealed-measurement map is written

$$\Phi_X(\rho) = \mathrm{Tr}_{\mathrm{SG}}\left(U\rho \otimes \rho_{\mathrm{SG}} U^\dagger\right), \tag{4.8.1}$$

with ρ_{SG} being the apparatus state. At this stage, the notion of collapsing measurement associated with the projectors X_j is replaced with correlations created between the systems by U. When the mass of the particle is large and the coupling with the field is weak, the correlations are nearly negligible, which legitimates us to replace the projective-measurement map Φ_X with the weak-measurement map [Dieguez (2018)]

$$M_X^\epsilon(\rho) \coloneqq (1-\epsilon)\,\rho + \epsilon\,\Phi_X(\rho), \tag{4.8.2}$$

where $0 \le \epsilon \le 1$. Being related to the effective strength of the measurement, ϵ can be shown [Dieguez (2018)] to emerge from the strength of the physical coupling. In the problem under scrutiny here, since the interaction is momentum conserving, then we can expect that $\epsilon \propto m_{\mathrm{SG}}/m$, where m (m_{SG}) is the mass of the particle (SG magnet). Upon the replacement $\Phi \longrightarrow M^\epsilon$, context incompatibility becomes

$$\mathscr{I}_{\mathbb{C}} = S\big(M_Y^\epsilon M_X^\epsilon(\rho)\big) - S\big(M_X^\epsilon(\rho)\big). \tag{4.8.3}$$

In the regime where $\epsilon \ll 1$, we have

$$M_Y^\epsilon M_X^\epsilon(\rho) \cong M_X^\epsilon(\rho) + \epsilon\big[\Phi_Y(\rho) - \rho\big], \tag{4.8.4}$$

which allows us to conclude that the large-mass limit, $(m_{\mathrm{SG}}/m) \longrightarrow 0$, implies that $\mathscr{I}_{\mathbb{C}} \longrightarrow 0$. When the particle is lightweight in comparison with the magnets, then the model can be improved as $\epsilon \propto 1 - \exp\left(-m_{\mathrm{SG}}/m\right)$, which correctly yields $\epsilon \longrightarrow 1$ for $(m_{\mathrm{SG}}/m) \longrightarrow \infty$ and also encapsulates the large-mass limit. Although the discussion has been conducted here for $d = 2$, similar arguments apply in general.

4.9 Comparison Between our Measures of Informational Incompatibility

Two quantifiers of quantum incompatibility for a context were introduced in this manuscript: the first, called Bayes incompatibility, based on Bayes' rule violations, and the second, the so-called informational incompatibility, related to information leakage in a communication task. The natural question would be what the difference is between them. To address this question, we consider again the case of a conical context for a qubit, which is defined by the constraint $\hat{\boldsymbol{x}} \cdot \boldsymbol{r} = \hat{\boldsymbol{x}} \cdot \boldsymbol{r}$, with $\boldsymbol{r}$ constituting the symmetry axis of the cone. As discussed previously, given the angular equivalence of $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ with respect to $\boldsymbol{r}$, which thus "remove" the role of the state from the context, this instance is expected to refer solely to the incompatibility between the couple $\{X, Y\}$. In other words, here our measure should be related with the purely algebraic notion of non-commutativity, which is captured by non-commutativity degree. To test these ideas, numerical simulations were performed comparing non-commutativity degree, Bayes' incompatibility and informational incompatibility of a context. In figure 4.2, we present parametric plots between informational incompatibility and non-commutativity degree, and between informational incompatibility and Bayes' incompatibility, as a function of $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}} \in [0, 1]$ for several values of $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{r}}$ (which is equal to $\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{r}}$). The results demonstrate, for the case of qubits, that the measures non-commutativity degree, Bayes incompatibility, and informational incompatibility are monotonically related to each another, which implies their essential equivalence. Although we demonstrated that informational incompatibility is a resource, due to time restriction it was not possible to demonstrate to Bayes' incompatibility. In this sense, informational incompatibility is more robust that Bayes' incompatibility. Figure 4.3 presents an instructive comparison between Bayes' incompatibility and a symmetrized version of informational incompatibility, which is defined as

