On Compatibility Properties of Quantum Observables represented by Positive Operator-Valued Measures

Bachelorthesis Mathematics
Radboud University Nijmegen

Yari Kraak
Supervisor: Klaas Landsman

July 2018

Abstract

In a rigorous post-von Neumann mathematical formalism of quantum mechanics, observables are represented by normalized positive operator-valued measures instead of self-adjoint operators. In this formalism, the notion of joint measurability of two observables is more complex then in the von Neumann formalism, where observables are jointly measurable if and only if they commute. We look into various notions of compatibility, among which joint measurability, coexistence, commutativity, and joint measurability of binarizations, and investigate for which classes of observables these are equivalent.

Contents

1 Introduction

In the literature, quantum-mechanical observables are usually described by self-adjoint operators, as introduced by von Neumann [20]. But in certain cases there are observables that cannot be fully described by self-adjoint operators [7, 9, 15, 17]. An example of such an observable is the *covariant phase space observable* in a single-mode optical field. Therefore, a new formalism of quantum mechanics has been formulated, where observables are represented by *positive operator-valued measures* (POVMs). This modern formalism of quantum mechanics, sometimes referred to as *operational quantum mechanics* or *quantum measurement theory* (see e.g., [5, 1, 3]) has the means to investigate these special observables. Moreover, it provides a powerful toolset to study sequential, joint, and approximate joint measurements. A subset of these POVMs are the *projection valued measures* (PVMs), which are in one-to-one correspondence to self-adjoint operators.

Measurement of multiple quantum observables is an important part of quantum mechanics. A well-known property, which lies at the heart of quantum mechanics, is the fact that in general two observables cannot be measured together. Following the famous heuristic work of Heisenberg, the mathematical analysis of this subject started with Von Neumann who characterised jointly measurable observables as commuting self-adjoint operators [20]. But when working in the new formalism, observables, represented by POVMs, can be jointly measurable also if they do not commute [16]. This introduced several notions describing the measurement of more than one quantum observable. We distinguish several properties of observables, among which commutativity, joint measurability, coexistence, and joint measurability of binarizations, some of which are easier to verify than others. Under some conditions theses notions are equivalent, but in general they differ.

Recently, for a broad class of observables, namely if one of the observables is extreme and discrete, it has been shown that these observables are coexistent if and only if they are jointly measurable [8]. In this thesis we further investigate the relation between these notions and take a look at the proof of this result. We will also look into the relation between coexistence and joint measurability of binarizations.

In section 2 we start by defining the needed mathematical knowledge about operator-valued measures. In section 3 a introduction into this post-von Neumann formalism is presented. In section 4 we will look into the results about the relations between coexistence, joint measurability, commutativity, and joint measurability of binarizations, based on the works of Lahti [13], Bush et al. [2], Reeb et al. [18], Pellonpää [16] and Haapasalo [8, 6]. We will clarify the proofs from [8] and correct a minor mistake in it.

We assume basic knowledge about functional analysis, measure theory and quantum mechanics, but in the appendix an introduction into most of the important mathematical concepts is given.

2 Positive Operator-Valued Measures

Before we can introduce the new quantum-mechanical formalism, we first have to introduce some mathematics. Some basic knowledge about functional analysis and measure theory will be assumed. In appendix A the important functional analytic concepts are introduced. Appendix B can be used as a reference for the measure theoretic definitions. Throughout this thesis$\mathcal{H}$ will be a separable Hilbert space and we will write $\mathcal{B}(\mathcal{H})$ for the set of bounded linear operators on $\mathcal{H}$. We use the symbol $\langle \cdot \mid \cdot \rangle$ for the inner product on any Hilbert space. By convention, we choose the inner product to be linear in its second argument. Finally, $\mathbb{N} = \{1, 2, 3, \ldots\}$.

We will now define the notion of an operator-valued measure, which will play a central role in the rest of this thesis. Instead of a positive real-valued measure $\mu : \mathcal{A} \to [0, \infty)$, as usual, we will define a *positive operator-valued measure* $E : \mathcal{A} \to \mathcal{B}(\mathcal{H})^+$.

Definition 2.1. Let $\mathcal{H}$ be a Hilbert space and $(\Omega, \mathcal{A})$ a measurable space.

(a) A function $E : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ is called a *positive operator-valued measure* (POVM) if the next two conditions hold:

  (i) $E(X) \geq 0$ for all $X \in \mathcal{A}$,

  (ii) E is weakly σ-additive, that is, for $\{X_n\} \subseteq \mathcal{A}$, a countable family of disjoint sets,

$$E\left(\bigcup_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} E(X_n),$$

  where the series on the right-hand side converges in the weak operator topology.[1]

(b) A POVM $E : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ is said to be *normalized* if $E(\Omega)$ equals the identity operator on $\mathcal{H}$. A normalized POVM is called a *semispectral measure.*

(c) A POVM $E : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ is called a *projection valued measure* (PVM) if $E(X)$ is an orthogonal projection on $\mathcal{H}$ for each $X \in \mathcal{A}$.

(d) A normalized PVM is called a *spectral measure.*

We will give some examples of POVMs.

Example 2.2. (a) Consider the Borel-measurable space $(\mathbb{R}, \mathcal{F}(\mathbb{R}))$ and let $\mathcal{H} = \mathrm{L}^2(\mathbb{R})$. For any measurable function $f : \mathbb{R} \to \mathbb{C}$, we define the multiplication operator M_f by

$$(M_f\phi)(x) = (f\phi)(x) = f(x)\phi(x) \quad \forall x \in \mathbb{R}, \phi \in \mathcal{D}(M_f) = \{\phi \in \mathrm{L}^2(\mathbb{R}) \mid f\phi \in \mathrm{L}^2(\mathbb{R}).\}$$

One can simply check that for $X \subseteq \mathbb{R}$, M_{χ_X} is a projection and thus we can define the PVM $M_\chi : \mathcal{F}(\mathbb{R}) \to \mathrm{L}^2(\mathbb{R})$ by

$$M_\chi : X \mapsto M_{\chi_X}.$$

Because $M_{\chi_\mathbb{R}}$ is the identity operator on $\mathrm{L}^2(\mathbb{R})$ we even see that M_χ is a spectral measure. This M_χ is called the *canonical spectral measure* in $\mathrm{L}^2(\mathbb{R})$.

[1] That is $\forall \phi, \psi \in \mathcal{H}$, $\left\langle \phi \,\middle|\, \left(\sum_{n=1}^{N} E(X_n)\right) \psi \right\rangle \to \left\langle \phi \,\middle|\, E\left(\bigcup_{n=1}^{\infty} X_n\right) \psi \right\rangle$, as $N \to \infty$.

(b) For our second example, consider the measurable space $(\{1,2,3\}, \mathcal{P}(\{1,2,3\}))$ and the Hilbert space $\mathcal{H} = \mathbb{C}^2$. Denote the standard orthogonal basis on $\mathbb{C}^2$,

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\},$$

by $\{|1\rangle, |2\rangle\}$. Operators on $\mathbb{C}^2$ are given by complex-valued two-by-two matrices, that is $\mathcal{B}(\mathbb{C}^2) = M_2(\mathbb{C})$. First, we define three operators,

$$\begin{aligned} A_1 &= \tfrac{1}{3} |1\rangle\langle 1| = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & 0 \end{pmatrix}, \\ A_2 &= \tfrac{1}{3} |2\rangle\langle 2| = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{3} \end{pmatrix}, \\ A_3 &= \boldsymbol{I} - A_1 - A_2 = \begin{pmatrix} \frac{2}{3} & 0 \\ 0 & \frac{2}{3} \end{pmatrix}, \end{aligned}$$

which are, as one can easily verify, positive and bounded by the identity. We can now construct a POVM $A : \mathcal{P}(\{1,2,3\}) \to M_2(\mathbb{C}^2)$ by

$$A(\{1\}) = A_1,\ A(\{2\}) = A_2,\ A(\{3\}) = A_3,$$

and the other outcomes of A follow by additivity. For example,

$$E(\{1,2\}) + E(\{1\}) + E(\{2\}) = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}.$$

Since $A(\{1,2,3\}) = \boldsymbol{I}$, we see that A is a semispectral measure, but because the operators appearing here are not projections, e.g.

$$A_1^2 = \begin{pmatrix} \frac{1}{9} & 0 \\ 0 & 0 \end{pmatrix} \neq A_1,$$

it follows that A is not a spectral measure. $\triangleleft$

We now give some standard results on POVMs.

Proposition 2.3. *Let $E : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ be a POVM. Then the following hold:*

(a) Let $X, Y \in \mathcal{A}$ such that $X \subseteq Y$. Then

$$\begin{aligned} E(Y \setminus X) &= E(Y) - E(X) \text{ and} \\ E(X) &\leq E(Y). \end{aligned}$$

(b) For any $X, Y \in \mathcal{A}$,

$$E(X \cup Y) + E(X \cap Y) = E(X) + E(Y).$$

(c) If $E(X) = E(X)^*$ or $\|E(X)\| \leq 1$ for all $X \in \mathcal{A}$, then the following are equivalent:

(i) $E(X \cap Y) = E(X)E(Y)\ \forall X, Y \in \mathcal{A}$;

(ii) $E(X) \in \mathbf{P}(\mathcal{H})\ \forall X \in \mathcal{A}$.

Proof. (a) Since for $X \subseteq Y$ one has $Y = X \cup (Y \setminus X)$, we have

$$E(Y \setminus X) = E(X) + E(Y \setminus X),$$

by additivity of E. And since E is positive, the second result follows immediately.

(b) For sets $X, Y \in \mathcal{A}$ one has

$$\begin{aligned} X \cup Y &= (X \setminus Y) \cup (X \cap Y) \cup (Y \setminus X), \\ X &= (X \setminus Y) \cup (X \cap Y), \\ Y &= (Y \setminus X) \cup (X \cap Y), \end{aligned}$$

and the result follows from the additivity of E.

(c) If (i) holds, then $E(X)^2 = E(X \cap X) = E(X)$, and combined with either $E(X) = E(X)^*$ or $\|E(X)\| \leq 1$ and with Proposition A.16, we conclude that $E(X)$ is a projection. Now if (ii) holds, then

$$E(X \cap Y) \leq E(X) \leq E(X \cup Y)$$

by (a), and so $E(X)E(X \cap Y) = E(X \cap Y)$ and $E(X)E(X \cup Y) = E(X)$ by Proposition A.18. Multiplying the equation of (b) by $E(X)$ from the left now yields

$$E(X) + E(X \cup Y) = E(X) + E(X)E(Y),$$

implying (i).

□

Definition 2.4 (POV Bimeasure)**.** A positive operator-valued function $B : \mathcal{A}_1 \times \mathcal{A}_2 \to \mathcal{B}(\mathcal{H})$ is called a *positive operator-valued bimeasure*, if for all $X \in \mathcal{A}_1, Y \in \mathcal{A}_2$ the functions

$$\begin{aligned} \mathcal{A}_2 \ni Y' &\mapsto B(X, Y'), \\ \mathcal{A}_1 \ni X' &\mapsto B(X', Y), \end{aligned}$$

are POVMs. If $B(\Omega_1, \Omega_2) = \boldsymbol{I}$, we call B a *semispectral bimeasure.*

2.1 Naimark dilation theorem

Now we will prove the important result that any semispectral measure $E : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ can be written as a dilation of a spectral measure $F : \mathcal{A} \to B(\mathcal{K})$ by an isometry, namely as $E(X) = J^* F(X) J \quad \forall X \in \mathcal{A}$. This result is called the *Naimark dilation theorem.* There are many ways to prove this theorem, and we follow the measure-theoretic proof in [12], since it is hands-on and direct. We first need to the define the concept of a *positive sesquilinear form measure* (PSFM), which is closely related to a POVM. For the rest of this section, let V be a vector space. We will write $\mathcal{S}(V)$ for the set of all sesquilinear forms on V and $\mathcal{S}^+(V)$ for the set of all positive sesquilinear forms.

Definition 2.5 (Positive sesquilinear form measure)**.** Let $(\Omega, \mathcal{A})$ a measurable space.

(a) Let $S : \mathcal{A} \to \mathcal{S}(V)$ be a function and write $S(X) = S_X$ for $X \in \mathcal{A}$. S is called a *positive sesquilinear form measure* (PSFM) if:

(i) $S(\mathcal{A}) \subseteq \mathcal{S}^+(V)$;

(ii) For all $\phi, \psi \in V$ the mapping $X \mapsto S_X(\phi, \psi)$ is σ-additive, i.e., a complex measure.

(b) A PSFM $S : \mathcal{A} \to \mathcal{S}^+(V)$ is said to be *strict* if $S_\Omega(\phi, \phi) > 0$ for all $\phi \in V \setminus \{0\}$.

Remark. If we have a POVM E, we can naturally identify it with a PSFM S by setting $S_X(\phi, \psi) = \langle \phi \,|\, E(X)\psi \rangle$. Using this identification, we will prove the Naimark dilation theorem first for PSFM's and immediately obtain the result for POVM's.

For the rest of this section we assume that $(\Omega, \mathcal{A})$ is a measurable space, that V is a vector space that has a countable infinite Hamel basis $(e_n)_{n=1}^\infty$. Let $\mathcal{F}$ denote the vector space of $\mathcal{A}$-simple V-valued functions on Ω, i.e. if $f \in \mathcal{F}$ we can write

$$f = \sum_{i=1}^{N} \phi_i \chi_{A_i},$$

with $\phi_i \in V, A_i \in \mathcal{A}\ \forall i \in \mathbb{N}$ and $N \in \mathbb{N}$. Here we write $\chi_A \phi$ for the function $x \mapsto \chi_A(x)\phi$, for $A \in \mathcal{A}$ and $\phi \in V$.

Theorem 2.6. *Let $S : \mathcal{A} \to \mathcal{S}^+(V)$ be a PSFM. Then there exists a Hilber space $\mathcal{K}$, a spectral measure $F : \mathcal{A} \to \mathcal{K}$, and a linear map $J : \mathcal{H} \to \mathcal{K}$ such that*

$$\langle J\phi \,|\, F(X)J\psi \rangle = S_X(\phi, \psi) \quad \forall X \in \mathcal{A}, \phi, \psi \in V.$$

Moreover, the linear span of the set $\{F(X)J\phi \mid X \in \mathcal{A}, \phi \in V\}$ is dense in $\mathcal{K}$.

Proof. Assume that $S : \mathcal{A} \to \mathcal{S}^+(V)$ is a PSFM. We now fix a sequence of positive numbers $(\alpha_i)_{i \in \mathbb{N}}$, such that $\sum_{n=1}^\infty \alpha_n < \infty$, and for $X \in \mathcal{A}$ we write

$$\mu(X) = \sum_{n=1}^{\infty} \alpha_n \frac{S_X(e_n, e_n)}{1 + S_\Omega(e_n, e_n)}.$$

Then μ is a finite positive measure, and with the Cauchy-Schwarz inequality we see that for $X \in \mathcal{A}$, $\mu(X) = 0$ if and only if we have $S_X(\phi, \psi) = 0$ for all $\phi, \psi \in V$. For any $\phi, \psi \in V$, with the Radon-Nikodým theorem (Theorem B.6) we can find a $C(\phi, \psi) \in \mathrm{L}^1(\Omega, \mathcal{A}, \mu)$ such that

$$S_X(\phi, \psi) = \int_X C(\phi, \psi) \,\mathrm{d}\mu \quad \forall X \in \mathcal{A}.$$

Clearly, the mapping $C : V \times V \to \mathbb{C}$ is sesquilinear, and for all $\phi \in V$ we have that $C(\phi, \phi) \geq 0$ μ-almost everywhere.

