Foundations of Physics, Vol. 20, No. 3, 1990

Nonideal Quantum Measurements

Hans Martens[1] **and Willem M. de Muynck**[1]

Received June 5, 1989; revised September 22, 1989

A partial ordering in the class of observables (∼ positive operator-valued measures, introduced by Davies and by Ludwig) is explored. The ordering is interpreted as a form of nonideality, and it allows one to compare ideal and nonideal versions of the same observable. Optimality is defined as maximality in the sense of the ordering. The framework gives a generalization of the usual (implicit) definition of self-adjoint operators as optimal observables (von Neumann), but it can, in contrast to this latter definition, be justified operationally. The nonideality notion is compared to other quantum estimation theoretic methods. Measures for the amount of nonideality are derived from information theory.

1. INTRODUCTION

In the "operational" approach to statistical theories, measurements are associated with measures that are affine functionals on the state space. For quantum mechanics this means that observables are represented by *positive operator-valued measures* (POVM's).[5,4;9;14]

A POVM is (for a countable or finite outcome set K) a family of bounded linear operators $\{\mathbf{M}_k\}_{k \in K}$ satisfying the relations

$$\forall_{k \in K} \mathbf{M}_k \geqslant \mathbf{0} \tag{1}$$

$$\sum_{k \in K} \mathbf{M}_k = \mathbf{1} \tag{2}$$

(Operators are boldfaced.)

If the object system is in a state represented by the density operator $\boldsymbol{\rho}$, a $\{\mathbf{M}_k\}_{k \in K}$ measurement will result in outcome k with probability $\mathrm{Tr}(\boldsymbol{\rho}\mathbf{M}_k)$.

The state change that a measurement procedure induces in the object

[1] Department of Theoretical Physics, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

can also be treated in the operational framework. Since we are here only concerned with the determinative aspects of measurement, we will in this paper ignore the state changes accompanying a measurement. We shall therefore not touch upon issues concerned with, for example, "measurement of the first kind."

Of special interest is the subclass of *simple observables*, represented by *projection-valued measures* (PVM's). PVM's obey, in addition to (1) and (2),

$$\forall_{k \in K} \mathbf{M}_k^2 = \mathbf{M}_k \tag{3}$$

PVM's correspond to orthogonal spectral resolutions of self-adjoint operators (spectral theorem).

The class of observables associated with POVM's in general is, however, too large in a certain sense. It contains many "bad" observables that mix information about the object system with noninformation coming from the measurement device. One can think of a kind of "randomization" or "noise"[9] affecting our meter. If this noise totally dominates our device's operation, it measures an *uninformative observable*. Such an observable is represented by a POVM of the form

$$\forall_{k \in K} \mathbf{M}_k = f_k \mathbf{1} \qquad \forall_{k \in K} f_k \geqslant 0 \qquad \sum_{k \in K} f_k = 1 \tag{4}$$

(f_k a scalar).

The outcome probability distribution of such an uninformative observable does not depend on the state of the object system at all.

There is therefore a need for a structure in the class of observables that gives us a subclass of observables maximally free from this noninformation, observables whose outcomes are maximally associated with the object system alone. We could call such observables "optimal." If it were known what observables are optimal, we could restrict ourselves to this class, since the properties of all measurements are characterized by those of only the optimal ones via the structure.

In textbooks it is usually assumed implicitly, following von Neumann and Dirac,[6,18] that PVM's represent these "optimal" observables. This identification, probably inspired by classical mechanics (cf. Sec. 3), is not satisfactory, since it is not at all supported by operational arguments.

As a consequence, attempts have been made to generalize it. One proposal is based on the only structure on the class of POVM's systematically investigated so far: *convexity*. If two POVM's $\{\mathbf{M}_k\}_{k \in K}$ and $\{\mathbf{N}_k\}_{k \in K}$ with the same outcome set K are given, the set of operators $\{\mathbf{O}_k\}_{k \in K}$ defined by

$$\mathbf{O}_k := \lambda \mathbf{M}_k + (1 - \lambda) \mathbf{N}_k \tag{5}$$

is also a POVM for all $\lambda \in (0, 1)$.

Within the class of POVM's with a given outcome set K there are POVM's $\{\mathbf{O}_k\}_{k \in K}$ that cannot be decomposed into other POVM's as in (5). Such *extreme* POVM's correspond to *pure* observables. It has been suggested that pure observables constitute the subclass of "optimal" observables referred to above (Ref. 14, p. 135). This class contains all simple observables[14] and more, unless the outcome set K consists of two elements (*binary* POVM's).[9]

The convexity structure is, however, not suitable for the task that is to be performed here. Suppose an experimenter wants to realize a given observable, say the one associated with the position-PVM. In general he will only be able to realize a "nonideal" position measurement. The convexity structure offers no clue as to which observables are "nonideal" position observables, since it relates an observable to at least *two* other observables. In other words: convexity does not lead to a concept of nonideality that can be used to characterize the difference between the observable the experimenter wanted to measure and the observable he was able to realize.

One of the important characteristics a useful nonideality concept should have is (partial) invertibility. It is after all not the nonideal measurement itself we are interested in. We want to use its results to say something about the measurement we wanted to perform. In the above example our nonideal position measurement will be used to deduce properties of the exact position distribution. Convexity cannot do this.

A third disadvantage of the convexity structure is its inability to relate POVM's in a label-independent way. As a consequence, there seems to be no natural way to connect two POVM's with different outcome sets via convexity. Since these outcome sets are merely a matter of convenience (Ref. 14, p. 97) (nothing physical in the measurement device is changed if we alter its scale), an acceptable structure should be insensitive to the labeling of the observables.

The identification of pure and "optimal" observables is therefore unsatisfactory. To gain a better understanding of problems involving nonideality, such as the search for "optimal" observables, we introduce a nonideality structure that does not have the disadvantages convexity suffers from. After a definition of the relations generating the structure, a derivation of a number of its properties for quantum mechanics and classical mechanics is given (Secs. 2 and 3). The quantum observables that are minimally nonideal ("optimal") in the sense of this structure are derived (Sec. 2). The name "nonideality") in the sense of this structure are derived (Sec. 2). The name "nonideality" is operationalized (Sec. 4). A nonideality measure is proposed (Sec. 5).

The structure is well suited for application in areas where typically quantum mechanical measurement inaccuracy occurs. One such area is the joint measurement of incompatible observables. In another paper[15] we shall study that problem in detail.

2. DEFINITIONS AND STRUCTURAL PROPERTIES

In this paper we shall study our nonideality notion on a finite-dimensional complex Hilbert space $\mathscr{H}$. We ignore superselection rules. In this way we can get an idea of the properties the structure such a concept generates without having to deal with too many mathematical technicalities. We will also restrict ourselves to POVM's with a countably infinite outcome set $\mathbb{F}$.

A POVM $\mathscr{M} = \{\mathbf{M}_k\}_{k \in K}$ is an element of $(\mathscr{B}^+)^{\mathbb{F}}$ satisfying (2). Here $\mathscr{B}^+$ is the set of positive (bounded) linear operators on $\mathscr{H}$. The set $K \subseteq \mathbb{F}$ is now taken to be the *support* of $\mathscr{M}$:

$$K := \{k \in \mathbb{F} \mid \mathbf{M}_k \neq \mathbf{0}\}$$

We define for two POVM's $\mathscr{M} = \{\mathbf{M}_k\}_{k \in K}$ and $\mathscr{N} = \{\mathbf{N}_l\}_{l \in L}$ the following relation:

Definition 1.

$$\mathscr{N} \rightarrow \mathscr{M} := \exists_{\{\lambda_{kl}\} \in \mathbb{R}^{K \times L}} \begin{cases} \sum_{k \in K} \lambda_{kl} = 1 \\ \lambda_{kl} \geqslant 0, \\ \mathbf{M}_k = \sum_{l \in L} \lambda_{kl} \mathbf{N}_l. \end{cases}$$

The matrix $\{\lambda_{kl}\}$ is a stochastic matrix.[19] It is a property of a $\mathscr{M}$-device (*not* of the object, since it has no relation to the density operator), characterizing its relation to the observable corresponding to $\mathscr{N}$, insofar as determinative aspects are concerned.

A more restrictive version of the relation $\rightarrow$ is defined (for two POVM's $\mathscr{M}$ and $\mathscr{N}$ as above) by

Definition 2.

$$\mathscr{N} \rightarrow^i \mathscr{M} := \mathscr{N} \rightarrow \mathscr{M} \wedge \exists_{\{\mu_{lk}\} \in \mathbb{R}^{L \times K}} \mathbf{N}_l = \sum_{k \in K} \mu_{lk} \mathbf{M}_k$$

The matrix summations in both Definition 1 and Definition 2, as well as later series, are required to be elementwise absolutely convergent. This is necessary because the order of summation should not be relevant for the result. If the summation concerns only positive elements (such as in Definition 1, but *not* in Definition 2), absolute convergence follows from convergence.

These definitions can be considered elaborations of a relation used by Davies and by Allcock (who also notes the possibility of invertibility (Definition 2) explicitly; cf. related work by Schroeck).[3,4;1;21;23] A systematic investigation has never been performed, however.

In Definition 2 the POVM $\mathscr{M}$ is to represent a nonideal ("smeared") version of $\mathscr{N}$. If it is the case that $\mathscr{N} \rightarrow^i \mathscr{M}$, the "smearing" can be undone in a certain sense. We will go into these physical aspect of the structure more closely in Sec. 4.

Define the equivalence relation:

Definition 3. $\mathscr{M} \leftrightarrow \mathscr{N} := \mathscr{M} \rightarrow \mathscr{N} \wedge \mathscr{M} \leftarrow \mathscr{N}$.

