From Incompatibility to Optimal Joint Measurability in Quantum Mechanics

Thomas Joseph Bullock

PhD

University of York

Mathematics

September, 2015

Abstract

This thesis is concerned with several topics related to concept of incompatibility of quantum observables. The operational description of quantum theory is given, in which incompatibility is expressed in terms of joint measurability. A connection between symmetric informationally complete positive operator-valued measures and mutually unbiased bases is given, and examples of this connection holding based on investigations in Mathematica are presented. An extension of the Arthurs-Kelly measurement model is then given, where the measured observable is calculated, thereby generalising the results given previously in the literature. It is shown that in the case of prior correlations between measurement probes there exists the possibility that a measurement of both probes leads to marginal observables with smaller statistical spread than if measurements are performed on the individual probes. This concept is then highlighted by considering two probe states that allow for this reduction in spread, and the required conditions for success are given. Finally, error-error relations for incompatible dichotomic qubit observables are considered in the case of state-dependent and independent error measures. Quantities that arise in the state-independent measures case, which were previously presented geometrically, have been given operational meaning, and optimal approximating schemes in both cases are compared. Limitations regarding the state-dependent optimal approximations, and experimental work built upon this construction are also discussed.

Contents

List of Figures

List of Tables

Acknowledgements

I am eternally grateful to a great number of people for their support during the past few years of my PhD. Firstly, this PhD was funded in full by the White Rose Studentship Network *Optimising Quantum Processes and Quantum Devices for future Digital Economy Applications*. Secondly, I must thank my family for their patience and help. For similar reasons, I thank the likes of Paul Beauchamp, Alex Keen, Francesca Holdrick, Mevan Babakar, Ed Donnellan, Ryan Morgan and Robert Anderson for keeping me sane and motivating me to continue through difficult times. I thank other students, including Mark Pearce, Oliver Goodbourn, Benjamin Lang, Spiros Kechrimparis, Chris Draper, Leon Loveridge and Dave Hunt for fuelling my interest in mathematics and quantum mechanics, amongst other things. Finally, I thank my supervisor Paul Busch and secondary supervisor Pieter Kok, both of whom have provided me with a great deal of insight throughout my PhD and have both given essential feedback during the writing of this thesis. I consider both of you to be dear friends, and I consider myself extremely fortunate to have been able to work with you.

Declaration

I declare that the work presented in this thesis is original and that it has not been previously submitted for a degree at this, or any other university. This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by explicit references.

The majority of work in this thesis has been published and, in the case of Chapter 5, is soon to be submitted for publication. The work in Chapter 3 forms the first half of the following paper:

R. Beneduci, T. Bullock, P. Busch, C. Carmeli, T. Heinosaari and A. Toigo, *Phys. Rev. A* **88**, 032312 (2013),

of which I am a co-author, whilst the work in Chapter 4 is contained in

T. Bullock and P. Busch, *Phys. Rev. Lett.* **113**, 120401 (2014),

where I am the main author.

Chapter 1

Introduction

Quantum theory is, to date, one of the most successful theories ever developed, providing us with considerable insights into the nature of that which is far too small for us to directly access. In doing so it has improved our understanding of biology, chemistry and physics, as well as significantly expanding our technological capabilities. Such developments are only possible as a result of the peculiar properties of the theory that differentiate it from the classical mechanics that preceded it, including its probabilistic and contextual nature. Indeed, this difference has resulted in a great deal of philosophical discussion on the reality behind the theory, with many competing interpretations arising.

We shall not focus on such discussions in this thesis. In what follows we will be considering quantum theory rather pragmatically, focussing on probability measures relating to measurement outcomes, with the objects of the theory being expressed operationally by way of these probabilities. This approach allows us to free ourselves from an interpretation of the reality behind the calculations and simply focus on the results we find at face value, much like the “shut up and calculate” approach.

This thesis shall be focussed on some topics related to another concept that distinguishes quantum theory from classical mechanics: the incompatibility of observables. By this we mean the inability to jointly measure any two observables on a system arbitrarily well. This concept is far removed from the classical situation, and presents limits on what can be performed on quantum systems, with the standard examples being the inability to measure the conjugate observables position and momentum without the first observable disturbing the outcomes of the second, and similarly for the spin-1/2 observables in the x and z direction, say. It is this concept of incompatibility that led to the introduction of uncertainty relations within the theory: either as a preparation relation where we cannot prepare a state arbitrarily well in order to measure two incompatible observables, or as an error-error (error-disturbance) relation in which we cannot acquire arbitrarily accurate measurement statistics for two incompatible observables measured jointly (sequentially). Such relations impose restrictions upon the extent to which we can perform accurate measurements in many different fields including, for example, quantum estimation theory and in the measuring of gravitational waves, where the accuracy of the interferometers is restricted by the standard quantum limit, a consequence of the preparation uncertainty relation for position and momentum.

In what follows we shall consider problems in both the finite and infinite-dimensional case where the incompatibility of the observables considered plays an important part. We begin by considering an operational link between two important tools in quantum metrology: Symmetric Informationally Complete Positive Operator-Valued Measures (SIC-POVMs) and Mutually Unbiased Bases (MUBs). MUBs correspond to maximally incompatible observables on a finite-dimensional quantum system, but we present a situation in which they can be made compatible, with a SIC-POVM forming the joint observable.

Secondly, we provide an extension to the Arthurs-Kelly measurement model. This model allows for an approximate measurement of the incompatible observables position and momentum by way of indirect measurements using probes coupled to the system. The extension discussed is the preparation of the probes in a completely arbitrary state, which allows for prior coupling between the probes (a possibility not considered operationally before).

Finally, we consider uncertainty relations for incompatible observables on two-dimensional quantum systems—the so-called "qubit" systems—and try to reconcile error bounds expressed in terms of state-dependent error measures with those in terms of state-independent measures. This is only possible in the case of dichotomic (two-valued) observables defined upon qubit systems, and in doing so we provide an expression of the relations for the state-independent measures that is independent of the representation used.

In Chapter 2 we shall cover the mathematical background that is relied upon throughout the remainder of the thesis. This will focus predominantly on functional analysis and Hilbert space theory, paying particular attention to certain subclasses of bounded linear operators. We also discuss the basics of topology, finite fields and measure theory, with the latter allowing us to introduce Positive Operator-Valued Measures (POVMs), including the special case of Projection-Valued Measures (PVMs), which leads on to the spectral decomposition of (not necessarily bounded) self-adjoint linear operators. We show that certain POVMs can be constructed by *smearing* PVMs. We also show that with each POVM we can associate a PVM acting on a larger Hilbert space via Naimark's dilation theorem.

These mathematical constructions are reconciled with the concepts within quantum theory. By considering measurements and the probabilities associated with measurement outcomes, we arrive at the concepts of the state of the system, observables and measurements. In terms of complex Hilbert space language, states take the form of positive operators of unit trace, whilst observables correspond to POVMs whose domains correspond to possible measurement outcomes. What we refer to as *sharp* observables correspond in this language to PVMs, whilst more general POVMs correspond to observables in which some statistical noise is present. From this, the concept of smearing makes sense as an introduction of statistical noise to an ideal measurement, whilst the Naimark dilation allows one to form a sharp observable by working in a larger Hilbert space. Note that we do not assume the commonly assumed notion that observables correspond to self-adjoint operators defined on a Hilbert space, but rather we define observables to be given by POVMs, and in the case of sharp observables these correspond to the PVMs associated with a self-adjoint operator's spectral decomposition. Measurement models are then discussed, which highlight the idea that we measure quantum systems indirectly, and give us an idea of the noise that arises within measurements of such systems. We then consider joint and sequential measurements, which allow us to form an operational meaning of what we mean by incompatible observables. The chapter is concluded with a discussion on two of the more commonly used measures of error within quantum theory, which will be used throughout: the first is a state-dependent error based on the concept of a noise operator, whilst the second is state-independent and is based on the Wasserstein 2-distance between probability distributions.

Chapter 3 defines and investigates an operational link between SIC-POVMs and MUBs, the results of which were published in [5]. MUBs are associated with PVMs whose overlaps are fixed and correspond to observables that are maximally incompatible. However, by taking margins of a SIC-POVM corresponding to mutually orthogonal Latin squares we are able to find instances in which the eigenvectors of distinct margin POVMs are mutually unbiased, thereby deriving MUBs from a SIC-POVM. Similarly, in Hilbert spaces of prime-power dimension, by starting with a collection of MUBs we show a construction from which we can in some instances derive a SIC-POVM. These constructions highlight a connection between these two constructions different from the geometric connections presented by Appleby *et al.*, and show how these incompatible observables can be modified in order to form compatible observables. We also provide a discussion of a construction of Wootters that postulates a similar connection involving affine planes. In his paper, Wootters mentions some issues with his construction that he was aware of, and we highlight some further ways in which his suggestion was incomplete in light of our own findings. We conclude the chapter with some discussion regarding investigations undertaken in Mathematica in low dimensions to explicitly find examples of our construction working. Whilst there are examples that are shown to work, we see that such examples are scarce, and as the dimension increases, the required computational power grows rapidly.

Chapter 4 provides an extension to the Arthurs-Kelly measurement model, which was published in [8]. The extension, suggested by Di Lorenzo in a 2013 Letter, is the inclusion of initial correlations between the measurement probes. We prepare the probes in an arbitrary state, thereby allowing for the possibility of initial correlations, and from this calculate the observable that is measured on the considered system. The measured observable is shown to be covariant under translations in phase space, with approximations of the position and momentum observables as margins. This result extends what was already known for probes prepared in uncorrelated states and shows this type of observable to be the case for all probe preparations in the Arthurs-Kelly measurement model. We then proceed to refute a claim made by Di Lorenzo that this measurement scheme, when initial correlations between the probes are included, allows for a violation of a Heisenberg-like error-disturbance relation. The reason for this supposed violation is shown to be the result of an unphysical definition of disturbance. We reconcile this definition with a measure of relative statistical spread between the margins of the observable derived from measuring both probes and observables derived from measuring each probe individually. We show that the presence of initial correlations between the probes can result in the marginal observables having a smaller statistical spread than their individual probe measurement counterparts; a phenomenon that we call *focussing*. The chapter is concluded with two examples of probe preparations that can allow for focussing to occur; the second example shows that classical correlations (i.e., preparing the probes in a mixed state) are sufficient to allow focussing.

In Chapter 5 we discuss error-error relations for incompatible dichotomic qubit observables when approximated by compatible observables. This scheme has been covered in the past by Hall, and in recent years by Branciard, Yu and Oh, and Busch, Lahti and Werner amongst others. In the case of Hall and Branciard the considerations were for more general observables and were based on the idea that any discrete observable could provide an approximate joint measurement of any pair of discrete observables. We refute the validity of this claim, as their construction relies solely upon post-processing of values, and therefore is little more than relabelling the scales of a measurement device. We then consider the error measures used by Branciard and Yu and Oh, which are the state-dependent measure and a rescaled version of the state-independent measure respectively, and the optimal error bounds that they provide. Branciard's bound holds for all finite-dimensional systems and is based on considerations on Euclidean space, whilst Yu and Oh base their considerations upon vectors on the Bloch sphere, and hence is exclusive to the qubit case. The form given by Yu and Oh is parametric and relies heavily on the geometry of the Bloch sphere, and so it is reformulated in terms of operational quantities. Optimal approximating observables in terms of each bound are considered, and it is shown that with the exception of maximally incompatible observables, we are unable to find approximating observables that meet both optimal bounds. Given that we are trying to optimise different bounds, this may not come as a surprise, but since they are both used to determine what should be an optimal approximation of an observable, a lack of agreement between them suggest that at least one measure is not doing what it is purported to do. We also discuss experimental work by Ringbauer *et al.* that claims to saturate Branciard's bound. This work is shown to be a special case of an example given by Branciard, and is subject to the same shortcomings as those found with Branciard and Hall's work.

We conclude with a brief summary of the work presented in this thesis.

Chapter 2

Mathematical Background

2.1 Topology

In this subsection we will briefly give the definition of a topology, and some further definitions that will be of use later on, in particular when mentioning operator topologies. For the interested reader, a more in-depth introduction to this section is given in, for example, [51]. In this section and throughout the rest of the thesis, we will make use of the standard notation for set theory, which we shall state for reference in page 152.

Consider a set X and its power set 2^X. A collection τ of subsets of X, i.e., $\tau \subset 2^X$, is a topology if

1. $X, \emptyset \in \tau$;
2. If $O_1, O_2 \in \tau$, then $O_1 \cap O_2 \in \tau$;
3. If $\{O_i\}_{i \in I}$ is a countable set (that is, $I \subseteq \mathbb{N}$) of subsets $O_i \in \tau$, then $\bigcup_{i \in I} O_i \in \tau$.

In other words, τ contains both the set X and the empty set, and is closed under finite intersections and countable unions of its elements. The pair (X, τ) is called a topological space, and the elements of τ are called open sets. The complements of open sets are called closed sets, and so we immediately note that X and $\emptyset$ are both open and closed (they are "clopen") as they are each other's complement. Further to this, from de Morgan's laws

$$(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c, \tag{2.1}$$

we can define τ instead in terms of closed sets.

Suppose that we possess a metric space (X, d); that is, a set X with a distance function $d : X \times X \to \mathbb{R}$ called a metric, which satisfies

1. $d(x, y) \geq 0$ for all $x, y \in X$, with equality iff $x = y$;
2. $d(x, y) = d(y, x)$;
3. $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

In such an eventuality, we can provide a more concrete definition of the sets used to define a topology on X:

(i) For $x \in X$, an *open ball* of radius r, $B(x, r)$ is the set $\{y \in X | d(x, y) < r\}$;

(ii) The set $O \subset X$ is *open* if, for all $x \in O$ there exists an $r > 0$ such that $B(x, r) \subset O$. In other words, O is an open set if every element of O possesses an open ball contained entirely in O;

(iii) Let $E \subset X$. The point/element $x \in X$ is a *limit point* of E if, for all $r > 0$, $B(x, r) \cap (E \backslash \{x\}) \neq \emptyset$. In other words, x is a limit point of E if E contains points that are arbitrarily close to x.

(iv) If E contains all of its limit points, then E is *closed.*

Any such topology that can be constructed in terms of a metric is said to be *metrisable.* Suppose that a set E, that is a subset of some larger set S, is not closed, then we may find its closure $\overline{E}$ by containing its limit points:

$$\overline{E} = E \cup \{x \in S | x \text{ is a limit point of } E\}. \tag{2.2}$$

A nontrivial subset $M \subset X$ is *dense* if its closure is equal to X, i.e., $\overline{M} = X$.

Suppose that we possess a topological space (X, τ) and we consider a subset $A \subset X$, then we may provide this set with a topology τ_A by

$$\tau_A = \{A \cap O | O \in \tau\}. \tag{2.3}$$

In other words, we essentially restrict the elements of the topology τ to A. That this is indeed a topology follows quickly from the distributive law for the set operations:

$$(A \cap O_1) \cup (A \cap O_2) = A \cap (O_1 \cup O_2), \tag{2.4}$$

etc. This topology is called the topology induced on A by τ.

Consider two topological spaces (X, τ) and (Y, σ) and a function $f : X \to Y$. This function is *continuous* if the inverse image of an open set on Y is open in X. In other words, if $O \in \sigma$, then

$$f^{-1}(O) = \{x \in X | f(x) \in O\} \in \tau. \tag{2.5}$$

This concept of continuous functions will be of considerable use later. Note that this is an equivalent expression of continuity to that used predominantly in physics: given the topological spaces (X, τ) and (Y, σ), where τ and σ are metrisable topologies, the function $f : X \to Y$ is continuous iff for any $\varepsilon > 0$ and $x \in X$, there is a $\delta > 0$ such that

$$d_X(x, y) < \delta \Rightarrow d_Y(f(x), f(y)) < \varepsilon, \tag{2.6}$$

where d_X and d_Y are the metrics defined on X and Y, respectively. The equivalence of these statements follows quickly from noting that Equation (2.6) is describing the existence of open balls in these metrisable topologies.

Consider the complete[1] metrisable topological space (X, τ) with a metric d. A subset $S \subset X$ is compact if every open cover; that is, a collection $\{O_i\}_{i \in I}$ of open sets $O_i \in \tau$

[1] See Page 32

satisfying

$$S \subseteq \bigcup_{i \in I} O_i, \tag{2.7}$$

admits a finite subcover. In other words, there is a subset $\{O_1, \dots, O_n\}$ such that

$$S \subseteq \bigcup_{i=1}^{n} O_i, \tag{2.8}$$

for each open cover of S. In the case of a complete metric space, such as (X, d), a subset S of X is compact iff it is closed and bounded [47, Theorem A4]. This property leads to the following result, which we will not prove here: for a compact set S and a continuous function $f : S \to \mathbb{R}$, f obtains its maximum and minimum values in S.

Assume that we possess a countable number of topological spaces $\{X_i, \tau_i\}_{i \in I}$ with index set $I \subset \mathbb{N}$. We form the Cartesian product of these sets,

$$X := \bigtimes_{i \in I} X_i, \tag{2.9}$$

and define the projections $p_i : X \to X_i$. For example, if $X = X_1 \times X_2 = \{(x_1, x_2) | x_1 \in X_1, x_2 \in X_2\}$, then $p_1(x_1, x_2) := x_1$, etc. The topology on X, called the product topology, is the coarsest topology (that is, the one with the fewest elements) for which each projection function is continuous.

Finally, we define a topological vector space. Consider a vector space X over $\mathbb{R}$ or $\mathbb{C}$, i.e., a vector space with coefficients belonging to $\mathbb{R}$ (resp. $\mathbb{C}$). Then X is a topological vector space if it can be given a topology τ such that

1. The map $(x, y) \mapsto x + y$ is a continuous function $+ : X \times X \to X$, where $X \times X$ possesses the product topology;

2. The map $(a, x) \mapsto a \cdot x$ is a continuous function $\cdot : \mathbb{R} \times X \to X$, where $\mathbb{R}$ possesses its standard topology as given above, and $\mathbb{R} \times X$ is the product topology.

In other words, X is a vector space for which the addition and scalar multiplication operations are continuous.

2.2 Finite fields

Here we will give an overview of fields, in particular finite fields. This mathematical construct provides useful results for SIC-POVMs and MUBs (see Chapter 3), and indeed has further uses within quantum information theory. We shall only provide the results needed for a usable knowledge of the subject for our purposes, and direct the reader to [35] for the proofs of results given here and further discussion.

2.2.1 Rings and fields

We begin by considering an abelian group G, i.e., a set with an associative binary operation $+ : G \times G \to G$ satisfying $a + b = b = a$ for any $a, b \in G$, such that G is closed under $+$, has a defined identity element 0 and each element a possesses an inverse element $-a$. We now allow for a second binary operation defined on this group.

Definition 2.2.1. A *Ring* $(R, +, \cdot)$ is a set R, with two binary operations, $+$ and $\cdot$, such that

1. R is an abelian group with respect to $+$;
2. $\cdot$ is associative, so for all $a, b, c \in R$, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$;
3. The *distributive law* holds, i.e., for all $a, b, c \in R$, $a \cdot (b + c) = a \cdot b + a \cdot c$, and similarly $(b + c) \cdot a = b \cdot a + c \cdot a$.

We will denote a ring by its set, so the ring $(R, +, \cdot)$ will simply be referred to as R. In keeping with the notation of abelian groups, 0 will refer to the identity element with respect to the operation $+$, and the inverse of a with respect to $+$ will be denoted by $-a$ such that $a + (-b)$ will be written as $a - b$. For the sake of brevity, we will concatenate the notation $a \cdot b$ to ab. As a result of the distributive law, it follows that for any $a, b \in R$

$$ab = a(b + 0) = ab + a0. \tag{2.10}$$

In other words, $a0 = 0$. Likewise,

$$a0 = a(b - b) = ab + a(-b) = 0, \tag{2.11}$$

so $a(-b) = -ab$.

This definition of a ring is quite open with regards to the second binary operation, and so we classify different types of rings.

Definition 2.2.2. Consider a ring R.

1. A *ring with identity* is a ring that contains a multiplicative identity e such that $ae = ea = a$ for all $a \in R$;
2. A ring is *commutative* if $\cdot$ is commutative;
3. An *integral domain* is a commutative ring with identity ($e \neq 0$) in which $ab = 0$ implies that $a = 0$ or $b = 0$;
4. A ring is a *division ring* if the elements $R \backslash \{0\}$ form a group under $\cdot$.
5. A commutative division ring is a *field.*

In other words, a field is a ring $(F, +, \cdot)$ for which F is an abelian group under the operation $+$ with an additive identity 0, and the set $F \backslash \{0\}$ is an abelian group under the operation $\cdot$ with a multiplicative identity e (also denoted by 1). These two operations are then linked by the distribution laws, as given above.

The condition of an integral domain—that $ab = 0$ implies that $a = 0$ or $b = 0$—is alternatively expressed by saying that there are no *zero divisors.* Fields have no zero divisors as a result of the group structure inherited by the operation $\cdot$. Indeed, let a and b belong to the field F. If $ab = 0$ and $a^{-1} \neq 0$, then $b = a^{-1}0 = 0$. That is not to say that any integral domain is a field, but we have the following theorem:

Theorem 2.2.1. Every finite integral domain is a field.

Similar to the case of groups, we define a *subring* of R, S, to be a subset $S \subset R$ that is closed under the additive and multiplicative operations of R, and still forms a ring.

Definition 2.2.3. A subring J of a ring R is an *ideal* if, for all $a \in J$ and all $r \in R$, $ar \in J$ and $ra \in J$.

From this, we may define a subclass of ideals:

Definition 2.2.4. Let R be a commutative ring. Denote by (a) the smallest ideal containing a. This ideal is of the form

$$(a) = \{ra + na | r \in R \text{ and } n \in \mathbb{Z}\}. \tag{2.12}$$

If R contains a multiplicative identity, then

$$(a) = \{ra | r \in R\}. \tag{2.13}$$

If, for an ideal J, there exists an a such that $J = (a)$, then J is said to be a *principal* ideal *generated* by a.

The ideals of a ring R form subgroups of the additive group of R since J is a subring of R. Further, since R is an abelian group under $+$, it follows that, for all $r \in R$, $r+J = J+r$, and so J forms a normal subgroup of the additive group of R. Hence, an ideal J of R defines a partition of R (as an additive group) into disjoint cosets, known as *residue classes* modulo J. For a given element $a \in R$, this residue class of R modulo J is denoted by $[a] = a + J$. Since J forms an abelian group under the additive operation, it follows that $J + J = J$, and so

$$[a] + [b] = (a + J) + (b + J) = (a + b) + J = [a + b]. \tag{2.14}$$

Similarly, since J is closed under the multiplicative operation, $JJ = J$. This, along with the definition of an ideal, leads to

$$\begin{aligned} [a][b] &= (a + J)(b + J) = ab + aJ + Jb + JJ \\ &= ab + J + J + J = ab + J \\ &= [ab]. \end{aligned} \tag{2.15}$$

From Equations (2.14) and (2.15), we see that the residue classes of R modulo J are closed under the additive and multiplicative operations of R, and indeed form a ring.

Definition 2.2.5. The ring of residue classes of the ring R modulo the ideal J, which satisfies Equations (2.14) and (2.15) is called the *factor ring* of R modulo J, and is denoted by R/J.

In the case of a commutative ring with identity, we can determine which ideals can lead to factor rings that form integral domains or fields. In order to do so, we must introduce some terminology from ring theory.

Consider a commutative ring R with identity. An element $a \in R$ is a *divisor* of $b \in R$ if there exists an element $c \in R$ such that $b = ac$. A *unit* of R is a divisor of the identity, and two elements $a, b \in R$ are *associates* if there exists a unit ε of R such that $a = b\varepsilon$. If an element $c \in R$ is not a unit and its only divisors are the units of R and its associates, then c is a *prime element.* An ideal $P \neq R$ of R is a *prime ideal* if, for any $a, b \in R$, $ab \in P$ only if $a \in P$ or $b \in P$. An ideal $M \neq R$ of R is a *maximal ideal* of R if for any ideal J of R, $M \subseteq J$ implies $J = R$ or $J = M$. Finally, R is a *principal ideal domain* if it is an integral domain whose every ideal is principal; i.e. for every ideal J of R there is an element $a \in R$ such that $J = (a) = \{ra | r \in R\}$.

Theorem 2.2.2. Let R be a commutative ring with identity 1. Then

(i) An ideal M of R is a maximal ideal iff R/M is a field;

(ii) An ideal P of R is a prime ideal iff R/P is an integral domain;

(iii) Every maximal ideal of R is a prime ideal;

(iv) If R is a principle ideal domain, then $R/(c)$ is a field iff $c \in R$ is a prime element of R.

Let us now consider the set of integers, $\mathbb{Z}$. We can provide this set with the multiplicative and additive operation, in which case we see that it forms an integral domain with additive identity 0 and multiplicative identity 1. That this is not a field can easily be seen from the absence of rational numbers from $\mathbb{Z}$, and so no element besides 1 possesses a multiplicative inverse. Let us now choose an element $n \in \mathbb{Z}$ and construct its principle ideal $(n) = \{zn | z \in \mathbb{Z}\}$. We then construct the factor ring of $\mathbb{Z}$ modulo n, $\mathbb{Z}_n = \mathbb{Z}/(n)$, as the ring of residue classes $[a] = a + (n)$, where $0 \leq a < n$. In other words,

$$\mathbb{Z}_n = \{[0], [1], \ldots, [n-1]\}. \tag{2.16}$$

Theorem 2.2.3. Let p be a prime number, then the factor ring of $\mathbb{Z}$ modulo the principal ideal of p, $\mathbb{Z}_p$, is a field.

A *ring homomorphism* φ is a map from the ring R to another ring S, such that for any $r, s, \in R$

$$\varphi(r+s) = \varphi(r) + \varphi(s), \qquad \varphi(rs) = \varphi(r)\varphi(s). \tag{2.17}$$

In other words, φ preserves the structure of both the additive and multiplicative operations of R. Hence, φ induces a homomorphism from the additive group of R to the additive group of S. With this in mind, we define the *kernel* of φ, $\ker\varphi$, as the set

$$\ker\varphi = \{a \in R | \varphi(a) = 0 \in S\}; \tag{2.18}$$

that is, $\ker\varphi$ is the set of elements mapped to the additive identity by φ. Since there is no guarantee that S, or indeed R, contains a multiplicative identity, we do not characterise a set of elements mapped to it by φ. Since this is an example of a group homomorphism, $\ker\varphi$ forms a (normal) subgroup of the additive group of R. The kernel provides an important role when considering homomorphisms, as is shown in the next theorem, known as the *first isomorphism theorem for rings* (note that there is an equivalent theorem for groups).

Theorem 2.2.4. Let R, S be rings and $\varphi : R \to S$ a ring homomorphism. Then the image of φ, im φ is a subring of S, the kernel of φ is an ideal of R, and φ induces the isomorphism:

$$R/\ker\varphi \cong \text{im } \varphi. \tag{2.19}$$

By forming a map $\varphi : R \to S$ from a ring R to a set S, we are capable of providing S with a structure that it otherwise does not possess. Suppose that φ is a one-to-one mapping from the ring R onto the set S, i.e., $\varphi(a) = \varphi(b)$ iff $a = b$, and for every $b \in S$ there exists an $a \in R$ such that $b = \varphi(a)$. Then, by Equation (2.17), φ provides S with the structure of a ring, and in doing so makes φ a ring isomorphism: let $s_1, s_2 \in S$, and choose $r_1, r_2 \in R$ uniquely via $s_1 = \varphi(r_1)$ and $s_2 = \varphi(r_2)$. Then, we let $s_1+s_2 = \varphi(r_1+r_2)$ and $s_1s_2 = \varphi(r_1r_2)$, and in doing so provide S with all of the necessary structure for it to be a ring and for φ to be a ring homomorphism. We therefore describe this structure on S as being *induced by φ*.

By inducing a structure via the map φ, we can form a more intuitive and convenient representation of $\mathbb{Z}/(p)$.

Definition 2.2.6. Let p be a prime number, $\mathbb{F}_p$ the set $\{0, 1, \ldots, p-1\}$, and $\varphi : \mathbb{Z}/(p) \to \mathbb{F}_p$ the map defined by $\varphi([a]) = a$ for all $a = 0, 1, \ldots, p-1$. The set $\mathbb{F}_p$, with the structure induced by the map φ, forms a finite field, called the *Galois field of order p*.

The map $\varphi : \mathbb{Z}/(p) \to \mathbb{F}_p$ is clearly one-to-one, and hence a ring isomorphism with $\varphi([a] + [b]) = \varphi([a]) + \varphi([b])$ and $\varphi([a][b]) = \varphi([a])\varphi([b])$. The structure of $\mathbb{F}_p$ is therefore identical to that of $\mathbb{Z}/(p)$, with the additive and multiplicative identities replaced by 0 and 1, respectively. The difference is that $\mathbb{F}_p$ has the benefit of its operations corresponding to regular arithmetic modulo p.

Definition 2.2.7. Consider the ring R and suppose that there exists a positive n such that, for all $r \in R$, $nr = 0$. The least such n is called the *characteristic* of R, and R is said to have positive characteristic n. If no such n exists, then R is said to have characteristic 0.

If we look at $\mathbb{F}_p \cong \mathbb{Z}/(p)$, we see that since this field is finite, there must exist such a finite integer n for which $na = 0$ for all $a \in \mathbb{F}_p$. Given that the elements of $\mathbb{F}_p$ are less than the prime number p, and so are coprime to it, the smallest number for which $na = 0$ mod p, i.e., $na = mp$ for some integer m, for all $a \in \mathbb{F}_p$ is p itself. Hence, $\mathbb{F}_p$ is a finite field of characteristic p. By contrast, the ring $\mathbb{Z}$ of integers is of characteristic 0.

Theorem 2.2.5. A ring $R \neq \{0\}$ of positive characteristic with an identity and no zero divisors must have prime characteristic.

Corollary 2.2.6. Every finite field has prime characteristic.

The following theorem is of use in the next section:

Theorem 2.2.7. Let R be a commutative ring of characteristic p. Then, for any $a, b \in R$ and $n \in \mathbb{N}$,

$$(a+b)^{p^n} = a^{p^n} + b^{p^n}, \qquad \text{and} \qquad (a-b)^{p^n} = a^{p^n} - b^{p^n}. \tag{2.20}$$

2.2.2 Polynomials over rings

Before we continue with fields, we want to briefly discuss polynomials over rings, as these provide assistance when finding certain results about fields, in particular regarding their dimensionality.

Consider a ring R. A *polynomial* over R is an expression

$$f(x) = \sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + \cdots + a_n x^n, \tag{2.21}$$

with n a nonnegative integer and the *coefficients* $a_i \in R$ for $0 \leq i \leq n$, whilst the *indeterminate* over R, x, is not an element of R. Note that, for brevity's sake, we will at times reduce $f(x)$ to f. Two polynomials over R

$$f(x) = \sum_{i=0}^{n} a_i x^i, \qquad g(x) = \sum_{i=0}^{n} b_i x^i, \tag{2.22}$$

are equal iff $a_i = b_i$ for all $0 \leq i \leq n$. The sum of two polynomials over R, as above, is given by

$$f(x) + g(x) = \sum_{i=0}^{n} (a_i + b_i) x^i, \tag{2.23}$$

whilst the product of two polynomials over R

$$f(x) = \sum_{i=0}^{n} a_i x^i, \qquad g(x) = \sum_{j=0}^{m} b_j x^j, \tag{2.24}$$

is given by

$$f(x)g(x) = \sum_{k=0}^{m+n} c_k x^k, \quad \text{where} \quad c_k = \sum_{i+j=k} a_i b_j. \tag{2.25}$$

Since the a_i and b_j are elements of R, the sum $a_i + b_i \in R$ for all nonnegative integers i, and similarly $a_i b_j \in R$ for all nonnegative integers i, j, hence $c_k \in R$ for all nonnegative integers k. Hence, the polynomials over R themselves form a ring.

Definition 2.2.8. The ring of polynomials over R with the operations defined in Equations (2.23) and (2.25) is called the *polynomial ring* over R, and is denoted by $R[x]$.

Since $R[x]$ is a ring, it must contain an additive identity, known as the zero polynomial 0, which has coefficients $a_i = 0$ for all i.

Definition 2.2.9. Let $f(x) = \sum_{i=0}^{n} a_i x^i \in R[x]$, with $a_n \neq 0$. The coefficient a_n is called the *leading coefficient* of $f(x)$, and n is the *degree* of $f(x)$, denoted by $n = \deg(f(x)) = \deg(f)$. The coefficient a_0 is called the *constant term.*

By convention, $\deg(0) = -\infty$, and polynomials with degrees ≤ 0 are called *constant polynomials.* If $1 \in R$, then polynomials $f(x) \in R[x]$ with leading coefficients equal to 1 are called *monic polynomials.*

We now state some results regarding the degree of the sum and product of two polynomials:

Theorem 2.2.8. Consider two polynomials $f, g \in R[x]$. Then

$$\begin{aligned} \deg(f+g) &\leq \max(\deg(f), \deg(g)), \\ \deg(fg) &\leq \deg(f) + \deg(g). \end{aligned} \tag{2.26}$$

In the case of R being an integral domain,

$$\deg(fg) = \deg(f) + \deg(g). \tag{2.27}$$

The first of these results is fairly intuitive, with the inequality arising in the case when f and g are of the same degree, and the leading coefficients are the additive inverse of each other. Similarly, the inequality in the second result is needed for the case where the product of the leading coefficients of f and g equals zero. In the third case, where R is an integral domain and so has no zero divisors, the product of the leading coefficients must be nonzero, and so we have equality in Equation (2.27).

We may identify the constant polynomials of $R[x]$ with the elements of R, and so R may be perceived as a subring of $R[x]$. This suggests that $R[x]$ may inherit some features from R, and indeed this is the case.

Theorem 2.2.9. Let R be a ring. Then

(i) $R[x]$ is commutative iff R is commutative;

(ii) $R[x]$ is a ring with identity iff R has an identity;

(iii) $R[x]$ is an integral domain iff R is an integral domain.

We now work in particular with the ring of polynomials over a field F, $F[x]$. With this in mind, we can apply the concept of division to the elements of $F[x]$: $g \in F[x]$ *divides* the polynomial $f \in F[x]$ if there exists a polynomial $h \in F[x]$ such that $f = gh$. In this case, g is a *divisor* of f, or alternatively, f is a *multiple* of g and so f is *divisible* by g. The units of $F[x]$ are the divisors of the constant polynomial 1, so the units of $F[x]$ are the nonzero constant polynomials, i.e., the elements of F, which are all invertible.

Theorem 2.2.10. Let $g \in F[x]$ be a nonzero polynomial. Then, for any $f \in F[x]$ there exist unique polynomials $q, r \in F[x]$ such that

$$f = qg + r, \quad \text{where} \deg(r) < \deg(g). \tag{2.28}$$

This is called the *division algorithm.*

Theorem 2.2.11. $F[x]$ is a principal ideal domain, and for every ideal $J \neq (0)$ of $F[x]$ there exists a uniquely determined monic polynomial $g \in F[x]$ such that $J = (g)$.

We now classify an important type of polynomial.

Definition 2.2.10. A polynomial $f \in F[x]$ is *irreducible over* F (or *irreducible in* $F[x]$) if f has positive degree and $f = bc$, with $b, c \in F[x]$, implies that b or c is a constant polynomial. Any polynomial in $F[x]$ which is not irreducible is therefore *reducible.*

Theorem 2.2.12. For $f \in F[x]$, the residue class ring $F[x]/(f)$ is a field iff f is irreducible over F.

As we would expect, if we take the indeterminate x in $f(x) = \sum_{i=0}^{n} a_i x^i \in F[x]$ and replace it with an element $b \in F$, then the result, $f(b) = \sum_{i=0}^{n} a_i b^i$ will again be an element of F. From this we consider a particular subset of elements of F:

Definition 2.2.11. Consider the polynomial $f(x) \in F[x]$. An element b of F is a *root* of f if $f(b) = 0$.

Theorem 2.2.13. An element $b \in F$ is a root of the polynomial $f \in F[x]$ iff $x - b$ is a divisor of $f(x)$.

Suppose that $b \in F$ is a root of the polynomial $f \in F[x]$. If the function $(x - b)^k$, where k is a positive integer, divides f, but $(x - b)^{k+1}$ does not, then k is the *multiplicity* of b. If $k = 1$, then the root is *simple*, but if $k \geq 2$, then it is a *multiple root*.

Consider a polynomial $f(x) \in F[x]$ of the form $f(x) = a_0 + a_1 x + \cdots + a_n x^n$. Its *derivative* $f'(x) \in F[x]$ is given by $f'(x) = a_1 + 2a_2 x + \cdots + (n-1)x^{n-1}$.

Theorem 2.2.14. An element $b \in F$ is a multiple root of $f(x) \in F[x]$ iff it is a root of both $f(x)$ and its derivative $f'(x)$.

2.2.3 Extensions of fields

Let us consider the field F, and a subset $K \subset F$. If K itself has the structure of a field, then K is a *subfield* of F, and F is an *extension field* of K. In the case that $K \neq F$, K is a *proper subfield* of F.

Suppose that K is a subfield of the Galois field $\mathbb{F}_p$, where p is, by necessity, a prime number. Since K is itself a field, it contains the additive and multiplicative identities, 0 and 1, and so by the closure of addition has to contain the remaining elements of $\mathbb{F}_p$. In other words, the Galois fields $\mathbb{F}_p$ for any prime p cannot contain any proper subfields.

Definition 2.2.12. A field containing no proper subfields is called a *prime field*.

By the above argument, $\mathbb{F}_p$ is prime, but that does not exhaust the set of prime fields. For example, $\mathbb{Q}$ is an infinite prime field.

Suppose now that we possess a collection of subfields of a field F. The intersection of these subfields will again be a subfield. We now extend this to the intersection of *all* subfields of F, thereby producing the *prime subfield* of F. This is the smallest subfield in F, since all subfields of F have already been intersected in order to produce it, and anything smaller implies that it wasn't included in the intersection. Therefore, the prime subfield of F is immediately a prime field.

Theorem 2.2.15. The prime subfield of a field F is isomorphic to $\mathbb{F}_p$ or $\mathbb{Q}$, depending on whether F has characteristic p or 0.

Definition 2.2.13. Consider a field F, a subfield K and a subset M of F. The field $K(M)$ is defined as the intersection of all subfields of F containing both K and M, and hence is the smallest such subfield. This subfield is called the extension field of K obtained by *adjoining* the elements in M. In the case of a finite set, i.e., $M = \{\theta_1, \ldots, \theta_n\}$, then $K(M) = K(\theta_1, \ldots, \theta_n)$, and similarly in the case of a single element $\theta \in F$, then $L = K(\theta)$ is said to be a *simple extension* of K, and θ is called a *defining element* of L over K.

Definition 2.2.14. Let K be a subfield of F, and $\theta \in F$. If θ is a root of a nontrivial polynomial $f \in K[x]$, then θ is *algebraic* over K. An extension L of K is *algebraic* over K, or an *algebraic extension* of K, if every element of L is algebraic over K.

Suppose that the element $\theta \in F$ is algebraic over K, and consider the set of polynomials $f \in K[x]$ satisfying $f(\theta) = 0$. Since, for any $h \in K[x]$, $h(\theta)f(\theta) = f(\theta)h(\theta) = 0$, it follows that this set is an ideal, denoted by $J_\theta := \{f \in K[x] | f(\theta) = 0\}$. By Theorem 2.2.11, $K[x]$ is a principal ideal domain, and so there exists a unique monic polynomial g such that $J_\theta = (g)$. The polynomial g is irreducible in K: firstly, because it contains θ as a root, it is of positive degree; secondly, if $g = h_1 h_2$ with $1 \leq \deg(h_i) < \deg(g)$ with $i = 1, 2$, then $g(\theta) = h_1(\theta)h_2(\theta) = 0$ and so either h_1 or h_2 belongs to J_θ and must therefore be divisible by g, which contradicts the preceding argument.

Definition 2.2.15. Let $\theta \in F$ be algebraic over K. The uniquely determined monic polynomial $g \in K[x]$ generating the ideal $J_\theta = \{f \in K[x] | f(\theta) = 0\}$ of $K[x]$ is called the *minimal polynomial* of θ over $K[x]$. The *degree* of θ is equal to the degree of g.

Theorem 2.2.16. Suppose $\theta \in F$ is algebraic over K. Then its minimal polynomial g over K satisfies the following properties:

(i) g is irreducible in $K[x]$;

(ii) For $f \in K[x]$, $f(\theta) = 0$ iff g divides f;

(iii) g is the monic polynomial in $K[x]$ of least degree having θ as a root.

Suppose that L is the extension field of K, then L may be considered as a vector space over K: Since L is itself a field, the elements form an abelian group under addition. Further to this, "scalar" multiplication is allowed, i.e., for any $r \in K$ and $\alpha \in L$, $r\alpha \in L$, since it is just multiplication of two elements of L. Similarly, $r(\alpha + \beta) = r\alpha + r\beta$, $(r + s)\alpha = r\alpha + s\alpha$ by the requirement of distributivity, $(rs)\alpha = r(s\alpha)$ and $1\alpha = \alpha$ for $r, s \in K$ and $\alpha, \beta \in L$

Definition 2.2.16. Let L be an extension field of K. When treated as a vector space over K, if L is finite dimensional, then it is a *finite extension* of K. The *degree* of L over K, denoted by $[L : K]$, is then the dimension of the vector space L over K.

Theorem 2.2.17. If L is a finite extension of K and M is a finite extension of L, then M is a finite extension of K satisfying

$$[M : K] = [M : L][L : K]. \tag{2.29}$$

Theorem 2.2.18. Every finite extension of K is algebraic over K.

Theorem 2.2.19. Let $\theta \in F$ be algebraic of degree n over K, and let g be the minimal polynomial of θ over K. Then

(i) $K(\theta)$ is isomorphic to $K[x]/(g)$;

(ii) $[K(\theta) : K] = n$ and $\{1, \theta, \ldots, \theta^{n-1}\}$ forms a basis of $K(\theta)$ over K;

(iii) Every $\alpha \in K(\theta)$ is algebraic over K and its degree over K is a divisor of n.

From this result, we have that any element of the simple algebraic extension $K(\theta)$ can be expressed in the form $a_0 + a_1\theta + \dots a_{n-1}\theta^{n-1}$ with the coefficients $a_i \in K$.

Note that in Theorem 2.2.19 we make the prior assumption of the existence of a larger field F containing both θ and K. However, we wish to remove this initial assumption.

Theorem 2.2.20. Let f be an irreducible polynomial in $K[x]$. Then there exists a simple algebraic extension of K with a root of f as a defining element.

Theorem 2.2.21. Let α and β be two roots of the polynomial $f \in K[x]$ that is irreducible in K. The simple extensions $K(\alpha)$ and $K(\beta)$ are isomorphic under an isomorphism mapping α to β and keeping the elements of K fixed.

We shall now define the extension field containing all roots of a polynomial.

Definition 2.2.17. Let $f \in K[x]$ be of positive degree and F an extension field of K. The polynomial f is said to *split in* F if it can be expressed as a product of linear factors in $F[x]$, i.e., if there exist $\alpha_1, \alpha_2, \dots, \alpha_n \in F$ such that

$$f(x) = a(x - \alpha_1)(x - \alpha_2)\dots(x - \alpha_n), \tag{2.30}$$

with a being the leading coefficient of f. The field F is a *splitting field* of f over K if it splits f in F and $F = K(\alpha_1, \alpha_2, \dots, \alpha_n)$.

By use of Theorems 2.2.20 and 2.2.21 we reach the following theorem:

Theorem 2.2.22. If K is a field and f is any polynomial of positive degree in $K[x]$, there there exists a splitting field of f over K. Any two splitting fields of f over K are isomorphic under an isomorphism which keeps the elements of K fixed and maps roots of f to each other.

2.2.4 Finite fields

At this point, we wish to restrict ourselves to finite fields, as they serve the greatest purpose to us.

Lemma 2.2.23. Let F be a finite field containing a subfield K with q elements. Then F has q^m elements, with $m = [F : K]$.

Theorem 2.2.24. Let F be a finite field with prime subfield F'. In which case, F has p^n elements, where p is the characteristic of the field, and n is the degree of F over its prime subfield, i.e., $n = [F : F']$.

If there exists a polynomial $f \in \mathbb{F}_p[x]$ of degree n, then we may create an extended field with p^n elements. However, we have not proven that this is possible for all primes p and natural numbers n, and that is our next goal. We begin with the following preliminary result:

Theorem 2.2.25. If F is a finite field with q elements, then every $a \in F$ satisfies $a^q = a$.

Lemma 2.2.26. If F is a finite field with q elements, and K is a subfield of F, then the polynomial $x^q - x \in K[x]$ factors in $F[x]$ as

$$x^q - x = \prod_{a \in F}(x - a), \tag{2.31}$$

and F is a splitting field of $x^q - x$ over K.

Theorem 2.2.27. For every prime p and every $n \in \mathbb{N}$ there exists a finite field with p^n elements. Any field with $p^n = q$ elements is isomorphic to the splitting field of $x^q - x$ over $\mathbb{F}_p$.

We may therefore consider there to be *one* field with $q = p^n$ elements, which is guaranteed to exist, and that is the Galois field $\mathbb{F}_q$.

We now consider a prime field $\mathbb{F}_p$ and the extension to $\mathbb{F}_q$ with $q = p^n$ as vector spaces over $\mathbb{F}_p$. With this in mind, we define the following map from $\mathbb{F}_q$ to $\mathbb{F}_p$.

Definition 2.2.18. For $\alpha \in \mathbb{F}_q$ with $q = p^n$, the *absolute trace* of α over the prime subfield $\mathbb{F}_p$ is given by

$$\mathrm{Tr}(\alpha) = \alpha + \alpha^p + \cdots + \alpha^{p^{n-1}}. \tag{2.32}$$

We can see that, for any $a \in \mathbb{F}_q$, $\mathrm{Tr}(a) \in \mathbb{F}_p$: By making use of Theorem 2.2.7 inductively, we have

$$\begin{aligned}(\mathrm{Tr}(a))^p &= (a + a^p + \cdots + a^{p^{n-1}})^p = (a + (a^p + \cdots + a^{p^{n-1}}))^p \\ &= a^p + (a^p + \ldots a^{p^{n-1}})^p = a^p + a^{p^2} + \ldots a^{p^n} \\ &= a^p + a^{p^2} + \ldots a = \mathrm{Tr}(a),\end{aligned} \tag{2.33}$$

and Equation (2.33) can have no more than p solutions.

Since $(\mathrm{Tr}(a))^p = \mathrm{Tr}(a)$, it follows that $\mathrm{Tr}(a) \in \mathbb{F}_p$ by Theorem 2.2.25. Similarly, we can show that Tr is a linear map from the vector spaces $\mathbb{F}_q$ to $\mathbb{F}_p$:

Theorem 2.2.28. Let Tr be the absolute trace from the finite field $\mathbb{F}_q$, with $q = p^n$ to its prime field $\mathbb{F}_p$. Then, for any $c, d \in \mathbb{F}_p$ and $\alpha, \beta \in \mathbb{F}_q$,

$$\mathrm{Tr}(c\alpha + d\beta) = c\mathrm{Tr}(\alpha) + d\mathrm{Tr}(\beta). \tag{2.34}$$

2.2.5 Characters of finite fields

This section deals with a concept that is also heavily used when discussing group theory. Given that a field forms an abelian group with regard to one of its binary operations, usually denoted by "+", this extension is natural.

Definition 2.2.19. For a finite abelian group G, a *character* χ is a homomorphism

$$\chi : G \to \mathbb{T} = \{z \in \mathbb{C} \mid |z| = 1\}. \tag{2.35}$$

Because χ is a homomorphism, it is required that

$$\chi(g)\chi(h) = \chi(gh), \quad g, h \in G, \tag{2.36}$$

and so it follows that

$$\chi(e) = \chi(ee) = \chi(e)^2, \quad \therefore\ \chi(e) = 1, \tag{2.37}$$

where e is the identity element of G. Further to this, because G is a finite group, it follows that $g^{|G|} = e$ for all elements $g \in G$, and so

$$\chi(g)^{|G|} = \chi\big(g^{|G|}\big) = \chi(e) = 1 \quad \forall\, g \in G. \tag{2.38}$$

In other words, the character χ of G is a complex $|G|^{\text{th}}$ root of unity for all elements of G.

If we consider Equations (2.36) and (2.37), along with the fact that $g \in G$ iff $g^{-1} \in G$, we see that

$$1 = \chi(e) = \chi(gg^{-1}) = \chi(g)\chi(g^{-1}), \quad \therefore\ \chi(g^{-1}) = \chi(g)^{-1} = \overline{\chi(g)} \quad \forall\, g \in G. \tag{2.39}$$

For any finite group G, there exists at least one character, called the trivial character χ_0, for which $\chi_0(g) = 1$ for all $g \in G$. Assuming there exists more than one character of G, denoted by $\chi_1, \chi_2, \ldots, \chi_n$, we can form the product character $\chi_1\chi_2 \ldots \chi_n$ on G via

$$\chi_1\chi_2 \ldots \chi_n(g) = \chi_1(g)\chi_2(g) \ldots \chi_n(g) \quad \forall\, g \in G. \tag{2.40}$$

With these properties, it follows that the set $G^\wedge$ of characters of the finite abelian group G is itself a group. Given that, for any $g \in G$ and $\chi_i \in G^\wedge$, $\chi_i(g)$ is a $|G|^{\text{th}}$ root of unity, it follows that $G^\wedge$ must be a finite group. Furthermore, since the characters $\chi_i \in G^\wedge$ map elements of G to values in $\mathbb{C}$, it follows that $G^\wedge$ must be abelian; indeed,

$$\chi_i\chi_j(g) = \chi_i(g)\chi_j(g) = \chi_j(g)\chi_i(g) = \chi_j\chi_i(g), \quad \forall\, \chi_i, \chi_j \in G^\wedge,\ g \in G. \tag{2.41}$$

In other words, the set of characters of the finite abelian group G, $G^\wedge$, is itself a finite abelian group. The extent to which these structures are similar is increased by the following theorems:

Theorem 2.2.29. Consider the finite abelian group G and the finite abelian group $G^\wedge$ of its characters.

a) If $\chi \in G^\wedge$ is a nontrivial character of G, then

$$\sum_{g \in G} \chi(g) = 0; \tag{2.42}$$

b) If $g \in G$ and $g \neq e$, then

$$\sum_{\chi \in G^\wedge} \chi(g) = 0. \tag{2.43}$$

Theorem 2.2.30. Let G be a finite abelian group, and $G^\wedge$ its associated group of characters. Then $|G| = |G^\wedge|$.

By making use of these two theorems, we can now define a function

$$\langle \cdot, \cdot \rangle : G^\wedge \times G^\wedge \to \mathbb{C}, \tag{2.44}$$

via

$$\langle \chi, \psi \rangle = \frac{1}{|G|} \sum_{g \in G} \chi(g)\overline{\psi(g)}, \tag{2.45}$$

which can be extended to an inner product on the complex vector space spanned by $G^{\wedge}$. Note that, unlike the inner product $\langle\cdot|\cdot\rangle$ used throughout the rest of this thesis, this inner product is linear in the first argument. This is the standard used within the field of group theory (amongst others), and so we shall obey it for this particular inner product.

Using this new inner product we may show the orthonormality of the characters of G; that is, for different characters $\psi, \chi \in G$, $\langle\chi, \psi\rangle = 0$ and $\langle\psi, \psi\rangle = \langle\chi, \chi\rangle = 1$:

$$\begin{aligned}\langle\chi, \psi\rangle &= \frac{1}{|G|}\sum_{g\in G}\chi(g)\overline{\psi(g)} = \frac{1}{|G|}\sum_{g\in G}(\chi\psi^{-1})(g)\\ &= \frac{1}{|G|}\sum_{g\in G}\mu(g),\end{aligned} \tag{2.46}$$

where $\mu = \chi\psi^{-1}$. From part a) of Theorem 2.2.29, it follows that this sum is equal to zero unless μ is the trivial character. In other words,

$$\langle\chi, \psi\rangle = \begin{cases} 1 & \text{if } \chi = \psi, \\ 0 & \text{if } \chi \neq \psi. \end{cases} \tag{2.47}$$

Similarly, we also have

$$\frac{1}{|G|}\sum_{\chi\in G^{\wedge}}\chi(g)\overline{\chi(h)} = \frac{1}{|G|}\sum_{\chi\in G^{\wedge}}\hat{g}(\chi)\overline{\hat{h}(\chi)} = \frac{1}{|G|}\sum_{\chi\in G^{\wedge}}\widehat{gh^{-1}}(\chi), \tag{2.48}$$

where we define the function $\hat{g}(\chi) := \chi(g)$ and similarly for $\hat{h}$. Again this sum is equal to zero unless $gh^{-1} = e$, i.e.,

$$\frac{1}{|G|}\sum_{\chi\in G^{\wedge}}\chi(g)\overline{\chi(h)} = \begin{cases} 1 & \text{if } g = h, \\ 0 & \text{if } g \neq h. \end{cases} \tag{2.49}$$

We can categorise the characters of a finite field into several different types. In particular, we shall now consider additive and multiplicative characters, and Gaussian sums, which are a class of characters used later that combine the previous two types.

2.2.5.1 Additive characters

Let us consider the finite field $\mathbb{F}_q$, with $q = p^n$. In particular, consider the additive group of the finite field $\mathbb{F}_q$ with characteristic p, i.e., for all $a \in \mathbb{F}_q$, $pa = 0$, where 0 is the additive identity of the group. The prime field of $\mathbb{F}_q$ is $\mathbb{F}_p \cong \mathbb{Z}/(p)$. Let $\mathrm{Tr} : \mathbb{F}_q \to \mathbb{F}_p$ be the absolute trace function given in Definition 2.2.18

$$\mathrm{Tr} : a \mapsto a + a^p + a^{p^2} + \cdots + a^{p^{n-1}} \in \mathbb{F}_p. \tag{2.50}$$

Then the function

$$\chi_1(c) = e^{2\pi i \mathrm{Tr}(c)/p}, \quad c \in \mathbb{F}_q, \tag{2.51}$$

is a character of the additive group. Since the trace satisfies $\mathrm{Tr}(c_1 + c_2) = \mathrm{Tr}(c_1) + \mathrm{Tr}(c_2)$, it follows that $\chi(c_1+c_2) = \chi(c_1)\chi(c_2)$, as we would expect with χ_1 being a homomorphism from the additive group of $\mathbb{F}_q$ to $\mathbb{C}$. The character χ_1 is the *canonical additive character* of $\mathbb{F}_q$.

Theorem 2.2.31. All additive characters of the additive group of $\mathbb{F}_q$ are of the form $\chi_b(c) = \chi_1(bc)$, with $b, c \in \mathbb{F}_q$.

2.2.5.2 Multiplicative characters

We now change our focus to the multiplicative group of $\mathbb{F}_q$, $\mathbb{F}_q^* = \mathbb{F}_q \backslash \{0\}$.

Theorem 2.2.32. $\mathbb{F}_q^*$ is a cyclic group.

Definition 2.2.20. A generator of $\mathbb{F}_q^*$ is called a *primitive element* of $\mathbb{F}_q$.

Since $\mathbb{F}_q^*$ is a cyclic group of order $q-1$, we obtain the following theorem:

Theorem 2.2.33. Let g be a fixed primitive element of $\mathbb{F}_q$. For each $j = 0, 1, \ldots, q-2$, the function ψ_j, where

$$\psi_j(g^k) = e^{2\pi ijk/(q-1)}, \quad k = 0, 1, \ldots, q-2, \tag{2.52}$$

defines a multiplicative character of $\mathbb{F}_q$, and every multiplicative character of $\mathbb{F}_q$ is obtained in this way.

By making use of the orthogonality conditions of characters given in Equations (2.47) and (2.49), we have the following properties:

Firstly, for additive characters,

(i) For two characters, χ_a and χ_b, of $\mathbb{F}_q$,

$$\sum_{c \in \mathbb{F}_q} \chi_a(c)\overline{\chi_b(c)} = \begin{cases} q & \text{if } a = b, \\ 0 & \text{if } a \neq b; \end{cases} \tag{2.53}$$

(ii) If we set $b = 0$ in (i), i.e., the second character is the trivial character, then

$$\sum_{c \in \mathbb{F}_q} \chi_a(c) = 0 \quad \text{if } a \neq 0; \tag{2.54}$$

(iii) If $c, d \in \mathbb{F}_q$, then

$$\sum_{b \in \mathbb{F}_q} \chi_b(c)\overline{\chi_b(d)} = \begin{cases} q & \text{if } c = d, \\ 0 & \text{if } c \neq d. \end{cases} \tag{2.55}$$

Secondly, for multiplicative characters,

(iv) For two multiplicative characters, ψ_j and ψ_k, of $\mathbb{F}_q$,

$$\sum_{c \in \mathbb{F}_q^*} \psi_j(c)\overline{\psi_k(c)} = \begin{cases} q-1 & \text{if } j = k, \\ 0 & \text{if } j \neq k; \end{cases} \tag{2.56}$$

(v) If we let $k = 0$ in (iv), then

$$\sum_{c \in \mathbb{F}_q^*} \psi_j(c) = 0 \quad \text{if } j \neq 0; \tag{2.57}$$

(vi) If $c, d \in \mathbb{F}_q^*$, then

$$\sum_{j \in \mathbb{F}_q^*} \psi_j(c)\overline{\psi_j(d)} = \begin{cases} q-1 & \text{if } c = d, \\ 0 & \text{if } c \neq d. \end{cases} \tag{2.58}$$

2.2.5.3 Gaussian sums

Using what we have found in the past two sections, if we consider the additive and multiplicative characters, χ and ψ, respectively, of $\mathbb{F}_q$, then we define the Gaussian sum $G(\psi, \chi)$ by

$$G(\psi, \chi) = \sum_{c \in \mathbb{F}_q^*} \psi(c)\chi(c). \tag{2.59}$$

Theorem 2.2.34. Let ψ be a multiplicative character and χ an additive character of $\mathbb{F}_q$. Then the Gaussian sum $G(\psi, \chi)$ satisfies

$$G(\psi, \chi) = \begin{cases} q-1 & \text{for } \psi = \psi_0, \chi = \chi_0, \\ 0 & \text{for } \psi \neq \psi_0, \chi = \chi_0, \\ -1 & \text{for } \psi = \psi_0, \chi \neq \chi_0, \end{cases} \tag{2.60}$$

where ψ_0 and χ_0 are the trivial multiplicative and additive characters, respectively. If, however, $\psi \neq \psi_0$ and $\chi \neq \chi_0$, then

$$|G(\psi, \chi)| = \sqrt{q}. \tag{2.61}$$

2.3 Functional Analysis: Hilbert spaces and operators

2.3.1 Hilbert and Banach spaces

The basic structure that we will rely on throughout is the Hilbert space. Consider a vector space $\mathcal{H}$ over the complex numbers with an inner product $\langle \cdot | \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ that satisfies the following properties for any vectors $\psi, \varphi, \xi \in \mathcal{H}$ and $c \in \mathbb{C}$:

1. $\langle \xi | \psi + c\varphi \rangle = \langle \xi | \psi \rangle + c \langle \xi | \varphi \rangle$;
2. $\langle \psi | \varphi \rangle = \overline{\langle \varphi | \psi \rangle}$;
3. $\langle \psi | \psi \rangle > 0$ for all $\psi \neq 0$.

In other words, the inner product is a positive-semidefinite symmetric sesquilinear form. A vector space with such an inner product is called an inner product space. Note that we use the physicist's convention of linearity in the *second* argument. It is also follows quickly that $\langle 0 | \psi \rangle = \langle 0 | \psi \rangle = 0$ for any vector ψ. For any two vectors ψ and φ that vary only by a complex phase, i.e., $\varphi = e^{i\theta}\psi$, $\theta \in [0, 2\pi)$, then $\langle\varphi|\varphi\rangle = \langle\psi|\psi\rangle$. We may define the equivalence relation

$$\psi \sim \varphi \Leftrightarrow \exists\, \theta \in [0, 2\pi), \varphi = e^{i\theta}\psi, \tag{2.62}$$

and hence consider the space $\mathcal{H}/\sim$. The equivalence classes $[\psi] = \{e^{i\theta}\psi|\theta \in [0, 2\pi)\}$, $\psi \in \mathcal{H}$, are called *rays* and the space $\mathcal{H}/\sim$ is called the *projective space* of $\mathcal{H}$.

An important result for inner product spaces is the Cauchy-Schwarz inequality:

$$|\langle\varphi|\psi\rangle|^2 \leq \langle\varphi|\varphi\rangle\, \langle\psi|\psi\rangle\,, \tag{2.63}$$

with equality iff the vectors ψ, φ are linearly dependent, i.e., if $\varphi = \lambda\psi$ for some $\lambda \in \mathbb{C}$. The proof of Equation (2.63) can be found, e.g., in [1].

Two inner product spaces $\mathcal{H}$ and $\mathcal{H}'$ are *isomorphic* if there exists a bijective linear map $U : \mathcal{H} \to \mathcal{H}'$ satisfying

$$\langle U\varphi|U\psi\rangle = \langle\varphi|\psi\rangle\,, \tag{2.64}$$

for any $\varphi, \psi \in \mathcal{H}$, i.e., the map U is isometric. Note that the inner product on the right hand side is the inner product defined on $\mathcal{H}'$ and so may not be the same inner product as on the left hand side. The map U is an isomorphism from $\mathcal{H}$ to $\mathcal{H}'$.