Assume f $= \sum_{i=1}^n \phi_i \chi_{A_i}$ and $g = \sum_{j=1}^m \psi_j \chi_{B_j}$ in $\mathcal{F}$, such that all the ϕ_i are distinct and all A_i pairwise disjoint. We can now define the unique positive sesquilinear form $\theta : \mathcal{F} \times \mathcal{F} \to \mathbb{C}$ by

$$\theta(f, g) = \sum_{i=1}^{n} \sum_{j=1}^{m} \int_{A_i \cap B_i} C(\phi_i, \psi_i) \,\mathrm{d}\mu,$$

since every $f \in \mathcal{F}$ can uniquely be written in such a way. Then θ almost is an inner product on $\mathcal{F}$, but it is not strict. To accomplish this, we write $\mathcal{N} = \{f \in \mathcal{F} \mid \theta(f, f) = 0\}$ and by the Cauchy-Schwarz inequality we see that $\mathcal{N}$ is a vector subspace of $\mathcal{F}$. We can now define a inner product on the quotient space $\mathcal{F}/\mathcal{N}$ via

$$\langle [f] \,|\, [g] \rangle = \theta(f, g),$$

where we have written $[f] = f + \mathcal{N}$. We denote the Hilbert space completion of this inner product space by $\mathcal{K}$ and call it the *associated Hilbert space* of S, relative to the basis (e_n) and sequence (α_n). For every $X \in \mathcal{A}$, $f \in \mathcal{F}$ we define the function $F_0 : \mathcal{F}/\mathcal{N} \to \mathcal{K}$ by

$$F_0(X)[f] = [\chi_X f].$$

For $g \in [f]$ we have

$$\|[\chi_X f - \chi_X g]\|^2 = \theta(\chi_X f - \chi_X g, \chi_X f - \chi_X g) \leq \theta(f - g, f - g) = 0,$$

and hence this definition for $F_0(X)$ is sound. Furthermore, because

$$\|F_0(X)[f]\|^2 = \langle F_0(X)[f] \,|\, F_0(X)[f] \rangle \leq \langle [f] \,|\, [f] \rangle = \|[f]\|^2,$$

we see that $F_0(X)$ is bounded and thus uniquely extends to a bounded linear operator $F(X) : \mathcal{K} \to \mathcal{K}$. We see that

$$\langle [\phi \chi_A] \,|\, F(X)[\psi \chi_B] \rangle = \int_{A \cap B \cap X} C(\phi, \psi) \, \mathrm{d}\mu \quad \forall A, B \in \mathcal{A}, \phi, \psi \in V,$$

and it can easily be verified that

$$F(X)^2 = F(X)^* = F(X),$$

for all $X \in \mathcal{A}$. To see that $X \mapsto F(X)$ is a spectral measure on $\mathcal{A}$, we first notice that $F(\Omega) = \boldsymbol{I}$. And since F is bounded, for weak σ-additivity it is sufficient to show that it is weakly σ-additive on a dense subset of $\mathcal{K}$. And since by definition of $\mathcal{K}$ the linear span of $\{[\phi \chi_A] \mid \phi \in V, A \in \mathcal{A}\}$ is dense in $\mathcal{K}$, we verify this by noting that

$$\mathcal{A} \ni X \mapsto \langle [\phi \chi_A] \,|\, F(X)[\psi \chi_B] \rangle = \int_{A \cap B \cap X} C(\phi, \psi) \, \mathrm{d}\mu$$

is σ-additive for all $A, B \in \mathcal{A}, \phi, \psi \in V$. Finally, we define the linear map $J : V \to \mathcal{K}$ by

$$J\phi = [\phi \chi_\Omega].$$

Now we see that for $\phi, \psi \in V, X \in \mathcal{A}$ we have

$$\langle J\phi \,|\, F(X) J\psi \rangle = \langle [\phi \chi_\Omega] \,|\, F(X)[\psi \chi_\Omega] \rangle = \int_X C(\phi, \psi) \, \mathrm{d}\mu = S_X(\phi, \psi).$$

Furthermore, by construction,

$$\operatorname{span}\{F(X) J\phi \mid X \in \mathcal{A}, \phi \in V\} = \operatorname{span}\{[\phi \chi_X] \mid X \in \mathcal{A}, \phi \in V\}$$

is dense in $\mathcal{K}$. $\square$

Proposition 2.7. *Let $S : \mathcal{A} \to \mathcal{S}^+(V)$ be a PSFM. The representation of S by the Hilbert space $\mathcal{K}$, the spectral measure $F : \mathcal{A} \to \mathcal{K}$ and the linear map $J : V \to \mathcal{K}$ from Theorem 2.6 is essentially unique. That is, if the triple $(\mathcal{K}', F', J')$ gives another representation with these properties, then there is a unique unitary map $U : \mathcal{K} \to \mathcal{K}'$ such that*

$$UF(X)J\phi = F'(X)J'\phi \ \forall X \in \mathcal{A}, \phi \in V.$$

In particular, we have

$$UJ\phi = J'\phi \ \forall \phi \in V,$$
$$UF(X) = F'(X)U \ \forall X \in \mathcal{A}.$$

Proof. We define $U_0 : \mathrm{span}\{F(X)J\phi \mid X \in \mathcal{A}, \phi \in V\} \to \mathcal{K}'$ by

$$U_0 F(X)J\phi = F'(X)J'\phi.$$

And since U_0 is defined on a dense subspace of $\mathcal{K}$, we can uniquely extend it to a bounded operator $U : \mathcal{K} \to \mathcal{K}'$. And because if $X_1, \ldots, X_n \in \mathcal{A}$ and $\phi_1, \ldots, \phi_n \in V$, we have

$$\begin{aligned} \left\| \sum_{i=1}^{n} F(X_i)J\phi_i \right\|^2 &= \sum_{i,j=1}^{n} \langle F(X_i)J\phi_i \,|\, F(X_j)J\phi_j \rangle = \sum_{i,j=1}^{n} \langle J\phi_i \,|\, F(X_i \cap X_j)J\phi_j \rangle \\ &= \sum_{i,j=1}^{n} S_{X_i \cap X_j}(\phi_i, \phi_j) = \left\| \sum_{i=1}^{n} F'(X_i)J'\phi_i \right\|^2, \end{aligned}$$

and we see that U is an isometry. Thus there is a well-defined isometry sending each $\sum_{i=1}^{n} F(X_i)J\phi_i$ to $\sum_{i=1}^{n} F'(X_i)J'\phi_i$, and this map extends by continuity to a unitary $U : \mathcal{K} \to \mathcal{K}'$. In particular,

$$UJ\phi = UF(\Omega)J\phi = F'(\Omega)J'\phi = J'\phi \ \forall \phi \in V.$$

Moreover, for all $X, Y \in \mathcal{A}$ and $\phi \in V$ we have

$$\begin{aligned} UF(X)F(Y)J\phi &= UF(X \cap Y)J\phi = F'(X \cap Y)J'\phi \\ &= F'(X)F'(Y)J'\phi = F'(X)UF(Y)J\phi, \end{aligned}$$

from which $UF(X) = F'(X)U$ immediately follows. □

We can now almost define the Naimark dilation of a semispectral measure. Let $(e_n)_{n=0}^{\infty}$ be an orthonormal basis of $\mathcal{H}$ and $V = \mathrm{span}\{e_n\}_{n=0}^{\infty}$ its finite linear span. Now assume that $E : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ is a semispectral measure and let $S : \mathcal{A} \to \mathcal{S}^+(V)$ be the corresponding PSFM, defined by

$$S_X(\phi, \psi) = \langle \phi \,|\, E(X)\psi \rangle,$$

for $X \in \mathcal{A}, \phi, \psi \in V$. Now let the triple $(\mathcal{K}, F, J)$ be as in Theorem 2.6. Now for all $\phi, \psi \in V, X \in \mathcal{A}$ we have

$$\langle \phi \,|\, E(X)\psi \rangle = S_X(\phi, \psi) = \langle J\phi \,|\, F(X)J\psi \rangle = \langle \phi \,|\, J^* F(X)J\psi \rangle.$$

In this case $J : V \to \mathcal{K}$ is an isometry, since, using the notation of Theorem 2.6, we have

$$\begin{aligned}\|J\phi\|^2 &= \|[\phi\chi_\Omega]\|^2 = \theta(\phi\chi_\Omega, \phi\chi_\Omega) \\ &= \int_\Omega C(\phi,\phi)\,\mathrm{d}\mu = S_\Omega(\phi,\phi) = \langle \phi \,|\, E(\Omega)\phi\rangle = \|\phi\|^2,\end{aligned}$$

since E is a semispectral measure. All constructions above have taken place in the vector space V, but since it is dense in $\mathcal{H}$ we can extend all constructions to the whole space $\mathcal{H}$. For the details we refer to [12]. Doing this, we conclude this section with our desired result, for which the proof follows from Theorem 2.6, Proposition 2.7 and our discussion above.

Theorem 2.8 (Minimal Diagonal Naimark Dilation)**.** *Let $E : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ be a semispectral measure.*

(a) There exists a Hilbert space $\mathcal{K}$, an isometry $J : \mathcal{H} \to \mathcal{K}$, and a spectral measure $F : \mathcal{A} \to B(\mathcal{K})$ such that

$$E(X) = J^*F(X)J \quad \forall X \in \mathcal{A}.$$

(b) The linear span of $\{F(X)J\phi \;\mid X \in \mathcal{A}, \phi \in \mathcal{H}\}$ is dense in $\mathcal{K}$.

(c) The triple $(\mathcal{K}, F, J)$ is essentially unique. That is if the triple $(\mathcal{K}', F', J')$ gives another representation with the properties from (a) and (b), then there is a unique unitary map $U : \mathcal{K} \to \mathcal{K}'$ such that

$$UF(X)J\phi = F'(X)J'\phi \;\forall X \in \mathcal{A}, \phi \in \mathcal{H}.$$

In particular, we have

$$\begin{aligned}UJ\phi &= J'\phi \;\forall \phi \in \mathcal{H}, \\ UF(X) &= F'(X)U \;\forall X \in \mathcal{A}.\end{aligned}$$

(d) F is the canonical spectral measure on $\mathcal{K}$, i.e. $\forall \psi \in \mathcal{K}$ and μ-almost all $\omega \in \Omega$ we have

$$(F(X)\psi)(\omega) = \chi_X(\omega)\psi(\omega).$$

Using the notation from the theorem above, we call the triple $(\mathcal{K}, F, J)$ a *Naimark representation* of a semispectral measure E if it satisfies (a) of the theorem. We call this dilation *minimal* if the linear span of $\{F(X)J\phi \;\mid X \in \mathcal{A}, \phi \in \mathcal{H}\}$ is dense in $\mathcal{K}$ and, a *diagonal minimal Naimark representation* if it is a minimal Naimark representation and F is the canonical spectral measure on $\mathcal{K}$. Theorem 2.8 thus states that every semispectral measure has a minimal diagonal Naimark representation. Not all Naimark representations are minimal, but these are the one most helpful to us, and thus we will not look into non-minimal Naimark representations.

3 States, Effects and Observables

In this section we define the notions of states, effects, and observables, which are the central concepts in the formalism of quantum measurement. A good understanding of functional analysis, in particular positive and traceclass operators, is necessary to comprehend these definitions. Also, convex sets and the boundary of convex sets are used throughout this section. In Appendices A and C, an introduction to these concepts is given.

Definition 3.1 (State space)**.** Let $\mathcal{H}$ be the Hilbert space associated with some quantum mechanical system. A *density operator* ρ of the quantum system is defined as a positive trace class operator ρ on $\mathcal{H}$ of trace one. The set of all density operators $\mathbf{S}(\mathcal{H})$ is therefore given by

$$\mathbf{S}(\mathcal{H}) = \{\rho \in \mathcal{B}_1(\mathcal{H}) \mid \rho \geq 0, \operatorname{tr}(\rho) = 1\}.$$

In this formalism, we identify *states* of the quantum mechanical system with *density operators.*

Remark. Sometimes another definition of a state is used. Namely, a state can also be defined as a linear map $\omega : \mathcal{B}(\mathcal{H}) \to \mathbb{C}$, satisfying

1. $\omega(A) \geq 0 \quad \forall A \in \mathcal{B}(\mathcal{H})^+$;
2. $\omega(\boldsymbol{I}) = 1$.

It can be shown that for a finite-dimensional Hilbert space $\mathcal{H}$ there is a one-to-one correspondence between states and density operators. But if $\mathcal{H}$ is infinite-dimensional, the set of states is strictly larger than the set of density operators; states that are in one-to-one correspondence with density operators are called *normal states.* In most systems non-normal states represent non-physical situations and can therefore be disregarded. For the rest of this thesis we will only consider normal states, which we represent by density operators. For more information about the difference between normal states and non-normal states, we refer to [14, §4.2].

Proposition 3.2. $\mathbf{S}(\mathcal{H})$ *is a convex subset of* $\mathcal{B}(\mathcal{H})$.

Proof. Let $\rho_1, \rho_2 \in \mathbf{S}(\mathcal{H})$ and $\lambda \in (0, 1)$ arbitrary. It is obvious that, since λ is positve, $\lambda\rho_1 + (1 - \lambda)\rho_2$ is positive and also

$$\operatorname{tr}(\lambda\rho_1 + (1 - \lambda)\rho_2) = \lambda \operatorname{tr}(\rho_1) + (1 - \lambda) \operatorname{tr}(\rho_2) = \lambda + (1 - \lambda) = 1,$$

using the linearity of the trace, and so $\lambda\rho_1 + (1 - \lambda)\rho_2 \in \mathbf{S}(\mathcal{H})$. □

This convex structure of $\mathbf{S}(\mathcal{H})$ reflects the physical possibility of mixing states and the extreme elements $\operatorname{ex}(\mathbf{S}(\mathcal{H}))$ are those states that cannot be realised by mixing states. These extreme elements are called the *pure states.*

Proposition 3.3. *The set of pure states of* $\mathbf{S}(\mathcal{H})$ *is equal to the set of one-dimensional projections on* $\mathcal{H}$*,* $\mathbf{P}_1(\mathcal{H})$*, that is*

$$\operatorname{ex}(\mathbf{S}(\mathcal{H})) = \mathbf{P}_1(\mathcal{H}).$$

Proof. Let $\rho \in \operatorname{ex}(\mathbf{S}(\mathcal{H}))$. By the spectral theorem for a compact self-adjoint operator, we know that we can write ρ as a weighted sum of one-dimensional projections:

$$\rho = \sum_n \lambda_n \left|\phi_n\right\rangle\!\left\langle\phi_n\right|,$$

where $(\phi_n)_{n\in\mathbb{N}}$ is an orthogonal sequence of eigenvectors and $(\lambda_n)_{n\in\mathbb{N}}$ is the set of corresponding eigenvalues. But because ρ is extreme, λ_n is zero for every n except one, and thus ρ is a one-dimensional projection. For the converse implication, assume $\rho \in \mathbf{P}_1(\mathcal{H})$, but now the only decomposition of ρ is the trivial decomposition $\rho = |\psi\rangle\langle\psi|$, for a normalized $\psi \in \mathcal{H}$, and we conclude that ρ is pure. □

Definition 3.4 (Effects)**.** A positive operator on $\mathcal{H}$, bounded by the identity operator, is called an *effect operator.* The set of all effect operators $\mathbf{E}(\mathcal{H})$ is given by

$$\mathbf{E}(\mathcal{H}) = \{A \in \mathcal{B}(\mathcal{H}) \mid 0 \leq A \leq \boldsymbol{I}\}.$$

Proposition 3.5. $\mathbf{E}(\mathcal{H})$ *is a convex subset of* $\mathcal{B}(\mathcal{H})$.

Proof. If $E_1, E_2 \in \mathbf{E}(\mathcal{H})$ and $t \in (0,1)$, positivity of $tE_1 + (1-t)E_2$ follows immediately. And from $\boldsymbol{I} - E_1 \geq 0$ and $\boldsymbol{I} - E_2 \geq 0$ if follows that

$$0 \leq t(\boldsymbol{I} - E_1) + (1-t)(\boldsymbol{I} - E_2) = \boldsymbol{I} - (tE_1 + (1-t)E_2),$$

and thus $tE_1 + (1-t)E_2 \leq \boldsymbol{I}$. □

We say that an effect E is *extreme* if it is an element of the extreme boundary of $\mathbf{E}(\mathcal{H})$.

Theorem 3.6. *The set of extreme effects is equal to the set set of projections, i.e.,*

$$\mathrm{ex}(\mathbf{E}(\mathcal{H})) = \mathbf{P}(\mathcal{H})$$

Proof. From the definition of a projection operator it is immediately clear that any projection is an effect operator. Now let $P \in \mathbf{P}(\mathcal{H})$, and let $A, B \in \mathbf{E}(\mathcal{H}), t \in (0,1)$ such that $P = tA + (1-t)B$. Choose $\phi \in \mathcal{H}$ such that $P\phi = 0$ (such ϕ can always be found). Then

$$\begin{aligned} 0 &= \|tA\phi + (1-t)B\phi\| \leq \left\|t\sqrt{A}\sqrt{A}\phi\right\| \leq t\left\|\sqrt{A}\phi\right\| \quad (\text{since } \left\|\sqrt{A}\right\|^2 \leq 1) \\ &= t\langle\phi \mid A\phi\rangle \leq t\langle\phi \mid A\phi\rangle + (1-t)\langle\phi \mid B\phi\rangle \\ &= \langle\phi \mid P\phi\rangle = 0, \end{aligned}$$

which implies $A\phi = 0$. Let now $\psi \in \mathcal{H}$ be such that $P\psi = \psi$, that is $(\boldsymbol{I} - P)\psi = 0$, but since $\boldsymbol{I} - P = t(\boldsymbol{I} - A) + (1-t)(\boldsymbol{I} - B)$, one also has $(\boldsymbol{I} - A)\psi = 0$. And since $\mathcal{H} = \ker(P) \oplus \mathrm{ran}(P)$ we conclude $A = P$, from which it immediately follows that $B = P$ and that P is an extreme effect.

Now consider an $A \in \mathbf{E}(\mathcal{H})$ that is not a projection. Then there is an $a \in \sigma(A) \cap (0,1)$. Let $f(x)$ be a continuous function on $[0,1]$ such that $0 \leq x \pm f(x) \leq 1 \ \forall x \in [0,1]$ with $f(a) \neq 0$. From the spectral theorem, Theorem A.44, we see that both $A_1 = A + f(A)$ and $A_2 = A - f(A)$ are effects and $A_1 \neq A \neq A_2$, but $A = \frac{1}{2}A_1 + \frac{1}{2}A_2$, from which we conclude that A is not an extreme effect. □

We have the following characterisation for effect operators.

Proposition 3.7. *For a positive operator* $E \in \mathcal{B}(\mathcal{H})^+$ *we have*

$$E \in \mathbf{E}(\mathcal{H}) \quad \Longleftrightarrow \quad \|E\| \leq 1 \quad \Longleftrightarrow \quad \mathrm{r}(E) \leq 1.$$

Proof. This is an immediate consequence of Corollary A.18.1 and Corollary A.38.1. □

Instead of identifying observables with self-adjoint operators as in the von-Neumann formalism, we identify observables with normalised POVMs .

Definition 3.8 (Observables)**.** For a measurable space $(\Omega, \mathcal{A})$ and a Hilbert space $\mathcal{H}$, the set of all observables $\mathbf{O}(\Omega, \mathcal{A}, \mathcal{H})$ is given by

$$\mathbf{O}(\Omega, \mathcal{A}, \mathcal{H}) = \{E : \mathcal{A} \to \mathbf{E}(\mathcal{H}) \mid E \text{ is a semispectral measure}\}.$$

The measurable space $(\Omega, \mathcal{A})$ is called the *outcome space* of E. A semispectral bimeasure is called a *bi-observable.*

Since an observable E is a normalised positive operator-valued measure, $E(X)$ is an effect operator for every $X \in \mathcal{A}$. Using this, we can relate observables to measurements by interpretating $\operatorname{tr}(\rho E(X))$ as a probability. Indeed, if $E(X) \in \mathbf{E}(\mathcal{H})$ and $\rho \in \mathbf{S}(\mathcal{H})$, we see from Proposition A.31 that $\rho A \in \mathcal{B}_1(\mathcal{H})$, and because $E(X)$ and ρ are positive, from Proposition A.34 we see that $\operatorname{tr}(\rho E(X)) \geq 0$. Using Proposition A.33 we conclude $\operatorname{tr}(\rho E(X)) \leq \|E(X)\| \operatorname{tr}(\rho) \leq 1$, and hence we have $0 \leq \operatorname{tr}(\rho E(X)) \leq 1$.