If we define $\leftrightarrow^i$ analogously, it is trivial to show that:

Theorem. $\mathscr{M} \leftrightarrow \mathscr{N} \Leftrightarrow \mathscr{N} \leftrightarrow^i \mathscr{M}$.

Thus, as is easily verified, both $\rightarrow$ and $\rightarrow^i$ define a partial order relation between the equivalence classes defined by $\leftrightarrow$. In a partial order structure it is natural to define:

Definition 4. $\mathscr{M}$ is *maximal* $:= \forall_{\text{POVM } \mathscr{N}}(\mathscr{N} \rightarrow \mathscr{M} \Rightarrow \mathscr{N} \leftrightarrow \mathscr{M})$.

Definition 5. $\mathscr{M}$ is *minimal* $:= \forall_{\text{POVM } \mathscr{N}}(\mathscr{N} \leftarrow \mathscr{M} \Rightarrow \mathscr{N} \leftrightarrow \mathscr{M})$.

Using $\rightarrow^i$, we can define *i-maximality* and *i-minimality* in a similar way. We introduce the following notations:

$$\mathbf{A} \sim \mathbf{B} := \exists_{c \in \mathbb{R} \setminus \{0\}} \mathbf{A} = c\mathbf{B}$$

$$L(\mathscr{M}) := \left\{ \mathbf{X} \mid \exists_{(\alpha_k) \in \mathbb{R}^K} \mathbf{X} = \sum_{k \in K} \alpha_k \mathbf{M}_k \right\} \quad \text{(cf. Refs. 3 and 23)}$$

$$K(\mathscr{M}) := \left\{ \mathbf{X} \mid \exists_{(\alpha_k) \in \mathbb{R}^K; \alpha_k \geq 0} X = \sum_{k \in K} \alpha_k \mathbf{M}_k \right\}$$

[It can be verified that both L and K are closed. Obviously, if $\mathscr{N} \rightarrow \mathscr{M}$, then $K(\mathscr{M}) \subseteq K(\mathscr{N})$. If $\mathscr{N} \rightarrow^i \mathscr{M}$, we have, in addition to this, $L(\mathscr{N}) = L(\mathscr{M})$.]

$$B(\mathscr{M}) := \{ \mathbf{X} \in K(\mathscr{M}) \mid \mathrm{Tr}(\mathbf{1X}) = 1 \}$$

[The functional $f(\mathbf{X}) = \mathrm{Tr}(\mathbf{1}\,\mathbf{X})$ is a strictly monotonic linear functional on the cone K($:=$ $f(\mathbf{X})>0$ for all nonzero $\mathbf{X}$ in K). Hence B is a base for the cone K ($:=$ there is an $\alpha>0$ such that $\alpha\mathbf{X}\in B$ for all nonzero $\mathbf{X}$ in K).][10]

$$K_{\max}(\mathcal{M}) := L(\mathcal{M}) \cap \mathcal{B}^+$$

$B_{\max}(\mathcal{M})$ is defined analogously to B, as the base of $K_{\max}$

The extreme elements of a convex set C are denoted by $\partial_E C$; the elements of the *extremal rays*[10] of a cone K are denoted by ∂K:

$$\partial K := \{\mathbf{X}\in K \mid \forall_{\mathbf{Y}\in K} \mathbf{X}-\mathbf{Y}\in K \Rightarrow \mathbf{Y}\sim\mathbf{X}\}$$

Also useful is:

Definition 6. The POVM $\mathcal{M}$ is *pairwise linearly independent.*

$$:=$$

$$\forall_{k,l\in K}(\mathbf{M}_k \sim \mathbf{M}_l \Rightarrow k=l)$$

In a pairwise linearly independent POVM, no two nonzero elements lie on the same ray. The support of a pairwise linearly independent POVM is not unnecessarily large.

Our main results regarding structural properties are (the POVM's $\mathcal{N}$ and $\mathcal{M}$ as above; proofs can be found at the end of each section):

Theorem 1. In every equivalence class there is a pairwise linearly independent POVM, unique up to labeling.

This theorem characterizes the content of the equivalence classes. Maximal POVM's are characterized by:

Theorem 2. $\mathcal{M}$ is maximal $\Leftrightarrow$ $\forall_{k\in K}\mathbf{M}_k \in \partial\mathcal{B}^+$.

The set $\partial\mathcal{B}^+$ consists precisely of operators that are up to a scalar factor one-dimensional projectors. Note that Theorem 2 implies that there is more than one equivalence class of maximal POVM's. It also implies that our definition reduces to the usual one (= nondegeneracy) for PVM's.

Theorem 3. $\mathcal{M}$ is i-maximal $\Leftrightarrow$ $\forall_{k\in K}\mathbf{M}_k \in \partial K_{\max}(\mathcal{M})$.

Theorem 4. $\mathcal{M}$ is minimal $\Leftrightarrow$ $\forall_{k\in K}M_k \sim \mathbf{1} \Leftrightarrow \mathcal{M} \leftrightarrow \mathcal{I}$.

Hence uninformative observables are represented precisely by minimal POVM's. The trivial POVM $\{\mathbf{1}\}$ is denoted by $\mathcal{I}$.
Evident is (Definition 2):

Theorem. $\mathcal{M}$ is maximal $\Rightarrow$ $\mathcal{M}$ is i-maximal; $\mathcal{M}$ is minimal $\Rightarrow$ $\mathcal{M}$ is i-minimal.

Less obvious are perhaps:

Theorem 5. If $\dim(\mathscr{H})=2$, a POVM is i-maximal iff it is either minimal or maximal. This is not true if $\dim(\mathscr{H})>2$.

Theorem 6. $\mathcal{M}$ is minimal $\Leftrightarrow \mathcal{M}$ is i-minimal;
$\mathcal{M}$ is i-maximal $\Leftarrow \mathcal{M}$ is i-minimal.

The structure is closed in the sense that:

Theorem 7. For every POVM $\mathcal{M}$ there is a maximal POVM $\mathcal{N} \rightarrow \mathcal{M}$; for every POVM $\mathcal{M}$ there is a i-maximal POVM $\mathcal{N} \rightarrow^{i} \mathcal{M}$; for every POVM $\mathcal{M}$ there is a minimal POVM $\mathcal{N} \leftarrow \mathcal{M}$; this last assertion is not true for i-minimality and ${}^{i}\leftarrow$.

Proof of Theorem 1. An arbitrary POVM $\mathcal{M}=\{\mathbf{M}_k\}_{k\in K}$ is given. Divide K into subsets:

$$\mathscr{K} := \{I \subseteq K \mid \forall_{k\in I}[\forall_{m\in K}\mathbf{M}_k \sim \mathbf{M}_m \Leftrightarrow m \in I]\}$$

Obviously $\mathscr{K}$ can be mapped into $\mathbb{F}$. We have

$$\forall_{k\in K}\exists!_{J\in\mathscr{K}}\, k\in J$$

This allows us to define the POVM $\mathcal{N}=\{\mathbf{N}_I\}_{I\in\mathscr{K}}$:

$$\mathbf{N}_I := \sum_{k\in I} \mathbf{M}_k$$

This POVM is pairwise linearly independent by construction, and obviously $\mathcal{M} \leftrightarrow \mathcal{N}$. We have thus shown that there is a pairwise linearly independent POVM in every equivalence class. We shall now proceed to show that it is unique up to its labeling.

Suppose two POVM's $\mathcal{N}=\{N_l\}_{l\in L}$ and $\mathcal{O}=\{\mathbf{O}_m\}_{m\in M}$ are given such that $\mathcal{N}\leftrightarrow\mathcal{O}$ and such that both POVM's are pairwise linearly independent. There exist nonideality matrices $\{\lambda_{lm}\}$ and $\{\mu_{ml}\}$ such that

$$\mathbf{N}_l=\sum_{m\in M}\lambda_{lm}\mathbf{O}_m \quad (l\in L), \qquad \mathbf{O}_m=\sum_{l\in L}\mu_{ml}\mathbf{N}_l \quad (m\in M)$$

Define

$$\gamma_{km} := \tfrac{1}{2}\left(\delta_{km}+\sum_{l\in L}\mu_{kl}\lambda_{lm}\right) \qquad (k, m\in M) \tag{6}$$

Then

$$\mathbf{O}_k = \sum_{m \in M} \gamma_{km} \mathbf{O}_m \qquad (k, m \in M) \tag{7}$$

If we use the notation $p_m = \mathrm{Tr}(\boldsymbol{\rho} \mathbf{O}_m)$ ($\boldsymbol{\rho}$ arbitrary), this amounts to the eigenvalue equation $p_k = \sum_{m \in M} \gamma_{km} p_m$ (eigenvalue 1). A matrix like $\{\gamma_{km}\}$ can be used to represent a Markov chain with stationary transition probabilities.[2] Seen from that point of view, Eq. (7) means that this chain has a summable stationary distribution (one for every $\boldsymbol{\rho}$).

If a nonempty set $J \subseteq M$ has the property that $\forall_{m \in J} \sum_{k \in J} \gamma_{km} = 1$, we call it *closed.* If the set J has no proper closed subsets, it is *minimal closed* (or, equivalently, *essential*).[2] It can be shown that, given that (7) is satisfied and that $\forall_{m \in M} \mathbf{O}_m \neq \mathbf{0}$, a nonempty proper subset of M is closed if its complement in M is either closed or empty.