Two vectors, $\psi, \varphi \in \mathcal{H}$, are orthogonal, denoted by $\psi \perp \varphi$, if their inner product is equal to zero, i.e., $\langle\psi|\varphi\rangle = 0$. If, for any $d \in \mathbb{N}$, an inner product space $\mathcal{H}$ contains a set of d mutually orthogonal vectors, i.e., any pair of vectors in this set are orthogonal, then $\mathcal{H}$ is an infinite dimensional inner product space. Otherwise, $\mathcal{H}$ is a finite dimensional inner product space, with the largest such d being the dimension of the space. Any d-dimensional inner product space is isomorphic to $\mathbb{C}^d$, the space of d-tuples $\{(x_1, x_2, \ldots, x_d)|x_i \in \mathbb{C}\}$, whose inner product between any two vectors $x = (x_1, \ldots, x_d)$ and $y = (y_1, \ldots, y_d)$ is given by

$$\langle x|y\rangle = \sum_{i=1}^{d} \overline{x_i} y_i. \tag{2.65}$$

An alternative but equivalent way of defining the dimension of an inner product space, in the finite case, is as the maximum number of linearly independent vectors that can exist in the space; that is, the largest size of the set $\{\varphi_1, \ldots, \varphi_n\}$ such that

$$\sum_i \alpha_i \varphi_i = 0 \Rightarrow \alpha_i = 0\, \forall i. \tag{2.66}$$

Every inner product space $\mathcal{H}$ defines a norm $\|\cdot\|$ on the space, with

$$\|\psi\| = \sqrt{\langle\psi|\psi\rangle}, \tag{2.67}$$

thus making $\mathcal{H}$ a normed space. A norm is a function from $\mathcal{H}$ to $\mathbb{R}$ that satisfies the following conditions, the first two of which are immediate consequences of the defining properties of the inner product:

1. $\|\psi\| \geq 0$, and $\|\psi\| = 0$ iff $\psi = 0$;

2. $\|c\psi\| = |c|\, \|\psi\|$ for $c \in \mathbb{C}$;

3. $\|\varphi + \psi\| \leq \|\varphi\| + \|\psi\|$.

The third property, called the triangle inequality, is a consequence of the inner product satisfying the Cauchy-Schwarz inequality. Note that if $\varphi \perp \psi$ or their overlap is strictly imaginary, i.e., $\mathrm{Re}(\langle\varphi|\psi\rangle) = 0$, then these vectors satisfy the Pythagorean formula

$$\|\varphi + \psi\|^2 = \|\varphi\|^2 + \|\psi\|^2 . \tag{2.68}$$

The norm defines a metric $d : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ via

$$d(x, y) := \|x - y\| . \tag{2.69}$$

Hence, $\mathcal{H}$ is also an example of a metric space with metric d.

Suppose that we possess a sequence $\{x_i\}$ of vectors on $\mathcal{H}$. This sequence is said to converge with respect to the metric d to the vector $x \in \mathcal{H}$ if, for every $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that

$$d(x_n, x) < \varepsilon \quad \forall\, n \geq N. \tag{2.70}$$

In what follows there will no ambiguity over the metric with which a sequence is convergent, so we will simply say that the sequence $\{x_n\}$ converges to x, which we can denote symbolically by $x_n \to x$, or $\lim_{n\to\infty} x_n = x$, and we say that x is the limit of the sequence. A weaker condition on the sequence is for it to be a Cauchy sequence. This means that, for every $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that

$$d(x_m, x_n) < \varepsilon \quad \forall\, m, n \geq N. \tag{2.71}$$

In other words, as the sequence increases, the elements get closer together. This is a strictly weaker condition, as convergence of a sequence implies that it is Cauchy, but there exist metric spaces for which Cauchy sequences are not convergent. A space where any Cauchy sequence is convergent is called a *complete* metric space. Any inner product space $\mathcal{H}$ that is also complete is called a *Hilbert space*, whilst a complete normed spaced is called a *Banach space.*

For a d-dimensional inner product space $\mathcal{H}$, suppose that we possess a set of d orthonormal vectors, that is normalised vectors that are pairwise orthogonal, $\{e_i\}_{i=1}^{d}$. This set forms a *basis* for $\mathcal{H}$, and with this basis we can express any vector $\psi \in \mathcal{H}$ in the form

$$\psi = \sum_{j=1}^{d} \psi_j e_j \tag{2.72}$$

where $\psi_j \in \mathbb{C}$. From the orthonormality of the basis vectors, $\psi_i = \langle e_i|\psi\rangle$ and so we may express ψ in the following way:

$$\psi = \sum_{i=1}^{d} \langle e_i|\psi\rangle\, e_i. \tag{2.73}$$

In the case of an infinite dimensional Hilbert space $\mathcal{H}$, Equation (2.72) again holds (with the summation running to infinity), where the proof relies on the completeness of the space and use of Bessel's inequality:

$$\sum_{i=1}^{n} |\langle e_i|\psi\rangle|^2 \leq \|\psi\|^2 . \tag{2.74}$$

The Hilbert spaces that we will consider are separable; that is, spaces that possess a countable dense subset. A Hilbert space is separable iff it possess a countable orthonormal basis. A (finite) d-dimensional Hilbert space is isomorphic to $\mathbb{C}^d$, whilst a countably infinite space—a space for which we can define a bijective map $f : \mathbb{N} \to \mathcal{H}$ via $f(n) = e_n \in \mathcal{H}$, where e_n is an element of an orthonormal basis—is isomorphic to the space $\ell^2(\mathbb{N})$ of square-summable functions $f : \mathbb{N} \to \mathbb{C}$, i.e., sequences, with the inner product on the space given by

$$\langle f|g\rangle = \sum_{j=0}^{\infty} \overline{f(j)}g(j), \tag{2.75}$$

and the norm $\|f\| = (\sum_{i=0}^{\infty} |f(j)|^2)^{1/2}$. A function $f : \mathbb{N} \to \mathbb{C}$ is an element of $\ell^2(\mathbb{N})$ if $\|f\|^2 < \infty$. By defining the Kronecker functions $\delta_j : \mathbb{N} \to \{0, 1\}$ via $\delta_j(k) = \delta_{jk}$, where δ_{jk} is the Kronecker delta, which is equal to 1 iff $j = k$ and 0 otherwise, we have provided the space with an orthonormal basis $\{\delta_k\}_{k\in\mathbb{N}}$, and the coefficients of $f \in \ell^2(\mathbb{N})$ are then $f_k = \langle \delta_k|f\rangle$.

Since all separable infinite-dimensional spaces are isomorphic, we can make use of additional structure that presents itself in the case of certain spaces when considering particular physical situations. The space that we will make the most use of is

$$L^2(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{C} \middle| \int_{\mathbb{R}} dx\, |f(x)|^2 < \infty \right\} / \sim, \tag{2.76}$$

i.e., the space of complex valued square-integrable functions with respect to the Lebesgue measure defined over the set $\mathbb{R}$, with the equivalence that we take the quotient over being $f \sim g$ iff the set of values x for which $f(x) \neq g(x)$ is of Lebesgue measure zero . This space possesses the inner product

$$\langle f|g\rangle = \int_{\mathbb{R}} d\omega\, \overline{f(\omega)}g(\omega), \tag{2.77}$$

and the norm

$$\|f\| = \left(\int_{\mathbb{R}} d\omega\, |f(\omega)|^2 \right)^{1/2} . \tag{2.78}$$

2.3.2 Bounded operators and dual spaces

2.3.2.1 Bounded operators

Consider a Hilbert space $\mathcal{H}$ and a linear map $T : \mathcal{H} \to \mathcal{H}$, i.e., for any $\psi, \varphi \in \mathcal{H}$ and $s \in \mathbb{C}$,

$$T(\psi + s\varphi) = T\psi + sT\varphi. \tag{2.79}$$

Such a map is called a (linear) *operator* on $\mathcal{H}$. The operator T is *bounded* if there exists a number $c \geq 0$ such that

$$\|T\psi\| \leq c\,\|\psi\| \qquad \forall \psi \in \mathcal{H}. \tag{2.80}$$

An operator is bounded iff it is a continuous linear mapping.

In the case of a finite-dimensional Hilbert space $\mathcal{H}_d$, every operator is bounded [47] but this is not the case for infinite-dimensional spaces. For any operator $T : \mathcal{H} \to \mathcal{H}$, we define

- The kernel of T: $\ker(T) = \{\varphi \in \mathcal{H} | T\varphi = 0\}$;
- The range of T: $\operatorname{ran}(T) = \{\psi \in \mathcal{H} | \psi = T\varphi \text{ for some } \varphi\}$;

Each of these form a linear subspace, which follows from the linearity of T, and the dimension of $\operatorname{ran}(T)$ is called the *rank* of T.

The set of bounded operators defined on $\mathcal{H}$ is a vector space, with

$$(S+T)\psi = S\psi + T\psi, \quad (cT)\psi = c(T\psi), \tag{2.81}$$

for all vectors $\psi \in \mathcal{H}$ and $c \in \mathbb{C}$. The zero element is given by the null operator O for which $O\psi = 0$ for all $\psi \in \mathcal{H}$. In addition, we denote by I the identity operator for which $I\psi = \psi$ for all $\psi \in \mathcal{H}$. From now on, we shall denote the vector space of bounded operators acting on the Hilbert space $\mathcal{H}$ by $\mathcal{L}(\mathcal{H})$. This space is also a normed space, with the operator norm defined as

$$\|T\| := \sup_{\|\psi\|=1} \|T\psi\|\,. \tag{2.82}$$

Equation (2.80) can now be rewritten: $\|T\|$ is clearly the smallest c for which $\|T\psi\| \leq c\,\|\psi\|$, and so if $T \in \mathcal{L}(\mathcal{H})$, then for every $\psi \in \mathcal{H}$,

$$\|T\psi\| \leq \|T\|\,\|\psi\|\,. \tag{2.83}$$

The space $\mathcal{L}(\mathcal{H})$ is complete with respect to the metric defined by the operator norm, so $\mathcal{L}(\mathcal{H})$ is a Banach space. Furthermore, $\mathcal{L}(\mathcal{H})$ is a *Banach algebra*: a vector space that is closed under multiplication of elements, and $\|ST\| \leq \|S\|\,\|T\|$ for any $S, T \in \mathcal{L}(\mathcal{H})$.

2.3.2.2 Linear functionals and dual spaces

Consider the linear continuous map $f : \mathcal{H} \to \mathbb{C}$. Such a map is called a linear functional. The space of all bounded linear functionals on $\mathcal{H}$ is called the *dual space* of $\mathcal{H}$, denoted by $\mathcal{H}^*$. This space has a norm defined on it: for $f \in \mathcal{H}^*$,

$$\|f\| := \sup_{\|\psi\|=1} |f(\psi)|\,. \tag{2.84}$$

By way of the inner product defined on $\mathcal{H}$, each vector φ defines a functional f_φ via

$$f_\varphi(\psi) = \langle\varphi|\psi\rangle\,, \tag{2.85}$$

which is bounded as a result of the Cauchy-Schwarz inequality. Indeed, it is the case that all bounded linear functionals take this form [24, p. 13]:

Theorem 2.3.1 (Fréchet-Riesz representation theorem). Any linear function $f : \mathcal{H} \to \mathbb{C}$ is bounded iff there exists a unique vector $\varphi \in \mathcal{H}$ satisfying Equation (2.85). Furthermore, $\|f\| = \|\varphi\|$.

Theorem 2.3.1 shows us a one-to-one correspondence between $\mathcal{H}$ and $\mathcal{H}^*$ (although the map is conjugate linear, as $c\varphi$ is mapped to $f_{c\varphi} = \overline{c} f_\varphi$). In other words, the Hilbert space is *self-dual*. At this point, if there is any risk of confusion, we make use of the Dirac convention of using kets $|\psi\rangle$ for elements of $\mathcal{H}$ and bras $\langle\varphi|$ for elements of $\mathcal{H}^*$.

2.3.2.3 Adjoints

For a bounded T acting on $\mathcal{H}$, we would like to see if there exists a corresponding operator acting on $\mathcal{H}^*$. The following theorem claims that this operator exists, and is unique.

Theorem 2.3.2. For every bounded operator $T \in \mathcal{L}(\mathcal{H})$, there exists a unique bounded operator, denoted T^* for which

$$\langle\varphi|T\psi\rangle = \langle T^*\varphi|\psi\rangle\,, \tag{2.86}$$

for all $\varphi, \psi \in \mathcal{H}$. Furthermore, $\|T^*\| = \|T\| = \|T^*T\|^{1/2}$.

The operator T^* defined by Equation (2.86) is called the *adjoint operator* of T. For all $S, T \in \mathcal{L}(\mathcal{H})$,

$$(ST)^* = T^*S^*, \tag{2.87a}$$

$$(S + cT)^* = S^* + \overline{c}T^*, \tag{2.87b}$$

$$(T^*)^* = T. \tag{2.87c}$$

In short, the space $\mathcal{L}(\mathcal{H})$ forms a C^*-algebra; that is,

- $\mathcal{L}(\mathcal{H})$ is a Banach space;
- $\mathcal{L}(\mathcal{H})$ forms an algebra (in particular, a Banach algebra);
- the map $* : T \mapsto T^*$ is conjugate linear and satisfies Equation (2.86);
- the operator norm on $\mathcal{L}(\mathcal{H})$ satisfies $\|T\| = \|T^*\|$ for all $T \in \mathcal{L}(\mathcal{H})$.

Suppose that we consider a finite dimensional Hilbert space $\mathcal{H}_d$ of dimension d. On this space all continuous linear operators are bounded, and so each possesses an adjoint. If we fix a basis $\{e_i\}_{i=1}^d$ for $\mathcal{H}_d$, then we may express T in terms of the $d \times d$ complex matrix $[T_{ij}]$, where $T_{ij} = \langle e_i|Te_j\rangle$. In which case, $T^*_{ij} = \overline{T_{ji}}$. In other words, as a finite-dimensional matrix, T^* is the conjugate transpose of the matrix describing T. In the case of an infinite-dimensional Hilbert space, we may express T as an infinite-dimensional matrix, but since not all operators in infinite-dimensional Hilbert spaces are bounded, there is some issue with deciding whether a given matrix on such a space is bounded. For the most part it is safer to just consider maps on these spaces.

2.3.2.4 Self-adjoint and positive operators

A particular class of bounded operators that we wish to pay attention to are *self-adjoint* operators. These are operators $A \in \mathcal{L}(\mathcal{H})$ for which $A^* = A$. From Equation (2.87b) we see that if we restrict to real linear combinations of self-adjoint operators, the adjoint mapping remains linear, and so the set of self-adjoint operators, denoted by $\mathcal{L}_s(\mathcal{H})$, forms a real vector space. However, $\mathcal{L}_s(\mathcal{H})$ does not form an algebra: if $A, B \in \mathcal{L}_s(\mathcal{H})$, then $(AB)^* = B^*A^* = BA$, and so $(AB)^* \neq AB$ unless $AB = BA$; that is, unless A and B *commute* (denoted $[A, B] = 0$). If we consider the bounded operator $T \in \mathcal{L}(\mathcal{H})$, we can decompose it into the sum of two self-adjoint operators: we define $T_R = (T + T^*)/2$ and $T_I = -i(T - T^*)/2$, which are readily verified as being self-adjoint, and from this we see that

$$T = T_R + iT_I. \tag{2.88}$$

From this, we see that another way of stating that an operator is self-adjoint is $T = T_R$. If an operator $T \in \mathcal{L}_s(\mathcal{H})$, then $\langle\psi|T\psi\rangle \in \mathbb{R}$ for all $\psi \in \mathcal{H}$. This allows us to consider a subset of $\mathcal{L}_s(\mathcal{H})$: a self-adjoint operator A is *positive* if, for all $\psi \in \mathcal{H}$, $\langle\psi|A\psi\rangle \geq 0$. We denote an operator A as positive by $A \geq 0$. As a simple example, consider $A = aI$, where $a \in [0, \infty)$. Similarly, any map from $\mathcal{L}(\mathcal{H})$ to $\mathcal{L}(\mathcal{H})$ is positive if it maps positive operators to positive operators. The concept of positivity of operators provides us with a natural partial ordering on $\mathcal{L}_s(\mathcal{H})$, where $S \geq T$ iff $S - T \geq 0$. This is indeed only a partial (and not a total) ordering, as for any two operators $S, T \in \mathcal{L}(\mathcal{H})$ there is no guarantee that either $S \geq T$ or $T \geq S$. Furthermore, if $R, S, T \in \mathcal{L}_s(\mathcal{H})$ then from the linearity of the inner product

- $S \geq R$ implies that $T + S \geq T + R$;
- $S \geq R$ implies $aS \geq aR$ for any $a \geq 0$.

In other words, $\mathcal{L}_s(\mathcal{H})$ forms a partially ordered vector space with regards to the relation $\geq$.

For any bounded operator $T \in \mathcal{L}(\mathcal{H})$, the product T^*T is positive, as

$$\langle\psi|T^*T\psi\rangle = \|T\psi\|^2 \geq 0, \tag{2.89}$$

for all ψ. Conversely, any positive operator S can be expressed in terms of a bounded operator and its adjoint, which is an immediate consequence of the following theorem (for a proof, see, e.g. [4, Theorem 23.2]):

Theorem 2.3.3. For every positive operator T there exists a unique positive operator $T^{1/2}$, called the *square root* of T, that satisfies $(T^{1/2})^2 = T$. If an operator $S \in \mathcal{L}(\mathcal{H})$ commutes with T, then it also commutes with $T^{1/2}$, and if T is invertible, then so is $T^{1/2}$ with $(T^{1/2})^{-1} = (T^{-1})^{1/2}$.

For any positive operator T, its square T^2 is also positive, and if $T \leq I$, then $\langle\psi|T^2\psi\rangle \leq \langle\psi|T\psi\rangle$ for all $\psi \in \mathcal{H}$. Therefore, for any positive operator T satisfying $0 \leq T \leq I$, its square satisfies $0 \leq T^2 \leq T$. We denote the space of such operators by $\mathcal{E}(\mathcal{H})$:

$$\mathcal{E}(\mathcal{H}) = \{T \in \mathcal{L}_s(\mathcal{H}) | 0 \leq T \leq I\}. \tag{2.90}$$

As a consequence of Theorem 2.3.3, we have the following: if $\langle\psi|T\psi\rangle = 0$, then $T\psi = 0$.

Much like how every bounded operator can be expressed as the sum of two self-adjoint operators, it is possible to express every self-adjoint operator in terms of two positive operators. In order to see this, we first note that for any $T \in \mathcal{L}_s(\mathcal{H})$, $-\|T\|\, I \leq T \leq \|T\|\, I$. With this inequality, we now define the two positive operators

$$T_+ = \frac{1}{2}(\|T\|\, I + T), \quad T_- = \frac{1}{2}(\|T\|\, I - T), \tag{2.91}$$

and so the self-adjoint operator T can be expressed as

$$T = T_+ - T_-. \tag{2.92}$$

As a result of this, we can express any bounded operator in terms of 4 positive operators.

We briefly also consider unbounded self-adjoint operators. An unbounded operator A defined on $\mathcal{H}$ with domain $\mathcal{D}(A)$ is *Hermitian* if $\langle\varphi|A\psi\rangle = \langle A\varphi|\psi\rangle$ for all $\varphi, \psi \in \mathcal{D}(A)$, and self-adjoint if $\mathcal{D}(A^*) = \mathcal{D}(A)$ and $A = A^*$ on their domain. Particular examples that we shall consider are the position and momentum operators Q and P, which satisfy

$$Q\psi(q) = q\psi(q), \qquad P\varphi(q) = -i\frac{d}{dx}\varphi(q), \tag{2.93}$$

for $\psi \in \mathcal{D}(Q), \varphi \in \mathcal{D}(P)$ (where we have set $\hbar = 1$ in the definition of P, as will be the standard later).

2.3.2.5 Unitary operators

We defined an isomorphic map U as a bilinear map between two inner product spaces that preserves the value of the inner product of the first space, as given in Equation (2.64). Suppose now that U is a map from a Hilbert space $\mathcal{H}$ to itself; in which case, Equation (2.64) takes the form

$$\langle U\varphi|U\psi\rangle = \langle U^*U\varphi|\psi\rangle = \langle\varphi|\psi\rangle\,. \tag{2.94}$$

A trivial example of an operator that satisfies this relation is the identity I, but there exist many more. These operators satisfy the following proposition (see e.g. [29, Proposition 1.47] for the proof):

Proposition 2.3.4. Let U be a linear mapping on $\mathcal{H}$, then the following are equivalent:

(i) U is an isomorphism;

(ii) U is a bijective isometry;

(iii) $U \in \mathcal{L}(\mathcal{H})$ and $UU^* = U^*U = I$.

The operators $U \in \mathcal{L}(\mathcal{H})$ forming such isomorphisms are said to be *unitary*, and we denote the set of unitary operators by $\mathcal{U}(\mathcal{H})$. In the finite-dimensional case, any operator satisfying either $UU^* = I$ or $U^*U = I$ is unitary: in this case $\det(UU^*) = \det(U)\det(U^*) = \det(I) = 1$, and so $\det(U) \neq 0$, which guarantees the existence of the inverse of U. However, for infinite-dimensional spaces, one must check that both $UU^* = I$ and $U^*U = I$ hold.

The set $\mathcal{U}(\mathcal{H})$ forms a group under the action of multiplication:

(a) $I \in \mathcal{U}(\mathcal{H})$;

(b) If $U, V \in \mathcal{U}(\mathcal{H})$, then $(UV)^*UV = V^*U^*UV = V^*V = I$ and $UV(UV)^* = UVV^*U^* = UU^* = I$, so $UV \in \mathcal{U}(\mathcal{H})$;

(c) Since $(U^*)^* = U$, if $U \in \mathcal{U}(\mathcal{H})$, $U^{-1} = U^* \in \mathcal{U}(\mathcal{H})$.

There is an important link between $\mathcal{U}(\mathcal{H})$ and $\mathcal{L}_s(\mathcal{H})$: For any $T \in \mathcal{L}(\mathcal{H})$, the Taylor series

$$e^T := \sum_{n=0}^{\infty} \frac{T^n}{n!}, \tag{2.95}$$

with $T^0 = I$, is known as the *exponential map* of T and satisfies (due to multiplication and the adjoint map being continuous in $\mathcal{L}(\mathcal{H})$)

$$e^{aT}e^{bT} = e^{(a+b)T}, \tag{2.96a}$$

$$\left(e^{aT}\right)^* = e^{\overline{a}T^*}, \tag{2.96b}$$

for any $T \in \mathcal{L}(\mathcal{H})$ and $a, b \in \mathbb{C}$. If we specify that $T \in \mathcal{L}_s(\mathcal{H})$ and $a = i\alpha$, where $\alpha \in \mathbb{R}$, then Equation (2.96b) is of the form

$$\left(e^{i\alpha T}\right)^* = e^{-i\alpha T}, \tag{2.97}$$

and so

$$\left(e^{i\alpha T}\right)^* e^{i\alpha T} = e^{-i\alpha T}e^{i\alpha T} = e^0 = I. \tag{2.98}$$

In other words, $\exp[i\alpha T] \in \mathcal{U}(\mathcal{H})$ for every $\alpha \in \mathbb{R}$, and each $T \in \mathcal{L}_s(\mathcal{H})$ defines a map $\alpha \mapsto \exp[i\alpha T]$ from $\mathbb{R}$ to $\mathcal{U}(\mathcal{H})$. The set $\{\exp[i\alpha T] | \alpha \in \mathbb{R}\} \subset \mathcal{U}(\mathcal{H})$ is a *one-parameter unitary group*, and there exists one for every self-adjoint operator T. Furthermore, a result by Stone [50] states that there exists a one-to-one correspondence between strongly continuous one-parameter unitary groups[2] and the space $\mathcal{L}_s(\mathcal{H})$, so for every such group $\alpha \mapsto U_\alpha$ where $U_\alpha^* = U_{-\alpha}$ there exists a self-adjoint operator T such that $U_\alpha = \exp[i\alpha T]$. Note that the results in this section hold even if the operator we consider is unbounded, and so we are free to consider the unitary operators $\exp[-i\lambda Q]$, etc., although additional machinery is needed to prove its validity, as the Taylor series will fail to converge.

2.3.2.6 The spectrum of an operator

Consider the bounded operator $T \in \mathcal{L}(\mathcal{H})$. A value $\lambda \in \mathbb{C}$ is an *eigenvalue* of T if there exists a nonzero vector $\psi \in \mathcal{H}$ satisfying $T\psi = \lambda\psi$, in which case ψ is the *eigenvector* of T with eigenvalue λ. More generally, λ is in the *spectrum* of T if $T - \lambda I$ is singular. We denote the spectrum of T by $\sigma(T)$.

Any eigenvector of T lies in the kernel of $T - \lambda I$, which means that the kernel contains nonzero vectors. This shows that $T - \lambda I$ is not an injective map, and so is not invertible,

[2] A unitary group $t \mapsto U_t$ is strongly continuous if as $t \to t_0$, $\|(U_t - U_{t_0})\varphi\| \to 0$ for all $\varphi \in \mathcal{H}$ and all t_0.

hence the eigenvalues of T belong to its spectrum. Indeed, in a d-dimensional space, the eigenvalues are the only solutions of the d-order polynomial equation $\det(T - \lambda I) = 0$, and every bounded operator possesses eigenvalues. This does not hold true in an infinite-dimensional system, but, regardless of the dimensionality of the space considered, every bounded operator $T \in \mathcal{L}(\mathcal{H})$ possesses a nonzero spectrum that is bounded by $\|T\|$: suppose that ψ is an eigenvector of T with eigenvalue λ, then $|\lambda| \leq \|T\|$. In the case of a self-adjoint operator T, $\sigma(T) \subset \mathbb{R}$, and for a positive operator A, the elements of its spectrum are non-negative and no larger than $\|A\|$, i.e., $\sigma(A) \subset [0, \|A\|]$. For a unitary operator $U \in \mathcal{U}(\mathcal{H})$, its eigenvalues λ satisfy $|\lambda| = 1$, and hence $\lambda = e^{i\alpha}$ for some $\alpha \in \mathbb{R}$. Since each self-adjoint operator generates a group of unitary operators, any eigenvector of the self-adjoint operator is also an eigenvector of the unitary operators generated by it.

2.3.2.7 Projections

A special class of operators that we will consider often are *projections*. A projection P is a self-adjoint operator that is also idempotent, i.e., $P \in \mathcal{L}_s(\mathcal{H})$ and $P^2 = P$. Furthermore, $P \geq 0$. We denote the set of projections on $\mathcal{H}$ by $\mathcal{P}(\mathcal{H})$, and shall consider the simplest example of a projection, as it is of great use to us: for any normalised vector $\varphi \in \mathcal{H}$, we define the operator P_φ by

$$P_\varphi \psi = \langle \varphi | \psi \rangle \, \varphi. \tag{2.99}$$

This is both self-adjoint and idempotent, so indeed it is a projection. The range of P_φ is the subspace $\{c\varphi | c \in \mathbb{C}\} = \mathbb{C}\varphi$, which is one-dimensional, and so P_φ is a one-dimensional projection.

Projections satisfy a series of important properties, which we shall make use of (the proofs of these Propositions are contained in [29, Section 1.2.3]).

Proposition 2.3.5. For any projection $P \in \mathcal{P}(\mathcal{H})$ where $0 \neq P \neq I$, $\|P\| = 1$, P only has eigenvalues 0 and 1 and any vector $\psi \in \mathcal{H}$ decomposes into the sum of two orthogonal vectors ψ_0 and ψ_1 such that $P\psi_0 = 0$, $P\psi_1 = \psi_1$.

Proposition 2.3.6. Let $P \in \mathcal{P}(\mathcal{H})$ and $\psi \in \mathcal{H}$, then the following are equivalent conditions:

(i) $\psi \in \operatorname{ran}(P)$;

(ii) $P\psi = \psi$;

(iii) $\|P\psi\| = \|\psi\|$.

Since $\mathcal{P}(\mathcal{H}) \subset \mathcal{L}_s(\mathcal{H})$, it inherits the partial ordering structure given by the relation $\geq$, which leads to the following proposition:

Proposition 2.3.7. Let $P, Q \in \mathcal{P}(\mathcal{H})$, then the following conditions are equivalent:

(i) $P \geq Q$;

(ii) $PQ = QP = Q$;

(iii) $P - Q$ is a projection.

For any $P \in \mathcal{P}(\mathcal{H})$, we define the new operator $P^{\perp} = I - P$ called the *complement* of P. Since $P \in \mathcal{P}(\mathcal{H}) \subset \mathcal{L}_s(\mathcal{H})$, it follows that $P^{\perp} \in \mathcal{L}_s(\mathcal{H})$. Furthermore, $P^{\perp} \in \mathcal{P}(\mathcal{H})$ and $(P^{\perp})^{\perp} = P$. If $P \geq Q$, then $Q^{\perp} \geq P^{\perp}$. Suppose for $P, Q \in \mathcal{P}(\mathcal{H})$ that $P \geq Q$ and $P^{\perp} \geq Q$, then $Q = 0$. In other words, if $Q \in \mathcal{P}(\mathcal{H})$ is ordered below the projection P and its complement, then $Q = 0$, and 0 is the infimum of a projection and its complement. With these properties the map $P \mapsto P^{\perp}$ on $\mathcal{P}(\mathcal{H})$ is an *orthocomplementation.*

For any two projections $P, Q \in \mathcal{P}(\mathcal{H})$, the sum $P + Q \in \mathcal{P}(\mathcal{H})$ iff $PQ = QP = 0$. In the case of two one-dimensional projections $P_{\varphi}, P_{\xi} \in \mathcal{P}(\mathcal{H})$, $P_{\varphi}P_{\xi} = 0$ iff $\varphi \perp \xi$.

Consider a projection $P \in \mathcal{P}(\mathcal{H})$ with $\mathrm{ran}(P)$ of dimension $r > 1$. Then P can be expressed as a sum of r one-dimensional projections: assuming that $\mathrm{ran}(P)$ is finite-dimensional, we fix an orthonormal basis $\{\varphi_i\}_{i=1}^{r}$ for $\mathrm{ran}(P)$ and define the one-dimensional projections $P_i = P_{\varphi_i}$. If $i \neq j$, P_i and P_j are orthogonal and so their sum will be a projection, namely, $\sum_{i=1}^{r} P_i = P$. For the infinite-dimensional case, note that if $\psi \in \mathrm{ran}(P)$, $P^{\perp}\psi = 0$, and so $P^{\perp}$ is a continuous map from $\mathrm{ran}(P)$ to the closed subspace $\{0\}$. Hence, $\mathrm{ran}(P)$ is a closed linear subspace of $\mathcal{H}$ and possesses an orthonormal basis $\{\varphi_i\}_{i=1}^{\infty}$, from which we define the rank one projections $P_i = P_{\varphi_i}$ as above, so the infinite sum $\sum_{i=1}^{\infty} P_i$ converges to P in the weak (and strong) operator topology (which we shall discuss in Section 2.3.2.10).

Proposition 2.3.8. Let $P \in \mathcal{P}(\mathcal{H})$ and T be a positive operator, then if $T \leq P$, it follows that $TP = PT = T$.

2.3.2.8 Rank-one operators

The one-dimensional projections that we have discussed can be expressed in the form

$$P_{\eta} = |\eta\rangle\langle\eta|\,, \tag{2.100}$$

where $(|\eta\rangle\langle\eta|)\psi = \langle\eta|\psi\rangle\,\eta$. Indeed, we define an entire class of operators of the form $|\varphi\rangle\langle\eta|$, where for any $\psi \in \mathcal{H}$, $(|\varphi\rangle\langle\eta|)\psi = \langle\eta|\psi\rangle\,\varphi$. By applying the Cauchy-Schwarz inequality we immediately see that $|\varphi\rangle\langle\eta| \in \mathcal{L}(\mathcal{H})$, we also see that its range is the one-dimensional subspace $\mathbb{C}\varphi = \{z\varphi | z \in \mathbb{C}\}$. We describe such an operator as a *rank-one operator.* If we apply the adjoint map $(|\varphi\rangle\langle\eta|)^* = |\eta\rangle\langle\varphi|$, we see that a rank-one operator is self-adjoint iff it is of the form $R = r\,|\eta\rangle\langle\eta|$, with $r \in \mathbb{R}$ and $\eta \in \mathcal{H}$. In other words, self-adjoint rank-one operators are multiples of a rank-one projection.

2.3.2.9 Trace class operators

In a d-dimensional space (which is therefore isomorphic to $\mathbb{C}^d$), the trace of an operator A is given by

$$\mathrm{tr}\,[A] = \sum_{i=1}^{d} \langle\varphi_i|A\varphi_i\rangle = \sum_{i=1}^{d} A_{ii}, \tag{2.101}$$

where $\{\varphi_i\}_{i=1}^{d}$ is an orthonormal basis for the space and $A_{ii} = \langle\varphi_i|A\varphi_i\rangle$. In this case, the trace is equal to the sum of the eigenvalues of A (including any repetitions of eigenvalues if A possesses any degenerate eigenvalues)[3]. However, if we wish to define the trace for an infinite-dimensional space we must be more careful. If we are starting with a separable infinite-dimensional Hilbert space with an orthonormal basis $\{\varphi_i\}_{i=1}^{\infty}$, then we may define the trace for a positive operator T in an analogous way:

$$\mathrm{tr}\,[T] = \sum_{i=1}^{\infty} \langle \varphi_i | T\varphi_i \rangle \,, \tag{2.102}$$

thereby defining a sum of non-negative numbers. In the case that this sum does not converge then we say that $\mathrm{tr}\,[T] = \infty$. As in the finite dimensional case, for a positive operator T the trace $\mathrm{tr}\,[T]$ is independent of the basis considered for the space. The map $\mathrm{tr}\,[\cdot]$ is map from the set of positive operators to $\mathbb{R}$, and for any positive operator T and $U \in \mathcal{U}(\mathcal{H})$, $\mathrm{tr}\,[UTU^*] = \mathrm{tr}\,[T]$.

In order to extend the concept of the trace to any bounded operator $T \in \mathcal{L}(\mathcal{H})$, we recall that T^*T is positive, and so possesses a unique positive square root operator. We denote this operator by $|T| := (T^*T)^{1/2}$. With this in mind, we say that a bounded operator $T \in \mathcal{L}(\mathcal{H})$ is *trace class* if $\mathrm{tr}\,[|T|] < \infty$, and we denote the set of trace class operators by $\mathcal{T}(\mathcal{H})$. This is a proper subset of $\mathcal{L}(\mathcal{H})$, as $I \notin \mathcal{T}(\mathcal{H})$. If an operator $T \in \mathcal{T}(\mathcal{H})$, then the trace given in Equation (2.102) satisfies the properties we require of it, in particular the fact its value is independent of the basis we measure it with.

The set $\mathcal{T}(\mathcal{H})$ forms a vector space, and the map $T \mapsto \mathrm{tr}\,[|T|] =: \|T\|_{\mathrm{tr}}$ defines the *trace norm* $\|\cdot\|_{\mathrm{tr}}$. Furthermore, we can define the *Hilbert-Schmidt norm* $\|\cdot\|_{HS} : \mathcal{L}(\mathcal{H}) \to \mathbb{R} \cup \{\infty\}$ via

$$\|T\|_{HS} := \left(\mathrm{tr}\,[T^*T]\right)^{1/2} . \tag{2.103}$$

Any operator A satisfying $\|A\|_{HS} < \infty$ is said to be a Hilbert-Schmidt operator, and the space of Hilbert-Schmidt operators is denoted by $HS(\mathcal{H})$. The space $\mathcal{T}(\mathcal{H})$ is dense in $HS(\mathcal{H})$ with respect to the metric defined by the Hilbert-Schmidt norm, and $HS(\mathcal{H})$ forms a Hilbert space with respect to the Hilbert-Schmidt inner product $\langle \cdot | \cdot \rangle_{HS} : HS(\mathcal{H}) \times HS(\mathcal{H}) \to \mathbb{C}$, where for any two operators $T, S \in HS(\mathcal{H})$

$$\langle T | S \rangle_{HS} := \mathrm{tr}\,[T^*S] \,. \tag{2.104}$$

The space $\mathcal{T}(\mathcal{H})$ is an ideal of $\mathcal{L}(\mathcal{H})$: for any $S \in \mathcal{L}(\mathcal{H})$ and $T \in \mathcal{T}(\mathcal{H})$, the products ST and TS are trace class, satisfy $\mathrm{tr}\,[ST] = \mathrm{tr}\,[TS]$ and

$$|\mathrm{tr}\,[TS]| \leq \|T\|_{\mathrm{tr}} \|S\| \,. \tag{2.105}$$

Any rank-one operator $|\varphi\rangle\langle\psi|$ is trace class, and the trace satisfies $\mathrm{tr}\,[|\varphi\rangle\langle\psi|] = \langle\psi|\varphi\rangle$. From this we can associate the value $\langle\eta|A\eta\rangle$ for any $A \in \mathcal{L}(\mathcal{H})$ with the rank-one projection $P_\eta \in \mathcal{P}(\mathcal{H})$ via $\mathrm{tr}\,[P_\eta A] = \langle\eta|A\eta\rangle$.

Trace class operators possess a discrete spectrum, and for a self-adjoint trace class

[3] Note that this also holds when A is not diagonalisable, i.e., $\mathrm{ran}(A)$ is a proper subset of $\mathbb{C}^d$. In this case, one may make use of the Jordan normal form J, which is an upper triangular matrix with the eigenvalues of A on the diagonal and ones on the superdiagonal, via $A = SJS^{-1}$, where S is an invertible matrix. The operators A and J share the same eigenvalues and trace, proving the statement.

operator T, the three norms presented can be expressed in terms of the eigenvalues $\{\lambda_j\}$ of T:

$$\|T\| = \max_j |\lambda_j|, \quad \|T\|_{\mathrm{tr}} = \sum_j |\lambda_j|\,, \quad \|T\|_{HS} = \big(\sum_j |\lambda_j|^2\big)^{1/2}, \tag{2.106}$$

and so the following hierarchy exists for the norms (which also holds for non-self-adjoint trace class operators):

$$\|T\| \leq \|T\|_{HS} \leq \|T\|_{\mathrm{tr}}\,. \tag{2.107}$$

If we consider $\mathcal{P}(\mathcal{H})$, the rank-one projections are trace one. Furthermore, projections $P, Q \in \mathcal{P}(\mathcal{H})$ are orthogonal iff $\langle P|Q\rangle_{HS} = 0$.

Much like we considered the dual space of $\mathcal{H}$, we also wish to consider the dual space $\mathcal{T}(\mathcal{H})^*$ of $\mathcal{T}(\mathcal{H})$ and in this case $\mathcal{T}(\mathcal{H})^* \cong \mathcal{L}(\mathcal{H})$: there exists a bijective linear mapping $S \mapsto f_S$ from $\mathcal{L}(\mathcal{H})$ to $\mathcal{T}(\mathcal{H})^*$, where $f_S(T) = \mathrm{tr}\,[ST]$ for all $T \in \mathcal{T}(\mathcal{H})$, such that $\|S\| = \|f_S\|$ for every $S \in \mathcal{L}(\mathcal{H})$, where

$$\|f_S\| = \sup_{A \in \mathcal{T}(\mathcal{H})} |f_S(A)|\,, \tag{2.108}$$

(see [25, Theorem 19.1]).

2.3.2.10 Operator topologies

There exist several topologies that we can define on $\mathcal{L}(\mathcal{H})$. The first that we have defined is given by the operator norm $\|\cdot\|$. A sequence $\{T_j\} \subset \mathcal{L}(\mathcal{H})$ converges to an operator T *in the operator norm topology*, or *uniformly*, if

$$\lim_j \|T - T_j\| = 0. \tag{2.109}$$

There exist additional topologies that we may define on $\mathcal{L}(\mathcal{H})$ that prove to be of some use. A sequence $\{T_j\} \subset \mathcal{L}(\mathcal{H})$ converges to an operator T *in the strong operator topology* if

$$\lim_j \|(T - T_j)\psi\| = 0, \tag{2.110}$$

for all $\psi \in \mathcal{H}$, and converges *in the weak operator topology* if

$$\lim_j \big|\langle\varphi|(T - T_j)\psi\rangle\big| = 0, \tag{2.111}$$

for all $\psi, \varphi \in \mathcal{H}$. By application of the Cauchy-Schwarz inequality,

$$\big|\langle\varphi|(T - T_j)\psi\rangle\big| \leq \|\varphi\|\,\|\psi\|\,\|T - T_j\|\,, \tag{2.112}$$

and so we possess the following hierarchy:

$$\text{Uniform convergence} \quad \Rightarrow \quad \text{Strong convergence} \quad \Rightarrow \quad \text{Weak convergence.} \tag{2.113}$$

2.3.2.11 Tensor products

Consider two Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$. We can form a product of these two spaces, called the *tensor product* of $\mathcal{H}$ and $\mathcal{K}$ that is itself a Hilbert space. For any vectors $\varphi \in \mathcal{H}$ and $\psi \in \mathcal{K}$, we denote the conjugate bilinear form $\varphi \otimes \psi$ that acts on the product $\mathcal{H}^* \times \mathcal{K}^*$ by

$$(\varphi \otimes \psi)(\xi^*, \eta^*) = \langle \xi|\varphi\rangle \langle \eta|\psi\rangle , \tag{2.114}$$

for all $\xi^* \in \mathcal{H}^*$ and $\eta^* \in \mathcal{K}^*$. We denote by ε the space of finite linear combinations of such linear forms and define the inner product $\langle\cdot|\cdot\rangle$ on ε by defining

$$\langle \xi \otimes \eta|\varphi \otimes \psi\rangle = \langle \xi|\varphi\rangle \langle \eta|\psi\rangle , \tag{2.115}$$

and extending by linearity. The completion of ε under the inner product $\langle\cdot|\cdot\rangle$ is called the tensor product of $\mathcal{H}$ and $\mathcal{K}$, and is denoted by $\mathcal{H} \otimes \mathcal{K}$. If $\{\varphi_i\}$ forms an orthonormal basis for $\mathcal{H}$ and $\{\psi_j\}$ forms an orthonormal basis for $\mathcal{K}$, then $\{\varphi_i \otimes \psi_j\}$ forms an orthonormal basis for $\mathcal{H} \otimes \mathcal{K}$.

In an analogous way, if $S \in \mathcal{L}(\mathcal{H})$ and $T \in \mathcal{L}(\mathcal{K})$, then $S \otimes T \in \mathcal{L}(\mathcal{H} \otimes \mathcal{K})$, with $(S \otimes T)(\psi \otimes \varphi) = (S\psi) \otimes (T\varphi)$. Note, however, that there exist operators in $\mathcal{L}(\mathcal{H} \otimes \mathcal{K})$ which cannot be decomposed in such a way.

Consider the tensor product $\mathcal{H} \otimes \mathcal{K}$ and any operator $T \in \mathcal{L}(\mathcal{H} \otimes \mathcal{K})$. We define the *partial trace* over $\mathcal{H}$ as the linear mapping

$$\mathrm{tr}_{\mathcal{H}} : \mathcal{T}(\mathcal{H} \otimes \mathcal{K}) \to \mathcal{T}(\mathcal{K}) \tag{2.116}$$

which satisfies

$$\mathrm{tr}\left[\mathrm{tr}_{\mathcal{H}}\left[T\right] E\right] = \mathrm{tr}\left[T(I \otimes E)\right] , \tag{2.117}$$

for any $E \in \mathcal{L}(\mathcal{K})$, and similarly for the partial trace over $\mathcal{K}$. As one would expect, this map corresponds to performing the trace over the one subsystem, and indeed for any state $\varphi \in \mathcal{K}$ and orthonormal basis $\{\psi_i\}$ for $\mathcal{H}$,

$$\langle \varphi|\mathrm{tr}_{\mathcal{H}}\left[T\right] \varphi\rangle = \sum_i \langle \psi_i \otimes \varphi|T\psi_i \otimes \varphi\rangle . \tag{2.118}$$

2.4 Quantum theory of measurement

In what follows we shall give an overview of the operational version of quantum physics, in particular the act of measuring quantum systems, as described by [26, 36, 34, 11, 31]. However, we shall not discuss possible interpretations of the underlying reality of the measurements, and instead shall only focus on quantum mechanics as a statistical theory.

Since we are considering the act of performing a measurement of a quantum system, we have to think about what happens in this process. We can divide it into three separate procedures: firstly, we *prepare* the system in such a way that we specify the initial conditions the system possesses; secondly, we *measure* the system with regards to the observable we wish to infer information about, for example, a particle's momentum or an electron's spin; finally, we *register* the measurement outcome that occurs for our given system. From repeated measurements, we are able to calculate the probability of a system prepared in a given way producing a specific outcome when we measure certain quantities. With this procedure in mind, we shall now elaborate on the two operational constructs that we require, namely states and observables.

2.4.1 States

The first part of the procedure of measuring that we need to consider is the preparation of the system being measured. In doing so we specify conditions that the system must possess no matter how many copies of it we wish to make, thereby ensuring that the statistics of what we are measuring are accurate. We may use one of several preparations that lead to the same initial conditions, and so they are statistically the same. We describe these statistically equivalent preparations mathematically with the same operator, known as the *state* of the system. There is the possibility that we may prepare an ensemble of systems in a mixture of two states, denoted by ρ_1 and ρ_2, say. If $\lambda \in [0,1]$ is the probability of preparing a system in the state ρ_1, and similarly $1-\lambda$ for ρ_2, then we describe the ensemble by the *mixed state* $\lambda\rho_1 + (1-\lambda)\rho_2$. More generally, we assume σ-convexity: for a sequence $\{\rho_j\}$ of states and a sequence $\{\lambda_k\}$ of equal length of positive numbers adding to one, then the sum $\sum_j \lambda_j \rho_j$ is also a valid state. We conclude that states form a convex set with *pure states* corresponding to extremal elements; that is, states for which the decomposition $\lambda\rho_1 + (1-\lambda)\rho_2$ with $\lambda \in [0,1]$ implies that $\rho_1 = \rho_2$. Any mixed state admits uncountably many convex decompositions, and so the pure states are the only states for which their decomposition is unique. The preceding discussion can be extended to any finite convex combinations of states.

In what follows, where we focus on Hilbert spaces exclusively, we denote the set of states on $\mathcal{H}$ by $\mathcal{S}(\mathcal{H})$. This set is a convex subset of $\mathcal{T}(\mathcal{H})$ composed of positive operators of unit trace:

$$\mathcal{S}(\mathcal{H}) = \{\rho \in \mathcal{T}(\mathcal{H}) | \rho \geq 0,\, \mathrm{tr}\,[\rho] = 1\}, \tag{2.119}$$

and in this case σ-convexity means that the sum $\sum_j \lambda_j \rho_j$ converges in the trace norm with the limit belonging to $\mathcal{S}(\mathcal{H})$. Since they are bounded operators, we will also refer to states as *density operators* at times. With this structure in mind, the extremal elements of $\mathcal{S}(\mathcal{H})$ are the rank-one projections, and we shall at times associate pure states to normalised rays in $\mathcal{H}$ (although this is simply a shorthand for their corresponding rank-one projections). This includes superpositions of pure states, too; so if $\psi_1, \psi_2 \in \mathcal{H}$ are states (that is, normalised vectors), then so is $(\alpha\psi_1 + \beta\psi_2)/\left\|\alpha\psi_1 + \beta\psi_2\right\|$ for any $\alpha, \beta \in \mathbb{C}$ with $|\alpha|^2 + |\beta|^2 = 1$ and $\alpha\psi_1 + \beta\psi_2 \neq 0$.

There will be instances where we are required to consider states of multiple systems, for example in the interaction between a pair of electrons. If we are dealing with two Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$ describing two systems individually, then we consider the tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$ for their combined system. For a generic state $\rho \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$, we may produce *reduced states* by performing the partial trace on the system, i.e., $\rho_1 = \mathrm{tr}_{\mathcal{H}_2}[\rho] \in \mathcal{S}(\mathcal{H}_1)$ and $\rho_2 = \mathrm{tr}_{\mathcal{H}_1}[\rho] \in \mathcal{S}(\mathcal{H}_2)$. A vector $\psi \in \mathcal{H}_1 \otimes \mathcal{H}_2$ is *separable* if there exist vectors $\varphi_1 \in \mathcal{H}_1$ and $\varphi_2 \in \mathcal{H}_2$ such that $\psi = \varphi_1 \otimes \varphi_2$. If no such vectors exist, then ψ is said to be *entangled*. In a similar fashion, a state $\rho \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is separable if it can be expressed as a convex combination of states of the form $\rho_1 \otimes \rho_2$ where $\rho_1 \in \mathcal{S}(\mathcal{H}_1)$ and $\rho_2 \in \mathcal{S}(\mathcal{H}_2)$ (such states are known as *factorised*), and is entangled if it is not separable.

2.4.2 Observables

2.4.2.1 Effects

Consider an ensemble of N systems prepared in the state ρ and we perform the measurement M, which for simplicity we shall assume has a discrete number of possible measurement outcomes ω_i. The number of times that each outcome occurs is given by $N(\omega_i)$, and so the relative frequency of this outcome occurring after N measurements is given by $N(\omega_i)/N$. As we let N get larger, this number is expected to tends towards $p(\omega_i|\rho, M)$, the probability of receiving outcome ω_i after performing a measurement of M with state ρ. This defines an affine functional[4] $E_i^{(M)} : \rho \mapsto p(\omega_i|\rho, M) \in [0,1]$ for every outcome ω_i. These functionals are called (measurement) *effects*, and in what follows we shall drop the superscript M. Three examples of effects worth mentioning are the identity effect $E_I(\rho) = 1$ for all ρ, the null effect $E_O(\rho) = 0$ for all ρ, and the set of trivial effects, which for any state ρ will always produce the same value $\lambda \in [0,1]$ (clearly, the identity and null effects are examples of trivial effects). The state independence of the trivial effects simply highlights that we are gaining no information about the state by performing this measurement; for example, if two outcomes are described by the effect $E_i : \rho \mapsto 1/2$, then we are simply performing a coin toss.

Similarly, we may begin with effects $\{E_i^{(M)}\}$ and then consider the probability $p(\omega_i|\rho, M)$ via the map $\mu_\rho : E_i^{(M)} \mapsto p(\omega_i|\rho, M)$ defined for ρ. This use of states to define positive linear functionals on effects and similarly the use of effects as affine functionals on states highlights the statistical duality between states and effects.

With the effects forming functionals on $\mathcal{S}(\mathcal{H}) \subset \mathcal{T}(\mathcal{H})$, it follows that the functionals must belong to $\mathcal{T}(\mathcal{H})^* = \mathcal{L}(\mathcal{H})$. Indeed, for every effect E there exists a bounded operator $\widetilde{E}$ such that $E(\rho) = \text{tr}\left[\widetilde{E}\rho\right]$ for all $\rho \in \mathcal{S}(\mathcal{H})$. With this in mind, the identity effect $E_I = I$ and null effect $E_O = O$. Similarly, since any trivial effect yields $E(\rho) = \lambda$ for all $\rho \in \mathcal{S}(\mathcal{H})$, it must take the form λI for $\lambda \in [0,1]$. As we require $0 \leq E(\rho) \leq 1$ for any $\rho \in \mathcal{S}(\mathcal{H})$, the effects must satisfy $O \leq \widetilde{E} \leq I$, and so the set of effects is $\mathcal{E}(\mathcal{H}) = \{E \in \mathcal{L}(\mathcal{H}) | O \leq E \leq I\}$ as given before. This set forms a convex space and its extremal elements coincide with $\mathcal{P}(\mathcal{H})$. From now on, we shall remove the tilde from the effect operators and simply refer to $\mathcal{E}(\mathcal{H})$ as the space of effects.

Much like with the idea of preparation, in which many different preparations lead to equivalent statistics (the equivalence class being described by the state of the system), many different measurements will lead to the same probability assignments. The effects correspond to the equivalence class of such probability assignments. We then express the equivalence class of statistically indistinguishable experiments; that is, experiments with the same collection of effects, by an *observable*, described by the mapping $i \mapsto E_i^{(M)}$. In order to fully describe observables, we must first briefly discuss measure theory.

[4] By affine we mean $E_i^{(M)}(\lambda\rho + (1-\lambda)\rho') = \lambda E_i^{(M)}(\rho) + (1-\lambda)E_i^{(M)}(\rho')$, $\lambda \in [0,1]$, for any outcome i.

2.4.2.2 Measure theory

In order to provide a sense of what we mean by an observable of a quantum system, we need to consider the concept of a measure (in the mathematical sense).

Consider a set Ω and its power set 2^{Ω}. A subset $\Sigma \subseteq 2^{\Omega}$ is a *σ-algebra* if Σ obeys the following properties:

1. $\Omega, \emptyset \in \Sigma$;
2. If $X \in \Sigma$, then $X^c \in \Sigma$;
3. If $\{X_i\}_{i\in I}$ is a countable sequence of sets $X_i \in \Sigma$, then $\bigcup_{i\in I} X_i \in \Sigma$.

In other words, Σ is a σ-algebra if it contains both Ω and $\emptyset$, and is closed under complementation and countable unions. The pair (Ω, Σ) forms a measurable space, with the elements of Σ called measurable sets, and upon this space we define a *real-valued (positive) measure* μ as a map $\mu : \Sigma \to [0, \infty]$ such that

1. $\forall X \in \Sigma$, $\mu(X) \geq 0$;
2. $\mu(\emptyset) = 0$;
3. For a countable sequence $\{X_i\}_{i\in I}$ of pairwise disjoint elements of Σ; i.e, $X_i \in \Sigma$ for all $i \in I$ and $X_i \cap X_j = \emptyset$ for all $i, j \in I$, $\mu(\cup_{i\in I} X_i) = \sum_{i\in I} \mu(X_i)$.

The triple (Ω, Σ, μ) is called a measure space. An important point that we shall make use of is the monotonicity of measures: if $A, B \in \Sigma$ and $B \subseteq A$, then $\mu(B) \leq \mu(A)$. To prove this, note that

$$A = (A \cap B) \cup (A \cap B^c) = B \cup (A \cap B^c), \tag{2.120}$$

since $B \subseteq A$. These are disjoint sets, so

$$\mu(A) = \mu\big((B) \cup (A \cap B^c)\big) = \mu(B) + \mu(A \cap B^c) \geq \mu(B), \tag{2.121}$$

from the positivity of the measure. Further to this, μ is subtractive: if $\mu(B) < \infty$, then $\mu(A \backslash B) = \mu(A \cap B^c) = \mu(A) - \mu(B)$.

Consider two measurable spaces (Ω_A, Σ_A) and (Ω_B, Σ_B). A function $f : \Omega_A \to \Omega_B$ is measurable if the inverse image of a measurable set is measurable; that is, if $Y \in \Sigma_B$, then $f^{-1}(Y) \in \Sigma_A$. Note the analogous form this has to continuous functions between topological spaces.

The most useful (and most commonly used in what follows) example of a measurable space is $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ is defined as

$$\mathcal{B}(\mathbb{R}) = \bigcap \{\Sigma | \Sigma \text{ is a } \sigma\text{-algebra containing all intervals in } \mathbb{R}\}. \tag{2.122}$$

From this definition, $\mathcal{B}(\mathbb{R})$ is clearly a σ-algebra: each Σ in the intersection contains both $\mathbb{R}$ and $\emptyset$, and so must $\mathcal{B}(\mathbb{R})$; if $X \in \mathcal{B}(\mathbb{R})$, then X belongs to each Σ in the intersection, hence $X^c \in \Sigma$ as they are all closed under complementation, and so $X^c \in \mathcal{B}(\mathbb{R})$; similarly, if a countable sequence of sets $\{X_i\}_{i\in I} \in \mathcal{B}(\mathbb{R})$, then $\{X_i\}_{i\in I} \in \Sigma$ for all Σ in the intersection, hence $\bigcup_{i\in I} X_i \in \Sigma$ for each Σ and so $\bigcup_{i\in I} X_i \in \mathcal{B}(\mathbb{R})$. Furthermore, $\mathcal{B}(\mathbb{R})$ is the smallest σ-algebra containing all intervals in $\mathbb{R}$: if there existed a smaller such σ-algebra, then $\mathcal{B}(\mathbb{R})$ would contain it as a subset, contradicting its construction as an intersection of all such σ-algebras. This space is known as the Borel σ-algebra over $\mathbb{R}$. Upon $\mathcal{B}(\mathbb{R})$ we use the *Lebesgue measure* m, which for any interval in $\mathbb{R}$ the measure is equal to the difference of the endpoints:

$$m([a,b]) = m((a,b)) = m((a,b]) = m([a,b)) = b - a, \quad b < a. \tag{2.123}$$

From the definition of a measure, it follows that $m(\emptyset) = 0$, but the empty set is not the only set to satisfy this property. For example, the point set $\{\alpha\}$ is contained in the open interval $(\alpha - \delta/2, \alpha + \delta/2)$, which has measure

$$m((\alpha - \delta/2, \alpha + \delta/2)) = (\alpha + \delta/2) - (\alpha - \delta/2) = \delta. \tag{2.124}$$

If we let δ tend to zero the measure of the set tends to zero and so from the monotonicity of the measure, $\mu(\{\alpha\}) \leq \mu((\alpha - \delta/2, \alpha + \delta/2))$, hence $\mu(\{\alpha\}) = 0$. Similarly, countable unions of point sets are also measure zero, as they are disjoint and hence are equal to the sum of measure zero sets. Any set $X \in \mathcal{B}(\mathbb{R})$ of measure zero is called a null set.

Suppose that we possess two measurable functions f, g mapping from $\mathbb{R}$ to some other measurable space. We say that $f = g$ almost everywhere (a.e.) if the functions are equal excluding a null set of values with respect to the measure μ:

$$f = g \text{ a.e.} \Leftrightarrow \mu\big(x \in \mathbb{R} | f(x) \neq g(x)\big) = 0. \tag{2.125}$$

If we possess two measures, $\mu, \mu' : \Omega \to \mathbb{R}$, where Ω is assumed to have a defined sum operation, then we define the *convolution*, denoted $\mu * \mu'$, as

$$\mu * \mu'(X) = (\mu \times \mu')(\{(x, x') | x + x' \in X\}), \tag{2.126}$$

for any $X \in \mathcal{B}(\mathbb{R})$ and $\mu \times \mu'$ is the product measure (see, e.g., [22, Chapter 6]).

2.4.2.3 POVMs

As was discussed in Section 2.4.2.1, if we consider the probability $p(\omega_i|\rho, M)$ of acquiring the outcome ω_i in the measurement M of a system in the state $\rho \in \mathcal{S}(\mathcal{H})$, then this may be expressed in terms of the effect $E_i^{(M)}$ via

$$p(\omega_i|\rho, M) = E_i^{(M)}(\rho) = \mathrm{tr}\left[\widetilde{E}_i^{(M)} \rho\right], \tag{2.127}$$

where $\widetilde{E}_i^{(M)} \in \mathcal{E}(\mathcal{H})$ is the unique operator corresponding to $E_i^{(M)}$. For each state ρ and measurement M we can then define the probability measure p_ρ^M via $p_\rho^M(\omega_i) = p(\omega_i|\rho, M)$, i.e.,

$$p_\rho^M(\omega_i) = \mathrm{tr}\left[\widetilde{E}_i^{(M)} \rho\right]. \tag{2.128}$$

If we assume that the measurement M has the measurable space (Ω, Σ), then the probability measure p^M_ρ satisfies the following conditions:

1. $0 \leq p^M_\rho(X) \leq 1$ for all $X \in \Sigma$;
2. $p^M_\rho(\Omega) = 1$;
3. If $\{X_i\}_{i \in I}$ is a sequence of disjoint sets in Σ, then $p^M_\rho(\cup_{i \in I} X_i) = \sum_{i \in I} p^M_\rho(X_i)$.

In the case of a probability measure, the measurable space is known as an outcome space and the measurable sets are known as events. We noted earlier that an observable is a map from the measurement outcomes to the effects, i.e., from Σ to $\mathcal{E}(\mathcal{H})$. If we label an observable describing the measurement M by E^M, say, then $\widetilde{E}_i^{(M)} =: \mathsf{E}^M(\omega_i)$. Assuming p^M_ρ has outcome space (Ω, Σ), Equation (2.128) can be rewritten as

$$p^M_\rho(X) = \mathrm{tr}\left[\mathsf{E}^M(X)\rho\right], \tag{2.129}$$

for any $X \in \Sigma$. From the requirements of the probability measure p^M_ρ for any state $\rho \in \mathcal{S}(\mathcal{H})$, we require that the map $\mathsf{E}^M : \Sigma \to \mathcal{E}(\mathcal{H})$ satisfies

1. $\mathsf{E}^M(X) \geq 0$ for all $X \in \Sigma$;
2. $\mathsf{E}^M(\Omega) = I$;
3. $\mathsf{E}^M(\cup_{i \in I} X_i) = \sum_i \mathsf{E}^M(X_i)$ for any sequence of disjoint sets $\{X_i\}_{i \in I}$ in Σ, where the series on the right converges in the weak operator topology.

Such a map is called a *(normalised) positive operator-valued measure* or *POVM*, and from now on we will use the terms "POVM" and "observable" interchangeably. In particular, instead of referring to a measurement M we will consider its POVM E^M.

Assuming the measurable space (Ω, Σ), we shall cover two particular examples of observables defined on the Hilbert space $\mathcal{H}$; namely, sharp and discrete observables. Sharp, or spectral, observables are given by projection-valued measures (PVMs). These are POVMs $\mathsf{P} : \Sigma \to \mathcal{P}(\mathcal{H})$ for which $\mathsf{P}(A)\mathsf{P}(B) = \mathsf{P}(A \cap B)$ for $A, B \in \Sigma$. In other words, P maps disjoint elements of Σ to orthogonal projections. For a spectral measure $\mathsf{P} : \mathcal{B}(\mathbb{R}) \to \mathcal{E}(\mathcal{H})$, we can make use of the measure $dp_\varphi(\lambda) = d\langle\varphi|\mathsf{P}(\lambda)\varphi\rangle$ for any $\varphi \in \mathcal{H}$ to define a unique self-adjoint operator T via

$$\langle\varphi|T\varphi\rangle = \int_{\mathbb{R}} \lambda\, d\langle\varphi|\mathsf{P}(\lambda)\varphi\rangle. \tag{2.130}$$

Note that T may be unbounded, in which case its domain is composed of pure states $\varphi \in \mathcal{H}$ for which $\int_{\mathbb{R}} \lambda^2\, d\langle\varphi|\mathsf{P}(\lambda)\varphi\rangle < \infty$. The converse of this statement is a significant result:

Theorem 2.4.1. For a self-adjoint operator T with domain $\mathcal{D}(T) \subset \mathcal{H}$, there exists a unique spectral measure $\mathsf{E}^T : \mathcal{B}(\mathbb{R}) \to \mathcal{E}(\mathcal{H})$ such that

$$\mathcal{D}(T) = \left\{\varphi \in \mathcal{H} \middle| \int_{\mathbb{R}} \lambda^2\, d\langle\varphi|\mathsf{E}^T(\lambda)\varphi\rangle < \infty\right\}, \tag{2.131}$$

and for any $\varphi \in \mathcal{D}(T)$ Equation (2.130) holds.