Definition 3.9 (Probability interpretation)**.** For a system prepared in a state ρ, we define the probability $p^E_\rho(X)$ to obtain an outcome in the set $X \in \mathcal{A}$ when the observable $E \in \mathbf{O}(\Omega, \mathcal{A}, \mathcal{H})$ is measured as

$$p^E_\rho(X) = E_\rho(X) = \operatorname{tr}(\rho E(X)).$$

We thus see that p^E_ρ or equivalently E_ρ is a probability measure on $\mathcal{A}$.

These outcome probabilities are to be interpreted as follows. If for a system in state ρ, N times a measurement of E is performed, and if the result measured is $N(X)$ times in the set X, then we have for large N

$$p^E_\rho(X) \approx \frac{N(X)}{N}.$$

And for larger N the measured probability approaches the predicted probability closer.

Example 3.10. Consider Example 2.2 (b), where the POVM $A : \mathcal{P}(\{1, 2, 3\}) \to M_2(\mathbb{C}^2)$ is defined. We can define a state ρ on $\mathbb{C}^2$ by giving a positive trace-one matrix, for example:

$$\rho = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{2}{3} \end{pmatrix}.$$

We can now calculate the probability to get the outcome 1, when measuring A for the state ρ:

$$p^A_\rho(\{1\}) = \operatorname{tr}(\rho A_1) = \operatorname{tr}\left(\begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{2}{3} \end{pmatrix}\begin{pmatrix} \frac{1}{3} & 0 \\ 0 & 0 \end{pmatrix}\right) = \operatorname{tr}\begin{pmatrix} \frac{1}{9} & 0 \\ 0 & 0 \end{pmatrix} = \frac{1}{9}.$$

In the same way, one can calculate to probabilities to get outcomes 2 and 3:

$$p^A_\rho(\{2\}) = \operatorname{tr}\begin{pmatrix} 0 & 0 \\ 0 & \frac{2}{9} \end{pmatrix} = \frac{2}{9}, \quad p^A_\rho(\{3\}) = \operatorname{tr}\begin{pmatrix} \frac{2}{9} & 0 \\ 0 & \frac{4}{9} \end{pmatrix} = \frac{6}{9}. \qquad \triangleleft$$

Spectral measures form a special class of observables, called *sharp* observables.

Definition 3.11 (Sharp observable)**.** Observables that are given by normalized PVM's, i.e. spectral measures, are called *sharp* observables.

Remark. In the von-Neumann formalism of quantum mechanics, we identify observables with self-adjoint operators. But, these observables correspond only with the sharp observables in the new formalism. To make this concrete, consider an observable $E \in \mathbf{O}(\mathbb{R}, \mathcal{F}(\mathbb{R}), \mathcal{H})$. Here $\mathcal{F}(\mathbb{R})$ denotes the Borel σ-algebra. For this observable, one can now define the first moment operator[2], $E[1]$, by

$$E[1] = \int_{\mathbb{R}} x \,\mathrm{d}E(x).$$

With the spectral theorem for (unbounded) self-adjoint operators, one can prove that if E is sharp, then the first momentum operator $E[1]$ is self-adjoint and for any self-adjoint operator A, there is a unique sharp observable E, such that $E[1] = A$. We thus have a one-to-one correspondence between sharp observables and self-adjoint operators. This correspondence is relevant to distinguish this new formalism from the standard one. Indeed, by only regarding self-adjoint operators (and thus sharp observables) as observables, one rejects the whole class of observables represented by semispectral measures that are not sharp. We refer to the Remark after Theorem 3.14 for some examples of these non-sharp observables.

We will now give the corresponding probability interpretation for (sharp) observables in the von Neumann formalism, which, as can be shown, coincides with Definition 3.9. For this example, let $\mathcal{H}$ be finite-dimensional, and let $E \in \mathbf{O}(\mathbb{R}, \mathcal{F}(\mathbb{R}), \mathcal{H})$ be sharp. We now only consider the first momentum operator $E[1] = A \in \mathcal{B}(\mathcal{H})$, which is self-adjoint. Let $\sigma(A)$ be its spectrum. For $\lambda \in \sigma(A)$, e_λ will denote the projection onto the eigenspace $H_\lambda = \{\psi_\lambda \in \mathcal{H} \mid A\psi_\lambda = \lambda\psi_\lambda\}$. We can now define the probability distribution p_A^ρ on $\sigma(A)$ for the state ρ by the *Born rule*

$$p_\rho^A(\lambda) = \mathrm{tr}(\rho e_\lambda),$$

and we interpret this as the probability that, when measuring the observable A, we get λ as a result. If in fact $\rho = |\phi\rangle\langle\phi|$ for some $\phi \in \mathcal{H}$ with $\|\phi\| = 1$ (i.e. ρ is a pure state) and we write p_ϕ^A for $p_{|\phi\rangle\langle\phi|}^A$, this simplifies to

$$p_\phi^A = \langle \phi \mid e_\lambda \phi \rangle \,.$$

If in addition H_λ is one dimensional, such that $e_\lambda = |\psi_\lambda\rangle\langle\psi_\lambda|$, the Born rule takes its standard form, in which it can be found in all standard introductions to quantum mechanics,

$$p_\phi^A(\lambda) = |\langle \phi \mid \psi_\lambda \rangle|^2.$$

For an extensive treatment of this formalism, the proofs of the statements above and the correspondings statements in an arbitrary Hilbert space, we refer to [14, Ch. 2, Ch. 4].

Remark. The term *sharp* suggest that a sharp operator is precisely defined or noiseless in a certain way. This is indeed the case. To make this concrete, consider again an observable $E \in \mathbf{O}(\mathbb{R}, \mathcal{F}(\mathbb{R}), \mathcal{H})$ and its first moment operators, as in the previous remark. We also define its second moment operator $E[2]$ as

$$E[2] = \int_{\mathbb{R}} x^2 \,\mathrm{d}E(x).$$

[2] This operator is unbounded and its domain consist of the $\phi \in \mathcal{H}$ such that x is integrable with respect to the complex measure $E_{\phi,\psi}$ for all $\psi \in \mathcal{H}$. But since unbounded operators are outside the scope of this thesis, we will not go into further detail here and do not pay attention to the relevant domains in the remainder of this section.

With these moment operators, for a given state ρ one can define the expectation and the variance of the probability measure E_ρ, by the integrals

$$\begin{aligned} \mathrm{Exp}(E_\rho) &= \int_{\mathbb{R}} x \,\mathrm{d}E_\rho(x) = \mathrm{tr}(\rho E[1]), \\ \mathrm{Var}(E_\rho) &= \int_{\mathbb{R}} x^2 \,\mathrm{d}E_\rho(x) - \left(\int_{\mathbb{R}} x \,\mathrm{d}E_\rho(x) \right)^2 = \mathrm{tr}(\rho E[2]) - (\mathrm{tr}(\rho E[1]))^2 . \end{aligned}$$

If we now define the *noise operator* $N(E) = E[2] - (E[1])^2$, we can rewrite the second equation as the sum of two non-negative terms:

$$\mathrm{Var}(E_\rho) = \big(\mathrm{tr}(\rho (E[1])^2) - \mathrm{tr}(\rho E[1])^2\big) + \mathrm{tr}(\rho N(E)).$$

This last *noise-term* is in general not zero; but it can be shown that for a sharp observable it is. In fact, an observable $E \in \mathbf{O}(\mathbb{R}, \mathcal{F}(\mathbb{R}), \mathcal{H})$ with self-adjoint first momentum operator is sharp if and only if its noise operator is zero. See for example [3, Thm. 8.5, Cor. 9.1].

3.1 Classes of observables

Now that we have laid down the basis of this formalism, we can introduce some special observables.

If $E_1, E_2 \in \mathbf{O}(\Omega, \mathcal{A}, \mathcal{H})$ and $t \in (0,1)$, then we define the observable $E = tE_1 + (1-t)E_2$ by

$$\mathcal{A} \ni X \mapsto E(X) = tE_1(X) + (1-t)E_2(X) \in \mathbf{E}(\mathcal{H}).$$

And thus we see that the set of observables is convex, induced by the convexity of $\mathbf{E}(\mathcal{H})$.

Definition 3.12 (Extreme Observable)**.** We say that $E \in \mathbf{O}(\Omega, \mathcal{A}, \mathcal{H})$ is an *extreme observable* if E is an element of the extreme boundary of $\mathbf{O}(\Omega, \mathcal{A}, \mathcal{H})$, i.e. if

$$E \in \mathrm{ex}(\mathbf{O}(\Omega, \mathcal{A}, \mathcal{H})).$$

Proposition 3.13. *Sharp observables are extreme.*

Proof. Let $P : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ be a sharp observable, that is, $P(X)$ is a projection for all $X \in \mathcal{A}$. Now assume $P = tA + (1-t)B$, for observables A and B, and $t \in (0,1)$. For every $X \in \mathcal{A}$ we have

$$P(X) = tA(X) + (1-t)B(X),$$

and since from Theorem 3.6 we know that $P(X)$ is extreme, we conclude

$$P(X) = A(X) = B(X).$$

Because this holds for every X, the claim follows. □

There is a convenient characterisation of extreme observables, given in the theorem below. Because the proof is slightly involved we will not give the complete proof here, but we refer to the original article.

Theorem 3.14. *Let $A \in \mathbf{O}(\Omega, \mathcal{A}, \mathcal{H})$, and let $(\mathcal{K}, P, J)$ be a minimal diagonal Naimark dilation of A. Then A is extreme if and only if for any bounded operator $D \in B(\mathcal{K})$ such that $[D, P(X)] = 0 \ \forall X \in \mathcal{A}$, the condition $J^* D J = 0$ implies $D = 0$.*

Sketch of proof. Theorem 2 of [15] gives the result for a bounded decomposable operator D. Then we use Proposition 8.11 in [3], stating that D is decomposable if and only if it commutes with every M_f, $f \in L^\infty(\mathcal{A}, \mu)$, where M_f is the multiplication operator with f. And since every M_f is a norm limit of a sequence of linear combination of M_{χ_X}, any D commuting with every M_{χ_X} commutes with every M_f. Now Theorem 2.8 shows that P is of the form $P(X) = M_{\chi_X}$. □

Remark. For an infinite-dimensional Hilbert space $\mathcal{H}$, there are observables that are extreme, but not sharp. These observables are not artificial mathematical constructions, but are of real physical relevance. All these examples are defined on infinite-dimensional Hilbert spaces, and the proofs are beyond the scope of this article. But the existence of these observables is the reason the questions in this article are relevant, since these non-sharp extreme observables are precisely the observables which could not be studied in the old formalism. Therefore, we will try to give some examples, which can be found in more detail in [7, 9, 15, 17]. For these examples, consider the single-mode optical field with the Hilbert space $\mathcal{H} \cong \mathrm{L}^2(\mathbb{R})$ spanned by the photon number states $\{|0\rangle, |1\rangle, |2\rangle, \ldots\}$, associated with the number operator

$$N = a^* a = \sum_{n=0}^{\infty} n\, |n\rangle\langle n|\,,$$

where $a = \sum_{n=0}^{\infty} \sqrt{n+1}\, |n\rangle\, \langle n+1|$. Define the position and momentum operator Q and P, respectively, as follows:

$$Q = \tfrac{1}{\sqrt{2}}(a^* + a), \quad P = \tfrac{i}{\sqrt{2}}(a^* - a),$$

which in position space behave as expected, $(Q\psi)(x) = x\psi(x), (P\psi)(x) = -i\frac{d\psi}{dx}(x)$. For $q, p \in \mathbb{R}$ define the Weyl operators

$$W(q,p) = e^{\frac{iqp}{2}} e^{-iqP} e^{ipQ},$$

and thus for all $\psi \in \mathcal{H} = \mathrm{L}^2(\mathbb{R})$ we have $((D(q,p)\psi)(x) = e^{\frac{iqp}{2}} e^{ipz} \psi(x-q)$. Consider the following two observables.

- For $\rho \in \mathbf{S}(\mathcal{H})$, define the *covariant phase space observable* G_ρ for $Z \subseteq \mathbb{C}$

  $$G_\rho(Z) = \frac{1}{2\pi} \int_Z W(\mathrm{Re}(z), \mathrm{Im}(z)) \rho W(\mathrm{Re}(z), \mathrm{Im}(z))^* dz,$$

  This observable represents the measurement in a so-called eight-port homodyne detector, with reference state ρ. A homodyne detector is used to extract information encoded in the phase of signal by comparing it to the signal that would have been transmitted if no information was encoded. If ρ is a pure state, i.e. $\rho = |\psi\rangle\langle\psi|$ for a normalized $\psi \in \mathcal{H}$, and $\langle \psi \,|\, W(q,p)\psi \rangle \neq 0$ for all $(q,p) \in \mathbb{R}^2$, it can be shown that G_ρ is a extreme observable, but not sharp, see [10, Thm. 1].

- The *canonical phase observable* $\mathit{\Phi}$ for $X \subseteq [0, 2\pi)$

  $$\mathit{\Phi}(X) = \sum_{n,m=0}^{\infty} \frac{1}{2\pi} \int_X e^{i(n-m)\theta}\, \mathrm{d}\theta\, |n\rangle\, \langle m|\,.$$

  For a discussion how to physically measure this observable, we refer to [17]. In [9] it shown that $\mathit{\Phi}$ is an extreme, but not a sharp, observable.

An important class of observables are the so called *discrete observables*. They occur in nature, but are also very useful for giving simple examples of observables. In Example 2.2 (b) we already saw an example of a discrete observable. Using a discrete observable, one can think of a *complete discrete observable*, since its definition is simpler, and in most situations sufficient.

Definition 3.15 (Completely discrete observable)**.** An observable $E \in \mathbf{O}(\Omega, \mathcal{A}, \mathcal{H})$ is called *completely discrete* if there exists a countable set $\Omega_0 \subseteq \Omega$, such that $\{\omega\} \in \mathcal{A}\ \forall \omega \in \Omega_0$ and $E(\Omega_0) = \boldsymbol{I}$.

Then for an observable as in the definition above,

$$E(X) = E(\Omega_0 \cap X) = \sum_{\omega_i \in \Omega_0 \cap X} E(\{\omega_i\}).$$

Now we can denote a completely discrete observable by its generating effects $E_i = E(\{\omega_i\})$, that is,

$$E = (E_1, E_2, \dots).$$

If Ω is finite, one can always denote an observable by $E = (E_1, E_2, \dots, E_n)$. In the case that $\omega_i \notin \mathcal{A}$ for some ω_i we have the somewhat weaker notion of a *discrete observable.*

Definition 3.16 (Discrete observabele)**.** An observabele $E \in \mathbf{O}(\Omega, \mathcal{A}, \mathcal{H})$ is called *discrete* if there exists a countable set $\Omega_0 \subseteq \Omega$ such that E is absolutely continuous with respect to the measure $\sum_{\omega \in \Omega} \delta_\omega$, where δ_x is the point measure concentrated on the point x.

This implies that one can identify an discrete observable A with the sequence $(A_i)_{i=1}^{\infty}$ of effects where $A_i = A(X_i)$, and $\{X_i\}_{i=1}^{\infty}$ is a disjoint collections of sets. This sequence can contain empty sets and it may happen that $A_i = 0$ for some i's. Now we can restrict A to the sub-σ-algebra generated by $\{X_i\}_{i=1}^{\infty}$.

Any effect induces a simple *binary observable* on a binary outcome space.

Definition 3.17 (Binarization)**.** Let $E \in \mathbf{E}(\mathcal{H})$ an effect operator. We define the binary observable

$$O^E : \{\emptyset, \{+1\}, \{-1\}, \{+1, -1\}\} \to \mathcal{B}(\mathcal{H})$$

by $O^E(\{+1\}) = E$, and hence $O^E(\{-1\}) = \boldsymbol{I} - E$.
For $A \in \mathbf{O}(\Omega, \mathcal{A}, \mathcal{H})$, we define its *binarization* associated to X as the binary observable $O^{A(X)}$. A binarization is sometimes also called a *partitioning.*

In the next definition, the concept of a weak Markov Kernel is used. Its definition can be found as Definition B.7 in Appendix B.

Definition 3.18 (Smearing)**.** Let $A \in \mathbf{O}(\Omega, \mathcal{A}, \mathcal{H}),\ M \in \mathbf{O}(\overline{\Omega}, \overline{\mathcal{A}}, \mathcal{H})$. We call A a *smearing* or *post-processing* of M, if there exist:

(a) a σ-finite measure $\mu : \overline{\mathcal{A}} \to [0, \infty]$, such that M is absolutely continuous with respect to it, i.e. $\mu(X) = 0$ implies $M(X) = 0$,

(b) a weak Markov Kernel $\beta : \overline{\Omega} \times \mathcal{A} \to \mathbb{R}$ with respect to μ, such that

$$A(X) = \int_{\overline{\Omega}} \beta(\omega, X)\, \mathrm{d}M(\omega).$$

4 Compatibility properties

The problem of measuring simultaneous two or more observables lies at the hart of quantum mechanics. Unlike in classical physcis, in quantum mechanics this is usually not possible. But under certain circumstances it is. It is well known that if observables are represented by self-adjoint operators, two observables are jointly measurable if and only if the operators commute. But in the formalism of quantum measurement, this only holds for sharp observables, as we shall see. So we will define other compatibility properties to characterise whether observables can be measured together or not. The most basic property will be *joint measurability*, which is essentially the property whether one can measure two observables A and B by measuring a single observable M. For an observable $A : \mathcal{A} \to \mathcal{B}(\mathcal{H})$, we denote its range by $\mathrm{ran}(A) = \{A(X) \mid X \in \mathcal{A}\}$. For the rest of this section, let $A \in \mathbf{O}(\Omega_1, \mathcal{A}_1, \mathcal{H})$, $B \in \mathbf{O}(\Omega_2, \mathcal{A}_2, \mathcal{H})$ be two observables.