Now introduce the index set I, which we shall assume to be nonempty:

$$I := \{m \in M \mid \gamma_{mm} < 1\}$$

Then $\bar{I}$, the complement of I in M, is either empty or closed in M, which means I is also closed. Look at the smallest closed set $J_i \subseteq I$ containing $\{i\}$ for some $i \in I$. Since I is closed and not empty, there is such a subset. Then on J_i the eigenvalue equation reduces to

$$p_k = \sum_{m \in J_i} \gamma_{km} p_m \qquad (k \in J_i) \tag{8}$$

The matrix $\{\gamma_{km}\}_{(k,m) \in J_i \times J_i}$ has, because of (6), a period[2] equal to 1. Since the set J_i is minimal closed by construction, the stationary summable distribution is unique up to a scalar factor (Ref. 2, Theorem I.7.1). (For finite J_i this is a consequence of the Perron–Frobenius theorem; Ref. 19, p. 219.) This means that (8) has a solution

$$p_m = \lambda a_m \qquad (m \in M; \lambda \geqslant 0)$$

for some fixed nonnegative sequence (a_m). This would imply for (7) that $\mathbf{O}_m \sim \mathbf{O}_k$ if $k, m \in J_i$. Since $\mathcal{O}$ was pairwise linearly independent by assumption, this means that J_i cannot contain other elements than i. But then the eigenvalue 1 can only be achieved if $\gamma_{ii} = 1$, contradicting the definition of I.

Since p was arbitrary, I does not contain any minimal closed subset and must be empty. Hence $\gamma_{km} = \sum_{l \in L} \mu_{kl} \lambda_{lm} = \delta_{km}$. Similarly, one can prove that $\sum_{m \in M} \lambda_{km} \mu_{ml} = \delta_{kl}$. It is not difficult to see that this implies that $\{\lambda_{lm}\}$ and $\{\mu_{ml}\}$ must be permutation matrices. ∎

This means that the way we constructed a pairwise linearly independent POVM in the earlier part of this proof is essentially the only way.

Proof of Theorem 2.

$\Rightarrow$

Since our Hilbert space is finite dimensional, there is an affine decomposition into (not necessarily orthogonal) one-dimensional projectors for every positive operator, and consequently also for every $\mathbf{M}_k \in \{\mathbf{M}_k\}_{k \in K}$:

$$\mathbf{M}_k = \sum_{m=0,1,\ldots} c_{km} |\psi_{km}\rangle\langle\psi_{km}| \qquad (k \in K; c_{km} \geqslant 0) \tag{9}$$

The set $L := \{(k, m) \in \mathbb{F} \times \mathbb{N} \mid c_{km} \neq 0\}$ can be mapped into $\mathbb{F}$. Consequently this decomposition gives us a POVM $\mathscr{N} = \{\mathbf{N}_{km}\}_{(k,m) \in L}$, $\mathscr{N} \to \mathscr{M}$:

$$\mathbf{N}_{km} := c_{km} |\psi_{km}\rangle\langle\psi_{km}| \qquad ((k, m) \in L)$$

Hence, for a given POVM $\mathscr{M}$ there is always a POVM $\mathscr{N} = \{\mathbf{N}_l\}_{l \in L}$ such that $\mathscr{N} \to \mathscr{M}$ and $\forall_{l \in L} \mathbf{N}_l \in \partial\mathscr{B}^+$. But since we assumed $\mathscr{M}$ to be maximal,

$$\mathbf{N}_l = \sum_{k \in K} \lambda_{lk} \mathbf{M}_k \qquad (l \in L; \{\lambda_{lk}\} \text{ a nonideality matrix}) \tag{10}$$

We have $\forall_{l \in L} \mathbf{N}_l \in \partial\mathscr{B}^+$, which means that all nonzero elements on the right-hand side of (10) must be $\sim \mathbf{N}_l$. Since there can be no $k \in K$ for which there is no $l \in L$ such that $\lambda_{lk} > 0$, it follows that $\forall_{k \in K} \mathbf{M}_k \in \partial\mathscr{B}^+$.

$\Leftarrow$

Suppose that $\mathscr{M}$ satisfies the premise (i.e., $\forall_{k \in K} \mathbf{M}_k \in \partial\mathscr{B}^+$). We can, just as in the proof of Theorem 1, construct a pairwise linearly independent POVM $\mathscr{M}' = \{\mathbf{M}'_m\}_{m \in M}$ such that $\mathscr{M}' \leftrightarrow \{\mathbf{M}_k\}_{k \in K}$ and $\forall_{m \in M} \mathbf{M}'_m \in \partial\mathscr{B}^+$. Suppose there is a pairwise linearly independent POVM $\mathscr{N} = \{\mathbf{N}_l\}_{l \in L}$ satisfying $\mathscr{N} \to \mathscr{M}'$. Then there exists a nonideality matrix $\{\lambda_{ml}\}$ such that

$$\mathbf{M}'_m = \sum_{l \in L} \lambda_{ml} \mathbf{N}_l \qquad (m \in M)$$

Since $\forall_{m \in M} \mathbf{M}'_m \in \partial\mathscr{B}^+$ and $\mathscr{M}'$ is pairwise linearly independent, there is for every $l \in L$ precisely one $m \in M$ such that $\mathbf{N}_l \sim \mathbf{M}'_m$. Moreover, since $\sum_{m \in M} \lambda_{ml} = 1$, for every $l \in L$ such that $\mathbf{N}_l \neq \mathbf{0}$ there is an $m \in M$ such that $\mathbf{N}_l = \mathbf{M}'_m$. Hence $\mathscr{M} \to \mathscr{N}$. ∎

Before we proceed with the other proofs, we shall introduce another property a POVM can have, and prove a lemma for it:

Definition 7. The POVM $\mathscr{M} = \{\mathbf{M}_k\}_{k \in K}$ is *self-extremal* $:= \forall_{k \in K} \mathbf{M}_k \in \partial K(\mathscr{M})$.

[In a self-extremal POVM no element can be written as a nontrivial convex sum of other elements. Note that i-maximal POVM's are self-extremal.]

Lemma 1. For every POVM $\mathscr{M}$ there is a self-extremal and pairwise linearly independent POVM $\mathscr{N} = \{\mathbf{N}_l\}_{l \in L}$ such that $\mathscr{N} \to^i \mathscr{M}$ and $\forall_{l \in L} \exists_{k \in K} \mathbf{N}_l \sim \mathbf{M}_k$.

Proof of Lemma 1. The set $B(\mathscr{M})$ is closed since $K(\mathscr{M})$ is closed, and it is bounded since

$$\sup_{\mathbf{X}, \mathbf{Y} \in B(\mathscr{M})} \|\mathbf{X} - \mathbf{Y}\| \leqslant 2 \sup_{\mathbf{X} \in B(\mathscr{M})} \|\mathbf{X}\|$$
$$\leqslant 2 \sup_{\mathbf{X} \in B(\mathscr{M})} \mathrm{Tr}(\mathbf{X}) = 2 < \infty$$

The set $B(\mathscr{M})$ is also convex. It is a subset of the space $\mathscr{B}(\mathscr{H})$ of linear operators on $\mathscr{H}$. Since $\mathscr{H}$ is finite dimensional, so is $\mathscr{B}(\mathscr{H})$. Therefore, using Minkowski's and Carathéodory's theorems,[11] we see that every element of $B(\mathscr{M})$ can be written as a convex sum of finitely many elements of $\partial_E B(\mathscr{M})$. It is not difficult to see that[10]

$$\partial_E B(\mathscr{M}) = B(\mathscr{M}) \cap \partial K(\mathscr{M})$$

For these reasons there is a countable set $\{X_l\}_{l \in L} \subseteq \partial_E B(\mathscr{M})$ such that we can write

$$\mathbf{M}_k = \sum_{l \in L} \alpha_{kl} \mathbf{X}_l \qquad (k \in K; \alpha_{kl} \geqslant 0)$$

We can assume that for every $l \in L$ there is a $k \in K$ such that $\alpha_{kl} > 0$. If this were untrue for some $l_0 \in L$, we would not have included $\mathbf{X}_{l_0}$ in $\{\mathbf{X}_l\}_{l \in L}$. Moreover, $\sum_{k \in K} \alpha_{kl}$ has to be finite because $\sum_{k \in K} \mathbf{M}_k = \mathbf{1}$ and $\mathbf{X}_l \neq \mathbf{0}$ for all $l \in L$. Now define

$$\mathbf{N}_l := \left(\sum_{k \in K} \alpha_{kl} \right) \mathbf{X}_l \qquad (l \in L)$$
$$\lambda_{kl} := \alpha_{kl} \left(\sum_{k \in K} \alpha_{kl} \right)^{-1} \qquad (k \in K; l \in L)$$

It follows that $\mathscr{N} = \{\mathbf{N}_l\}_{l \in L}$ is a self-extremal and pairwise linearly independent POVM, and that $\{\lambda_{kl}\}$ is a nonideality matrix. Hence $\mathscr{N} \to \mathscr{M}$.

Since $\forall_{l \in L} \mathbf{N}_l \in K(\mathscr{M})$, there must be a nonnegative matrix $\{\beta_{lk}\}$ such that

$$\mathbf{N}_l = \sum_{k \in K} \beta_{lk} \mathbf{M}_k \qquad (l \in L)$$

Since also $\forall_{l \in L} \mathbf{N}_l \in \partial K(\mathcal{M})$, all nonzero terms on the right-hand side must be $\sim \mathbf{N}_l$. Consequently for every $m \in L$ there is an $l \in L$ such that $\mathbf{N}_m \sim \mathbf{M}_l$ ∎

Note that, although the relation of Lemma 1 is a stronger relation than $\to^i$, we cannot in general achieve equivalence.

Proof of Theorem 3.