$$\mathscr{I}_s := \frac{\mathscr{I}_{\{\boldsymbol{r}, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}\}}}{\eta_1} + \frac{\mathscr{I}_{\{\boldsymbol{r}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{x}}\}}}{\eta_2}. \tag{4.9.1}$$

where,

$$\eta_1 = \lim_{\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{r}} \to 1} \mathscr{I}_{\{\boldsymbol{r}, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}\}} \tag{4.9.2}$$

and

$$\eta_2 = \lim_{\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{r}} \to 1} \mathscr{I}_{\{\boldsymbol{r}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{x}}\}}. \tag{4.9.3}$$

In this case, we fixed $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}} = 1/2$ and let $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{r}}$ vary independently in the interval $[0, 1]$. We see that both measures monotonically increase with $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{r}}$.

This result is valid for qubits. For instance, systems with higher dimensions are inconclusive, needing further investigation. However, this is an interesting comparison; it validates what is expected, quantifiers are equal footing.

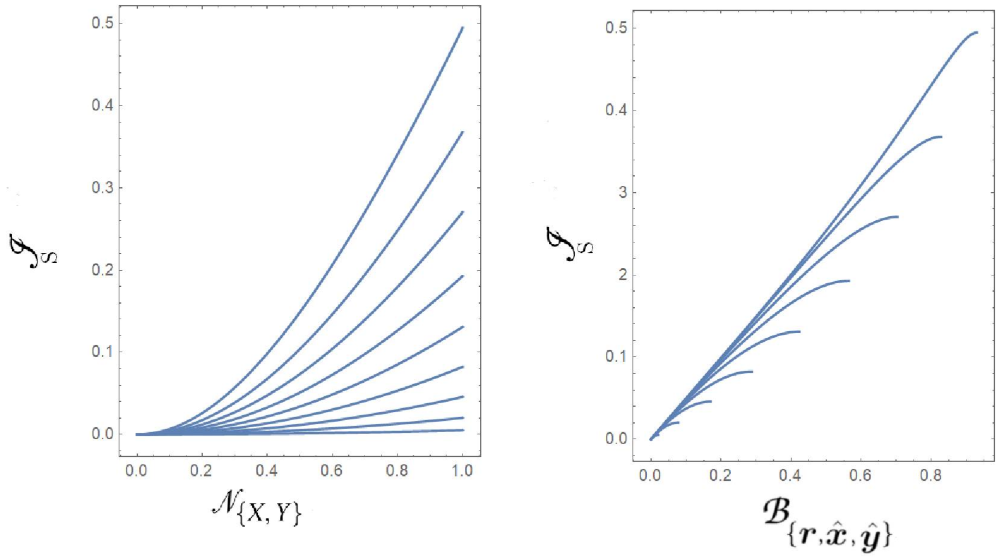


Figure 4.2: In the graphs, each blue curve refers to a specific value of $\hat{\boldsymbol{x}} \cdot \boldsymbol{r}$ (which is equal to $\hat{\boldsymbol{y}} \cdot \boldsymbol{r}$) in the set $\{0.1, ..., 0.9\}$ with increments of 0.1, so that the lower curve corresponds to 0.1 and the upper one to 0.9. On the left-hand side, a graph symmetric informational incompatibility of a physical context versus the non-commutativity degree. On the right-hand side, a graph informational incompatibility of a physical context versus Bayes' incompatibility. Both graphs for the conical context, namely, $\hat{\boldsymbol{x}} \cdot \boldsymbol{r} = \hat{\boldsymbol{x}} \cdot \boldsymbol{r}$. The plot takes the same values as figure 3.7.