Definition 4.1 (Joint measurability)**.** Let $\mathcal{A}_1 \otimes \mathcal{A}_2$ denote the σ-algebra generated by $\mathcal{A}_1 \times \mathcal{A}_2$. The observables A and B are called *jointly measurable* if there exists a POVM $M : \mathcal{A}_1 \otimes \mathcal{A}_2 \to \mathcal{B}(\mathcal{H})$ such that for all $X \in \mathcal{A}_1$ and all $Y \in \mathcal{A}_2$,

$$A(X) = M(X \times \Omega_2), \qquad B(Y) = M(\Omega_1 \times Y).$$

To verify that two observables are jointly measurable on this definition, one has to explicitly construct the joint observable, and to prove that two observables are not jointly measurable one has to show that such an observable does not exist. In many cases this is not an easy task. Therefore, other definitions of compatibility, equivalent to or weaker than joint measurability, have been formulated, hoping these are easier to verify. We will first give three definitions, which, as we will show, are equivalent to joint measurability.

Definition 4.2 (Common bi-observable)**.** The observables A and B are said to have a *common bi-observable* if there is a bi-observable $M : \mathcal{A}_1 \times \mathcal{A}_2 \to \mathcal{B}(\mathcal{H})$ such that for all $X \in \mathcal{A}_1, Y \in \mathcal{A}_2$,

$$A(X) = M(X, \Omega_2), \quad B(Y) = M(\Omega_1, Y).$$

Definition 4.3 (Functions of an observable)**.** The observables A and B are called *functions of an observable* if there is an observable $M \in \mathbf{O}(\overline{\Omega}, \overline{\mathcal{A}}, \mathcal{H})$, and two measurable functions $f_1 : \overline{\Omega} \to \Omega_1$, $f_2 : \overline{\Omega} \to \Omega_2$, such that for all $X \in \mathcal{A}_1, Y \in \mathcal{A}_2$

$$A(X) = M(f_1^{-1}(X)), \quad B(Y) = M(f_2^{-1}(Y)).$$

Definition 4.4 (Smearing of an observable)**.** The observables A and B are said to be *smearings* of an observable $M \in \mathbf{O}(\overline{\Omega}, \overline{\mathcal{A}}, \mathcal{H})$, if there are weak Markov kernels $\beta_1 : \overline{\Omega} \times \mathcal{A}_1 \to \mathbb{R}$ and $\beta_2 : \overline{\Omega} \times \mathcal{A}_2 \to \mathbb{R}$, such that A is a smearing of M with respect to β_1 and B is a smearing of M with respect to β_2.

The next theorem states that under the right topological assumptions on the outcome spaces the previous four compatibility properties are equivalent. These topological requirements are reasonable and for example hold if Ω is a subset of $(\mathbb{R}^n, \mathcal{F}(\mathbb{R}^n))$ (see [3, prop. 4.9]). Here $\mathcal{F}(\mathbb{R}^n)$ indicates the Borel σ-algebra. Because these topological details are beyond the scope of this article, we refer to [3, Section 4.7] for the proofs concerning these properties. In the remainder of this theorem, we assume these conditions when necessary.

Theorem 4.5. *Assume the outcome spaces* $(\Omega_1, \mathcal{A}_1)$ *and* $(\Omega_2, \mathcal{A}_2)$ *of* A *and* B*, respectively, have the right topological properties.*[3] *Then the following conditions are equivalent:*

(a) A *and* B *have a common bi-observable;*

(b) A *and* B *are jointly measurable;*

(c) A *and* B *are functions of a third observable;*

(d) A *and* B *are smearings of a third observable.*

Proof. For the proof of $(a) \implies (b)$ we refer to [3, Thm. 4.2]. The implication $(b) \implies (c)$ is quite trivial. Indeed, for a joint observable $M : \mathcal{A}_1 \otimes \mathcal{A}_2 \to \mathcal{B}(\mathcal{H})$ of A and B, functions $f_1 : \Omega_1 \times \Omega_2 \to \Omega_1, f_2 : \Omega_1 \times \Omega_2 \to \Omega_2$ given by $f_1(x, y) = x, f_2(x, y) = y$ such that $f_1^{-1}(X) = X \times \Omega_2,\ f_2^{-1}(Y) = \Omega_1 \times Y$ make A and B functions of the observable M.
For the implications $(c) \implies (d)$, assume that A and B are functions of a third observables M via functions $f_i : \overline{\Omega} \to \Omega_i, i = 1, 2$. Then the Markov kernels β_1, β_2, defined by

$$\beta_1 : \overline{\Omega} \times \mathcal{A}_1 \to \mathbb{R} \qquad\qquad \beta_2 : \overline{\Omega} \times \mathcal{A}_2 \to \mathbb{R}$$
$$\beta_1(\omega, X) = \chi_X(f_1(\omega)) \qquad\qquad \beta_2(\omega, Y) = \chi_Y(f_2(\omega)),$$

make A and B smearings of M.
Now we consider the last implication: $(d) \implies (a)$. Let $M \in \mathbf{O}(\overline{\Omega}, \overline{\mathcal{A}}, \mathcal{H})$ be an observable such that A and B are smearings of M by means of Markov kernels β_1, β_2. Consider the positive operator bimeasure $O : \mathcal{A}_1 \times \mathcal{A}_2 \to \mathcal{B}(\mathcal{H})$ defined by

$$O(X, Y) = \int_{\overline{\Omega}} \beta_1(\omega, X)\beta_2(\omega, Y)\, \mathrm{d}M(\omega).$$

Since $A(X) = O(X, \Omega_2)\ \forall X \in \mathcal{A}_1$ and $B(Y) = O(\Omega_1, Y)\ \forall X \in \mathcal{A}_2$, we see that O is a bi-observable for A and B. □

But also these compatibility properties are usually not easy to verify, since they essentially require one to construct some kind of joint observable. This is the reason why people have come up with weaker notions of compatibility, namely *commutativity*, *coexistence*, and *joint measurability of binarizations* which are easier to verify. Commutativity is the easiest one, since it doesn't require the construction of another observable, but only imposes requirements on the given observables themselves.

Definition 4.6 (Commutativity)**.** Two POVMs A and B are said to *commute* if

$$[A(X), B(Y)] = A(X)B(Y) - B(Y)A(X) = 0 \quad \forall X \in \mathcal{A}_1,\ \forall Y \in \mathcal{A}_2.$$

Notice that this is a somewhat non-standard definition of commutativity, since with this definition, a POVM does not need to commute with itself.

We will now give the equivalent of the well-known property that, when identifying observables with self-adjoint observables, observables are jointly measurable if and only if they commute.

[3] We assume Ω is a Hausdorff space and there is a ring $\mathcal{R}$ generating the σ-algebra $\mathcal{A}$, such that Ω is the union of a countable collection of members of $\mathcal{R}$. Moreover, we assume that every σ-additive set function $\mu : \mathcal{R} \to [0, \infty)$ is such that for each $L \in \mathcal{R}$ and $\epsilon > 0$ there is a set $O \in \mathcal{R}$ with a compact set K such that $O \subseteq K \subseteq L$ and $\mu(L) - \mu(O) < \epsilon$.

Proposition 4.7. *Let $A \in \mathbf{O}(\Omega_1, \mathcal{A}_1, \mathcal{H}), B \in \mathbf{O}(\Omega_2, \mathcal{A}_2, \mathcal{H})$ be two observables and assume A is sharp. Then A and B commute if and only if they are jointly measurable. In that case they have a unique joint observable M, determined by*

$$M(X \times Y) = A(X)B(Y) \quad \forall X \in \mathcal{A}_1, Y \in \mathcal{A}_2.$$

Proof. Let $A \in \mathbf{O}(\Omega_1, \mathcal{A}_1, \mathcal{H})$ be sharp and $B \in \mathbf{O}(\Omega_2, \mathcal{A}_2, \mathcal{H})$. First assume A and B commute. We claim that $M : \mathcal{A}_1 \times \mathcal{A}_2 \to \mathcal{B}(\mathcal{H})$, defined by

$$M(X, Y) = A(X)B(Y),$$

is a bi-observable for A and B. For M to be an bi-observable it is sufficient to check that it is positive, since the other properties follow immediately from the fact that A and B are observables. To see that M is positive, take $X \in \mathcal{A}_1, Y \in \mathcal{A}_2$ arbitrarily and consider

$$\begin{aligned} \langle \phi \,|\, M(X,Y)\phi \rangle &= \langle \phi \,|\, A(X)B(Y)\phi \rangle = \left\langle \phi \,\middle|\, A(X)^2 B(Y)\phi \right\rangle \text{ (since A(X) is a projection)} \\ &= \langle A(X)\phi \,|\, A(X)B(Y)\phi \rangle \\ &= \langle A(X)\phi \,|\, B(Y)A(X)\phi \rangle \geq 0. \end{aligned}$$

Now it follows immediately that M is a bi-observable for A and B, and with Theorem 4.5 we conclude that A and B are jointly measurable.

To prove the other implication, let $M \in \mathbf{O}(\overline{\Omega}, \overline{\mathcal{A}}, \mathcal{H})$ be the joint observable of A and B, and let $X \in \mathcal{A}_1, Y \in \mathcal{A}_2$ be arbitrary. Since

$$M(X \times Y) \leq M(X \times \Omega_2) = A(X),$$

it follows that $\operatorname{ran} M \subseteq \operatorname{ran} A$. Because $A(X)$ is a projection, from Proposition A.18 one has

$$A(X)M(X \times Y) = M(X \times Y)A(X) = M(X \times Y).$$

Applying this to X^c, we obtain

$$A(X)M(X^c \times Y) = (\boldsymbol{I} - A(X^c))M(X^c \times Y) = 0 \quad \text{and} \quad M(X^c \times Y)A(X) = 0.$$

It now follows that

$$A(X)B(Y) = A(X)M(\Omega_1 \times Y) = A(X)((M(X \times Y) + M(X^c \times Y)) = M(X \times Y),$$

and similarly

$$B(Y)A(X) = M(X \times Y).$$

We conclude that A and B commute, and since M is determined by its values on product sets, it is unique. ☐

Coexistence is another a weaker property than joint measurability, since it only requires the existence of a encompassing observable, the so called *mother observable* of the two observables, but not an explicit way to obtain the measurement outcome of the original observables, when we have measured the mother observable.

Definition 4.8 (Coexistence)**.** Two POVMs A and B are called *coexistent* if there exists a POVM M such that

$$\operatorname{ran}(A) \cup \operatorname{ran}(B) \subseteq \operatorname{ran}(M).$$

Such a POVM M is called a *mother observable* to A and B.

We will extensively study the relationship between coexistence and joint measurability in the next section. But first, we introduce the last compatibility property which we will study. It has been introduced by Heinosaari et al. in [11], and concerns the binarizations of observables, instead of the observables themselves. Since binarizations are simple observables, with only two non-trivial outcomes, it is often easier to study them instead of the original observables.

Definition 4.9 (Joint Measurability of Binarizations)**.** The observables A and B are said to have *jointly measurable binarizations* if for all $X \in \mathcal{A}_1$ and $Y \in \mathcal{A}_2$ the binarizations $O^{A(X)}$ and $O^{B(Y)}$ are jointly measurable.

As we shall see in the next few pages, for sharp observables all compatibility properties are equivalent, including coexistence and joint measurability of binarizations. In the next part of this paper, we will be looking for weaker notions than sharpness for the these last two properties to be equivalent with joint measurability. That would be really helpful, because then for these classes of observables it is sufficient to verify one of these simpler notions to prove joint measurability. In section 4.1 we study the link between coexistence and joint measurability, and in section 4.2 we study the relationship between joint measurability of binarizations and joint measurabiliy.

4.1 Coexistence and Joint Measurability

It is obvious that joint measurability of two observables implies their coexistence, since the POVM M in Definition 4.1 is clearly a mother observable of A and B. But the converse does not hold in general, as proved by Reeb et al. [18]. We will now give their counterexample, which is a simple construction concerning two observables.

Proposition 4.10. *Coexistence does not imply joint measurability.*

Proof. We will construct observables A and B that are coexistent but not jointly measurable. Let $\Omega_A = \{1,2,3\}$ and $\Omega_B = \{1,2\}$. Let $\{|1\rangle, |2\rangle, |3\rangle\}$ be an orthonormal basis of $\mathcal{H} = \mathbb{C}^3$ and let $|\phi\rangle = \frac{1}{\sqrt{3}}(|1\rangle, + |2\rangle + |3\rangle)$. Define the following effects:

$$A_i = \tfrac{1}{2}(\boldsymbol{I} - |i\rangle\langle i|), \quad i \in \Omega_A,$$

$$B_1 = \tfrac{1}{2}|\phi\rangle\langle\phi|, \quad B_2 = \boldsymbol{I} - B_1.$$

To show that A and B are coexistent we have to construct an observable whose range contains the ranges of A and B. Now consider the observable M on the outcome space $\Omega_M = \{1,2,3,4,5\}$ given by

$$M = \left(\tfrac{1}{2}|1\rangle\langle 1|, \tfrac{1}{2}|2\rangle\langle 2|, \tfrac{1}{2}|3\rangle\langle 3|, B_1, \tfrac{1}{2}\boldsymbol{I} - B_1\right).$$

We can now see that M is a mother observable for A and B by noting

$$A_i = M(\{2,3\}), A_2 = M(\{1,3\}), A_3 = M(\{1,2\}), B_1 = M(\{4\}) \text{ and } B_2 = M(\{1,2,3,5\}),$$

and conclude that A and B are coexistent.
To prove that A and B aren't jointly measurable, we argue by contradiction. If they would

be jointly measurable, there should exist an observable $J = (J_{11}, J_{12}, J_{21}, J_{22}, J_{31}, J_{32})$ such that

$$A_i = \sum_{j=1}^{2} J_{ij} \ \forall i \in \Omega_A, \quad \text{and } B_j = \sum_{j=1}^{3} J_{ij} \ \forall j \in \Omega_B.$$

Because the range of the projection operator B_1 is one-dimensional, $J_{i1} = \lambda_i B_1$ must hold for all $i \in \Omega_A$ for some $\lambda_i \geq 0$. And hence, by the equations above $A_i = \lambda_i B_1 + J_{i2}$. But then

$$0 = \langle i|A_i|i\rangle = \frac{\lambda_i}{2} |\langle i \mid \phi\rangle|^2 + \langle i|J_{i2}|i\rangle \quad \forall i \in \Omega_A,$$

which implies $\lambda_i = 0 \ \forall i$, since $|\langle i \mid \phi\rangle|^2 = \frac{1}{3}$. Then however $J_{i1} = 0$ for all i and hence $B_1 = 0$, which is the desired contradiction. □

Recently Haapasalo et al. proved that if one of the two observables A, B is extreme and discrete, then coexistence and joint measurability are equivalent [8]. We will now give a slightly revised and more detailed version of their proof. We have also corrected a minor mistake in the proof of the following lemma.

Lemma 4.11. *Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces and let $A \in B(\mathcal{H}, \mathcal{K})$ and $B \in \mathcal{B}(\mathcal{H})$. Then $0 \leq B \leq A^*A$ if and only if there exists a $C \in B(\mathcal{K})$, $0 \leq C \leq \boldsymbol{I}_\mathcal{K}$, such that $B = A^*CA$. Furthermore, C is unique if and only if* $\operatorname{ran}(A)$ *is dense in $\mathcal{K}$.*

Proof. First assume that $0 \leq B \leq A^*A$. For an arbitrary $\phi \in \mathcal{H}$, we define $C_0(A\phi) = \sqrt{B}\phi$. First we show this is well defined. If $A\phi = A\psi$, i.e. $A\phi_- = 0$ for $\phi_- = \phi - \psi$, then

$$0 \leq \|C_0(A\phi_-)\|^2 = \left\|\sqrt{B}\phi_-\right\|^2 = \langle \phi_- \mid B\phi_-\rangle \leq \langle \phi_- \mid A^*A\phi_-\rangle = \|A\phi_-\| = 0,$$

and with linearity we can conclude $C_0(A\phi) = C_0(A\psi)$. If $\eta \in ran(A)^\perp$ then define $C_0(\eta) = 0$. Now we define $C = C_0^* C_0$, and we see that

$$\langle \phi \mid A^*CA\phi\rangle = \langle \phi \mid A^*C_0^*C_0A\phi\rangle = \langle C_0A\phi \mid C_0A\phi\rangle = \left\langle \sqrt{B}\phi \,\middle|\, \sqrt{B}\phi \right\rangle = \langle \phi \mid B\phi\rangle,$$

and so $B = A^*CA$. Furthermore, from $\langle A\phi \mid (\boldsymbol{I}_\mathcal{K} - C)A\phi\rangle = \langle \phi \mid A^*A\phi\rangle - \langle \phi \mid B\phi\rangle$ it directly follows that $0 \leq C \leq \boldsymbol{I}_\mathcal{K}$. Now assume that there is a $0 \leq C \leq \boldsymbol{I}_\mathcal{K}$, such that $B = A^*CA$. For any $\phi \in \mathcal{H}$ we have

$$\langle \phi \mid B\phi\rangle = \langle \phi \mid A^*CA\phi\rangle = \langle A\phi \mid CA\phi\rangle \geq 0,$$

since C is positive. But also because $C \leq \boldsymbol{I}_\mathcal{K}$, one has

$$0 \leq \langle A\phi \mid (\boldsymbol{I}_\mathcal{K} - C)A\phi\rangle = \langle A\phi \mid A\phi\rangle - \langle A\phi \mid CA\phi\rangle = \langle \phi \mid A^*A\phi\rangle - \langle \phi \mid B\phi\rangle,$$

we have verified that $0 \leq B \leq A^*A$.