$\Rightarrow$

We can, for a given POVM $\mathcal{M}$, construct a pairwise linearly independent POVM $\mathcal{N} = \{\mathbf{N}_l\}_{l \in L}$ such that $\mathcal{N} \to^i \mathcal{M}$ and $\forall_{l \in L} \mathbf{N}_l \in \partial K_{\max}(\mathcal{M})$, using an algorithm similar to the one we used in the proof of Lemma 1. Assume $\mathcal{M}$ is i-maximal. We must then have

$$\mathbf{N}_l = \sum_{k \in K} \lambda_{lk} \mathbf{M}_k \qquad (l \in L;\ \{\lambda_{lk}\} \text{ a nonideality matrix}) \tag{11}$$

Since $\forall_{l \in L} \mathbf{N}_l \in \partial K_{\max}(\mathcal{M})$, all nonzero terms on the right-hand side of (11) must be $\sim \mathbf{N}_l$. There can be no $k \in K$ for which there is no $l \in L$ such that $\lambda_{lk} > 0$, so it follows that $\forall_{k \in K} \mathbf{M}_k \in \partial K_{\max}(\mathcal{M})$.

$\Leftarrow$

Suppose that $\mathcal{M}$ satisfies the premise (i.e., $\forall_{k \in K} \mathbf{M}_k \in \partial K_{\max}(\mathcal{M})$). We can, just as in the proof of Theorem 1, construct a pairwise linearly independent POVM $\mathcal{M}' = \{\mathbf{M}'_m\}_{m \in M}$ satisfying $\forall_{m \in M} \mathbf{M}'_m \in \partial K_{\max}(\mathcal{M}) = \partial K_{\max}(\mathcal{M}')$ and $\mathcal{M}' n \mathcal{M}$. Suppose there is a pairwise linearly independent POVM $\mathcal{N} = \{\mathbf{N}_l\}_{l \in L}$ satisfying $\mathcal{N} \to^i \mathcal{M}$. Then there exists a nonideality matrix $\{\lambda_{ml}\}$ such that

$$\mathbf{M}'_m = \sum_{l \in L} \lambda_{ml} \mathbf{N}_l \qquad (m \in M)$$

Since $\forall_{m \in M} \mathbf{M}'_m \in \partial K_{\max}(\mathcal{M}')$ and $\mathcal{M}'$ is pairwise linearly independent, we have $\forall_{l \in L} \exists!_{m \in M} \mathbf{N}_l \sim \mathbf{M}'_m$. Combining this with $\sum_{m \in M} \lambda_{ml} = 1$ gives $\forall_{l \in L} \exists_{m \in M} \mathbf{N}_l = \mathbf{M}'_m$. Hence $\mathcal{M} \to \mathcal{N}$. ∎

The proof of Theorem 4 is simple, if one realizes that every POVM $\mathcal{N}$ such that $\mathcal{N} \leftarrow \mathcal{I}$ is equivalent to $\mathcal{I}$, and that for every POVM $\mathcal{N}$ it is true that $\mathcal{I} \leftarrow \mathcal{N}$.

Proof of Theorem 5. If $\dim(\mathcal{H}) = 2$, we only have to show that if $\mathcal{M}$ is i-maximal and not minimal, it is maximal. We can always write

$$\mathbf{M}_k = c_{k1} \mathbf{E}_k + c_{k2} \mathbf{E}_k^{\perp} = (c_{k1} - c_{k2})\, \mathbf{E}_k + c_{k2} \mathbf{1} \qquad (k \in K)$$

Here $\mathbf{E}_k$ is a projector onto a one-dimensional subspace. We assume (without loss of generality) that $c_{k1} \geqslant c_{k2}$ for all $k \in K$. There must be at least one $k_0 \in K$ for which $c_{k_0 1} > c_{k_0 2}$, since $\mathscr{M}$ is not minimal (Theorem 4). This means that

$$c_{m1} = c_{m2}(k_0 \neq m \in K) \Rightarrow \mathbf{M}_m \in L(\mathscr{I}) \subseteq K(\mathscr{M}) \backslash \partial K(\mathscr{M})$$

Since $\mathscr{M}$ is i-maximal, there can be no such m (Theorem 3). Consequently, we must have $c_{k1} > c_{k2}$ for all $k \in K$. Then it follows that

$$\partial K(\mathscr{M}) = \{\mathbf{X} \mid \exists_{k \in K} \mathbf{0} \leqslant \mathbf{X} \sim \mathbf{E}_k\}$$

Since $\forall_{k \in K}\{\mathbf{E}_k, \mathbf{E}_k^{\perp}\} \subseteq \partial\mathscr{B}^+$ and an i-maximal POVM is self-extremal, this means (Theorem 2) that $\mathscr{M}$ is maximal.

If $\dim(\mathscr{H}) > 2$, the PVM $\{\mathbf{E}, \mathbf{E}^{\perp}\}$ $(\mathbf{E} \neq \mathbf{0}, \mathbf{E} \neq \mathbf{1})$ is i-maximal, but neither maximal nor minimal. ∎

Proof of Theorem 6. For the first part we only need to prove that an i-minimal POVM is minimal Suppose $\mathscr{M}$ is i-minimal, but not minimal. Choose a POVM $\mathscr{N}$ with the same support as $\mathscr{M}$, such that $\forall_{k \in K} \mathbf{N}_k = c_k \mathbf{1} \neq \mathbf{0}$. Define

$$\mathbf{O}_k := \lambda \mathbf{M}_k + (1 - \lambda)\, \mathbf{N}_k \qquad (k \in K; 0 < \lambda < 1)$$

Obviously $\mathscr{O} = \{\mathbf{O}_k\}_{k \in K}$ is a POVM satisfying $\mathscr{M} \to^i \mathscr{O}$. Because of our assumption, we must also have $\mathscr{O} \to \mathscr{M}$, or

$$\mathbf{M}_k = \sum_{l \in K} \gamma_{kl} \mathbf{M}_l \qquad (k \in K)$$

with

$$\gamma_{kl} = \lambda \mu_{kl} + (1 - \lambda) \left(\sum_{m \in K} \mu_{km} c_m \right)$$

$$\{\mu_{km}\} \text{ a nonideality matrix for } \mathscr{O} \to \mathscr{M}$$

It can be seen that for the matrix $\{\gamma_{kl}\}$ the whole index set K is minimal closed. Hence we can use reasoning similar to that of the proof of Theorem 1 to complete this proof.

The second part of the theorem is immediately clear if one considers Theorem 4, Definition 2 and the above. ∎

Proof of Theorem 7. First note that for a given POVM $\mathscr{M}$ one can find a maximal POVM $\mathscr{N} \to \mathscr{M}$ simply by using (9). Because the decomposition (9) is in general not unique, there exist in general many non-equivalent maximal POVM's related to a given POVM.

We now proceed with the second assertion of the theorem. For a given POVM $\mathscr{M}$ one can find an i-maximal POVM $\mathscr{N} \to^i \mathscr{M}$ by repeating the construction of the proof of Lemma 1, using $K_{\max}(\mathscr{M})$ instead of $K(\mathscr{M})$. Because the corresponding base $B_{\max}(\mathscr{M})$ is not always a simplex, this i-maximal POVM is in general also not unique.

The third assertion of the theorem is trivial. The fourth is a straightforward corollary of Theorem 6. ∎

3. MISCELLANEOUS RESULTS

In this section we shall give some other results characterizing the structure. Moreover, these results can be used to compare it to approaches that are more common in the literature.

The connection between the structure induced by $\to$ and the convexity structure is indicated by the following theorems:

Theorem 8. If the POVM $\mathscr{M}$ is extreme, then it is i-maximal and pairwise linearly independent and $B(\mathscr{M})$ is a simplex. The converse is true only if $\dim(\mathscr{H}) = 2$.

Theorem 9. If a POVM $\mathscr{M}$ is maximal or minimal and it is pairwise linearly independent and $B(\mathscr{M})$ is a simplex, then it is extreme. The converse is true only if $\dim(\mathscr{H}) = 2$.

Since PVM's are always extreme,[9,14] PVM's are i-maximal and pairwise linearly independent. A maximal PVM is the PVM associated with a complete orthonormal basis. We can therefore conclude that a maximal POVM on $\mathbb{C}^n$ must have at least n elements.

In special cases POVM's can be related to PVM's via the nonideality structure (Ref. 7, p. 87; Ref. 8):

Theorem 10. For every binary POVM $\mathscr{M} = \{\mathbf{M}_1, \mathbf{M}_2\}$, there is a PVM $\mathscr{E} = \{E_k\}_{k \in K}$ such that $\mathscr{E} \to \mathscr{M}$. Suppose $\mathscr{M} = \{\mathbf{M}_k\}_{k \in K}$ is a POVM. Then $\forall_{k,l \in K}[\mathbf{M}_k, \mathbf{M}_l]_- = \mathbf{0} \Leftrightarrow \exists_{\mathrm{PVM}\,\mathscr{E}}\, \mathscr{E} \to \mathscr{M}$.

This theorem is easily proven (the first assertion follows from the fact that $[\mathbf{M}_1, \mathbf{M}_2]_- = [\mathbf{M}_1, 1 - \mathbf{M}_1]_- = \mathbf{0}$).