This result, known as the *spectral decomposition* of T, allows us to consider the PVM associated with T, and express T in the form

$$T = \int_{\mathbb{R}} \lambda \, d\mathsf{E}^T(\lambda), \tag{2.132}$$

where we are implicitly understanding this in terms of Equation (2.130). Furthermore, if we suppose that f is a real-valued measurable function defined on the support of E^T, then there exists a unique self-adjoint operator $f(T)$ of the form

$$f(T) = \int_{\mathbb{R}} f(\lambda) d\mathsf{E}^T(\lambda). \tag{2.133}$$

Similar to the spectral decomposition, for a POVM $\mathsf{E} : \mathcal{B}(\mathbb{R}) \to \mathcal{E}(\mathcal{H})$ we can define the Hermitian *first moment* operator

$$\mathsf{E}[1] = \int_{\mathbb{R}} \lambda d\mathsf{E}(\lambda). \tag{2.134}$$

However, since we no longer assume that E is projection-valued, we will find that $f(\mathsf{E}[1])$ is not of the form given in Equation (2.133); indeed, the k^{th} moment operator $\mathsf{E}[k] = \int_{\mathbb{R}} \lambda^k d\mathsf{E}(\lambda)$ is in general not equal to $\mathsf{E}[1]^k$.

Finally, for an unbounded self-adjoint operator there exists a similar spectral decomposition: for any unbounded self-adjoint operator A with domain $\mathcal{D}(A)$ dense in $\mathcal{H}$, then for any vector $\varphi \in \mathcal{D}(A)$ there exists a unique PVM E^A for which

$$\mathrm{tr}\left[P_\varphi A\right] = \int_{\mathbb{R}} \lambda \mathrm{tr}\left[P_\varphi d\mathsf{E}^A(\lambda)\right]. \tag{2.135}$$

In particular, for the position and momentum operators Q and P we can define the PVMs E^Q and E^P. These spectral measures correspond to the eigenbases $\{|q\rangle\}$ and $\{|p\rangle\}$, respectively, but a word of warning is needed here: despite providing representations for the Hilbert space $L^2(\mathbb{R})$ via $\psi(q) = \langle q|\psi\rangle$ and $\widetilde{\psi}(p) = \langle p|\psi\rangle$, etc., these states do not belong to $L^2(\mathbb{R})$. To see this, let ψ be an eigenvector of Q with eigenvalue x_0. In which case

$$x_0\psi(q) = x_0 \langle q|\psi\rangle = \langle q|Q\psi\rangle = \langle Qq|\psi\rangle = q\psi(q), \tag{2.136}$$

for all $q \in \mathbb{R}$, and so $\psi(q)$ is proportional to $\delta(q - x_0) \notin L^2(\mathbb{R})$. However, these pseudo-eigenbases do serve a purpose, and we will utilise them, all the while bearing in mind their shortcomings. In particular, they make sense when we wish to calculate the value

$$\langle\psi|Q\psi\rangle = \int_{\mathbb{R}} q \left|\langle\psi|q\rangle\right|^2 dq = \int_{\mathbb{R}} q \left|\psi(q)\right|^2 dq, \tag{2.137}$$

which is a well defined quantity, known as the *expectation value* of Q with respect to the state ψ. The quantity $|\psi(X)|^2$, $X \in \mathcal{B}(\mathbb{R})$ is the probability $p_\psi^Q(X)$, and so $\langle\psi|Q\psi\rangle = \int_{\mathbb{R}} q dp_\psi^Q(q)$ is the mean value of Q when measured in the state ψ. Note that at times we will use the shorthand $\langle A\rangle_\psi$ to denote the expectation value $\langle\psi|A\psi\rangle$. Similarly to the bounded case, we may consider $f(A)$ for an unbounded real-valued function f, but we must be careful as the domain depends on f.

From any sharp observable $\mathsf{P} : \mathcal{B}(\mathbb{R}) \to \mathcal{E}(\mathcal{H})$ we can form a POVM E by *smearing*. In general, we would express the smearing of P into E in the form

$$\mathsf{E}(X) = \int_{\mathbb{R}} k(\lambda, X) d\mathsf{P}(\lambda), \tag{2.138}$$

where $k : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to [0, 1]$, known as a *Markov kernel*, is such that $k(\lambda, \cdot)$ is a probability measure and $k(\cdot, X)$ is a measurable function. Conceptually, k is responsible for adding an additional uncertainty into the measurement outcomes of E compared to the measurement of P.

There exists another way in which we can connect POVMs and PVMs. Suppose that P is a PVM on $\mathcal{H}'$ with outcome space (Ω, Σ) and an isometry $V : \mathcal{H} \to \mathcal{H}'$ with $\mathcal{H} \subset \mathcal{H}'$. In which case, the map $\mathsf{E} : \Sigma \to \mathcal{E}(\mathcal{H})$

$$\mathsf{E}(X) = V^* \mathsf{P}(X) V, \tag{2.139}$$

for all $X \in \Sigma$, is clearly a POVM on $\mathcal{H}$ with outcome space (Ω, Σ). It is a result of Naimark's, known as *Naimark's dilation theorem*, that for any POVM $\mathsf{E} : \Sigma \to \mathcal{E}(\mathcal{H})$, there exists a Hilbert space $\mathcal{H}' \supset \mathcal{H}$, a linear isometry $V : \mathcal{H} \to \mathcal{H}'$ and a PVM $\mathsf{P} : \Sigma \to \mathcal{E}(\mathcal{H}')$ satisfying Equation (2.139) [38]. A special example of a Naimark dilation is the measurement model, which we shall discuss in Section 2.4.3.

Discrete observables are POVMs for which the set Ω is finite, i.e., $\Omega = \{x_1, \dots, x_n\}$. In such cases, we tend not to consider the σ-algebra and instead simply focus on the power set 2^{Ω}. Considering the discrete POVM $\mathsf{E} : \{x_1, \dots, x_n\} \to \mathcal{E}(\mathcal{H})$, we relabel the effects $\mathsf{E}(x_i) =: \mathsf{E}(i)$, and so we have $I = \mathsf{E}(\Omega) = \sum_i \mathsf{E}(i)$. It is common in the literature to denote a discrete POVM E by its range, i.e., the set $\{\mathsf{E}(i)\}_{i=1}^{n}$. However, in what follows we shall restrict ourselves to treating a POVM as a map from a σ-algebra to the set of effects defined on the Hilbert space. In the case of a discrete spectral observable P with range $\{P_i = \mathsf{P}(i)\}$, we tend to express its associated self-adjoint operator A in terms of its eigenvalues $\{a_i\}$, i.e., $A = \sum_i a_i P_i$.

As we have stated above, we associate observables with POVMs. This is in contrast with the traditional view of quantum mechanics, in which self-adjoint operators correspond to observables, with measurement outcomes given by their eigenvalues, and their associated spectral measure determining the probability of outcomes occurring. Whilst this restriction to PVMs is already too limited, this viewpoint also possesses conceptual difficulties. Within a given experimental setup, a single measurement provides an outcome, and from repeated measurements we derive probability distributions allowing us to infer the likelihood of a given outcome depending on what state we prepare the system in. At this point what we are *observing* is how an input state alters the probability distributions we find within this setup, with this being described mathematically by POVMs. By comparison, what self-adjoint operators tell us is how the *average* of the distributions vary with the state the system is prepared in, which does not give us the full story of what is happening in the experiment.

Consider a measurable space (Ω, Σ) and group G with a group action $\alpha : G \times \Omega \to \Omega$,

$\alpha(g,x) := \alpha_g(x)$, such that Ω is a homogeneous G-space, i.e., $\alpha_{gh} = \alpha_g\alpha_h$ for any two $g,h \in G$, $\alpha_e = \iota_\Omega$ where e is the identity element of G and ι_Ω is the identity map on Ω, and for any two elements $\omega,\omega' \in \Omega$ there exists a $g \in G$ such that $\omega' = \alpha_g(\omega)$. An observable $\mathsf{E} : \Sigma \to \mathcal{E}(\mathcal{H})$ is *covariant* with respect to the group G if there exists a pair (E, U), known as a *system of covariance*, with $U : G \to \mathcal{U}(\mathcal{H})$ a unitary representation of G such that

$$U_g\mathsf{E}(Y)U_g^* = \mathsf{E}(\alpha_g^{-1}(Y)) \tag{2.140}$$

for all $g \in G$ and $Y \in \Sigma$. This action can be expressed in terms of the commutative diagram:

$$\begin{tikzcd}
\Omega \arrow[r, "\alpha_g"] \arrow[d, "\mathsf{E}"'] & \Omega \arrow[d, "\mathsf{E}"] \\
\mathcal{E}(\mathcal{H}) \arrow[r, "U_g"'] & \mathcal{E}(\mathcal{H})
\end{tikzcd}$$

2.4.3 Measurement models

Measurement models provide a more in-depth mathematical description of the measurement process compared to working with just a POVM. The idea behind them is that we do not measure a quantum system directly, but rather we couple them to an apparatus, such as a meter or dial, from which we read out a result and infer the system's measurement result. As an example, consider a double-slit experiment; we know the final location of the photons or electrons being measured by their location on the detecting plate, which we measure by direct observation.

Consider a quantum system described by the Hilbert space $\mathcal{H}$, whose state is given by the density operator $\rho \in \mathcal{S}(\mathcal{H})$. We then couple this system to the measuring apparatus $\mathcal{A}$, called the *probe*, with an associated Hilbert space $\mathcal{H}_\mathcal{A}$ and state $\sigma \in \mathcal{S}(\mathcal{H}_\mathcal{A})$. This coupling is performed via a channel—that is, a linear completely positive[5] trace-preserving map—$\mathcal{V} : \mathcal{T}(\mathcal{H}\otimes\mathcal{H}_\mathcal{A}) \to \mathcal{T}(\mathcal{H}\otimes\mathcal{H}_\mathcal{A})$, and the measurement on the probe is given by the observable Z with outcome space $(\Omega_\mathcal{A}, \Sigma_\mathcal{A})$. In order to accommodate the possibility that the pointer and the system possess different "scales", we define a *pointer function* $f : \Omega_\mathcal{A} \to \Omega$, which is both bijective and measurable. This measurement scheme, denoted by the quintuple $\mathcal{M} = \langle \mathcal{H}_\mathcal{A}, \sigma, \mathcal{V}, \mathsf{Z}, f\rangle$, produces an observable E with outcome space (Ω, Σ) via

$$\mathrm{tr}\left[\rho\mathsf{E}(X)\right] = \mathrm{tr}\left[\mathcal{V}(\rho\otimes\sigma)(I\otimes\mathsf{Z}(f^{-1}(X)))\right], \tag{2.141}$$

for all $\rho \in \mathcal{S}(\mathcal{H})$ and $X \in \Sigma$. We can rearrange this to find a form for E:

$$\begin{aligned}
\mathrm{tr}\left[\rho\mathsf{E}(X)\right] &= \mathrm{tr}\left[\mathcal{V}(\rho\otimes\sigma)(I\otimes\mathsf{Z}(f^{-1}(X)))\right] \\
&= \mathrm{tr}\left[(\rho\otimes\sigma)\mathcal{V}^*((I\otimes\mathsf{Z}(f^{-1}(X))))\right] \\
&= \mathrm{tr}\left[(\rho\otimes I)(I\otimes\sigma)\mathcal{V}^*((I\otimes\mathsf{Z}(f^{-1}(X))))\right] \\
&= \mathrm{tr}\left[\rho\,\mathrm{tr}_{\mathcal{H}_\mathcal{A}}\left[(I\otimes\sigma)\mathcal{V}^*(I\otimes\mathsf{Z}(f^{-1}(X)))\right]\right],
\end{aligned} \tag{2.142}$$

[5] A map $\mathcal{V} : \mathcal{T}(\mathcal{H}) \to \mathcal{T}(\mathcal{H})$ is completely positive if the map $\mathcal{V}\otimes I_{\mathcal{H}'} : \mathcal{T}(\mathcal{H}\otimes\mathcal{H}') \to \mathcal{T}(\mathcal{H}\otimes\mathcal{H}')$ is positive for any finite-dimensional additional Hilbert space $\mathcal{H}'$.

where $\mathcal{V}^*$ denotes the dual channel $\mathcal{V}^* : \mathcal{L}_s(\mathcal{H} \otimes \mathcal{H}_{\mathcal{A}}) \to \mathcal{L}_s(\mathcal{H} \otimes \mathcal{H}_{\mathcal{A}})$ defined by

$$\mathrm{tr}\left[\mathcal{V}(T)E\right] = \mathrm{tr}\left[T\,\mathcal{V}^*(E)\right]. \tag{2.143}$$

This map is necessarily completely positive and unital, i.e., $\mathcal{V}^*(I_{\mathcal{H}} \otimes I_{\mathcal{H}_{\mathcal{A}}}) = I_{\mathcal{H}} \otimes I_{\mathcal{H}_{\mathcal{A}}}$. Since $(\mathcal{V} \otimes I)^* = \mathcal{V}^* \otimes I$, the complete positivity of $\mathcal{V}^*$ is equivalent to the complete positivity of $\mathcal{V}$. From this consideration, we may express the observable E as:

$$\mathsf{E}(\cdot) = \mathrm{tr}_{\mathcal{H}_{\mathcal{A}}}\left[(I \otimes \sigma)\mathcal{V}^*(I \otimes \mathsf{Z}(f^{-1}(\cdot)))\right], \tag{2.144}$$

Every observable defined on $\mathcal{H}$ defines a class of measurement models that satisfy Equation (2.144), and it is a result of Ozawa's [39] that, for every observable on $\mathcal{H}$, there exists a measurement model such that σ is a pure state, $\mathcal{V}$ is a unitary channel and Z is a sharp observable. In other words, for any POVM E we can find a measurement model $\mathcal{M} = \langle \mathcal{H}_{\mathcal{A}}, P_{\varphi}, U, \mathsf{Z}, f \rangle$ such that

$$\mathsf{E}(\cdot) = \mathrm{tr}_{\mathcal{H}_{\mathcal{A}}}\left[(I \otimes P_{\varphi})U^*(I \otimes \mathsf{Z}(f^{-1}(\cdot)))U\right]. \tag{2.145}$$

With this in mind, when we come to discuss measurement models in Chapter 4, and we encounter models where the coupling is provided by a unitary channel and the probes are measured by sharp observables, we will not be restricting the range of measurement models that we consider.

2.4.4 Joint and sequential measurements

As will feature often in later chapters, we may wish to measure more than a single observable in a given measurement setup. This leads to the concepts of joint and sequential measurements. There is already extensive work on these topics in the literature [15, 11, 41], so we shall only briefly highlight the aspects that are of use to us.

Two observables $\mathsf{E} : \Sigma_1 \to \mathcal{E}(\mathcal{H})$ and $\mathsf{F} : \Sigma_2 \to \mathcal{E}(\mathcal{H})$ with measurable spaces (Ω_1, Σ_1), (Ω_2, Σ_2), respectively, are jointly measurable if there exists an observable $\mathsf{J} : \Sigma_1 \times \Sigma_2 \to \mathcal{E}(\mathcal{H})$ such that

$$\mathsf{E}(X) = \mathsf{J}(X \times \Omega_2), \quad \mathsf{F}(Y) = \mathsf{J}(\Omega_1 \times Y), \tag{2.146}$$

for any $X \in \Sigma_1$ and $Y \in \Sigma_2$. Such an observable J is called a *joint observable* of E and F, and conversely E and F are called *marginal observables* (or simply margins) of J. A simple example of a joint observable arises when E and F commute; that is, when any pair of effects $\mathsf{E}(X)$ and $\mathsf{F}(Y)$ satisfy $[\mathsf{E}(X), \mathsf{F}(Y)] = 0$. In this case, a valid joint observable for E and F is

$$\mathsf{J}(X \times Y) = \mathsf{E}(X)\mathsf{F}(Y), \tag{2.147}$$

where the commutativity is required in order to guarantee that $\mathsf{J}(X \times Y)$ is self-adjoint for any $X \in \Sigma_1$ and $Y \in \Sigma_2$. This highlights an important point: if two observables commute, then they are jointly measurable. Whilst this is a sufficient condition to guarantee joint measurability, it is not necessary unless at least one of the observables considered is sharp.

As has been alluded to above, quantum mechanics allows for the existence of observables that cannot be measured jointly, as opposed to the case in classical mechanics where any pair of observables are jointly measurable. We call such observables *incompatible*, and for two incompatible observables we are unable to build an experiment such that we can acquire information about both arbitrarily well. An alternative way of representing this is by showing that a sequential measurement of one observable followed by a second disturbs the measurement statistics of the second observable.

In order to discuss sequential measurements, we must first mention *instruments* (see [29, p. 226-232], [11, p. 37-39] or [26, p. 17-19] for details). Assuming that we have performed a repeatable/nondestructive measurement scheme $\mathcal{M}$ (detecting photons on a photographic plate counts as an example of a non-repeatable measurement) associated with an observable E, we expect that the state of the system is changed in some way depending on the outcome of the measurement. Suppose that we have performed a measurement of the observable E on the state ρ via the measurement scheme $\mathcal{M} = \langle \mathcal{H}_{\mathcal{A}}, \sigma, \mathcal{V}, \mathsf{Z}, f \rangle$ and measured the outcome $X \in \Sigma$. By expanding Equation (2.141), we see that

$$\begin{aligned} \mathrm{tr}\left[\mathsf{E}(X)\rho\right] &= \mathrm{tr}\left[\mathcal{V}(\rho \otimes \sigma)(I \otimes \mathsf{Z}(f^{-1}(X)))\right] \\ &= \mathrm{tr}\left[\mathrm{tr}_{\mathcal{H}_{\mathcal{A}}}\left[\mathcal{V}(\rho \otimes \sigma)(I \otimes \mathsf{Z}(f^{-1}(X)))\right] I\right] \\ &= \mathrm{tr}\left[\mathcal{I}^{\mathcal{M}}_X(\rho)\right], \end{aligned} \tag{2.148}$$

where $\mathcal{I}^{\mathcal{M}}_X(\rho) = \mathrm{tr}_{\mathcal{H}_{\mathcal{A}}}\left[\mathcal{V}(\rho \otimes \sigma)(I \otimes \mathsf{Z}(f^{-1}(X)))\right]$ describes the evolution of the state after obtaining the measurement outcome $X \in \Sigma$. In general, we give the change of the state of the system after a measurement described by the scheme $\mathcal{M}$ via the map $\mathcal{I}^{\mathcal{M}} : \Sigma \to \mathcal{L}(\mathcal{T}(\mathcal{H}))$, known as an *instrument*, with $\mathcal{I}^{\mathcal{M}}_X := \mathcal{I}^{\mathcal{M}}(X) : \mathcal{T}(\mathcal{H}) \to \mathcal{T}(\mathcal{H})$ for $X \in \Sigma$. An instrument must satisfy the following properties:

(i) For each $X \in \Sigma$, $\mathcal{I}^{\mathcal{M}}_X$ is linear, completely positive and trace nonincreasing;

(ii) $\mathrm{tr}\left[\mathcal{I}^{\mathcal{M}}_\Omega(\rho)\right] = 1$ and $\mathcal{I}^{\mathcal{M}}_\emptyset(\rho) = O$ for all $\rho \in \mathcal{S}(\mathcal{H})$;

(iii) If $\{X_j\} \subset \Sigma$ is a disjoint sequence, then for any $\rho \in \mathcal{S}(\mathcal{H})$

$$\mathrm{tr}\left[\mathcal{I}^{\mathcal{M}}_{\cup_j X_j}(\rho)\right] = \sum_j \mathrm{tr}\left[\mathcal{I}^{\mathcal{M}}_{X_j}(\rho)\right]. \tag{2.149}$$

For each measurement scheme $\mathcal{M}$ we may define an instrument $\mathcal{I}^{\mathcal{M}}$. Indeed, for every instrument $\mathcal{I}$ there exists a measurement scheme $\mathcal{M}$ such that $\mathcal{I} = \mathcal{I}^{\mathcal{M}}$, hence there is a correspondence between measurement schemes and instruments, which is many-to-one. Physically, this means that there exist a number of ways of setting up a measurement that will affect the state in the same way. If we define the dual of the instrument via

$$\mathrm{tr}\left[\mathcal{I}_X(\rho)A\right] = \mathrm{tr}\left[\rho\,\mathcal{I}^*_X(A)\right] \tag{2.150}$$

for any $\rho \in \mathcal{S}(\mathcal{H})$ and $A \in \mathcal{L}(\mathcal{H})$, then if we set $A = I$ then we can retrieve the observable E using the relation $\mathcal{I}^*_X(I) = \mathsf{E}(X)$. Hence, each instrument defines an observable. Similarly to the correspondence between measurement schemes and instruments, the correspondence between instruments and observables is many to one, which expresses the idea that we can measure an observable in different ways, but the output state will be different depending on which method we choose.

Since, for an instrument $\mathcal{I}$ we find that $\operatorname{tr}[\mathcal{I}_X(\rho)] = \operatorname{tr}[\mathsf{E}(X)\rho]$ for some observable E and any state $\rho \in \mathcal{S}(\mathcal{H})$, we will often find that $\mathcal{I}$ maps states to subnormalised states. With this in mind, assuming $\operatorname{tr}[\mathsf{E}(X)\rho] \neq 0$, we define the *conditional state*

$$\widetilde{\rho}_X = \frac{\mathcal{I}_X(\rho)}{\operatorname{tr}[\mathcal{I}_X(\rho)]}. \tag{2.151}$$

With this in mind, we can now consider a sequential measurement. Suppose that we first perform $\mathsf{E} : \Sigma_1 \to \mathcal{E}(\mathcal{H})$ and measure the outcome $X \in \Sigma_1$, leaving the system in the conditional state $\widetilde{\rho}_X$, and then measure $\mathsf{F} : \Sigma_2 \to \mathcal{E}(\mathcal{H})$. The conditional probability that we get the measurement outcome belonging to $Y \in \Sigma_2$ given that we first received the outcome belonging to $X \in \Sigma_1$ is given by

$$p_\rho(\mathsf{F} \in Y | \mathsf{E} \in X) = \operatorname{tr}[\widetilde{\rho}_X \mathsf{F}(Y)] = \frac{1}{\operatorname{tr}[\mathcal{I}_X(\rho)]} \operatorname{tr}[\mathcal{I}_X(\rho)\mathsf{F}(Y)]. \tag{2.152}$$

By making use of the dual instrument and Bayes' theorem, we find the joint probability to be

$$p_\rho(\mathsf{F} \in Y \& \mathsf{E} \in X) = p_\rho(\mathsf{F} \in Y | \mathsf{E} \in X) p_\rho(\mathsf{E} \in X) = \operatorname{tr}[\rho \mathcal{I}^*_X(\mathsf{F}(Y))]. \tag{2.153}$$

In doing so, we have defined a joint observable

$$\mathsf{J}(X \times Y) = \mathcal{I}^*_X(\mathsf{F}(Y)), \tag{2.154}$$

which has margins

$$\begin{aligned} \mathsf{E}'(X) &= \mathsf{J}(X \times \Omega_2) = \mathcal{I}^*_X(\mathsf{F}(\Omega_2)) = \mathcal{I}^*_X(I) = \mathsf{E}(X), \\ \mathsf{F}'(Y) &= \mathsf{J}(\Omega_1 \times Y) = \mathcal{I}^*_{\Omega_1}(\mathsf{F}(Y)). \end{aligned} \tag{2.155}$$

In other words, if E and F are incompatible, then we can build a joint measurement such that one can be perfectly measured, and we are left with some approximation of the second. In this sense, we may consider sequential measurements to be a special case of joint measurements, and in what follows we shall do as much.

2.4.5 Error measures within quantum theory

Throughout we will be addressing the problem of approximating sharp observables, commonly denoted by A or B, via alternative (usually unsharp) observables, denoted by C or D. In such instances we require some means of determining how faithful an approximation C (or D) is to A (B). Such a measure should be non-negative, with a value of zero implying that the approximating observable faithfully approximates the sharp observable in some sense determined by the measure. There exist several such measures, which are collectively referred to as *error measures*.

The two error measures that we discuss here have both been provided as extensions, for the case of quantum observables, to the root-mean square deviation used in statistics. In both cases, we suppose that an observable C is being used to approximate an ideal measurement A with associated first moment operator $A = \mathsf{A}[1]$.

The first measure that we discuss was introduced by Ozawa [40] and is a state-dependent error. Suppose that our approximate measurement $\mathsf{C} : \mathcal{B}(\mathbb{R}) \to \mathcal{E}(\mathcal{H})$ is given by the measurement scheme $\langle \mathcal{K}, \xi, U, \mathsf{Z}, \iota \rangle$, where $\xi \in \mathcal{K}$ is a pure state, $\mathsf{Z} : \mathcal{B}(\mathbb{R}) \to \mathcal{E}(\mathcal{K})$ is a sharp observable and the pointer function is the identity map $\iota : \mathbb{R} \to \mathbb{R}$, then the error of C as an approximation of A with respect to the pure state $\psi \in \mathcal{H}$ is given by

$$\varepsilon(\mathsf{C}, \mathsf{A}, \psi)^2 = \left\langle \psi \otimes \xi \middle| (U^*(I \otimes \mathsf{Z}[1])U - A \otimes I)^2 \psi \otimes \xi \right\rangle . \tag{2.156}$$

Following Arthurs and Goodman [2], by defining the *noise operator*

$$N_A := U^*(I \otimes \mathsf{Z}[1])U - A \otimes I, \tag{2.157}$$

we see that $\varepsilon(\mathsf{C}, \mathsf{A}, \psi)^2 = \left\langle N_A^2 \right\rangle_{\psi \otimes \xi}$, and so we will henceforth refer to this measure as the *noise measure.*

As may be inferred, this measure arises from considering self-adjoint operators as observables, but we may express it more operationally (see [12, Appendix A]):

$$\begin{aligned} \varepsilon(\mathsf{C}, \mathsf{A}, \psi)^2 &= \left\langle \psi \middle| (\mathsf{C}[2] - \mathsf{C}[1]A - A\mathsf{C}[1] + A^2)\psi \right\rangle \\ &= \left\langle \psi \middle| (\mathsf{C}[1] - A)^2 \psi \right\rangle + \left\langle \psi \middle| (\mathsf{C}[2] - \mathsf{C}[1]^2)\psi \right\rangle . \end{aligned} \tag{2.158}$$

The first term in this expression can be perceived as a measure of the relative noise between C and A, whilst the second term is a measure of the intrinsic noise of C, as it is zero iff C is a sharp observable [13]. This first term also highlights that this is an example of an error based on *value deviation* between the two observables. Note that in order for $\mathsf{C}[1] - A$ to possess any operational meaning we require $\mathsf{C}[1]$ and A to commute, otherwise we are unable to measure them within the same measurement scheme. In the case of $\mathsf{C}[1]$ and A commuting we can express $\varepsilon(\mathsf{C}, \mathsf{A}, \psi)^2$ as

$$\begin{aligned} \varepsilon(\mathsf{C}, \mathsf{A}, \psi)^2 &= \left\langle \psi \middle| (\mathsf{C}[2] - 2\mathsf{C}[1]A + A^2)\psi \right\rangle \\ &= \int_{\mathbb{R}^2} (x - y)^2 \left\langle \psi | d\mathsf{C}(x) d\mathsf{A}(y) \psi \right\rangle . \end{aligned} \tag{2.159}$$

In this form we see how $\varepsilon(\mathsf{C}, A, \psi)^2$ forms the mean value of the deviation of the random variables x and y with respect to the probability bi-measure $p(dx, dy) := \langle \psi | d\mathsf{C}(x) d\mathsf{A}(y) \psi \rangle$. Alternatively, if the state ψ is an eigenvector of A with eigenvalue a, then

$$\varepsilon(\mathsf{C}, A, \psi) = \left(\int_{\mathbb{R}} (x - a)^2 \left\langle \psi | d\mathsf{C}(x) \psi \right\rangle \right)^{1/2} , \tag{2.160}$$

and thus ε reduces to the standard deviation from the value a with respect to the probability measure $p(dx) = \langle \psi | d\mathsf{C}(x) \psi \rangle$, as would be classically used to determine the error of an approximate observable. However, these two instances are rare cases, and in general there are issues with measuring the noise error.

The second measure, given by Busch, Lahti and Werner [16], is state-independent, unlike Ozawa's measure, and is based on the Wasserstein 2-distance (an extension of the Monge-Kantorovich "earth mover's" distance, which is discussed in detail in [53]) and in what follows we shall refer to this as the *BLW error* for the sake of brevity.

For a pair of (possibly incompatible) observables $\mathsf{C}, \mathsf{A} : \mathcal{B}(\mathbb{R}) \to \mathcal{E}(\mathcal{H})$ (with C approximating A) and a state $\rho \in \mathcal{S}(\mathcal{H})$ we define a *coupling* $\gamma : \mathbb{R} \times \mathbb{R} \to [0, 1]$ to be a probability measure such that

$$\gamma(X \times \mathbb{R}) = p_\rho^{\mathsf{C}}(X), \qquad \gamma(\mathbb{R} \times Y) = p_\rho^{\mathsf{A}}(Y). \tag{2.161}$$

With this definition, the Wasserstein 2-distance $\Delta_\rho(\mathsf{C}, \mathsf{A})^2$ between two observables with respect to a state ρ is given by

$$\Delta_\rho(\mathsf{C}, \mathsf{A})^2 = \inf_{\gamma \in \Gamma(\mathsf{C},\mathsf{A})} \int_{\mathbb{R}^2} (x - y)^2 d\gamma(x, y), \tag{2.162}$$

where $\Gamma(\mathsf{C}, \mathsf{A})$ is the space of all couplings between p_ρ^{C} and p_ρ^{A}. Note that in the case that, again, ρ is an eigenfunction of $A = \mathsf{A}[1]$ with eigenvalue a, then p_ρ^{A} reduces to the point measure δ_a, so $\gamma = p_\rho^{\mathsf{C}} \times \delta_a$, and thus

$$\Delta_\rho(\mathsf{C}, \mathsf{A}) = \left(\int_{\mathbb{R}} (x - a)^2 dp_\rho^{\mathsf{C}}(x) \right)^{1/2}. \tag{2.163}$$

In other words, we again arrive at the standard deviation in the case that the state we consider is an eigenfunction of A. The state-dependent measure Δ_ρ gives a distance between the two distributions p_ρ^{C} and p_ρ^{A} with regards to a best case scenario, thereby seeing what the smallest possible difference is between the two distributions while still requiring that the distributions can be compared. It should be noted that, unlike the noise measure, this scheme does not require that we are measuring both observables at the same time and comparing values; rather their distributions are compared from separate measurements, with the proviso that the distributions are still comparable. We then define the error $\Delta(\mathsf{C}, \mathsf{A})$ to be the worst case scenario over all states:

$$\Delta(\mathsf{C}, \mathsf{A}) := \sup_\rho \Delta_\rho(\mathsf{C}, \mathsf{A}). \tag{2.164}$$

In essence, this defines an upper limit on the possible error one may find whilst measuring C as an approximation of A over all states. In other words, if someone were to perform C as an approximation of A in any state $\rho \in \mathcal{S}(\mathcal{H})$, then the error in the approximation would be no greater than $\Delta(\mathsf{C}, \mathsf{A})$.

Chapter 3

An Operational Link between SIC-POVMs and MUBs

In recent years two classes of observables have gained significance, particularly in terms of quantum state determination: symmetric informationally complete POVMs (SIC-POVMs) and mutually unbiased PVMs, or alternatively mutually unbiased bases (MUBs). A family of MUBs represents a family of observables as incompatible as is possible, and yet it is possible to show that they can, in some instances, be related to the marginal observables of a SIC-POVM, which are by their very definition compatible observables. In this chapter we will present the construction that allows for this, and give examples found via Mathematica showing how it works for low dimensions.

3.1 Preliminaries

We begin by detailing the quantum information theoretic constructs that we will be investigating, namely SIC-POVMs and MUBs. We also provide an overview of the combinatorial concept of mutually orthogonal Latin squares, which, as we will see in Section 3.2.2, are linked to our connection between SIC-POVMs and MUBs.

3.1.1 Mutually Unbiased Bases (MUBs)

For this section we assume that we are dealing with a finite d-dimensional Hilbert space, i.e., $\mathcal{H}_d := \mathbb{C}^d$. For such a Hilbert space, two bases $\{\psi_i\}_{i=1}^d$ and $\{\varphi_j\}_{j=1}^d$ are said to be *mutually unbiased* if the magnitude of the inner product between any two elements of these bases is fixed:

$$\left|\langle\psi_i|\varphi_j\rangle\right|^2 = c \quad \forall\, i, j, \tag{3.1}$$

where c is a constant. Given the normalisation of the basis vectors, we quickly determine the value of c:

$$1 = |\psi_i|^2 = \sum_j \left|\langle\psi_i|\varphi_j\rangle\right|^2 = dc \quad \Rightarrow \quad c = \frac{1}{d}. \tag{3.2}$$

For a Hilbert space of dimension d, there cannot exist more than $d+1$ mutually unbiased bases. This can be seen by considering the d^2-dimensional state space of $\mathcal{H}_d$, $\mathcal{S}(\mathcal{H}_d)$. Taking into account the unit-trace property, each state is determined by $d^2 - 1$

linearly independent operators in the space of traceless self-adjoint operators. Each MUB $\{\psi_i^{(m)}\}$ defines a collection of $d-1$ linearly independent operators $\{P_i^{(m)} - (1/d)I\}_{i=0}^{d-1}$, where $P_i^{(m)} = \left|\psi_i^{(m)}\right\rangle\left\langle\psi_i^{(m)}\right|$, which therefore span a $d-1$-dimensional subspace of the space of traceless self-adjoint operators. From the mutually unbiased condition, the $d-1$-dimensional subspaces associated with different MUBs are orthogonal, and so there can exist at most $(d^2-1)/(d-1) = d+1$ MUB bases.

In the case that d is a prime-power number, i.e., $d = p^n$ for some prime number p and natural number n, it is known that one can construct a family of $d+1$ MUBs. The proof goes as follows [32, 56]:

1. $d = 2$: In the Bloch representation, the projections $\mathsf{P}(\pm)$ associated with a given basis $\{\psi_\pm\}$ are described by their Bloch vector

$$\mathsf{P}(\pm) = |\psi_\pm\rangle\langle\psi_\pm| = \frac{1}{2}(1 \pm \boldsymbol{r}\cdot\boldsymbol{\sigma}), \tag{3.3}$$

where $\boldsymbol{r} \in S^2 = \{\boldsymbol{a} \in \mathbb{R}^3 |\, \|\boldsymbol{a}\| = 1\}$, and $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is the vector whose components are the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.4}$$

Another basis $\{\varphi_\pm\}$, with associated projections $\mathsf{Q}(\pm) = \frac{1}{2}(1 \pm \boldsymbol{s}\cdot\boldsymbol{\sigma})$, is mutually unbiased to $\{\psi_\pm\}$ iff

$$\frac{1}{2} = |\langle\psi_\pm|\varphi_\pm\rangle|^2 = \mathrm{tr}\,[\mathsf{P}(\pm)\mathsf{Q}(\pm)] = \frac{1}{2}(1 \pm \boldsymbol{r}\cdot\boldsymbol{s}), \tag{3.5}$$

where we have used the identity $(\boldsymbol{r}\cdot\boldsymbol{\sigma})(\boldsymbol{s}\cdot\boldsymbol{\sigma}) = (\boldsymbol{r}\cdot\boldsymbol{s})I + i(\boldsymbol{r}\times\boldsymbol{s})\cdot\boldsymbol{\sigma}$ and the traceless property of the Pauli matrices to arrive at the final equality. Clearly, in order to achieve equality, the vectors $\boldsymbol{r}$ and $\boldsymbol{s}$ must be orthogonal. Given the number of mutually orthogonal vectors is bounded above by the dimension of the space they exist in, we cannot find more than 3 such vectors. In other words, there cannot exist more than 3 mutually unbiased bases for $d = 2$.

2. d an odd prime: We begin by defining a "computational basis"

$$v^{(0)} = \left\{v_k^{(0)}\right\}_{k=0}^{d-1}, \tag{3.6}$$

where the ℓ^{th} component of $v_k^{(0)}$ is

$$\left(v_k^{(0)}\right)_\ell = \delta_{k\ell}. \tag{3.7}$$

We now define the d^{th} root of unity $\omega = e^{2\pi i/d}$, and the bases

$$v^{(r)} = \left\{v_k^{(r)}\right\}_{k=0}^{d-1}, \tag{3.8}$$

with $r = 1, \ldots, d$ via

$$\left(v_a^{(r)}\right)_\ell = \frac{1}{\sqrt{d}}\omega^{(r\ell^2 + a\ell)}. \tag{3.9}$$

The inner product of any two of these states is of the form

$$\left|\left\langle v_a^{(r)} \middle| v_b^{(s)} \right\rangle\right| = \left|\frac{1}{d}\sum_\ell \omega^{(s-r)\ell^2+(b-a)\ell}\right|. \tag{3.10}$$

If $r = s$, then (3.10) reduces to $\frac{1}{d}\sum_\ell \omega^{(b-a)\ell} = \delta_{ab}$ due to the orthogonality of the characters of the finite field $\mathbb{F}_d$, confirming that $v^{(r)}$ forms an orthonormal basis. If, on the other hand, $r \neq s$, then we note that for any $\ell, m \in \mathbb{F}_d$ with $\ell \neq m$, we may write $\ell = m + \alpha$, with α a nonzero element of $\mathbb{F}_d$. With this in mind, removing the factor of $1/d$, we calculate the square of the absolute quantity in Equation (3.10):

$$\begin{aligned}\left|\sum_\ell \omega^{(s-r)\ell^2+(b-a)\ell}\right|^2 &= \sum_{\ell,m} \omega^{(s-r)(\ell^2-m^2)+(b-a)(\ell-m)} \\ &= \sum_m \omega^{0(s-r)+0(b-a)} + \sum_m \sum_{\alpha>0} \omega^{(s-r)(\alpha^2+2\alpha m)+(b-a)\alpha} \\ &= d + \left(\sum_{\alpha>0} \omega^{(s-r)\alpha^2+(b-a)\alpha}\left(\sum_m \omega^{2\alpha(s-r)m}\right)\right).\end{aligned} \tag{3.11}$$

Since $r \neq s$ and $\alpha > 0$, the summand over m is equal to zero, and so we are left with just the first term. Hence,

$$\left|\sum_\ell \omega^{(s-r)\ell^2+(b-a)\ell}\right| = \sqrt{d}, \tag{3.12}$$

for all $a, b \in \mathbb{Z}_d$. Therefore, for $r \neq s$,

$$\left|\left\langle v_a^{(r)} \middle| v_b^{(s)} \right\rangle\right| = \left|\frac{1}{d}\sum_\ell \omega^{(s-r)\ell^2+(b-a)\ell}\right| = \frac{1}{\sqrt{d}}, \tag{3.13}$$

thus proving that these bases are mutually unbiased.

3. d a prime-power, i.e. $d = p^n$: This step is of the same form as the previous case. We start with the computational basis

$$v^{(0)} = \left\{v_k^{(0)}\right\}_{k=0}^{d-1}. \tag{3.14}$$

We now define $\omega = e^{2\pi i/p}$, and construct the d bases

$$v^{(r)} = \left\{v_k^{(r)}\right\}_{k=0}^{d-1}, \tag{3.15}$$

with $r = 1, \ldots, d$ via

$$\left(v_a^{(r)}\right)_\ell = \frac{1}{\sqrt{d}}\omega^{\mathrm{Tr}\left(r\ell^2 + a\ell\right)}, \tag{3.16}$$

where Tr () denotes the absolute trace as given in Definition 2.2.18 given in Section

2.2. From here, the proof is the same as for the previous case.

In the case of $d \neq p^n$, the existence of a complete set of MUBs is not known. Despite both algebraic [6] and numerical efforts [21], it has not yet been possible to construct a set of 4 MUBs for $d = 6$, let alone a complete set of 7. For the non-prime dimension d with prime factorisation $d = p_1^{n_1} p_2^{n_2} \dots p_r^{n_r}$, with the primes $p_1 \dots p_r$ increasing in value, there is at most a guarantee of there existing $p_1^{n_1} + 1$ MUBs by step 2 given above.

3.1.2 Symmetric Informationally Complete POVMs (SIC-POVMS)

Instead of relying on several bases for $\mathcal{H}_d$, we could see if there exists an observable acting on $\mathcal{S}(\mathcal{H}_d)$ that can tell any two states apart, i.e., a POVM G such that for any two states $\rho, \sigma \in \mathcal{S}(\mathcal{H}_d)$, where $\rho \neq \sigma$, there exists an effect $\mathsf{G}(k)$, say, where

$$\mathrm{tr}\left[\mathsf{G}(k)\rho\right] \neq \mathrm{tr}\left[\mathsf{G}(k)\sigma\right]. \tag{3.17}$$

Such an observable is said to be *informationally complete.* In order to satisfy this condition, we require the sample space of G to have at least d^2 elements. To prove this, suppose that G is an informationally complete observable defined on $\mathcal{H}_d$ with $n < d^2$ elements in its sample sample. The space of self-adjoint operators is d^2-dimensional, so there must exist a non-zero operator $T \notin \mathrm{span}(\{\mathsf{G}(i)\})$ satisfying $\mathrm{tr}\left[T\mathsf{G}(i)\right] = 0$ for all i. Since G is a POVM, $\sum_{i=1}^{n} \mathsf{G}(i) = I$ and so $\mathrm{tr}\left[T\right] = 0$. We define the state $\rho = (I + T/\left\|T\right\|)/d$ (this is indeed both positive and of unit trace) and see that for all i, $\mathrm{tr}\left[\rho\mathsf{G}(i)\right] = \mathrm{tr}\left[\mathsf{G}(i)I/d\right]$. In other words, G cannot distinguish ρ from I/d, contradicting its informational completeness. For the sake of simplicity, in what follows we restrict the sample space to exactly d^2 elements. We further demand that the effects of G are rank-one operators and so each of these effects corresponds to a multiple of a rank-one projection, i.e., $\mathsf{G}(i) = c_i P_i$ for some rank-one projection P_i. This restriction is again for ease of use, although there is work on categorising generalised SIC-POVMs [42].

We now simplify this POVM by demanding symmetry between the elements; that is, each effect has the same trace:

$$\mathrm{tr}\left[\mathsf{G}(i)\right] = c_i = \alpha \quad \forall\, i, \tag{3.18}$$

and for any two effects $\mathsf{G}(i)$ and $\mathsf{G}(j)$ the overlap is fixed:

$$\mathrm{tr}\left[\mathsf{G}(i)\mathsf{G}(j)\right] = \beta \quad \forall\, i, j \neq i. \tag{3.19}$$

An observable whose effects satisfy these conditions is described mathematically by a *symmetric informationally complete POVM (SIC-POVM).* The values α and β can be readily calculated: Firstly, given $\sum_i \mathsf{G}(i) = I$,

$$d = \mathrm{tr}\left[I\right] = \sum_i \mathrm{tr}\left[\mathsf{G}(i)\right] = d^2\,\alpha \quad \Rightarrow \quad \alpha = \frac{1}{d}. \tag{3.20}$$

From (3.20) we can find the value of β:

$$\frac{1}{d} = \text{tr}\,[\mathsf{G}(i)] = \sum_j \text{tr}\,[\mathsf{G}(i)\mathsf{G}(j)] = \frac{1}{d^2} + (d^2-1)\beta. \tag{3.21}$$

By rearranging for β we find

$$\beta = \frac{1}{d^2(d+1)}. \tag{3.22}$$

To confirm that G is indeed informationally complete, we require it to form a basis for $\mathcal{S}(\mathcal{H})$, that is, it must form a collection of d^2 linearly independent operators. Consider the d^2 operators

$$T_i = \sqrt{d(d+1)}\mathsf{G}(i) - \frac{1}{d^{3/2}}(\sqrt{d+1}-1)I. \tag{3.23}$$

The inner product of any two of these operators is

$$\begin{aligned}\text{tr}\,[T_iT_j] &= d(d+1)\text{tr}\,[\mathsf{G}(i)\mathsf{G}(j)] - \left(\frac{d+1}{d} - \frac{\sqrt{d+1}}{d}\right)\text{tr}\,[\mathsf{G}(i)+\mathsf{G}(j)] \\ &\quad + \left(\sqrt{\frac{d+1}{d^3}} - \frac{1}{d^{3/2}}\right)^2 \text{tr}\,[I] \\ &= \frac{d\delta_{ij}+1}{d} - \frac{2}{d^2}(d+1-\sqrt{d+1}) + \frac{1}{d^2}(d+2-2\sqrt{d+1}) \\ &= \delta_{ij}.\end{aligned} \tag{3.24}$$

These d^2 operators are mutually orthonormal, hence linearly independent, and so form a basis for $\mathcal{S}(\mathcal{H}_d)$. We can express the effects $\mathsf{G}(i)$ in terms of the T_i, i.e.,

$$\mathsf{G}(i) = aT_i + bI, \tag{3.25}$$

where

$$a = \frac{1}{\sqrt{d(d+1)}} \quad \text{and} \quad b = \frac{\sqrt{d+1}-1}{d^2\sqrt{d+1}}. \tag{3.26}$$

From this, we can see that they are linearly independent: Suppose that

$$\sum_i \alpha_i \mathsf{G}(i) = 0, \tag{3.27}$$

then by taking the trace of this we get that $\sum_i \alpha_i = 0$. If we expand Equation (3.27) in terms of Equation (3.25), multiply both sides by T_j and then take the trace, we find that

$$\begin{aligned}0 &= \sum_i \alpha_i\big(a\,\text{tr}\,[T_iT_j] + b\,\text{tr}\,[T_j]\big) = \sum_i \alpha_i\big(a\delta_{ij} + b/\sqrt{d}\big) \\ &= a\,\alpha_j + (b/\sqrt{d})\sum_i \alpha_i = a\,\alpha_j.\end{aligned} \tag{3.28}$$

In other words, $\alpha_j = 0$, and since this is an arbitrary coefficient in the sum, it must be that the sum equals zero iff each component is zero. The effects $\mathsf{G}(i)$ must therefore be a collection of d^2 linearly independent operators, and so form a (non-orthogonal) basis for $\mathcal{S}(\mathcal{H})$, which is to say that G must be an informationally complete observable.

It is conjectured [58] that there exists a SIC-POVM for every dimension $d \geq 2$, however currently the largest dimension shown to contain one is $d = 67$ (with numerical evidence for all dimensions less than 67) [48]. The most commonly considered type of SIC-POVM, and the type used for the investigations just discussed, are *Weyl-Heisenberg covariant SIC-POVMs*: Consider the computational basis for the Hilbert Space $\mathcal{H}_d$, $\{|n\rangle\}_{n=0}^{d-1}$, and define the *shift operator* X and the *phase operator* Z via

$$X|n\rangle = |n \oplus 1\rangle, \tag{3.29a}$$

$$Z|n\rangle = \omega^n |n\rangle, \tag{3.29b}$$

where, as before, $\omega = e^{2\pi i/d}$ and $\oplus$ denotes addition modulo d. Note that, unlike in the case of MUBs, we are not restricting d to being a prime power, and so the powers of ω are not an image of the prime field $\mathbb{F}_d$.

We define the set $W(d)$ as the set of all products of the phase and shift operators with a factor of ω:

$$W(d) = \{\omega^\alpha X^\beta Z^\gamma | \alpha, \beta, \gamma = 0, 1, \ldots, d-1\}. \tag{3.30}$$

From (3.29a) and (3.29b), we quickly see that $ZX = \omega XZ$, and so

$$\begin{aligned}(\omega^\alpha X^p Z^q)\left(\omega^\beta X^r Z^s\right) &= \omega^{\alpha+\beta} X^p Z^q X^r Z^s \\ &= \omega^{\alpha+\beta+rq} X^{p+r} Z^{q+s} \in W(d).\end{aligned} \tag{3.31}$$

In other words, $W(d)$ is closed under multiplication of its elements. Further to this, the identity matrix I is contained in $W(d)$ and, if we multiply an element of $W(d)$ by its adjoint,

$$\omega^\alpha X^\beta Z^\gamma (\omega^\alpha X^\beta Z^\gamma)^* = \omega^\alpha X^\beta Z^\gamma (\omega^{-\alpha} Z^{-\gamma} X^{-\beta}) = I. \tag{3.32}$$

In other words, the elements of $W(d)$ are unitary, with their inverses also belonging to $W(d)$. Hence, $W(d)$ forms a group under multiplication of operators, called the *Weyl-Heisenberg group*, and is a subgroup of $U(d)$, the group of $d \times d$ unitary matrices.

We now define a subset of elements of $W(d)$: the discrete Weyl-Heisenberg operators

$$W_{jk} = \omega^{2^{-1}jk} X^j Z^k, \tag{3.33}$$

where $j, k \in \mathbb{Z}_d$, and

$$2^{-1} = \begin{cases} (d+1)/2 & \text{if } d \text{ is odd,} \\ 1/2 & \text{if } d \text{ is even.} \end{cases} \tag{3.34}$$

Note that in the case of $d = 2$ the term $\omega^{2^{-1}}$ simply reduces to i. If we let the outcome space of G be $\mathbb{Z}_{d^2}$, and decompose each $i \in \mathbb{Z}_{d^2}$ as $i = i_1 d + i_2$ with $i_1, i_2 \in \mathbb{Z}_d$, then G is covariant under the action of $W(d)$ if

$$W_{qp} \mathsf{G}(i) W_{qp}^* = \mathsf{G}((i_1 \oplus q)d + (i_2 \oplus p)). \tag{3.35}$$

Given (3.35) and an extension of (3.31), namely,

$$\begin{aligned} W_{qp}W_{jk} &= \omega^{2^{-1}qp}X^qZ^p\omega^{2^{-1}jk}X^jZ^k = \omega^{2^{-1}(jk+qp+2pj)}X^{q+j}Z^{p+k} \\ &= \omega^{2^{-1}(pj-qk)}\omega^{2^{-1}(q+j)(p+k)}X^{q+j}Z^{p+k} \\ &= \omega^{2^{-1}(pj-qk)}W_{(q+j)(p+k)}, \end{aligned} \tag{3.36}$$

it follows that the effects take the form

$$\mathsf{G}(i) = \mathsf{G}(i_1d + i_2) = W_{i_1i_2}\mathsf{G}(0)W^*_{i_1i_2} = \frac{1}{d}W_{i_1i_2}P_\varphi W^*_{i_1i_2}, \tag{3.37}$$

where φ is known as the *fiducial vector* for the SIC-POVM. This vector determines the SIC-POVM we are dealing with, and searches for SIC-POVMs in a given dimension correspond to finding a valid fiducial vector in that dimension.

Example 3.1.1. As a simple example, in the qubit case, the four Weyl-Heisenberg operators are simply the identity and the 3 Pauli operators, i.e.,

$$W_{00} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad W_{01} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad W_{10} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad W_{11} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{3.38}$$

If we let $\varphi = (a, b)^T$ be our fiducial vector, where $|a|^2 + |b|^2 = 1$, then from the requirement

$$\begin{aligned} \frac{1}{12} &= \operatorname{tr}\left[\frac{1}{4}P_\varphi W_{01}P_\varphi W^*_{01}\right] \\ &= \frac{1}{4}\operatorname{tr}\left[\begin{pmatrix} |a|^2 & a\bar{b} \\ \bar{a}b & |b|^2 \end{pmatrix}\begin{pmatrix} |a|^2 & -a\bar{b} \\ -\bar{a}b & |b|^2 \end{pmatrix}\right] \\ &= \frac{1}{4}\left(|a|^4 - 2|a|^2|b|^2 + |b|^4\right) \\ &= \frac{1}{4}(|a|^2 - |b|^2)^2 \\ &= \frac{1}{4}(2|a|^2 - 1)^2, \end{aligned} \tag{3.39}$$

we find that

$$(2|a|^2 - 1)^2 = \frac{1}{3}; \quad \text{i.e., } |a|^2 = \frac{3 \pm \sqrt{3}}{6}, \tag{3.40}$$

and hence

$$|b|^2 = 1 - |a|^2 = \frac{3 \mp \sqrt{3}}{6}. \tag{3.41}$$

At this point, without loss of generality, we assume that a has no complex phase, and so

$$a = \frac{\sqrt{3 \pm \sqrt{3}}}{\sqrt{6}}, \qquad b = e^{i\beta}\frac{\sqrt{3 \mp \sqrt{3}}}{\sqrt{6}}. \tag{3.42}$$

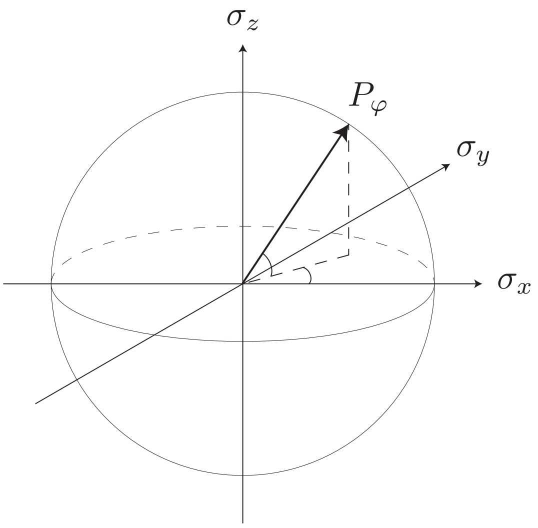


Figure 3.1: The location of the Projection operator P_φ, with φ the fiducial vector given in Equation (3.44), on the Bloch sphere.

In order to calculate β, we consider a second requirement, namely

$$\begin{aligned}\mathrm{tr}\left[\frac{1}{4}P_\varphi W_{10} P_\varphi W_{10}^*\right] &= \frac{1}{4}\mathrm{tr}\left[\begin{pmatrix}|a|^2 & a\bar{b}\\ \bar{a}b & |b|^2\end{pmatrix}\begin{pmatrix}|b|^2 & \bar{a}b\\ a\bar{b} & |a|^2\end{pmatrix}\right] = \frac{1}{4}\left(2\,|a|^2\,|b|^2 + (a\bar{b})^2 + (\bar{a}b)^2\right)\\ &= \frac{1}{4}\left(\frac{1}{18}(3\pm\sqrt{3})(3\mp\sqrt{3}) + \frac{1}{36}(3\pm\sqrt{3})(3\mp\sqrt{3})(e^{-2i\beta}+e^{2i\beta})\right) \\ &= \frac{1}{4}\left(\frac{1}{3}+\frac{1}{3}\cos(2\beta)\right) = \frac{1}{12}(1+\cos(2\beta)) = \frac{1}{12}.\end{aligned} \tag{3.43}$$

In other words, we require $\cos(2\beta) = 0$ if we wish to preserve the symmetry of these operators, and so $\beta = \pi/4$ or $\beta = 3\pi/4$. With this in mind, an example of a fiducial vector for a qubit Weyl-Heisenberg covariant SIC-POVM would be

$$\varphi = \frac{1}{\sqrt{6}}\begin{pmatrix}\sqrt{3+\sqrt{3}}\\ e^{i\pi/4}\sqrt{3-\sqrt{3}}\end{pmatrix}. \tag{3.44}$$

The projection operator for this state, P_φ, is of the form

$$P_\varphi = \frac{1}{2}\left(I + \frac{1}{\sqrt{3}}(\sigma_x+\sigma_y+\sigma_z)\right), \tag{3.45}$$

and its location on the Bloch sphere is show in Figure 3.1

When we start performing investigations in Mathematica, we will work exclusively with Weyl-Heisenberg covariant SIC-POVMs.

3.1.3 Mutually Orthogonal Latin Squares (MOLS)

We now introduce the concept of a Latin square, so named in honour of Euler, who introduced such structures, using Latin characters, in his attempts to solve what is known as the thirty-six officers problem.

Consider a $d \times d$ array $A : (i,j) \mapsto A_{ij}$, where $i,j = 1,\ldots,d$, and A_{ij} is the symbol that appears on the i^{th} row and j^{th} column. We shall restrict ourselves to symbols $A_{ij} =$

$1, \ldots, d$.

A particular class of $d \times d$ arrays are *Latin squares of order d*. These are arrays where $A_{ij} \neq A_{ik} \, \forall \, k \neq j$, and similarly $A_{ij} \neq A_{\ell j} \, \forall \, \ell \neq i$. This guarantees that every symbol appears in each and line and column of the array. An example of a Latin square of order 5 is given in Figure 3.2.

$$\begin{array}{ccccc}
1 & 2 & 3 & 4 & 5 \\
2 & 3 & 4 & 5 & 1 \\
3 & 4 & 5 & 1 & 2 \\
4 & 5 & 1 & 2 & 3 \\
5 & 1 & 2 & 3 & 4
\end{array}$$

Figure 3.2: A Latin square of order 5.

We now consider two Latin squares of order d, $A : (i, j) \mapsto A_{ij}$ and $B : (i, j) \mapsto B_{ij}$, and construct a new array of ordered pairs $C : (i, j) \mapsto (A_{ij}, B_{ij})$. The Latin squares A and B are *mutually orthogonal* if every pair in C is unique; that is, no ordered pair in C is repeated. This means that every symbol in A is paired with every symbol in B somewhere in C. An example of two mutually orthogonal Latin squares of order 5 is given in Figure 3.3, along with the array of ordered pairs formed from them.

$$\begin{array}{ccccc}
1 & 2 & 3 & 4 & 5 \\
2 & 3 & 4 & 5 & 1 \\
3 & 4 & 5 & 1 & 2 \\
4 & 5 & 1 & 2 & 3 \\
5 & 1 & 2 & 3 & 4
\end{array}
\qquad
\begin{array}{ccccc}
1 & 2 & 3 & 4 & 5 \\
3 & 4 & 5 & 1 & 2 \\
5 & 1 & 2 & 3 & 4 \\
2 & 3 & 4 & 5 & 1 \\
4 & 5 & 1 & 2 & 3
\end{array}
\quad \rightarrow \quad
\begin{array}{ccccc}
(1,1) & (2,2) & (3,3) & (4,4) & (5,5) \\
(2,3) & (3,4) & (4,5) & (5,1) & (1,2) \\
(3,5) & (4,1) & (5,2) & (1,3) & (2,4) \\
(4,2) & (5,3) & (1,4) & (2,5) & (3,1) \\
(5,4) & (1,5) & (2,1) & (3,2) & (4,3)
\end{array}$$

Figure 3.3: Two mutually orthogonal Latin squares of order 5. As can be checked, no ordered pair in the rightmost array is repeated twice.

As can be seen, every pair (i, j), $i, j = 1, \ldots, 5$ is formed in the array of ordered pairs. There is a strict maximum on the number of possible mutually orthogonal Latin squares of order d, as given by the following Lemma:

Lemma 3.1.1. There can be no more than $d - 1$ mutually orthogonal Latin squares of order d.

Proof. Suppose that we possess n mutually orthogonal Latin squares $L_1, \ldots, L_n$, and place the symbols $1, 2, \ldots, d$ in the first row of each Latin square in the order given, that is, the symbol i appears in the i^{th} column of the first row, but all other rows are so far undecided, like so

$$L_1 = \begin{array}{cccc}
1 & 2 & \ldots & d \\
- & - & \ldots & - \\
\vdots & \vdots & & \vdots \\
- & - & \ldots & -
\end{array}, \quad
L_2 = \begin{array}{cccc}
1 & 2 & \ldots & d \\
- & - & \ldots & - \\
\vdots & \vdots & & \vdots \\
- & - & \ldots & -
\end{array}, \quad \ldots, \quad
L_n = \begin{array}{cccc}
1 & 2 & \ldots & d \\
- & - & \ldots & - \\
\vdots & \vdots & & \vdots \\
- & - & \ldots & -
\end{array}.$$

We now decide what must be placed in the first column of the second row. Since they all contain 1 in the first column we must choose one of the remaining $d-1$ symbols. Without loss of generality we let the element in the first column of the second row of L_1 contain 2:

$$L_1 = \begin{array}{cccc} 1 & 2 & \dots & d \\ 2 & - & \dots & - \\ \vdots & \vdots & & \vdots \\ - & - & \dots & - \end{array}.$$

However, this means that the next Latin square cannot have 2 in the same location, as this would mean that the ordered pair $(2,2)$ would be repeated between them, contradicting their mutual orthogonality. Hence, L_2 has one of $d-2$ possible symbols to choose from. Following this logic through to L_n, we are left with one possible symbol after $d-1$ steps, and so we conclude that there can be no more than $d-1$ mutually orthogonal Latin squares of order d. □

Despite the existence of this upper bound, it is only reached for prime power values of d. Otherwise, there are at least two mutually orthogonal Latin squares for all values of d, with the exception of $d = 1, 2$ and 6, for which there exist one (this last case being a definitive negative answer for the thirty-six officers problem of Euler) [49].

3.2 Producing MUBs from a SIC-POVM

3.2.1 Motivation: Qubit example

Before going through the construction of finding a collection of MUBs from a SIC-POVM in a general finite-dimensional space, we will consider the case of a qubit system, where the connection is more transparent and the construction suggests the way for higher dimensions.

For a qubit system, a general SIC-POVM (we are not restricting ourselves to Weyl-Heisenberg covariant SIC-POVMs at this point) has 4 effects $\mathsf{G}(i) = \frac{1}{2}P_i$, $i \in \mathbb{Z}_4$, with

$$\operatorname{tr}[\mathsf{G}(i)\mathsf{G}(j)] = \frac{2\delta_{ij}+1}{12}. \tag{3.46}$$

By using the Bloch representation $\mathsf{G}(i) = \frac{1}{4}(I + \boldsymbol{s}_i \cdot \boldsymbol{\sigma})$, where $\|\boldsymbol{s}_i\| = 1$ for all i , we can infer a relationship between the Bloch vectors:

$$\begin{aligned} \frac{1}{12} &= \frac{1}{16}\operatorname{tr}\left[(1+\boldsymbol{s}_i\cdot\boldsymbol{s}_j)I + (\boldsymbol{s}_i+\boldsymbol{s}_j+i\boldsymbol{s}_i\times\boldsymbol{s}_j)\cdot\boldsymbol{\sigma}\right] \\ &= \frac{1}{8}(1+\boldsymbol{s}_i\cdot\boldsymbol{s}_j) \\ &\Rightarrow \boldsymbol{s}_i\cdot\boldsymbol{s}_j = -\frac{1}{3},\ i\neq j. \end{aligned} \tag{3.47}$$

The Bloch vectors correspond to the vertices of a tetrahedron embedded in the Bloch sphere, as shown in figure 3.4. The fact that the vectors corresponding to the vertices of a regular tetrahedron centred at the origin add to the zero vector guarantees the normalisation of G.