Concerning uniqueness of C, we notice that C is determined by A and B only on the closure of the range of A. So if and only if $\operatorname{ran}(A)$ is dense in $\mathcal{K}$, i.e. $\operatorname{ran}(A)^\perp = \{0\}$, and hence C is unique. □

Theorem 4.12. *Let $A \in \mathbf{O}(\Omega, \mathcal{A}, \mathcal{H})$ and $M \in \mathbf{O}(\overline{\Omega}, \overline{\mathcal{A}}, \mathcal{H})$ If A is discrete and extreme and such that* $\operatorname{ran}(A) \subseteq \operatorname{ran}(M)$*, then A and M are jointly measurable.*

Proof. Let $A \in \mathbf{O}(\Omega, \mathcal{A}, \mathcal{H})$ and $M \in \mathbf{O}(\overline{\Omega}, \overline{\mathcal{A}}, \mathcal{H})$ and assume that there are $X \in \Sigma$ and $Z_X \in \overline{\Sigma}$ such that $M(Z_X) = A(X)$. Pick a minimal diagonal Naimark dilation $(\mathcal{K}, P, J)$ for A. For all $Z \in \overline{\Sigma}$,

$$M(Z \cap Z_X) \leq M(Z_X) = A(X) = (P(X)J)^* P(X)J,$$

and Lemma 4.11 implies that there exists a $C_X(Z) \in B(\mathcal{K})$ such that $0 \leq C_X(Z) \leq \boldsymbol{I}_{\mathcal{K}}$, and

$$M(Z \cap Z_X) = (P(X)J)^* C_X(Z) P(X)J = J^* P(X) C_X(Z) P(X)J.$$

From this we see that we can choose $C_X(Z)$ such that $C_X(Z)\mathcal{K} \subseteq P(X)\mathcal{K}$, and with Proposition A.18 we conclude $0 \leq C_X(Z) \leq P(X)$ and

$$M(Z \cap Z_X) = J^* C_X(Z) P(X) J. \tag{1}$$

Assume now that A is discrete, that is, there exists a disjoint sequence $\{X_i\}_{i=1}^{\infty}$ such that one can identify A with the sequence $(A_i)_{i=1}^{\infty}$ where $A_i = A(X_i)$. Similarly, we write $P_i = P(X_i)$. Moreover, we assume that A is extreme and that $\operatorname{ran} A \subseteq \operatorname{ran} M$. Now, for all $X \in \mathcal{A}$ there is a $Z_X \in \overline{\mathcal{A}}$ such that $M(Z_X) = A(X)$ and, as above, we have the positive operators $C_X(Z) \leq P(X)$ such that (1) holds for all $X \in \mathcal{A}$ and $Z \in \overline{\mathcal{A}}$. From now on, denote $C_i(Z) = C_{X_i}(Z)$ and $Z_i = Z_{X_i}$ for any i and $Z \in \overline{\mathcal{A}}$. Note that since $\boldsymbol{I} \in \operatorname{ran}(A)$ and $M(\overline{\Omega}) = \boldsymbol{I}$ we must have $\cup_{i=1}^{\infty} Z_i = \overline{\Omega}$. It follows that $0 \leq C_i(Z) \leq P_i$ for any i and $Z \in \overline{\mathcal{A}}$, and with Proposition A.18 we conclude $C_i(Z) = P_i C_i(Z) = C_i(Z) P_i$. For any $X_i \neq X_j$ such that $A_i \neq 0 \neq A_j$ one has from (1),

$$M(Z_i \cap Z_j) = J^* C_i(Z_j) J = J^* C_j(Z_i) J.$$

By defining $D = C_i(Z_j) - C_j(Z_i)$ we get

$$J^* D J = 0, \quad [D, P(X)] = 0 \ \forall X \in \mathcal{A},$$

implying $D = 0$ by Theorem 3.14, so that, since the operator $C_i(Z_j)$ and $C_j(Z_i)$ are supported on the orthogonal subspaces $P_i\mathcal{H}$ and $P_j\mathcal{H}$, we have $C_i(Z_j) = 0 = C_j(Z_i)$ and $M(Z_i \cap Z_j) = 0$. From Equation (1) we get

$$M(Z \cap Z_i) = J^* C_i(Z) J.$$

Claim: For all i, $C_i : \overline{\mathcal{A}} \to B(P_i\mathcal{K})$ is a unique normalized POVM.
Assume that $(W_n)_{n=1}^{\infty}$ is a disjoint sequence in $\mathcal{A}$. For any $N \in \mathbb{N}$ the following equality holds for any i:

$$J^* C_i\Big(\bigcup_{n=1}^{N} W_n\Big) J = M\Big(\bigcup_{n=1}^{N} W_n \cap Z_i\Big) = \sum_{n=1}^{N} M(W_n \cap Z_i) = \sum_{n=1}^{N} J^* C_i(W_n) J.$$

Since the operator $C_i\big(\bigcup_{n=1}^{N} W_n\big) - \sum_{n=1}^{N} C_i(W_n)$ commutes with the spectral measure P, by using the previous equality together with the extremality of A we can conclude, with Theorem 3.14, that

$$\sum_{n=1}^{N} C_i(W_n) = C_i\Big(\bigcup_{n=1}^{N} W_n\Big) \leq P_i.$$

Since all the summands are positive, the sequence $\left(\sum_{n=1}^{N} C_i(W_n)\right)_{N=1}^{\infty} = \left(C_i\left(\bigcup_{n=1}^{N} W_n\right)\right)_{N=1}^{\infty}$ is increasing, as well as bounded from above by P_i, implying that

$$\sum_{n=1}^{\infty} C_i(W_n) = \text{w-lim}_{N\to\infty} \sum_{n=1}^{N} C_i(W_n) = \sup_{N\in\mathbb{N}} C_i\Big(\bigcup_{n=1}^{N} W_n\Big)$$

is well-defined, and since the limit is weak, the operator $\sum_{n=1}^{\infty} C_i(W_n)$ commutes with $P(X)\ \forall X \in \mathcal{A}$. Again by using Theorem 3.14, one now has

$$C_i\big(\bigcup_{n=1}^{\infty} W_n\big) = \sum_{n=1}^{\infty} C_i(W_n),$$

since the operator $C_i\left(\bigcup_{n=1}^{\infty} W_n\right) - \sum_{n=1}^{\infty} C_i(W_n)$ commutes with the P_j's. Hence for all i, the map $C_i : \overline{\mathcal{A}} \to B(P_i\mathcal{K})$ is weakly σ-additive. Similarly, $C_i(\overline{\Omega}) = P_i = \boldsymbol{I}_{P_i\mathcal{K}}$, proving the claim. Note that $\mathcal{K} = \oplus_{i=1}^{\infty} P_i\mathcal{K}$.
Now one can define a joint observable $N : \mathcal{A} \otimes \overline{\mathcal{A}} \to \mathcal{B}(\mathcal{H})$ via

$$N(X_i \times Z) = J^* C_i(Z) J,$$

satisfying

$$\begin{aligned} A(X_i) &= J^* P_i J = N(X_i \times \overline{\Omega}), \\ M(Z) &= \sum_i M(Z \cap Z_i) = \sum_i J^* C_i(Z) J = N(\Omega \times Z). \end{aligned} \tag{2}$$

This proves that A and M are jointly measurable. □

Corollary 4.12.1. *Any discrete and extreme $A \in \mathbf{O}(\Omega_1, \mathcal{A}_1, \mathcal{H})$ and any $B \in \mathbf{O}(\Omega_2, \mathcal{A}_2, \mathcal{H})$ are jointly measurable if and only if they are coexistent.*

Proof. We use the assumptions and notation of Theorem 4.12. Assume also that $\operatorname{ran} B \subseteq \operatorname{ran} M$, i.e. for each $Y \in \mathcal{A}_2$ there exists a $W_Y \in \overline{\mathcal{A}}$ such that $M(W_Y) = B(Y)$. Then, by Equation (2) we have

$$B(Y) = M(W_Y) = \sum_{i=1}^{\infty} J^* C_i(W_Y) J.$$

We now claim that

$$C_i' : \mathcal{A}_2 \to \mathcal{B}(P_i\mathcal{K}), \quad C_i'(Y) = C_i(W_Y)$$

is a semispectral measure for all $i \in \mathbb{N}$. Namely, let $(Y_n)_{n=1}^{\infty} \subseteq \mathcal{A}_2$ be a disjoint sequence, and write $Y^N = \bigcup_{n=1}^{N} Y_n$ for any $N \in \mathbb{N}$. Now one has

$$\sum_i J^* C_i(W_{Y^N}) J = B(Y^N) = \sum_{n=1}^{N} B(Y_n) = \sum_i \sum_{n=1}^{N} J^* C_i(W_{Y_n}) J,$$

and because A is extreme, from Theorem 3.14 we can conclude

$$\sum_i C_i(W_{Y^N}) = \sum_i \sum_{n=1}^{N} C_i(W_{Y_n}).$$

Since the images of C_i are supported on mutually orthogonal subspaces, it follows that

$$C'_i\left(\bigcup_{n=1}^{N} Y_n\right) = C_i(W_{Y^N}) = \sum_{n=1}^{N} C_i(W_{Y_n}) = \sum_{n=1}^{N} C'_i(Y_n),$$

for all i and N, and, as in the previous proof, the sequence of finite partial sums in increasing and bounded from above, and one can show that C'_i is weakly σ-additive. Positivity of C'_i follows immediately from the positivity of C_i. And since $B(\Omega_2) = \boldsymbol{I} = M(W_{\Omega_2})$, we have $W_{\Omega_2} = \overline{\Omega}$, and we see that C'_i is normalised. This proves our claim that C'_i is a semispectral measure. And thus we see that

$$(X_i, Y) \mapsto J^* C_i(W_Y) J$$

is a common bi-observable of A and B, and by Theorem 4.5 we conclude that A and B are jointly measurable. □

4.2 Joint Measurability of Binarizations and Coexistence

We will now discuss the relationship between joint measurability of binarizations and coexistence of two quantum observables. We first note that for binary observables, coexistence and joint measurability are equivalent, as proved in [13].

Proposition 4.13. *Let $O^E, O^F \in \mathbf{O}(\Omega = \{+1,-1\}, \mathcal{P}(\Omega), \mathcal{H})$ be two binary observables for effects $E, F \in \mathbf{E}(\mathcal{H})$. O^E, O^F are jointly measurable if and only if they are coexistent.*

Proof. Let $O^E, O^F \in \mathbf{O}(\Omega = \{+1,-1\}, \mathcal{P}(\Omega), \mathcal{H})$ and let $M \in \mathbf{O}(\overline{\Omega}, \overline{\mathcal{A}}, \mathcal{H})$ a mother observable of O^E, O^F, such that $M(X) = E$, $M(Y) = F$. Now consider the following partition of $\overline{\Omega}$

$$\mathcal{R} = \{X \cap Y, X^c \cap Y, X \cap Y^c, X^c \cap Y^c\},$$

and define the observable $M^{\mathcal{R}} : \{1,2,3,4\} \to \mathcal{B}(\mathcal{H})$ by

$$(M_1^{\mathcal{R}}, M_2^{\mathcal{R}}, M_3^{\mathcal{R}}, M_4^{\mathcal{R}}) = \big(M(X \cap Y), M(X^c \cap Y), M(X \cap Y^c), M(X^c \cap Y^c)\big).$$

We now define the functions f_1, f_2, by

$$\begin{aligned} &f_1 : \{1,2,3,4\} \to \{+1,-1\} && f_2 : \{1,2,3,4\} \to \{+1,-1\}, \\ &f_1 : 1,3 \mapsto 1;\ 2,4 \mapsto -1 && f_2 : 1,2 \mapsto 1;\ 3,4 \mapsto -1, \end{aligned}$$

and hence we can write

$$E = M(X) = M((X \cap Y) \cup (X \cap Y^c)) = M(X \cap Y) + M(X \cap Y^c) = M^{\mathcal{R}}(f_1^{-1}(1)),$$

$$F = M(Y) = M((X \cap Y) \cup (X^c \cap Y)) = M(X \cap Y) + M(X^c \cap Y) = M^{\mathcal{R}}(f_2^{-1}(1)),$$

which shows that O^E, O^F are functions of the observable $M^{\mathcal{R}}$ and thus by Theorem 4.5 they are jointly measurable. □

We also see that coexistence implies joint measurability of binarizations.

Theorem 4.14. *If $A \in \mathbf{O}(\Omega_1, \mathcal{A}_1, \mathcal{H})$ and $B \in \mathbf{O}(\Omega_2, \mathcal{A}_2, \mathcal{H})$ are coexistent, they have jointly measurable binarizations.*

Proof. Let $A \in \mathbf{O}(\Omega_1, \mathcal{A}_1, \mathcal{H})$, and $B \in \mathbf{O}(\Omega_2, \mathcal{A}_2, \mathcal{H})$, and let $M \in \mathbf{O}(\overline{\Omega}, \overline{\mathcal{A}}, \mathcal{H})$ be their mother observable, and choose $X \in \mathcal{A}_1$ and $Y \in \mathcal{A}_2$ arbitrary. It is clear that M is also a mother observable for $O^{A(X)}$ and $O^{B(Y)}$. Thus there is a $Z \in \overline{\mathcal{A}}$ such that

$$O^{A(X)}(\{+1\}) = A(X) = M(Z) = \int_{\overline{\Omega}} \beta(\omega, \{+1\}) \, \mathrm{d}M(\omega),$$

where $\beta(\omega, \{+1\}) = \chi_Z(\omega)$. We can now extend β to a Markov kernel by defining for all $X \in \{\emptyset, \{+1\}, \{-1\}, \{+1, -1\}\}$,

$$\beta(\omega, X) = \chi_{f^{-1}(X)}(\omega),$$

where $f : \overline{\Omega} \to \{+1, -1\}$ is a measurable function such that

$$f^{-1}(\{+1\}) = Z,$$
$$f^{-1}(\{-1\}) = \overline{\Omega} \setminus Z,$$

and now we see that $O^{A(X)}$ is a smearing of M. In exactly the same way $O^{B(Y)}$ is a smearing of M as well. With Theorem 4.5 we conclude that $O^{A(X)}$ and $B^{B(Y)}$ are jointly measurable. □

As Haapsalo et al. has showed in [8], joint measurability of binarizations is a weaker property than coexistence.

Proposition 4.15. *Joint measurability of binarizations does not imply coexistence.*

Proof. We will show this by constructing two observables A and B that have jointly measurable binarizations but are not coexistent. Let $\Omega_A = \{1, 2, 3\}$ and $\Omega_B = \{1, 2\}$. Let $\{|1\rangle, |2\rangle\}$ be the standard orthonormal basis of $\mathcal{H} = \mathbb{C}^2$ and let $|\phi\rangle = \frac{1}{\sqrt{2}}(|1\rangle - |2\rangle)$. Define the following effects

$$A_1 = \tfrac{4}{7} |1\rangle\langle 1|, \quad A_2 = \tfrac{4}{7} |2\rangle\langle 2| \quad A_3 = \boldsymbol{I} - A_1 - A_2,$$

$$B_1 = \tfrac{4}{7} |\phi\rangle\langle\phi|, \quad B_2 = \boldsymbol{I} - B_1,$$

and, from these, define the observables $A = (A_1, A_2, A_3)$ and $B = (B_1, B_2)$.

If A and B are coexistent, there exists a mother observable $M : \overline{\mathcal{A}} \to B(\mathbb{C}^2)$, such that $\operatorname{ran} A \cup \operatorname{ran} B \subseteq \operatorname{ran} M$. This implies that there exist sets $X, Y, Z \in \overline{\mathcal{A}}$ such that $A_1 = M(X), A_2 = M(Y), B_1 = M(Z)$. Since these effects are rank-1 and we always have $M(X \cap Y) \leq M(X)$, it follows that $M(X \cap Y) = 0$. Because this has to hold for all pairwise intersections of X, Y, and Z, and $M(X \cap Y \cap Z) \leq M(X \cap Y) = 0$, we have

$$M(X \cup Y \cup Z) = A_1 + A_2 + B_1 \leq \boldsymbol{I},$$

where the inequality has to hold because the range of an observable only consists of effect operators. But this gives a contradiction with Proposition 3.7, since the greatest eigenvalue of $A_1 + A_2 + B_1 = \begin{pmatrix} 6/7 & -2/7 \\ -2/7 & 6/7 \end{pmatrix}$ is $\frac{8}{7}$ and thus the observables A and B are not coexistent.

To see that all binarizations of A and B are jointly measurable, because of Proposition 4.13 it is sufficient to show that all the binarizations are coexistent, which is the case if the following three conditions hold:

$$A_i + B_1 \leq \boldsymbol{I} \quad \forall i \in \Omega_A.$$

Indeed, consider the binarization O^{A_1} whose range is $\{A_1, \boldsymbol{I} - A_1\}$. If $A_1 + B_1 \leq \boldsymbol{I}$, we can define a mother observable $M : \overline{\mathcal{A}} \to B(\mathbb{C}^2)$ by $M(X) = A_1$, $M(Y) = B_1$ for some disjoint $X, Y \in \overline{\mathcal{A}}$.

By a straightforward calculation, we then find that the eigenvalues of the the operators $A_i + B_1$ are $\lambda_1 = \frac{-2\sqrt{2}+4}{7}, \lambda_2 = \frac{2\sqrt{2}+4}{7}$ for $i = 1, 2$ and $\lambda_1 = \frac{3}{7}, \lambda_2 = 1$ for $i = 3$. Now with Proposition 3.7 one can see that the conditions $A_i + B_1 \leq \boldsymbol{I}$ hold. □

But as Heinosaari et al. have showed in [11], if one of the observables A and B is sharp, joint measurability of binarizations even implies joint measurability of A and B.

Theorem 4.16. *Let $A \in \mathbf{O}(\Omega_1, \mathcal{A}_1, \mathcal{H})$ and $B \in \mathbf{O}(\Omega_2, \mathcal{A}_2, \mathcal{H})$ and let A be sharp. If A and B have jointly measurable binarizations, then A and B are jointly measurable.*

Proof. Let $A \in \mathbf{O}(\Omega_1, \mathcal{A}_1, \mathcal{H})$ be a sharp observable and $B \in \mathbf{O}(\Omega_2, \mathcal{A}_2, \mathcal{H})$. Assume that A and B have jointly measurable binarizations. Since A is sharp, every binarization of A is. But from Proposition 4.7 we now known that for all $X \in \mathcal{A}_1, Y \in \mathcal{A}_2$ $O^{A(X)}$ and $O^{B(Y)}$ commute. But this means that A and B commute, since for two observables to commute we have to have $[A(X), B(Y)] = [O^{A(X)}(+1), O^{B(Y)}(+1)] = 0 \; \forall X \in \mathcal{A}_1, Y \in \mathcal{A}_2$. And thus again by Proposition 4.7, we conclude that A and B are jointly measurable. □

5 Conclusion

We have studied different compatibility properties and found that they are equivalent in certain cases. If one of the two studied observables is sharp, i.e. a spectral measure, all given compatibility properties are equivalent[4], as represented in the following scheme. Here we have written JMOB for joint measurability of binarizations.