An interesting feature of the relation $\to$ is its connection to the representation of a POVM in terms of a PVM on a larger Hilbert space. This procedure is common in quantum estimation theory. It is based on a theorem due to Naĭmark (for the proof of which see, e.g., the book by Holevo[9]):

Theorem 11. For every POVM $\{\mathbf{M}_k\}_{k\in K}$ on $\mathscr{H}$ there is a Hilbert space $\mathscr{H}'$, a density operator $\boldsymbol{\rho}'$ on $\mathscr{H}'$, and a PVM $\{\mathbf{E}_k\}_{k\in K}$ on $\mathscr{H}\otimes\mathscr{H}'$ such that

$$\mathbf{M}_k = \mathrm{Tr}_{\mathscr{H}'}(\boldsymbol{\rho}'\mathbf{E}_k) \tag{12}$$

The link is given by:

Theorem 12. Suppose $\mathscr{F} = \{\mathbf{F}_a\}_{a\in A}$ is given as the PVM of $\mathbf{A} = \sum_{a\in A} a\mathbf{F}_a$ on $\mathscr{H}$. If $\{\mathbf{E}_b\}_{b\in B}$ on $\mathscr{H}\otimes\mathscr{H}'$ is the PVM of $\mathbf{B} = \sum_{b\in B} b\mathbf{E}_b$ and $\mathbf{B}$ can be written as a function $b(\mathbf{A}, \mathbf{C})$ of $\mathbf{A}$ and an Hermitian operator $\mathbf{C}$ on $\mathscr{H}'$, then $\mathscr{M} = \{\mathbf{M}_b\}_{b\in B}$ defined as in (12) has the property $\mathscr{F} \to \mathscr{M}$ for every density operator $\boldsymbol{\rho}'$ on $\mathscr{H}'$.

The proof of Theorem 12 is straightforward, using the spectral theorem on $\mathscr{H}'$.

Besides Naïmark's theorem, *covariance* and *unbiasedness* are key concepts in quantum estimation theory (Ref. 7; Ref. 9, Ch. 3; cf. also Refs. 12 and 17). We will see whether these concepts are related to our non-ideality relation. Define on $\mathscr{H} = \mathbb{C}^n$ a PVM $\mathscr{E} = \{\mathbf{E}_k\}_{k\in K}$:

$$\mathbf{E}_k := |x_k\rangle\langle x_k| \qquad k\in K = \{0,\dots, n-1\}$$

where $(|x_k\rangle)_{k\in K}$ is an orthonormal base for $\mathscr{H}$, and

$$\mathbf{S}_x := \sum_{k\in K} |x_{[[k+1]]}\rangle\langle x_k| \qquad [[k]] := k \bmod n$$

$$\mathbf{X} := \sum_{k\in K} k\mathbf{E}_k$$

Covariance and unbiasedness are defined by

Definition 8. The POVM $\mathscr{O} = \{\mathbf{O}_k\}_{k\in K}$ is covariant

$$:=$$

$$\forall_{k\in K}\mathbf{S}_x\mathbf{O}_k\mathbf{S}_x^\dagger = \mathbf{O}_{[[k+1]]}$$

The POVM $\mathscr{N} = \{\mathbf{N}_l\}_{l\in L}$ with the labeling $(f_l)_{l\in L}$ is unbiased

$$:=$$

$$\sum_{l\in L} f_l\mathbf{N}_l = \mathbf{X}$$

($\mathbf{S}_x$ and K as above; L arbitrary)
Then we have the following theorem:

Theorem 13. For $\mathscr{E}$ and K as above and $\mathcal{O}=\{\mathbf{O}_k\}_{k\in K}$ a POVM,

$$\left.\begin{aligned}&\forall_{k\in K}\mathbf{S}_x\mathbf{O}_k\mathbf{S}_x^\dagger=\mathbf{O}_{[[k+1]]}\\ &\exists_{(f_k)_{k\in K}}\sum_k f_k\mathbf{O}_k=\mathbf{X}\end{aligned}\right\}\Rightarrow\mathscr{E}\rightarrow^i\mathcal{O}$$

Note that neither covariance nor unbiasedness (for some labeling) is by itself sufficient for $\rightarrow$ or $\rightarrow^i$. There is, however, a stronger relation between $\rightarrow$ and another symmetry property. To show this, we introduce a second orthonormal base $(|y_k\rangle)_{k\in K}$ on $\mathbb{C}^n$, such that

$$\langle x_k|\,y_l\rangle=\frac{1}{\sqrt{n}}\exp\left(i\frac{2\pi}{n}kl\right)$$

and a shift operator $\mathbf{S}_y$ analogous to $\mathbf{S}_x$. Then

Theorem 14. For $\mathscr{E}$ and $\mathbf{S}_y$ as above and $\mathcal{O}=\{\mathbf{O}_l\}_{l\in L}$ a POVM,

$$\mathscr{E}\rightarrow\mathcal{O}\Leftrightarrow\forall_{l\in L}\mathbf{S}_y\mathbf{O}_l\mathbf{S}_y^\dagger=\mathbf{O}_l\qquad(\textit{invariance})$$

We see that symmetry properties are quite usable within this structure, although its definition does not depend on them. In particular, we have shown that a covariant observable that can be labeled such that it is unbiased, satisfies Definition 2. The demands of Definition 8 represent only a very special kind of nonideal measurement, however. Hence the quantum estimation theoretic approach to certain problems through Definition 8 is not definitive.

Note that, since the bases $(|x_k\rangle)_{k\in K}$ and $(|y_k\rangle)_{k\in K}$ are finite-dimensional analogs of position and momentum (we have, e.g., $\mathbf{S}_y^c\mathbf{S}_x^d=\mathbf{S}_x^d\mathbf{S}_y^c\exp(i(2\pi/n)\,cd)$ for all $c,d\in\mathbb{Z}$[22]) results like Theorem 14 are also of interest outside this finite-dimensional context.

At this point it may be of interest to investigate whether the aspects of the structure considered so far are purely quantal. In a classical analog (Ref. 9, Ch. I) we have a finite phase space $\Omega:=\{\omega_1,\ldots,\omega_n\}$. The states are given by probability distributions $(p_\omega)_{\omega\in\Omega}$; $\forall_{\omega\in\Omega}p_\omega\geqslant 0$; $\sum_{\omega\in\Omega}p_\omega=1$. An observable is represented in such a classical model by a set $\{f_k(\omega)\}_{k\in K}$ of functions (a *positive function-valued measure*, PFVM) that satisfies

$$\forall_{k\in K,\omega\in\Omega}f_k(\omega)\geqslant 0\qquad\forall_{\omega\in\Omega}\sum_{k\in K}f_k(\omega)=1$$

The probability of outcome k is given by $\sum_{\omega\in\Omega}p_\omega f_k(\omega)$

The most characteristic difference with the quantum case is that here the equivalence class of maximal PFVM's is *unique*. All maximal PFVM's are equivalent to the PFVM $\{g_k(\omega)\}_{k\in K}$; $g_k(\omega):=\delta_{k\omega}$; $K=\Omega$. This reflects the fact that in classical models any measurement can be related to a determination of a property of the object system. The "maximal" property is the system's "state." Therefore the function $f_k(\omega)$ is interpretable as a conditional probability. The uniqueness of the equivalence class of maximal PFVM's also implies the compatibility of all classical observables.[15]

The structure induced by $\rightarrow^i$ for the classical case is simpler in similar ways. This is a consequence of the fact that the set of PFVM's on Ω is isomorphic with respect to $\rightarrow^i$ and $\rightarrow$ (and convexity) to the set of POVM's $\mathscr{M} \leftarrow \mathscr{N}$; $\mathscr{N} = \{\mathbf{N}_1,..., \mathbf{N}_n\}$ a fixed POVM with $\forall_k \|\mathbf{N}_k\| = 1$ (e.g., a PVM; cf. the second part of Theorem 10).

We can therefore conclude that most of the intricacies the structure has in quantum theory are indeed unclassical, much as this is the case for the convexity structure.[9]

Note that, if $\dim(\mathscr{H}) = 2$, the conditions of Theorem 8 and Theorem 9 coincide because of Theorem 5. Hence, we do not need to prove explicitly the converse part of Theorem 8 and Theorem 9 if $\dim(\mathscr{H}) = 2$. We shall also use the fact that i-maximal POVM's are self-extremal (Definition 7).

Proof of Theorem 8. First the counterexample. If $\dim(\mathscr{H}) > 2$, there is a PVM $\{\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3\}$ $(\forall_i \mathbf{E}_i \neq 0)$. Define $\mathscr{M} = \{\mathbf{M}_1, \mathbf{M}_2\} := \{\mathbf{E}_1 + \frac{1}{2}\mathbf{E}_2, \frac{1}{2}\mathbf{E}_2 + \mathbf{E}_3\}$. This binary POVM is i-maximal, pairwise linearly independent, and $B(\mathscr{M})$ is a simplex. It is not extreme.

The major part of the proof consist of four stages:

(a) First assume that $\mathscr{M}$ is not pairwise linearly independent. Then (proof of Theorem 1, first part) there is a pairwise linearly independent POVM $\mathscr{O} = \{\mathbf{O}_l\}_{l \in L}$ such that $\mathscr{O} \rightarrow \mathscr{M}$ and

$$\mathbf{M}_k = \sum_{l \in L} \lambda_{kl} \mathbf{O}_l (k \in K)$$

Here $\{\lambda_{kl}\}$ is a nonideality matrix such that $\forall_k \exists!_l \lambda_{kl} > 0$. The matrix $\{\lambda_{kl}\}$ cannot consist entirely of 1's and 0's because $\mathscr{M}$ is not pairwise linearly independent. Such a matrix $\{\lambda_{kl}\}$ can be written as a convex sum of two different nonideality matrices. Consequently $\mathscr{M}$ can be written as a convex sum of two other POVM's, and is not extreme.