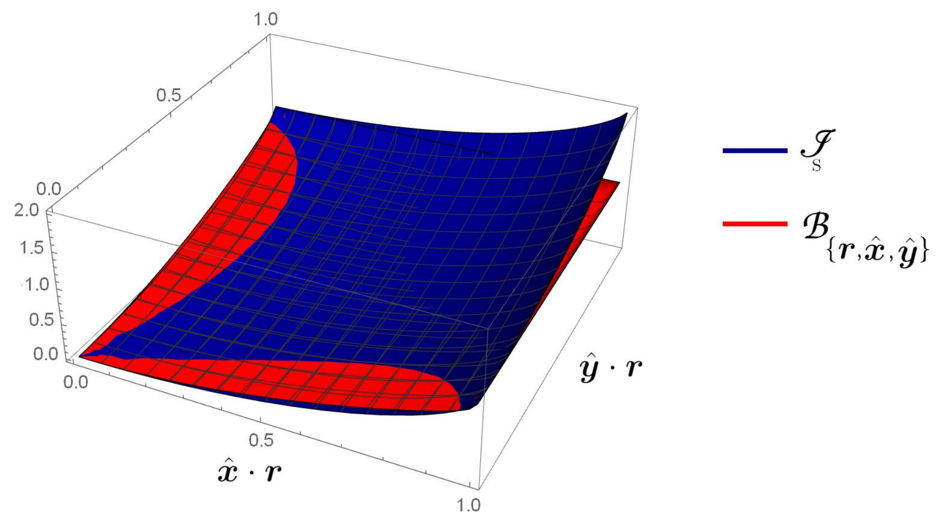


Figure 4.3: Bayes incompatibility (red surface) and symmetrized informational incompatibility (blue surface) as a function of $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{r}}$ for $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}} = 1/2$.

CONCLUSION AND OUTLOOK

In this work a notion of incompatibility was introduced whose quantifier does not rely on commutators, but rather on physical context $\mathbb{C} = \{\rho, X, Y\}$ and in its underlying probability distributions. Besides allowing one to describe the disappearance of incompatibility in the classical regime, our results defines incompatibility as violation of Bayes' rule and an information-based task in space-time, rather than an algebraic construction in the Hilbert space.

The first proposed measure Bayes' incompatibility brings together two different scientific subjects to treat a topic not very well defined in the academic literature. From one side, Bayes' rule, which is one of the most fundamental concepts in classical theory. From the other, the KL divergence—a central tool in information theory. Combined, these concepts yield a measure to quantify quantum incompatibility for a context. When applied to a simple system, as a qubit, this measure shows up very intuitive, admitting even geometrical interpretations. In particular, as far as position and momentum are concerned, our measure leads to the expected result, namely, maximum incompatibility for states that are eigenstates of one of the observables. However, it is not always an easy task to compute it in an analytical form, since the number of variables gets bigger and bigger with the space dimension.

On the other hand, informational incompatibility is easily computable and yet admits a norm-based estimate. Remarkably, the informational incompatibility is shown to be a resource, with particular application to a protocol devised to test information leakage, makes contact with the notion of measurement incompatibility, and admits a geometrical interpretation in a vector space of arbitrary dimension. Our results give rise to some noteworthy research lines. The first one concerns the extension of our approach to contexts involving more than two (eventually continuous) observables. The second refers to the use of our easily computable quantifier of measurement incompatibility for MUB searching, a long-lasting intricate problem in quantum physics.

It has recently been shown that the notion of incompatibility can be extended to more general scenarios, as for instance the one involving quantum operations and quantum channels [Ziman (2016)]. Little is known, however, about the connections with foundational and quantum-information quantities such as quantum entanglement, quantum nonlocality and quantum irrealism. In addition, the research concerning the relevance of quantum incompatibility in the interface between quantum mechanics and relativity, or even problems involving quantum field theory, is rather incipient [Savi (2020)]. All this together defines a formidable research program for future works.

REFERENCES

[1] W. Heisenberg, "Über den Anschaulichen Inhalt der Quantentheoretischen Kinematik und Mechanik", *Z. Phys.* **43**, 172 (1928).

[2] N. Bohr, "The Quantum Postulate and the Recent Development of Atomic Theory", *Nature* **121**, 580 (1928).

[3] R. Gomatam, "Niels Bohr's Interpretation and the Copenhagen Interpretation — Are the two incompatible?", *Philos. Sci.* **74:5**, 736 (2007).