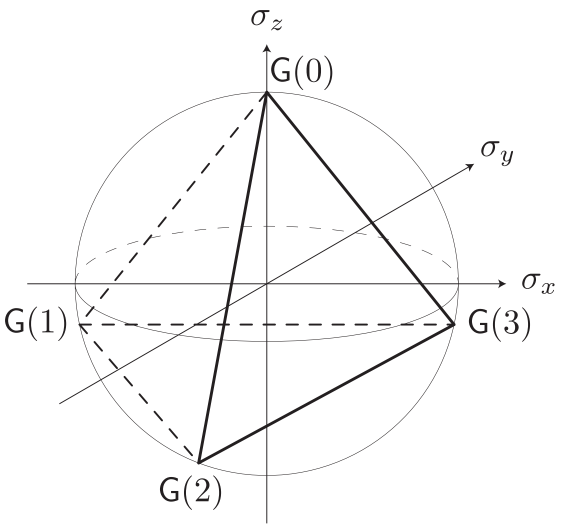


Figure 3.4: The points associated with the Bloch vectors of a qubit SIC-POVM form the vertices of a tetrahedron embedded in the Bloch sphere.

We can add these vectors together, leading to new operators. In particular, we define the vectors

$$\boldsymbol{m}_i = \frac{1}{2}(\boldsymbol{s}_0 + \boldsymbol{s}_i), \quad i = 1, 2, 3, \tag{3.48}$$

which satisfy the following condition:

$$\begin{aligned} \boldsymbol{m}_i \cdot \boldsymbol{m}_j &= \frac{1}{4}(\|\boldsymbol{s}_0\|^2 + \boldsymbol{s}_0 \cdot (\boldsymbol{s}_i + \boldsymbol{s}_j) + \boldsymbol{s}_i \cdot \boldsymbol{s}_j) \\ &= \frac{1}{4}\left[1 - \frac{2}{3} + \frac{1}{3}(4\delta_{ij} - 1)\right] \\ &= \frac{1}{3}\delta_{ij}. \end{aligned} \tag{3.49}$$

With (3.49) at hand, the associated effects

$$\mathsf{E}^k(\pm) = \frac{1}{2}(I \pm \boldsymbol{m}_k \cdot \boldsymbol{\sigma}), \tag{3.50}$$

which correspond to smearings of G:

$$\begin{aligned} &\mathsf{E}^1(+) = \mathsf{G}(0) + \mathsf{G}(1), \qquad \mathsf{E}^1(-) = \mathsf{G}(2) + \mathsf{G}(3), \\ &\mathsf{E}^2(+) = \mathsf{G}(0) + \mathsf{G}(2), \qquad \mathsf{E}^2(-) = \mathsf{G}(1) + \mathsf{G}(3), \\ &\mathsf{E}^3(+) = \mathsf{G}(0) + \mathsf{G}(3), \qquad \mathsf{E}^3(-) = \mathsf{G}(1) + \mathsf{G}(2), \end{aligned} \tag{3.51}$$

can be seen to satisfy the following properties:

$$\operatorname{tr}\left[\mathsf{E}^k(\pm)\mathsf{E}^\ell(\pm)\right] = \frac{1}{2}(1 \pm \boldsymbol{m}_k \cdot \boldsymbol{m}_\ell) = \frac{1}{2}, \quad k \neq \ell. \tag{3.52}$$

In other words, the POVMs E^k, $k = 1, 2, 3$ are mutually unbiased. Given the subnormalisation of the vectors $\boldsymbol{m}_k$, we could express them in terms of the normalised vectors

$\boldsymbol{n}_k = \sqrt{3}\,\boldsymbol{m}_k$. Hence,

$$\begin{aligned} \mathsf{E}^k(\pm) &= \frac{1}{2}\left(I \pm \frac{1}{\sqrt{3}}\boldsymbol{n}_k \cdot \boldsymbol{\sigma}\right) = \frac{1}{\sqrt{3}}\left[\frac{1}{2}(I \pm \boldsymbol{n}_k \cdot \boldsymbol{\sigma})\right] + \frac{\sqrt{3}-1}{2\sqrt{3}} I \\ &= \frac{\sqrt{3}+1}{2\sqrt{3}}\left[\frac{1}{2}(I \pm \boldsymbol{n}_k \cdot \boldsymbol{\sigma})\right] + \frac{\sqrt{3}-1}{2\sqrt{3}}\left[\frac{1}{2}(I \mp \boldsymbol{n}_k \cdot \boldsymbol{\sigma})\right] \\ &= \frac{\sqrt{3}+1}{2\sqrt{3}}\mathsf{P}^k(\pm) + \frac{\sqrt{3}-1}{2\sqrt{3}}\mathsf{P}^k(\mp), \end{aligned} \tag{3.53}$$

where P^k is the PVM projecting onto the eigenbasis of E^k. Given the effects $\mathsf{P}^k(\pm)$ are determined by the Bloch vectors $\boldsymbol{n}_k$, it follows immediately that

$$\operatorname{tr}\left[\mathsf{P}^k(\pm)\mathsf{P}^\ell(\mp)\right] = \frac{1}{2}, \quad k \neq \ell, \tag{3.54}$$

i.e., the eigenbases of the POVMs E^k, $k = 1, 2, 3$ are mutually unbiased, and so we have constructed a complete set of MUBs from a SIC-POVM in the qubit case.

3.2.2 d-partitions and the one-overlap property

In order to construct MUBs from a SIC-POVM in spaces $\mathcal{H}_d$, where $d > 2$, we must first consider how we want to create marginal observables from our SIC-POVM G, as in the qubit case. This corresponds to forming partitions—disjoint subsets whose union equals the whole set—of the d^2 effects of G, and with this in mind we present the following definition:

Definition 3.2.1. For a set of d^2 elements $A = \{a_1, \ldots, a_{d^2}\}$, a *$d$-partition* $\mathcal{P}$ is a partition of A into d disjoint bins $\{\mathcal{P}_1, \ldots, \mathcal{P}_d\}$, each containing d elements.

Further to this, we introduce the following property:

Definition 3.2.2. Two d-partitions $\mathcal{P}^1$, $\mathcal{P}^2$ of a set satisfy the *one-overlap property* if any two of their bins have just one element in common, i.e., for any two bins $\mathcal{P}^1_\mu \in \mathcal{P}^1$, $\mathcal{P}^2_\nu \in \mathcal{P}^2$, with $\mu, \nu = 1, 2, \ldots, d$, $\left|\mathcal{P}^1_\mu \cap \mathcal{P}^2_\nu\right| = 1$.

As a simple example of both of these concepts, consider a set of 9 elements; we can place these elements in a 3×3 array, and from this we immediately identify two 3-partitions by splitting the array into its rows and columns, as shown in Figure 3.5. In general, we can take a d^2-element set $A = \{1, 2, \ldots, d^2\}$, and form a $d \times d$ array with the first d elements $\{1, 2, \ldots, d\}$ forming the first row, etc. The corresponding row and column partitions, denoted by $\mathcal{P}^R$ and $\mathcal{P}^C$ respectively, are then given via

$$\mathcal{P}^R_\mu = \{(\mu - 1)d + 1, (\mu - 1)d + 2, \ldots, \mu d\}, \tag{3.55a}$$

$$\mathcal{P}^C_\nu = \{\nu, \nu + d, \ldots, \nu + (d - 1)d\}. \tag{3.55b}$$

These partitions, which we shall refer to as *Cartesian partitions*, immediately satisfy the one-overlap property.

Lemma 3.2.1. For a given set of d^2 elements, the number of d-partitions that satisfy the one-overlap property is at least 3 and at most $d + 1$.

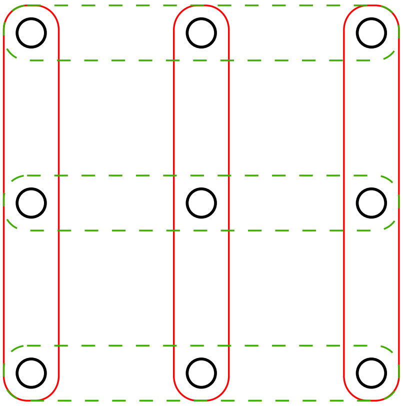


Figure 3.5: The Cartesian partitions of a set of 9 elements immediately satisfy the one-overlap property.

Proof. Suppose that there exist n partitions of a set of d^2 elements $\{a_1, a_2, \ldots, a_{d^2}\}$, and consider the bin in each partition containing the element a_1. Since each bin shares this element, no two bins can share any other element. There exist $d^2 - 1$ elements in the subset $\{a_2, a_3, \ldots, a_{d^2}\}$, and each bin contains $d-1$ elements that are not repeated in any of the other bins considered. Therefore $n \leq (d^2-1)/(d-1) = d+1$.

We now express the d^2 elements in a $d \times d$ array. The Cartesian partitions have already been seen to be d-partitions that satisfy the one-overlap property, so all that we must do is show that there exists at least one more, labelled $\mathcal{P}^{(3)}$. We begin by relabelling the array elements into matrix form, (a_{ij}). The Cartesian partitions span the rows and columns of the array, respectively, and so each bin of $\mathcal{P}^{(3)}$ can have only one element from each row and column. The first bin is composed of the main diagonal: $\mathcal{P}^{(3)}_1 = \{a_{11}, a_{22}, \ldots, a_{dd}\}$. To compose the remaining bins we consider the diagonals parallel to the main diagonal defining $\mathcal{P}^{(3)}_1$. We define $S_{1,j}$ to be the diagonal starting with the element a_{1j}, where $j \geq 2$, and $S_{d+2-i,1}$ to be the diagonal starting with a_{i1} with $i \geq 2$. The disjoint bins $\mathcal{P}^{(3)}_1$ and $\mathcal{P}^{(3)}_\nu = S_{1,\nu} \cup S_{\nu,1}$, $\nu \in \{2, 3, \ldots, d\}$, form a d-partition that shares the one-overlap property with the Cartesian partitions. An example of $\mathcal{P}^{(3)}$ is given for a 9-element set in Figure 3.6, with bins $\mathcal{P}^{(3)}_1 = \{a_{11}, a_{22}, a_{33}\}$ (shown via a red line), $\mathcal{P}^{(3)}_2 = \{a_{12}, a_{23}, a_{31}\}$ (shown via a green dashed line) and $\mathcal{P}^{(3)}_3 = \{a_{13}, a_{21}, a_{32}\}$ (unlabelled). □

Whilst Lemma 3.2.1 provides an upper and lower bound for the number of possible d-partitions that satisfy the one-overlap property, it does not give us any indication of how many such partitions are actually obtainable. A more useful value is given via the following proposition:

Proposition 3.2.2. There is a one-to-one correspondence between the set of d-partitions $\{\mathcal{P}^k\}$ that satisfy the one-overlap property with respect to the Cartesian partitions and the set of Latin squares of order d. Furthermore, two d-partitions which also satisfy the one-overlap property with respect to each other correspond to mutually orthogonal Latin

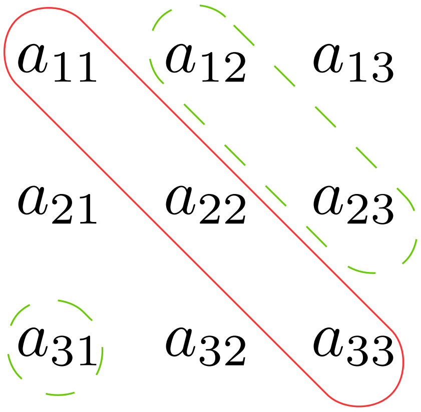


Figure 3.6: The partition $\mathcal{P}^{(3)}$ for a 9-element set. The bins $\mathcal{P}_1^{(3)}$ and $\mathcal{P}_2^{(3)}$, are highlighted by red full and green dashed lines, respectively.

squares.

Proof. Consider a d^2-element set $A = \{1, 2, \dots, d^2\}$ placed in a $d \times d$ array such that the first d elements $\{1, 2, \dots, d\}$ form the first row, etc., as above. The Cartesian partitions, $\mathcal{P}^R$ and $\mathcal{P}^C$, are therefore given via Equations (3.55a) and (3.55b). Hence, any d-partition $\mathcal{P}^k$ satisfying the one-overlap property with respect to $\mathcal{P}^R$ and $\mathcal{P}^C$ cannot possess a bin containing two elements of the form $(\mu-1)d+i$ with μ fixed, or two of the form $\nu+jd$ with ν fixed. In terms of the array, a given d-partition of A acts as a collection of paths through the array such that only one path coincides with each point on the array (due to the disjoint nature of the bins). With this interpretation, that the partition $\mathcal{P}^k$ satisfies the one-overlap property with $\mathcal{P}^R$ and $\mathcal{P}^C$ means that no bin $\mathcal{P}_\mu^k \in \mathcal{P}^k$ can contain two elements from the same row or column. For each point in the array (i, j), which corresponds to the element $(i-1)d+j \in A$, we can assign to it the value $\mu_{(i,j)} \in \{1, 2, \dots, d\}$ corresponding to the bin $\mathcal{P}_\mu^k$ that element of A belongs to. Since no two elements of a given row or column can be in the same bin belonging to $\mathcal{P}^k$, each row and column of the array $\{\mu_{(i,j)}\}$ must contain every value in the set $\{1, 2, \dots, d\}$ once. Hence, this array is a Latin square of order d.

Conversely, any Latin square of order d details a collection of d disjoint bins, each containing d elements, that form a partition of a set of d^2 elements which satisfy the one-overlap property the Cartesian partitions.

Consider now two d-partitions of A, $\mathcal{P}^k$ and $\mathcal{P}^\ell$, that satisfy the one-overlap property with respect to the Cartesian partitions (hence correspond to Latin squares of order d), and also satisfy the one-overlap property with respect to each other. We now construct a $d \times d$ array with each point (i, j) being assigned the ordered pair $(\mu_{(i,j)}, \nu_{(i,j)}) \in \{1, 2, \dots, d\} \times \{1, 2, \dots, d\}$ corresponding to the bins $\mathcal{P}_\mu^k$ and $\mathcal{P}_\nu^\ell$ that the value $(i-1)d + j \in A$ belongs to in each partition. Since these partitions satisfy the one-overlap property, any two bins coincide only once, and hence every possible ordered pair $(\mu, \nu) \in \{1, 2, \dots, d\} \times \{1, 2, \dots, d\}$ appears in this array. Therefore, the Latin squares corresponding to these partitions are mutually orthogonal. □

Corollary 3.2.3. Number of d-partitions satisfying the one-overlap property is equal to 2 plus the number of mutually orthogonal Latin squares of order d.

3.2.3 Mutually unbiased SIC-compatible observables

Suppose that we have a SIC-POVM G on $\mathcal{H}_d$, with d^2 effects $\{\mathsf{G}(i)\}_{i=0}^{d^2-1}$. For every partition $\mathcal{P}^k = \{\mathcal{P}_1^k, \ldots, \mathcal{P}_d^k\}$ of the set of effects we define the POVM E^k with effects

$$\mathsf{E}^k(\mu) = \sum_{\mathsf{G}(i)\in\mathcal{P}_\mu^k} \mathsf{G}(i). \tag{3.56}$$

Using the defining properties of our SIC-POVM, we arrive at the following theorem

Theorem 3.2.4. For two POVMs $\mathsf{E}^k, \mathsf{E}^\ell$ arising from the d-partitions $\mathcal{P}^k$ and $\mathcal{P}^\ell$, respectively, of the SIC-POVM G that satisfy the one-overlap property, the following conditions are satisfied:

1. $$\operatorname{tr}\left[\mathsf{E}^k(\mu)\right] = 1 \quad \forall\, \mu = 1, \ldots, d; \tag{3.57}$$

2. $$\operatorname{tr}\left[\mathsf{E}^k(\mu)^2\right] = \frac{2}{d+1} \quad \forall\, \mu = 1, \ldots, d; \tag{3.58}$$

3. $$\operatorname{tr}\left[\mathsf{E}^k(\mu)\mathsf{E}^k(\nu)\right] = \frac{1}{d+1} \quad \forall\, \mu = 1, \ldots, d,\, \nu \neq \mu; \tag{3.59}$$

4. $$\operatorname{tr}\left[\mathsf{E}^k(\mu)\mathsf{E}^\ell(\nu)\right] = \frac{1}{d} \quad \forall\, \mu, \nu \;\text{ if } k \neq \ell. \tag{3.60}$$

Proof. We will address these results in order.

1. Property 1 is trivial: it is a sum of d operators, each of trace $1/d$;

2. $\operatorname{tr}\left[\mathsf{E}^k(\mu)^2\right]$ is composed of d terms of the form $\operatorname{tr}\left[\mathsf{G}(i)^2\right]$ and $d(d-1)$ terms of the form $\operatorname{tr}\left[\mathsf{G}(i)\mathsf{G}(j)\right]$, where $i \neq j$. Therefore, using equations (3.20) and (3.22),

$$\operatorname{tr}\left[\mathsf{E}^k(\mu)^2\right] = d\left(\frac{1}{d^2}\right) + d(d-1)\left(\frac{1}{d^2(d+1)}\right) = \frac{d+1+d-1}{d(d+1)} = \frac{2}{d+1}; \tag{3.61}$$

3. $\operatorname{tr}\left[\mathsf{E}^k(\mu)\mathsf{E}^k(\nu)\right]$, where $\mu \neq \nu$, contains just d^2 terms of the form $\operatorname{tr}\left[\mathsf{G}(i)\mathsf{G}(j)\right]$, where $i \neq j$, and so from equation (3.22)

$$\operatorname{tr}\left[\mathsf{E}^k(\mu)\mathsf{E}^k(\nu)\right] = d^2\left(\frac{1}{d^2(d+1)}\right) = \frac{1}{d+1}; \tag{3.62}$$

4. Since the partitions $\mathcal{P}^k$ and $\mathcal{P}^\ell$ satisfy the one-overlap property, there exists one SIC-POVM effect $\mathsf{G}(i)$, say, that is shared by the effects $\mathsf{E}^k(\mu)$ and $\mathsf{E}^\ell(\nu)$, whilst the remaining SIC-POVM effects are unique to either $\mathsf{E}^k(\mu)$ or $\mathsf{E}^\ell(\nu)$. Therefore,

$\mathrm{tr}\left[\mathsf{E}^k(\mu)\mathsf{E}^\ell(\nu)\right]$ contains one term of the form $\mathrm{tr}\left[\mathsf{G}(i)^2\right]$ and d^2-1 terms of the form $\mathrm{tr}\left[\mathsf{G}(m)\mathsf{G}(n)\right]$, where $m \neq n$ and at most one of them is equal to i. Hence,

$$\mathrm{tr}\left[\mathsf{E}^k(\mu)\mathsf{E}^\ell(\nu)\right] = \left(\frac{1}{d^2}\right) + (d^2-1)\left(\frac{1}{d^2(d+1)}\right) = \frac{1+d-1}{d^2} = \frac{1}{d}. \tag{3.63}$$

This concludes the proof. □

Definition 3.2.3. A collection of d-outcome POVMs are *SIC-compatible* if they arise as margins of a common SIC-POVM.

Remark. From Lemma 3.2.1, we can have at least 3 and up to $d+1$ SIC-compatible observables satisfying the properties given by Equations (3.57)-(3.60). This highlights an interesting feature of these SIC-compatible observables: For a general set of n jointly measurable d-outcome observables, we would naïvely expect the joint observable to possess d^n outcomes, and yet here we have up to $d+1$ jointly measurable observables whose joint observable, the SIC-POVM G possesses only d^2 outcomes instead of d^{d+1}.

We now properly label what we saw in the qubit case, and also in equation (3.60):

Definition 3.2.4. Two POVMs, E^k and E^ℓ, acting on $\mathcal{H}_d$ are *mutually unbiased* if

$$\mathrm{tr}\left[\mathsf{E}^k(i)\mathsf{E}^\ell(j)\right] = \frac{1}{d} \quad \forall\, i,j. \tag{3.64}$$

Instead of simply being sufficient, the construction of d-partitions and the one-overlap property are necessary if we want mutually unbiased SIC-compatible POVMS:

Proposition 3.2.5. If a set of at least 3 mutually unbiased POVMs E^k are SIC-compatible, then their associated partitions are d-partitions, $\mathcal{P}^k$, that satisfy the one-overlap property.

Proof. We begin with our collection of margin POVMs E^k. We assume nothing about the number of effects each POVM possesses, nor do we assume anything about the number of SIC-POVM effects used to form a given effect of any of the E^k. In terms of the partitions $\mathcal{P}^k$ forming each of the POVMs, this means that we assume nothing about the number of bins each partition possesses, or the number of elements in each bin. Further to this, we assume nothing about how much any two bins from different partitions intersect.

The first two POVMs that we consider, E^k and E^ℓ, are given by partitions $\mathcal{P}^k$ and $\mathcal{P}^\ell$, which contain m_k and m_ℓ bins, respectively. The number of elements in a given bin of $\mathcal{P}^k$ is given by $\left|\mathcal{P}^k_\mu\right| = n^k_\mu$, and similarly $n^\ell_\nu = \left|\mathcal{P}^\ell_\nu\right|$. It follows immediately that $\sum_\mu n^k_\mu = \sum_\nu n^\ell_\nu = d^2$. We denote the overlap between two bins by $a^{k,\ell}_{\mu,\nu} = \left|\mathcal{P}^k_\mu \cap \mathcal{P}^\ell_\nu\right|$. Since every element of $\mathcal{P}^k_\mu$ belongs to a bin in $\mathcal{P}^\ell$, and similarly for the elements of $\mathcal{P}^\ell_\nu$ in $\mathcal{P}^k$, it follows that $\sum_\mu a^{k,\ell}_{\mu,\nu} = n^\ell_\nu$ and $\sum_\nu a^{k,\ell}_{\mu,\nu} = n^k_\mu$.

Looking at the mutual unbiasedness of these POVMs in this light, we see

$$\begin{aligned}\frac{1}{d} = \mathrm{tr}\left[\mathsf{E}^k(\mu)\mathsf{E}^\ell(\nu)\right] &= \frac{1}{d^2}a^{k,\ell}_{\mu,\nu} + \frac{1}{d^2(d+1)}(n^k_\mu n^\ell_\nu - a^{k,\ell}_{\mu,\nu}) \\ &= \frac{1}{d^2(d+1)}(d a^{k,\ell}_{\mu,\nu} + n^k_\mu n^\ell_\nu).\end{aligned} \tag{3.65}$$

If we sum over ν:

$$\begin{aligned}\frac{m_\ell}{d} &= \sum_\nu \frac{1}{d} = \frac{1}{d^2(d+1)}(d\sum_\nu a^{k,\ell}_{\mu,\nu} + n^k_\mu \sum_\nu n^\ell_\nu) \\ &= \frac{1}{d^2(d+1)} d(d+1) n^k_\mu \\ &= \frac{n^k_\mu}{d}.\end{aligned} \tag{3.66}$$

In other words, $n^k_\mu = m_\ell$ for all μ, and similarly $n^\ell_\nu = m_k$ for all ν. Since n^k_μ is fixed for all μ, we shorten the notation to n^k, and similarly we change n^ℓ_ν to n^ℓ, and so we now have $n^k m_k = n^\ell m_\ell = m_k m_\ell = d^2$.

We now introduce the third margin E^j satisfying tr $\left[\mathsf{E}^j(\sigma)\mathsf{E}^k(\mu)\right] =$ tr $\left[\mathsf{E}^j(\sigma)\mathsf{E}^\ell(\nu)\right] = 1/d$ for all σ, μ, ν. By repeating the preceding argument, we find that $n^k = n^\ell = m_j$, and so $d = m_k = n_k$, and likewise for the other partitions. In other words, mutually unbiased SIC-compatible POVMs arise from d-partitions. Further to that, from Equation (3.65),

$$a^{k,\ell}_{\mu,\nu} = d + 1 - \frac{n^k_\mu n^\ell_\nu}{d} = d + 1 - d = 1. \tag{3.67}$$

That is, the partitions satisfy the one-overlap property, which concludes the proof. □

Corollary 3.2.6. There exist at least 3 and at most $d+1$ mutually unbiased SIC-compatible POVMs defined on $\mathcal{H}_d$. In the case of dimension 6, there cannot exist more than 3 mutually unbiased SIC-compatible POVMs.

The first point here is a direct result of Lemma 3.2.1, whilst the second follows from Corollary 3.2.3 and the non-existence of two mutually orthogonal Latin squares of order 6.

3.2.4 Commutative mutually unbiased SIC-compatible POVMs

We now restrict our consideration to cases where the mutually unbiased SIC-compatible POVMs we work with are commutative. As will be highlighted in Section 3.5, this is indeed a restriction, but it allows us to present some important results.

By commutative POVMs we mean POVMs E whose effects commute, i.e., $[\mathsf{E}(\mu), \mathsf{E}(\nu)] = 0$ for all μ, ν in the POVM's outcome space. Such POVMs possess a common eigenbasis, and can therefore be expressed in the following form

$$\mathsf{E}(\nu) = \sum_k \lambda_{\nu,k} \mathsf{P}(k), \tag{3.68}$$

where $\mathsf{P}(k)$ is the projection onto the shared eigenstate of the effects, and the $\lambda_{\nu,k}$ is the respective eigenvalue, i.e. $\mathsf{E}(\nu)\mathsf{P}(k) = \lambda_{\nu,k}\mathsf{P}(k)$. If we denote by $\mathbf{P}$ the vector of projections $\mathsf{P}(k)$, i.e., $\mathbf{P} = (\mathsf{P}(1), \mathsf{P}(2), \dots, \mathsf{P}(d))$, and by $\boldsymbol{\lambda}_\nu$ the vector of eigenvalues for the effect $\mathsf{E}(\nu)$, then Equation (3.68) can be rewritten in terms of the scalar product

$$\mathsf{E}(\nu) = \boldsymbol{\lambda}_\nu \cdot \mathbf{P}. \tag{3.69}$$

Using Equation (3.68), and the linear independence of the projections $\mathsf{P}(k)$, the normalisation of E leads to

$$I = \sum_\nu \mathsf{E}(\nu) = \sum_\nu \sum_k \lambda_{\nu,k} \mathsf{P}(k) = \sum_k \left(\sum_\nu \lambda_{\nu,k} \right) \mathsf{P}(k). \tag{3.70}$$

In other words,

$$\sum_\nu \lambda_{\nu,k} = 1 \;\forall\; k, \;\text{ or }\; \sum_\nu \boldsymbol{\lambda}_\nu = \mathbb{I}, \tag{3.71}$$

where $\mathbb{I} = (1, 1, \ldots, 1)$ is the d-dimensional unit vector. If we now consider the SIC-compatible POVMs from the preceding discussions, then, by Theorem 3.2.4, we can impose further restrictions on the $\lambda_{\nu,k}$. From Equation (3.57):

$$1 = \operatorname{tr}\left[\mathsf{E}(\nu)\right] = \sum_k \lambda_{\nu,k} \operatorname{tr}\left[\mathsf{P}(k)\right] = \sum_k \lambda_{\nu,k} = \boldsymbol{\lambda}_\nu \cdot \mathbb{I} \;\forall\; \nu. \tag{3.72}$$

In other words, the matrix $\Lambda = [\lambda_{\nu,k}]$ of eigenvalues for the commutative SIC-compatible POVM E is doubly stochastic, as well as lying on the $d-1$-dimensional hyperplane $\boldsymbol{\lambda}\cdot\mathbb{I} = 1$. Continuing with the results of Theorem 3.2.4, if we now consider Equation (3.58):

$$\frac{2}{d+1} = \operatorname{tr}\left[\mathsf{E}(\nu)^2\right] = \sum_{k,\ell} \lambda_{\nu,k}\lambda_{\nu,\ell} \operatorname{tr}\left[\mathsf{P}(k)\mathsf{P}(\ell)\right] = \sum_k \lambda_{\nu,k}^2 = \left\|\boldsymbol{\lambda}_\nu\right\|^2. \tag{3.73}$$

Equation (3.73) tells us that the eigenvalue vectors $\boldsymbol{\lambda}_\nu$ lie on the surface of a $d-1$-sphere in $\mathbb{R}^d$ centred at the origin with radius $\sqrt{2/(d+1)}$. Given that the effects of E all have positive eigenvalues, these vectors all lie within the positive region of $\mathbb{R}^d$. Finally, from Equation (3.59):

$$\frac{1}{d+1} = \operatorname{tr}\left[\mathsf{E}(\nu)\mathsf{E}(\mu)\right] = \sum_{k,\ell} \lambda_{\nu,k}\lambda_{\mu,\ell} \operatorname{tr}\left[\mathsf{P}(k)\mathsf{P}(\ell)\right] = \sum_k \lambda_{\nu,k}\lambda_{\mu,k} = \boldsymbol{\lambda}_\nu \cdot \boldsymbol{\lambda}_\mu. \tag{3.74}$$

At this point we define the vector

$$\boldsymbol{r}_\nu = \boldsymbol{\lambda}_\nu - \frac{1}{d}\mathbb{I}. \tag{3.75}$$

These vectors satisfy the following two properties:

$$\begin{aligned} \left\|\boldsymbol{r}_\nu\right\|^2 &= \left\|\boldsymbol{\lambda}_\nu\right\|^2 + \frac{1}{d^2}\left\|\mathbb{I}\right\|^2 - \frac{2}{d}\boldsymbol{\lambda}_\nu \cdot \mathbb{I} \\ &= \frac{2}{d+1} - \frac{1}{d} = \frac{d-1}{d(d+1)}, \qquad (3.76a) \\ \boldsymbol{r}_\nu \cdot \boldsymbol{r}_\mu &= \boldsymbol{\lambda}_\nu \cdot \boldsymbol{\lambda}_\mu + \frac{1}{d^2}\left\|\mathbb{I}\right\|^2 - \frac{1}{d}\mathbb{I}\cdot\left(\boldsymbol{\lambda}_\nu + \boldsymbol{\lambda}_\mu\right) \\ &= \frac{1}{d+1} - \frac{1}{d} = -\frac{1}{d(d+1)}, \qquad (3.76b) \end{aligned}$$

where we have made use of Equation (3.72) in both parts, as well as the fact that $\left\|\mathbb{I}\right\|^2 = d$.

From Equation (3.76b) we have

$$-\frac{1}{d(d+1)} = \boldsymbol{r}_\nu \cdot \boldsymbol{r}_\mu = \|\boldsymbol{r}_\nu\|\,\|\boldsymbol{r}_\mu\| \cos\theta = \frac{d-1}{d(d+1)}\cos\theta. \tag{3.77}$$

In other words, $\cos\theta = -1/(d-1)$ for any two $\boldsymbol{r}$ vectors defined above. This is a characteristic of the vertices of a regular d-simplex, and so the eigenvalue vectors point to the vertices of a regular d-simplex centred at the point $(1/d)\mathbb{I}$. We summarise these results in the following proposition:

Proposition 3.2.7. Consider a commutative POVM E with effects $\mathsf{E}(\nu) = \sum_k \lambda_{\nu,k}\mathsf{P}(k) = \boldsymbol{\lambda}_\nu \cdot \mathbf{P}$ that are of unit trace and satisfy $\mathrm{tr}\,[\mathsf{E}(\nu)\mathsf{E}(\mu)] = 1/(d+1)$. The eigenvalue vectors $\boldsymbol{\lambda}_\nu \in \mathbb{R}^d$ correspond to the vertices of a regular $d-1$-simplex centred at $(1/d)\mathbb{I}$ that is embedded in the intersection of the $d-1$-dimensional hyperplane $\boldsymbol{\lambda}\cdot\mathbb{I} = 1$ and the $d-1$-sphere centred at the origin with radius $\sqrt{2/(d+1)}$.

It is with these geometric properties of the eigenvalue vectors that we now prove the following theorem:

Theorem 3.2.8. Consider two commutative POVMs E^k and E^ℓ, whose effects have unit trace and satisfy Equation (3.59). If their effects have the spectral decompositions given by Equation (3.68), i.e.,

$$\mathsf{E}^k(\nu) = \sum_i \lambda^k_{\nu,i}\mathsf{P}^k(i), \quad \mathsf{E}^\ell(\mu) = \sum_j \lambda^\ell_{\mu,j}\mathsf{P}^\ell(j), \tag{3.78}$$

then the following equivalence holds:

$$\mathrm{tr}\left[\mathsf{E}^k(\nu)\mathsf{E}^\ell(\mu)\right] = \frac{1}{d}\ \forall\, \mu,\nu \quad \Longleftrightarrow \quad \mathrm{tr}\left[\mathsf{P}^k(i)\mathsf{P}^\ell(j)\right] = \frac{1}{d}\ \forall\, i,j. \tag{3.79}$$

Proof. The trivial part of this proof comes when we begin by assuming that $\mathrm{tr}\left[\mathsf{P}^k(i)\mathsf{P}^\ell(j)\right] = 1/d$ for all i,j. In which case

$$\mathrm{tr}\left[\mathsf{E}^k(\nu)\mathsf{E}^\ell(\mu)\right] = \sum_{i,j}\lambda^k_{\nu,i}\lambda^\ell_{\mu,j}\mathrm{tr}\left[\mathsf{P}^k(i)\mathsf{P}^\ell(j)\right] = \frac{1}{d}\left(\sum_i \lambda^k_{\nu,i}\right)\left(\sum_j \lambda^\ell_{\mu,j}\right) = \frac{1}{d}. \tag{3.80}$$

We now assume that $\mathrm{tr}\left[\mathsf{E}^k(\nu)\mathsf{E}^\ell(\mu)\right] = 1/d$ for all ν,μ. This can be expressed differently:

$$\frac{1}{d} = \mathrm{tr}\left[\mathsf{E}^k(\nu)\mathsf{E}^\ell(\mu)\right] = \sum_{i,j}\lambda^k_{\nu,i}\lambda^\ell_{\mu,j}\mathrm{tr}\left[\mathsf{P}^k(i)\mathsf{P}^\ell(j)\right] =: \sum_{i,j}\lambda^k_{\nu,i}q^{k,\ell}_{i,j}\lambda^\ell_{\mu,j}, \tag{3.81}$$

where $q^{k,\ell}_{i,j} = \mathrm{tr}\left[\mathsf{P}^k(i)\mathsf{P}^\ell(j)\right]$. This is equivalent to

$$\Lambda^k Q^{k,\ell}(\Lambda^\ell)^T = \frac{1}{d}U, \tag{3.82}$$

where $\Lambda^k = [\lambda^k_{\nu,i}]$ as above, $Q^{k,\ell} = [q^{k,\ell}_{i,j}]$ and $U_{ij} = 1$ for all i,j. Because the rows of Λ^k and Λ^ℓ correspond to the vertices of a regular simplex in a hyperplane not containing the origin, they must be linearly independent, and hence both Λ^k and Λ^ℓ are invertible.

Therefore there must exist two matrices, Γ^k and Γ^ℓ, such that

$$\Lambda^k\Gamma^k = \Lambda^\ell\Gamma^\ell = I, \tag{3.83}$$

and so

$$Q^{k,\ell} = \frac{1}{d}\Gamma^k U(\Gamma^\ell)^T. \tag{3.84}$$

A matrix A has rows which add to unity iff $A\mathbb{I}^T = \mathbb{I}^T$, with $\mathbb{I}^T = (1,1,\ldots,1)^T$. This property hence follows for matrices which are stochastic in their rows. From the row stochasticity of Λ^k, it follows that

$$\mathbb{I}^T = I\mathbb{I}^T = (\Gamma^k\Lambda^k)\mathbb{I}^T = \Gamma^k\mathbb{I}^T, \tag{3.85}$$

and similarly, due to the column stochasticity of Λ^ℓ, and hence the row stochasticity of $(\Lambda^\ell)^T$,

$$\mathbb{I}^T = (\Gamma^\ell)^T\mathbb{I}^T. \tag{3.86}$$

Hence, both Γ^k and $(\Gamma^\ell)^T$ have rows which add to unity (although they will not usually be stochastic, as they will not be positive definite). From Equations (3.85) and (3.86) it follows immediately that

$$\Gamma^k U = U = U(\Gamma^\ell)^T, \tag{3.87}$$

and so

$$Q^{k,\ell} = \frac{1}{d}\Gamma^k U(\Gamma^\ell)^T = \frac{1}{d}U. \tag{3.88}$$

In other words,

$$q^{k,\ell}_{i,j} = \mathrm{tr}\left[\mathsf{P}^k(i)\mathsf{P}^\ell(j)\right] = \frac{1}{d}\ \forall\ i,j, \tag{3.89}$$

which concludes the proof. □

It is from this theorem that we can conclude the main result for this section of our investigation:

Corollary 3.2.9. Assume that G is a SIC-POVM, which possesses a collection of commutative margins E^k arising from d-partitions satisfying the one-overlap property, i.e., the E^k are commutative mutually unbiased SIC-compatible POVMs. The eigenbases of the associated margins are then mutually unbiased.

With Theorem 3.2.8 and Corollary 3.2.9 we have shown that by starting with a SIC-POVM G and by constructing d-partitions of the effects of G, we are capable of, in some instances, finding up to a complete set of $d+1$ MUBs. In other words, we have shown an example of how compatible observables – the margins of G – can be intricately linked to the most incompatible observables on the system – the set of MUBs.

3.3 Producing a SIC-POVM from MUBs

3.3.1 Qubit example

Much like in the previous section, we will begin by considering the qubit case in order to get an understanding for the construction that follows.

We begin with 3 MUBs $\mathsf{P}^k(\pm) = (1/2)(I \pm \boldsymbol{n}_k \cdot \boldsymbol{\sigma})$, and apply some smearing parameter $\lambda \in (0,1)$ to create the mutually unbiased POVMs

$$\mathsf{E}^k(\pm) = \lambda \mathsf{P}^k(\pm) + (1-\lambda)\frac{1}{2}I = \frac{1}{2}(1+\lambda)\mathsf{P}^k(\pm) + \frac{1}{2}(1-\lambda)\mathsf{P}^k(\mp). \tag{3.90}$$

As can be seen, the mutual unbiasedness of the E^k is independent of the value of λ, and is a result of the effects being of unit trace and the mutual unbiasedness of the eigenbases, as was shown in the proof of Theorem 3.2.8. Further to this, these effects satisfy

$$\begin{aligned}
\mathrm{tr}\left[\mathsf{E}^k(\pm)^2\right] &= \lambda^2 \mathrm{tr}\left[\mathsf{P}^k(\pm)^2\right] + \lambda(1-\lambda)\mathrm{tr}\left[\mathsf{P}^k(\pm)\right] + (1-\lambda)^2\frac{1}{4}\mathrm{tr}\left[I\right] \\
&= \lambda^2 + \lambda(1-\lambda) + \frac{1}{2}(1-\lambda)^2 = \frac{1}{2}(1+\lambda^2), && (3.91a) \\
\mathrm{tr}\left[\mathsf{E}^k(\pm)\mathsf{E}^k(\mp)\right] &= \lambda^2 \mathrm{tr}\left[\mathsf{P}^k(\pm)\mathsf{P}^k(\mp)\right] + \frac{1}{2}\lambda(1-\lambda)\mathrm{tr}\left[\mathsf{P}^k(\pm) + \mathsf{P}^k(\mp)\right] + (1-\lambda)^2\frac{1}{4}\mathrm{tr}\left[I\right] \\
&= \lambda(1-\lambda) + \frac{1}{2}(1-\lambda)^2 = \frac{1}{2}(1-\lambda^2). && (3.91b)
\end{aligned}$$

We now presuppose the existence of a set of four operators $\{\mathsf{G}(i)\}_{i=0}^3$ such that the effects of these mutually unbiased POVMs can be derived in the form of Equation (3.51). In such a case, the $\mathsf{G}(i)$ operators would be recovered by summing over these effects and noting that $\sum_{i=0}^3 \mathsf{G}(i) = I$:

$$\begin{aligned}
\mathsf{E}^1(+) + \mathsf{E}^2(+) + \mathsf{E}^3(+) &= 3\mathsf{G}(0) + \mathsf{G}(1) + \mathsf{G}(2) + \mathsf{G}(3) = 2\mathsf{G}(0) + I, \\
\mathsf{E}^1(+) + \mathsf{E}^2(-) + \mathsf{E}^3(-) &= 3\mathsf{G}(1) + \mathsf{G}(0) + \mathsf{G}(2) + \mathsf{G}(3) = 2\mathsf{G}(1) + I, \\
\mathsf{E}^1(-) + \mathsf{E}^2(+) + \mathsf{E}^3(-) &= 3\mathsf{G}(2) + \mathsf{G}(0) + \mathsf{G}(1) + \mathsf{G}(3) = 2\mathsf{G}(2) + I, \\
\mathsf{E}^1(-) + \mathsf{E}^2(-) + \mathsf{E}^3(+) &= 3\mathsf{G}(3) + \mathsf{G}(0) + \mathsf{G}(1) + \mathsf{G}(2) = 2\mathsf{G}(3) + I.
\end{aligned} \tag{3.92}$$

In other words, the operators $\mathsf{G}(i)$ can be expressed in the following form:

$$\begin{aligned}
\mathsf{G}(0) &= \frac{1}{2}(\mathsf{E}^1(+) + \mathsf{E}^2(+) + \mathsf{E}^3(+) - I), \\
\mathsf{G}(1) &= \frac{1}{2}(\mathsf{E}^1(+) + \mathsf{E}^2(-) + \mathsf{E}^3(-) - I), \\
\mathsf{G}(2) &= \frac{1}{2}(\mathsf{E}^1(-) + \mathsf{E}^2(+) + \mathsf{E}^3(-) - I), \\
\mathsf{G}(3) &= \frac{1}{2}(\mathsf{E}^1(-) + \mathsf{E}^2(-) + \mathsf{E}^3(+) - I).
\end{aligned} \tag{3.93}$$

Using Equation (3.90), the $\mathsf{G}(i)$ can be expressed in terms of the MUB projections $\mathsf{P}^k(\pm)$, and further to that expressed in terms of their Bloch representation:

$$\begin{aligned}\mathsf{G}(i) &= \frac{1}{2}(\mathsf{E}^1(\pm) + \mathsf{E}^2(\pm) + \mathsf{E}^3(\pm) - I) \\ &= \frac{1}{2}\left(\lambda(\mathsf{P}^1(\pm) + \mathsf{P}^2(\pm) + \mathsf{P}^3(\pm)) + \frac{3}{2}(1-\lambda)I - I\right) \\ &= \frac{1}{2}\left(\lambda\left(\frac{3}{2}I + \frac{1}{2}(\pm\boldsymbol{n}_1 \pm \boldsymbol{n}_2 \pm \boldsymbol{n}_3)\cdot\boldsymbol{\sigma}\right) + \frac{1}{2}(1-3\lambda)I\right) \\ &= \frac{1}{4}(I + \lambda\boldsymbol{s}\cdot\boldsymbol{\sigma}),\end{aligned} \tag{3.94}$$

where $\boldsymbol{s} = \pm\boldsymbol{n}_1 \pm \boldsymbol{n}_2 \pm \boldsymbol{n}_3$. Because the projections $\mathsf{P}^k(\pm)$ are mutually unbiased for different k, it follows that their respective Bloch vectors must be perpendicular. As a result of this, the vector $\boldsymbol{s}$ has square norm $\|\boldsymbol{s}\|^2 = \sum_{i=1}^{3}\|\boldsymbol{n}_i\|^2 = 3$. The operator $\mathsf{G}(i)$ therefore has eigenvalues $(1/4)(1 \pm \sqrt{3}\lambda)$, and so whilst the $\mathsf{G}(i)$ are always self-adjoint, they are positive iff $\lambda \leq 1/\sqrt{3}$.

By checking Equation (3.93), we can see that

$$\sum_{i=0}^{3}\mathsf{G}(i) = \sum_{k=1}^{3}\mathsf{E}^k(+) + \sum_{k=1}^{3}\mathsf{E}^k(-) - 2I = I, \tag{3.95}$$

as we expected (and, indeed, relied upon). We next note that any two of the $\mathsf{G}(i)$ share just one effect, whilst the remaining effects are different. The product of any two therefore contains one term of the form $\mathsf{E}^k(\pm)^2$, two of the form $\mathsf{E}^k(\pm)\mathsf{E}^k(\mp)$ and 6 of the form $\mathsf{E}^k(\pm)\mathsf{E}^\ell(\mp)$, where $\ell \neq k$, as well as 6 instances of a product of an effect with the identity and one of the identity with itself. Hence,

$$\mathrm{tr}\left[\mathsf{G}(i)\mathsf{G}(j)\right] = \frac{1}{4}\left(\frac{1}{2}(1+\lambda^2) + (1-\lambda^2) + \frac{6}{2} - 6 + 2\right) = \frac{1}{8}(1-\lambda^2). \tag{3.96}$$

From this, we see that the symmetric property $\mathrm{tr}\left[\mathsf{G}(i)\mathsf{G}(j)\right] = 1/12$ is satisfied iff $\lambda = 1/\sqrt{3}$. In this instance, the effects $\mathsf{E}^k(\pm)$ are positive, and so the $\mathsf{G}(i)$ form not only a POVM, but a SIC-POVM.

3.3.2 An additional combinatorial construction

In order to generalise the construction used in the qubit case, we first need to consider one further combinatorial structure.

Consider a $(d+1) \times d$ array of points that possesses a system of paths. Each path is of length $d+1$—that is, each path coincides with $d+1$ points in the array—and contains one element from each row of the array. Further to this, we assume that the paths share an analogue of the one-overlap property: any two paths through this array intersect in exactly one point.

Lemma 3.3.1. Let $\varepsilon = \{\varepsilon_{ij}\}$ be a $(d+1) \times d$ array of points. There exist at most d^2 paths p_i through ε containing one point from each row such that any two paths satisfy the one-overlap property.

Proof. Consider every path beginning at the j^{th} point on the first row; that is, at the point

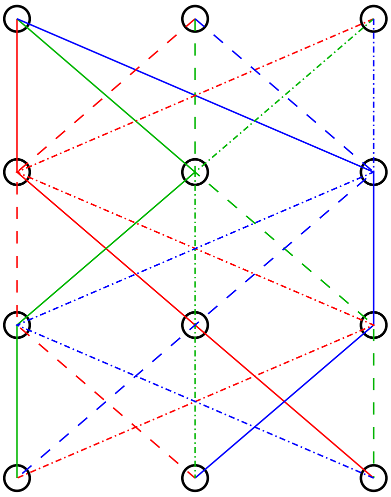


Figure 3.7: An example of 9 possible paths that overlap once through a 4×3 array such that every path coincides with each row once. Note that no further path could be added that would only overlap with each other path once.

ε_{1j}. Each of these paths shares ε_{1j} in common and after this these paths can no longer intersect. In which case, these paths must coincide with different points on the second row, of which there are d, i.e., there can be at most d paths of the type described starting at ε. We may perform the same argument for any of the points on the first row of ε, of which there are d, and so we must have at most d^2 such paths through ε. □

An example of this construction for $d = 3$ is given in Figure 3.7. Given the numbers considered here, namely $d(d+1)$ and d^2, we may suspect that there is some connection between this array and the $d \times d$ arrays considered in Section 3.2.2. This is indeed the case:

Proposition 3.3.2. Consider a $d \times d$ array A with a complete set of $d+1$ d-partitions $\{\mathcal{P}^k\}$ satisfying the one-overlap property and a $(d+1) \times d$ array ε with a complete set of d^2 paths of length $d+1$ that coincide with one element from each row and any two paths overlap once. Then A and ε are equivalent, with the points and paths in A corresponding to the paths and points in ε, respectively.

Proof. We shall start with the $d \times d$ array $A = \{a_{ij}\}$ and the $d+1$ d-partitions $\{\mathcal{P}^k\}$ satisfying the one-overlap property. We construct a $(d+1) \times d$ array $\varepsilon = \{\varepsilon_{ij}\}$ in the following way: each point in ε corresponds to a bin of a partition of A and the d bins of a given partition of A form a row of ε, i.e., the point $\varepsilon_{i\mu}$ corresponds to the bin $\mathcal{P}^i_\mu$ and the bins of the partition $\mathcal{P}^j$ form the j^{th} row of ε.

Since each element $a_{ij} \in A$ belongs to a bin in each partition, we may construct a path p_{ij} connecting every bin containing a_{ij}. We denote by

$$r_k : \{1, 2, \ldots, d\} \times \{1, 2, \ldots, d\} \to \{1, 2, \ldots, d\}, \tag{3.97}$$

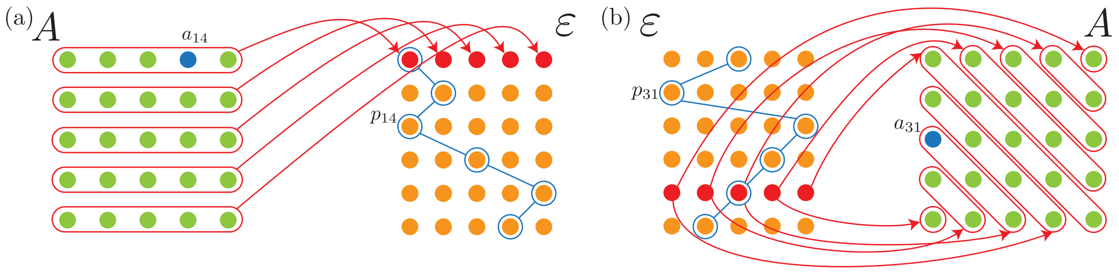


Figure 3.8: The equivalence between a $d \times d$ array A with $d+1$ d-partitions satisfying the one-overlap property and a $(d+1) \times d$ array ε with d^2 downward paths satisfying the one-overlap property. (a) Starting with the $d \times d$ array A, for each d-partition we associate with each bin a point in the $(d+1) \times d$ array ε, with a given d-partition forming a row of ε. For each point a_{ij} in A we then define a path p_{ij} through ε that passes through each point corresponding to a bin containing a_{ij}. (b) If we now start with the $(d+1) \times d$ array ε, we associate the path p_{ij} through ε, which is the j^{th} path to start at the point ε_{1i}, with the point a_{ij} in the $d \times d$ array A. Every point ε_{ij} is then associated with the j^{th} bin of i^{th} d-partition of A corresponding to the paths through ε that pass through ε_{ij}.

with $k = 1, \ldots, d+1$, the function that performs the map for which

$$a_{ij} \in \mathcal{P}^{k}_{r_k(i,j)}. \tag{3.98}$$

In other words, r_k maps (i,j) to the bin in the k^{th} partition that contains the element a_{ij}. Note that, since a_{ij} can only belong in one bin per partition, due to the disjoint nature of partitions, there is no ambiguity when we talk about the bin a_{ij} belongs to for a given partition. With this notation, we denote the path p_{ij} through ε as follows:

$$p_{ij} = \left\{ \mathcal{P}^{1}_{r_1(i,j)}, \mathcal{P}^{2}_{r_2(i,j)}, \ldots, \mathcal{P}^{d+1}_{r_{d+1}(i,j)} \right\}. \tag{3.99}$$

Since each element of A belongs to only one bin per partition, every path constructed in this way coincides with each row only once. This construction is given in Figure 3.8 (a).

A given path p_{ij} on ε coincides with $d+1$ bins that, by way of the one-overlap property between these bins, can only have the element a_{ij} in common. Each bin in this path hence contains $d-1$ elements of A that are distinct from a_{ij} and the elements of the other bins in the path. As a result, the path contains $(d+1)(d-1)+1 = d^2$ distinct elements of A, i.e., every element of A is contained within each path on ε, and for every $a_{ij} \in A$ and path $p_{\mu\nu}$ on ε, there is a $k = 1, \ldots, d+1$ such that

$$a_{ij} \in \mathcal{P}^{k}_{r_k(\mu,\nu)}. \tag{3.100}$$

Because of this, any two paths must overlap at least once: trivially, $a_{ij} \in \mathcal{P}^{\ell}_{r_\ell(i,j)}$ for all $\ell = 1, \ldots, d+1$, and there must exist a k such that $a_{ij} \in \mathcal{P}^{k}_{r_k(\mu,\nu)}$ by the preceding argument. In which case,

$$\left| \mathcal{P}^{k}_{r_k(i,j)} \cap \mathcal{P}^{k}_{r_k(\mu,\nu)} \right| \neq \emptyset, \tag{3.101}$$

and so the sets must coincide due to the disjoint nature of partitions. Hence, the paths

p_{ij} and $p_{\mu\nu}$ overlap on the k^{th} row.

Let us now suppose that two paths, p_{ij} and $p_{\mu\nu}$, overlap twice, i.e., there exist k and ℓ such that

$$\mathcal{P}^k_{r_k(i,j)} = \mathcal{P}^k_{r_k(\mu,\nu)} \quad \text{and} \quad \mathcal{P}^\ell_{r_\ell(i,j)} = \mathcal{P}^\ell_{r_\ell(\mu,\nu)}. \tag{3.102}$$

By the definition of these paths, this is equivalent to the existence of two elements $a_{ij}, a_{\mu\nu} \in A$ such that

$$a_{ij}, a_{\mu\nu} \in \mathcal{P}^k_{r_k(i,j)} \quad \text{and} \quad a_{ij}, a_{\mu\nu} \in \mathcal{P}^\ell_{r_\ell(i,j)}. \tag{3.103}$$

However, this violates the one-overlap property that these bins were assumed to satisfy, and so it cannot be that these paths p_{ij} and $p_{\mu\nu}$ overlap more than once. Hence, we have constructed d^2 paths of length $d+1$ on ε that coincide with one element from each row and overlap with any other path only once.

Consider now the $(d+1) \times d$ array $\varepsilon = \{\varepsilon_{ij}\}$ with d^2 paths as described above. We construct a $d \times d$ array $A = \{a_{ij}\}$ by associating each point $a_{i\mu}$ with the path $p_{i\mu}$, where $p_{i\mu}$ corresponds to the μ^{th} path that starts at the point $\varepsilon_{1i} \in \varepsilon$, i.e., the first row of A corresponds to the paths that start at ε_{11}, etc.

For each point $\varepsilon_{ij} \in \varepsilon$, there are d paths that coincide with it, and none of these paths will coincide with any other element from the i^{th} row of ε. Using a similar notation as before, we denote by

$$\mu_k : \{1, 2, \dots, d+1\} \times \{1, 2, \dots, d\} \to \{1, 2, \dots, d\}, \tag{3.104}$$

with $k = 1, 2, \dots, d$, the map for which

$$\varepsilon_{ij} \in p_{k\mu_k(i,j)}. \tag{3.105}$$

That is, μ_k maps (i, j) to the path through ε that starts at the point ε_{1k} and contains ε_{ij}. We therefore have associated with each point $\varepsilon_{ij} \in \varepsilon$, with $i \neq 1$, the set

$$\mathcal{P}^i_j = \{p_{1\mu_1(i,j)}, \dots, p_{d\mu_d(i,j)}\}, \tag{3.106}$$

and for $i = 1$ we have

$$\mathcal{P}^1_j = \{p_{j1}, \dots, p_{jd}\}. \tag{3.107}$$

These sets each contain d paths and, for a fixed i, the sets $\mathcal{P}^i_k$ and $\mathcal{P}^i_\ell$ are disjoint, otherwise they would contain paths through ε which coincide with the i^{th} row twice, which is not possible by construction. In other words, the set $\{\mathcal{P}^i_k\}_{k=1}^d$ for a fixed i forms a partition of the paths through ε, which, by construction, correspond to the points in A, i.e., we have constructed d-partitions of A. As a trivial example, the set $\{\mathcal{P}^1_k\}_{k=1}^d$ forms a partition of A into horizontal bins. Given each row corresponds to a partition of A, and ε possesses $d+1$ rows, we have a set of $d+1$ d-partitions of A. This process is illustrated in Figure 3.8 (b).

From the one-overlap property for paths on ε, we find a one-overlap property between the bins of the partitions on A. Suppose that the intersection of two bins from different partitions of A, $\mathcal{P}^i_\mu$ and $\mathcal{P}^j_\nu$, contains two elements. This means that there are two paths through ε that share two elements in common, namely $\varepsilon_{i\mu}$ and $\varepsilon_{j\nu}$. This is, however, a violation of the one-overlap property for these paths, and so it must be that these bins from different partitions of A cannot share more than one element in common. If, on the other hand, the intersection of two bins from different partitions is empty, then this corresponds to two points in ε on different rows that are not connected by a path. However, consider the paths that intersect at any point in ε: there are d such paths that cannot intersect again, and so on every subsequent (and preceding) row they must be dispersed separately amongst the d elements on each of these rows. This means that if two points in ε on different rows are not connected by a path, then there exists a point on one of these rows upon which two paths intersect for a second time, which we have already shown not to be possible. Hence, the bins from these partitions must intersect once and only once, and so we have shown that the $d+1$ d-partitions of A satisfy the one-overlap property between bins. This concludes the proof. □

3.3.3 SIC systems from mutually unbiased POVMs

We now generalise the results for the qubit case, and see that further restrictions are required in order to guarantee that what we derive is indeed a SIC-POVM.

Let us begin with an ensemble of $d+1$ d-outcome POVMs E^k, $k = 1, \ldots, d+1$. By denoting the μ^{th} effect of E^k by $\mathsf{E}^k(\mu)$, we construct the $(d+1) \times d$ array $\varepsilon = \{\varepsilon_{ij}\}$ with $\varepsilon_{k\mu} = \mathsf{E}^k(\mu)$:

$$\varepsilon = \begin{matrix} \mathsf{E}^1(1) & \ldots & \mathsf{E}^1(d) \\ \vdots & \ddots & \vdots \\ \mathsf{E}^{d+1}(1) & \ldots & \mathsf{E}^{d+1}(d) \end{matrix} . \tag{3.108}$$

As we did in the qubit case, we presuppose that the POVMs E^k are margins of a d^2-outcome observable $\mathsf{G} = \{\mathsf{G}(i)\}_{i=0}^{d^2-1}$ such that each POVM is associated with a d-partition of G and the partitions satisfy the one-overlap property with respect to each other. As a result of this, no effects from the same POVM can share an operator $\mathsf{G}(i)$, whilst any two effects from different POVMs must share just one. We can associate with each effect $\mathsf{G}(i)$ a path p_i through ε such that p_i coincides with the effects that contain $\mathsf{G}(i)$. We immediately have the following properties for these paths:

1. Each effect $\mathsf{G}(i)$ occurs in only one effect for each of these POVMs, so its associated path must be strictly downwards, i.e., p_i coincides with only one point per row in ε. Hence, each path p_i is of length $d+1$;

2. Since any two effects of these margin POVMs share only one operator $\mathsf{G}(i)$ in common, the paths through ε must only intersect once: suppose that the paths p_i and p_j intersect twice, then there would exist two effects from different POVMs that share two operators, namely $\mathsf{G}(i)$ and $\mathsf{G}(j)$, in common, which violates the requirements of the construction.

This is clearly the combinatorial structure that we have just discussed, and so, by Proposition 3.3.2, there must exist d^2 downward paths of length $d+1$ in ε satisfying the one-overlap property.

Consider the path p_i, which connects all effects in the array ε that contain the operator $\mathsf{G}(i)$. We now define the operator

$$E_i = \sum_k \mathsf{E}^k(\nu(i,k)), \tag{3.109}$$

where, similarly to the previous section, $\nu(i,k)$ is a map such that $\mathsf{E}^k(\nu(i,k))$ is the effect that contains $\mathsf{G}(i)$ in the k^{th} POVM. Each effect included in the path p_i has $\mathsf{G}(i)$ in common, but otherwise the operators $\mathsf{G}(j)$, $j \neq i$, contained in the effects are distinct. This path is of length $d+1$ and each effect corresponds to a bin of a d-partition, so the path must contain $d-1$ operators distinct from $\mathsf{G}(i)$ and the operators in every other effect in the path. Hence, we require $(d+1)(d-1) = d^2 - 1$ operators other than $\mathsf{G}(i)$ in order to describe E_i. Since the POVMs E^k arise from partitions of G, it follows that

$$\sum_{i=0}^{d^2-1} \mathsf{G}(i) = \sum_{\mu} \mathsf{E}^k(\mu) = I, \tag{3.110}$$

for all k. Hence, the d^2-1 operators $\mathsf{G}(j)$, with $j \neq i$, must add up to $I - \mathsf{G}(i)$, and so

$$\begin{aligned} E_i &= \sum_k \mathsf{E}^k(\nu(i,k)) \\ &= (d+1)\mathsf{G}(i) + \sum_{j\neq i} \mathsf{G}(j) \\ &= (d+1)\mathsf{G}(i) + (I - \mathsf{G}(i)) \\ &= d\mathsf{G}(i) + I. \end{aligned} \tag{3.111}$$

By rearranging we are led to the following proposition:

Proposition 3.3.3. Assume that a d^2-outcome POVM G possesses $d+1$ margins E^k associated with a complete system of d-partitions that satisfy the one-overlap property with respect to each other. We may retrieve the operators $\mathsf{G}(i)$ from the margin POVMs via

$$\mathsf{G}(i) = \frac{1}{d}(E_i - I), \tag{3.112}$$

with E_i given by Equation (3.109).

Theorem 3.3.4. Consider a family of $d+1$ POVMs E^k that are mutually unbiased, i.e.,

$$\operatorname{tr}\left[\mathsf{E}^k(\mu)\mathsf{E}^\ell(\nu)\right] = \frac{1}{d}, \quad k,\ell = 1,\dots,d+1,\ \mu,\nu = 1,\dots,d, \tag{3.113}$$

whenever $k \neq \ell$, and whose effects satisfy

$$\operatorname{tr}\left[\mathsf{E}^k(\mu)\mathsf{E}^k(\nu)\right] = \frac{1}{d+1}, \quad k = 1,\dots,d+1,\ \nu \neq \mu. \tag{3.114}$$

Assume that there exist d^2 sets p_i, $i \in \mathbb{Z}_{d^2}$, each composed of $d+1$ effects with one taken from each POVM such that the one-overlap property is satisfied. If we denote by E_i the sum of effects in a given set, then the d^2 operators $\mathsf{G}(i) := (1/d)(E_i - I)$ form a SIC-POVM iff the $\mathsf{G}(i)$ are positive.