(b) Now assume that $\mathscr{M}$ is pairwise linearly independent, but not self-extremal (Definition 7). Then (Lemma 1) there is a pairwise linearly independent self-extremal POVM $\mathscr{O} = \{\mathbf{O}_l\}_{l \in L}$ such that $\mathscr{O} \rightarrow \mathscr{M}$ and

$$\mathbf{M}_k = \sum_{l \in L} \lambda_{kl} \mathbf{O}_l \qquad (k \in K)$$

$$\forall_{l \in L} \exists_{k \in K} \forall_{m \neq l} \lambda_{km} = 0$$

The matrix $\{\lambda_{kl}\}$ cannot consist entirely of 1's and 0's because $\mathcal{M}$ is not self-extremal. We can use the above reasoning (stage a) to complete stage b.

(c) Suppose $\mathcal{M}$ is self-extremal and pairwise linearly independent, but $B(\mathcal{M})$ is not a simplex. Since $L(\mathcal{M})$ must be finite dimensional, there must be a finite set $J \subseteq K$ such that $B(\{\mathbf{M}_k\}_{k\in J})$ is not a simplex ($\Leftrightarrow$ the number of elements of J is greater than $\dim\{L(\{\mathbf{M}_k\}_{k\in J})\}$). Then there is an operator $\mathbf{X} \in K(\{\mathbf{M}_k\}_{k\in J})$ that can be written as $\mathbf{X} = \sum_{k\in J} v_k \mathbf{M}_k$ for *two* distinct nonnegative finite sequences $(v_k)_{k\in J}$. Hence there is a nontrivial bounded sequence $(a_k)_{k\in K}$ such that $\sum_{k\in K} a_k \mathbf{M}_k = 0$. Define

$$\mathbf{M}_k^{\pm} := (1 \pm \lambda a_k)\, \mathbf{M}_k \qquad (k \in K;\, 0 < \lambda < (\sup_{k\in K} (|a_k|))^{-1})$$

Under these conditions $\{\mathbf{M}_k^{\pm}\}_{k\in K}$ are different POVM's, and

$$\mathbf{M}_k = \tfrac{1}{2}(\mathbf{M}_k^{+} + \mathbf{M}_k^{-}) \qquad (k \in K)$$

Hence $\mathcal{M}$ is not extreme.

(d) Now suppose $B(\mathcal{M})$ is a simplex, $\mathcal{M}$ is self-extremal and pairwise linearly independent, but not i-maximal. Then [proof of Theorem 7 (Part 2), Lemma 1] there is a pairwise linearly independent i-maximal POVM $\mathcal{O} = \{\mathbf{O}_l\}_{l\in L}$ such that $\mathcal{O} \rightarrow^{i} \mathcal{M}$. If the matrix $\{\lambda_{kl}\}$ in $\mathbf{M}_k = \sum_{l\in L} \lambda_{kl} \mathbf{O}_l$ does not consist entirely of 1's and 0's, we can use the reasoning of stage a to see that $\mathcal{M}$ is not extreme. If this matrix $\{\lambda_{kl}\}$ does consist entirely of 1's and 0's, the set $B(\mathcal{O})$ cannot be a simplex because $L(\mathcal{O}) = L(\mathcal{M})$ and $B(\mathcal{O})$ contains more extreme elements than $B(\mathcal{M})$. Hence $\mathcal{O}$ is not an extreme POVM (stage c) and can be written as the convex sum of two different POVM's. These lead, through this same $\{\lambda_{kl}\}$, to two different POVM's of which $\mathcal{M}$ is a convex sum. Consequently under these circumstances $\mathcal{M}$ is not extreme either. ∎

Proof of Theorem 9. First the counterexample. A PVM $\{\mathbf{E}, \mathbf{E}^{\perp}\}$ $(\mathbf{E} \neq 0, \mathbf{E} \neq 1)$ is extreme, but neither minimal nor maximal, if $\dim(\mathcal{H}) > 2$.

If $\mathcal{M}$ is minimal, self-extremal, and pairwise linearly independent, it must be equal to $\mathcal{I}$ up to labeling. Such a POVM is clearly extreme. A more meaningful statement results if we assume that $\mathcal{M}$ is maximal and also satisfies the other criteria. We then have (using Theorem 2)

$$\mathbf{M}_k = \lambda \mathbf{M}_k^{(1)} + (1-\lambda)\, \mathbf{M}_k^{(2)} \Rightarrow (\lambda = 0 \vee \mathbf{M}_k^{(1)} \sim \mathbf{M}_k)$$

for all POVM's $\{\mathbf{M}_k^{(1)}\}_{k\in K}$ and $\{\mathbf{M}_k^{(2)}\}_{k\in K}$. Therefore there is a nonnegative sequence $(v_k)_{k\in K}$ such that $\mathbf{M}_k^{(1)} = v_k \mathbf{M}_k$. But since $\mathcal{M}$ is self-extremal, pairwise linearly independent, and $B(\mathcal{M})$ is a simplex, there can be only one sequence $(v_k)_{k \in K}$ such that $\sum_{k \in K} v_k \mathbf{M}_k = \mathbf{1}$, namely $v_k = 1$. Hence it follows that $\mathbf{M}_k^{(1)} = \mathbf{M}_k$. ▮

Proof of Theorem 13.

$$\mathbf{S}_x \mathbf{X} \mathbf{S}_x^\dagger = [[\mathbf{X} - 1]] = \mathbf{X} - \mathbf{1} + n\mathbf{E}_0$$

$$\mathbf{S}_x \mathbf{X} \mathbf{S}_x^\dagger = \mathbf{S}_x \left(\sum_{k \in K} f_k \mathbf{O}_k \right) \mathbf{S}_x^\dagger = \sum_{k \in K} f_{[[k-1]]} \mathbf{O}_k$$

Combining these two equations, we obtain

$$\mathbf{E}_0 = \sum_{k \in K} \frac{1}{n} (1 - f_k + f_{[[k-1]]}) \mathbf{O}_k$$

Analogously, we have

$$\mathbf{S}_x^2 \mathbf{X} (\mathbf{S}_x^\dagger)^2 = [[\mathbf{X} - 2]] = \mathbf{X} - 2 + n(\mathbf{E}_0 + \mathbf{E}_1)$$

and, accordingly,

$$\mathbf{E}_1 = \sum_{k \in K} \frac{1}{n} (2 - f_k + f_{[[k-2]]}) \mathbf{O}_k - \mathbf{E}_0$$

$$= \sum_{k \in K} \frac{1}{n} (1 - f_{[[k-1]]} + f_{[[k-2]]}) \mathbf{O}_k$$

In this way we prove that there is a matrix $\{\mu_{kl}\}$ such that $\mathbf{E}_k = \sum_{l \in K} \mu_{kl} \mathbf{O}_l$. Since both $\mathscr{E}$ and $\mathscr{O}$ consist of n elements, $\{\mu_{kl}\}$ is an $n \times n$ matrix. The range over which the vector (x_k), $x_k = \mathrm{Tr}(\boldsymbol{\rho} \mathbf{E}_k)$, varies for variable $\boldsymbol{\rho}$ is n-dimensional. Hence $\{\mu_{kl}\}$ is invertible. There is a matrix $\{\lambda_{kl}\}$ such that $\mathbf{O}_k = \sum_{l \in K} \lambda_{kl} \mathbf{E}_l$. Since $\mathscr{E}$ is a PVM, $\{\lambda_{kl}\}$ has to have the properties of a nonideality matrix if $\mathscr{O}$ is to be a POVM. ▮

Proof of Theorem 14.

$\Rightarrow$

This follows easily from the fact that

$$\forall_{k \in K} \mathbf{S}_y \mathbf{E}_k \mathbf{S}_y^\dagger = \mathbf{E}_k$$

$\Leftarrow$

Every operator can be written as

$$\mathbf{G} = \sum_{a=0}^{n-1} \sum_{b=0}^{n-1} g_{ab} |x_a\rangle\langle x_b|$$

so that we can write

$$\mathbf{S}_y \mathbf{G} \mathbf{S}_y^\dagger = \sum_{a=0}^{n-1} \sum_{b=0}^{n-1} g_{ab} \, |x_a\rangle\langle x_b| \exp\left(i \frac{2\pi}{n}(a-b)\right)$$

Combining these two equations gives $\mathbf{S}_y \mathbf{G} \mathbf{S}_y^\dagger = \mathbf{G} \Leftrightarrow g_{ab} = \tilde{g}_a \delta_{ab}$. Hence $\mathbf{G}$ is diagonal in $\mathbf{X}$-representation. If we now substitute $\mathbf{O}_l$ for $\mathbf{G}$, the fact that $\mathscr{E}$ is a PVM implies that the coefficients involved constitute a nonideality matrix. ∎

4. OPERATIONALIZATION

The structure we have introduced in the preceding section does in no way depend on the state of the object (preparation). It relates two observables (measurement procedures). One of these is interpreted as a nonideal version of the other. This means that we are dealing with relative nonideality rather than a form of absolute nonideality. Moreover, this nonideality deals exclusively with measurement; the object plays no part in it.

More concretely, if two POVM's $\mathscr{N} = \{\mathbf{N}_l\}_{l \in L}$ and $\mathscr{M} = \{\mathbf{M}_k\}_{k \in K}$ are given:

If $\mathscr{N} \to \mathscr{M}$, we say that $\mathscr{M}$ is associated with a nonideal $\mathscr{N}$ measurement.

If $\mathscr{N} \to^i \mathscr{M}$, we say that $\mathscr{M}$ is associated with an invertibly nonideal $\mathscr{N}$ measurement.

A meter realizing $\mathscr{M}$ is called a "nonideal $\mathscr{N}$-meter." The matrix $\{\lambda_{kl}\}$ is then a property of this nonideal $\mathscr{N}$-meter, characterizing the nonideality in a mathematically precise way, allowing us to deduce from the $\mathscr{M}$-results certain information about what the $\mathscr{N}$-results would have been like. This information will, however, in general be not as good as when one would have been able to realize $\mathscr{N}$ directly. This shows, for example, when the measurement is used for state separation: the POVM $\mathscr{N}$ separates the states $\boldsymbol{\rho}$ at least as well as $\mathscr{M}$ (Theorem 15). The separation is equally good in the case of equivalence (Theorem 16).