[4] D. Saha, M. Oszmaniec, and M. L. Czekaj, "Operational Foundations for Complementarity and Uncertainty Relations Fundamental Principles of Quantum Theory", *Phys. Rev. A* **101**, 052104 (2020).

[5] P. A. M. Dirac, *The Principles of Quantum Mechanics*, (Oxford University Press, Oxford) 1930.

[6] G. Temple, *The General Principles of Quantum Theory*, (1st edition, John Wiley and Sons) 1934.

[7] A. Messiah, *Quantum Mechanics*, Vol. 1, (1st edition, John Wiley and Sons, Inc.) 1961.

[8] R. Shankar, *Principles of Quantum Mechanics*, (2nd edition, Springer) 1994.

[9] J. J. Sakurai and J. J. Napolitano, *Modern Quantum Mechanics*, (2nd edition, Addison Wesley Longman) 1994.

[10] E. Merzbacher, *Quantum Mechanics*, (3rd edition, Wiley) 1997.

[11] T. Banks, *Quantum Mechanics: An Introduction*, (1st edition, CRC Press) 2018.

[12] R. H. Dicke and J. P. Wittke, *Introduction to Quantum Mechanics*, (1st edition, Addison-Wesley Pub. Co.) 1960.

[13] F. Schwabl, *Quantum Mechanics*, (4th edition, Springer) 2007.

[14] V. Moretti, *Fundamental Mathematical Structures of Quantum Theory*, (2nd edition, Springer) 2019.

[15] F. Mandl, *Quantum Mechanics*, (1st edition, John Wiley and Sons) 1992.

[16] D. J. Griffiths and D. F. Schroeter, *Introduction to Quantum Mechanics*, (3rd edition, Cambridge University Press, Cambridge) 2018.

[17] H. J. W. Müller-Kirsten, *Introduction to Quantum Mechanics - Schrodinger Equation and Path Integral*, (2nd edition, World Scientific Publishing Company) 2006.

[18] L. Schiff, *Quantum Mechanics*, (3rd edition, McGraw-Hill Education) 1968.

[19] S. Weinberg, *Lectures on Quantum Mechanics*, (2nd edition, Cambridge University Press, Cambridge) 2015.

[20] T. Heinosaari, J. Kiukas, D. Reitzner, and J. Schultz, "Incompatibility Breaking Quantum Channels", *J. Phys. A: Math. Theor.* **48**, 435301 (2015).

[21] F. Buscemi, E. Chitambar, and W. Zhou, "Complete Resource Theory of Quantum Incompatibility as Quantum Programmability", *Phys. Rev. Lett.* **124**, 120401 (2020).

[22] S. Designolle, M. Farkas, and J. Kaniewski, "Incompatibility Robustness of Quantum Measurements: a Unified Framework", *New J. Phys.* **21**, 113053 (2019).

[23] W. H. Zurek, "Decoherence and the Transition from Quantum to Classical — Revisited", in *Quantum Decoherence: Poincaré Seminar 2005*, Editors B. Duplantier, J.-M. Raimond, and V. Rivasseau, 1 (Birkhäuser Basel, Basel, 2007).

[24] P. R. Dieguez and R. M. Angelo, "Information-reality complementarity: The Role of Measurements and Quantum Reference Frames", *Phys. Rev. A* **97**, 022107 (2018).

[25] C. Cohen-Tannoudji, B. Dui, and F. Laloe, *Quantum Mechanics* Vol. 1, (1st edition, Wiley) 1991.

[26] P. Busch, "Unsharp Reality and Joint Measurements of Spin Observables", *Phys. Rev. D* **33**, 2253 (1986).

[27] T. Heinosaari, D. Reitzner, and P. Stano, Notes on Joint Measurability of Quantum Observables, *Found. Phys.* **38**, 1133 (2008).

[28] T. Heinosaari, T. Miyadera, and M. Ziman, "An Invitation to Quantum Incompatibility", *J. Phys. A: Math. Theor.* **49**, 123001 (2016).

[29] A. Fine, "Hidden Variables, Joint Probability, and the Bell Inequalities", *Phys. Rev. Lett.* **48**, 291 (1982).