Proof. The first thing we note is that from Equations (3.113) and (3.114), we also have that

$$\mathrm{tr}\left[\mathsf{E}^k(\mu)\right] = 1, \quad \mathrm{tr}\left[\mathsf{E}^k(\mu)^2\right] = \frac{2}{d+1}, \tag{3.115}$$

for all $k = 1, \ldots, d+1$ and $\mu = 1, \ldots, d$. The first of these comes from Equation (3.113) and the normalisation of these POVMs:

$$\mathrm{tr}\left[\mathsf{E}^k(\mu)\right] = \mathrm{tr}\left[\mathsf{E}^k(\mu)I\right] = \sum_{\nu} \mathrm{tr}\left[\mathsf{E}^k(\mu)\mathsf{E}^\ell(\nu)\right] = d \cdot \frac{1}{d} = 1, \tag{3.116}$$

meanwhile the second is a result of Equations (3.114) and (3.116):

$$\mathrm{tr}\left[\mathsf{E}^k(\mu)^2\right] = \mathrm{tr}\left[\mathsf{E}^k(\mu)\left(I - \sum_{\nu\neq\mu} \mathsf{E}^k(\nu)\right)\right] = 1 - \frac{d-1}{d+1} = \frac{2}{d+1}. \tag{3.117}$$

As a result of Equation (3.116), the operators E_i must have trace $d+1$, and so

$$\mathrm{tr}\left[\mathsf{G}(i)\right] = \frac{1}{d}\mathrm{tr}\left[E_i - I\right] = \frac{1}{d}(d+1) - 1 = \frac{1}{d}. \tag{3.118}$$

Next, in calculating $\mathrm{tr}\left[\mathsf{G}(i)^2\right]$ we note that $\mathrm{tr}\left[E_i^2\right]$ contains $d+1$ terms of the form $\mathrm{tr}\left[\mathsf{E}^k(\nu)^2\right] = 2/(d+1)$ and $d(d+1)$ of the form $\mathrm{tr}\left[\mathsf{E}^k(\mu)\mathsf{E}^\ell(\nu)\right] = 1/d$. Hence,

$$\begin{aligned} \mathrm{tr}\left[\mathsf{G}(i)^2\right] &= \frac{1}{d^2}\mathrm{tr}\left[E_i^2 - 2E_i + I\right] \\ &= \frac{1}{d^2}\left(\frac{2(d+1)}{d+1} + \frac{d(d+1)}{d} - 2(d+1) + d\right) \\ &= \frac{1}{d^2}. \end{aligned} \tag{3.119}$$

If we now consider the product $\mathsf{G}(i)\mathsf{G}(j)$, the trace of the product E_iE_j is only slightly more complicated: since these operators correspond to two sets, p_i and p_j, which share just one element in common, we have one term of the form $\mathrm{tr}\left[\mathsf{E}^k(\mu)^2\right] = 2/(d+1)$ and d of the form $\mathrm{tr}\left[\mathsf{E}^k(\mu)\mathsf{E}^k(\nu)\right] = 1/(d+1)$, followed by $d(d+1)$ of the form $\mathrm{tr}\left[\mathsf{E}^k(\mu)\mathsf{E}^\ell(\nu)\right] = 1/d$, as in the previous case. Hence,

$$\begin{aligned} \mathrm{tr}\left[\mathsf{G}(i)\mathsf{G}(j)\right] &= \frac{1}{d^2}\mathrm{tr}\left[E_iE_j - E_i - E_j + I\right] \\ &= \frac{1}{d^2}\left(\frac{2}{d+1} + \frac{d}{d+1} + \frac{d(d+1)}{d} - 2(d+1) + d\right) \\ &= \frac{1}{d^2(d+1)}. \end{aligned} \tag{3.120}$$

These trace conditions are exactly those found with the effects of a SIC-POVM. Further to this, we have already shown that $\sum_i \mathsf{G}(i) = I$, so the $\mathsf{G}(i)$ are both symmetric and, by being d^2 linearly independent operators, informationally complete. However, at no point have we assumed positivity of these operators, nor have shown it, and so $\mathsf{G} = \{\mathsf{G}(i)\}$ is a SIC-POVM iff the operators $\mathsf{G}(i)$ are all positive. □

As is concluded at the end of Theorem 3.3.4, there is no reason to assume that the

$\mathsf{G}(i)$ will be positive. We therefore provide the following definition

Definition 3.3.1. Consider a set of d^2 operators $\{\mathsf{G}(i)\}$ satisfying

$$\mathrm{tr}\left[\mathsf{G}(i)\right] = \frac{1}{d}, \quad \mathrm{tr}\left[\mathsf{G}(i)^2\right] = \frac{1}{d^2}, \quad \mathrm{tr}\left[\mathsf{G}(i)\mathsf{G}(j)\right] = \frac{1}{d^2(d+1)}, \tag{3.121}$$

and $\sum_i \mathsf{G}(i) = I$. Such a set of operators is called a *SIC system.* If positivity can be guaranteed for all d^2 operators, then the SIC system forms a SIC-POVM.

Hence, Theorem 3.3.4 tells us that from a complete set of mutually unbiased POVMs we may always construct a SIC system.

Let us now suppose that we possess a collection of MUBs $\{\varphi_i^k\}_{i=1}^d$, with $k = 1, \dots, n$ with associated PVMs $\mathsf{P}^k(i) = \left|\varphi_i^k\right\rangle\left\langle\varphi_i^k\right|$. We now, for each PVM, take a stochastic matrix $\Lambda^k = [\lambda_{ij}^k]$ and smear the PVMs, as in Equation (3.68), to create d-valued POVMs:

$$\mathsf{E}^k(\nu) = \sum_i \lambda_{\nu i}^k \mathsf{P}^k(i). \tag{3.122}$$

Proposition 3.3.5. The POVMs E^k obtained by smearing mutually unbiased PVMs via Equation (3.68) are themselves mutually unbiased iff the effects $\mathsf{E}^k(\nu)$ of each POVM are unit trace (this corresponds to each of the Λ^k being a doubly stochastic matrix).

Proof. If we assume that the effects are unit trace, then by Theorem 3.2.8 we have that they are also mutually unbiased.

Conversely, if the d-valued POVMs E^k and E^ℓ are mutually unbiased, then from the normalisation of these POVMs,

$$\mathrm{tr}\left[\mathsf{E}^k(\nu)\right] = \sum_\mu \mathrm{tr}\left[\mathsf{E}^k(\nu)\mathsf{E}^\ell(\mu)\right] = \sum_\mu \frac{1}{d} = 1, \tag{3.123}$$

which concludes the proof. □

By making use of Proposition 3.3.5 and Theorem 3.3.4, we see that by starting with a complete set of MUBs, we may smear them using doubly stochastic matrices so that the resultant POVMs are mutually unbiased. If, further to this, the effects of a given POVM satisfy Equation 3.114, then we may construct a SIC system from these mutually unbiased POVMs, and even, in some instances, a SIC-POVM if the operators derived via Equation (3.112) are positive.

3.4 A comparison with the work of Wootters.

As part of a 2006 Festschrift honouring Asher Peres [55], Wootters published a paper describing a similar structure to the one presented in this chapter. We will begin here by describing his construction and how it can relate to SIC-POVMs and MUBs. We will then highlight where the differences lie between our constructions, and see that whilst what we have presented here is less simple than the work of Wootters, it does not possess the same shortcomings.

3.4.1 From mutually unbiased striations to MUBs

The geometric construction considered by Wootters is defined as follows:

Definition 3.4.1. Consider a set A of d^2 points. A *striation* on A is a partition of the points into d *parallel lines* such that no point is contained in more than one line. Two striations are *mutually unbiased* if any two lines from the different striations coincide at only one point.

This structure is very similar in nature to the definition of d-partitions that we have given above, and so we know that for a set A of d^2 points we can construct up to $d+1$ mutually unbiased striations. Further, we can place these points in an array such that we may construct "Cartesian striations", and so any striations that are mutually unbiased to these will correspond to a Latin square of order d, with mutually unbiased striations leading to mutually orthogonal Latin squares.

Where our work differs from that of Wootters is in the assignment of the points and lines: each point $\alpha \in A$ corresponds to a self-adjoint operator C_α/d, whilst every line λ belonging to a striation corresponds to a rank-one projection $\mathsf{P}(\lambda)$. The operators $\{C_\alpha\}$ satisfy the following properties:

1. $\mathrm{tr}\,[C_\alpha/d] = 1/d$ for all $\alpha \in A$;
2. $\mathrm{tr}\,[(C_\alpha/d)(C_\beta/d)] = (1/d)\delta_{\alpha\beta}$;
3. $\sum_{\alpha\in\lambda} C_\alpha/d = \mathsf{P}(\lambda)$.

From these properties, we have that the $\mathsf{P}(\lambda)$ associated with a given striation are normalised and mutually orthogonal, and therefore linearly independent. Since we have d such rank-one projection operators, their associated vector states form an orthonormal basis for $\mathcal{H}_d$. Furthermore, given that any two lines, λ and μ, from different striations coincide at just one point, the trace of their respective projections, $\mathsf{P}(\lambda)$ and $\mathsf{P}(\mu)$, must equal $1/d$. Therefore the bases associated with mutually unbiased striations are themselves mutually unbiased.

There are similarities between our construction and the one of Wootters. Whilst his construction is more immediate in its derivation of MUBs—appearing as lines in mutually unbiased striations as opposed to ours, where the bins of partitions correspond to mutually unbiased POVM effects that lead to MUBs if the POVMs are commutative—the degree to which the operators C_α is known beforehand is much less. Whilst he possesses some trace properties that the operators must satisfy, there is not greater detail available from this construction, and it is not in general obvious how the C_α are constructed. By contrast, we know that the operators that we partition are the effects of a SIC-POVM, and so we already have the starting operators constructed.

3.4.2 Affine planes

Before moving onto the construction for SIC-POVMs, we consider a final combinatorial structure, of which a set of d^2 points and $d+1$ mutually unbiased striations are an example.

Definition 3.4.2. An *affine plane of order d* is a set A of d^2 points and $d(d+1)$ lines such that

1. For any two points there exists exactly one line that passes through both;
2. For any point $\alpha \in A$ and line λ, there exists a single line that is parallel (non-intersecting) to λ and coincides with α;
3. There exist three points that are non-collinear, i.e., there exist 3 points such that there does not exist a line that is coincident with all 3.

Given the existence of complete sets of mutually unbiased striations for prime power dimensions, we know that affine planes must exist for prime power orders. Otherwise, there are only certain details known for non-prime power dimensions, including the lack of existence of an affine plane of order 6[1].

3.4.3 Construction of a SIC-POVM

We will now try to use this construction to find a SIC-POVM G associated with a set A of points, and lines defined on A such that the sum of the elements of a line form a rank-one projection $\mathsf{P}(i) = d\mathsf{G}(i)$, where $\mathsf{G}(i)$ is an effect belonging to G. At this point, we do not know how many elements are in A, and in order to determine the number of points needed, both belonging to a line and in A in general, Wootters made use of an idea given by Zauner in his thesis [58]. This idea provides a bridge between the cardinality of a set of points and the trace of the operator corresponding to that set: If M is an operator corresponding to a subset of elements of a geometric construct, then let that subset be denoted by S_M. For example, in the case of the array given in Section 3.4.1, the set corresponding to the operator C_α would be $S_{C_\alpha} = \{\alpha\} \subset A$. We now consider relations of the form

$$|S_M| = k\,\mathrm{tr}\,[M] \quad \text{and} \quad |S_{M_1} \cap S_{M_2}| = k\,\mathrm{tr}\,[M_1 M_2]\,, \tag{3.124}$$

for some constant k. The rank-one projection $\mathsf{P}(\lambda)$ associated with the line λ must be of unit trace, and so must contain k points within it. From the symmetry property for the effects, $\mathsf{G}(\mu)$ and $\mathsf{G}(\nu)$, of a SIC-POVM, $\mathrm{tr}\,[\mathsf{P}(\mu)\mathsf{P}(\nu)] = 1/(d+1)$ where $\mathsf{P}(\mu) = d\mathsf{G}(\mu)$, etc. Hence, the lines μ and ν must intersect $k/(d+1)$ times, and so k must be a multiple of $d+1$. We will restrict ourselves to $k = d+1$, and so any two lines intersect once through this set of points. We assume that the operator associated with the entire set of points is the identity, and that every point coincides with the same number of lines as any other. Hence, $|S_A| = k\mathrm{tr}\,[I] = d(d+1)$ and so every point is contained within d lines.

From this reasoning, we have seen that we require $d(d+1)$ points in our set, and any two of the d^2 lines must intersect at one point. If we were to relabel the points and lines, then this construction produces an affine plane of order d. This parallelism between the constructions of MUBs and SIC-POVMs, as presented by Wootters, is similar to that shown for our constructions, in particular Proposition 3.3.2. However, this holds only in

[1] This is due to the result that there exists an affine plane of order n iff there exists $n-1$ mutually orthogonal Latin squares of order n, which we have already stated does not exist for $n = 6$.

the case $k = d + 1$, which was an assumption that we chose to make, but comes naturally in the construction we presented in the preceding discussions.

Given this connection with affine planes, we are restricted in the dimensions that this construction works. Assuming that we can indeed form an affine plane of order d, and hence form a set of $d(d+1)$ points and d^2 lines intersecting only once, what properties do the associated operators D_α need to satisfy so that the lines correspond to projections arising from SIC-POVM effects? From the trace properties of these projections, we require that

1. $\mathrm{tr}\left[D_\alpha^2\right] = d/(d+1)^2$;
2. $\mathrm{tr}\left[D_\alpha D_\beta\right] = 1/d(d+1)^2$ if $\alpha \neq \beta$ and they share a line;
3. $\mathrm{tr}\left[D_\alpha D_\beta\right] = -1/(d+1)^2$ if $\alpha \neq \beta$ but they do not share a line.

Similar to the case for constructing MUBs from mutually unbiased striations, this construction is more immediate in finding the effects of a SIC-POVM than the method that we have presented in this chapter. However, there is again the issue that there is no obvious construction for the D_α operators, with the simplest method being to start with a SIC-POVM and work backwards. By contrast, we are aware of the operators that we begin with, although the derivation of a SIC-POVM via downward paths requires a greater amount of work near the end of the process.

3.5 Investigations in Mathematica

We will now address some results that have arisen from applying the constructions discussed in this chapter in Mathematica. Although this is only a preliminary investigation, it has provided some interesting evidence for these constructions and, as has been mentioned at times in this chapter, while what has been presented thus far is indeed capable of working, there exist issues, including the scarcity of positive results.

Due to their well-known structure, and their relative ease to construct, we restrict ourselves to Weyl-Heisenberg covariant SIC-POVMs, as we introduced in Section 3.1.2. Indeed, by defining the phase, Z, and shift, X, operators accordingly, we may construct as many SIC-POVMs for a given dimension as there are fiducial vectors, of which there exists an exhaustive list for dimensions up to 67 [48].

Given the comparative ease of solving it analytically, we omit the case of qubit SIC-POVMs from these investigations.

3.5.1 $d = 3$

In the case of $d = 3$, there exists a continuous range of fiducial vectors that can be used to produce SIC-POVMs [44]. By considering the Clifford group $C(d)$—the normaliser of the Weyl-Heisenberg group $W(d)$, i.e., $C(d)$ is the subgroup of $U(d)$ such that, for any $U \in C(d)$, $UW(d)U^* = W(d)$—one finds that the fiducial vectors split into 13 classes of inequivalent SIC-POVMs [52].

In what follows, we restrict ourselves to constructing Weyl-Heisenberg covariant SIC-POVMs using fiducial vectors according to [44]: Let $r_0 \in (1/\sqrt{2}, \sqrt{2/3}]$ and define

$$r_\pm(r_0) = \frac{1}{2}r_0 \pm \frac{1}{2}\sqrt{2-3r_0^2}. \tag{3.125}$$

The range of possible fiducial vectors, in the computations basis, are of the form

$$\begin{aligned} &\left\{ \begin{pmatrix} r_0 \\ r_+e^{i\theta_1} \\ r_-e^{i\theta_2} \end{pmatrix} \text{ and all permutations } \middle| \theta_1, \theta_2 \in \left\{\frac{\pi}{3}, \pi, \frac{5\pi}{3}\right\}, r_0 \in \left(\frac{1}{\sqrt{2}}, \sqrt{\frac{2}{3}}\right] \right\} \\ &\qquad \bigcup \left\{ \begin{pmatrix} 1/\sqrt{2} \\ e^{i\theta}/\sqrt{2} \\ 0 \end{pmatrix} \text{ and all permutations } \middle| \theta \in [0, 2\pi) \right\}. \end{aligned} \tag{3.126}$$

In these preliminary investigations 5 fiducial vectors were chosen, corresponding to the following choices of r_0:

$$\begin{gathered} \varphi_1 : r_0 = \sqrt{\frac{2}{3}}, \qquad \varphi_2 : r_0 = \frac{4}{5}, \qquad \varphi_3 : r_0 = \frac{1}{2}\left(\frac{1}{\sqrt{2}} + \sqrt{\frac{2}{3}}\right), \\ \varphi_4 : r_0 = \frac{3}{4}, \qquad \varphi_5 : r_0 = \frac{1}{\sqrt{2}}. \end{gathered} \tag{3.127}$$

Note that the choice of angles used in the fiducial vector does not change the SIC-POVM found, but does change the order of the effects.

Starting with one of these fiducial vectors, now simply denoted by φ, the 9 effects of the SIC-POVM $\{\mathsf{G}(i,j)\}_{i,j=1}^3$ were placed in a 3×3 array such that the point (i,j) of the array corresponded to the element

$$\mathsf{G}(i-1, j-1) = \frac{1}{3}W_{i-1,j-1}P_\varphi W_{i-1,j-1}^*, \tag{3.128}$$

where $P_\varphi = |\varphi\rangle\langle\varphi|$ is the projection onto φ. The Cartesian partitions were then constructed, as in Figure 3.5, and were shown for every fiducial vector to produce commutative mutually unbiased POVMs. Following this, the 3-partitions that correspond to the mutually orthogonal Latin squares

$$\begin{matrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{matrix} \quad \text{and} \quad \begin{matrix} 1 & 2 & 3 \\ 3 & 1 & 2, \\ 2 & 3 & 1 \end{matrix} \tag{3.129}$$

were created, where the label at each point corresponds to the bin that SIC-POVM effect belongs to. We shall refer to these partitions as *Latin partitions.* These Latin squares are mutually orthogonal and, for each fiducial vector, the margin POVMs corresponding to the Latin partitions were found to be commutative. It should be noted that this is the maximum number of mutually orthogonal Latin squares of order 3. Further to this, despite the squares being equivalent under a permutation of the rows, doing so in this context would also correspond to a permutation of the effects of the SIC-POVM located at each point in the array, thereby still forming different partitions. For almost every fiducial vector—and hence every SIC-POVM—considered, 6 additional 3-partitions were found that produced commutative marginal observables. These 3-partitions split into two sets of 3 partitions satisfying the one-overlap property, i.e., for each SIC-POVM there existed 2 additional sets of 3 mutually unbiased commutative margin POVMs. These partitions are non-Latin, but they all satisfy the one-overlap property with one of the Latin partitions.

For every complete set of mutually unbiased commutative POVMs—derived from Cartesian and Latin partitions, or otherwise—the eigenvalues of each effect were recorded. For almost every complete set, the effects of 3 of the POVMs possessed the same spectrum, whilst the effects of the remaining POVM would have spectra $(1/2, 1/2, 0)$. The spectrum that the effects of the 3 POVMs share would depend on the value of r_0 used in the fiducial vector, but it is not obvious what function would describe this dependency. The POVM whose effects have spectrum $(1/2, 1/2, 0)$ corresponds to the Latin partition that completes the 2 sets of mutually unbiased POVMs made from non-Latin partitions.

The only fiducial vector that did not produce commutative margin POVMs in the same way as described above was φ_5, i.e., when $r_0 = 1/\sqrt{2}$. In this case, every possible effect created from a 3-partition belonged to a commutative margin POVM, and so 9 complete sets of commutative margin POVMs could be constructed for this SIC-POVM. Every margin POVM arising from either a Cartesian or Latin partition appears in 3 different complete sets, including the set composed solely from Cartesian and Latin partitions, whilst every other margin POVM appears in just one complete set. It was not possible to replicate this result for any other SIC-POVMs that were considered. Further to this, the spectrum $(1/2, 1/2, 0)$ was found for every effect from both the Cartesian and Latin partitions.

For each SIC-POVM, every possible margin effect—that is, every possible bin of a 3-partition of the SIC-POVM—was calculated, of which there are $\binom{9}{3} = 84$, and their spectra were calculated. For the fiducial vectors with $r_0 \neq 1/\sqrt{2}$ there existed several different spectra, but 54 effects possessed the spectrum $(0.646564, 0.275451, 0.0779852)$. These effects did not belong to a commutative margin POVM, whilst any other spectrum would lead to the discovery of a commutative margin POVM. For the φ_5 case, as mentioned above, the effects of the Cartesian and Latin partitions have spectra $(1/2, 1/2, 0)$, whilst all other possible effects possess the spectrum $(0.646564, 0.275451, 0.0779852)$. Given that every effect can be associated with a commutative margin POVM for this fiducial vector, this shows a further difference from the margin POVMs associated with the other fiducial vectors. This distribution of spectra also means that for any complete set of mutually unbiased margin POVMs containing POVMs derived from non-Latin partitions, there will be two different spectra appearing, and this means that these sets are inequivalent to the complete set formed from the Cartesian and Latin partitions. The decomposition of the spectra for the considered SIC-POVMs is given in Table 3.1.

From this work we may draw one obvious conclusion: with the exception of some special fiducial vectors like φ_5—and, indeed, φ_5 may be the only exception—whilst we may construct a complete set of mutually unbiased commutative margin POVMs from a SIC-POVM for $d = 3$, the majority of margin POVMs constructed from non-Cartesian and

r_0	Cartesian & Latin partitions	Non-Latin partitions	Non-commuting margin POVM effects
$\sqrt{\frac{2}{3}}$	$\begin{pmatrix}0.666667\\0.166667\\0.166667\end{pmatrix}$ (9), $\begin{pmatrix}0.5\\0.5\\0\end{pmatrix}$ (3)	$\begin{pmatrix}0.588681\\0.391216\\0.0201025\end{pmatrix}$ (18)	$\begin{pmatrix}0.646564\\0.275451\\0.00779852\end{pmatrix}$ (54)
$\frac{4}{5}$	$\begin{pmatrix}0.64\\0.293137\\0.0668629\end{pmatrix}$ (9), $\begin{pmatrix}0.5\\0.5\\0\end{pmatrix}$ (3)	$\begin{pmatrix}0.652227\\0.257931\\0.0898423\end{pmatrix}$ (9), $\begin{pmatrix}0.51524\\0.48428\\0.000479605\end{pmatrix}$ (9)	$\begin{pmatrix}0.646564\\0.275451\\0.00779852\end{pmatrix}$ (54)
$\frac{1}{2}\left(\frac{1}{\sqrt{2}}+\sqrt{\frac{2}{3}}\right)$	$\begin{pmatrix}0.580342\\0.403668\\0.0159904\end{pmatrix}$ (9), $\begin{pmatrix}0.5\\0.5\\0\end{pmatrix}$ (3)	$\begin{pmatrix}0.666425\\0.177774\\0.155801\end{pmatrix}$ (9), $\begin{pmatrix}0.596651\\0.37868\\0.0246684\end{pmatrix}$ (9)	$\begin{pmatrix}0.646564\\0.275451\\0.00779852\end{pmatrix}$ (54)
$\frac{3}{4}$	$\begin{pmatrix}0.5625\\0.428381\\0.00911863\end{pmatrix}$ (9), $\begin{pmatrix}0.5\\0.5\\0\end{pmatrix}$ (3)	$\begin{pmatrix}0.664479\\0.200777\\0.134744\end{pmatrix}$ (9), $\begin{pmatrix}0.611512\\0.353291\\0.0351972\end{pmatrix}$ (9)	$\begin{pmatrix}0.646564\\0.275451\\0.00779852\end{pmatrix}$ (54)
$\frac{1}{\sqrt{2}}$	$\begin{pmatrix}0.5\\0.5\\0\end{pmatrix}$ (12)	$\begin{pmatrix}0.646564\\0.275451\\0.00779852\end{pmatrix}$ (72)	(0)

Table 3.1: Spectra of all possible margin POVM effects for the 5 considered fiducial vectors, determined by r_0, sorted according to which partitions they belong to in order to form commutative margin POVMs. The number of elements with a given spectrum is denoted in brackets next to it.

non-Latin partitions will not be commutative. Further to this, in general the effects in a complete set of mutually unbiased commutative margin POVMs will possess two different spectra for $d = 3$.

A SIC-POVM was then constructed from a complete set of MUBs. The MUBs chosen, with $v^{(0)}$ denoting the computational basis and $v^{(i)}$ with $i = 1, 2, 3$ the remaining 3 bases, may be expressed as the following matrices, where each basis element corresponds to a column of the matrix:

$$\begin{aligned}
v^{(0)} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, & v^{(1)} &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{pmatrix}, \\
v^{(2)} &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ \omega & \omega^2 & 1 \\ \omega & 1 & \omega^2 \end{pmatrix}, & v^{(3)} &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ \omega^2 & \omega & 1 \\ \omega^2 & 1 & \omega \end{pmatrix},
\end{aligned} \tag{3.130}$$

with $\omega = \exp(2\pi i/3)$.

These MUBs, as sharp observables, were then smeared, as in Equation (3.68), in order to create mutually unbiased POVMs. The spectra by which these MUBs were smeared correspond to the spectra found when taking the Cartesian and Latin partitions of the SIC-POVMs just considered. These mutually unbiased POVMs were then placed in a 4×3 array, and the paths p_i described in Section 3.3.3 were constructed. From these paths the operators E_i given by Equation (3.109) were calculated, and from that the operators $\mathsf{G}(i)$ via Equation (3.112). There exist a total of $3^4 = 81$ paths through this array that are strictly downward in the sense of Section 3.3.3, and the eigenvalues of each of the corresponding $\mathsf{G}(i)$ were calculated. In the case that the eigenvalues for a given operator were all non-negative, then it was considered a potential SIC-POVM effect. It should noted that for $d = 3$, up to unitary equivalence, there exists only one complete set of MUBs, and so the only things that are changed in these investigations are the spectra used to smear the MUBs, and the number of spectra used.

Led by the investigations starting with SIC-POVMs, up to two spectra were used to smear the MUBs at any time. If two were used, then three of the MUBs would be smeared by one spectrum, whilst the final MUB would be smeared with the spectrum $(1/2, 1/2, 0)$. In both cases, of the 81 operators associated with downward paths, only 9 would be positive operators, and their associated paths would only overlap once. As a result, by Theorem 3.3.4, these 9 operators form a SIC-POVM for $d = 3$. The paths producing SIC-POVM effects vary depending on the spectra used, and so it is not sufficient to find a SIC-POVM with one collection of spectra, change the spectra, and then assume that the new operators will form a SIC-POVM.

This result highlights that although a SIC-POVM can be constructed from a complete set of smeared MUBs in the way described above, the number of possible operators that we may construct that are positive is far outweighed by the operators that contain negative eigenvalues. Indeed, for a given spectrum, or pair of spectra, we are able to determine a unique SIC-POVM.

3.5.2 $d = 5$

We may reproduce the method used for $d = 3$ for $d = 5$. By beginning with the fiducial vector [48]

$$\varphi=\begin{pmatrix} 0.39104489402214774638257588694092612854 \\ -0.28486558319586666154004262263615905414-0.64712933282796239400249892482493465752i \\ -0.23188384736899577826443055554623498413-0.19820390755555243362222174127453043647i \\ 0.13193857997561206409072011157833671545-0.10939599651964327439560846264470517498i \\ 0.43096743921096438598841830212227390108+0.19747405581954685663309013848038222203i \end{pmatrix}, \tag{3.131}$$

we produce a SIC-POVM, after which we place the 25 effects in a 5×5 array. The Cartesian margins were calculated and were indeed commutative. Given that there exist $\binom{25}{5} = 53130$ possible bins corresponding to effects of margin POVMs, an extensive search of all possible 5-partitions, and the spectra of the respective effects was not conducted, and so we restricted to Latin partitions. The 5-partitions constructed correspond to the following Latin squares:

$$\begin{matrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \\ 3 & 4 & 5 & 1 & 2 \\ 4 & 5 & 1 & 2 & 3 \\ 5 & 1 & 2 & 3 & 4 \end{matrix}, \quad \begin{matrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 1 & 2 \\ 5 & 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 & 1 \\ 4 & 5 & 1 & 2 & 3 \end{matrix}, \quad \begin{matrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 1 & 2 & 3 \\ 2 & 3 & 4 & 5 & 1 \\ 5 & 1 & 2 & 3 & 4 \\ 3 & 4 & 5 & 1 & 2 \end{matrix} \quad \text{and} \quad \begin{matrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 2 & 3 & 4 \\ 4 & 5 & 1 & 2 & 3 \\ 3 & 4 & 5 & 1 & 2 \\ 2 & 3 & 4 & 5 & 1 \end{matrix}. \tag{3.132}$$

This is not an exhaustive list of Latin squares of order 5, and other Latin partitions will exist, although they cannot be mutually orthogonal to all 4 of these Latin squares. These Latin partitions and the Cartesian partitions therefore form a complete set of six 5-partitions satisfying the one-overlap property. These Latin partitions produce margin POVMs that are all commutative, and the effects of these POVMs possess one of two spectra, which appear in one of two possible orders:

1. (a) $(0.152916, 0.499925, 0.0930549, 0.0293753, 0.244729)$;

   (b) $(0.152916, 0.244729, 0.0293753, 0.0930549, 0.499925)$;

2. (a) $(0.0584088, 0235772, 0.492705, 0.0399745, 0.17314)$;

   (b) $(0.0584088, 0.0399745, 0.235772, 0.17314, 0.492705)$.

Spectra 1. (a) and 2. (a) were both found in two margin POVMs each, whilst 1. (b) and 2. (b) were found in one margin POVM each. By calculating the eigenvectors of these margin POVMs, we then confirm that they form a complete set of MUBs.

Starting now with a complete set of 6 MUBs, given by the matrices

$$
\begin{aligned}
v^{(0)} &= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, &
v^{(1)} &= \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ \omega & \omega^2 & \omega^3 & \omega^4 & 1 \\ \omega^4 & \omega & \omega^3 & 1 & \omega^2 \\ \omega^4 & \omega^2 & 1 & \omega^3 & \omega \\ \omega & 1 & \omega^4 & \omega^3 & \omega^2 \end{pmatrix}, \\
v^{(2)} &= \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ \omega^2 & \omega^3 & \omega^4 & 1 & \omega \\ \omega^3 & 1 & \omega^2 & \omega^4 & \omega \\ \omega^3 & \omega & \omega^4 & \omega^2 & 1 \\ \omega^2 & \omega & 1 & \omega^4 & \omega^3 \end{pmatrix}, &
v^{(3)} &= \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ \omega^3 & \omega^4 & 1 & \omega & \omega^2 \\ \omega^2 & \omega^4 & \omega & \omega^3 & 1 \\ \omega^2 & 1 & \omega^3 & \omega & \omega^4 \\ \omega^3 & \omega^2 & \omega & 1 & \omega^4 \end{pmatrix}, \\
v^{(4)} &= \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ \omega^4 & 1 & \omega & \omega^2 & \omega^3 \\ \omega & \omega^3 & 1 & \omega^2 & \omega^4 \\ \omega & \omega^4 & \omega^2 & 1 & \omega^3 \\ \omega^4 & \omega^3 & \omega^2 & \omega & 1 \end{pmatrix}, &
v^{(5)} &= \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & \omega & \omega^2 & \omega^3 & \omega^4 \\ 1 & \omega^2 & \omega^4 & \omega & \omega^3 \\ 1 & \omega^3 & \omega & \omega^4 & \omega^2 \\ 1 & \omega^4 & \omega^3 & \omega^2 & \omega^1 \end{pmatrix},
\end{aligned}
\tag{3.133}
$$

and by smearing the associated sharp observables using the spectra given above, taking note to use each one the correct number of times, we create 6 mutually unbiased commutative POVMs. The effects of these POVMs were then placed in a 6 × 5 array, and we calculated all possible operators corresponding to downward paths through the array. The spectra of these operators were calculated, and only those that were positive were printed. Of the $5^6 = 15625$ possible operators, only 25 were positive, and their associated paths overlap only once. In other words, the only positive operators form a SIC-POVM.

In the case of $d = 5$, the point that must be highlighted is that the ordering of the spectra matters, which was not the case for $d = 3$. This is a point that must be taken into account in higher dimensions.

3.5.3 Issues in dimensions $d = 4, 7$

In dimensions 4 and 7, issues arose within the investigation. In dimension 4, we worked with the fiducial vector described as follows [44]: We define the four constants[2]

$$
r_0 = \frac{\sqrt{1 - 1/\sqrt{5}}}{2\sqrt{2 - \sqrt{2}}}, \quad r_1 = (\sqrt{2} - 1)r_0, \quad r_{\pm} = \frac{1}{2}\sqrt{1 + 1/\sqrt{5} \pm \sqrt{1/5 + 1/\sqrt{5}}}, \tag{3.134}
$$

and the angles

$$
a = \arccos \frac{2}{\sqrt{5 + \sqrt{5}}}, \qquad b = \arcsin \frac{2}{\sqrt{5}}. \tag{3.135}
$$

[2] The value for r_0 given here is different to that given in [44] (the overall square root in the numerator is missing in their version), and this is to ensure that the resultant vector is normalised and, indeed, the fiducial vector for a SIC-POVM.

With these values we define the set

$$\begin{aligned}\Omega = \{ \, &((-1)^m(a/2+b/4)+\pi(m+2n+7j+1)/4, \pi(2k+1)/2 \, , \\ &(-1)^m(-a/2+b/4)+\pi(m+2n+3j+4k+1)/4) \mid j,k,m=0,1, \; n \in \mathbb{Z}_4\},\end{aligned} \tag{3.136}$$

and from this the fiducial vectors for our SIC-POVM, in the computational basis, are given by

$$\left\{ \begin{pmatrix} r_0 \\ r_+e^{i\theta_+} \\ r_1e^{i\theta_1} \\ r_-e^{i\theta_-} \end{pmatrix}, \begin{pmatrix} r_0 \\ r_-e^{i\theta_-} \\ r_1e^{i\theta_1} \\ r_+e^{i\theta_+} \end{pmatrix}, \text{ plus all permutations} \;\middle|\; (\theta_+,\theta_1,\theta_-) \in \Omega \right\}. \tag{3.137}$$

Using the simplest case; i.e, $j=k=m=n=0$, the SIC-POVM was constructed in Mathematica and the Cartesian partitions were made. The corresponding margin POVMs were found to be commutative, and so we subsequently focussed on Latin partitions. There exist 24 inequivalent Latin squares of order 4, and the POVM corresponding to each was tested. Unfortunately, of these 24 POVMs, only 2 were commutative, and the corresponding Latin squares are not mutually orthogonal, so this method is not sufficient in this case.

An alternative method to consider, as was done for $d=3$, is to choose the first bin of a partition and then exhaust all possible complementary bins with corresponding operators that commute with the operator corresponding to the first bin. However, assuming that we fix this first bin to contain the first element of the SIC-POVM, there exist $\binom{15}{3}=455$ possible initial bins that can be considered, and subsequently $\binom{12}{4}=495$ possible second bins, so there are upwards of 225225 possible pairs of bins to check for commutativity. This is an unfeasible number to check, and so whilst a complete set of commutative margin POVMs may exist for $d=4$, we do not know of any, and subsequently any spectra which would allow for the construction of a SIC-POVM from a complete set of MUBs.

For dimension 7, the two fiducial vectors given in [48] were considered, namely

$$\begin{aligned}
\varphi_1 &= \begin{pmatrix} 0.34160748940869677901847702221939960794 \\ 0.55148563849933599729808437186250132025-0.34867715271314252440447697902337375141i \\ -0.11648580220876396683451558464316567078-0.26166619656998671045928442919728701366i \\ -0.07531290658117406012144376218066882451\!5-0.27525817515698021623612242207232334852i \\ -0.07325737634520240670332850107094522629\!0-0.10603759291025605066417926940588166460i \\ -0.34887242536080270339160710148674089447+0.23627220962007852786364601038178637997i \\ -0.27916461741208963926566644470038031212-0.14844155524225577344987795977973540454i \end{pmatrix}, \\
\varphi_2 &= \begin{pmatrix} 0.41526356195047080423032850597006010210 \\ -0.07788677567440470750681177077239629809\!6-0.38900572497454333820768884319198370735i \\ -0.58205250739832009441182422104971379149-0.26698341367949812425899377170619235952i \\ 0.04282013260966922441846438570089085625\!2-0.05705702761089401226075608738988398123\!8i \\ 0.13748747041004807562464450190666624759-0.11091890130114900823149188171139550779i \\ -0.22922883526414535564906529306017970666-0.29118397371245900866263335536522313692i \\ 0.29359695336668205329426389130467259030+0.01646492781073352558176274412400031430\!4i \end{pmatrix}.
\end{aligned} \tag{3.138}$$

In both instances a SIC-POVM was constructed, as were the Cartesian margins, which were found to be commutative. The 6 remaining margins correspond to the Latin squares

$$\begin{array}{ccccccc}
1&2&3&4&5&6&7\\
2&3&4&5&6&7&1\\
3&4&5&6&7&1&2\\
4&5&6&7&1&2&3\\
5&6&7&1&2&3&4\\
6&7&1&2&3&4&5\\
7&1&2&3&4&5&6
\end{array},\quad
\begin{array}{ccccccc}
1&2&3&4&5&6&7\\
3&4&5&6&7&1&2\\
5&6&7&1&2&3&4\\
7&1&2&3&4&5&6\\
2&3&4&5&6&7&1\\
4&5&6&7&1&2&3\\
6&7&1&2&3&4&5
\end{array},\quad
\begin{array}{ccccccc}
1&2&3&4&5&6&7\\
4&5&6&7&1&2&3\\
7&1&2&3&4&5&6\\
3&4&5&6&7&1&2\\
6&7&1&2&3&4&5\\
2&3&4&5&6&7&1\\
5&6&7&1&2&3&4
\end{array},$$
$$\begin{array}{ccccccc}
1&2&3&4&5&6&7\\
5&6&7&1&2&3&4\\
2&3&4&5&6&7&1\\
6&7&1&2&3&4&5\\
3&4&5&6&7&1&2\\
7&1&2&3&4&5&6\\
4&5&6&7&1&2&3
\end{array},\quad
\begin{array}{ccccccc}
1&2&3&4&5&6&7\\
6&7&1&2&3&4&5\\
4&5&6&7&1&2&3\\
2&3&4&5&6&7&1\\
7&1&2&3&4&5&6\\
5&6&7&1&2&3&4\\
3&4&5&6&7&1&2
\end{array},\quad
\begin{array}{ccccccc}
1&2&3&4&5&6&7\\
7&1&2&3&4&5&6\\
6&7&1&2&3&4&5\\
5&6&7&1&2&3&4\\
4&5&6&7&1&2&3\\
3&4&5&6&7&1&2\\
2&3&4&5&6&7&1
\end{array}. \tag{3.139}$$

As in the case of $d = 5$, this does not exhaust the possible inequivalent Latin squares of order 7, and other complete sets of partitions satisfying the one-overlap property can be found. Again, similar to $d = 5$, the spectra of every possible effect were not calculated, as there exist $\binom{49}{7} = 85900584$ such effects for each SIC-POVM to consider. However, the partitions that were considered did indeed lead to commutative margin POVMs, so we stopped our search with them.

For the constructed margin POVMs, the effects possessed the following spectra:

1. (a) $(0.425712, 0.177537, 0.116696, 0.0999678, 0.0820381, 0.0814391, 0.0166106)$;

(b) $(0.382799, 0.217579, 0.210489, 0.0908925, 0.0419947, 0.0384167, 0.0178294)$;

(c) $(0.0540971, 0.285421, 0.285421, 0.0298802, 0.285421, 0.0298802, 0.0298802)$;

(d) $(0.419906, 0.150834, 0.150834, 0.0425309, 0.150834, 0.0425309, 0.0425309)$;

2. (a) $(0.410065, 0.172444, 0.157392, 0.137334, 0.0864703, 0.0312058, 0.00508907)$;

   (b) $(0.0800943, 0.0225843, 0.0225843, 0.284051, 0.0225843, 0.284051, 0.284051)$;

   (c) $(0.445903, 0.0923495, 0.0923495, 0.0923495, 0.0923495, 0.0923495, 0.0923495)$.

Spectra 1. (a) and 1. (b) are both present in the effects of 3 of the margin POVMs, whilst spectra 1. (c) and 1. (d), which contain degenerate eigenvalues, are present in just one POVM each. This is the case for spectra 2. (b) and 2. (c) as well, with spectrum 2. (a) appearing in the effects of 6 margin POVMs.

The presence of degenerate eigenvalues presented problems for Mathematica, as it was unable to directly provide a shared eigenbasis for a POVM whose effects had such spectra, despite these effects commuting and hence possessing one. This was resolved by collecting eigenvectors that correspond to the one non-degenerate eigenvalue for each effect, and in doing so an eigenbasis was found for each such POVM that was shown to be mutually unbiased to the other eigenbases calculated by Mathematica.

The difficulty in resolving this issue highlights problems that occur when degenerate eigenvalues appear in the spectra of the effects of the margin POVMs. Further to that, the significance of the ordering of the spectra used for smearing was affirmed when attempts to construct a SIC-POVM from a complete set of MUBs proved fruitless. It should be possible to overcome this issue, but more work on the subject would be needed.

Chapter 4

Generalising the Arthurs-Kelly Measurement Model with Correlated Probes and Focussing

The classic example of a pair of incompatible quantum observables is the position E^Q and momentum E^P of a particle, expressed most succinctly by the preparation uncertainty relation:

$$\mathrm{Var}\,(Q,\psi)\,\mathrm{Var}\,(P,\psi) \geq \frac{1}{4}, \tag{4.1}$$

where Var () denotes variance, $\hbar$ has been set to 1, $Q = \mathsf{E}^Q[1]$ is the position operator and $P = \mathsf{E}^P[1]$ the momentum. This relation sets a limit on how precisely a state ψ can be prepared with regards to the statistics of one of these observables, and what payoff the statistics of the second observable faces.

An alternative means of highlighting the incompatibility of these two observables lies in the fact that one cannot directly measure both on the same system without greatly disturbing the statistics of one. Indeed, performing an accurate measurement of the system's position will lead to a highly localised state with a momentum that, if subsequently measured, is statistically widely spread out.

A theoretical method, using additional "probe" systems, was given by Arthurs and Kelly [3] to allow for an indirect measurement of these two observables at the expense of statistical noise required by Equation (4.1). An extension to this model was proposed by Di Lorenzo [27], in which initial correlations are allowed to exist between the probes. This extension, in association with his given definition of error and disturbance, was claimed to allow for a violation of an error-disturbance relation of a form similar to Equation (4.1).

In this chapter we shall analyse the effect these correlations have on the measured observable, and refute the claim that any such violation of a physically relevant error-disturbance relation occurs. We will further show how these correlations lead to the phenomenon of *focussing*, which can allow for more precise measurements when using less than ideal equipment. Note that whilst we relied on finite dimensional Hilbert spaces in Chapter 3, we will assume the Hilbert spaces in what follows are infinite dimensional. In particular, each system we will consider will be described by the Hilbert space $L^2(\mathbb{R})$, the space of real-valued square-integrable complex functions. In other words, we will be considering systems with one continuous degree of freedom.

4.1 The Arthurs-Kelly measurement model

We take the concept of measurement models introduced in Section 2.4.3 and extend it to include two probes, one performing a (sharp) position measurement and the second a momentum measurement. The system we wish to measure, described by the Hilbert space $\mathcal{H}$ with state $\rho \in \mathcal{S}(\mathcal{H})$, is coupled to two probes with associated Hilbert spaces $\mathcal{K}_1$, $\mathcal{K}_2$ and states $\sigma_1 \in \mathcal{S}(\mathcal{K}_1)$, $\sigma_2 \in \mathcal{S}(\mathcal{K}_2)$. The coupling is given by the interaction Hamiltonian

$$H_{int} = \lambda Q P_1 - \mu P Q_2 + \frac{\lambda\mu}{2}\kappa P_1 Q_2, \tag{4.2}$$

where the subscript on the position and momentum operators denote which probe they pertain to (and if there is no subscript then they pertain to the measured system)[1]. The positive constants λ, μ and κ are coupling parameters between the system and the first probe, the system and the second probe, and between the two probes, respectively. These constants are assumed to be large enough that the free evolution of the system and two probes may be ignored in what follows. We assume an impulsive interaction; that is, a short-time interaction described by the unitary $U_{int} = \exp(-iH_{int})$ (we may also perform the interaction for a time t and then have the value absorbed into the coupling parameters). By making use of the Baker-Campbell-Hausdorff formula

$$\exp[A + B] = \exp[A]\exp[B]\exp\left[-\frac{1}{2}[A, B]\right], \tag{4.3}$$

where A and B satisfy $[A, [A, B]] = [B, [A, B]] = 0$, thereby removing any further terms, we may decompose U_{int} into three terms:

$$\begin{aligned} U_{int} &= U_\mu U_\lambda U_\kappa \\ &= \exp(i\mu P Q_2)\exp(-i\lambda Q P_1)\exp\left[-i\frac{\lambda\mu}{2}(\kappa - 1)P_1 Q_2\right]. \end{aligned} \tag{4.4}$$

Equation (4.4) highlights that this coupling can be perceived as a sequence of interactions between pairs of systems. If we wish, we can also rearrange these interactions, which leads to the same unitary, albeit with a slightly different decomposition:

$$\begin{aligned} U_{int} &= U_\lambda U_\mu U'_\kappa \\ &= \exp(-i\lambda Q P_1)\exp(i\mu P Q_2)\exp\left[-i\frac{\lambda\mu}{2}(\kappa + 1)P_1 Q_2\right]. \end{aligned} \tag{4.5}$$

The difference is present in the form of the interaction between the two probes, but results in the same post-coupling state.

After this coupling we perform ideal measurements on the probes. We measure the ideal position observable E^{Q_1} on the first probe, and measure the ideal momentum observable E^{P_2} on the second. After the system has been coupled to a probe that probe

[1] In [3], Arthurs and Kelly set κ to zero, and use P_2 as the shift operator on the second probe.

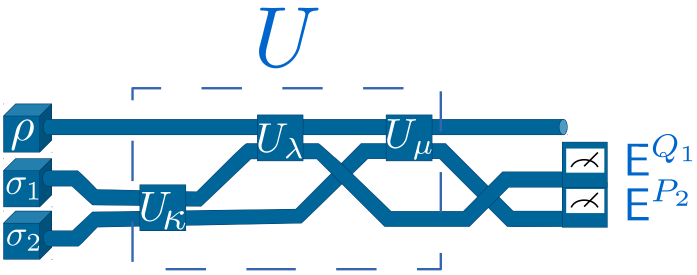


Figure 4.1: The standard form of the Arthurs-Kelly measurement model. Two probes in states σ_1 and σ_2 are coupled to the measured system in state ρ via the unitary U. After the coupling, an ideal measurement of the position of the first probe is performed, and similarly an ideal measurement of momentum for the second probe. From these measurements we infer information about the position and momentum of the system we are interested in.

is then measured, so the decomposition given in Equation (4.4) can be perceived as a sequential measurement of the first probe's position followed by the second probe's momentum, whilst the order is reversed for the decomposition given in Equation (4.5). The decomposition given by Equation (4.4) is shown in Figure 4.1. In the case that $|\kappa| = 1$, the coupling leads to a strictly sequential measurement (where the order of the measurements varies depending on the sign of κ).

We shall begin by allowing the system and probes to be prepared in pure states; in other words,

$$\rho = P_\psi, \quad \sigma_1 = P_{\varphi_1}, \quad \sigma_2 = P_{\varphi_2}, \tag{4.6}$$

where $\psi \in \mathcal{H}$, $\varphi_1 \in \mathcal{H}_1$ and $\varphi_2 \in \mathcal{H}_2$. Preparing the probes in pure states is done for ease of calculations, and can be readily extended to mixed states, as will be shown, due to the convex nature of the state spaces. In order to determine the state of the system and two probes after the coupling, we make use of the position representation $\psi(q) = \langle q|\psi\rangle$, where $|q\rangle \in \mathcal{H}$ is a position pseudo-eigenvector. We also use the identity

$$\begin{aligned}(e^{-i\lambda x P}\psi)(q) &= \left\langle q \middle| e^{-i\lambda x P}\psi \right\rangle \\ &= \left\langle e^{i\lambda x P} q \middle| \psi \right\rangle \\ &= \int_{\mathbb{R}} dp\, e^{-i\lambda x p} \langle q|p\rangle \langle p|\psi\rangle \\ &= \int_{\mathbb{R}} dp\, \frac{1}{\sqrt{2\pi}} e^{ip(q-\lambda x)} \widetilde{\psi}(p) \\ &= \psi(q - \lambda x),\end{aligned} \tag{4.7}$$

where we have used the spectral decomposition of the momentum operator

$$P = \int_{\mathbb{R}} dp\, p\, |p\rangle \langle p|\,, \tag{4.8}$$

with the pseudo eigenvectors satisfying

$$\langle q|p\rangle = \frac{1}{\sqrt{2\pi}} e^{iqp}, \tag{4.9}$$

and the momentum representation of ψ, $\widetilde{\psi}(p) = \langle p|\psi\rangle$. Using Equation (4.7), the post-coupling state of the system and two probes is given by the state Ψ, which in the position representation (with q_1 the position coordinate for the first probe and likewise for q_2) is equal to

$$\begin{aligned}
\Psi(q,q_1,q_2) &= \langle q,q_1,q_2|U_{int}(\psi\otimes\varphi_1\otimes\varphi_2)\rangle \\
&= \left\langle e^{i\frac{\lambda\mu}{2}(\kappa-1)P_1Q_2}e^{i\lambda QP_1}e^{-i\mu Pq_2}(q,q_1,q_2)\middle|\psi\otimes\varphi_1\otimes\varphi_2\right\rangle \\
&= \left\langle e^{i\frac{\lambda\mu}{2}(\kappa-1)P_1Q_2}e^{i\lambda(q+\mu\,q_2)P_1}(q+\mu\,q_2,q_1,q_2)\middle|\psi\otimes\varphi_1\otimes\varphi_2\right\rangle \\
&= \left\langle e^{i\frac{\lambda\mu}{2}(\kappa-1)P_1q_2}(q+\mu\,q_2,q_1-\lambda(q+\mu\,q_2),q_2)\middle|\psi\otimes\varphi_1\otimes\varphi_2\right\rangle \\
&= \left\langle q+\mu\,q_2,q_1-\lambda\left(q+\mu\,q_2+\frac{\mu}{2}(\kappa-1)q_2\right),q_2\middle|\psi\otimes\varphi_1\otimes\varphi_2\right\rangle \\
&= \psi(q+\mu\,q_2)\varphi_1\left(q_1-\lambda\left(q+\frac{\mu}{2}(\kappa+1)q_2\right)\right)\varphi_2(q_2).
\end{aligned} \tag{4.10}$$

Upon the first probe we shall perform an ideal position measurement, and on the second we perform an ideal momentum measurement. To accommodate this, we will perform a Fourier transformation on the final argument and thereby acquire the momentum representation for the second probe:

$$\Psi(q,q_1,q_2)\mapsto\widetilde{\Psi}(q,q_1,w_2) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} dq_2\, e^{-iq_2w_2}\Psi(q,q_1,q_2), \tag{4.11}$$

where in what follows we will use w_i to denote the momentum variable on the i^{th} probe. From this, the effective observable G measured on the system is given by

$$\text{tr}\left[P_\psi\mathsf{G}^{(\lambda,\mu)}(X\times Y)\right] = \text{tr}\left[P_\Psi\big(I\otimes\mathsf{E}^{Q_1}(\lambda X)\otimes\mathsf{E}^{P_2}(\mu Y)\big)\right], \tag{4.12}$$

where $X,Y\in\mathcal{B}(\mathbb{R})$, $\lambda X = \{\lambda x|x\in X\}$, etc., and we have used the pointer functions $g: x\mapsto\lambda^{-1}x$ and $h: y\mapsto\mu^{-1}y$ on the first and second probe, respectively, as given in Equation (2.141).

4.2 Extension of the Arthurs-Kelly model and derived observables

4.2.1 The effective observable derived on the system

We will now allow for the probes to be prepared in an arbitrary state, simply denoted by $\varphi\in\mathcal{H}_1\otimes\mathcal{H}_2$. This is shown in Figure 4.2, where the unitary channel V provides initial correlations between the probes prior to interacting with the system.

The post-coupling state, in the position representation, is now of the form

$$\Psi(q,q_1,q_2) = \langle q,q_1,q_2|U_{int}(\psi\otimes\varphi)\rangle = \psi(q+\mu\,q_2)\varphi\left(q_1-\lambda\left(q+\frac{\mu}{2}(\kappa+1)q_2\right),q_2\right), \tag{4.13}$$

upon which we again perform a Fourier transform on the final argument to produce the state $\widetilde{\Psi}(q,q_1,w_2)$, as given in Equation (4.11). The effective observable $\mathsf{G}^{(\lambda,\mu)}$ measured on the system is again given by Equation (4.12). As shown in Appendix A.1, $\mathsf{G}^{(\lambda,\mu)}$ takes

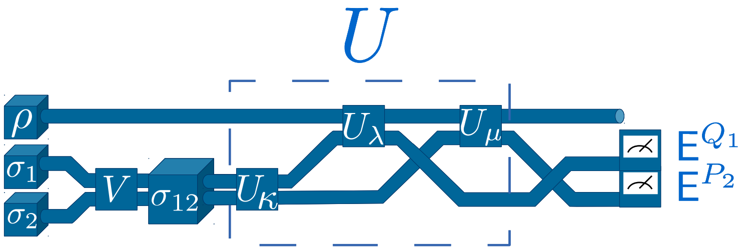


Figure 4.2: The extension of the Arthurs-Kelly model that we present here is one that allows for arbitrary probe preparations. This is expressed here by the unitary channel V that transforms $\sigma_1 \otimes \sigma_2$ into the state σ_{12}, which may be entangled or mixed.

the form

$$\mathsf{G}^{(\lambda,\mu)}(X \times Y) = \int_{X\times Y} dq\, dp\, \big(K_{qp}^{(\lambda,\mu)}\big)^* K_{qp}^{(\lambda,\mu)}, \tag{4.14}$$

where $K_{qp}^{(\lambda,\mu)}$ has the kernel

$$K_{qp}^{(\lambda,\mu)}(x,x') = \sqrt{\frac{\lambda}{2\pi\mu}} e^{ip(x-x')} \varphi\left(\lambda\big(q - \frac{1}{2}((1-\kappa)x + (1+\kappa)x')\big), \frac{1}{\mu}(x'-x)\right). \tag{4.15}$$

We can rewrite $K_{qp}^{(\lambda,\mu)}$ in the following way:

$$\begin{aligned} K_{qp}^{(\lambda,\mu)} &= \sqrt{\tfrac{\lambda}{2\pi\mu}} \int_{\mathbb{R}^2} dx\, dx'\, e^{ip(x-x')} \varphi\big(\lambda\big(q - \tfrac{1}{2}\big((1-\kappa)x + (1+\kappa)x'\big)\big), \tfrac{1}{\mu}(x'-x)\big) \, |x\rangle\langle x'| \\ &= \sqrt{\tfrac{\lambda}{2\pi\mu}} \int_{\mathbb{R}^2} dx\, dx'\, e^{ip(x-x')} \varphi\big(- \tfrac{1}{2}\lambda\big((1-\kappa)(x-q) + (1+\kappa)(x'-q)\big), \tfrac{1}{\mu}(x'-x)\big) \, |x\rangle\langle x'| \\ &= \sqrt{\tfrac{\lambda}{2\pi\mu}} \int_{\mathbb{R}^2} dx\, dx'\, e^{ip(x-x')} \varphi\big(- \tfrac{1}{2}\lambda\big((1-\kappa)x + (1+\kappa)x'\big), \tfrac{1}{\mu}(x'-x)\big) \, |x+q\rangle\langle x'+q| \\ &= e^{-iqP} \big(\sqrt{\tfrac{\lambda}{2\pi\mu}} \int_{\mathbb{R}^2} dx\, dx'\, e^{ip(x-x')} \varphi\big(- \tfrac{1}{2}\lambda\big((1-\kappa)x + (1+\kappa)x'\big), \tfrac{1}{\mu}(x'-x)\big) \, |x\rangle\langle x'| \,\big) e^{iqP} \\ &= e^{-iqP} e^{ipQ} \big(\sqrt{\tfrac{\lambda}{2\pi\mu}} \int_{\mathbb{R}^2} dx\, dx'\, \varphi\big(- \tfrac{1}{2}\lambda\big((1-\kappa)x + (1+\kappa)x'\big), \tfrac{1}{\mu}(x'-x)\big) \, |x\rangle\langle x'| \,\big) e^{-ipQ} e^{iqP} \\ &= W_{qp} K_{00}^{(\lambda,\mu)} W_{qp}^*, \end{aligned} \tag{4.16}$$

where the W_{qp} are the Weyl operators $W_{qp} = \exp[iqp/2]\exp[-iqP]\exp[ipQ]$ that generate shifts in phase space, and are the continuous analogue of the Weyl-Heisenberg operators given in Equation (3.33). By making use of Equation (4.16) we see that

$$\begin{aligned} W_{qp} \mathsf{G}^{(\lambda,\mu)}(X \times Y) W_{qp}^* &= W_{qp} \left(\int_{X\times Y} dq'\, dp' \big(K_{q'p'}^{(\lambda,\mu)}\big)^* K_{q'p'}^{(\lambda,\mu)} \right) W_{qp}^* \\ &= \int_{X\times Y} dq'\, dp' \left(W_{qp} K_{q'p'}^{(\lambda,\mu)} W_{qp}^* \right)^* W_{qp} K_{q'p'}^{(\lambda,\mu)} W_{qp}^* \\ &= \int_{(X+q)\times(Y+p)} dq'\, dp' \big(K_{q'p'}^{(\lambda,\mu)}\big)^* K_{q'p'}^{(\lambda,\mu)} \\ &= \mathsf{G}^{(\lambda,\mu)}((X+q) \times (Y+p)), \end{aligned} \tag{4.17}$$

where $X + q = \{x + q | x \in X\}$, etc. In other words, $\mathsf{G}^{(\lambda,\mu)}$ is a covariant phase space observable. Furthermore, we can extend this result by initially preparing our probes in a mixed state, i.e., in the state $\sigma = \sum_i p_i P_{\varphi_i}$, where $\varphi_i \in \mathcal{H}_1 \otimes \mathcal{H}_2$. Each of these pure states leads to a covariant phase space observable G_i, and so, from the linearity of the trace, the effective observable H is of the form

$$\mathsf{H}(X \times Y) = \sum_i p_i \mathsf{G}_i(X \times Y), \tag{4.18}$$

which satisfies

$$\begin{aligned} W_{qp}\mathsf{H}(X \times Y)W_{qp}^* &= \sum_i p_i W_{qp} \mathsf{G}_i(X \times Y) W_{qp}^* \\ &= \sum_i p_i \mathsf{G}_i((X+q) \times (Y+p)) \\ &= \mathsf{H}((X+q) \times (Y+p)). \end{aligned} \tag{4.19}$$

That is, H is also a covariant phase space observable. We are therefore able, having already proved it, to state the main result of this chapter:

Theorem 4.2.1. The observable $\mathsf{G}^{(\lambda,\mu)}$ measured in an Arthurs-Kelly-like measurement scheme, where the probes are prepared in an arbitrary state, is always a covariant phase space observable.

This theorem extends a result given in [9], where the probes were prepared in only pure, separable states, i.e., $\varphi = \varphi_1 \otimes \varphi_2$.

Covariant phase space observables have been studied in great depth elsewhere (see, for example, [31, 54, 23, 33]), and it is a well known result that for any $Z \subseteq \mathbb{R}^2$, a covariant phase space observable G may be expressed via

$$\mathsf{G}(Z) = \frac{1}{2\pi} \int_Z dq\, dp\, W_{qp}\, \tau\, W_{qp}^*, \tag{4.20}$$

where τ is a positive operator of unit trace. In other words, τ is mathematically a density operator, but does not correspond to a physical state in the system. Indeed, in what we have presented, $\tau = 2\pi (K_{00}^{(\lambda,\mu)})^* K_{00}^{(\lambda,\mu)}$, which is indeed positive and trace one. The positivity is immediate: for any state $\psi \in \mathcal{H}$,

$$\left\langle \psi \middle| (K_{00}^{(\lambda,\mu)})^* K_{00}^{(\lambda,\mu)} \psi \right\rangle = \left\| K_{00}^{(\lambda,\mu)} \psi \right\|^2 \geq 0, \tag{4.21}$$

by the positivity requirement of the norm defined on $\mathcal{H}$. With regards to the trace:

$$\begin{aligned} 2\pi \operatorname{tr}\left[(K_{00}^{(\lambda,\mu)})^* K_{00}^{(\lambda,\mu)} \right] &= 2\pi \int_{\mathbb{R}} dx\, \left((K_{00}^{(\lambda,\mu)})^* K_{00}^{(\lambda,\mu)} \right)(x,x) \\ &= \frac{\lambda}{\mu} \int_{\mathbb{R}^2} dx\, dx' \left| \varphi\left(-\frac{1}{2}\lambda((1-\kappa)x' + (1+\kappa)x), \frac{1}{\mu}(x - x') \right) \right|^2 \\ &= \frac{\lambda}{\mu} \int_{\mathbb{R}^2} dx\, dy' \left| \varphi\left(\lambda \left(\frac{1}{2}(1-\kappa)y' - x \right), \frac{1}{\mu} y' \right) \right|^2 \\ &= \frac{\lambda}{\mu} \int_{\mathbb{R}^2} dy\, dy' \left| \varphi\left(\lambda y, \frac{1}{\mu} y' \right) \right|^2 \\ &= \int_{\mathbb{R}^2} dy\, dy'\, |\varphi(y,y')|^2 \\ &= 1, \end{aligned} \tag{4.22}$$

where in the third equality we have made the substitution $y' = x - x'$ and $y = \frac{1}{2}((1-\kappa)y' - x$ in the fourth.