State separation is not always important. It is often convenient to use the nonideal measurement to estimate one or several parameters (linear functionals) of the probability distribution of the observable one wanted to measure. For a given labeling, one might think of the mean value, or a higher moment, or the probability that the outcome lies in some interval.

In general this job can be summarized as follows:

Estimate the expectation value of a given operator $\mathbf{F} \in L(\mathscr{N})$, using a measurement of $\mathscr{M}$, $\mathscr{N} \to \mathscr{M}$.

A measure for the number of repetitions of the experiment needed to estimate this parameter with a given reliability is the variance: the larger the variance, the larger the number of repetitions needed; Ref. 7, Sec. I.4 (we shall assume the variance to be finite). We shall show that, if the estimate is possible at all using an $\mathscr{M}$ measurement, the number of repetitions is necessarily at least as large as the number one would have needed, had one been able to realize $\mathscr{N}$ (Theorem 17).

In the case of the relation $\rightarrow^i$ we speak of "invertibility" because it is possible to estimate $\langle \mathbf{F} \rangle$ for any $\mathbf{F} \in L(\mathscr{N})$ (Theorem 18). This means that it is possible to calculate the entire $\mathscr{N}$ distribution, if the $\mathscr{M}$ distribution is given. Then any question that can be answered using an $\mathscr{N}$ measurement can also be answered using an $\mathscr{M}$ measurement, although in the latter case it may take more repetitions of the experiment.

In the case of equivalence, no reason exists to prefer either measurement over the other (Theorem 19).

Theorem 15.

$$\mathscr{N} \rightarrow \mathscr{M} \Rightarrow \forall_{\boldsymbol{\rho}_1, \boldsymbol{\rho}_2} \sum_{l \in L} |\mathrm{Tr}((\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2) \mathbf{N}_l)| \geqslant \sum_{k \in K} |\mathrm{Tr}((\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2) \mathbf{M}_k)|$$

where $\boldsymbol{\rho}_1$, $\boldsymbol{\rho}_2$ are density operators.

Theorem 16.

$$\mathscr{N} \leftrightarrow \mathscr{M} \Rightarrow \forall_{\boldsymbol{\rho}_1, \boldsymbol{\rho}_2} \sum_{l \in L} |\mathrm{Tr}((\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2) \mathbf{N}_l)| = \sum_{k \in K} |\mathrm{Tr}((\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2) \mathbf{M}_k)|$$

where $\boldsymbol{\rho}_1$, $\boldsymbol{\rho}_2$ are density operators.

We have here used the $\mathscr{L}^1$ norm to quantify the distinguishability of two probability distributions. Similar results may hold for other distinguishability measures. For instance, for the "statistical distance" advocated by Uffink and Hilgevoord,[26,27] theorems like the above can be derived from Minkowski's inequality.

Theorem 17. Suppose that $\mathscr{N} \rightarrow \mathscr{M}$ and that $\mathbf{F} \in L(\mathscr{N})$ is given. Then one of the following alternatives is true:

(i) $\mathbf{F} \notin L(\mathscr{M})$;

(ii) For every sequence $(g_k)_{k \in K}$ such that $\sum_{k \in K} g_k \mathbf{M}_k = \mathbf{F}$ there is a sequence $(f_l)_{l \in L}$ such that $\sum_{l \in L} f_l \mathbf{N}_l = \mathbf{F}$ and

$$\sum_{k \in K} g_k^2 \mathbf{M}_k \geqslant \sum_{l \in L} f_l^2 \mathbf{N}_l \qquad (\geqslant \mathbf{F}^2;\ \text{Ref. 9, p. 88 and Ref. 12})$$

Theorem 18. Suppose that $\mathcal{N} \to^i \mathcal{M}$ and that $\mathbf{F} \in L(\mathcal{N})$ is given. Then $\mathbf{F} \in L(\mathcal{M})$. For every sequence $(g_k)_{k \in K}$ such that $\sum_{k \in K} g_k \mathbf{M}_k = \mathbf{F}$ there is a sequence $(f_l)_{l \in L}$ such that $\sum_{l \in L} f_l \mathbf{N}_l = F$ and

$$\sum_{k \in K} g_k^2 \mathbf{M}_k \geqslant \sum_{l \in L} f_l^2 N_l \qquad .$$

Theorem 19. Suppose that $\mathcal{N} \leftrightarrow \mathcal{M}$ and that $\mathbf{F} \in L(\mathcal{N})$ is given. Then $\mathbf{F} \in L(\mathcal{M})$. For every sequence $(f_l)_{l \in L}$ such that $\mathbf{F} = \sum_{l \in L} f_l \mathbf{N}_l$ one of the following alternatives is true:

(i) There is a sequence $(f'_l)_{l \in L}$ such that $\sum_{l \in L} f'_l \mathbf{N}_l = \mathbf{F}$ and

$$\sum_{l \in L} f_l'^2 N_l \begin{Bmatrix} \leqslant \\ \neq \end{Bmatrix} \sum_{l \in L} f_l^2 \mathbf{N}_l$$

(ii) There is a sequence $(g_k)_{k \in K}$ such that $\sum_{k \in K} g_k \mathbf{M}_k = \mathbf{F}$ and

$$\sum_{k \in K} g_k^2 \mathbf{M}_k = \sum_{l \in L} f_l^2 \mathbf{N}_l$$

On a more conceptual level we may say that, if $\mathcal{N} \to \mathcal{M}$, an $\mathcal{M}$ measurement result is to be interpreted as a *fuzzy* $\mathcal{N}$ measurement result. The matrix $\{\lambda_{kl}\}$ has in this connection been called a *confidence function* by Prugovečki,[20] meaning that a particular $\mathcal{M}$ result corresponds to an $\mathcal{N}$ result with a confidence proportional to λ_{kl}. Such an interpretation is close to a likelihood interpretation of the nonideality matrix.

It might be tempting to substantiate this by claiming that the *probability* that the real (in the naive sense of the word) result (of the $\mathcal{N}$ measurement, which unfortunately could not be performed) was l, given that our nonideal measurement (i.e., of $\mathcal{M}$) gave k, is proportional to λ_{kl}. The other way round, λ_{kl} could represent the probability that the nonideal measurement gives result k where an ideal measurement would have given l. But of course all such statements cannot be taken literally (let alone that an interpretation of the structure can be based on them), since there is generally no event corresponding to a "real value," so that any talk of "probability" is in this connection at best a figure of speech.

An aspect of the structure that is appealing from a physical point of view is the fact that its definition is not associated with the labeling of the outcomes in any way. This is nice because, as we noted in the first section, such a labeling, however convenient, is generally of no *fundamental* importance (that is why the introduction of PVM's via Hermitian operators is not satisfactory[14]). If we change a measurement device merely by replacing its measurement scale by another one, the device is not changed *physically*. This is conveniently represented in this structure by the fact that two POVM's differing only in labeling are members of the same equivalence class. An equivalence class consists of POVM's representing devices that measure physically identical quantities equally well.

We have restricted ourselves here to one outcome set $\mathbb{F}$. But a consequence of the above label independence is that, as is easily seen, Definitions 1 and 2 can be extended to the case where $\mathscr{M}$ and $\mathscr{N}$ have very different outcome sets. It is, for example, perfectly possible to have $\mathscr{N} \to \mathscr{M}$, where $\mathscr{M}$ has outcome set {red, yellow, blue} and $\mathscr{N}$ has outcome set $\mathbb{N}$ (or vice versa). Hence, as we required in Sec. 1, the nonideality relation structures the class of POVM's as a whole, irrespective of outcome set, whereas two POVM's with outcome sets as different as in this example cannot be connected through convexity in a natural way.

These considerations show that *i*-maximal (or even maximal) pairwise linearly independent POVM's are likely candidates to represent the "optimal" observables referred to in the first section. The observables of most textbooks ($\sim$PVM's) are optimal in this sense, just like pure observables (which were the observables that appeared optimal from the convexity point of view). There are many others, however (an interesting example is given by Helstrom; Ref. 7, p. 74). Moreover, calling observables corresponding to *i*-maximal POVM's optimal has some operational justification: they cannot be improved upon in the sense of Theorems 15 through 19. Such justification is not available for the other two classes mentioned.

We end this section by giving an example, showing that the relation $\to$ is not only of academic interest. We consider photodetection for a single-mode optical field.[13] (This is an example on an infinite-dimensional Hilbert space so that it is, strictly speaking, not a proper example. It is, however, only intended as an illustration of $\to$'s physical content.) Introducing the number states $|n\rangle$, it is well known that a detector with quantum efficiency $\eta = 1$ realizes the PVM $\mathscr{E} = \{\mathbf{E}_n\}_{n \in \mathbb{N}}$; $\mathbf{E}_n = |n\rangle\langle n|$.