[30] Ll. Masanes, A. Acin, and N. Gisin, "General Properties of Nonsignaling Theories", *Phys. Rev. A* **73**, 012112 (2006).

[31] M. M. Wolf, D. Perez-Garcia, and C. Fernandez, "Measurements Incompatible in Quantum Theory Cannot Be Measured Jointly in any other No-Signaling Theory", *Phys. Rev. Lett.* **103**, 230402 (2009).

[32] M. T. Quintino, J. Bowles, F. Hirsch, and N. Brunner, "Incompatible Quantum Measurements Admitting a Local-Hidden-Variable Model", *Phys. Rev. A* **93**, 052115 (2016).

[33] F. Hirsch, M. T. Quintino, and N. Brunner, "Quantum measurement incompatibility does not imply Bell nonlocality", *Phys. Rev. A* **97**, 012129 (2018).

[34] E. Bene and T. Vértesi, "Measurement Incompatibility Does Not Give Rise to Bell Violation in General", *New J. Phys.* **20**, 013021 (2018).

[35] M. T. Quintino, T. Vértesi, and N. Brunner, "Joint Measurability, Einstein-Podolsky-Rosen Steering, and Bell Nonlocality", *Phys. Rev. Lett.* **113**, 160402 (2014).

[36] R. Uola, T. Moroder, and O. Gühne, "Joint Measurability of Generalized Measurements Implies Classicality", *Phys. Rev. Lett.* **113**, 160403 (2014).

[37] R. Uola, C. Budroni, O. Gühne, and J. P. Pellonpää, "One-to-One Mapping between Steering and Joint Measurability Problems", *Phys. Rev. Lett.* **115**, 230402 (2015).

[38] P. Busch, T. Heinosaari, J. Schultz, and N. Stevens, "Comparing the Degrees of Incompatibility Inherent in Probabilistic Physical Theories", *Europhys. Lett.* **103**, 10002 (2013).

[39] T. Heinosaari, J. Kiukas, and D. Reitzner, "Noise Robustness of the Incompatibility of Quantum Measurements", *Phys. Rev. A* **92**, 022115 (2015).

[40] E. Haapasalo, "Robustness of Incompatibility for Quantum Devices", *J. Phys. A: Math. Theor.* **48**, 255303 (2015).

[41] D. Cavalcanti and P. Skrzypczyk, "Quantitative Relations between Measurement Incompatibility, Quantum Steering, and Nonlocality", *Phys. Rev. A* **93**, 052112 (2016).

[42] S.-L. Chen, C. Budroni, Y.-C. Liang, and Y.-N. Chen, "Natural Framework for Device-Independent Quantification of Quantum Steerability, Measurement Incompatibility, and Self-Testing", *Phys. Rev. Lett.* **116**, 240401 (2016).

[43] S.-L. Chen, C. Budroni, Y.-C. Liang, and Y.-N. Chen, "Exploring the Framework of Assemblage Moment Matrices and its Applications in Device-Independent Characterizations", *Phys. Rev. A* **98**, 042127 (2018).

[44] C. Carmeli, T. Heinosaari, and A. Toigo, "State Discrimination with Postmeasurement Information and Incompatibility of Quantum Measurements", *Phys. Rev. A* **98**, 012126 (2018).

[45] C. Carmeli, T. Heinosaari, and A. Toigo, "Quantum Incompatibility Witnesses", *Phys. Rev. Lett.* **122**, 130402 (2019).

[46] R. Uola, T. Kraft, J. Shang, X.-D. Yu, and O. Gühne, "Quantifying Quantum Resources with Conic Programming", *Phys. Rev. Lett.* **122**, 130404 (2019).

[47] P. Skrzypczyk, I. Šupić, and D. Cavalcanti, "All Sets of Incompatible Measurements give an Advantage in Quantum State Discrimination", *Phys. Rev. Lett.* **122**, 130403 (2019).

[48] M. Horodecki, P. Horodecki, and J. Oppenheim, "Reversible Transformation from Pure to Mixed States, and the Unique Measure of Information", *Phys. Rev. A* **67**, 062104 (2003).