4.2.2 The marginal observables and joint measurement error relations

So far we have found that our effective observable is a covariant phase space observable, but our original goal was to use this measurement scheme to make (approximate) joint measurements of the system's position and momentum. We perform this by taking the margins of $\mathsf{G}^{(\lambda,\mu)}$ to create the jointly measurable observables $\mathsf{E}^{(\lambda,\mu)}$ and $\mathsf{F}^{(\lambda,\mu)}$:

$$\mathsf{E}^{(\lambda,\mu)}(X) = \mathsf{G}^{(\lambda,\mu)}(X \times \mathbb{R}), \qquad \mathsf{F}^{(\lambda,\mu)}(Y) = \mathsf{G}^{(\lambda,\mu)}(\mathbb{R} \times Y). \tag{4.23}$$

As is shown in Appendix A.2, these marginal observables can be expressed in the form

$$\mathsf{E}^{(\lambda,\mu)}(X) = (\chi_X * e^{(\lambda,\mu)})(Q), \tag{4.24a}$$

$$\mathsf{F}^{(\lambda,\mu)}(Y) = (\chi_Y * f^{(\lambda,\mu)})(P), \tag{4.24b}$$

where χ_A is the characteristic function for the set $A \subset \mathcal{B}(\mathbb{R})$, and $e^{(\lambda,\mu)}$ and $f^{(\lambda,\mu)}$ are probability distributions given by

$$e^{(\lambda,\mu)}(q) = \frac{\lambda}{\mu} \int_{\mathbb{R}} dq' \left| \varphi \left(\lambda \left(\frac{1}{2}(1-\kappa)q' - q \right), \frac{1}{\mu} q' \right) \right|^2, \tag{4.25a}$$

$$f^{(\lambda,\mu)}(p) = \frac{\lambda}{\mu} \int_{\mathbb{R}} dp' \left| \widetilde{\varphi} \left(\lambda p', \frac{1}{\mu} \left(\frac{p'}{2}(\kappa+1) - p \right) \right) \right|^2, \tag{4.25b}$$

where again $\widetilde{\varphi}$ corresponds to the momentum representation of φ. In other words, these marginal observables are smeared versions of the ideal position and momentum observables with the smearing being provided by the probability distributions $e^{(\lambda,\mu)}$ and $f^{(\lambda,\mu)}$, respectively. Indeed, these probability distributions correspond to the probability distributions for the ideal position and momentum observables, respectively, with regards to the state $\tau = 2\pi (K_{00}^{(\lambda,\mu)})^* K_{00}^{(\lambda,\mu)}$, as introduced in Equation (4.20):

$$e^{(\lambda,\mu)}(X) = p_\tau^Q(X) = \int_X \operatorname{tr}\left[d\mathsf{E}^Q(q)\tau \right], \tag{4.26a}$$

$$f^{(\lambda,\mu)}(Y) = p_\tau^P(Y) = \int_Y \operatorname{tr}\left[d\mathsf{E}^P(p)\tau \right]. \tag{4.26b}$$

Note that this is a general result for all covariant phase space observables, as can be quickly verified.

Since we are using the observables $\mathsf{E}^{(\lambda,\mu)}$ and $\mathsf{F}^{(\lambda,\mu)}$ as approximations of position and momentum, respectively, we would like some means of determining how accurate an approximation each observable is. To that end we shall calculate the value of the noise measure and BLW error, as described in Section 2.4.5, for these observables. These measures possess several important differences, the most notable of which being the state-dependence of the noise measure and the state-independence of the BLW error, and the question of which corresponds to a physically valid measure of error has lead to considerable heated debate. However, this debate is not a focus of this work, and instead we simply recall (see Appendix B) that in the case of dealing with the margins of covariant phase space observables they coincide; in particular, for the two marginal observables $\mathsf{E}^{(\lambda,\mu)}$ and $\mathsf{F}^{(\lambda,\mu)}$ their errors, according to these measures, are equal to the second moments of the probability distributions over which they are smeared. That is,

$$\begin{aligned}\varepsilon(\mathsf{E}^{(\lambda,\mu)},\mathsf{E}^{Q},\rho)^2 &= \Delta(\mathsf{E}^{(\lambda,\mu)},\mathsf{E}^{Q})^2 = e^{(\lambda,\mu)}[2] = \int_{\mathbb{R}} q^2 e^{(\lambda,\mu)}(q)dq,\\ \varepsilon(\mathsf{F}^{(\lambda,\mu)},\mathsf{E}^{P},\rho)^2 &= \Delta(\mathsf{F}^{(\lambda,\mu)},\mathsf{E}^{P})^2 = f^{(\lambda,\mu)}[2] = \int_{\mathbb{R}} p^2 f^{(\lambda,\mu)}(p)dp,\end{aligned} \tag{4.27}$$

for all $\rho \in \mathcal{S}(\mathcal{H})$. The second moment of this probability distribution is greater than or equal to its variance, and so from these error measures and Equation (4.26) we are able to provide a lower bound for these marginal observables:

$$\Delta(\mathsf{E}^{(\lambda,\mu)},\mathsf{E}^{Q})^2\Delta(\mathsf{F}^{(\lambda,\mu)},\mathsf{E}^{P})^2 \geq \mathrm{Var}\,(Q,\tau)\,\mathrm{Var}\,(P,\tau) \geq \frac{1}{4}, \tag{4.28}$$

by Equation (4.1). That is, the product of the errors is bounded from below by the preparation uncertainty relation with respect to the state τ that defines the observable $\mathsf{G}^{(\lambda,\mu)}$. It should be noted that, since Equations (4.24) and (4.26) hold for the margins of any covariant phase space observable, Equation (4.28) is a general result for the margins of a covariant phase observable.

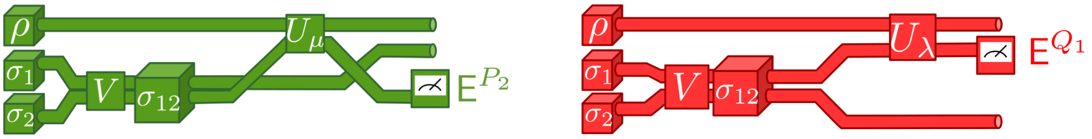
4.3 Focussing

We concluded the previous section by showing that our marginal observables $\mathsf{E}^{(\lambda,\mu)}$, $\mathsf{F}^{(\lambda,\mu)}$, and indeed all marginal observables of a covariant phase space observable, satisfy Equation (4.28) when using either the noise or BLW error measure. Despite this result, it was claimed by Di Lorenzo [27] that the measurement scheme presented, when combined with his measure of error and disturbance, could lead to a lower bound of their products that could be made arbitrarily small or even negative. In this section, we make sense of the definition of disturbance given by Di Lorenzo (his definition of error is of a similar form, but not identical), and show that it in fact highlights the phenomenon of focussing. Note that whilst Di Lorenzo focusses on sequential measurements, we express the output in terms of joint measurements for reasons argued in Section 2.4.4.

4.3.1 Individual measurements and Di Lorenzo's disturbance

In his paper, Di Lorenzo defines the disturbance caused by an observable in the following way: One prepares a measurement set up as we have discussed so far, in which we couple a system in the pure state ψ with two probes in the pure state φ_{12} (for simplicity's sake) via a unitary U_{int}. However, at this point we switch one of the coupling constants, either λ or μ, to zero.

Suppose we set λ to zero, so the system is not directly coupled to the first probe and the coupling unitary reduces to U_μ. We now perform the ideal momentum measurement on the second probe, thereby leading to the effective observable $\mathsf{F}^{(0,\mu)}$ on the system. This is shown schematically in Figure 4.3a. The observable $\mathsf{F}^{(0,\mu)}$, as shown in Appendix C.1,



(a) The measurement scheme leading to the effective observable $\mathsf{F}^{(0,\mu)}$.

(b) The measurement scheme leading to the effective observable $\mathsf{E}^{(\lambda,0)}$.

Figure 4.3: Single measurement schemes.

is a smearing of the ideal momentum observable:

$$\mathsf{F}^{(0,\mu)}(Y) = (\chi_Y * f^{(0,\mu)})(P), \tag{4.29}$$

where the probability distribution $f^{(0,\mu)}$ is of the form

$$f^{(0,\mu)}(p) = \frac{1}{\mu}\int_{\mathbb{R}} dw \left|\widetilde{\varphi}(w, -\frac{1}{\mu}p)\right|^2. \tag{4.30}$$

Similarly, we can consider the case where we let $\mu = 0$, thereby only directly coupling the system to the first probe via the unitary U_λ. In this case we only measure the first probe, performing an ideal position measurement, resulting in the effective observable $\mathsf{E}^{(\lambda,0)}$ being measured on the system. This is presented schematically in Figure 4.3b, and it is shown in Appendix C.2 that $\mathsf{E}^{(\lambda,0)}$ is of the form

$$\mathsf{E}^{(\lambda,0)}(X) = (\chi_X * e^{(\lambda,0)})(Q), \tag{4.31}$$

with

$$e^{(\lambda,0)}(q) = \lambda \int_{\mathbb{R}} dq' \left|\varphi(-\lambda q, q')\right|^2. \tag{4.32}$$

With these two single-probe measurements defined, Di Lorenzo's definition of disturbance η_{DL}^2 is given by the difference of the variances of one of the marginal observables and its single-probe measurement counterpart with respect to a state $\rho \in \mathcal{S}(\mathcal{H})$:

$$\eta_{\mathrm{DL}}(Q,\rho)^2 = \mathrm{Var}\left(\mathsf{E}^{(\lambda,\mu)}, \rho\right) - \mathrm{Var}\left(\mathsf{E}^{(\lambda,0)}, \rho\right), \tag{4.33a}$$

$$\eta_{\mathrm{DL}}(P,\rho)^2 = \mathrm{Var}\left(\mathsf{F}^{(\lambda,\mu)}, \rho\right) - \mathrm{Var}\left(\mathsf{F}^{(0,\mu)}, \rho\right). \tag{4.33b}$$

Given the form of $\mathsf{E}^{(\lambda,0)}$ and $\mathsf{F}^{(0,\mu)}$, and the fact that the variance of a convolution is equal to the sum of their variances, it follows that these disturbances are in fact state-independent:

$$\eta_{\mathrm{DL}}(Q,\rho)^2 = \mathrm{Var}\left(e^{(\lambda,\mu)}\right) - \mathrm{Var}\left(e^{(\lambda,0)}\right) =: \eta_{\mathrm{DL}}(Q), \tag{4.34a}$$

$$\eta_{\mathrm{DL}}(P,\rho)^2 = \mathrm{Var}\left(f^{(\lambda,\mu)}\right) - \mathrm{Var}\left(f^{(0,\mu)}\right) =: \eta_{\mathrm{DL}}(P). \tag{4.34b}$$

By making use of Equations (A.16), (A.23), (C.11) and (C.17), we can express $\eta_{\mathrm{DL}}(Q)$

and $\eta_{\mathrm{DL}}(P)$ in the form

$$\eta_{\mathrm{DL}}(Q) = \frac{\mu^2}{4}(1-\kappa)^2 \mathrm{Var}\,(Q_2, \varphi) - \frac{\mu}{\lambda}(1-\kappa)\mathrm{Cov}(Q_1, Q_2, \varphi), \tag{4.35a}$$

$$\eta_{\mathrm{DL}}(P) = \frac{\lambda^2}{4}(1+\kappa)^2 \mathrm{Var}\,(P_1, \varphi) - \frac{\lambda}{\mu}(1+\kappa)\mathrm{Cov}(P_1, P_2, \varphi), \tag{4.35b}$$

where $\mathrm{Cov}(Q_1, Q_2, \varphi) = \langle \varphi | \frac{1}{2}(Q_1 Q_2 + Q_2 Q_1)\varphi \rangle - \langle \varphi | Q_1 \varphi \rangle \langle \varphi | Q_2 \varphi \rangle$ is the covariance between Q_1 and Q_2 with respect to the state φ, etc. We can replace the pure state $\varphi \in \mathcal{H}_1 \otimes \mathcal{H}_2$ with the mixed state $\sigma \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ for our probes and arrive at the same form of η_{DL} for both Q and P.

As was noted in Di Lorenzo's paper, these quantities can be made arbitrarily small or even negative, and this only occurs if the probes are prepared in a non-separable or a mixed state (so that the covariance terms do not immediately go to zero). In the paper, a negative value of η_{DL} is considered to be a reduction in the uncertainty in the margin observable compared to its single-measure counterpart. However, given that disturbance is an absolute value (an observable or state is either disturbed or it is not), a negative value cannot make sense. Furthermore, this definition is specific to the measurement schemes outlined, and is therefore not extendible to additional situations. With this construction, negative values are not implausible, as one begins by performing a sub-optimal measurement, and then use the statistical spread of this measurement as a benchmark for subsequent indirect measurements. If one had begun by performing the ideal position measurement, say, then $\eta_{\mathrm{DL}}(Q)$ would reduce to the variance of $e^{(\lambda,\mu)}$ and we would arrive at a similar form of error value for $\mathsf{E}^{(\lambda,\mu)}$.

As a result, η_{DL} does not correspond to a physically relevant definition of disturbance, but does highlight the phenomenon of *focussing*, where $\mathsf{E}^{(\lambda,\mu)}$, say, is more precise than $\mathsf{E}^{(\lambda,0)}$. In other words, by performing a joint/sequential measurement we have found marginal observables that are less statistically spread out than if we had performed just the measurements on single probes. As hinted at in the previous paragraph, this is only possible if we allow for prior correlations between the probes; that is, if the probes are prepared in either an entangled or mixed state.

4.3.2 Examples of focussing

We shall conclude this chapter by showing some examples of probe states that can lead to focussing in both margins. One example will be an entangled state, whilst the other will be a mixed state. Given that we now know η_{DL} does not correspond to a physically valid disturbance measure, we shall relabel it by $\mathcal{F}$.

We shall first consider the two-level Gaussian pure state

$$\varphi(x,y) = \left(\frac{4 \det D}{\pi^2}\right)^{1/4} \exp[-(x,y)D(x,y)^T] = \left(\frac{4 \det D}{\pi^2}\right)^{1/4} \exp[-(ax^2 + 2bxy + dy^2)], \tag{4.36}$$

where D is the matrix

$$D = \begin{pmatrix} a & b \\ b & d \end{pmatrix}. \tag{4.37}$$

Similarly, we can express φ in terms of its momentum representation

$$\begin{aligned}\widetilde{\varphi}(w,z) &= \left(\frac{1}{4\pi^2 \det D}\right)^{1/4} \exp\left[-\frac{1}{4}(w,z)D^{-1}(w,z)^T\right] \\ &= \left(\frac{1}{4\pi^2 \det D}\right)^{1/4} \exp\left[-\frac{1}{4\det D}(dw^2 - 2bwz + az^2)\right].\end{aligned} \tag{4.38}$$

From these two forms one quickly verifies that

$$D = \begin{pmatrix} \langle P_1^2\rangle_\varphi & \langle P_1P_2\rangle_\varphi \\ \langle P_1P_2\rangle_\varphi & \langle P_2^2\rangle_\varphi \end{pmatrix} = 4\det D \begin{pmatrix} \langle Q_2^2\rangle_\varphi & -\langle Q_1Q_2\rangle_\varphi \\ -\langle Q_1Q_2\rangle_\varphi & \langle Q_1^2\rangle_\varphi \end{pmatrix}, \tag{4.39}$$

where $\langle P_1\rangle_\varphi = \langle \varphi | P_1 \varphi\rangle$, etc. This state is unbiased, i.e., $\langle Q_i\rangle_\varphi = \langle P_i\rangle_\varphi = 0$ for $i = 1, 2$, and so the variances and covariances reduce, hence

$$D = \begin{pmatrix} \mathrm{Var}\,(P_1,\varphi) & \mathrm{Cov}(P_1,P_2,\varphi) \\ \mathrm{Cov}(P_1,P_2,\varphi) & \mathrm{Var}\,(P_2,\varphi) \end{pmatrix} = 4\det D \begin{pmatrix} \mathrm{Var}\,(Q_2,\varphi) & -\mathrm{Cov}(Q_1,Q_2,\varphi) \\ -\mathrm{Cov}(Q_1,Q_2,\varphi) & \mathrm{Var}\,(Q_1,\varphi) \end{pmatrix}. \tag{4.40}$$

With this in mind, the expressions for $\mathcal{F}(Q)$ and $\mathcal{F}(P)$ can be expressed in terms of a, b and d:

$$\mathcal{F}(Q) = \frac{1}{4\det D}\left(\frac{\mu^2}{4}(1-\kappa)^2 a + \frac{\mu}{\lambda}(1-\kappa)b\right), \tag{4.41a}$$

$$\mathcal{F}(P) = \frac{\lambda^2}{4}(1+\kappa)^2 a - \frac{\lambda}{\mu}(1+\kappa)b. \tag{4.41b}$$

In order for both $\mathcal{F}(Q)$ and $\mathcal{F}(P)$ to be negative we require that

$$-\frac{(1-\kappa)}{\lambda}\mu b > \frac{(1-\kappa)^2}{4}\mu^2 a > 0, \tag{4.42a}$$

$$\frac{(1+\kappa)}{\mu}\lambda b > \frac{(1+\kappa)^2}{4}\lambda^2 a > 0. \tag{4.42b}$$

This restricts us to the case where $|\kappa| > 1$, otherwise we would require $b < 0$ from Equation (4.42a) and $b > 0$ from Equation (4.42b), which clearly cannot be achieved simultaneously. (Note that we have relied here on the requirement that $\lambda, \mu > 0$.)

Suppose that we let $\kappa > 1$. In this case $b > 0$ in Equation (4.42b) and

$$\frac{b}{a} > \frac{(1+\kappa)}{4}\lambda\mu. \tag{4.43}$$

Similarly, if $\kappa < -1$, then $b < 0$ in Equation (4.42a) and

$$\frac{|b|}{a} > \frac{(1-\kappa)}{4}\lambda\mu. \tag{4.44}$$

We can therefore find focussing in both margins if we prepare the probes in the two-level Gaussian state provided $|\kappa| > 1$ and

$$\frac{|b|}{a} > \frac{(1+|\kappa|)}{4}\lambda\mu. \tag{4.45}$$

The second example that we consider is the mixed state $\sigma = p\sigma_1 + (1-p)\sigma_2$, with $p \in (0,1)$. Both states σ_1 and σ_2 are pure separable states composed of one-level Gaussian states centred on the same point (x_i, w_i) in phase space. In other words, the state σ_i is of the form

$$\sigma_i = P_{\varphi_i^{(1)} \otimes \varphi_i^{(2)}}, \tag{4.46}$$

such that $\langle Q_1 \rangle_{\sigma_i} = \langle Q_2 \rangle_{\sigma_i} = x_i$ and $\langle P_1 \rangle_{\sigma_i} = \langle P_2 \rangle_{\sigma_i} = w_i$. Further to this, we assume that the pure states have a fixed variance S with respect to the position operators and R with respect to the momentum operator:

$$\mathrm{Var}\,(Q_1, \sigma_i) = \mathrm{Var}\,(Q_2, \sigma_i) = S, \qquad \mathrm{Var}\,(P_1, \sigma_i) = \mathrm{Var}\,(P_2, \sigma_i) = R, \tag{4.47}$$

for $i = 1, 2$. The covariances for this state are then given by

$$\begin{aligned} \mathrm{Cov}(Q_1, Q_2, \sigma) &= \langle Q_1 Q_2 \rangle_\sigma - \langle Q_1 \rangle_\sigma \langle Q_2 \rangle_\sigma \\ &= \mathrm{tr}\,[Q_1 Q_2 (p\sigma_1 + (1-p)\sigma_2)] \\ &\quad - \mathrm{tr}\,[Q_1 (p\sigma_1 + (1-p)\sigma_2)]\,\mathrm{tr}\,[Q_2 (p\sigma_1 + (1-p)\sigma_2)] \\ &= px_1^2 + (1-p)x_2^2 - (px_1 + (1-p)x_2)^2 \\ &= (p - p^2)(x_1 - x_2)^2. \end{aligned} \tag{4.48}$$

Similarly,

$$\mathrm{Cov}(P_1, P_2, \sigma) = (p - p^2)(w_1 - w_2)^2. \tag{4.49}$$

In both cases the covariance is strictly non-negative, which means, by comparing with Equation (4.35), that if we want to be able to make both $\mathcal{F}(Q)$ and $\mathcal{F}(P)$ negative we require $|\kappa| < 1$, thereby ensuring that both $(1+\kappa)$ and $(1-\kappa)$ are positive.

We can also calculate the variance of Q_2, say, with respect to σ:

$$\begin{aligned} \mathrm{Var}\,(Q_2, \sigma) &= \mathrm{tr}\left[(p\sigma_1 + (1-p)\sigma_2)Q_2^2\right] - \mathrm{tr}\,[(p\sigma_1 + (1-p)\sigma_2)Q_2]^2 \\ &= p(S + x_1^2) + (1-p)(S + x_2^2) - (px_1 + (1-p)x_2)^2 \\ &= S + px_1^2 + (1-p)x_2^2 - (px_1 + (1-p)x_2)^2 \\ &= S + \mathrm{Cov}(Q_1, Q_2, \sigma), \end{aligned} \tag{4.50}$$

where in the second equality we used

$$\mathrm{tr}\left[\sigma_i Q_2^2\right] = \mathrm{Var}\,(Q_2, \sigma_i) + \mathrm{tr}\,[\sigma_i Q_2]^2 = S + x_i^2. \tag{4.51}$$

Given the structure of the states we have considered, we can immediately infer that

$$\mathrm{Var}\,(P_1, \sigma) = R + \mathrm{Cov}(P_1, P_2, \sigma). \tag{4.52}$$

If we position σ_1 and σ_2 far apart in phase space, then the covariances will keep increasing in value, and by Equations (4.50) and (4.52) we know that the variance is guaranteed to satisfy $\mathrm{Var}\,(Q_1, \sigma)\,\mathrm{Var}\,(Q_2, \sigma) \geq \mathrm{Cov}(Q_1, Q_2, \sigma)$, etc., as required. We can now rewrite

Equation (4.35) as

$$\begin{aligned}\mathcal{F}(Q) &= \frac{\mu^2}{4}(1-\kappa)^2\mathrm{Var}\,(Q_2,\varphi) - \frac{\mu}{\lambda}(1-\kappa)\mathrm{Cov}(Q_1,Q_2,\varphi) \\ &= \frac{\mu^2}{4}(1-\kappa)^2(S+\mathrm{Cov}(Q_1,Q_2,\sigma)) - \frac{\mu}{\lambda}(1-\kappa)\mathrm{Cov}(Q_1,Q_2,\varphi) \\ &= \frac{\mu^2}{4}(1-\kappa)^2 S + \frac{\mu}{\lambda}(1-\kappa)\mathrm{Cov}(Q_1,Q_2,\sigma)\left(\frac{(1-\kappa)}{4}\lambda\mu - 1\right),\end{aligned} \tag{4.53a}$$

$$\begin{aligned}\mathcal{F}(P) &= \frac{\lambda^2}{4}(1+\kappa)^2\mathrm{Var}\,(P_1,\varphi) - \frac{\lambda}{\mu}(1+\kappa)\mathrm{Cov}(P_1,P_2,\varphi) \\ &= \frac{\lambda^2}{4}(1+\kappa)^2 R + \frac{\lambda}{\mu}(1+\kappa)\mathrm{Cov}(P_1,P_2,\sigma)\left(\frac{(1+\kappa)}{4}\lambda\mu - 1\right).\end{aligned} \tag{4.53b}$$

We can therefore see that if we prepare our probes in the mixed state σ and set $|\kappa| < 1$, then we are capable of finding both $\mathcal{F}(Q)$ and $\mathcal{F}(P)$ to be negative if the covariances are sufficiently large and

$$\frac{(1+|\kappa|)}{4}\lambda\mu < 1, \tag{4.54}$$

which is a requirement of a very similar form to Equation (4.45). This condition restricts the size of the coupling parameters that we may use if we wish to witness focussing, and so the assumption of ignoring the free evolution of the system and two probes may not be justified.

Chapter 5

Incompatibility and Error Relations for Qubit Observables

Incompatibility is a ubiquitous feature of quantum theory, with even the simplest quantum systems possessing observables that cannot be measured together. One way of mitigating the impossibility of realising accurate joint measurements of two incompatible quantities, denoted throughout this chapter by A and B, is to approximate them by two observables, denoted by C and D, that are jointly measurable, as shown in Figure 5.1. The degree to which an approximating observable is an accurate representation of the initial observable, the so-called *target observable*, is given by a measure of error. When attempting to approximate the two target observables, the requirement of joint measurability that we impose on the approximating observables places constraints on the possible error values that we may obtain. For example, performing a perfect approximation of the first target observable prohibits performing an equally accurate approximation of the second, as these approximating observables would be incompatible.

Suppose that we consider an (arbitrary) error measure D; for this measure we can plot the range of possible pairs of error values $(D(\mathsf{C},\mathsf{A}), D(\mathsf{D},\mathsf{B}))$. The joint measurability condition has the following consequence: for any value of $D(\mathsf{C},\mathsf{A})$, say, we have a (usually non-zero) minimum value for $D(\mathsf{D},\mathsf{B})$ that we may obtain whilst still ensuring C and D are jointly measurable. This error tradeoff is quantified by so-called *measurement uncertainty relations*, and are dependent on the level of incompatibility between A and B. In their most general form, measurement uncertainty relations can be expressed as

$$f\big(D(\mathsf{C},\mathsf{A}), D(\mathsf{D},\mathsf{B})\big) \geq g(\mathsf{A},\mathsf{B}), \tag{5.1}$$

where $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathcal{L}_s(\mathcal{H}) \to \mathbb{R}$ are functions that describe the relation. (In the case of state-dependent relations, for example, g often takes the form of the expectation value of the commutator of $\mathsf{A}[1]$ and $\mathsf{B}[1]$ divided by $2i$.) Any pair of error values that satisfy Equation (5.1) are said to lie in the so-called *admissible region*, whilst any pair of errors that saturate the inequality, and therefore correspond to a pair of the smallest possible error values under the constraint of joint measurability, form the lower bound of the admissible region. It is possible that the lower bound of an admissible region, as described by a given uncertainty relation, cannot be reached. However, when the lower

$$\begin{tikzcd}
\mathsf{A} \arrow[d, no head, squiggly] & & \mathsf{B} \arrow[d, no head, squiggly] \\
\mathsf{C} & & \mathsf{D} \\
& \mathsf{J} \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

Figure 5.1: Two incompatible observables A and B are approximated via jointly measurable observables C and D, with joint observable J. The error measures act as a merit of how well C (D) approximates A (B).

bound can be met, that is, an uncertainty relation can be saturated, then we refer to it as a *tight error bound*.

In this chapter we discuss two reported tight error bounds for approximating two incompatible sharp dichotomic (two-valued) qubit observables via jointly measurable observables, as discussed above. A qualitative description of a bound for this kind of observable has previously been given [14], with the error measure being the distance between the observables considered, and a linear tight approximation of the bound was provided. The first error bound that we shall discuss, given by Branciard [7], is in terms of the noise measure introduced by Ozawa [40] and is shown to be an attainable lower bound for any two incompatible observables in any dimension of Hilbert space considered. The second bound, presented by Yu and Oh [57], is given strictly for dichotomic qubit observables and is based on the error measure considered by Busch, Lahti and Werner [16]. This bound, while again shown to be attainable, was originally given parametrically and the operational meaning of the terms given were not immediately clear.

The original goal of the work in this chapter was to attempt to reconcile these two bounds, such that the bound given by Yu and Oh may be extended to higher dimensional systems. However, this task was quickly seen to be impossible, so we then focussed on a comparison of the relative worth of each bound and which classes of observables provided optimal approximations, that is, which observables provided error values that saturated the bound for a given error measure. Examples of optimal approximating observables were given in each paper.

These two papers address the problem from slightly different directions, and so we need to reword some ideas so that they may be more easily compared. In what follows we examine Branciard's method of producing jointly measurable approximating observables, which is based on the work of Hall [28] and is expressed in terms of first moment operators. We also spell out the meaning of the terms given in Yu and Oh's minimum error quantities that are wrapped up in trigonometric quantities and provide a derivation of the error bound in Appendix E, as this was not given with much detail in the original letter and allows us to uncover and fix a shortcoming in their given argument. We then provide a comparison of these two bounds, and show that whilst there exist certain instances where the optimal approximators of one bound are also optimal for the other, these are indeed special cases. Furthermore, we show that the class of observables that saturate the Branciard's bound for the noise measure is much greater than that for Yu and Oh's bound for the BLW error measure, which suggests that the noise measure provides misleading assessments of error. Indeed, the combination of Hall and Branciard's joint measurement scheme and the bound given for the noise measure lead to a questionable form of "optimal" joint measurement of two incompatible observables.

5.1 Branciard's joint observable construction

In order to allow for a comparison between Branciard's work and the work of Yu and Oh, we need to express Branciard's construction of jointly measurable approximating observables in terms of the language presented thus far (Yu and Oh's work is already presented within a compatible language).

In his paper [7], adopting the method of Hall [28], Branciard considers a Hilbert space $\mathcal{H}$ and an ancillary system $\mathcal{K}$ in a fixed state ξ, upon which he measures the discrete sharp observable $\mathsf{M} : \{1, \ldots, N\} \to \mathcal{E}(\mathcal{H} \otimes \mathcal{K})$ associated with the self-adjoint operator $M = \mathsf{M}[1] = \sum_{m=1}^{N} m\, \mathsf{M}(m) \in \mathcal{L}_s(\mathcal{H} \otimes \mathcal{K})$. It is this self-adjoint operator that Branciard considers to be his observable, and not the PVM that describes it. As a result his discussion on approximating two observables is in fact focussed on approximating the two self-adjoint operators associated with two incompatible POVMs, $A = \mathsf{A}[1]$ and $B = \mathsf{B}[1]$, and not the POVMs themselves. This approximation is performed by applying two functions, f and g, on M such that for each eigenvalue m, $f(m)$ approximates an eigenvalue of the first self-adjoint operator and $g(m)$ approximates an eigenvalue of the second. This then leads to our two approximating self-adjoint operators

$$f(M) = \sum_{m=1}^{N} f(m)\mathsf{M}(m), \qquad g(M) = \sum_{m=1}^{N} g(m)\mathsf{M}(m). \tag{5.2}$$

Since these operators share a common eigenbasis (the elements of the PVM M) they must commute.

What is discussed above is an example of a Naimark dilation, and we define the POVM $\mathsf{E} : \{1, \ldots, N\} \to \mathcal{E}(\mathcal{H})$ by performing the partial trace over the ancillary system (cf. Equation (2.145) with $U = I$, $\sigma = P_\xi$ and $\mathsf{Z} = \mathsf{M}$),

$$\mathsf{E}(m) = \mathrm{tr}_{\mathcal{K}} \left[(I \otimes P_\xi)\mathsf{M}(m) \right]. \tag{5.3}$$

Whilst not considered in Hall and Branciard's work, we can construct the POVMs C and D that approximate A and B, respectively, via

$$\mathsf{C}(k) = \sum_{m \in f^{-1}(k)} \mathsf{E}(m), \qquad \mathsf{D}(\ell) = \sum_{m \in g^{-1}(\ell)} \mathsf{E}(m), \tag{5.4}$$

where $k \in \mathrm{ran}(f)$ and $\ell \in \mathrm{ran}(g)$. In other words, the functions f and g form partitions of the set of effects of M, and C and D satisfy

$$\langle \psi | \mathsf{C}[1] \psi \rangle = \langle \psi \otimes \xi | f(M) \psi \otimes \xi \rangle, \qquad \langle \psi | \mathsf{D}[1] \psi \rangle = \langle \psi \otimes \xi | g(M) \psi \otimes \xi \rangle. \tag{5.5}$$

The bins of the partitions defined by f and g will necessarily overlap to some extent and, by including null effects when needed, we may define the joint observable J for C and D as

$$\mathsf{J}(k,\ell) = \sum_{m} \chi_{f^{-1}(k)\cap g^{-1}(\ell)}(m)\mathsf{E}(m), \tag{5.6}$$

where χ denotes the characteristic function. The positivity and normalisation of these effects is given, and we quickly see that by summing over the outcome space of D or C we retrieve effects of C or D, respectively.

Whilst we have shown that recovering these observables as margins of J is indeed possible, we must stress that Hall's method of deriving a joint approximation for two self-adjoint operators is physically lacking. This method was originally highlighted by Hall based on the idea that "any measurement [in the sense presented above] is considered to provide a joint measurement of any two observables" if one simply rescales the values of the possible measurement outcomes via the functions f and g. However, as we show in Appendix D, such functions can lead to suboptimal approximating observables possessing noise measure values equal to zero.

5.2 Qubit error measures

We will consider the noise measure and the BLW error measure, as discussed in Section 2.4.5. The target observables A and B are sharp qubit observables, i.e., two-valued observables with effects of the form

$$\mathsf{A}(\pm) = \frac{1}{2}(I \pm \boldsymbol{a}\cdot\boldsymbol{\sigma}), \qquad \mathsf{B} = \frac{1}{2}(I \pm \boldsymbol{b}\cdot\boldsymbol{\sigma}), \tag{5.7}$$

where $\|\boldsymbol{a}\| = \|\boldsymbol{b}\| = 1$. The first moment operators associated with these observables are therefore $\mathsf{A}[1] = \boldsymbol{a}\cdot\boldsymbol{\sigma}$ and $\mathsf{B}[1] = \boldsymbol{b}\cdot\boldsymbol{\sigma}$. The observables C and D approximating A and B, respectively, that we will consider are also dichotomic and covariant under value swaps, i.e., there exists a unitary operator U such that

$$U\mathsf{C}(\pm)U^* = \mathsf{C}(\mp), \qquad U\mathsf{D}(\pm)U^* = \mathsf{D}(\mp). \tag{5.8}$$

Such observables are necessarily unbiased; that is, their effects are of the form

$$\mathsf{C}(\pm) = \frac{1}{2}(I \pm \boldsymbol{c}\cdot\boldsymbol{\sigma}), \qquad \mathsf{D}(\pm) = \frac{1}{2}(I \pm \boldsymbol{d}\cdot\boldsymbol{\sigma}), \tag{5.9}$$

thereby ensuring that both effects for either observable possess the same spectrum.

The reason for this choice of approximating observables is threefold: firstly, it ensures that the target and approximating observables possess the same value space, which is a feature of use later; secondly, there exists a simple Bloch-geometric requirement for ensuring the joint measurability of dichotomic observables; thirdly, as we will shortly see, the noise measure possesses an interesting property for the case of dichotomic observables covariant under value swaps that will allow us to compare it against the BLW error measure.

The condition of covariance given in Equation (5.8) is equivalent to the Bloch vector $\boldsymbol{u}$ satisfying $U = \boldsymbol{u}\cdot\boldsymbol{\sigma}$ being perpendicular to both $\boldsymbol{c}$ and $\boldsymbol{d}$. (In the case that $\boldsymbol{c} \perp \boldsymbol{d}$, U

amounts to a Heisenberg-Weyl shift operator.) The first moment operators of C and D are of the form

$$\mathsf{C}[1] = \boldsymbol{c}\cdot\boldsymbol{\sigma}, \qquad \mathsf{D}[1] = \boldsymbol{d}\cdot\boldsymbol{\sigma}, \tag{5.10}$$

whilst the second moment for both observables is the identity. The squares of the first moment operators are equal to the identity times the squared norm of their respective Bloch vectors, i.e., $\mathsf{C}[1]^2 = \|\boldsymbol{c}\|^2 I$, etc. Using Equation (2.158) we can readily calculate $\varepsilon(\mathsf{C},\mathsf{A},\psi)^2$ and $\varepsilon(\mathsf{D},\mathsf{B},\psi)^2$:

$$\begin{aligned}\varepsilon(\mathsf{C},\mathsf{A},\psi)^2 &= \langle\psi|(\mathsf{C}[1]-\mathsf{A}[1])^2\psi\rangle + \langle\psi|(\mathsf{C}[2]-\mathsf{C}[1]^2)\psi\rangle \\ &= \langle\psi|[(\boldsymbol{a}-\boldsymbol{c})\cdot\boldsymbol{\sigma}]^2\psi\rangle + \left\langle\psi\middle|(1-\|\boldsymbol{c}\|^2)I\psi\right\rangle \\ &= \|\boldsymbol{a}-\boldsymbol{c}\|^2 + 1 - \|\boldsymbol{c}\|^2\,,\end{aligned} \tag{5.11}$$

where we have made use of the identity $(\boldsymbol{a}\cdot\boldsymbol{\sigma})(\boldsymbol{b}\cdot\boldsymbol{\sigma}) = (\boldsymbol{a}\cdot\boldsymbol{b})I + i(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{\sigma}$, and similarly,

$$\varepsilon(\mathsf{D},\mathsf{B},\psi)^2 = \|\boldsymbol{b}-\boldsymbol{d}\|^2 + 1 - \|\boldsymbol{d}\|^2\,. \tag{5.12}$$

A point of note is that both of these quantities are independent of the state considered, and that this is a consequence of considering covariant approximating observables. Given that the noise measure is state-independent in the case of these observables, we shall use the shorthands $\varepsilon_A := \varepsilon(\mathsf{C},\mathsf{A},\psi)$ and $\varepsilon_B := \varepsilon(\mathsf{D},\mathsf{B},\psi)$ in what follows.

We now briefly calculate the values $\Delta(\mathsf{C},\mathsf{A})^2$ and $\Delta(\mathsf{D},\mathsf{B})^2$ using equations (2.162) and (2.164), following the method given in [18]. Supposing that we possess a system in a state $\rho = (I + \boldsymbol{r}\cdot\boldsymbol{\sigma})/2$, a general coupling γ that we can use for the probability distributions p^{A}_ρ and p^{C}_ρ is of the form

$$\begin{aligned}&\gamma(+,+) = \alpha, &&\gamma(-,+) = p^{\mathsf{A}}_\rho(+) - \alpha,\\ &\gamma(+,-) = p^{\mathsf{C}}_\rho(+) - \alpha, &&\gamma(-,-) = 1 - p^{\mathsf{C}}_\rho(+) - p^{\mathsf{A}}_\rho(+) + \alpha.\end{aligned} \tag{5.13}$$

Using this coupling in Equation (2.162) gives us

$$\begin{aligned}\Delta_\rho(\mathsf{C},\mathsf{A})^2 &= \sum_{x\in\{-1,1\}}\sum_{y\in\{-1,1\}} (x-y)^2\,\gamma(x,y)\\ &= 4\big(\gamma(+,-)+\gamma(-,+)\big)\\ &= 4(p^{\mathsf{C}}_\rho(+) + p^{\mathsf{A}}_\rho(+) - 2\alpha).\end{aligned} \tag{5.14}$$

In order to minimise this quantity, whilst still ensuring its positivity, we require that $\alpha = \min\{p^{\mathsf{C}}_\rho(+), p^{\mathsf{A}}_\rho(+)\}$, and so

$$\begin{aligned}\Delta_\rho(\mathsf{C},\mathsf{A})^2 &= 4\left|p^{\mathsf{C}}_\rho(+) - p^{\mathsf{A}}_\rho(+)\right|\\ &= 2\,|(\boldsymbol{c}-\boldsymbol{a})\cdot\boldsymbol{r}|\,.\end{aligned} \tag{5.15}$$

The quantity $\Delta(\mathsf{C},\mathsf{A})^2$ is found by taking the supremum over all states, which in this case comes from choosing the Bloch vector $\boldsymbol{r}$ parallel to $\boldsymbol{c}-\boldsymbol{a}$, i.e., $\boldsymbol{r} = (\boldsymbol{c}-\boldsymbol{a})/\,\|\boldsymbol{c}-\boldsymbol{a}\|$.

Hence,

$$\Delta(\mathsf{C},\mathsf{A})^2 = 2\left\|\boldsymbol{a}-\boldsymbol{c}\right\|, \tag{5.16}$$

and, similarly,

$$\Delta(\mathsf{D},\mathsf{B})^2 = 2\left\|\boldsymbol{b}-\boldsymbol{d}\right\|. \tag{5.17}$$

By introducing the *unsharpness* $U(\mathsf{C})^2$ of the observable C

$$U(\mathsf{C})^2 = 1 - \left\|\boldsymbol{c}\right\|^2, \tag{5.18}$$

and likewise for $U(\mathsf{D})^2$, we can rewrite ε_A^2 and ε_B^2 in the case of covariant dichotomic observables as

$$\begin{aligned} \varepsilon_A^2 &= \frac{1}{4}\Delta(\mathsf{C},\mathsf{A})^4 + U(\mathsf{C})^2, \\ \varepsilon_B^2 &= \frac{1}{4}\Delta(\mathsf{D},\mathsf{B})^4 + U(\mathsf{D})^2. \end{aligned} \tag{5.19}$$

Further to this, in the special case that C is derived from A via a trivial smearing, i.e., $\boldsymbol{c} = \gamma\boldsymbol{a}$ where $0 < \gamma < 1$, the quantity ε_A^2 reduces:

$$\begin{aligned} \varepsilon_A^2 &= 1 + \left((1-\gamma)^2 - \gamma^2\right)\left\|\boldsymbol{a}\right\|^2 \\ &= 2(1-\gamma) = 2(1-\gamma)\left\|\boldsymbol{a}\right\| \\ &= 2\left\|\boldsymbol{a}-\boldsymbol{c}\right\| \\ &= \Delta(\mathsf{C},\mathsf{A})^2. \end{aligned} \tag{5.20}$$

These quantities can also be compared to the probabilistic distance $\mathfrak{D}$ introduced in [14],

$$\mathfrak{D}(\mathsf{C},\mathsf{A}) = \max_X \sup_{\rho\in\mathcal{S}(\mathcal{H})} \left|\mathrm{tr}\left[\rho\mathsf{A}(X)\right] - \mathrm{tr}\left[\rho\mathsf{C}(X)\right]\right| = \max_X \left\|\mathsf{A}(X) - \mathsf{C}(X)\right\|, \tag{5.21}$$

where we have assumed that A and C possess the same outcome space and $X \in \Omega_{\mathsf{A}}$. For our dichotomic observables this difference of effects is of the form

$$\mathsf{A}(\pm) - \mathsf{C}(\pm) = \pm\frac{1}{2}(\boldsymbol{a}-\boldsymbol{c})\cdot\boldsymbol{\sigma}, \tag{5.22}$$

and so the probabilistic distance between C and A is

$$\mathfrak{D}(\mathsf{C},\mathsf{A}) = \frac{1}{2}\left\|\boldsymbol{a}-\boldsymbol{c}\right\| = \frac{1}{4}\Delta(\mathsf{C},\mathsf{A})^2. \tag{5.23}$$

If we then compare this to the noise measure we see that

$$\varepsilon_A^2 = 4\mathfrak{D}(\mathsf{C},\mathsf{A})^2 + U(C)^2. \tag{5.24}$$

In the next section we shall see how these errors lead to bounds on how well two incompatible observables can be approximated in a joint measurement scheme.

5.3 The error bounds

The motivations leading to the minimum error bounds will be given here. Given that the details for the derivation of Yu and Oh's bound were omitted in the original paper, we shall include them here in Appendix E. The bounds presented here describe the smallest possible errors that may be obtained during a joint measurement of C and D when approximating A and B, respectively, with respect to their given measure of error.

5.3.1 Branciard's bound

Branciard's bound, which is based on the noise measure ε and was originally given for any finite-dimensional system, is derived by first considering the following pure states on $\mathcal{H} \otimes \mathcal{K}$:

$$\begin{aligned} \alpha &= \frac{A \otimes I - \langle A \rangle_\psi}{\Delta(A)}\, \psi \otimes \xi, & \beta &= \frac{B \otimes I - \langle B \rangle_\psi}{\Delta(B)}\, \psi \otimes \xi, \\ \gamma &= \frac{f(M) - \langle A \rangle_\psi}{\Delta(A)}\, \psi \otimes \xi, & \delta &= \frac{g(M) - \langle B \rangle_\psi}{\Delta(B)}\, \psi \otimes \xi, \end{aligned} \tag{5.25}$$

where the standard deviations

$$\Delta(A) = \sqrt{\langle A^2 \rangle_\psi - \langle A \rangle_\psi^2}, \qquad \Delta(B) = \sqrt{\langle B^2 \rangle_\psi - \langle B \rangle_\psi^2}, \tag{5.26}$$

are assumed to be nonzero. Indeed, Branciard at this point considers the case where $\langle A \rangle_\psi = \langle B \rangle_\psi = 0$ and $A^2 = B^2 = I$, leading to $\Delta(A) = \Delta(B) = 1$, and thereby reducing the above vectors to

$$\begin{aligned} \alpha &= A \otimes I(\psi \otimes \xi), & \beta &= B \otimes I(\psi \otimes \xi), \\ \gamma &= f(M)(\psi \otimes \xi), & \delta &= g(M)(\psi \otimes \xi). \end{aligned} \tag{5.27}$$

A further assumption made by Branciard is that the approximating observables possess the same spectrum as the observables that they are approximating, i.e., C and D are also dichotomic observables, and therefore $f(M)^2 = g(M)^2 = I$. This, along with the condition $A^2 = B^2 = I$ given above, ensures that the four vectors above are normalised. From here, by decomposing the previous states in terms of any orthonormal basis $\{\varphi_j\}_{j=1}^d$ for $\mathcal{H} \otimes \mathcal{K}$, where $d = \dim(\mathcal{H} \otimes \mathcal{K})$, the following vectors are given on the Euclidean space $\mathbb{R}^{2d}$:

$$\begin{aligned} \boldsymbol{e} &= \begin{pmatrix} \mathrm{Re}(\alpha) \\ \mathrm{Im}(\alpha) \end{pmatrix}, & \boldsymbol{f} &= \begin{pmatrix} \mathrm{Im}(\beta) \\ -\mathrm{Re}(\beta) \end{pmatrix}, \\ \boldsymbol{g} &= \begin{pmatrix} \mathrm{Re}(\gamma) \\ \mathrm{Im}(\gamma) \end{pmatrix}, & \boldsymbol{h} &= \begin{pmatrix} \mathrm{Im}(\delta) \\ -\mathrm{Re}(\delta) \end{pmatrix}. \end{aligned} \tag{5.28}$$

In other words, if we consider the j^{th} component of α, for example, then we can express it in terms of the components of $\boldsymbol{e}$ via

$$\alpha_j = \langle \varphi_j | \alpha \rangle = e_j + i\, e_{j+d}. \tag{5.29}$$

The normalisation of the states given in Equation (5.27) guarantees the vectors in Equation (5.28) are also normalised. We see that

$$\|\boldsymbol{e}-\boldsymbol{g}\|^2 = \|(A\otimes I - f(M))\psi\otimes\xi\|^2 = \varepsilon(\mathsf{C},\mathsf{A},\psi)^2, \tag{5.30}$$

where we have made use of the form of the noise measure given in Equation (2.156), and associated it with the observable C via Equation (5.5). Note that in this situation we are dealing with generic observables and so the noise measure is still assumed to be state-dependent. Similarly, we also have

$$\|\boldsymbol{f}-\boldsymbol{h}\|^2 = \varepsilon(\mathsf{D},\mathsf{B},\psi)^2. \tag{5.31}$$

If we take the scalar product of the two vectors $\boldsymbol{g}$ and $\boldsymbol{h}$:

$$\boldsymbol{g}\cdot\boldsymbol{h} = \mathrm{Re}(\gamma)^T\mathrm{Im}(\delta) - \mathrm{Im}(\gamma)^T\mathrm{Re}(\delta), \tag{5.32}$$

and compare it to the inner product of γ and δ:

$$\begin{aligned}\langle\gamma|\delta\rangle &= \big(\mathrm{Re}(\gamma)+i\,\mathrm{Im}(\gamma)\big)^*\big(\mathrm{Re}(\delta)+i\,\mathrm{Im}(\delta)\big)\\ &= \big(\mathrm{Re}(\gamma)^T\mathrm{Re}(\delta)+\mathrm{Im}(\gamma)^T\mathrm{Im}(\delta)\big)\\ &\quad + i\big(\mathrm{Re}(\gamma)^T\mathrm{Im}(\delta)-\mathrm{Im}(\gamma)^T\mathrm{Re}(\delta)\big),\end{aligned} \tag{5.33}$$

we see that

$$\boldsymbol{g}\cdot\boldsymbol{h} = \mathrm{Im}(\langle\gamma|\delta\rangle) = \frac{1}{2i}(\langle\gamma|\delta\rangle - \langle\delta|\gamma\rangle) = \frac{1}{2i}\langle\psi\otimes\xi|[f(M),g(M)]\,\psi\otimes\xi\rangle = 0. \tag{5.34}$$

Similarly,

$$\boldsymbol{e}\cdot\boldsymbol{f} = \mathrm{Im}(\langle\alpha|\beta\rangle) = \frac{1}{2i}\langle\psi|[A,B]\,\psi\rangle =: C_{AB}, \tag{5.35}$$

which is the measure of incompatibility of the two sharp observables that is used in the paper (and, as we shall see, reduces to the measure used by Yu and Oh). Finally, Branciard defines two further scalar quantities

$$\begin{aligned}e_\perp^2 &:= 1-(\boldsymbol{e}\cdot\boldsymbol{g})^2 = \|\boldsymbol{e}-\boldsymbol{g}\|^2\left(1-\frac{\|\boldsymbol{e}-\boldsymbol{g}\|^2}{4}\right),\\ f_\perp^2 &:= 1-(\boldsymbol{f}\cdot\boldsymbol{h})^2 = \|\boldsymbol{f}-\boldsymbol{h}\|^2\left(1-\frac{\|\boldsymbol{f}-\boldsymbol{h}\|^2}{4}\right),\end{aligned} \tag{5.36}$$

which, by comparing with Equations (5.30) and (5.31), we see can be rewritten as

$$e_\perp^2 = \varepsilon(\mathsf{C},\mathsf{A},\psi)^2\left(1-\frac{\varepsilon(\mathsf{C},\mathsf{A},\psi)^2}{4}\right),\qquad f_\perp^2 = \varepsilon(\mathsf{D},\mathsf{B},\psi)^2\left(1-\frac{\varepsilon(\mathsf{D},\mathsf{B},\psi)^2}{4}\right). \tag{5.37}$$

By direct application of a geometric inequality defined on a Euclidean space, as shown in [7], where we require the vectors $\boldsymbol{e}$ and $\boldsymbol{f}$ to be normalised and $\boldsymbol{g}$ and $\boldsymbol{h}$ to be orthonormal:

$$e_\perp^2 + f_\perp^2 + 2\sqrt{1-(\boldsymbol{e}\cdot\boldsymbol{f})^2}\,e_\perp f_\perp \geq (\boldsymbol{e}\cdot\boldsymbol{f})^2, \tag{5.38}$$

we arrive at the Branciard inequality for the noise measure:

$$\begin{aligned}&\varepsilon(\mathsf{C},\mathsf{A},\psi)^2\left(1-\frac{\varepsilon(\mathsf{C},\mathsf{A},\psi)^2}{4}\right)+\varepsilon(\mathsf{D},\mathsf{B},\psi)^2\left(1-\frac{\varepsilon(\mathsf{D},\mathsf{B},\psi)^2}{4}\right)\\&+2\varepsilon(\mathsf{C},\mathsf{A},\psi)\varepsilon(\mathsf{D},\mathsf{B},\psi)\sqrt{1-C_{AB}^2}\sqrt{\left(1-\frac{\varepsilon(\mathsf{C},\mathsf{A},\psi)^2}{4}\right)\left(1-\frac{\varepsilon(\mathsf{D},\mathsf{B},\psi)^2}{4}\right)}\geq C_{AB}^2.\end{aligned}\tag{5.39}$$

As has been stated before, this inequality holds for any finite-dimensional system, and is trivially satisfied when $C_{AB}=0$, but we can give a more useful form for a qubit system. In this case, the first moments of A and B are described by the Bloch vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, i.e., $A=\boldsymbol{a}\cdot\boldsymbol{\sigma}$ and $B=\boldsymbol{b}\cdot\boldsymbol{\sigma}$. Given that we have assumed that $A^2=B^2=I$, we require that the Bloch vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are normalised, and, in order for the state ψ of the system to satisfy the conditions $\langle A\rangle_\psi=\langle B\rangle_\psi=0$, we require that the Bloch vector $\boldsymbol{r}$ that describes the pure state ψ—that is, $P_\psi=(I+\boldsymbol{r}\cdot\boldsymbol{\sigma})/2$—must be orthogonal to both $\boldsymbol{a}$ and $\boldsymbol{b}$. Assuming that $[A,B]\neq 0$, hence $\boldsymbol{b}\neq\pm\boldsymbol{a}$, the vector $\boldsymbol{r}$ is of the form $\boldsymbol{r}=\boldsymbol{a}\times\boldsymbol{b}/\sin\theta$, where $\sin\theta=\|\boldsymbol{a}\times\boldsymbol{b}\|$. We assume that $\theta\in(0,\pi/2]$, as for any angle $\theta\in(\pi/2,\pi)$ we could replace the Bloch vector $\boldsymbol{b}$, say, with $-\boldsymbol{b}$. This would result in a value-swapped version of the observable $\mathsf{B}'(\pm)=\mathsf{B}(\mp)$, but would otherwise lead to the same problem being solved, with the angle $\pi-\theta\in(0,\pi/2)$ between the Bloch vectors. Furthermore, if we consider any angle $\theta\in[\pi,2\pi]$ then we would find $\sin\theta\leq 0$, contradicting our initial definition of $\sin\theta$.

In this case the quantity C_{AB} takes the form

$$C_{AB}=\frac{1}{2i}\left\langle\psi|[A,B]\,\psi\right\rangle=(\boldsymbol{r}\cdot\boldsymbol{\sigma})\big((\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{\sigma}\big)=\frac{\|\boldsymbol{a}\times\boldsymbol{b}\|^2}{\sin\theta}=\sin\theta.\tag{5.40}$$

Using the unbiased dichotomic approximating observables C,D, so that the noise measure becomes state-independent, we can express Equation (5.39) as

$$\varepsilon_A^2\left(1-\frac{\varepsilon_A^2}{4}\right)+\varepsilon_B^2\left(1-\frac{\varepsilon_B^2}{4}\right)+2\varepsilon_A\varepsilon_B\cos\theta\sqrt{\left(1-\frac{\varepsilon_A^2}{4}\right)\left(1-\frac{\varepsilon_B^2}{4}\right)}\geq\sin^2\theta,\tag{5.41}$$

where the positivity of $\cos\theta$ comes from the restriction of $\theta\in(0,\pi/2]$. If we consider the non-negative quantities x,y, where

$$x^2:=\varepsilon_A^2\left(1-\frac{\varepsilon_A^2}{4}\right),\qquad y^2:=\varepsilon_B^2\left(1-\frac{\varepsilon_B^2}{4}\right),\tag{5.42}$$

then we see that the lower bound for Equation (5.41) describes an ellipse centred at the origin:

$$x^2+y^2+2xy\cos\theta=\sin^2\theta.\tag{5.43}$$

As we vary the level of incompatibility between A and B—done here by changing θ—we change the shape of the ellipse: As we approach $\theta=0$, Equation (5.43) reduces to $(x+y)^2=0$. If we considered the full range of possible values of x and y we would find that this defines a line running from $(-1,1)$ to $(1,-1)$, but since we are considering positive quantities the only point on the line we would consider is $(0,0)$. This corresponds to the admissible region becoming larger until any pair of approximating observables are

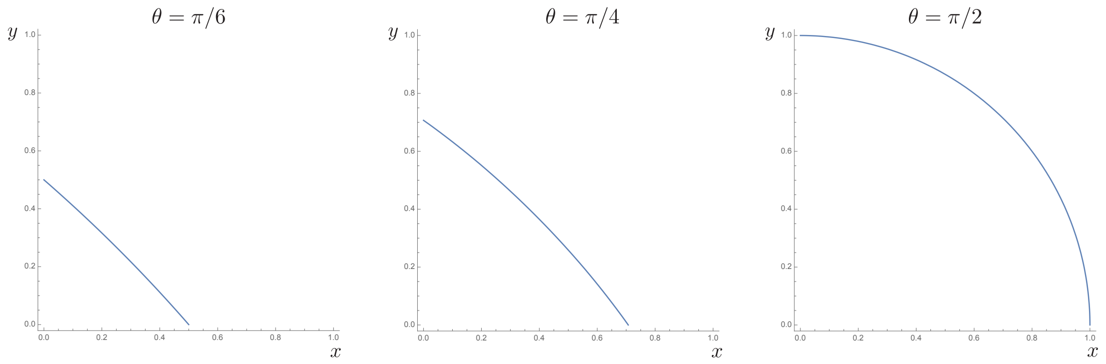


Figure 5.2: A plot of Equation (5.43) for $\theta = \pi/6, \pi/4$ and $\pi/2$.

possible. As we increase θ the length of the semi-major axis reduces, whilst the semi-minor axis increases, until we reach the angle $\theta = \pi/2$, where Equation(5.43) reduces to $x^2 + y^2 = 1$, describing a circle of radius one centred at the origin. A plot of Equation (5.43) for $\theta = \pi/6, \pi/4$ and $\pi/2$ is given in Figure 5.2.

For any given value of θ, we can quickly see from Equation (5.43) that if either x^2 or y^2 equals 0, then y^2 or x^2 becomes equal to $\sin^2\theta$, respectively. Given the forms of x^2 and y^2 in Equation (5.42), the outcome $x^2 = 0$, say, can occur in two possible ways: either $\varepsilon_A^2 = 0$ or $\varepsilon_A^2 = 4$. By expressing (5.11) in a different way:

$$\varepsilon_A^2 = \|\boldsymbol{a} - \boldsymbol{c}\|^2 + 1 - \|\boldsymbol{c}\|^2 = 2(1 - \boldsymbol{a}\cdot\boldsymbol{c}), \tag{5.44}$$

where we have made use of the normalisation of $\boldsymbol{a}$, we can express the quantity $\varepsilon_A^2(1 - \varepsilon_A^2/4)$ in a much simpler form:

$$\varepsilon_A^2\left(1 - \frac{\varepsilon_A^2}{4}\right) = 2(1 - \boldsymbol{a}\cdot\boldsymbol{c})\left(\frac{1 + \boldsymbol{a}\cdot\boldsymbol{c}}{2}\right) = 1 - (\boldsymbol{a}\cdot\boldsymbol{c})^2, \tag{5.45}$$

and similarly for $\varepsilon_B^2(1 - \varepsilon_B^2/4)$. With these in mind, we can alternatively express Equation (5.41) as

$$2 - (\boldsymbol{a}\cdot\boldsymbol{c})^2 - (\boldsymbol{b}\cdot\boldsymbol{d})^2 + 2\cos\theta\sqrt{\left(1 - (\boldsymbol{a}\cdot\boldsymbol{c})^2\right)\left(1 - (\boldsymbol{b}\cdot\boldsymbol{d})^2\right)} \geq \sin^2\theta. \tag{5.46}$$

5.3.2 Yu and Oh's bound

The full derivation of Yu and Oh's bound is given in Appendix E, so we shall only give a brief overview in this section.

The paper deals with a rescaled version of the BLW error measure given in Equation (5.16):

$$D(\mathsf{C}, \mathsf{A}) = 2\mathfrak{D}(\mathsf{C}, \mathsf{A}) = \frac{1}{2}\Delta(\mathsf{C}, \mathsf{A})^2 = \|\boldsymbol{a} - \boldsymbol{c}\|, \tag{5.47}$$

and similarly for $D(\mathsf{D}, \mathsf{B})$. It is assumed from the outset that the approximating observables possess the smallest possible values for $D(\mathsf{C}, \mathsf{A})$ and $D(\mathsf{D}, \mathsf{B})$, subject to the condition that they are jointly measurable. Since these observables are covariant, this condition is equivalent [10] to their Bloch vectors $\boldsymbol{c}$ and $\boldsymbol{d}$ satisfying the condition

$$\|\boldsymbol{c}+\boldsymbol{d}\|+\|\boldsymbol{c}-\boldsymbol{d}\|\leq 2. \tag{5.48}$$

Geometrically, if we fix the vector $\boldsymbol{d}$, say, then we are attempting to find the smallest circle of radius $D(\mathsf{C},\mathsf{A})$ centred at $\boldsymbol{a}$ that will intersect with the ellipsoid defined by Equation (5.48). By minimising $D(\mathsf{C},\mathsf{A})$ and $D(\mathsf{D},\mathsf{B})$ subject to Equation (5.48) we find a minimum error bound given by the quantities

$$\begin{aligned} D(\mathsf{C},\mathsf{A}) &= \frac{\sin\varphi+\sin\theta\cos\varphi}{\sqrt{1+\sin\theta\sin 2\varphi}}-\sin\varphi, \\ D(\mathsf{D},\mathsf{B}) &= \frac{\cos\varphi+\sin\theta\sin\varphi}{\sqrt{1+\sin\theta\sin 2\varphi}}-\cos\varphi, \end{aligned} \tag{5.49}$$

where $\sin\theta=\|\boldsymbol{a}\times\boldsymbol{b}\|$ $(=\left\|\frac{1}{2i}[A,B]\right\|)$ as in the case of Branciard's bound, and φ is defined via

$$\sin\varphi=\sqrt{\frac{1-\|\boldsymbol{d}\|^2}{1-(\boldsymbol{c}\cdot\boldsymbol{d})^2}}. \tag{5.50}$$

Unfortunately, given that the bound described in Equation (5.49) is expressed in terms of the angle φ, it is in a form that is dependent upon the Bloch representation, and not expressed in terms of quantities with immediate operational meaning. However, this can be rectified.