Commutativity

⇕

Joint measurability ⟺ Coexistence ⟺ JMOB

⇕

Common bi-observable

⇕

Smearings of a third observable

⇕

Functions of third observable

[4]Under the right topological assumptions on the outcome space.

If one of the observables if extreme and discrete, then the relations between the compatibility properties are given by the following scheme.

$$
\begin{array}{ccccc}
\text{Commutativity} & & & & \\
\not\Updownarrow & & & & \\
\text{Joint measurability} & \Longleftrightarrow & \text{Coexistence} & \overset{?}{\Longleftrightarrow} & \text{JMOB} \\
\Updownarrow & & & & \\
\text{Common bi-observable} & & & & \\
\Updownarrow & & & & \\
\text{Smearings of a third observable} & & & & \\
\Updownarrow & & & & \\
\text{Functions of third observable} & & & &
\end{array}
$$

There is one unknown in this scheme, namely whether joint measurability of binarizations is equivalent to coexistence under these conditions. Further research is needed here. There is also the hope that the rather arbitrary condition of discreteness of one of the observables can be lifted. If this is the case, then it will be much easier to verify joint measurability for extreme observables, the most important class of observables, from which all other observables can be obtained by affine combinations.

Appendices

A Functional Analysis

In this section we give a brief introduction to functional analysis, focusing on positive and trace-class operators. We also give an introduction to and state the spectral theorem. Although most fundamental definitions are given, a basic understanding of (functional) analysis is assumed. We will always write $\mathcal{H}$ for a separable Hilbert space. For some (simple) results we have omitted the proof and refer to any good introduction to functional analysis, such as [19] and [3, Ch. 2,3].

Definition A.1 (Inner product)**.** Let V be a vector space. We say that a mapping $h : V \times V \to \mathbb{C}$ is a *sesquilinear form*, if for all $\phi, \psi, \eta \in V$ and $\alpha, \beta \in \mathbb{C}$ we have

(a) $h(\phi, \alpha\psi + \beta\eta) = \alpha h(\phi, \psi) + \beta h(\phi, \eta)$;

(b) $h(\alpha\phi + \beta\psi, \eta) = \overline{\alpha} h(\phi, \eta) + \overline{\beta} h(\psi, \eta)$;

We denote the set of all sesquilinear forms on V by $\mathcal{S}(V)$. If we also have

$$h(\phi, \phi) \geq 0 \quad \forall \phi \in V$$

we call h a *positive sesquilinear form.* We denote the set of all positive sesquilinear forms on V by $\mathcal{S}^+(V)$. Furthermore, if for a postive sesquilinear form h we have

$$h(\phi, \phi) = 0 \iff \phi = 0,$$

we call h *strict* or *positive definite.* A strict positive sequilinear form is called an *inner product* on the vector space V and we call V equipped with h, an *inner product space.*

For an inner product space V we define its norm by $\|x\| = \sqrt{\langle x \,|\, x \rangle}\ \forall x \in V$. And for any normed linear space W we can define a metric $d : W \times W \to [0, \infty)$ by $d(x, y) = \|x - y\|$, which makes W into a metric space. We now quickly recapitulate the definitions of Banach and Hilbert spaces.

Definition A.2.

- We say that a metric space X is *complete* if every Cauchy sequence in X converges to an element in X.
- If X is a complete normed vector space, then we say that X is a *Banach space.*
- If $\mathcal{K}$ is a complete inner product space, we say that $\mathcal{K}$ is a *Hilbert space.*
- If $\mathcal{H}$ is a Hilbert space and admits a countable orthonormal basis, then we say that $\mathcal{H}$ is a *separable* Hilbert space.

Definition A.3 (Bounded linear operator)**.** Let V and W be normed linear spaces. A map $T : V \to W$ is called *linear* if

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y) \quad \forall x, y \in V, \alpha, \beta \in \mathbb{C}.$$

We say that a linear map T is a *bounded linear operator* from V to W if there is a finite constant C such that

$$\|Tx\|_W \leq C\|x\|_V.$$

We denote the set of bounded linear operators $T : V \to W$, by $\mathcal{B}(V, W)$ and we write $\mathcal{B}(V)$ for $\mathcal{B}(V, V)$.

Definition A.4. The norm of a bounded operator $T : V \to W$ is given by

$$\|T\| = \sup_{\xi \in V, \|\xi\| \leq 1} \|T\xi\|.$$

This norm is called the *operator norm.*

Proposition A.5. *On an inner product space V with induced norm $\|v\| = \sqrt{\langle v \mid v \rangle}$, the norm of $v \in V$ is equal to*

$$\|v\| = \sup\{|\langle v \mid w \rangle| \mid w \in V, \|w\| \leq 1\}.$$

We now immediately have the following corollary for the operator norm.

Corollary A.5.1. *The norm of an operator $T \in \mathcal{B}(\mathcal{H})$ is equal to*

$$\|T\| = \sup\{|\langle \psi \mid T\phi \rangle| \mid \psi, \phi \in \mathcal{H}, \|\psi\| \leq 1, \|\phi\| \leq 1\}.$$

Proposition A.6. *Let V be an inner product space.*

(a) If $\phi_1, \ldots, \phi_n \in V$ are vectors satisfying $\langle \phi_i \mid \phi_j \rangle = 0$, whenever $i \neq j$, then

$$\left\| \sum_{k=1}^{n} \phi_k \right\|^2 = \sum_{k=1}^{n} \|\phi_k\|^2.$$

(b) For all $\phi, \psi \in V$,

$$\|\phi + \psi\|^2 + \|\phi - \psi\|^2 = 2\|\phi\|^2 + 2\|\psi\|^2.$$

(c) If W is an complex inner product space, then for any $\phi, \psi \in W$,

$$\langle \phi \mid \psi \rangle = \frac{1}{4} \left(\|\phi + \psi\|^2 - \|\phi - \psi\|^2 + i\|\phi - i\psi\|^2 - i\|\phi + i\psi\|^2 \right).$$

All of the above equations are proved by simple calculations. The equation in (a) is the inner product space version of the Pythagorean theorem. The equation in (b) is called *parallelogram law* and the one in (c) is the *polarisation identity*, which shows that the norm of an inner product space completely determines the inner product from which it is induced.

Proposition A.7 (Parseval's identity)**.** *If $K \subseteq \mathcal{H}$ is an orthonormal basis, then*

$$\|\phi\|^2 = \sum_{\xi \in K} |\langle \xi \mid \phi \rangle| \quad \forall \phi \in \mathcal{H}.$$

A well known fact is that for any bounded operator T there is an operator T^*, called the *adjoint* of T.

Proposition A.8. *For any $T \in \mathcal{B}(\mathcal{H})$ there is an unique operator T^* such that for all $\phi, \psi \in \mathcal{H}$,*

$$\langle \phi \,|\, T\psi \rangle = \langle T^* \phi \,|\, \psi \rangle \,.$$

We always have $T^{**} = T$.

Definition A.9. We say that an operator $T \in \mathcal{B}(\mathcal{H})$ is *self-adjoint* if $T^* = T$.

Proposition A.10. *For an operator $T \in \mathcal{B}(\mathcal{H})$ we have the following equivalence,*

$$T \text{ self-adjoint} \quad \Longleftrightarrow \quad \langle \phi \,|\, T\phi \rangle \in \mathbb{R} \;\forall \phi \in \mathcal{H}.$$

Proof. For the implication to the right, we note that for $T \in \mathcal{B}(\mathcal{H})$ self-adjoint we have

$$\langle \phi \,|\, T\phi \rangle = \langle T\phi \,|\, \phi \rangle = \overline{\langle \phi \,|\, T\phi \rangle} \in \mathbb{C},$$

and conclude $\langle \phi \,|\, T\phi \rangle \in \mathbb{R}$. For the other implication, we note that $\langle \phi \,|\, T\phi \rangle = \langle T\phi \,|\, \phi \rangle$ follows immediately from its realvaluedness. Now by using the following variation of the polarization identity

$$\begin{aligned} \langle \phi \,|\, T\psi \rangle = \langle \phi + \psi \,|\, T(\phi + \psi) \rangle - \langle \phi - \psi \,|\, T(\phi - \psi) \rangle \\ + i \langle \phi - i\psi \,|\, T(\phi - i\psi) \rangle - i \langle \phi + i\psi \,|\, T(\phi + i\psi) \rangle \,, \end{aligned}$$

self-adjointness follows. □

Proposition A.11. *If $T \in \mathcal{B}(\mathcal{H})$ is self-adjoint, then*

$$\|T\| = \sup_{\|\phi\| \leq 1} |\langle \phi \,|\, T\phi \rangle|.$$

Proof. Using the polarisation identity and the parallelogram law, we obtain for $\phi, \psi \in \mathcal{H}$ with $\|\phi\| \leq 1, \|\psi\| \leq 1$,

$$\begin{aligned} |\mathrm{Re}\, \langle \phi \,|\, T\psi \rangle| &= \frac{1}{4} |\langle \psi + \phi \,|\, T(\psi + \phi) \rangle - \langle \psi - \phi \,|\, T(\psi - \phi) \rangle| \\ &\leq \frac{1}{4} M(\|\psi + \phi\|^2 + \|\psi - \phi\|^2) = \frac{1}{2} M(\|\psi\|^2 + \|\phi\|^2) \leq M, \end{aligned}$$

where $M = \sup_{\|\phi\| \leq 1} |\langle \phi \,|\, T\phi \rangle|$. Suppose now that $\|\phi\| \leq 1$ and $\|\psi\| \leq 1$. Choose $\alpha \in \mathbb{C}$ such that $|\alpha| = 1$ and $|\langle \phi \,|\, T\psi \rangle| = \alpha \langle \phi \,|\, T\psi \rangle = \langle \phi \,|\, T\alpha\psi \rangle$. Applying the first part of the proof to the $a\psi$ and ϕ yields $|\langle \phi \,|\, T\psi \rangle| = \mathrm{Re}\, \langle \phi \,|\, T\alpha\psi \rangle \leq M$, so that

$$\|T\| = \sup \{ |\langle \phi \,|\, T\psi \rangle| \mid \|\phi\| \leq 1, \|\psi\| \leq 1 \} \leq M.$$

On the other hand, $|\langle \phi \,|\, T\phi \rangle| \leq \|\phi\| \|T\| \leq \|T\|$ if $\|\phi\| \leq 1$, and so $M \leq \|T\|$, proving their equality. □

We will now introduce some classes of operators which are important in quantum mechanics.

Proposition A.12. *Any $T \in \mathcal{B}(\mathcal{H})$ can be written uniquely as $T = A + iB$, with $A, B \in \mathcal{B}(\mathcal{H})$ self-adjoint.*

Proof. If $T = A + iB$, and A and B are self-adjoint, we have $A = \frac{1}{2}(T + T^*)$ and $B = \frac{1}{2i}(T - T^*)$ and these operators always exist and are self-adjoint if $T \in \mathcal{B}(\mathcal{H})$. $\square$

Definition A.13 (Isometry)**.** An operator $T \in \mathcal{B}(\mathcal{H})$ is called an *isometry* if it satisfies

$$T^*T = \boldsymbol{I}_{\mathcal{H}}\,.$$

An equivalent definition for an isometry T is $\|T\psi\| = \|\psi\|$ for all $\psi \in \mathcal{H}$ and from this definition we immediately see that an isometry in injective, by noting $\|T\phi - T\psi\| = \|\phi - \psi\|$. Whenever an operator is an isometry and also is surjective we call it an *unitary operator*, which is equivalent with the following definition.

Definition A.14 (Unitary operator)**.** An operator $U \in \mathcal{B}(\mathcal{H})$ is called an *unitary operator* if it statisfies

$$U^*U = UU^* = \boldsymbol{I}_{\mathcal{H}}\,.$$

Since injectivity implies surjectivity in a finite dimension case, any isometry is an unitary operator in a finite dimensional Hilbert space .

Definition A.15 (Orthogonal projection)**.** An operator $P \in \mathcal{B}(\mathcal{H})$ is called an *(orthogonal) projection* if

$$P = P^2 = P^*.$$

We denote the set of all projection operators on $\mathcal{H}$ by $\mathbf{P}(\mathcal{H}) = \{P \in \mathcal{B}(\mathcal{H}) \mid P = P^2 = P^*\}$.

Proposition A.16. *For a linear map $P : \mathcal{H} \to \mathcal{H}$ satisfying $P = P^2$ we have*

$$P = P^* \quad \iff \quad \|P\| \leq 1.$$

Proof. Assume $P = P^*$ and $\phi \in \mathcal{H}$. By noticing

$$\|P\phi\|^2 = \langle P\phi \,|\, P\phi\rangle = \left\langle \phi \,\middle|\, P^2\phi \right\rangle = \langle \phi \,|\, P\phi\rangle \leq \|\phi\|\|P\phi\|,$$

we obtain $\|P\phi\| \leq \|\phi\|$, proving the implication to the right. For the other implication we refer to [3, Thm. 2.10]. $\square$

Definition A.17 (Positive operator)**.** We say that an operator $T \in \mathcal{B}(\mathcal{H})$ is *positive* if

$$\langle \phi \,|\, T\phi\rangle \geq 0 \quad \forall \phi \in \mathcal{H}.$$

We write $\mathcal{B}(\mathcal{H})^+$ for the set of all positive operators on $\mathcal{H}$. For $T, S \in \mathcal{B}(\mathcal{H})$ self-adjoint we write $S \leq T$ if $T - S \geq 0$.

Since $\langle \phi \,|\, T\phi\rangle \in \mathbb{R}$ implies self-adjointness, as proven in Proposition A.10, we see that a positive operator is self-adjoint.

Proposition A.18. *If $P \in \mathbf{P}(\mathcal{H})$ and $T \in \mathcal{B}(\mathcal{H})^+$, then the following conditions are equivalent:*

(a) $T(\mathcal{H}) \subseteq P(\mathcal{H})$ and $\|T\| \leq 1$;

(b) $T \leq P$.

In that case $PT = TP = T$.

Proof. First assume (a). For all $\phi \in \mathcal{H}$ we have

$$\langle \phi \,|\, T\phi \rangle = \langle \phi \,|\, PT\phi \rangle = \langle P\phi \,|\, TP\phi \rangle \leq \|P\phi\| \|T\| \|P\phi\| \leq \|P\phi\|^2 = \langle \phi \,|\, P\phi \rangle ,$$

since $PT = T$, so that $T = T^* = T^*P^* = TP$.
Now assume (b). We first note that $\|P\| \leq 1$, by observing

$$\|P\phi\|^2 = \langle P\phi \,|\, P\phi \rangle = \left\langle \phi \,\middle|\, P^2\phi \right\rangle = \langle \phi \,|\, P\phi \rangle \leq \|\phi\| \|P\phi\|,$$

and thus $\|P\phi\| \leq \|\phi\|$. Since the map $(\phi, \psi) \mapsto \langle \phi \,|\, T\psi \rangle$ is a positive sesquilinear form, for all $\xi, \phi \in \mathcal{H}$ we have by the Cauchy-Swartz inequality

$$|\langle \xi \,|\, T\phi \rangle|^2 \leq \langle \xi \,|\, T\xi \rangle \langle \phi \,|\, T\phi \rangle \leq \langle \xi \,|\, P\xi \rangle \langle \phi \,|\, P\phi \rangle .$$

In particular,

$$\|T\|^2 = \sup_{\xi \leq 1, \phi \leq 1} |\langle \xi \,|\, T\phi \rangle|^2 \leq \sup_{\xi \leq 1, \phi \leq 1} \|P\xi\| \|P\phi\| \leq 1,$$

and so $\|T\| \leq 1$. Furthermore, $T\phi = 0$ if $P\phi = 0$ and so $\langle \xi \,|\, T\phi \rangle = \langle T\xi \,|\, \phi \rangle = 0$ whenever $\phi \in \mathcal{H}$ and $\xi \in \mathbf{P}(\mathcal{H})^\perp$, implying $T\phi \in \mathbf{P}(\mathcal{H})^{\perp\perp} = \mathbf{P}(\mathcal{H})$, proving $T(\mathcal{H}) \subseteq P(\mathcal{H})$. □

Corollary A.18.1. *For $T \in \mathcal{B}(\mathcal{H})^+$, one has*

$$\|T\| \leq 1 \quad \iff \quad T \leq \boldsymbol{I} .$$

Proof. Apply Proposition A.18 to the identity $\boldsymbol{I}$, which is of course a projection onto $\mathcal{H}$. □

For a positive operator T one can define the *square root* of that operator, usually denoted by $\sqrt{T}$ or $T^{\frac{1}{2}}$.

Lemma A.19. *For a positive operator $T \in \mathcal{B}(\mathcal{H})^+$ there is a unique positive operator $\sqrt{T}$, satisfying $(\sqrt{T})^2 = T$.*

The square root enables us to state the following proposition.