[49] E. Chitambar and G. Gour, "Quantum Resource Theories", *Rev. Mod. Phys.* **91**, 025001 (2019).

[50] A. C. S. Costa and R. M. Angelo, “Information-Based Approach Towards a Unified Resource Theory”, preprint arXiv 2001.07489 (2020).

[51] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, “Quantum Cryptography”, *Rev. Mod. Phys.* **74**, 145 (2002).

[52] C. A. Fuchs and A. Peres, “Quantum State Disturbance vs. Information Gain: Uncertainty Relations for Quantum Information”, *Phys. Rev. A* **53**, 2038 (1996).

[53] P. Busch, T. Heinosaari, J. Schultz, and N. Stevens, “No Information Without Disturbance”: Quantum Limitations of Measurement”, in *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle. The Western Ontario Series in Philosophy of Science*, Editors W. C. Myrvold and J. Christian, **73**, 229 (Springer, Dordrecht, 2009).

[54] D. Aerts and M. S. de Bianchi, “The Extended Bloch Representation of Quantum Mechanics and the Hidden-Measurement Solution to the Measurement Problem”, *Ann. Phys.* **351**, 975 (2014).

[55] D. M. Appleby, “Symmetric Informationally Complete Measurements of Arbitrary Rank”, *Opt. Spectrosc.* **103**, 416–428 (2007).

[56] A. Kolmogorov, *Foundations of the Theory of Probability*, (Chelsea Publishing Company), 1950.

[57] S. M. Ross, *A First Course in Probability*, (9th edition, Pearson), 2012

[58] E. Witten, “A Mini-Introduction to Information Theory” *Riv. del Nuovo Cim.* **43**, 187 (2020).

[59] J. Preskill, Lecture notes (2019), http://www.theory.caltech.edu/~preskill/ph219/ph219_2019-20.

[60] C. E. Shannon, “A Mathematical Theory of Communication”, *B. Syst. Tech. J.* **27**, 3 (1948).

[61] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, (Cambridge University Press, Cambridge) 2000.

[62] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, (2nd edition, Wiley-Interscience) 2006.

[63] J. von Neumann, “Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik [Probabilistic Theory of Quantum Mechanics]”, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* **1**, 245 (1927).

[64] G. Landi, “Quantum Information and Quantum Noise” Lecture notes (2018), http://www.fmt.if.usp.br/~gtlandi/courses/qinfo.html.

[65] W. H. Zurek, “Decoherence, Einselection, and the Quantum Origins of the Classical”, *Rev. Mod. Phys.* **75**, 715 (2003).

[66] T. Baumgratz, M. Cramer, and M. B. Plenio, “Quantifying Coherence”, *Phys. Rev. Lett.* **113**, 140401 (2014).

[67] I. S. Freire and R. M. Angelo, "Quantifying Continuous-Variable Irrealism", preprint arXiv 1904.02490v1 (2019).

[68] M. K. Shukla, R. Mundra, A. K Pati, I. Chakrabarty, and J. Wu, "Quantumness of Relative Incompatibility", preprint arXiv 1909.12534 (2019).

[69] E. Martins, M. F. Savi, and R. M. Angelo, "Quantum Incompatibility of a Physical Context", preprint arXiv 2004.01008 (2020).

[70] T. Durt, B.-G. Englert, I. Bengtsson, and K. Życzkowski, "On Mutually Unbiased Bases", *Int. J. Quant. Info.* **8**, 535 (2010).

[71] A. Streltsov, G. Adesso, and M. B. Plenio, "Quantum Coherence as a Resource", *Rev. Mod. Phys.* **89**, 041003 (2017).

[72] T. Heinosaari and M. Wolf, "Nondisturbing Quantum Measurements", *J. Math. Phys.* **51**, 092201 (2010).

[73] P. R. Dieguez and R. M. Angelo, "Information-Reality Complementarity: The Role of Measurements and Quantum Reference Frames", *Phys. Rev. A* **97**, 022107 (2018).

[74] M. F. Savi and R. M. Angelo, "Quantum Resources Covariance", preprint arXiv 2005.09612 (2020).