First, by comparing with Equation (5.18), we see that the numerator in Equation (5.50) is the square root of the unsharpness of the observable D, $U(\mathsf{D})$. Secondly, we note that, as is shown in Appendix E, the above bound is found by making the Bloch vectors $\boldsymbol{c}$ and $\boldsymbol{d}$ saturate the inequality (5.48). This condition—as is seen by squaring both sides, rearranging and then squaring again—is equivalent to

$$\|\boldsymbol{c}\|^2+\|\boldsymbol{d}\|^2=1+(\boldsymbol{c}\cdot\boldsymbol{d})^2. \tag{5.51}$$

By making use of Equation (5.51) we can rewrite the denominator of Equation (5.50) as

$$1-(\boldsymbol{c}\cdot\boldsymbol{d})^2=2-\|\boldsymbol{c}\|^2-\|\boldsymbol{d}\|^2=U(\mathsf{C})^2+U(\mathsf{D})^2. \tag{5.52}$$

Hence, we can express $\sin\varphi$ solely in terms of the unsharpness of the two approximating observables:

$$\sin\varphi=\frac{U(\mathsf{D})}{\sqrt{U(\mathsf{C})^2+U(\mathsf{D})^2}}, \tag{5.53}$$

and similarly,

$$\cos\varphi=\frac{U(\mathsf{C})}{\sqrt{U(\mathsf{C})^2+U(\mathsf{D})^2}}. \tag{5.54}$$

These quantities are positive, and so $\varphi\in[0,\pi/2]$. By combining these two together, we can also express $\sin 2\varphi$ in terms of $U(\mathsf{C})$ and $U(\mathsf{D})$:

$$\sin 2\varphi=2\sin\varphi\cos\varphi=\frac{2U(\mathsf{C})U(\mathsf{D})}{U(\mathsf{C})^2+U(\mathsf{D})^2}. \tag{5.55}$$

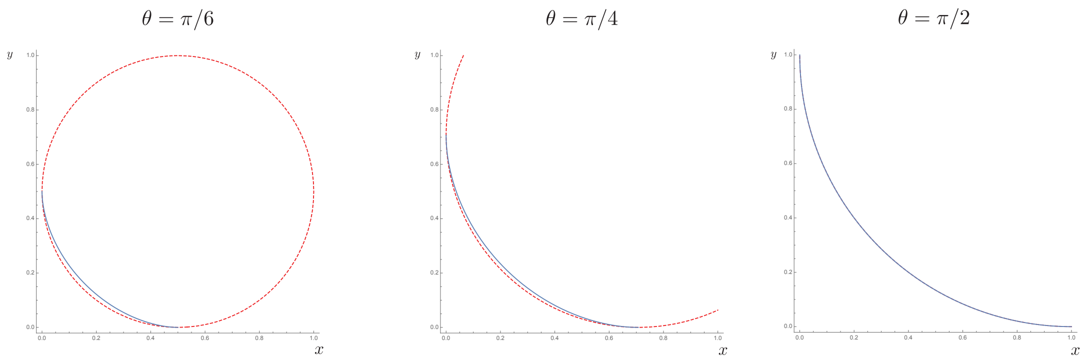


Figure 5.3: A comparison of Yu and Oh's minimum error bound (blue line) with the circle $(x - \sin\theta)^2 + (y - \sin\theta)^2 = \sin^2\theta$ (red dashed line) for the angles $\theta = \pi/6, \pi/4$ and $\pi/2$.

Using these identities in Equation (5.49) we can express the bounds in terms of the unsharpness of C and D:

$$\begin{aligned} D(\mathsf{C},\mathsf{A}) &= \frac{U(\mathsf{D}) + U(\mathsf{C})\sin\theta}{\sqrt{(U(\mathsf{D}) + U(\mathsf{C})\sin\theta)^2 + U(\mathsf{C})^2\cos^2\theta}} - \frac{U(\mathsf{D})}{\sqrt{U(\mathsf{C})^2 + U(\mathsf{D})^2}}, \\ D(\mathsf{D},\mathsf{B}) &= \frac{U(\mathsf{C}) + U(\mathsf{D})\sin\theta}{\sqrt{(U(\mathsf{C}) + U(\mathsf{D})\sin\theta)^2 + U(\mathsf{D})^2\cos^2\theta}} - \frac{U(\mathsf{C})}{\sqrt{U(\mathsf{C})^2 + U(\mathsf{D})^2}}. \end{aligned} \tag{5.56}$$

Note that for the work we present here we will interchange between the forms given in Equation (5.49) and (5.56), depending on which form is of greater use at the time.

Much like the case of Branciard's bound, the shape of the bound described by Equation (5.49) is determined by the angle θ that quantifies the incompatibility between A and B. In the case of compatible observables ($\theta = 0$), the optimal errors are both zero, as is expected, and when we consider maximally incompatible observables ($\theta = \pi/2$) we arrive at the form

$$D(\mathsf{C},\mathsf{A}) = 1 - \sin\varphi, \qquad D(\mathsf{D},\mathsf{B}) = 1 - \cos\varphi, \tag{5.57}$$

which for the domain that we are considering, $\varphi \in [0, \pi/2]$, defines the lower left quadrant of a unit circle centred at $(1, 1)$. For all other values of θ, this bound defines a curve that lies slightly above a circle of radius $\sin\theta$ centred at $(\sin\theta, \sin\theta)$. A comparison of these curves is shown in Figure 5.3 for $\theta = \pi/6, \pi/4$ and $\pi/2$.. If we let $\varphi = 0$ in Equation (5.49) we see that $D(\mathsf{C},\mathsf{A}) = \sin\theta$ and $D(\mathsf{D},\mathsf{B}) = 0$, whilst if $\varphi = \pi/2$ we find $D(\mathsf{C},\mathsf{A}) = 0$ and $D(\mathsf{D},\mathsf{B}) = \sin\theta$.

5.4 A comparison of optimal approximators

In their respective papers, Branciard and Yu and Oh provide examples of approximating observables C and D that saturate their given bounds. In this section we shall compare these optimal approximators and show that, whilst they saturate their respective bound, they do not in general saturate the other's.

We will also discuss the experimental work by Ringbauer *et al.* [45], who have constructed jointly measurable approximators of the sharp observables associated with the Pauli operators σ_x and σ_z via photon polarisation measurements. Using the noise measure, their approximators near the bound given in Equation (5.41). Their paper discusses two methods of measurement, the so-called three-state measurement and weak measurement methods. The conceptual shortcomings of the three-state method have been discussed elsewhere [18, 19], and so we restrict ourselves here to the discussion on the weak measurement method presented. As will be shown, the method used by Ringbauer *et al.* is very similar in nature to that presented by Branciard in [7], and as a result whilst their approximators approach the bound in Equation (5.41), they do not near the bound in Equation (5.49). This is in contrast with the experiment of Rozema *et al.* [46], which (as is shown in the supplemental material of [20]) does lead to the optimal bound for $D(\mathsf{C},\mathsf{A})$ and $D(\mathsf{D},\mathsf{B})$ when A and B are maximally incompatible.

5.4.1 Branciard's optimal approximators

We shall begin with the optimal approximators given by Branciard. The sharp observables that Branciard wishes to approximate are represented by Bloch unit vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ that lie within the x–y plane of the Bloch sphere, and are characterised by the angles ϕ_a and ϕ_b, respectively:

$$A := \mathsf{A}[1] = \cos\phi_a\,\sigma_x + \sin\phi_a\,\sigma_y, \qquad B := \mathsf{B}[1] = \cos\phi_b\,\sigma_x + \sin\phi_b\,\sigma_y, \tag{5.58}$$

where it is assumed that $\phi_a \leq \phi_b$. By identifying the angle $\theta := \phi_b - \phi_a$ we see that the commutator of these two operators is $[A,B] = 2i\sin\theta\,\sigma_z$, and so for these two observables C_{AB} is equal to (cf. Equation (5.35))

$$C_{AB} = \sin\theta\,\langle\sigma_z\rangle_\psi\,, \tag{5.59}$$

with ψ denoting the state of the system we are interested in. Since we are interested in saturating the bound in Equation (5.41), we require that the state ψ satisfies $\langle A\rangle_\psi = \langle B\rangle_\psi = 0$, and so ψ must be an eigenstate of σ_z. Given the eigenvalues of σ_z are ± 1, it follows that the degree of incompatibility between A and B is

$$C_{AB}^2 = \sin^2\theta. \tag{5.60}$$

The observables A and B are approximated via an indirect measurement scheme: The system of interest is coupled to an ancillary qubit system described by the Hilbert space $\mathcal{K} = \mathbb{C}^2$ and prepared in a state $\xi \in \mathcal{K}$ via a unitary transformation U, after which the ancillary system is measured via a sharp dichotomic observable M, thereby approximating A, and subsequently we directly measure B. The self-adjoint operator $M := \mathsf{M}[1]$ associated with the observable M is of the form

$$M = \cos\varphi\,\sigma_x + \sin\varphi\,\sigma_y = \boldsymbol{m}\cdot\boldsymbol{\sigma}, \tag{5.61}$$

where $\varphi \in [\phi_a, \phi_b]$, whilst the coupling unitary U is of the form

$$U = (U_R \otimes I)U_{\text{copy}}. \tag{5.62}$$

The operator U_R is a rotation operator around the z axis,

$$U_R = \exp\left[-i\frac{\phi_b - \varphi}{2}\sigma_z\right] = \cos\left(\frac{\phi_b - \varphi}{2}\right) I - i\sin\left(\frac{\phi_b - \varphi}{2}\right)\sigma_z, \tag{5.63}$$

whilst U_{copy} maps the state of the ancillary system to one of the eigenvectors of M,

$$U_{\text{copy}}(\mathsf{M}(\pm)\otimes P_\xi)U_{\text{copy}}^* = \mathsf{M}(\pm)\otimes\mathsf{M}(\pm). \tag{5.64}$$

The joint observable for this measurement scheme is given by $\mathsf{J} : \{\pm 1\}\times\{\pm 1\}\to\mathcal{E}(\mathbb{C}^2)$ via

$$\operatorname{tr}\left[P_\psi \mathsf{J}(k,\ell)\right] = \operatorname{tr}\left[U(P_{\psi\otimes\xi})U^*\mathsf{B}(k)\otimes\mathsf{M}(\ell)\right], \tag{5.65}$$

where $k, \ell = \pm 1$. In other words, the observable J is given by

$$\mathsf{J}(k,\ell) = \operatorname{tr}_{\mathcal{K}}\left[U_{\text{copy}}(I\otimes P_\xi)U_{\text{copy}}^*(U_R^*\,\mathsf{B}(k)\,U_R)\otimes\mathsf{M}(\ell)\right]. \tag{5.66}$$

By decomposing the identity into a sum of the effects of M, we see that

$$U_{\text{copy}}(I\otimes P_\xi)U_{\text{copy}}^* = \mathsf{M}(+)\otimes\mathsf{M}(+) + \mathsf{M}(-)\otimes\mathsf{M}(-), \tag{5.67}$$

whilst the second quantity contained within the trace, $U_R^*\mathsf{B}(k)U_R\otimes\mathsf{M}(\ell)$, is resolved by first noting that

$$\begin{aligned}
U_R^* B U_R &= \left(\cos\left(\frac{\phi_b - \varphi}{2}\right) I + i\sin\left(\frac{\phi_b - \varphi}{2}\right)\sigma_z\right)(\cos\phi_b\,\sigma_x + \sin\phi_b\,\sigma_y)\\
&\quad\times\left(\cos\left(\frac{\phi_b - \varphi}{2}\right) I - i\sin\left(\frac{\phi_b - \varphi}{2}\right)\sigma_z\right)\\
&= \left(\cos^2\left(\frac{\phi_b - \varphi}{2}\right) - \sin^2\left(\frac{\phi_b - \varphi}{2}\right)\right)(\cos\phi_b\,\sigma_x + \sin\phi_b\,\sigma_y)\\
&\quad + 2\sin\left(\frac{\phi_b - \varphi}{2}\right)\cos\left(\frac{\phi_b - \varphi}{2}\right)(\sin\phi_b\,\sigma_x - \cos\phi_b\,\sigma_y)\\
&= (\cos(\phi_b - \varphi)\cos\phi_b + \sin(\phi_b - \varphi)\sin\phi_b)\sigma_x\\
&\quad + (\cos(\phi_b - \varphi)\sin\phi_b - \sin(\phi_b - \varphi)\cos\phi_b)\sigma_y\\
&= \cos\varphi\,\sigma_x + \sin\varphi\,\sigma_y = M,
\end{aligned} \tag{5.68}$$

and so we can immediately see that $U_R^*\mathsf{B}(k)U_R = \mathsf{M}(k)$. By combining these results we can find the joint observable that is measured:

$$\begin{aligned}
\mathsf{J}(k,\ell) &= \operatorname{tr}_{\mathcal{K}}\left[\big(\mathsf{M}(+)\otimes\mathsf{M}(+) + \mathsf{M}(-)\otimes\mathsf{M}(-)\big)(\mathsf{M}(k)\otimes\mathsf{M}(\ell))\right]\\
&= \mathsf{M}(k)\mathsf{M}(\ell).
\end{aligned} \tag{5.69}$$

Clearly, this is a trivial example of a joint observable where both margins are the same observable; that is, $\mathsf{C} = \mathsf{D} = \mathsf{M}$. Conceptually, what is being measured here is a single sharp observable as an approximator for both A and B, with its respective Bloch vector $\boldsymbol{m}$ lying somewhere in between $\boldsymbol{a}$ and $\boldsymbol{b}$ on the circumference of the circle in the plane spanned by $\boldsymbol{a}$ and $\boldsymbol{b}$. In the language of Branciard and Hall's joint measurement scheme,

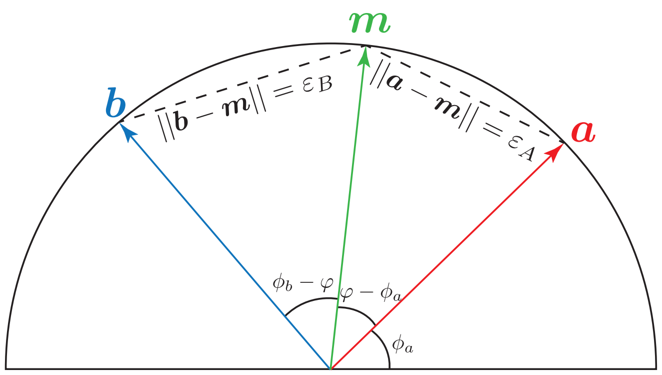

$\boldsymbol{m}$
$\boldsymbol{b}$
$\boldsymbol{a}$
$\|\boldsymbol{b}-\boldsymbol{m}\| = \varepsilon_B$
$\|\boldsymbol{a}-\boldsymbol{m}\| = \varepsilon_A$
$\phi_b - \varphi$
$\varphi - \phi_a$
ϕ_a

Figure 5.4: An example of Branciard's optimal approximating observable M, characterised by the angle $\varphi \in [\phi_a, \phi_b]$. The Bloch vector $\boldsymbol{m}$ characterising M is normalised, and so $\varepsilon_A^2 = \|\boldsymbol{a}-\boldsymbol{m}\|$, etc. Any normalised Bloch vector lying between $\boldsymbol{a}$ and $\boldsymbol{b}$ provides an optimal approximating observable with respect to the noise measure.

this is done by setting $f(m) = g(m) = m$ for all m within the spectrum of M. The value of the noise measures ε_A^2 and ε_B^2 are found via Equation (5.11):

$$\begin{aligned} \varepsilon_A^2 &= \|\boldsymbol{a}-\boldsymbol{m}\|^2 = 2(1-\boldsymbol{a}\cdot\boldsymbol{m}) = 2(1-\cos(\varphi-\phi_a)) = 4\sin^2\left(\frac{\varphi-\phi_a}{2}\right), \\ \varepsilon_B^2 &= \|\boldsymbol{b}-\boldsymbol{m}\|^2 = 4\sin^2\left(\frac{\phi_b-\varphi}{2}\right). \end{aligned} \tag{5.70}$$

From this the expression $\varepsilon_A^2(1-\varepsilon_A^2/4)$ can be readily found,

$$\begin{aligned} \varepsilon_A^2\left(1-\frac{\varepsilon_A^2}{4}\right) &= 4\sin^2\left(\frac{\varphi-\phi_a}{2}\right)\left(1-\sin^2\left(\frac{\varphi-\phi_a}{2}\right)\right) \\ &= 4\sin^2\left(\frac{\varphi-\phi_a}{2}\right)\cos^2\left(\frac{\varphi-\phi_a}{2}\right) \\ &= \sin^2(\varphi-\phi_a), \end{aligned} \tag{5.71}$$

and similarly for $\varepsilon_B^2(1-\varepsilon_B^2/4)$. Making use of the value of C_{AB}^2 given in Equation (5.60), the left hand side of Equation (5.41), with the above values $\varepsilon_A^2(1-\varepsilon_A^2/4)$ and $\varepsilon_B^2(1-\varepsilon_B^2/4)$, is now equal to

$$\sin^2(\varphi-\phi_a) + \sin^2(\phi_b-\varphi) + 2\cos\theta\sin(\varphi-\phi_a)\sin(\phi_b-\varphi). \tag{5.72}$$

At this point we make use of the identity $\phi_b = \phi_a + \theta$, from which it follows that

$$\sin^2(\varphi-\phi_a) + \sin^2(\phi_b-\varphi) + 2\cos\theta\sin(\varphi-\phi_a)\sin(\phi_b-\varphi) = \sin^2\theta. \tag{5.73}$$

In other words, for any value of ϕ_a and ϕ_b, where $\phi_a \leq \phi_b$, and any $\varphi \in [\phi_a, \phi_b]$, the measurement scheme outlined here produces approximating observables that saturate the bound given in Equation (5.41), i.e., this bound is saturated by any sharp observable whose Bloch unit vector lies in between $\boldsymbol{a}$ and $\boldsymbol{b}$ in the plane spanned by them (see Figure 5.4). Furthermore, suppose that we consider the unit vector $\boldsymbol{m}$, and then define the vectors $\boldsymbol{c} = (\boldsymbol{a}\cdot\boldsymbol{m})\boldsymbol{a}$ and $\boldsymbol{\lambda} = \lambda\,\boldsymbol{c} + (1-\lambda)\boldsymbol{m}$, where $\lambda \in [0, 2]$ so that $\|\boldsymbol{\lambda}\| \leq 1$. In this case

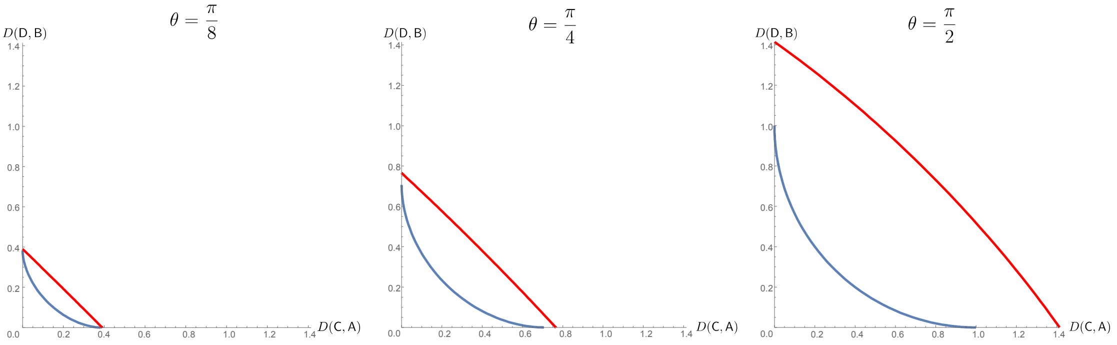


Figure 5.5: A comparison of the D values for the approximating observables that arise from Branciard's measurement scheme (red curve) against the optimal curve given in Equation (5.49) (blue curve) for several values of θ, which determines the level of incompatibility of the observables A and B. In the case of small θ the limits of these curves are close to coinciding, but for all other points—corresponding to values of the angle φ not equal to ϕ_a or ϕ_b—there is a notable disparity between Branciard's approximators and the optimal case, and indeed for larger values of θ even the limit cases vary greatly from the optimal situation.

any dichotomic observable approximating A described by the Bloch vector $\boldsymbol{\lambda}$ satisfies (cf. Equation (5.44))

$$\varepsilon_A^2 = 2(1 - \boldsymbol{a}\cdot\boldsymbol{\lambda}) = 2(1 - \boldsymbol{a}\cdot\boldsymbol{m}). \tag{5.74}$$

In other words, for any vector on the line segment between $\boldsymbol{m}$ and $\boldsymbol{c}$, we find that the noise error is fixed, although the level to which these vectors approximate A varies. Indeed, in the case of maximum incompatibility, when $\phi_b - \phi_a = \pi/2$, if we let $\varphi = \phi_a + \pi/4$, then any vector on the line segment between $\boldsymbol{m}$ and $\boldsymbol{c} = \boldsymbol{a}/\sqrt{2}$ will have the same noise error. However, as we shall discuss in Section 5.4.3, in the case of the D and BLW error measures, only the Bloch vector $\boldsymbol{c}$ corresponds to an optimal approximating observable in the case of maximal incompatibility.

If we now use the vector-norm measure D on this approximation scheme, we find that it is suboptimal. Indeed, using Equations (5.47) and (5.70), we see that

$$\begin{aligned} D(\mathsf{C},\mathsf{A}) &= \|\boldsymbol{a}-\boldsymbol{m}\| = 2\sin\left(\frac{\varphi-\phi_a}{2}\right), \\ D(\mathsf{D},\mathsf{B}) &= 2\sin\left(\frac{\phi_b-\varphi}{2}\right) = 2\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\varphi-\phi_a}{2}\right) - 2\cos\left(\frac{\theta}{2}\right)\sin\left(\frac{\varphi-\phi_a}{2}\right). \end{aligned} \tag{5.75}$$

At the limit $\varphi = \phi_a$ we see that $D(\mathsf{C},\mathsf{A}) = 0$ and $D(\mathsf{D},\mathsf{B}) = 2\sin(\theta/2)$, and vice versa when $\varphi = \phi_b$. By comparison, the values at the limit for the optimal case, as shown in Section 5.3.2 are $(0, \sin\theta)$ and $(\sin\theta, 0)$. Since $\sin\theta = 2\sin(\theta/2)\cos(\theta/2) \leq 2\sin(\theta/2)$ for $\theta \in [0,\pi/2]$, it follows that the limiting values given by Branciard's scheme are greater than the optimal values for all cases except $\theta = 0$. Indeed, by plotting the curve given by Equation (5.75) against the optimal bound given by Equation (5.49) (see Figure 5.5) we see that, when using the vector-norm measure D as the measure of error, Branciard's scheme for approximating A and B is clearly suboptimal.

5.4.2 Ringbauer *et al.*'s experiment

Following on from the work presented by Branciard, Ringbauer *et al.* [45] performed an experimental verification of a measurement scheme which nears the bound (5.41). Based on the discussion presented in Branciard's paper, they work towards this bound by means of two separate methods: the three-state method and the weak measurement approach. As stated above, we shall not dwell on the three-state method[1] and instead focus on the weak measurement scheme.

The reason for using this method stems from expanding the general form of the noise measure $\varepsilon(\mathsf{C},\mathsf{A},\psi)^2$:

$$\begin{aligned}\varepsilon(\mathsf{C},\mathsf{A},\psi)^2 &= \big\langle\psi\big|(\mathsf{C}[1]-\mathsf{A}[1])^2\psi\big\rangle + \big\langle\psi\big|(\mathsf{C}[2]-\mathsf{C}[1]^2)\psi\big\rangle \\ &= \big\langle\psi\big|(\mathsf{A}[1]^2+\mathsf{C}[2])\psi\big\rangle - \langle\psi|(\mathsf{C}[1]\mathsf{A}[1]+\mathsf{A}[1]\mathsf{C}[1])\psi\rangle \\ &= \big\langle\psi\big|(\mathsf{A}[1]^2+\mathsf{C}[2])\psi\big\rangle - 2\mathrm{Re}\,\langle\psi|\mathsf{C}[1]\mathsf{A}[1]\psi\rangle\,.\end{aligned} \tag{5.76}$$

In our particular case, both $\mathsf{A}[1]^2$ and $\mathsf{C}[2]$ are equal to the identity, and so $\varepsilon(\mathsf{C},\mathsf{A},\psi)^2 = 2(1-\mathrm{Re}\,\langle\psi|\mathsf{C}[1]\mathsf{A}[1]\psi\rangle)$. In most instances it is assumed that the first moment operators $\mathsf{A}[1]$ and $\mathsf{C}[1]$ do not commute, and so strictly taking the real part of the expectation value is a necessity here. Following the work of Lund and Wiseman [37], Ringbauer *et al.* identify the quantity $\mathrm{Re}\,\langle\psi|\mathsf{C}[1]\mathsf{A}[1]\psi\rangle$ as the expectation value of a joint quasi-probability distribution $p^w_{\mathsf{A}}(k,\ell) = \mathrm{Re}\,\langle\psi|\mathsf{C}(k)\mathsf{A}(\ell)\psi\rangle$. However, whilst the setup of Lund and Wiseman was designed to measure both error and disturbance within a sequential measurement scheme—utilising two probe systems in the process—Ringbauer *et al.* consider a scheme using a single probe and instead simply calculate the error in approximating either A or B.

In their experimental setup, the incompatible observables being approximated are $\mathsf{A} = \mathsf{X}$ and $\mathsf{B} = \mathsf{Z}$, the PVMs associated with the Pauli operators σ_x and σ_z, respectively, and the observable being used to approximate them is the sharp observable M with associated self-adjoint operator $M = \cos\varphi\,\sigma_z + \sin\varphi\,\sigma_x = \boldsymbol{m}\cdot\boldsymbol{\sigma}$. Within the context of the paper, a weak measurement of A is performed, followed by a measurement of M, thereby allowing them to calculate the quasi-probability distribution p^w_{A} and hence the value of the noise measure for approximating A via M (and similarly in the case of B).

However, this method is not necessary for our analysis. The weak measurement is solely there to aid in the calculation of the value of the noise measure from the measurement statistics, and does not alter the statistics of the observable M that is being used to approximate A and B. The observable M is of the same form as that used by Branciard in his paper [7], and thus follows the same analysis as presented in the preceding section. Indeed, this experiment is centred on a much simpler version of the case given above, and

[1] The basis of the argument against the three-state method is that, in general, the quantity $\varepsilon(\mathsf{C},\mathsf{A},\psi)^2$ is designed to not only be state-dependent, but is also based on a value comparison of measurement outcomes. Given the probabilistic nature of measurements in quantum mechanics, a direct comparison of the measurement outcomes of three systems prepared in different states is not possible. As a result, the three-state method does not correspond to a "direct test" as given in [20].

the value of the noise measure for M approximating A and B is equal to

$$\begin{aligned}\varepsilon(\mathsf{M},\mathsf{A},\psi)^2 &= \|\boldsymbol{a}-\boldsymbol{m}\|^2 = 2(1-\sin\varphi) = 4\sin^2\left(\frac{\pi/2-\varphi}{2}\right),\\ \varepsilon(\mathsf{M},\mathsf{B},\psi)^2 &= \|\boldsymbol{b}-\boldsymbol{m}\|^2 = 2(1-\cos\varphi) = 4\sin^2\left(\frac{\varphi}{2}\right).\end{aligned} \tag{5.77}$$

Given that these are just a special case of Equation (5.70), where $\phi_a = \pi/2$ and $\phi_b = 0$, we can immediately infer that this scheme saturates the bound in Equation (5.41). Similarly, we immediately see that

$$\begin{aligned}D(\mathsf{M},\mathsf{A}) &= 2\sin\left(\frac{\pi/2-\varphi}{2}\right),\\ D(\mathsf{M},\mathsf{B}) &= 2\sin\left(\frac{\varphi}{2}\right).\end{aligned} \tag{5.78}$$

Since we are considering a case where A and B are maximally incompatible, we must have $\sin\theta = \|\boldsymbol{a}\times\boldsymbol{b}\| = 1$ and so the optimal bound for D is given by Equation (5.57),

$$\begin{aligned}D(\mathsf{C},\mathsf{A}) &= 1-\sin\varphi,\\ D(\mathsf{D},\mathsf{B}) &= 1-\cos\varphi.\end{aligned} \tag{5.57}$$

In the two limiting cases ($\varphi = 0$ and $\varphi = \pi/2$) we see that $D(\mathsf{M},\mathsf{A}) > D(\mathsf{C},\mathsf{A})$ and $D(\mathsf{M},\mathsf{B}) > D(\mathsf{D},\mathsf{B})$, and indeed in general the values of $D(\mathsf{M},\mathsf{A})$ and $D(\mathsf{M},\mathsf{B})$ follow the pattern shown in the rightmost plot in Figure 5.5.

5.4.3 Yu and Oh's optimal approximators

The optimal approximators given by Yu and Oh arise naturally within the derivation of their optimal bound (see Appendix E), and so we shall simply state them in this section.

The joint observable that provides the optimal approximating observables as margins is of the form

$$\mathsf{J}(k,\ell) = \frac{(1+k\ell M)I + (k\boldsymbol{c}+\ell\boldsymbol{d})\cdot\boldsymbol{\sigma}}{4}, \tag{5.79}$$

with $k,\ell = \pm 1$ and

$$M = \frac{\cos\theta}{\sqrt{1+\sin\theta\sin 2\varphi}}. \tag{5.80}$$

The marginal observables C and D are hence of the form

$$\begin{aligned}\mathsf{C}(\pm) &= \frac{I\pm\boldsymbol{c}\cdot\boldsymbol{\sigma}}{2},\\ \mathsf{D}(\pm) &= \frac{I\pm\boldsymbol{d}\cdot\boldsymbol{\sigma}}{2},\end{aligned} \tag{5.81}$$

where the Bloch vectors are of the form

$$\begin{aligned}\boldsymbol{c} &= \frac{(D(\mathsf{D},\mathsf{B}) + (1-M^2)\cos\varphi)\sin\varphi\,\boldsymbol{a} + MD(\mathsf{C},\mathsf{A})\cos\varphi\,\boldsymbol{b}}{\sin\theta},\\ \boldsymbol{d} &= \frac{(D(\mathsf{C},\mathsf{A}) + (1-M^2)\sin\varphi)\cos\varphi\,\boldsymbol{b} + MD(\mathsf{D},\mathsf{B})\sin\varphi\,\boldsymbol{a}}{\sin\theta}.\end{aligned} \tag{5.82}$$

Due to their reliance on the angle φ it is not very clear what the form of these vectors

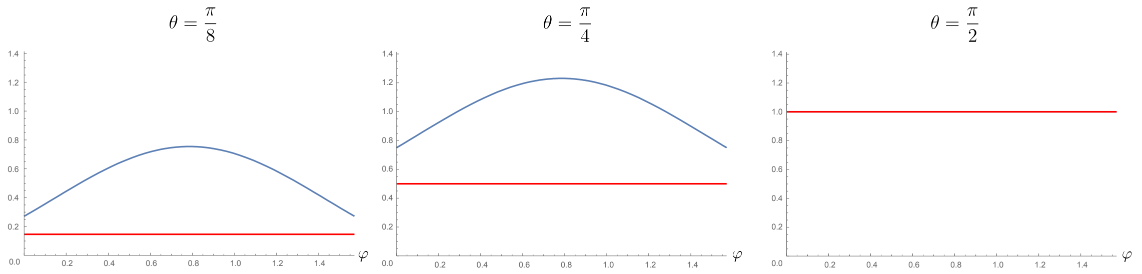


Figure 5.6: A plot of the left-hand side of Equation (5.46) in the case of Yu and Oh's optimal approximating observables for various values of $\varphi \in [0, \pi/2]$, shown by a blue curve, compared against the lower bound ($\sin^2\theta$), which is the red line. With the exception of $\theta = \pi/2$ the blue curve is always a greater value than the red line, highlighting that Yu and Oh's optimal approximators are suboptimal in terms of the noise measure (excluding the maximally incompatible case).

is, but it is instructive to see that in the case $\theta = \pi/2$, where A and B are maximally incompatible and the quantity M is equal to zero, $\boldsymbol{c}$ and $\boldsymbol{d}$ reduce to

$$\begin{aligned} \boldsymbol{c} &= (D(\mathsf{D},\mathsf{B}) + \cos\varphi)\sin\varphi\,\boldsymbol{a} = \sin\varphi\,\boldsymbol{a}, \\ \boldsymbol{d} &= (D(\mathsf{C},\mathsf{A}) + \sin\varphi)\cos\varphi\,\boldsymbol{b} = \cos\varphi\,\boldsymbol{b}, \end{aligned} \tag{5.83}$$

where we have made use of Equation (5.57). In other words, the observables C and D are smeared versions of the sharp observables A and B, respectively, which coincides with what we would expect for maximally incompatible observables when we use the BLW error measure [18]. Furthermore, we see that $\varepsilon_A^2 = 2(1 - \boldsymbol{a}\cdot\boldsymbol{c}) = 2(1 - \sin\varphi)$ and $\varepsilon_B^2 = 2(1 - \cos\varphi)$ for $\theta = \pi/2$, and so by making use of these values in the left hand side of Equation (5.41), where now $C_{AB}^2 = \sin^2\theta = 1$, we see immediately that this case optimises the bound given by Branciard.

However, these approximators only saturate the bound in the case $\theta = \pi/2$. If we plot the left-hand side of Equation (5.46), using the Bloch vectors in Equation (5.82), against the right-hand side for different values of θ—see Figure 5.6—then we see that these observables are suboptimal with the exception of $\theta = \pi/2$.

5.5 Different measures lead to different optimisers

As we have shown in the preceding section, the particular choice of measure we use to quantify error leads to a different set of optimising observables. From a purely mathematical perspective this should not come as a surprise: we are trying to optimise over two different functionals, and so we would not in general expect to arrive at the same optimal observables. However, given that both measures are intended to lead us to an optimal joint measurement for two incompatible observables, we must concede that they are not equally good at this task. This brings us to the question: Which of these measures (and therefore bounds) is more reliable for leading us to good approximating observables?

Whilst the bound given by Branciard is certainly well-defined, the noise measure is misleading. As we have shown, in the case of maximally incompatible observables both the optimal approximators for Yu and Oh and the PVMs whose normalised Bloch vectors lie between $\boldsymbol{a}$ and $\boldsymbol{b}$ have noise values that saturate the bound of Equation (5.41). Indeed, as was shown in Section 5.4.1, for any convex combination of one such normalised Bloch vector $\boldsymbol{m}$ and the vector $(\boldsymbol{m}\cdot\boldsymbol{a})\boldsymbol{a}$ the noise measure is fixed at $2(1-\boldsymbol{m}\cdot\boldsymbol{a})$ when treating it as an approximate measurement of A. However, in no way are all observables within this continuous range equally good approximations of A, and by comparison if we consider the BLW or D measure we see that the POVM characterised by $(\boldsymbol{m}\cdot\boldsymbol{a})\boldsymbol{a}$ is a better approximation than the PVM described by the vector $\boldsymbol{m}$. A second issue arising from this scheme is Hall and Branciard's method of joint measurement, which is little more than the relabelling of scales in order to best suit the noise measure. However, as highlighted in Appendix D, we can fix an appropriately poor approximation of an observable to have a zero noise measure by this method, despite having very different characteristics to the observable it is approximating. That such a construction is possible makes this a questionable method of approximating.

Despite its (initial) lack of operational meaning, we believe that the optimal bound presented by Yu and Oh is a more significant lower bound for approximating observables. This is because it is based upon a measure that does not fall prey to the shortcomings presented by the noise measure, in particular with regards to the number of observables that optimise it, as expressed above. Furthermore, unlike the noise measure, the measure D (and similarly the BLW error measure) does not require a direct value comparison, so joint measurability of the sharp observables and the observables that are approximating them is not needed. Such a requirement is highly restrictive, and in the qubit case means that the approximating observables must be smeared versions of the sharp observables they are approximating, which is only optimal when A and B are maximally incompatible. That methods such as the three-state and weak measurement schemes have been used as means of circumventing this problem highlights the limitations that joint measurability imposes on the experimental testing of the noise measure. By comparison, for the D or BLW error measure, only distributions are compared, and as a result the approximators do not need to be jointly measurable with the sharp observables, which allows for a much greater class of approximating observables to be considered.

Chapter 6

Summary

We shall now briefly summarise the main results presented in the thesis. In the first part, motivated by the qubit case, we discussed the connection between SIC-POVMs and MUBs. For a given SIC-POVM we are able to construct d-partitions, each forming a margin observable of the SIC-POVM. If we restrict the d-partitions to satisfy the one-overlap property, i.e., any two bins from different partitions share just one element in common, then these margin POVMs are mutually unbiased. A collection of d-partitions satisfying the one-overlap property with respect to each other and the so-called Cartesian partitions was shown to be equivalent to a collection of mutually orthogonal Latin squares. As a result, it was shown that the maximum number of such d-partitions was equal to two plus the maximum number of mutually orthogonal Latin squares of order d. For any two mutually unbiased POVMs, if they were also commutative, then the common eigenbases were also mutually unbiased, in which case MUBs were found from the margins of a SIC-POVM. In Mathematica, complete sets of MUBs were derived from SIC-POVMs for dimensions 3 and 5, but in dimension 4 the only two Latin squares that were found to produce commutative POVMs were not mutually orthogonal. Meanwhile, in dimension 7 Mathematica struggled to produce eigenbases when forced to deal with effects possessing degenerate eigenvalues.

Starting then from a complete set of MUBs, we showed that by smearing in order to create mutually unbiased POVMs, placing the effects in a $d+1 \times d$ array and then forming downward paths of length $d+1$ satisfying a one-overlap property, we could reconstruct d^2 operators that satisfied the trace properties required of SIC-POVM effects and would also sum to the identity. In other words, we were capable of constructing a SIC system. The only issue that needed careful consideration in order to find a SIC-POVM was ensuring positivity, which depended on the ordering of the spectra whilst smearing and on the paths taken through the $d + 1 \times d$ array. The requirement of a complete set of MUBs meant that we could only currently perform this in prime-power dimensions, as complete sets of MUBs in other dimensions are yet to have been found. This construction was performed in Mathematica and SIC-POVMs were constructed in dimensions 3 and 5, where it was found that of all possible paths through the array, the only ones corresponding to positive operators were indeed SIC-POVM effects, highlighting how scarce these positive results are.

In the second part of the thesis we calculated the observable associated with the

Arthurs-Kelly measurement model in the case of arbitrary (possibly correlated) probe preparation and showed it to be covariant under phase space translations. Its margins are approximate position and momentum observables that are known to satisfy a Heisenberg-like error-disturbance relationship, in conflict with Di Lorenzo's claims to the contrary. His definition of disturbance is shown to in fact be a measure of relative statistical spread between the marginal observables and the observables measured on the individual probes, and it is shown that by allowing for correlations between the probes one can make this value negative, in which case the joint measurement is a focussing of the individual probe measurements. This is a result which is not possible in the case where the probes are prepared in a separable pure state. Following on from this, we showed two examples, one based on the probes being prepared in an entangled state and the other in a mixed state, where focussing could possibly occur and provide conditions required for this to happen. The mixed state example answers the question asked by Di Lorenzo in his Letter of whether classical correlations would be sufficient to demonstrate what we have presented here as focussing, thereby extending his discussion.

In the final part of the thesis we considered the error-error relations for incompatible observables when approximated by jointly measurable observables in the case of dichotomic qubit observables. We highlighted the flaws in the scheme considered separately by Hall and Branciard, who claim that any discrete observable can act as a joint of any two discrete observables, where one simply post-processes the spectrum of the joint observable for each margin, and use this post-processing to find optimal approximations with respect to the noise measure. This post-processing is simply a relabelling of the gauges used whilst measuring, and so nothing different is actually being measured. Further to this, we presented an example of a suboptimal approximating observable that could be made to give a zero noise measure value via their optimising method. We discussed the optimal bound found by Branciard for the noise measure, and the optimal bound given by Yu and Oh for their rescaled version of the BLW error. Given that Yu and Oh's bound is expressed in terms of trigonometric functions defined on the Bloch sphere, we rewrote them in terms of operational quantities, namely the unsharpness of the approximating observables. Following this, we presented optimal approximating observables for both bounds. In the case of the Branciard bound, we found that any normalised Bloch vector that lay between the two vectors describing the incompatible observables corresponds to an optimal approximating observable for the two, although in the case of the BLW measure this corresponds to a suboptimal approximation scheme. This result was also extended to the experimental work of Ringbauer *et al.*, who provided an example of a setup that neared Branciard's optimal bound. We also showed that the optimal approximating observables for Yu and Oh's bound, with the exception of the maximally incompatible case, are suboptimal in terms of Branciard's bound. Given that we are considering two bounds that optimise different functionals, this should perhaps not come as a surprise, but given their mutual goal of finding optimal approximators for incompatible observables, the discrepancy highlights that these are not equally good measures, with the exception of the maximally incompatible case.

Appendices

A The measured effective observable

A.1 Derivation of $\mathsf{G}^{(\lambda,\mu)}$

We begin with the probes prepared in the arbitrary state $\varphi \in \mathcal{H}_1 \otimes \mathcal{H}_2$ that, after performing the coupling unitary U_{int} with the state of the measured system, give the state Ψ with the position representation given by Equation (4.13):

$$\Psi(q, q_1, q_2) = \psi(q + \mu\, q_2)\varphi\left(q_1 - \lambda\left(q + \frac{\mu}{2}(\kappa + 1)q_2\right), q_2\right).$$

With this state, we find the effective observable $\mathsf{G}^{(\lambda,\mu)}$ via Equation (4.12):

$$\mathrm{tr}\left[P_\psi \mathsf{G}^{(\lambda,\mu)}(X \times Y)\right] = \mathrm{tr}\left[P_\Psi\big(I \otimes \mathsf{E}^{Q_1}(\lambda X) \otimes \mathsf{E}^{P_2}(\mu Y)\big)\right], \tag{A.1}$$

where we apply a Fourier transform on the final probe state

$$\widetilde{\Psi}(q, q_1, p_2) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} dq_2 e^{-iq_2p_2}\Psi(q, q_1, q_2).$$

Expanding Equation (4.12):

$$\begin{aligned}\mathrm{tr}\left[\mathsf{G}^{(\lambda,\mu)}(X \times Y)P_\psi\right] =& \frac{1}{2\pi}\int_{\mathbb{R}^8} dq\, dq'\, dq_1\, dq_1'\, dq_2\, dq_2'\, dp_2\, dp_2'\, e^{i(p_2'q_2' - p_2q_2)} \\ & \times \overline{\psi(q' + \mu q_2')\varphi(q_1' - \lambda q' - \tfrac{\lambda\mu}{2}(\kappa+1)q_2', q_2')} \\ & \times \psi(q + \mu q_2)\varphi(q_1 - \lambda q - \tfrac{\lambda\mu}{2}(\kappa+1)q_2, q_2) \\ & \times \langle q'|q\rangle \langle q_1'|\mathsf{E}^{Q_1}(\lambda X)q_1\rangle \langle p_2'|\mathsf{E}^{P_2}(\mu Y)p_2\rangle .\end{aligned} \tag{A.2}$$

After expressing $\mathsf{E}^{Q_1}(\lambda X)$ and $\mathsf{E}^{P_2}(\mu Y)$ in terms of pseudo-eigenvectors of Q_1 and P_2, respectively

$$\mathsf{E}^{Q_1}(\lambda X) = \int_{\lambda X} dq_1\ |q_1\rangle\langle q_1| = \lambda \int_X dq_1\ |\lambda q_1\rangle\langle\lambda q_1|, \tag{A.3a}$$

$$\mathsf{E}^{P_2}(\mu Y) = \int_{\mu Y} dp_2\ |p_2\rangle\langle p_2| = \mu \int_Y dp_2\ |\mu p_2\rangle\langle\mu p_2|, \tag{A.3b}$$

the right hand side of (A.2) reduces to

$$
\begin{aligned}
\operatorname{tr}\left[\mathsf{G}^{(\lambda,\mu)}(X\times Y)P_\psi\right] =& \frac{\lambda\mu}{2\pi}\int_{X\times Y} dq_1\,dp_2 \int_{\mathbb{R}^4} dq\,dq'\,dq_2\,dq_2'\,e^{-i\mu p_2(q_2-q_2')} \\
&\times \overline{\psi(q'+\mu q_2')\varphi(\lambda(q_1-q'-\tfrac{\mu}{2}(\kappa+1)q_2'),q_2')} \\
&\times \psi(q+\mu q_2)\varphi(\lambda(q_1-q-\tfrac{\mu}{2}(\kappa+1)q_2),q_2)\,\langle q'|q\rangle \\
=& \int_{X\times Y} dq_1\,dp_2 \\
&\times \left(\sqrt{\frac{\lambda\mu}{2\pi}}\int_{\mathbb{R}^2} dq\,dq_2\,e^{-i\mu p_2 q_2}\psi(q+\mu q_2)\varphi(\lambda(q_1-q-\tfrac{\mu}{2}(\kappa+1)q_2),q_2)\,|q\rangle\right)^* \\
&\times \sqrt{\frac{\lambda\mu}{2\pi}}\int_{\mathbb{R}^2} dq\,dq_2\,e^{-i\mu p_2 q_2}\psi(q+\mu q_2)\varphi(\lambda(q_1-q-\tfrac{\mu}{2}(\kappa+1)q_2),q_2)\,|q\rangle\,.
\end{aligned}
\tag{A.4}
$$

We define $q' = q + \mu\, q_2$, so $q_2 = \frac{1}{\mu}(q'-q)$, $dq_2 = \frac{1}{\mu}dq'$ and $q + \frac{\mu}{2}(\kappa+1)q_2 = \frac{1}{2}\big((1-\kappa)q + (1+\kappa)q'\big)$. Therefore

$$
\begin{aligned}
\operatorname{tr}\left[\mathsf{G}^{(\lambda,\mu)}(X\times Y)P_\psi\right] =& \int_{X\times Y} dq_1\,dp_2 \\
&\times \Bigg(\sqrt{\frac{\lambda}{2\pi\mu}}\int_{\mathbb{R}^2} dq\,dq'\,e^{ip_2(q-q')}\psi(q') \\
&\times \varphi\big(\lambda\big(q_1-\tfrac{1}{2}\big((1-\kappa)q+(1+\kappa)q'\big)\big),\tfrac{1}{\mu}(q'-q)\big)\,|q\rangle\Bigg)^* \\
&\times \sqrt{\frac{\lambda}{2\pi\mu}}\int_{\mathbb{R}^2} dq\,dq'\,e^{ip_2(q-q')}\psi(q') \\
&\times \varphi\big(\lambda\big(q_1-\tfrac{1}{2}\big((1-\kappa)q+(1+\kappa)q'\big)\big),\tfrac{1}{\mu}(q'-q)\big)\,|q\rangle \\
=& \int_{X\times Y} dq_1\,dp_2\,\left(\int_{\mathbb{R}^2} dq\,dq'\,K_{q_1p_2}(q,q')\psi(q')\,|q\rangle\right)^* \\
&\times \int_{\mathbb{R}^2} dq\,dq'\,K_{q_1p_2}(q,q')\psi(q')\,|q\rangle \\
=& \left\langle \psi \middle| \left(\int_{X\times Y} dq_1\,dp_2\,K^*_{q_1p_2}K_{q_1p_2}\right)\psi\right\rangle\,.
\end{aligned}
\tag{A.5}
$$

We have therefore found our effective observable:

$$
\mathsf{G}^{(\lambda,\mu)}(X\times Y) = \int_{X\times Y} dq\,dp\,K^*_{qp}K_{qp},
\tag{A.6}
$$

where K_{qp} has the kernel

$$
K_{qp}(x,x') = \sqrt{\frac{\lambda}{2\pi\mu}}e^{ip(x-x')}\varphi\big(\lambda\big(q-\tfrac{1}{2}\big((1-\kappa)x+(1+\kappa)x'\big)\big),\tfrac{1}{\mu}(x'-x)\big),
\tag{A.7}
$$

as given in Equation (4.15). As was shown in Section 4.2.1, the K_{qp} satisfy $K_{qp} = W_{qp}K_{00}W^*_{qp}$, with $W_{qp} = \exp[iqp/2]\exp[-iqP]\exp[ipQ]$ being the generators of shifts in phase space, and so the effective observable $\mathsf{G}^{(\lambda,\mu)}$ found by preparing the probes in an arbitrary pure state is covariant under phase space translations.

Next, we consider the case of mixed states $\sigma = \sum_i p_i\sigma_i$, where the σ_i are arbitrary pure states. The post-coupling state is now given by $U(P_\psi\otimes\sigma)U^* = \sum_i p_i U(P_\psi\otimes\sigma_i)U^*$, and the effective observable is now found as follows:

$$\begin{aligned}
\operatorname{tr}\left[\mathsf{H}^{(\lambda,\mu)}(X\times Y)P_\psi\right] &= \operatorname{tr}\left[U(P_\psi\otimes\sigma)U^*(I\otimes\mathsf{E}^{Q_1}(\lambda X)\otimes\mathsf{E}^{P_2}(\mu Y))\right]\\
&= \sum_i p_i \operatorname{tr}\left[U(P_\psi\otimes\sigma_i)U^*(I\otimes\mathsf{E}^{Q_1}(\lambda X)\otimes\mathsf{E}^{P_2}(\mu Y))\right]\\
&= \operatorname{tr}\left[\left(\sum_i p_i \mathsf{G}_i^{(\lambda,\mu)}(X\times Y)\right)P_\psi\right],
\end{aligned} \tag{A.8}$$

where $\mathsf{G}_i^{(\lambda,\mu)}$ is the covariant phase space observable associated with the probes prepared in the pure state σ_i. Since this holds for any state $\rho\in\mathcal{S}(\mathcal{H})$ in place of P_ψ, by the convexity of $\mathcal{S}(\mathcal{H})$, we have the linearity condition

$$\mathsf{H}^{(\lambda,\mu)}(X\times Y) = \sum_i p_i\mathsf{G}_i^{(\lambda,\mu)}(X\times Y), \tag{A.9}$$

which we used in Section 4.12 to show that any state preparation of the probes leads to a covariant phase space observable being measured on our considered system.

A.2 Marginal observables of $\mathsf{G}^{(\lambda,\mu)}$

We find the margins of the observable $\mathsf{G}^{(\lambda,\mu)}$, $\mathsf{E}^{(\lambda,\mu)}$ and $\mathsf{F}^{(\lambda,\mu)}$, by integrating over the outcome space of the other variable (this may be seen as projecting down to a one-dimensional subspace of phase space):

$$\mathsf{E}^{(\lambda,\mu)}(X) = \mathsf{G}^{(\lambda,\mu)}(X\times\mathbb{R}), \tag{A.10a}$$

$$\mathsf{F}^{(\lambda,\mu)}(Y) = \mathsf{G}^{(\lambda,\mu)}(\mathbb{R}\times Y). \tag{A.10b}$$

Considering the case where the probes are prepared in the pure state $\varphi\in\mathcal{H}_1\otimes\mathcal{H}_2$, we first calculate $\mathsf{E}^{(\lambda,\mu)}$:

$$\begin{aligned}
\mathsf{E}^{(\lambda,\mu)}(X) =& \int_{X\times\mathbb{R}} dq\,dp\,K_{qp}^*K_{qp}\\
=&\frac{\lambda}{\mu}\int_X dq\int_{\mathbb{R}^3} dx\,dx'\,dy'\left(\frac{1}{2\pi}\int_{\mathbb{R}} dp\,e^{ip(y'-x')}\right)\\
&\times\overline{\varphi(\lambda(q-\tfrac{1}{2}((1-\kappa)x+(1+\kappa)y')),\tfrac{1}{\mu}(y'-x))}\\
&\times\varphi(\lambda(q-\tfrac{1}{2}((1-\kappa)x+(1+\kappa)x')),\tfrac{1}{\mu}(x'-x))\,|y'\rangle\langle x'|\,.\\
=&\frac{\lambda}{\mu}\int_X dq\int_{\mathbb{R}^2} dx\,dx'\left|\varphi(\lambda(q-\tfrac{1}{2}((1-\kappa)x+(1+\kappa)x')),\tfrac{1}{\mu}(x'-x))\right|^2|x'\rangle\langle x'|\,,
\end{aligned} \tag{A.11}$$

where we have used the identity $\int_{\mathbb{R}} dk\,\exp(ikx) = 2\pi\delta(x)$. We define $q' = x'-x$, so $x = x'-q'$, $dx = -dq'$ and $(1-\kappa)x+(1+\kappa)x' = 2x'-(1-\kappa)q'$. $\mathsf{E}^{(\lambda,\mu)}$ then takes the form:

$$\begin{aligned}\mathsf{E}^{(\lambda,\mu)}(X) &= \frac{\lambda}{\mu}\int_X dq \int_{\mathbb{R}^2} dq'\,dx' \left|\varphi(\lambda(\tfrac{1}{2}(1-\kappa)q' - (x'-q)), \tfrac{1}{\mu}q')\right|^2 |x'\rangle\langle x'| \\ &= \int_X dq \int_{\mathbb{R}} dx' e^{(\lambda,\mu)}(x'-q)\,|x'\rangle\langle x'| \\ &= \int_{\mathbb{R}} dq\,\chi_X(q) e^{(\lambda,\mu)}(Q-q) \\ &= (\chi_X * e^{(\lambda,\mu)})(Q),\end{aligned} \tag{A.12}$$

as is given in Equation (4.24a). The probability distribution $e^{(\lambda,\mu)}$, which characterizes the noise in the measurement of $\mathsf{E}^{(\lambda,\mu)}$, is of the form

$$e^{(\lambda,\mu)}(q) = \frac{\lambda}{\mu}\int_{\mathbb{R}} dq' \left|\varphi(\lambda(\tfrac{1}{2}(1-\kappa)q' - q), \tfrac{1}{\mu}q')\right|^2, \tag{A.13}$$

with first and second moments

$$\begin{aligned}e^{(\lambda,\mu)}[1] &= \int_{\mathbb{R}} dq\,q\,e^{(\lambda,\mu)}(q) \\ &= \frac{1}{\mu}\int_{\mathbb{R}^2} dq\,dq'(\tfrac{1}{2}(1-\kappa)q' - \tfrac{1}{\lambda}q)\left|\varphi\big(q, \tfrac{1}{\mu}q'\big)\right|^2 \\ &= \int_{\mathbb{R}^2} dq\,dq'(\tfrac{\mu}{2}(1-\kappa)q' - \tfrac{1}{\lambda}q)\left|\varphi(q,q')\right|^2 \\ &= \frac{\mu}{2}(1-\kappa)\langle Q_2\rangle_\varphi - \frac{1}{\lambda}\langle Q_1\rangle_\varphi,\end{aligned} \tag{A.14}$$

$$\begin{aligned}e^{(\lambda,\mu)}[2] &= \int_{\mathbb{R}} dq\,q^2\,e^{(\lambda,\mu)}(q) \\ &= \int_{\mathbb{R}^2} dq\,dq'(\tfrac{\mu}{2}(1-\kappa)q' - \tfrac{1}{\lambda}q)^2\left|\varphi(q,q')\right|^2 \\ &= \frac{\mu^2}{4}(1-\kappa)^2\langle Q_2^2\rangle_\varphi + \frac{1}{\lambda^2}\langle Q_1^2\rangle_\varphi - \frac{\mu}{\lambda}(1-\kappa)\langle Q_1Q_2\rangle_\varphi,\end{aligned} \tag{A.15}$$

where $\langle Q_1\rangle_\varphi = \mathrm{tr}\,[Q_1P_\varphi]$, etc. Using (A.14) and (A.15), the variance of $e^{(\lambda,\mu)}$ is

$$\mathrm{Var}\left(e^{(\lambda,\mu)}\right) = \frac{1}{\lambda^2}\mathrm{Var}\,(Q_1,\varphi) + \frac{\mu^2}{4}(1-\kappa)^2\mathrm{Var}\,(Q_2,\varphi) - \frac{\mu}{\lambda}(1-\kappa)\mathrm{Cov}(Q_1,Q_2,\varphi), \tag{A.16}$$

where $\mathrm{Cov}(Q_1,Q_2,\varphi) = \langle Q_1Q_2\rangle_\varphi - \langle Q_1\rangle_\varphi\langle Q_2\rangle_\varphi$ is the covariance of Q_1 and Q_2 with respect to φ. In a similar fashion, we derive an explicit form for $\mathsf{F}^{(\lambda,\mu)}$, the first step of which is to perform a Fourier transform on φ:

$$\begin{aligned}\varphi(\lambda(q - \tfrac{1}{2}((1-\kappa)x + (1+\kappa)x')), \tfrac{1}{\mu}(x'-x)) =& \frac{\mu}{2\pi\lambda}\int_{\mathbb{R}^2} dw\,dz\,e^{iw(q-\frac{1}{2}((1-\kappa)x+(1+\kappa)x'))} \\ &\times e^{iz(x'-x)}\widetilde{\varphi}(\tfrac{w}{\lambda}, \mu z),\end{aligned} \tag{A.17a}$$

$$\begin{aligned}\overline{\varphi(\lambda(q - \tfrac{1}{2}((1-\kappa)x + (1+\kappa)y')), \tfrac{1}{\mu}(y'-x))} =& \frac{\mu}{2\pi\lambda}\int_{\mathbb{R}^2} dw'\,dz'\,e^{-iw'(q-\frac{1}{2}((1-\kappa)x+(1+\kappa)y'))} \\ &\times e^{-iz'(y'-x)}\overline{\widetilde{\varphi}(\tfrac{w'}{\lambda}, \mu z')},\end{aligned} \tag{A.17b}$$

and so we find

$$\begin{aligned}
\mathsf{F}^{(\lambda,\mu)}(Y) =&\frac{\mu}{2\pi\lambda}\int_{\mathbb{R}\times Y} dq\,dp \int_{\mathbb{R}^3} dx\,dx'\,dy'\,e^{ip(y'-x')}\int_{\mathbb{R}^4}\frac{dw\,dw'\,dz\,dz'}{4\pi^2}e^{iq(w-w')}e^{-\frac{iw}{2}((1-\kappa)x+(1+\kappa)x')}\\
&\times e^{\frac{iw'}{2}((1-\kappa)x+(1+\kappa)y')}e^{ix(z'-z)}e^{izx'}e^{-iz'y'}\,\widetilde{\varphi}(\tfrac{w}{\lambda},\mu z)\overline{\widetilde{\varphi}(\tfrac{w'}{\lambda},\mu z')}\,|y'\rangle\langle x'|\\
=&\frac{\mu}{2\pi\lambda}\int_{Y} dp \int_{\mathbb{R}^3} dx\,dx'\,dy'\,e^{ip(y'-x')}\int_{\mathbb{R}^4}\frac{dw\,dw'\,dz\,dz'}{2\pi}\left(\frac{1}{2\pi}\int_{\mathbb{R}}dq\,e^{iq(w-w')}\right)\\
&\times e^{-\frac{iw}{2}((1-\kappa)x+(1+\kappa)x')}e^{\frac{iw'}{2}((1-\kappa)x+(1+\kappa)y')}e^{ix(z'-z)}e^{izx'}e^{-iz'y'}\\
&\times \widetilde{\varphi}(\tfrac{w}{\lambda},\mu z)\overline{\widetilde{\varphi}(\tfrac{w'}{\lambda},\mu z')}\,|y'\rangle\langle x'|\\
=&\frac{\mu}{2\pi\lambda}\int_{Y} dp \int_{\mathbb{R}^3} dw\,dz\,dz' \int_{\mathbb{R}^2}dx'\,dy'\,e^{i(p+\frac{w}{2}(1+\kappa))(y'-x')}\left(\frac{1}{2\pi}\int_{\mathbb{R}}dx\,e^{ix(z'-z)}\right)\\
&\times e^{izx'}e^{-iz'y'}\,\widetilde{\varphi}(\tfrac{w}{\lambda},\mu z)\overline{\widetilde{\varphi}(\tfrac{w}{\lambda},\mu z')}\,|y'\rangle\langle x'|\\
=&\frac{\mu}{2\pi\lambda}\int_{Y} dp \int_{\mathbb{R}^2} dx'\,dy' \int_{\mathbb{R}^2}dw\,dz\,e^{i(p+\frac{w}{2}(1+\kappa)-z)(y'-x')}\left|\widetilde{\varphi}(\tfrac{w}{\lambda},\mu z)\right|^2|y'\rangle\langle x'|\\
=&\frac{\mu}{\lambda}\int_{Y} dp \int_{\mathbb{R}^2} dw\,dz\,\left|\widetilde{\varphi}(\tfrac{w}{\lambda},\mu z)\right|^2\left(\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}dy'\,e^{iy'(p+\frac{w}{2}(\kappa+1)-z)}\,|y'\rangle\right)\\
&\qquad\times\left(\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}dx'\,e^{ix'(p+\frac{w}{2}(\kappa+1)-z)}\,|x'\rangle\right)^*\\
=&\frac{\mu}{\lambda}\int_{Y} dp \int_{\mathbb{R}^2} dw\,dz\,\left|\widetilde{\varphi}(\tfrac{w}{\lambda},\mu z)\right|^2\left|p+\tfrac{w}{2}(\kappa+1)-z\right\rangle\left\langle p+\tfrac{w}{2}(\kappa+1)-z\right|.
\end{aligned}\tag{A.18}$$

Defining $p' = p + \frac{w}{2}(\kappa+1) - z$, so $z = p - p' + \frac{w}{2}(\kappa+1)$ and $dz = -dp'$, $\mathsf{F}^{(\lambda,\mu)}$ takes the form

$$\begin{aligned}
\mathsf{F}^{(\lambda,\mu)}(Y) &= \frac{\mu}{\lambda}\int_Y dp\int_{\mathbb{R}^2} dw\,dp'\,\left|\widetilde{\varphi}\big(\tfrac{w}{\lambda},\mu(p-p'+\tfrac{w}{2}(\kappa+1))\big)\right|^2\left|p'\right\rangle\left\langle p'\right|\\
&= \int_Y dp\int_{\mathbb{R}} dp'\,f^{(\lambda,\mu)}(p'-p)\left|p'\right\rangle\left\langle p'\right|\\
&= \int_{\mathbb{R}} dp\,\chi_Y(p)f^{(\lambda,\mu)}(P-p)\\
&= (\chi_Y * f^{(\lambda,\mu)})(P),
\end{aligned}\tag{A.19}$$

as is given in Equation (4.24b). The probability distribution $f^{(\lambda,\mu)}$ is of the form

$$f^{(\lambda,\mu)}(p) = \frac{\mu}{\lambda}\int_{\mathbb{R}} dw\,\left|\widetilde{\varphi}\big(\tfrac{w}{\lambda},\mu(\tfrac{w}{2}(\kappa+1)-p)\big)\right|^2. \tag{A.20}$$

Following the same method used to derive (A.14) and (A.15), we find the first and second moments of $f^{(\lambda,\mu)}$:

$$f^{(\lambda,\mu)}[1] = \frac{\lambda}{2}(1+\kappa)\left\langle P_1\right\rangle_\varphi - \frac{1}{\mu}\left\langle P_2\right\rangle_\varphi, \tag{A.21}$$

$$f^{(\lambda,\mu)}[2] = \frac{\lambda^2}{4}(1+\kappa)^2\left\langle P_1^2\right\rangle_\varphi + \frac{1}{\mu^2}\left\langle P_2^2\right\rangle_\varphi - \frac{\lambda}{\mu}(1+\kappa)\left\langle P_1P_2\right\rangle_\varphi. \tag{A.22}$$

From these, the variance of $f^{(\lambda,\mu)}$ is given by

$$\mathrm{Var}\left(f^{(\lambda,\mu)}\right) = \frac{\lambda^2}{4}(1+\kappa)^2\mathrm{Var}\left(P_1,\varphi\right) + \frac{1}{\mu^2}\mathrm{Var}\left(P_2,\varphi\right) - \frac{\lambda}{\mu}(1+\kappa)\mathrm{Cov}(P_1,P_2,\varphi). \tag{A.23}$$

Note that it is now clear why we use the scaled sets λX and μY in Equation (4.12): the scaling is such that the marginal observables $\mathsf{E}^{(\lambda,\mu)}$ and $\mathsf{F}^{(\lambda,\mu)}$ are direct smearings of position and momentum, rather than of scaled versions.