Proposition A.20. *Any operator $T \in \mathcal{B}(\mathcal{H})$ can be written as a linear combination of four unitary operators.*

Proof. Let $T \in \mathcal{B}(\mathcal{H})$. From Proposition A.12 we know that there are $A, B \in \mathcal{B}(\mathcal{H})$ self-adjoint, such that $T = A + iB$, by dividing this by $m = \max(\|A\|, \|B\|)$ we obtain $\frac{T}{m} = C + iD$ with $C, D \in \mathcal{B}(\mathcal{H})$ self-adjoint and $\|C\| \leq 1, \|D\| \leq 1$. Define $U = C + i\sqrt{\boldsymbol{I} - C^2}$. Then $U^* = C - i\sqrt{\boldsymbol{I} - C^2}$ and we have $UU^* = U^*U = \boldsymbol{I}$ and $A = \frac{m}{2}(U + U^*)$. In the same way, B is the linear combination of two unitary operators and thus we conclude that T can be written as a linear combination of four unitary operators. □

By making the observation that $\langle \phi \,|\, T^*T\phi \rangle = \langle T\phi \,|\, T\phi \rangle = \|T\phi\|^2$ we see that for any bounded operator T, T^*T is a positive operator. This makes the following definition sound.

Definition A.21 (Absolute value)**.** For any $T \in \mathcal{B}(\mathcal{H})$, the positive operator $\sqrt{T^*T}$ is denoted by $|T|$ and is called the *absolute value* of T.

A.1 Trace-class operators

In this section we extend the definition of the trace of a matrix (i.e. the sum of diagonal elements) to a subset of bounded operators, the so-called trace-class operators.

Lemma A.22. *Let K and L be orthonormal bases in the Hilbert space $\mathcal{H}$. If $T \in \mathcal{B}(\mathcal{H})$, then:*

(a) $\displaystyle\sum_{\xi \in K} \langle \xi \,|\, T^*T\xi \rangle = \sum_{\eta \in K} \langle \eta \,|\, T^*T\eta \rangle \,.$

(b) $\displaystyle\sum_{\xi \in K} \|T\xi\|^2 = \sum_{\eta \in L} \|T\eta\|^2 = \sum_{\eta \in L} \|T^*\eta\|^2$

Proof. (a) Using Proposition A.7 and the fact that we may exchange the summation in double series with positive terms, we obtain:

$$\begin{aligned}\sum_{\xi \in K} \langle \xi \,|\, T^*T\xi \rangle &= \sum_{\xi \in K} \|T\xi\|^2 = \sum_{\xi \in K}\sum_{\eta \in L} |\langle \eta \,|\, T\xi \rangle|^2 \\ &= \sum_{\eta \in L}\sum_{\xi \in K} |\langle T^*\eta \,|\, \xi \rangle|^2 = \sum_{\eta \in L} \|T^*\eta\|^2 = \sum_{\eta \in K} \langle \eta \,|\, T^*T\eta \rangle \,.\end{aligned}$$

(b) The equality of the outer sides of the equation was already shown in the identity above. Taking $K = L$ we also obtain the remaining equality.

□

Corollary A.22.1. *Let K and L be orthonormal bases in the Hilbert space $\mathcal{H}$. If $T \in \mathcal{B}(\mathcal{H})^+$, then:*

$$\sum_{\xi \in K} \langle \xi \,|\, T\xi \rangle = \sum_{\eta \in K} \langle \eta \,|\, T\eta \rangle \,.$$

Proof. Apply Lemma A.22 on $T = \sqrt{T}\sqrt{T} = \sqrt{T}^*\sqrt{T} = \sqrt{T}\sqrt{T}^*$. □

Using this corollary, we can now define the *trace* base-independently for positive operators.

Definition A.23 (Trace)**.** We define the *trace* of an operator $T \in \mathcal{B}(\mathcal{H})^+$ by

$$\operatorname{tr}(T) = \sum_{\xi \in K} \langle \xi \,|\, T\xi \rangle \,.$$

Note that $\operatorname{tr}(T^*T) = \sum_{\xi \in K} \|T\xi\|^2$. In rest of this section, let K be an orthonormal basis of $\mathcal{H}$. Using the result above, we will define the set *Hilbert-Schmidt operators* and the set of *traceclass operators*, both independent of the orthonormal basis K.

Definition A.24 (Hilbert-Schimdt operator)**.** If $T \in \mathcal{B}(\mathcal{H})$ and $\operatorname{tr}(T^*T) < \infty$ we say that T is a *Hilbert-Schimdt operator*, and we write:

$$\mathcal{B}_2(\mathcal{H}) = \{T \in \mathcal{B}(\mathcal{H}) \mid \operatorname{tr}(T^*T) < \infty\},$$

for the set of all such operators. For a $T \in \mathcal{B}_2(\mathcal{H})$ we define the Hilbert-Schimdt norm $\|\cdot\|_2$ as follows:

$$\|T\|_2 = \sqrt{\operatorname{tr}(T^*T)}$$

Theorem A.25. *For all $S, T \in \mathcal{B}_2(\mathcal{H})$ and any orthonormal basis $\{\xi_i\}_{i\in\mathbb{N}}$ in $\mathcal{H}$, $\{\langle S\xi_i \mid T\xi_i\rangle\}_{i\in\mathbb{N}}$ is summable and*

$$\langle S \mid T\rangle = \mathrm{tr}(S^*T)$$

is independent of the choice of orthonormal basis K. Then $\mathcal{B}_2(\mathcal{H})$ is a linear subspace of $\mathcal{B}(\mathcal{H})$ and the mapping $\langle \cdot \mid \cdot \rangle$ as defined above is an inner product, with respect to which $\mathcal{B}_2(\mathcal{H})$ is a Hilbert space.

For a proof we refer to [3, p. 45-46].

Proposition A.26. *If $S \in \mathcal{B}_2(\mathcal{H}), T \in \mathcal{B}(\mathcal{H})$, then S^*, ST and TS are in $\mathcal{B}_2(\mathcal{H})$. Moreover*

$$\|ST\|_2 \leq \|S\|_2\|T\|,$$
$$\|TS\|_2 \leq \|T\|\|S\|_2 \text{ and}$$
$$\|S^*\|_2 = \|S\|_2.$$

Proof. We have

$$\sum_{\xi\in K} \|TS\xi\|^2 \leq \|T\|^2 \sum_{xi\in K} \|S\xi\|^2 = \|T\|^2\|S\|_2^2,$$

which shows that $TS \in \mathcal{B}_2(\mathcal{H})$ and proves the second inequality. From Lemma A.22 (b) we see that if $S \in \mathcal{B}_2(\mathcal{H})$ then $S^* \in \mathcal{B}_2(\mathcal{H})$ and $\|S^*\|_2 = \|S\|_2$. With this we ascertain $ST = (T^*S^*)^* \in \mathcal{B}_2(\mathcal{H})$. In addition

$$\|ST\|_2 = \|T^*S^*\|_2 \leq \|T^*\|\|S^*\|_2 = \|S\|_2\|T\|,$$

which proves the first inequality. □

Using Corollary A.22.1 we can define the trace class, also independent of choice of the orthonormal basis K.

Definition A.27 (Trace-class)**.** We write

$$\mathcal{B}_1(\mathcal{H}) = \{T \in \mathcal{B}(\mathcal{H}) \mid \mathrm{tr}(|T|) < \infty\}$$

and call this set the *trace class*.

In the proof of the next lemma, the *polar decomposition* of an operator is used. In short this entails that any operator $T \in \mathcal{B}(\mathcal{H})$ can be written as $T = U|T|$, where U is a bounded operator with some additional requirements.[5]

Lemma A.28. *We have the following identities for the trace class:*

(a) $\mathcal{B}_1(\mathcal{H}) = \{ST \mid S, T \in \mathcal{B}_2(\mathcal{H})\}$;

(b) $\mathcal{B}_1(\mathcal{H}) \subseteq \mathcal{B}_2(\mathcal{H})$.

[5]In fact U is an partially isometric operator whose initial projection is $\mathrm{supp}(|T|)$.

Proof. (a) We will prove both inclusions. So first assume $S,T \in \mathcal{B}_2(\mathcal{H})$ and let $U|ST|$ be the polar decomposition of ST. Now we see that

$$\sum_{\xi\in K} \langle \xi \,|\, |ST|\xi\rangle = \sum_{\xi\in K} \langle \xi \,|\, U^*ST\xi\rangle = \sum_{\xi\in K} = \sum_{\xi\in K} \langle (U^*S)^*\xi \,|\, T\xi\rangle \leq \|U^*S\|_2\|T\|_2,$$

where in the last step we have used the Cauchy-Schwartz inequality for the Hilbert-Schmidt norm. This shows that $ST \in \mathcal{B}_1(\mathcal{H})$. For the other inclusion, assume that $Q \in \mathcal{B}_1(\mathcal{H})$ and let $V|Q|$ its polar decomposition, so that $Q = V|Q|^{\frac{1}{2}}|Q|^{\frac{1}{2}}$. Now we see that $|Q|^{\frac{1}{2}} \in \mathcal{B}_2(\mathcal{H})$, since $Q \in \mathcal{B}_1(\mathcal{H})$ implies

$$\left\|(|Q|^{\frac{1}{2}})\right\|_2^2 = \sum_{\xi\in K} \left\langle |Q|^{\frac{1}{2}}\xi \,\middle|\, |Q|^{\frac{1}{2}}\xi \right\rangle = \sum_{\xi\in K} \langle \xi \,|\, |Q|\xi\rangle < \infty,$$

and $W|Q|^{\frac{1}{2}} \in \mathcal{B}_2(\mathcal{H})$ by Proposition A.26.

(b) This follows immediately from (a) and Proposition A.26.

□

Lemma A.29. *If K and L are orthonormal bases of $\mathcal{H}$, $T \in \mathcal{B}_1(\mathcal{H})$, then the sequences $(\langle \xi \,|\, T\xi\rangle)_{\xi\in K}$, $(\langle \eta \,|\, T\eta\rangle)_{\eta\in L}$ are summable, and*

$$\sum_{\xi\in K} \langle \xi \,|\, T\xi\rangle = \sum_{\eta\in L} \langle \eta \,|\, T\eta\rangle\,.$$

Proof. Let $T \in \mathcal{B}_1(\mathcal{H})$. From Lemma A.28 we know that there are $A, B \in \mathcal{B}_2(\mathcal{H})$ such that $T = AB$. By writing $\langle \xi \,|\, T\xi\rangle = \langle A^*\xi \,|\, B\xi\rangle$ and noting that $A^* \in \mathcal{B}_2(\mathcal{H})$, the result follows from Theorem A.25. □

With this lemma we can extend the trace, as defined in Definition A.23 base-independently to all operators in the trace class. We can also define the *trace norm.*

Definition A.30 (Trace norm)**.** For $T \in \mathcal{B}_1(\mathcal{H})$, we define the trace norm $\|\cdot\|_1$ by

$$\|T\|_1 = \operatorname{tr}(|T|).$$

Notice that $\operatorname{tr}(|T|) = \left\| |T|^{\frac{1}{2}} \right\|_2^2$.

Proposition A.31. *If $S \in \mathcal{B}_1(\mathcal{H}), T \in \mathcal{B}(\mathcal{H})$, then S^*, ST, and TS are in $\mathcal{B}_1(\mathcal{H})$.*

Proof. Let $T, S \in \mathcal{B}_1(\mathcal{H})$, from Lemma A.28 it follows that $T, S \in \mathcal{B}_2(\mathcal{H})$ and thus again with the same lemma one has that $TS, ST \in \mathcal{B}_1(\mathcal{H})$. Again from this lemma one can find operators $A, B \in \mathcal{B}_2(\mathcal{H})$ such that $S = AB$. From Proposition A.26 it follows that $A^*, B^* \in \mathcal{B}_2(\mathcal{H})$ and we conclude $S^* = B^*A^* \in \mathcal{B}_1(\mathcal{H})$. □

Proposition A.32. *The trace is a linear functional on $\mathcal{B}_1(\mathcal{H})$. Besides, for any $T \in \mathcal{B}_1(\mathcal{H})$ and $U \in \mathcal{B}(\mathcal{H})$ unitary we have*

$$\operatorname{tr}(T) = \operatorname{tr}(U^*TU).$$

Moreover, for any $S \in \mathcal{B}(\mathcal{H})$ we have

$$\operatorname{tr}(ST) = \operatorname{tr}(TS).$$

Proof. Linearity of the trace follows immediately from linearity of the inner product and the definition of the trace. Let $T \in \mathcal{B}_1(\mathcal{H})$ and $U \in \mathcal{B}(\mathcal{H})$ be unitary and K an orthonormal basis in K, then $UK = \{U\xi \mid \xi \in K\}$ is also an orthonormal basis of $\mathcal{H}$ and hence with Lemma A.29 we have

$$\operatorname{tr}(U^*TU) = \sum_{\xi \in K} \langle U\xi \,|\, TU\xi\rangle = \sum_{\eta \in UK} \langle \eta \,|\, T\eta\rangle = \operatorname{tr}(T).$$

Thus for $U \in \mathcal{B}(\mathcal{H})$ unitary we have

$$\operatorname{tr}(UT) = \operatorname{tr}(U^*UTU) = \operatorname{tr}(TU),$$

and since by Proposition A.20 any operator is a linear combination of four unitary operators, it follows from linearity of the trace that for any $S \in \mathcal{B}(\mathcal{H})$ we have $\operatorname{tr}(ST) = \operatorname{tr}(TS)$. □

Proposition A.33. *For all $S \in \mathcal{B}(\mathcal{H}), T \in \mathcal{B}_1(\mathcal{H})$ one has*

$$|\operatorname{tr}(ST)| \leq \|S\| \; \|T\|_1.$$

Proof. Let $T \in \mathcal{B}_1(\mathcal{H})$ and $S \in \mathcal{B}(\mathcal{H})$. By writing $U|T|$ for the polar decomposition of T, we have

$$\begin{aligned}|\operatorname{tr}(ST)| = |\operatorname{tr}(SU|T|)| = \left|SU|T|^{\frac{1}{2}}|T|^{\frac{1}{2}}\right| &= \left|\sum_{\xi \in K} \left\langle |T|^{\frac{1}{2}}U * S^*\xi \,\middle|\, |T|^{\frac{1}{2}}\xi \right\rangle\right| \\ &\leq \left\||T|^{\frac{1}{2}}U^*S^*\right\|_2 \left\||T|^{\frac{1}{2}}\right\|_2 \leq \left\||T|^{\frac{1}{2}}\right\|_2 \|U^*\|\|S^*\| \left\||T|^{\frac{1}{2}}\right\|_2 \\ &= \|S\|\|T\|_1,\end{aligned}$$

where for the first inequality we have used Cauchy-Swartz for the Hilbert-Schmidt inner-product from Theorem A.25. For the second inequality we used Proposition A.26 and in the final step we used the fact that for the polar decomposition $\|U\| = 1$ holds. □

Proposition A.34. *If $T \in \mathcal{B}_1(\mathcal{H}) \cap \mathcal{B}(\mathcal{H})^+$ and $S \in \mathcal{B}(\mathcal{H})^+$ then* $\operatorname{tr}(TS) > 0$.

Proof. Let $T \in \mathcal{B}_1(\mathcal{H})$ and $T, S \in \mathcal{B}(\mathcal{H})^+$. Since S is positive we can write $S = \sqrt{S}\sqrt{S}$. Now by Proposition A.32 we have

$$\operatorname{tr}(TS) = \operatorname{tr}(T\sqrt{S}\sqrt{S}) = \operatorname{tr}(\sqrt{S}T\sqrt{S}).$$

By observing

$$\left\langle \phi \,\middle|\, \sqrt{S}T\sqrt{S}\phi \right\rangle = \left\langle \sqrt{S}\phi \,\middle|\, \sqrt{T}\sqrt{T}\sqrt{S}\phi \right\rangle = \left\langle \sqrt{T}\sqrt{S}\phi \,\middle|\, \sqrt{T}\sqrt{S}\phi \right\rangle \geq 0 \quad \forall \phi \in \mathcal{H},$$

we see that $\sqrt{S}T\sqrt{S}$ is a positive operator and from the definition of the trace we conclude $\operatorname{tr}(TS) = \operatorname{tr}(\sqrt{S}T\sqrt{S}) \geq 0$. □

A.2 Spectral Theorem

We now give a brief introduction into the functional calculus and the spectral theorem.

Definition A.35 (Spectrum)**.** We define the *spectrum* $\sigma(T)$ of an operator $T \in \mathcal{B}(\mathcal{H})$ as

$$\sigma(T) = \{\lambda \in \mathbb{C} \mid T - \lambda \boldsymbol{I} \text{ is not invertible}\}.$$

Definition A.36 (Spectral radius)**.** For an operator $T \in \mathcal{B}(\mathcal{H})$ we define its *spectral radius*, $\mathrm{r}(T)$ as

$$r(T) = \max\{|\lambda| \mid \lambda \in \sigma(T)\}$$

There is a remarkably simple formula for the spectral radius.

Theorem A.37 (Spectral Radius Formula)**.** *For $T \in \mathcal{B}(\mathcal{H})$, one has,*

$$\mathrm{r}(T) = \lim_{n\to\infty} \|T^n\|^{\frac{1}{n}}.$$

Proposition A.38. *For a self-adjoint operator $T \in \mathcal{B}(\mathcal{H})$, we have the following identity for the spectral radius:*

$$\mathrm{r}(T) = \|T\|.$$

Proof. Let $T \in \mathcal{B}(\mathcal{H})$ be a self-adjoint operator and $\phi \in \mathcal{H}$ such that $\|\phi\| = 1$. Then one has

$$\|T\phi\|^2 = \langle T\phi \mid T\phi \rangle = \langle T^2\phi \mid \phi \rangle \leq \|T^2\phi\|\|\phi\| = \|T^2\phi\|,$$

from which we conclude: $\|T^2\| = \|T\|^2$. With induction we obtain $\|T^{2n}\| = \|T\|^{2n}$ for all $n \in \mathbb{N}$ and so

$$\mathrm{r}(T) = \lim_{n\to\infty} \|T^n\|^{\frac{1}{n}} = \lim_{n\to\infty} \|T\| = \|T\|.$$

□

Corollary A.38.1. *For $T \in \mathcal{B}(\mathcal{H})^+$, one has*

$$T \leq \boldsymbol{I} \iff \mathrm{r}(T) \leq 1.$$

Proof. This is an immeditate consequence of Corollary A.18.1 and Proposition A.38. □

We will now define the integral of a function with respect to a positive operator-valued measure (POVM). The definition of a POVM is given in Definition 2.1. For the rest of this section, assume that $(\Omega, \mathcal{A})$ is a measurable space. We say that $f = \sum_{i=1}^n \alpha_i \chi_{X_i}$ is an $\mathcal{A}$-simple function if $n \in \mathbb{N}$, $X_i \in \mathcal{A}, \alpha_i \in \mathbb{C}$ for all $i \in \{1, 2, \ldots, n\}$. When the measurable space is clear from the context, we just speak of simple functions.