We will now return to the case of the mixed state $\sigma = \sum_i p_i \sigma_i$, where $\sigma_i = P_{\varphi_i}$ are arbitrary pure states. The margins of the effective observable $\mathsf{H}^{(\lambda,\mu)}$ are now given in terms of the margins of the effective observables derived from the σ_i:

$$\begin{aligned}\mathsf{E}^{(\lambda,\mu)}(X) &= \mathsf{H}^{(\lambda,\mu)}(X\times\mathbb{R}) = \sum_i p_i \mathsf{G}_i^{(\lambda,\mu)}(X\times\mathbb{R}) = \sum_i p_i \mathsf{M}_i^{(\lambda,\mu)}(X) \\ &= \sum_i p_i(\chi_X * m_i^{(\lambda,\mu)})(Q) = (\chi_X * e^{(\lambda,\mu)})(Q),\end{aligned} \tag{A.24a}$$

$$\begin{aligned}\mathsf{F}^{(\lambda,\mu)}(Y) &= \mathsf{H}^{(\lambda,\mu)}(\mathbb{R}\times Y) = \sum_i p_i \mathsf{G}_i^{(\lambda,\mu)}(\mathbb{R}\times Y) = \sum_i p_i \mathsf{N}_i^{(\lambda,\mu)}(Y) \\ &= \sum_i p_i(\chi_Y * n_i^{(\lambda,\mu)})(P) = (\chi_Y * f^{(\lambda,\mu)})(P).\end{aligned} \tag{A.24b}$$

As we see, these margins again have the form as given in Equations (4.24a) and (4.24b), with the probability distributions

$$e^{(\lambda,\mu)}(q) = \sum_i p_i m_i^{(\lambda,\mu)}(q) = \frac{\lambda}{\mu}\sum_i p_i \int_{\mathbb{R}} dq' \left|\varphi_i(\lambda(\tfrac{1}{2}(1-\kappa)q' - q), \tfrac{1}{\mu}q')\right|^2, \tag{A.25a}$$

$$f^{(\lambda,\mu)}(p) = \sum_i p_i n_i^{(\lambda,\mu)}(p) = \frac{\mu}{\lambda}\sum_i p_i \int_{\mathbb{R}} dw \left|\widetilde{\varphi}_i\big(\tfrac{w}{\lambda}, \mu(\tfrac{w}{2}(\kappa+1) - p)\big)\right|^2. \tag{A.25b}$$

From here, and by using equations (A.14), (A.15), (A.21), (A.22), the first and second moments of these distributions can be readily calculated:

$$\begin{aligned}e^{(\lambda,\mu)}[1] &= \int_{\mathbb{R}} dq\, q\, e^{(\lambda,\mu)} = \sum_i p_i \int_{\mathbb{R}} dq\, q\, m_i^{(\lambda,\mu)}(q) = \sum_i p_i m^{(\lambda,\mu)}[1] \\ &= \sum_i p_i \left(\frac{\mu}{2}(1-\kappa)\langle Q_2\rangle_{\varphi_i} - \frac{1}{\lambda}\langle Q_1\rangle_{\varphi_i}\right) \\ &= \frac{\mu}{2}(1-\kappa)\langle Q_2\rangle_\sigma - \frac{1}{\lambda}\langle Q_1\rangle_\sigma,\end{aligned} \tag{A.26}$$

$$e^{(\lambda,\mu)}[2] = \frac{\mu^2}{4}(1-\kappa)^2\langle Q_2^2\rangle_\sigma + \frac{1}{\lambda^2}\langle Q_1^2\rangle_\sigma - \frac{\mu}{\lambda}\langle Q_1Q_2\rangle_\sigma, \tag{A.27}$$

$$f^{(\lambda,\mu)}[1] = \frac{\lambda}{2}(1+\kappa)\langle P_1\rangle_\sigma - \frac{1}{\mu}\langle P_2\rangle_\sigma, \tag{A.28}$$

$$f^{(\lambda,\mu)}[2] = \frac{\lambda^2}{4}(1+\kappa)^2\langle P_1^2\rangle_\sigma + \frac{1}{\mu^2}\langle P_2^2\rangle_\sigma - \frac{\lambda}{\mu}(1+\kappa)\langle P_1P_2\rangle_\sigma, \tag{A.29}$$

and so these probability distributions have variances of the form given in (A.16) and (A.23)

$$\mathrm{Var}\left(e^{(\lambda,\mu)}\right) = \frac{1}{\lambda^2}\mathrm{Var}\,(Q_1,\sigma) + \frac{\mu^2}{4}(1-\kappa)^2\mathrm{Var}\,(Q_2,\sigma) - \frac{\mu}{\lambda}(1-\kappa)\mathrm{Cov}(Q_1,Q_2,\sigma), \tag{A.30}$$

$$\mathrm{Var}\left(f^{(\lambda,\mu)}\right) = \frac{\lambda^2}{4}(1+\kappa)^2\mathrm{Var}\,(P_1,\sigma) + \frac{1}{\mu^2}\mathrm{Var}\,(P_2,\sigma) - \frac{\lambda}{\mu}(1+\kappa)\mathrm{Cov}(P_1,P_2,\sigma). \tag{A.31}$$

Note that even if we had specified that the pure states σ_i were product states, we would still find the covariance terms appearing as a result of the classical correlations between them.

B Error values for the margins of covariant phase space observables

In this appendix we shall consider the noise and Wasserstein 2-distance measures covered in Section 2.4.5 and apply them to the margins of covariant phase space observables. In doing so, we will show that for this type of observable the value of the measures coincide, and are equal to the value given in Equation (4.27).

As was shown in Appendix A.2, both margins take the form of a convolution of a sharp observable with a probability distribution. We shall therefore consider the general case of such an observable, denoted here by $\mathsf{Q}_\nu = \nu * \mathsf{Q}$. This form coincides with $\mathsf{E}^{(\lambda,\mu)}$ if we make the identities $\mathsf{Q} = \mathsf{E}^Q$ and $\nu(x) = e^{(\lambda,\mu)}(-x)$.

We begin by considering the noise measure, in particular in the form given by Equation (2.158). The first moment operator $\mathsf{Q}_\nu[1]$ can be readily calculated:

$$\begin{aligned}\mathsf{Q}_\nu[1] &= \int_{\mathbb{R}} x(\nu * \mathsf{Q})(dx)\\ &= \int_{\mathbb{R}^2} dx\, dy\, x\, \nu(y)\mathsf{Q}(x-y)\\ &= \int_{\mathbb{R}^2} dx\, dy\, (x+y)\nu(y)\mathsf{Q}(x)\\ &= \mathsf{Q}[1] + \nu[1],\end{aligned} \tag{B.1}$$

and similarly

$$\mathsf{Q}_\nu[2] = \mathsf{Q}[2] + \nu[2] + 2\mathsf{Q}[1]\nu[1]. \tag{B.2}$$

Since Q is sharp, $\mathsf{Q}[2] = \mathsf{Q}[1]^2$ and so

$$\mathsf{Q}_\nu[2] - \mathsf{Q}_\nu[1]^2 = \nu[2] - \nu[1]^2 = \mathrm{Var}\,(\nu)\,. \tag{B.3}$$

The value of the noise measure for Q_ν as an approximation of Q with respect to the state ψ is therefore

$$\begin{aligned}\varepsilon(\mathsf{Q}_\nu, \mathsf{Q}, \psi)^2 &= \langle\psi|(\mathsf{Q}_\nu[1] - \mathsf{Q}[1])^2\psi\rangle + \langle\psi|(\mathsf{Q}_\nu[2] - \mathsf{Q}_\nu[1]^2)\psi\rangle\\ &= \nu[1]^2 + \mathrm{Var}\,(\nu) = \nu[2].\end{aligned} \tag{B.4}$$

We have shown the first part of Equation (4.27) as well as the fact that for any observable that arises as a convolution of a sharp observable the value of noise measure is state-independent when it is seen as an approximation of that sharp observable.

The proof of the case for the BLW error is given in Lemma 7 of [17], where instead of restricting to the Wasserstein 2-distance, the authors consider the α-distance with $\alpha \in [1, \infty]$, and denote the distance D_α. By setting $\alpha = 2$ we obtain the intended result $\Delta(\nu * \mathsf{Q}, \mathsf{Q})^2 = \Delta(\nu, \delta_0)^2 = \nu[2]$.

C Derivation of $\mathsf{E}^{(\lambda,0)}$ and $\mathsf{F}^{(0,\mu)}$

We begin as we did in Appendix A.1 with the system in the pure state $\psi \in \mathcal{H}$ and the probes in the arbitrary pure state $\varphi \in \mathcal{H}_1 \otimes \mathcal{H}_2$. However, we now set one of the coupling constants λ or μ to zero.

C.1 Setting $\lambda = 0$

If we first set λ to zero, then our coupling unitary U_{int} reduces to $U_\mu = \exp[i\mu P Q_2]$, and our post-coupling state is

$$\Psi_\mu(q, q_1, q_2) := U_\mu(\psi \otimes \varphi)(q, q_1, q_2) = \psi(q + \mu q_2)\varphi(q_1, q_2). \tag{C.1}$$

We perform the ideal momentum measurement on the second probe (with the associated pointer function $h : y \mapsto \mu^{-1}y$ as before), and again perform a Fourier transform on the final argument of Ψ_μ:

$$\Psi_\mu(q, q_1, q_2) \mapsto \widetilde{\Psi}_\mu(q, q_1, w_2) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} dq_2\, e^{-iw_2 q_2}\Psi_\mu(q, q_1, q_2). \tag{C.2}$$

From this we derive the effective observable $\mathsf{F}^{(0,\mu)}$ measured on the system:

$$\begin{aligned}
\left\langle \mathsf{F}^{(0,\mu)}(Y) \right\rangle_\psi &= \langle \mathsf{E}^{P_2}(\mu Y) \rangle_{\Psi_\mu} \\
&= \frac{1}{2\pi} \int_{\mathbb{R}^8} dq\, dq'\, dq_1\, dq_1'\, dq_2\, dq_2'\, dw_2\, dw_2'\, e^{-iw_2 q_2} e^{iw_2' q_2'} \psi(q + \mu q_2)\varphi(q_1, q_2) \\
&\qquad \times \overline{\psi(q' + \mu q_2')\varphi(q_1', q_2')}\, \langle q', q_1' | q, q_1 \rangle \int_{\mu Y} dw\, \langle w_2' | w \rangle \langle w | w_2 \rangle \\
&= \frac{\mu}{2\pi} \int_Y dw \int_{\mathbb{R}^5} dq\, dq'\, dq_1\, dq_2\, dq_2'\, e^{-i\mu w(q_2 - q_2')} \psi(q + \mu q_2)\varphi(q_1, q_2) \\
&\qquad\qquad \times \overline{\psi(q' + \mu q_2')\varphi(q_1, q_2')}\, \langle q' | q \rangle \\
&= \int_Y dw \int_{\mathbb{R}} dq_1 \left(\sqrt{\frac{\mu}{2\pi}} \int_{\mathbb{R}^2} dq_2\, dq\, e^{-i\mu w q_2} \psi(q + \mu q_2)\varphi(q_1, q_2) \, |q\rangle \right)^* \\
&\qquad \times \left(\sqrt{\frac{\mu}{2\pi}} \int_{\mathbb{R}^2} dq_2\, dq\, e^{-i\mu w q_2} \psi(q + \mu q_2)\varphi(q_1, q_2) \, |q\rangle \right).
\end{aligned} \tag{C.3}$$

We now let $q' = q + \mu q_2$, so $q_2 = \frac{1}{\mu}(q' - q)$ and $dq_2 = -\frac{1}{\mu} dq'$. Hence,

$$\begin{aligned}
\left\langle \mathsf{F}^{(0,\mu)}(Y) \right\rangle_\psi &= \int_Y dw \int_{\mathbb{R}} dq_1 \left(\sqrt{\frac{1}{2\pi\mu}} \int_{\mathbb{R}^2} dq\, dq'\, e^{iw(q - q')} \psi(q')\varphi(q_1, \tfrac{1}{\mu}(q' - q))\, |q\rangle \right)^* \\
&\qquad \times \left(\sqrt{\frac{1}{2\pi\mu}} \int_{\mathbb{R}^2} dq\, dq'\, e^{iw(q - q')} \psi(q')\varphi(q_1, \tfrac{1}{\mu}(q' - q))\, |q\rangle \right) \\
&= \int_Y dw \int_{\mathbb{R}} dq_1 \left(\int_{\mathbb{R}^2} dq\, dq'\, K_{q_1 w}(q, q')\psi(q')\, |q\rangle \right)^* \\
&\qquad \times \left(\int_{\mathbb{R}^2} dq\, dq'\, K_{q_1 w}(q, q')\psi(q')\, |q\rangle \right) \\
&= \left\langle \left(\int_{\mathbb{R} \times Y} dq_1\, dw\, K^*_{q_1 w} K_{q_1 w} \right) \right\rangle_\psi .
\end{aligned} \tag{C.4}$$

That is,

$$\mathsf{F}^{(0,\mu)}(Y) = \int_{\mathbb{R}\times Y} dq\, dp\, K_{qp}^* K_{qp}, \tag{C.5}$$

where

$$K_{qp}(x, x') = \sqrt{\frac{1}{2\pi\mu}} e^{ip(x-x')} \varphi(q, \tfrac{1}{\mu}(x'-x)). \tag{C.6}$$

We can explicitly expand to find the form of $\mathsf{F}^{(0,\mu)}$:

$$\begin{aligned}
\mathsf{F}^{(0,\mu)}(Y) =& \frac{1}{2\pi\mu} \int_{\mathbb{R}\times Y} dq\, dp \int_{\mathbb{R}^4} dx\, dx'\, dy\, dy'\, e^{-ip(x-x')} e^{ip(y-y')} \overline{\varphi(q, \tfrac{1}{\mu}(x'-x))} \\
& \qquad\qquad\qquad\qquad \times \varphi(q, \tfrac{1}{\mu}(y'-y)) \left|x'\right\rangle \left\langle x|y\right\rangle \left\langle y'\right| \\
=& \frac{1}{2\pi\mu} \int_{\mathbb{R}\times Y} dq\, dp \int_{\mathbb{R}^3} dx\, dx'\, dy'\, e^{ip(x'-y')} \overline{\varphi(q, \tfrac{1}{\mu}(x'-x))} \varphi(q, \tfrac{1}{\mu}(y'-x)) \left|x'\right\rangle \left\langle y'\right| \\
=& \frac{\mu}{2\pi} \int_{\mathbb{R}\times Y} dq\, dp \int_{\mathbb{R}^3} dx\, dx'\, dy'\, e^{ip(x'-y')} \\
& \times \int_{\mathbb{R}^4} \frac{dw\, dw'\, dz\, dz'}{4\pi^2} e^{-iq(w'-w)} e^{ix(z'-z)} e^{izy'} e^{-iz'x'} \overline{\widetilde{\varphi}(w', \mu z')} \widetilde{\varphi}(w, \mu z) \left|x'\right\rangle \left\langle y'\right| \\
=& \frac{\mu}{2\pi} \int_{Y} dp \int_{\mathbb{R}^3} dx\, dx'\, dy'\, e^{ip(x'-y')} \\
& \times \int_{\mathbb{R}^3} \frac{dw\, dz\, dz'}{2\pi} e^{ix(z'-z)} e^{i(zy'-z'x')} \overline{\widetilde{\varphi}(w, \mu z')} \widetilde{\varphi}(w, \mu z) \left|x'\right\rangle \left\langle y'\right| \\
=& \frac{\mu}{2\pi} \int_{Y} dp \int_{\mathbb{R}^2} dx'\, dy'\, e^{i(p-z)(x'-y')} \int_{\mathbb{R}^2} dw\, dz\, \left|\widetilde{\varphi}(w, \mu z)\right|^2 \left|x'\right\rangle \left\langle y'\right| \\
=& \mu \int_{Y} dp \int_{\mathbb{R}^2} dw\, dz \left|\widetilde{\varphi}(w, \mu z)\right|^2 \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} dx' e^{ix'(p-z)} |x'\rangle \right) \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} dy' e^{iy'(p-z)} |y'\rangle \right)^* \\
=& \mu \int_{Y} dp \int_{\mathbb{R}^2} dw\, dz\, \left|\widetilde{\varphi}(w, \mu z)\right|^2 \left|p-z\right\rangle \left\langle p-z\right| .
\end{aligned} \tag{C.7}$$

We let $p' = p - z$, so $z = p - p'$ and $dz = -dp'$. Therefore,

$$\begin{aligned}
\mathsf{F}^{(0,\mu)}(Y) =& \mu \int_{Y} dp \int_{\mathbb{R}^2} dw\, dp'\, \left|\widetilde{\varphi}(w, \mu(p-p'))\right|^2 \left|p'\right\rangle \left\langle p'\right| \\
=& \int_{Y} dp \int_{\mathbb{R}} dp'\, f^{(0,\mu)}(p'-p) \left|p'\right\rangle \left\langle p'\right| \\
=& \int_{\mathbb{R}} dp\, \chi_Y(p) f^{(0,\mu)}(P-p) \\
=& (\chi_Y * f^{(0,\mu)})(P).
\end{aligned} \tag{C.8}$$

In other words, $\mathsf{F}^{(0,\mu)}$ is a smearing of the ideal momentum observable, with the probability distribution $f^{(0,\mu)}$ of the from

$$f^{(0,\mu)}(p) = \mu \int_{\mathbb{R}} dw \left|\widetilde{\varphi}(w, -\mu p)\right|^2 . \tag{C.9}$$

The first and second moments of $f^{(0,\mu)}$ are of the form

$$f^{(0,\mu)}[1] = \mu \int_{\mathbb{R}^2} dp\, dw\, p \left|\widetilde{\varphi}(w, -\mu p)\right|^2 = \frac{1}{\mu} \int_{\mathbb{R}^2} dp\, dw\, p \left|\widetilde{\varphi}(w, p)\right|^2 = \frac{1}{\mu} \left\langle P_2 \right\rangle_{\varphi}, \tag{C.10a}$$

$$f^{(0,\mu)}[2] = \mu \int_{\mathbb{R}^2} dp\, dw\, p^2 \left|\widetilde{\varphi}(w, -\mu p)\right|^2 = \frac{1}{\mu^2} \int_{\mathbb{R}^2} dp\, dw\, p^2 \left|\widetilde{\varphi}(w, p)\right|^2 = \frac{1}{\mu^2} \left\langle P_2^2 \right\rangle_{\varphi}, \tag{C.10b}$$

and so $f^{(0,\mu)}$ has variance

$$\mathrm{Var}\left(f^{(0,\mu)}\right) = f^{(0,\mu)}[2] - f^{(0,\mu)}[1]^2 = \frac{1}{\mu^2}\mathrm{Var}\left(P_2, \varphi\right). \tag{C.11}$$

C.2 Setting $\mu = 0$

We shall now consider the case where we set μ to zero. In this case $U_{int} = U_\lambda$, and we express the coupled state $\Psi_\lambda := U_\lambda(\psi \otimes \varphi)$ as

$$\Psi_\lambda = U_\lambda(\psi \otimes \varphi) = \int_{\mathbb{R}} d\mathsf{E}^Q(q)\psi \otimes e^{-i\lambda q P_1}\varphi = \int_{\mathbb{R}} d\mathsf{E}^Q(q)\psi \otimes \varphi_{\lambda q}, \tag{C.12}$$

where we have expanded $U_\lambda = \exp[-i\lambda Q P_1] = \int_{\mathbb{R}} d\mathsf{E}^Q(q) \otimes \exp[-i\lambda q P_1]$ and defined the state $\varphi_{\lambda q}(q_1, q_2) = \varphi(q_1 - \lambda q, q_2)$. We now perform an ideal position measurement on the first probe, using the pointer function $g : x \mapsto \lambda^{-1}x$, in order to derive the observable $\mathsf{E}^{(\lambda,0)}$ measured on the system:

$$\begin{aligned}
\left\langle \mathsf{E}^{(\lambda,0)}(X)\right\rangle_\psi =& \left\langle \mathsf{E}^{Q_1}(\lambda X)\right\rangle_{\Psi_\lambda} \\
=& \int_{\mathbb{R}^2} \left\langle d\mathsf{E}^Q(q) d\mathsf{E}^Q(q')\right\rangle_\psi \int_{\lambda X} \left\langle d\mathsf{E}^{Q_1}(x)\right\rangle_{\varphi_{\lambda q}} \\
=& \lambda \int_{\mathbb{R}} \left\langle d\mathsf{E}^Q(q)\right\rangle_\psi \int_X \left\langle d\mathsf{E}^{Q_1}(\lambda x)\right\rangle_{\varphi_{\lambda q}} \\
=& \lambda \int_{\mathbb{R}} \left\langle d\mathsf{E}^Q(q)\right\rangle_\psi \int_{\mathbb{R}^4} dq_1\, dq_1'\, dq_2\, dq_2' \\
& \times \varphi(q_1 - \lambda q, q_2)\overline{\varphi(q_1' - \lambda q, q_2)} \left\langle q_2' | q_2 \right\rangle \int_X \left\langle q_1' \middle| d\mathsf{E}^{Q_1}(\lambda x) q_1 \right\rangle \\
=& \int_X dx \int_{\mathbb{R}^2} dq_2\, \lambda \left|\varphi(-\lambda(q - x), q_2)\right|^2 \left\langle d\mathsf{E}^Q(q)\right\rangle_\psi \\
=& \int_X dx \int_{\mathbb{R}} e^{(\lambda,0)}(q - x) \left\langle d\mathsf{E}^Q(q)\right\rangle_\psi \\
=& \left\langle (\chi_X * e^{(\lambda,0)})(Q)\right\rangle_\psi .
\end{aligned} \tag{C.13}$$

In other words,

$$\mathsf{E}^{(\lambda,0)}(X) = (\chi_X * e^{(\lambda,0)})(Q), \tag{C.14}$$

so $\mathsf{E}^{(\lambda,0)}$ is a smeared version of the ideal position observable, much like $\mathsf{F}^{(0,\mu)}$ was with momentum, where

$$e^{(\lambda,0)}(q) = \lambda \int_{\mathbb{R}} dq' \left|\varphi(-\lambda q, q')\right|^2. \tag{C.15}$$

We can again quickly find the first and second moments of the probability distribution $e^{(\lambda,0)}$:

$$e^{(\lambda,0)}[1] = \lambda \int_{\mathbb{R}^2} dq\, dq' q \left|\varphi(-\lambda q, q')\right|^2 = \frac{1}{\lambda}\int_{\mathbb{R}^2} dq\, dq' q \left|\varphi(q, q')\right|^2 = \frac{1}{\lambda}\left\langle Q_1\right\rangle_\varphi, \tag{C.16a}$$

$$e^{(\lambda,0)}[2] = \lambda \int_{\mathbb{R}^2} dq\, dq' q^2 \left|\varphi(-\lambda q, q')\right|^2 = \frac{1}{\lambda^2}\int_{\mathbb{R}^2} dq\, dq' q^2 \left|\varphi(q, q')\right|^2 = \frac{1}{\lambda^2}\left\langle Q_1^2\right\rangle_\varphi, \tag{C.16b}$$

and so $e^{(\lambda,0)}$ has variance

$$\mathrm{Var}\left(e^{(\lambda,0)}\right) = \frac{1}{\lambda^2}\mathrm{Var}\left(Q_1, \varphi\right). \tag{C.17}$$

Finally, it should be noted that for both $\mathsf{E}^{(\lambda,0)}$ and $\mathsf{F}^{(0,\mu)}$, we can prepare the probes in a mixed state $\sigma \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ and, much like in Appendix A.2, the resultant observables would be the same form, as would the variances, with φ replaced by σ.

D Hall's optimisation and suboptimal approximations with zero noise measure value

Within his paper [28] Hall derives the optimal post-selection functions f and g needed for his observable M, with first moment operator $\mathsf{M}[1] =: M$, to minimise the noise measure when approximating $A = \mathsf{A}[1]$ and $B = \mathsf{B}[1]$ via $f(M)$ and $g(M)$, respectively. Note that, unlike Branciard [7], Hall does not assume an ancillary system in his construction, hence the approximating observable C is of the form

$$\mathsf{C}(k) = \sum_{m \in f^{-1}(k)} \mathsf{M}(m), \tag{D.1}$$

where one assumes M to be unsharp in general. The n^{th} moment operator $\mathsf{C}[n]$ is then given by

$$\mathsf{C}[n] = \sum_k k^n \mathsf{C}(k) = \sum_m f(m)^n \mathsf{M}(m). \tag{D.2}$$

By expanding the form of the noise measure given in Equation (2.158), where ψ has been replaced by the density operator ρ, we find

$$\begin{aligned}
\varepsilon(\mathsf{C}, \mathsf{A}, \rho)^2 &= \operatorname{tr}\left[(\mathsf{C}[2] - \mathsf{C}[1]A - A\mathsf{C}[1] + A^2)\rho\right] \\
&= \sum_k k^2 \operatorname{tr}\left[\mathsf{C}(k)\rho\right] - \sum_k k \operatorname{tr}\left[(\mathsf{C}(k)A + A\mathsf{C}(k))\rho\right] + \operatorname{tr}\left[A^2\rho\right] \\
&= \sum_m f(m)^2 \operatorname{tr}\left[\mathsf{M}(m)\rho\right] - \sum_m f(m) \operatorname{tr}\left[(\mathsf{M}(m)A + A\mathsf{M}(m))\rho\right] + \operatorname{tr}\left[A^2\rho\right] \\
&= \sum_m \operatorname{tr}\left[\mathsf{M}(m)\rho\right] \left(f(m) - \frac{\operatorname{tr}\left[(\mathsf{M}(m)A + A\mathsf{M}(m))\rho\right]}{2\operatorname{tr}\left[\mathsf{M}(m)\rho\right]} \right)^2 \\
&\quad - \sum_m \frac{\operatorname{tr}\left[(\mathsf{M}(m)A + A\mathsf{M}(m))\rho\right]^2}{4\operatorname{tr}\left[\mathsf{M}(m)\rho\right]} + \operatorname{tr}\left[A^2\rho\right].
\end{aligned} \tag{D.3}$$

The first term is non-negative and the last two terms are independent of $f(m)$. Therefore, the noise measure is minimised by setting

$$f(m) = \frac{\operatorname{tr}\left[(\mathsf{M}(m)A + A\mathsf{M}(m))\rho\right]}{2\operatorname{tr}\left[\mathsf{M}(m)\rho\right]}. \tag{D.4}$$

However, just because a minimising function f can be found, this does not mean that the observable C defined by it is accurately approximating the observable A. As an example, consider the observable A associated with the self-adjoint operator

$$A = \mathsf{A}[1] = \frac{\gamma}{2}(\sigma_x - \sigma_y), \tag{D.5}$$

where σ_x and σ_y denote the regular Pauli matrices, and $\gamma = 2 - \sqrt{2}$. We note here the useful identity $\gamma = \sqrt{2}(1 - \gamma)$. The observable M that we shall measure, and whose outcomes we shall post-process to approximate A, is a 3-outcome observable with effects

$$\begin{aligned}
\mathsf{M}(1) &= \frac{\gamma}{2}(I+\sigma_x),\\
\mathsf{M}(2) &= \frac{\gamma}{2}(I+\sigma_y),\\
\mathsf{M}(3) &= 2(1-\gamma)\frac{1}{2}\left(I-\frac{1}{\sqrt{2}}(\sigma_x+\sigma_y)\right).
\end{aligned}\tag{D.6}$$

These are all positive rank-one operators, and $\mathsf{M}(1)+\mathsf{M}(2)+\mathsf{M}(3)=I$. Given that this observable is not dichotomic the noise measure retains its state-dependence, and so we choose to measure in the state

$$\rho = \frac{1}{2}\left(I-\frac{1}{\sqrt{2}}(\sigma_x+\sigma_y)\right).\tag{D.7}$$

With this choice of state we find the optimal post-selection function to be of the form

$$\begin{aligned}
f(1) &= \frac{\operatorname{tr}\left[(\mathsf{M}(1)A+A\mathsf{M}(1))\rho\right]}{2\operatorname{tr}\left[\mathsf{M}(1)\rho\right]} = 1,\\
f(2) &= -1,\\
f(3) &= 0.
\end{aligned}\tag{D.8}$$

With these values, as shown in [19], we find that the approximating observable C defined in Equation (D.1) satisfies

$$\begin{aligned}
\mathsf{C}[1] &= \sum_m f(m)\mathsf{M}(m) = \mathsf{M}(1)-\mathsf{M}(2) = A,\\
\mathsf{C}[1]^2 &= A^2 = \frac{\gamma^2}{2}I,\\
\mathsf{C}[2] &= \mathsf{M}(1)+\mathsf{M}(2) = I-\mathsf{M}(3),\\
\mathsf{C}[2]-\mathsf{C}[1]^2 &= 2(1-\gamma)\frac{1}{2}\left(I+\frac{1}{\sqrt{2}}(\sigma_x+\sigma_y)\right).
\end{aligned}\tag{D.9}$$

Hence, the noise measure is equal to $\varepsilon(\mathsf{C},\mathsf{A},\rho)^2=0$, but in no way does C provide an accurate approximation of A, with the probability distributions p_ρ^{A} and p_ρ^{C} being very different:

$$p_\rho^{\mathsf{A}}(\pm) = \frac{1}{2}, \qquad p_\rho^{\mathsf{C}}(\pm) = \frac{\gamma^2}{4}, \quad p_\rho^{\mathsf{C}}(0) = 2(1-\gamma).\tag{D.10}$$

E Derivation of Yu and Oh's minimum error curves

E.1 The error curves

In this appendix we derive the minimum error curve given by Yu and Oh. The curves arise from a geometric consideration of the quantities $\Delta(\mathsf{C},\mathsf{A})^2$ and $\Delta(\mathsf{D},\mathsf{B})^2$ on the Bloch sphere, with the proviso that C and D are on the boundary of being jointly measurable. For ease of notation, in what follows we will consider the quantities $D(\mathsf{C},\mathsf{A}) = \frac{1}{2}\Delta(\mathsf{C},\mathsf{A})^2 = \|\boldsymbol{a}-\boldsymbol{c}\|$, etc..

We begin by assuming the optimal case: the approximating observables C and D considered possess values $D(\mathsf{C},\mathsf{A})$ and $D(\mathsf{D},\mathsf{B})$ that are as small as possible. By demanding these observables be jointly measurable we require that their Bloch vectors satisfy

$$\|\boldsymbol{c}+\boldsymbol{d}\| + \|\boldsymbol{c}-\boldsymbol{d}\| \leq 2. \tag{E.1}$$

This may be expressed equivalently by saying that the vector $\boldsymbol{d}$ lies within the area enclosed by the ellipsoid

$$E_{\boldsymbol{c}} = \{\boldsymbol{d} \,|\, \|\boldsymbol{c}+\boldsymbol{d}\| + \|\boldsymbol{c}-\boldsymbol{d}\| = 2\}, \tag{E.2}$$

and similarly $\boldsymbol{c}$ lies within the area enclosed by the ellipsoid

$$E_{\boldsymbol{d}} = \{\boldsymbol{c} \,|\, \|\boldsymbol{c}+\boldsymbol{d}\| + \|\boldsymbol{c}-\boldsymbol{d}\| = 2\}. \tag{E.3}$$

Since we have assumed that $D(\mathsf{C},\mathsf{A})$ and $D(\mathsf{D},\mathsf{B})$ are at their smallest possible values, this would mean the distance between $\boldsymbol{c}$ and $\boldsymbol{a}$ (resp. $\boldsymbol{d}$ and $\boldsymbol{b}$) are as small as possible whilst still requiring $\boldsymbol{c}$ ($\boldsymbol{d}$) lies within the area enclosed by $E_{\boldsymbol{d}}$ ($E_{\boldsymbol{c}}$). By making use of the concept of Lagrange multipliers, we find that $\boldsymbol{c}$ lies on the ellipsoid $E_{\boldsymbol{d}}$; in other words,

$$\|\boldsymbol{c}+\boldsymbol{d}\| + \|\boldsymbol{c}-\boldsymbol{d}\| = 2, \tag{E.4}$$

and the vector $\boldsymbol{a}-\boldsymbol{c}$ is parallel to the gradient of $E_{\boldsymbol{d}}$, and similarly for $\boldsymbol{b}-\boldsymbol{d}$:

$$\begin{aligned} \boldsymbol{a}-\boldsymbol{c} &\propto \boldsymbol{\nabla}_{\boldsymbol{c}}(\|\boldsymbol{c}+\boldsymbol{d}\| + \|\boldsymbol{c}-\boldsymbol{d}\|) = \frac{\boldsymbol{c}+\boldsymbol{d}}{\|\boldsymbol{c}+\boldsymbol{d}\|} + \frac{\boldsymbol{c}-\boldsymbol{d}}{\|\boldsymbol{c}-\boldsymbol{d}\|}, \\ \boldsymbol{b}-\boldsymbol{d} &\propto \boldsymbol{\nabla}_{\boldsymbol{d}}(\|\boldsymbol{c}+\boldsymbol{d}\| + \|\boldsymbol{c}-\boldsymbol{d}\|) = \frac{\boldsymbol{c}+\boldsymbol{d}}{\|\boldsymbol{c}+\boldsymbol{d}\|} - \frac{\boldsymbol{c}-\boldsymbol{d}}{\|\boldsymbol{c}-\boldsymbol{d}\|}. \end{aligned} \tag{E.5}$$

The next step is to find $\boldsymbol{a}$ and $\boldsymbol{b}$ in terms of $\boldsymbol{c}$ and $\boldsymbol{d}$. We begin by first making use of $D(\mathsf{C},\mathsf{A}) = \|\boldsymbol{a}-\boldsymbol{c}\|$ and $D(\mathsf{D},\mathsf{B}) = \|\boldsymbol{b}-\boldsymbol{d}\|$.

Since Equation (E.4) holds, we may make use of an equivalent expression, which comes from expanding it twice:

$$\|\boldsymbol{c}\|^2 + \|\boldsymbol{d}\|^2 = 1 + (\boldsymbol{c}\cdot\boldsymbol{d})^2 = 1 + M^2, \tag{E.6}$$

where we have introduced the term $M = \boldsymbol{c} \cdot \boldsymbol{d}$. As a result,

$$\begin{aligned} \|\boldsymbol{c} \pm \boldsymbol{d}\| &= \Big(\|\boldsymbol{c}\|^2 + \|\boldsymbol{d}\|^2 \pm 2M\Big)^{1/2} \\ &= \big(1 + M^2 \pm 2M\big)^{1/2} \\ &= 1 \pm M, \end{aligned} \tag{E.7}$$

where the positivity of the final term is guaranteed by $|\boldsymbol{c} \cdot \boldsymbol{d}| \leq 1$. With this identity at our disposal, we may simplify the norms of the expressions in Equation (E.5):

$$\begin{aligned} D(\mathsf{C}, \mathsf{A})^2 = \|\boldsymbol{a} - \boldsymbol{c}\|^2 \;&\propto\; 2\left(1 + \frac{(\boldsymbol{c} + \boldsymbol{d}) \cdot (\boldsymbol{c} - \boldsymbol{d})}{1 - M^2}\right) \\ &= 2\left(\frac{1 - M^2 + \|\boldsymbol{c}\|^2 - \|\boldsymbol{d}\|^2}{1 - M^2}\right) \\ &= 4\frac{1 - \|\boldsymbol{d}\|^2}{1 - M^2}. \end{aligned} \tag{E.8}$$

Since $0 \leq \|\boldsymbol{d}\|^2 \leq 1$, and similarly for $\boldsymbol{c}$, it follows that $M^2 = (\boldsymbol{c} \cdot \boldsymbol{d})^2 \leq \|\boldsymbol{d}\|^2$ via the Cauchy-Schwarz inequality, and so $1 - \|\boldsymbol{d}\|^2 \leq 1 - M^2$. We now introduce the angle φ satisfying

$$\sin\varphi = \sqrt{\frac{1 - \|\boldsymbol{d}\|^2}{1 - M^2}}, \tag{E.9}$$

and therefore

$$\cos\varphi = \sqrt{1 - \frac{1 - \|\boldsymbol{d}\|^2}{1 - M^2}} = \sqrt{\frac{\|\boldsymbol{d}\|^2 - M^2}{1 - M^2}} = \sqrt{\frac{1 - \|\boldsymbol{c}\|^2}{1 - M^2}}. \tag{E.10}$$

As a result, we have

$$\|\boldsymbol{a} - \boldsymbol{c}\| = D(\mathsf{C}, \mathsf{A}) \;\propto\; 2\sin\varphi, \tag{E.11}$$

and so

$$\begin{aligned} \boldsymbol{a} - \boldsymbol{c} &= \frac{\mu D(\mathsf{C}, \mathsf{A})}{2\sin\varphi}\left(\frac{\boldsymbol{c} + \boldsymbol{d}}{\|\boldsymbol{c} + \boldsymbol{d}\|} + \frac{\boldsymbol{c} - \boldsymbol{d}}{\|\boldsymbol{c} - \boldsymbol{d}\|}\right) \\ &= \frac{\mu D(\mathsf{C}, \mathsf{A})}{2(1 - M^2)\sin\varphi}\left((1 - M)(\boldsymbol{c} + \boldsymbol{d}) + (1 + M)(\boldsymbol{c} - \boldsymbol{d})\right) \\ &= \frac{\mu D(\mathsf{C}, \mathsf{A})(\boldsymbol{c} - M\boldsymbol{d})}{(1 - M^2)\sin\varphi}. \end{aligned} \tag{E.12}$$

where $\mu = \pm 1$. In other words,

$$\boldsymbol{a} = \boldsymbol{c} + \frac{\mu D(\mathsf{C}, \mathsf{A})(\boldsymbol{c} - M\boldsymbol{d})}{(1 - M^2)\sin\varphi}, \tag{E.13}$$

and, by a very similar calculation,

$$\boldsymbol{b} = \boldsymbol{d} + \frac{\nu D(\mathsf{D}, \mathsf{B})(\boldsymbol{d} - M\boldsymbol{c})}{(1 - M^2)\cos\varphi}, \tag{E.14}$$

where $\nu = \pm 1$. Note that since $\|\boldsymbol{a} - \boldsymbol{c}\| = D(\mathsf{C}, \mathsf{A})$, it follows that $(1 - M^2)\sin\varphi =$

$\|\boldsymbol{c} - M\boldsymbol{d}\|$, and similarly $(1 - M^2)\cos\varphi = \|\boldsymbol{d} - M\boldsymbol{c}\|$.

In order to calculate the minimum values $D(\mathsf{C}, \mathsf{A})$ and $D(\mathsf{D}, \mathsf{B})$ we utilise the normality of the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.

$$\begin{aligned} 1 = \|\boldsymbol{a}\|^2 &= \|\boldsymbol{c}\|^2 + D(\mathsf{C}, \mathsf{A})^2 + 2\mu D(\mathsf{C}, \mathsf{A}) \frac{\|\boldsymbol{c}\|^2 - M^2}{(1 - M^2)\sin\varphi} \\ &= \|\boldsymbol{c}\|^2 + D(\mathsf{C}, \mathsf{A})^2 + 2\mu D(\mathsf{C}, \mathsf{A}) \frac{1 - \|\boldsymbol{d}\|^2}{(1 - M^2)\sin\varphi} \\ &= \|\boldsymbol{c}\|^2 + D(\mathsf{C}, \mathsf{A})^2 + 2\mu D(\mathsf{C}, \mathsf{A})\sin\varphi \\ &= (D(\mathsf{C}, \mathsf{A}) + \mu\sin\varphi)^2 + \|\boldsymbol{c}\|^2 - \sin^2\varphi \\ &= (D(\mathsf{C}, \mathsf{A}) + \mu\sin\varphi)^2 + \frac{\|\boldsymbol{c}\|^2 (1 - M^2) - 1 + \|\boldsymbol{d}\|^2}{1 - M^2} \\ &= (D(\mathsf{C}, \mathsf{A}) + \mu\sin\varphi)^2 + M^2 \frac{1 - \|\boldsymbol{c}\|^2}{1 - M^2} \\ &= (D(\mathsf{C}, \mathsf{A}) + \mu\sin\varphi)^2 + M^2\cos^2\varphi. \end{aligned} \tag{E.15}$$

In other words,

$$(D(\mathsf{C}, \mathsf{A}) + \mu\sin\varphi)^2 = 1 - M^2\cos^2\varphi \geq \sin^2\varphi, \tag{E.16}$$

where the last inequality arises from $M^2 \leq 1$. Since we require a positive value for $D(\mathsf{C}, \mathsf{A})$, we must take the positive root of the above expression, and so

$$D(\mathsf{C}, \mathsf{A}) = \sqrt{1 - M^2\cos^2\varphi} - \mu\sin\varphi. \tag{E.17}$$

In the exact same way, by starting with $\|\boldsymbol{b}\|^2 = 1$, we arrive at the expression

$$D(\mathsf{D}, \mathsf{B}) = \sqrt{1 - M^2\sin^2\varphi} - \nu\cos\varphi. \tag{E.18}$$

All that remains is to find an expression for M, which we derive by considering $\boldsymbol{a}\cdot\boldsymbol{b} = \cos\theta$, where θ is the angle between the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. In order to simplify our calculation, we first note that

$$(\boldsymbol{c} - M\boldsymbol{d}) \cdot (\boldsymbol{d} - M\boldsymbol{c}) = (1 - \|\boldsymbol{c}\|^2 - \|\boldsymbol{d}\|^2 + M^2) = 0, \tag{E.19}$$

hence

$$\begin{aligned} \cos\theta = \boldsymbol{a}\cdot\boldsymbol{b} &= \left(\boldsymbol{c} + \frac{\mu D(\mathsf{C}, \mathsf{A})(\boldsymbol{c} - M\boldsymbol{d})}{(1 - M^2)\sin\varphi}\right) \cdot \left(\boldsymbol{d} + \frac{\nu D(\mathsf{D}, \mathsf{B})(\boldsymbol{d} - M\boldsymbol{c})}{(1 - M^2)\cos\varphi}\right) \\ &= M\left(1 + \frac{\mu D(\mathsf{C}, \mathsf{A})(1 - \|\boldsymbol{d}\|^2)}{(1 - M^2)\sin\varphi} + \frac{\nu D(\mathsf{D}, \mathsf{B})(1 - \|\boldsymbol{c}\|^2)}{(1 - M^2)\cos\varphi}\right) \\ &= M\left(1 + \mu D(\mathsf{C}, \mathsf{A})\sin\varphi + \nu D(\mathsf{D}, \mathsf{B})\cos\varphi\right). \end{aligned} \tag{E.20}$$

Making use of Equations (E.17) and (E.18), and recalling that $\mu^2 = \nu^2 = 1$, Equation

(E.20) takes the form

$$\begin{aligned}\cos\theta &= M(1+\mu\sin\varphi\sqrt{1-M^2\cos^2\varphi}-\sin^2\varphi+\nu\cos\varphi\sqrt{1-M^2\sin^2\varphi}-\cos^2\varphi)\\ &= M(\mu\sin\varphi\sqrt{1-M^2\cos^2\varphi}+\nu\cos\varphi\sqrt{1-M^2\sin^2\varphi}).\end{aligned} \tag{E.21}$$

Squaring both sides we find

$$\cos^2\theta = M^2\left(1-\tfrac{1}{2}M^2\sin^2 2\varphi+\mu\nu\sin 2\varphi\sqrt{1-M^2+\tfrac{1}{4}M^4\sin^2 2\varphi}\right), \tag{E.22}$$

which by rearranging and squaring both sides again we find

$$\begin{aligned}\left(\cos^2\theta+\tfrac{1}{2}M^4\sin^2 2\varphi-M^2\right)^2 &= \cos^4\theta+\tfrac{1}{4}M^8\sin^4 2\varphi+M^4(1+\cos^2\theta\sin^2 2\varphi)\\ &\quad -2M^2\cos^2\theta-M^6\sin^2 2\varphi\\ &= M^4\sin^2 2\varphi\,(1-M^2+\tfrac{1}{4}M^4\sin^2 2\varphi)\\ &= M^4\sin^2 2\varphi-M^6\sin^2 2\varphi+\tfrac{1}{4}M^8\sin^4 2\varphi.\end{aligned} \tag{E.23}$$

The terms on both sides of order M^6 and higher cancel, and so we rearrange to find

$$\begin{aligned}0 &= \cos^4\theta-2M^2\cos^2\theta+M^4(1-\sin^2\theta\sin^2 2\varphi)\\ &= (1-\sin^2\theta\sin^2 2\varphi)\left(M^2-\frac{\cos^2\theta}{1-\sin^2\theta\sin^2 2\varphi}\right)^2+\cos^4\theta\left(1-\frac{1}{1-\sin^2\theta\sin^2 2\varphi}\right)\\ &= (1-\sin^2\theta\sin^2 2\varphi)\left(M^2-\frac{\cos^2\theta}{1-\sin^2\theta\sin^2 2\varphi}\right)^2-\frac{\cos^4\theta\sin^2\theta\sin^2 2\varphi}{1-\sin^2\theta\sin^2 2\varphi}.\end{aligned} \tag{E.24}$$

At this point we move the second term over, divide through on both sides by $1-\sin^2\theta\sin^2 2\varphi$ and then take the square root, which leads to

$$M^2 = \frac{\cos^2\theta}{1-\sin^2\theta\sin^2 2\varphi}\left(1\pm\sin\theta\sin 2\varphi\right). \tag{E.25}$$

Introducing the quantity $\kappa=\pm 1$, we find the final expression for M^2:

$$M^2 = \frac{\cos^2\theta}{1+\kappa\sin\theta\sin 2\varphi}. \tag{E.26}$$

It should be noted that in the paper by Yu and Oh κ is given as $\mu\nu$, which leads them to quickly assume that the minimum errors are given when $\mu=\nu=1$. However, here we do not see this immediate necessity, and some further analysis is required to reach this conclusion. Leaving κ as undecided for the moment, if we now put the expression for M^2

into Equation (E.17) we find

$$\begin{aligned} D(\mathsf{C},\mathsf{A}) &= \sqrt{1-\frac{\cos^2\theta\cos^2\varphi}{1+\kappa\sin\theta\sin 2\varphi}}-\mu\sin\varphi \\ &= \sqrt{\frac{1+\kappa\sin\theta\sin 2\varphi-\cos^2\theta\cos^2\varphi}{1+\kappa\sin\theta\sin 2\varphi}}-\mu\sin\varphi \\ &= \sqrt{\frac{\sin^2\varphi+\sin^2\theta\cos^2\varphi+2\kappa\sin\theta\sin\varphi\cos\varphi}{1+\kappa\sin\theta\sin 2\varphi}}-\mu\sin\varphi \\ &= \pm\frac{\sin\varphi+\kappa\sin\theta\cos\varphi}{\sqrt{1+\kappa\sin\theta\sin 2\varphi}}-\mu\sin\varphi, \end{aligned} \tag{E.27}$$

where we have made use of the identity $1=\sin^2\varphi+\cos^2\varphi(\cos^2\theta+\sin^2\theta)$. By a similar method we find that Equation (E.18) can be rewritten as

$$D(\mathsf{D},\mathsf{B}) = \pm\frac{\cos\varphi+\kappa\sin\theta\sin\varphi}{\sqrt{1+\kappa\sin\theta\sin 2\varphi}}-\nu\cos\varphi. \tag{E.28}$$

In the case that $\kappa=-1$, the first term of $D(\mathsf{C},\mathsf{A})$ and $D(\mathsf{D},\mathsf{B})$ can go negative and so we must use the absolute values. However, in the case of $\kappa=-1$, if we let μ or ν be equal to 1 then we can again get a negative value, and if we let μ or ν equal to -1, then the value is greater than the case of $\kappa=\mu=\nu=+1$, which is also smaller than $\kappa=-\mu=-\nu=1$. As a result, we conclude that the minimum possible values, whilst still being positive, occurs when $\kappa=\mu=\nu=+1$, and so the minimum error bound is given by

$$\begin{aligned} D(\mathsf{C},\mathsf{A}) &= \frac{\sin\varphi+\sin\theta\cos\varphi}{\sqrt{1+\sin\theta\sin 2\varphi}}-\sin\varphi, \\ D(\mathsf{D},\mathsf{B}) &= \frac{\cos\varphi+\sin\theta\sin\varphi}{\sqrt{1+\sin\theta\sin 2\varphi}}-\cos\varphi. \end{aligned} \tag{E.29}$$

E.2 The optimal approximating observables

We are also able to derive the Bloch vectors of the optimal approximating observables from what we have covered in this Section. The derivations of $\boldsymbol{c}$ and $\boldsymbol{d}$ follow identical steps, and so we shall just focus on the derivation of $\boldsymbol{c}$ here.

We begin by first introducing the shorthands

$$\gamma=\frac{D(\mathsf{C},\mathsf{A})}{(1-M^2)\sin\varphi}, \qquad \delta=\frac{D(\mathsf{D},\mathsf{B})}{(1-M^2)\cos\varphi}, \tag{E.30}$$

thereby allowing us to rewrite Equation (E.13) and (E.14) in the simpler forms

$$\boldsymbol{a}=\boldsymbol{c}+\gamma(\boldsymbol{c}-M\boldsymbol{d}), \tag{E.13'}$$

$$\boldsymbol{b}=\boldsymbol{d}+\delta(\boldsymbol{d}-M\boldsymbol{c}). \tag{E.14'}$$

By rearranging Equation (E.14') we can express $\boldsymbol{d}$ in terms of $\boldsymbol{b}$ and $\boldsymbol{c}$:

$$\boldsymbol{d}=\frac{\boldsymbol{b}+\delta M\boldsymbol{c}}{1+\delta}, \tag{E.31}$$

which we may use in Equation (E.13'),

$$\boldsymbol{a} = \frac{[1+\gamma+\delta+\gamma\delta(1-M^2)]\boldsymbol{c} - \gamma M\boldsymbol{b}}{1+\delta}. \tag{E.32}$$

By rearranging, we arrive at an expression for $\boldsymbol{c}$ in terms of $\boldsymbol{a}$ and $\boldsymbol{b}$:

$$\boldsymbol{c} = \frac{(1+\delta)\boldsymbol{a} + \gamma M\boldsymbol{b}}{1+\gamma+\delta+\gamma\delta(1-M^2)}. \tag{E.33}$$

In order to evaluate the denominator, we note that, using Equations (E.13') and (E.14'),

$$\|\boldsymbol{a}\times\boldsymbol{b}\| = \|\boldsymbol{c}\times\boldsymbol{d}\| \big(1+\gamma+\delta+\gamma\delta(1-M^2)\big), \tag{E.34}$$

and so

$$1+\gamma+\delta+\gamma\delta(1-M^2) = \frac{\|\boldsymbol{a}\times\boldsymbol{b}\|}{\|\boldsymbol{c}\times\boldsymbol{d}\|} = \frac{\sin\theta}{\|\boldsymbol{c}\times\boldsymbol{d}\|}. \tag{E.35}$$

The quantity $\|\boldsymbol{c}\times\boldsymbol{d}\| = \sqrt{\|\boldsymbol{c}\|^2\,\|\boldsymbol{d}\|^2 - M^2}$ can be resolved by noting, via Equations (E.9) and (E.10), that

$$\|\boldsymbol{c}\|^2 = 1-(1-M^2)\cos^2\varphi, \qquad \|\boldsymbol{d}\|^2 = 1-(1-M^2)\sin^2\varphi, \tag{E.36}$$

from which it follows that

$$\begin{aligned}\|\boldsymbol{c}\times\boldsymbol{d}\| &= \sqrt{(1-(1-M^2)\cos^2\varphi)(1-(1-M^2)\sin^2\varphi) - M^2} \\ &= (1-M^2)\sin\varphi\cos\varphi,\end{aligned} \tag{E.37}$$

where positivity is guaranteed from $\varphi\in[0,\pi/2]$. Hence,

$$1+\gamma+\delta+\gamma\delta(1-M^2) = \frac{\sin\theta}{(1-M^2)\sin\varphi\cos\varphi}. \tag{E.38}$$

Making use of this quantity in Equation (E.33) we find that the observable C that minimises the error $D(\mathsf{C},\mathsf{A})$ has the Bloch vector

$$\begin{aligned}\boldsymbol{c} &= \frac{(1-M^2)\sin\varphi\cos\varphi}{\sin\theta}\left(\frac{(1-M^2)\cos\varphi + D(\mathsf{D},\mathsf{B})}{(1-M^2)\cos\varphi}\boldsymbol{a} + \frac{MD(\mathsf{C},\mathsf{A})}{(1-M^2)\sin\varphi}\boldsymbol{b}\right) \\ &= \frac{(D(\mathsf{D},\mathsf{B}) + (1-M^2)\cos\varphi)\sin\varphi\,\boldsymbol{a} + MD(\mathsf{C},\mathsf{A})\cos\varphi\,\boldsymbol{b}}{\sin\theta},\end{aligned} \tag{E.39}$$

and, similarly,

$$\boldsymbol{d} = \frac{(D(\mathsf{C},\mathsf{A}) + (1-M^2)\sin\varphi)\cos\varphi\,\boldsymbol{b} + MD(\mathsf{D},\mathsf{B})\sin\varphi\,\boldsymbol{a}}{\sin\theta}. \tag{E.40}$$

Table of commonly used notation

"x belongs to the set A"	$x \in A$
"x does not belong to A"	$x \notin A$
"B is a subset of A"	$B \subset A$
"C is a set of elements that satisfy property p"	$C = \{x \| x = p\}$
The empty set	$\emptyset = \{\}$
The complement of subset B of A	$B \backslash A$ or $B^c = \{x \in A \| x \notin B\}$
Power set of A	$2^A = \{B \subset A\}$
Union	$A \cup B = \{x \| x \in A$ or $x \in B\}$
Intersection	$A \cap B = \{x \| x \in A$ and $x \in B\}$
Cartesian product of C and D	$C \times D = \{(x, y) \| x \in C, y \in D\}$
"There exists"	$\exists$
"For all"	$\forall$
A function f from A to B	$f : A \to B$
"Maps to"	$\mapsto$
Domain of a function $f : A \to B$	$\mathcal{D}_f = \{a \in A \| \exists b \in B, b = f(a)\}$
Range of a function $f : A \to B$	$\mathcal{R}_f = \{b \in B \| \exists a \in A, b = f(a)\}$
Inverse image of a set $Y \subset B$ with respect to f	$f^{-1}(Y) = \{a \in A \| f(a) \in Y\}$
Composition of functions $f : A \to B$ and $g : B \to C$	$g \circ f : A \to C$, $(g \circ f)(x) = g(f(x))$
Characteristic function for a set A	χ_A, $\chi_A(x) = 1$ if $x \in A$ and 0 otherwise
Hilbert Spaces	$\mathcal{H}$ or $\mathcal{K}$
Vector states in the Hilbert Space $\mathcal{H}$	$\psi, \xi \in \mathcal{H}$
Self-adjoint operators on $\mathcal{H}$	$A, B \in \mathcal{L}_s(\mathcal{H})$
Density operators on $\mathcal{H}$	$\rho, \sigma \in \mathcal{S}(\mathcal{H})$
Unitary operators on $\mathcal{H}$	$U, V \in \mathcal{U}(\mathcal{H})$
POVMs	A, B

Bibliography

[1] N. I. Akhiezer and I. M. Glazman. *Theory of Linear Operators in Hilbert Space.* Dover Books on Mathematics, New York, 1993.

[2] E. Arthurs and M. S. Goodman. *Phys. Rev. Lett.* **60**, 2447 (1988).

[3] E. Arthurs and J. L. Kelly. *Bell Syst. Tech. J.* **44**, 725 (1965).

[4] G. Bachman and L. Narici. *Functional Analysis.* Dover Books on Mathematics, New York, 2000.

[5] R. Beneduci, T. Bullock, P. Busch, C. Carmeli, T. Heinosaari and A. Toigo. *Phys. Rev. A* **88**, 032312 (2013).

[6] I. Bengtsson, W. Bruzda, Å. Ericsson, J.-Å. Larsson, W. Tadej and K. Życzkowski. *J. Math. Phys.* **48**, 052106 (2007).

[7] C. Branciard. *Proc. Nat. Acad. Sci.* **110**, 6742 (2013).

[8] T. Bullock and P. Busch. *Phys. Rev. Lett.* **113**, 120401 (2014).

[9] P. Busch. *Int. J. Theor. Phys.* **24**, 63 (1985).

[10] P. Busch. *Phys. Rev. D* **33**, 2253 (1986).

[11] P. Busch, M. Grabowski and P. Lahti. *Operational Quantum Physics.* Second ed., Springer-Verlag, Berlin, 1997.

[12] P. Busch, T. Heinonen and P. Lahti. *Phys. Lett. A* **320**, 261 (2004).

[13] P. Busch, T. Heinonen and P. Lahti. *J. Phys. Rep.* **452**, 155 (2007).

[14] P. Busch and T. Heinosaari. *Quant. Inf. Comp.* **8**, 0797 (2008).

[15] P. Busch, P. Lahti and P. Mittelstaedt. *The Quantum Theory of Measurement.* Second ed., Springer-Verlag, Berlin, 1996.

[16] P. Busch, P. Lahti and R. F. Werner. *Phys. Rev. Lett.* **111**, 160405 (2013).

[17] P. Busch, P. Lahti and R. F. Werner. *J. Math. Phys.* **55**, 042111 (2014).

[18] P. Busch, P. Lahti and R. F. Werner. *Phys. Rev. A* **89**, 012129 (2014).

[19] P. Busch, P. Lahti and R. F. Werner. *Rev. Mod. Phys.* **86**, 1261 (2014).

[20] P. Busch and N. Stevens. *Phys. Rev. Lett.* **114**, 070402 (2015).

[21] P. Butterley and W. Hall. *Phys. Lett. A* **369**, 5 (2007).

[22] M. Capiński and E. Kopp. *Measure, Integral and Probability.* Second ed., Springer-Verlag, London, 2005.

[23] G. Cassinelli, E. De Vito and A. Toigo. *J. Math. Phys.* **44**, 4768 (2003).

[24] J. B. Conway. *A Course in Functional Analysis.* Springer, New York, 1990.

[25] J. B. Conway. *A Course in Operator Theory.* American Mathematical Society, Providence, Rhode Island, 2000.

[26] E. B. Davies. *Quantum Theory of Open Systems.* Academic Press, London, 1976.

[27] A. Di Lorenzo. *Phys. Rev. Lett.* **110**, 120403 (2013).

[28] M. J. W. Hall. *Phys. Rev. A* **69**, 052113 (2004).

[29] T. Heinosaari and M. Ziman. *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement.* Cambridge University Press, 2011.

[30] A. S. Holevo. *Rep. Math. Phys.* **16**, 385 (1979).

[31] A. S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory.* Scuola Normale Superiore, Pisa, 2011.

[32] I. D. Ivanović. *J. Phys. A* **14**, 3241 (1981).

[33] J. Kiukas, P. Lahti and K. Ylinen. *J. Math. Anal. Appl.* **319**, 783 (2006).

[34] K. Kraus. *States, Effects and Operations.* Springer-Verlag, Berlin, 1983.

[35] R. Lidl and H. Niederreiter. *Introduction to finite fields and their applications.* Cambridge University Press, 1986.

[36] G. Ludwig. *Foundations of Quantum Mechanics I.* Springer-Verlag, New York, 1983.

[37] A. Lund and H. M. Wiseman. *New J. Phys.* **12**, 093011 (2010).

[38] M. A. Naimark. *Dokl. Akad. Nauk SSSR* **41**, 359 (1943).

[39] M. Ozawa. *J. Math. Phys.* **25**, 79 (1984).

[40] M. Ozawa. *Phys. Lett. A* **299**, 1 (2002).

[41] M. Ozawa. *Ann. Phys.* **311**, 350 (2004).

[42] A. E. Rastegin. *Phys. Scr.* **89**, 085101 (2014).

[43] M. Reed and B. Simon. *Methods of Modern Mathematical Physics, Vol I: Functional Analysis.* Academic Press, London, 1981.

[44] J. M. Renes, R. Blume-Kohout, A. J. Scott and C. M. Caves. *J. Math. Phys.* **45**, 2171 (2004).

[45] M. Ringbauer, D. N. Biggerstaff, M. A. Broome, A. Fedrizzi, C. Branciard and A. G. White. *Phys. Rev. Lett.* **112**, 020401 (2014).

[46] L. A. Rozema, A. Darabi, D. H. Mahler, A. Hayat, Y. Soudagar and A. M. Steinberg. *Phys. Rev. Lett.* **109**, 100404 (2012).

[47] W. Rudin. *Functional Analysis.* Second ed., McGraw-Hill, New York, 1991.

[48] A. J. Scott and M. Grassl. *arXiv:* 0910.5784v2 (2009).

[49] D. R. Stinson. *J. Combin. Theory Ser. A* **36**, 373 (1984).

[50] M. H. Stone. *Ann. Math.* **33**, 643 (1932).

[51] W. A. Sutherland. *Introduction to Metric & Topological Spaces.* Second ed., Oxford University Press, 2009.

[52] G. N. M. Tabia and D. M. Appleby. *arXiv:*1304.8075 (2013).

[53] C. Villani. *Optimal Transport: Old and New.* Springer, New York, 2009.

[54] R. F. Werner. *J. Math. Phys.* **25**, 1404 (1984).

[55] W. K. Wootters. *Found. Phys.* **36**, 112 (2006).

[56] W. K. Wootters and B. D. Fields. *Ann. Phys.* **191**, 363 (1989).

[57] S. Yu and C.H. Oh. *arXiv:* 1402.3785 (2014).

[58] G. Zauner. *Int. J. Quant. Inf.* **9**, 445 (2011).