Definition A.39. The *integral* of a $\mathcal{A}$-simple function $f = \sum_{i=1}^n \alpha_i \chi_{X_i}$ with respect to a POVM $E : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ is defined as

$$\int_\Omega f \, \mathrm{d}E = \sum_{i=1}^n \alpha_i E(X_i).$$

As one can simply check, the integral is independent of the way f is represented as a linear combination of characteristic functions. Now let $\mathcal{F}_\mathcal{A}$ denote the space of all bounded $\mathcal{A}$-measurable functions $f : \Omega \to \mathbb{C}$, and equip it with the supremum norm $\|f\|_\infty = \sup_{\omega\in\Omega} |f(\omega)|$. It can be shown that $\mathcal{F}_\mathcal{A}$ with this norm is a Banach space.

Lemma A.40. *If $f \in \mathcal{F}_{\mathcal{A}}$, then there exists a sequence $\{f_n\}$ of simple functions on Ω converging uniformly to f, such that $|f_n(\omega)| \leq |f(\omega)|\ \forall \omega \in \Omega$. In particular, the space of simple functions is dense in $\mathcal{F}_{\mathcal{A}}$.*

Proof. For each $n \in \mathbb{N}$, we consider $\{z \in \mathbb{C} \mid |z| \leq \|f\|\}$ as the union of a finite number of disjoint Borel sets $B_1^n, B_2^n, \ldots, B_{k_n}^n$ having diameter of at most $\frac{1}{n}$. For every $i \in \{1, 2, \ldots, k_n\}$, we can choose an element $z_i^n \in \overline{B_i^n}$ such that $|z_i^n| \leq |z|$ for all $z \in B_i^n$, since the closure $\overline{B_i^n}$ is compact. Now we define the simple function $f_n = \sum_{i=1}^{k_n} z_i^n \chi_{f^{-1}(B_i^n)}$ and we see that $\|f_n - f\| \leq \frac{1}{n}$ and $|f_n(\omega)| \leq |f(\omega)|$ for all $\omega \in \Omega$. □

Using this lemma, we can extend the operator integral of simple functions to all bounded measurable functions. In this proof we use the convention to write $E_{\phi,\psi}(X)$ for the positive measure $\langle \phi \mid E(X)\psi \rangle$ for a POVM $E : \mathcal{A} \to \mathcal{B}(\mathcal{H})$.

Proposition A.41. *Let $E : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ be a POVM. For any simple function $f : \Omega \to \mathbb{C}$, we have*

$$\left\| \int_\Omega f \,\mathrm{d}E \right\| \leq 2\|E(\Omega)\| \sup_{\omega \in \Omega} |f(\omega)|.$$

Moreover, the mapping $f \mapsto \int_\Omega f \,\mathrm{d}E$ on the space of simple functions on Ω can be uniquely extended to a bounded linear map $L : \mathcal{F}_{\mathcal{A}} \to \mathcal{B}(\mathcal{H})$.

Proof. First assume that f is a real valued simple function. If $\phi \in \mathcal{H}$ and $\|\phi\| \leq 1$, we have

$$\begin{aligned}\left|\left\langle \phi \,\middle|\, \left(\int_\Omega f \,\mathrm{d}E\right) \phi \right\rangle\right| &= \left|\left\langle \phi \,\middle|\, \left(\sum\nolimits_{i=1}^n \alpha_i E(X_i)\right) \phi \right\rangle\right| = \left|\sum\nolimits_{i=1}^n \alpha_i \langle \phi \mid E(X_i)\phi \rangle\right| \\ &= \left|\int_\Omega f \,\mathrm{d}E_{\phi,\phi}\right| \leq \|f\|_\infty E_{\phi,\phi}(\Omega) = \|f\|_\infty \langle \phi \mid E(\Omega)\phi \rangle\,,\end{aligned}$$

and since $\int_\Omega f \,\mathrm{d}E$ is self-adjoint, with Proposition A.11 we conclude $\left\|\int_\Omega f \,\mathrm{d}E\right\| \leq \|f\|_\infty \|E(\Omega)\|$. Now if $f = f_1 + i f_2$ for two real valued simple functions f_1 and f_2, we have

$$\begin{aligned}\left\|\int_\Omega f \,\mathrm{d}E\right\| &\leq \left\|\int_\Omega f_1 \,\mathrm{d}E\right\| + \left\|\int_\Omega f_2 \,\mathrm{d}E\right\| \\ &\leq (\|f_1\| + \|f_2\|)\|E(\Omega)\| \leq 2\|f\|\|E(\Omega).\|\end{aligned}$$

The second claim follows from Lemma A.40 and the completeness of $\mathcal{B}(\mathcal{H})$. □

We can now give the following definition.

Definition A.42. In the situation of Proposition A.41, for any measurable function f we call the bounded operator $L(f)$ the *integral* of the function f with respect to the POVM $E : \mathcal{A} \to \mathcal{B}(\mathcal{H})$, and we write

$$L(f) = L(f, E) = \int_\Omega f \,\mathrm{d}E = \int f \,\mathrm{d}E = \int_\Omega f(\omega) \,\mathrm{d}E(\omega).$$

Before we can state the spectral theorem, we first have to define $f(A)$ for a function f and self-adjoint operator A. If f is a polynomial, i.e. $f(x) = \sum_{i=0}^N \alpha_i x^i$, it is natural to define

$$f(A) = \sum_{i=0}^\infty \alpha_i A^i.$$

The next theorem states that we can extend this uniquely to any bounded Borel-measurable function f.

Theorem A.43 (Borel functional calculus)**.** *Let A be a bounded self-adjoint operator, and let $\mathcal{F}_\infty(\sigma(A))$ denote the set of bounded Borel-measurable functions on the spectrum of A. There exists a unique map*

$$\gamma : \mathcal{F}_\infty(\sigma(A)) \to \mathcal{B}(\mathcal{H}), \quad \gamma(f) = f(A),$$

such that:

1. $\gamma(\alpha f) = \alpha\gamma(f)$;
2. $\gamma(f + g) = \gamma(f) + \gamma(g)$;
3. $\gamma(fg) = \gamma(f)\gamma(g)$;
4. $f(A)^* = \overline{f(A)}$;
5. *if* $f(x) = x$*, then* $f(A) = A$;
6. $\|f(A)\| \leq \|f\|_\infty$ *with equality for f continuous;*
7. $\sigma(f(A)) = \overline{\{f(\lambda) \mid \lambda \in \sigma(A)\}}$.

The assignment $f \mapsto f(A)$ is called the *the Borel functional calculus.* We will now state the spectral theorem for bounded self-adjoint operators.

Theorem A.44 (Spectral Theorem for bounded self-adjoint operators)**.** *Let $\mathcal{H}$ is a separable Hilbert space and let A be a bounded self-adjoint operator in $\mathcal{B}(\mathcal{H})$. Let $\mathcal{F}$ denote the Borel subsets of $\sigma(A)$. There is a unique spectral measure $F : \mathcal{F} \to \mathcal{B}(\mathcal{H})$ with*

$$A = \int_{\mathcal{F}} z \, \mathrm{d}F.$$

Moreover, given any bounded Borel-measurable function f on $\sigma(A)$ we have

$$f(A) = \int_{\mathcal{F}} f \, \mathrm{d}F,$$

where on the left-hand side $f(A)$ is defined by the Borel functional calculus.

B Measure theory

In this section we will state the important definitions for measure theory. It is intended only as a reference. For a good introduction to measure theory we refer to [4].

Definition B.1 (σ-algebra)**.** Let Ω be an arbitrary set. A collection $\mathcal{A}$ of subsets of Ω is called a *σ-algebra* on Ω if:

(a) $\Omega \in \mathcal{A}$;

(b) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$;

(c) $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$;

(d) $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A} \implies \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$.

The pair $(\Omega, \mathcal{A})$ is called a *measurable space.*

Definition B.2 (Measure)**.** Let $(\Omega, \mathcal{A})$ be a measurable space and let $\mu : \mathcal{A} \to [-\infty, \infty]$ be a function.

(a) μ is called *σ-additive* (or *countably additive*) if, for any $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$, we have

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

(b) μ is called an *signed measure* if it is *σ-additive* and $\mu(\emptyset) = 0$.

(c) If μ is a signed measure and $\operatorname{ran}(\mu) \subseteq [0, +\infty]$, μ is called a *positive measure.*

(d) If μ is a signed measure and $\operatorname{ran}(\mu) \subseteq \mathbb{R}$, μ is called a *finite measure.*

(e) If μ is a signed measure, $\operatorname{ran}(\mu) \subseteq [0, 1]$, and $\mu(\Omega) = 1$, then μ is called a *probability measure.*

(f) If μ is a positive measure and Ω is the union of a sequence $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$ such that

$$\mu(A_i) \leq +\infty \quad \forall i \in \mathbb{N},$$

then μ is called a *σ-finite measure.*

(g) A σ-additive function $\nu : \mathcal{A} \to \mathbb{C}$ such that $\mu(\emptyset) = 0$, is called an *complex measure.*

If we say a function μ is a measure, we always mean a positive measure. In this case the triple $(\Omega, \mathcal{A}, \mu)$ is called a *measure space.*

Definition B.3 (Almost everywhere)**.** Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. A property of elements in Ω holds *μ-almost everywhere* if there is a set $N \in \mathcal{A}$ with $\mu(N) = 0$ that contains every point at which the property fails to hold.

We often write *μ-a.e.* for μ-almost everywhere, and if the measure μ is clear from the context, the expressions *almost everywhere* and *a.e.* are also used.

Definition B.4. Let $(\Omega, \mathcal{A})$ be a measurable space, let μ and ν be a positive measures on $(\Omega, \mathcal{A})$. We say that ν is *absolutely continuous with respect to* μ and write $\nu \ll \mu$ if

$$\mu(A) = 0 \implies \nu(A) = 0 \quad \forall A \in \mathcal{A}.$$

Definition B.5. Let $(\Omega, \mathcal{A})$ be a measurable space and let $A \in \mathcal{A}$. A function $f : A \to [-\infty, +\infty]$ is called *$\mathcal{A}$-measurable* (or *measurable with respect to* $\mathcal{A}$) if

$$\{x \in A \mid f(x) \leq t\} \in \mathcal{A} \quad \forall t \in \mathbb{R}.$$

Theorem B.6 (Radon-Nikodym Theorem)**.** *Let $(\Omega, \mathcal{A})$ be a measurable space, and let μ and ν be σ-finite measures on $(\Omega, \mathcal{A})$. If ν is absolutely continuous with respect to μ, then there is a $\mathcal{A}$-measurable function $g : \Omega \to [0, \infty)$ such that*

$$\nu(A) = \int_A g \, \mathrm{d}\mu \quad \forall A \in \mathcal{A}.$$

The function g is unique up to μ-almost everywhere equality.

A proof can be found in [4, Thm. 4.2.2, p. 123].

Definition B.7 (Weak Markov Kernel)**.** Let $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be measurable spaces and let μ be a σ-finite postive measure on $\mathcal{A}_1$. We say that $\beta : \Omega_1 \times \mathcal{A}_2 \to \mathbb{R}$ is a *weak Markov kernel* with respect to μ if:

(a) $\Omega_1 \ni \omega \mapsto \beta(\omega, X) \in \mathbb{R}$ is $\mathcal{A}$-measurable for all $X \in \mathcal{A}_2$;

(b) $\forall X \in \mathcal{A}_2$, $0 \leq \beta(\omega, X) \leq 1$ for μ-almost all $\omega \in \Omega_1$;

(c) $\beta(\omega, \mathcal{A}_2) = 1$ and $\beta(\omega, \emptyset) = 0$ for μ-almost all $\omega \in \Omega_1$;

(d) if $\{X_i\}_{i=1}^{\infty} \subseteq \mathcal{A}_2$ is a disjoint sequence, then $\beta(\omega, \bigcup_i X_i) = \sum_i \beta(\omega, X_i)$ for μ-almost all $\omega \in \Omega_1$.

Definition B.8 (Markov Kernel)**.** Let $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be measurable spaces and let μ be a σ-finite postive measure on $\mathcal{A}_1$. We say that $\beta : \Omega_1 \times \mathcal{A}_2 \to \mathbb{R}$ is a *Markov kernel* if it fullfils the conditions of a weak Markov kernel, with the modification that (b) - (d) hold for all $\omega \in \Omega_1$, instead for almost all ω. That is, for every $\omega \in \Omega_1$,

$$\beta(\omega, \cdot) : \mathcal{A}_2 \to \mathbb{R}, \quad X \mapsto \beta(\omega, X)$$

is a probability measure on A_2.

C Convexity

Definition C.1 (Convex set)**.** A subset C of a vector space V is called *convex* if for all $v, w \in C$ and $t \in (0, 1)$, one has

$$tv + (1 - t)w \in C.$$

Definition C.2. The *(extreme) boundary* of a convex set C, denoted by $\mathrm{ex}(C)$, is given by

$$\mathrm{ex}(C) = \{u \in C \mid u = tv + (1 - t)w \text{ for } v, w \in C, t \in (0, 1) \text{ implies } u = v = w\}.$$

Elements of the boundary are called *extreme elements.*

Example C.3. A classic example of a convex set is the three-dimensional unit ball

$$B^3 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid \|(x_1, x_2, x_3)\| \leq 1\}.$$

Indeed, if $x, y \in B^3$ and $s \in (0, 1)$ then

$$\begin{aligned}\|tx + (1 - t)y\| &\leq \|tx\| + \|(1 - t)y\| \\ &= t\|x\| + (1 - t)\|y\| \leq t + (1 - t) = 1,\end{aligned}$$

where the first inequality holds because of the triangle inequality.
Furthermore, the extreme boundary of B^3 is the 2-dimensional unit sphere

$$S^2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid \|(x_1, x_2, x_3)\| = 1\}.$$

To see this, let $x \in B^3$ such that $\|x\| = 1$ and assume there are $t \in (0, 1)$ and $y, z \in B^3$ such that

$$x = ty + (1 - t)z.$$

Assuming that $x = y = z$ does not hold, by taking norms and using the triangle inequality we see

$$1 < t\|y\| + (1-t)\|z\| \leq 1,$$

yielding a contradiction. Furthermore, if we assume that $x \in B^3$ with $\|x\| < 1$, we can always find an $0 < \epsilon < 1$ such that $\|(1+\epsilon)x\| < 1$, and thus

$$x = t((1+\epsilon)x) + (1-t)(\epsilon x)$$

holds for $t = 1 - \epsilon$.

◁

References

[1] Paul Busch, Marian Grabowski, and Pekka J Lahti. *Operational quantum physics*, volume 31. Springer, 1997.

[2] Paul Busch, Jukka Kiukas, and Pekka Lahti. On the notion of coexistence in quantum mechanics. *Mathematica Slovaca*, 60(5):665–680, 2010.

[3] Paul Busch, Pekka Lahti, Juha-Pekka Pellonpää, and Kari Ylinen. *Quantum measurement*. Springer, 2016.

[4] Donald L Cohn. *Measure theory*, volume 165. Springer, 1980.

[5] Edward Brian Davies. *Quantum theory of open systems*. Academic Press London, 1976.

[6] Erkka Haapasalo et al. *Extreme Observables and Optimal Measurements in Quantum Theory*. PhD thesis, University of Turku, 2015.

[7] Erkka Haapasalo and Juha-Pekka Pellonpää. Optimal quantum observables. *Journal of Mathematical Physics*, 58(12):122104, 2017.

[8] Erkka Haapasalo, Juha-Pekka Pellonpää, and Roope Uola. Compatibility properties of extreme quantum observables. *Letters in Mathematical Physics*, 105(5):661–673, 2015.

[9] Teiko Heinosaari and Juha-Pekka Pellonpää. Canonical phase measurement is pure. *Physical Review A*, 80(4):040101, 2009.

[10] Teiko Heinosaari and Juha-Pekka Pellonpää. Generalized coherent states and extremal positive operator valued measures. *Journal of Physics A: Mathematical and Theoretical*, 45(24):244019, 2012.

[11] Teiko Heinosaari, Daniel Reitzner, and Peter Stano. Notes on joint measurability of quantum observables. *Foundations of Physics*, 38(12):1133–1147, 2008.

[12] Tuomas Hytönen, Juha-Pekka Pellonpää, and Kari Ylinen. Positive sesquilinear form measures and generalized eigenvalue expansions. *Journal of Mathematical Analysis and Applications*, 336(2):1287–1304, 2007.

[13] Pekka Lahti. Coexistence and joint measurability in quantum mechanics. *International Journal of Theoretical Physics*, 42(5):893–906, 2003.

[14] Klaas Landsman. *Foundations of Quantum Theory: From Classical Concepts to Operator Algebras*. Springer, 2017.

[15] Juha-Pekka Pellonpää. Complete characterization of extreme quantum observables in infinite dimensions. *Journal of Physics A: Mathematical and Theoretical*, 44(8):085304, 2011.

[16] Juha-Pekka Pellonpää. On coexistence and joint measurability of rank-1 quantum observables. *Journal of Physics A: Mathematical and Theoretical*, 47(5):052002, 2014.

[17] Juha-Pekka Pellonpää and Jussi Schultz. On measurements of the canonical phase observable. *arXiv preprint arXiv:1303.0127*, 2013.

[18] David Reeb, Daniel Reitzner, and Michael M Wolf. Coexistence does not imply joint measurability. *Journal of Physics A: Mathematical and Theoretical*, 46(46):462002, 2013.

[19] Walter Rudin. *Functional analysis.* McGraw-Hill, Inc., New York, 1991.

[20] John von Neumann. *Mathematical Foundations of Quantum Mechanics.* Princeton University Press, 1955.