In a more realistic situation, however, we have $\eta < 1$. In that case we have a POVM $\mathscr{M} = \{\mathbf{M}_k\}_{k \in \mathbb{N}}$ given by (Ref. 13, p. 240):

$$\mathbf{M}_k = \sum_{n=k}^{\infty} \mathbf{E}_n \binom{n}{k} \eta^k (1-\eta)^{n-k} = \sum_{n \in \mathbb{N}} \lambda_{kn} \mathbf{E}_n$$

with

$$\lambda_{kn} = \begin{cases} 0 & \text{if } n < k \\ \binom{n}{k} \eta^k (1-\eta)^{n-k} & \text{otherwise} \end{cases}$$

Since $\lambda_{kn} \geqslant 0$ and, as is easily seen, $\sum_k \lambda_{kn} = 1$, we have $\mathscr{E} \to \mathscr{M}$. Moreover, it is true that $\mathscr{E} \to^i \mathscr{M}$, since

$$\mathbf{E}_n = \sum_{n \in \mathbb{N}} \mu_{nk} \mathbf{M}_k \qquad \mu_{nk} = \begin{cases} 0 & \text{if } k < n \\ \binom{k}{n} \eta^{-k}(\eta - 1)^{k-n} & \text{otherwise} \end{cases}$$

(Note that not all elements of the matrix $\{\mu_{nk}\}$ are nonnegative.) In this example the nonideality is apparent: $\eta < 1$. It shows how nonideality does not imply that any information is irretrievably lost.

Proof of Theorem 15.

$$\begin{aligned}\sum_{k \in K} |\mathrm{Tr}((\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2)\,\mathbf{M}_k)| &= \sum_{k \in K} \left| \mathrm{Tr}\left((\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2) \sum_{l \in L} \lambda_{kl} \mathbf{N}_l \right) \right| \\ &\leqslant \sum_{k \in K} \sum_{l \in L} |\mathrm{Tr}((\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2)\, \lambda_{kl} \mathbf{N}_l)| \\ &= \sum_{k \in K} \sum_{l \in L} \lambda_{kl} |\mathrm{Tr}((\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2)\,\mathbf{N}_l)| \\ &= \sum_{l \in L} |\mathrm{Tr}((\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2)\,\mathbf{N}_l)| \qquad \blacksquare\end{aligned}$$

Theorem 16 is a simple consequence of Theorem 15 and Definition 3. Theorems 17 and 18 are corollaries of the following lemma:

Lemma 2. Suppose that $\mathscr{N}$ and $\mathscr{M}$ are two POVM's as above, such that $\mathscr{N} \to \mathscr{M}$ with nonideality matrix $\{\lambda_{kl}\}$. A sequence $(g_k)_{k \in K}$ is given such that $\sum_{k \in K} g_k \mathbf{M}_k$ and $\sum_{k \in K} g_k^2 \mathbf{M}_k$ converge. Define

$$f_l := \sum_{k \in K} g_k \lambda_{kl} \tag{13}$$

Then

$$\sum_{l \in L} f_l \mathbf{N}_l = \sum_{k \in K} g_k \mathbf{M}_k$$

and

$$\sum_{l \in L} f_l^2 \mathbf{N}_l \leqslant \sum_{k \in K} g_k^2 \mathbf{M}_k$$

Proof of Lemma 2. It is easily seen that the first part is true. We shall proceed with the proof of the second part:

$$\sum_{k\in K} g_k^2\mathbf{M}_k - \sum_{l\in L} f_l^2\mathbf{N}_l$$
$$= \sum_{l\in L} \mathbf{N}_l \left\{ \sum_{k\in K} g_k^2\lambda_{kl} - \left(\sum_{k\in K} g_k\lambda_{kl} \right)^2 \right\}$$
$$= \sum_{l\in L} \mathbf{N}_l \sum_{k\in K} \sum_{m\in K} g_k\{\delta_{km}\lambda_{kl} - \lambda_{kl}\lambda_{ml}\}\, g_m$$
$$= \sum_{l\in L} \mathbf{N}_l \sum_{k,m\in K} (g_k\sqrt{\lambda_{kl}})\, \{\delta_{km} - \sqrt{\lambda_{kl}}\sqrt{\lambda_{ml}}\}\, (g_m\sqrt{\lambda_{ml}})$$

If we regard $\{\delta_{km} - \sqrt{\lambda_{kl}}\sqrt{\lambda_{ml}}\}$ as a matrix-valued function of $l\in L$, it is seen to be equal to the identity matrix minus the projector onto the vector $(x_k^{(l)})_{k\in K}$; $x_k^{(l)} = \sqrt{\lambda_{kl}}$. This vector has norm 1, since

$$\sum_{k\in K} (x_k^{(l)})^2 = \sum_{k\in K} \lambda_{kl} = 1$$

Therefore the summation over k and m always results in a number $\geqslant 0$, so that

$$\sum_{k\in K} g_k^2\mathbf{M}_k - \sum_{l\in L} f_l^2\mathbf{N}_l \geqslant \mathbf{0}$$ ∎

Proof of Theorem 19. The first part of the theorem is obvious. We therefore proceed with the proof of the second part. We can construct, just as in the proof of Theorem 1, two pairwise linearly independent POVM's $\mathscr{M}'$ and $\mathscr{N}'$ such that $\mathscr{M}' \leftrightarrow \mathscr{M}$ and $\mathscr{N}' \leftrightarrow \mathscr{N}$. If a sequence $(f_l)_{l\in L}$ for $\mathscr{N}$ is such that there is none better [alternative (i) false], this sequence must have the property that $f_l = f_m$ for all $l, m\in L$ for which $\mathbf{N}_l \sim \mathbf{N}_m$. Through $\mathscr{M} \leftrightarrow \mathscr{M}' \leftrightarrow \mathscr{N}' \leftrightarrow \mathscr{N}$, we can then for every such sequence explicitly construct a sequence $(g_k)_{k\in K}$ for $\mathscr{M}$ that satisfies the demands. ∎

Note that Lemma 2 implies that alternative (i) of Th. 19 is true iff there is a sequence $(g_k)_{k\in K}$ such that

$$\sum_{k\in K} g_k\mathbf{M}_k = \mathbf{F} \wedge \sum_{k\in K} g_k^2\mathbf{M}_k \begin{Bmatrix} \leqslant \\ \neq \end{Bmatrix} \sum_{l\in L} f_l^2\mathbf{N}_l$$

5. NONIDEALITY MEASURES

For certain purposes it is convenient to express how nonideal a nonideal $\mathscr{N}$-meter is, as a real number. What we need is thus a mapping from the set of nonideality matrices to $\mathbb{R}^+$, satisfying certain consistency requirements (see Theorem 20). Matrices like $\{\lambda_{kl}\}$ should be well known to readers familiar with information theory. There they are used to represent discrete memoryless channels. If we restrict ourselves to the case of finite outcome sets, information theory also supplies us with a simple measure for the nonideality represented by a matrix $\{\lambda_{kl}\}$: Shannon's channel capacity (Ref. 24, Ch. I). (Compare Shannon's ordering of communication channels,[25] similar to $\rightarrow$.) Note that we do *not* take over the interpretation of $\{\lambda_{kl}\}$ from information theory; we are here only interested in finding suitable nonideality measures. Of course there are many other measures than the ones we derive here.

Suppose two POVM's $\mathcal{N} = \{\mathbf{N}_l\}_{l\in L}$ and $\mathcal{M} = \{\mathbf{M}_k\}_{k\in K}$ are given such that $\mathcal{N} \rightarrow \mathcal{M}$ with nonideality matrix $\{\lambda_{kl}\}; L = \{l_1,..., l_n\}$. For a given probability distribution $(p_l)_{l\in L}$, define

$$I(\{\lambda_{kl}\}; (p_l)) := \sum_{k\in K} \sum_{l\in L} q_{kl} \log(q_{kl}/(r_k p_l))$$

with

$$q_{kl} := \lambda_{kl} p_l \quad r_k := \sum_{l\in L} \lambda_{kl} p_l \qquad (k\in K; l\in L)$$

The quantity $I(\{\lambda_{kl}\}; (p_l))$ is called the mutual information. The channel capacity $C(\{\lambda_{kl}\})$ is defined by

Definition 9. $C(\{\lambda_{kl}\}) := \sup_{(p_l)_{l\in L}} \{I(\{\lambda_{kl}\}; (p_l))\}$.

The following properties of the capacity are easily verified[16]:

(i) $0 \leqslant C(\{\lambda_{kl}\}) \leqslant \log(n)$;

(ii) $C(\{\lambda_{kl}\}) = 0 \Leftrightarrow \lambda_{kl} = \bar{\lambda}_k$ $(k\in K; l\in L)$;

(iii) $C(\{\lambda_{kl}\}) = \log(n)$ iff there is for every $k\in K$ at most one $l\in L$ such that $\lambda_{kl} > 0$.

The capacity shares with $\rightarrow$ the property that it is insensitive to the labeling of the elements of the POVM. Its interpretation in this context is straightforward: the larger $C(\{\lambda_{kl}\})$ is, the smaller the nonideality $\{\lambda_{kl}\}$ represents.

Up to this point we have, by associating C with a given $\rightarrow$, implicitly assumed that the matrix $\{\lambda_{kl}\}$ is unique. This condition is in general fulfilled only if $B(\mathcal{N})$ is a simplex and $\mathcal{N}$ is both self-extremal and pairwise linearly independent. PVM's satisfy this condition, but not all POVM's do. Hence C is not guaranteed to be compatible with $\rightarrow$. This is especially clear if we take $\mathscr{M}$ minimal. In that case we would expect from a reasonable nonideality measure that it equals zero for all $\mathscr{N}$. If $\{\lambda_{kl}\}$ is not unique, C does not necessarily have this property.

Therefore we define a capacity tailored to fit a given relation $\mathscr{N} \to \mathscr{M}$. Define the set Λ of all matrices $\{\lambda_{kl}\}$ connecting the POVM's $\mathscr{N}$ and $\mathscr{M}$:

$$\Lambda_{\mathscr{N} \to \mathscr{M}} := \left\{ \{\lambda_{kl}\} \,\middle|\, \lambda_{kl} \geqslant 0 \wedge \sum_{k \in K} \lambda_{kl} = 1 \wedge \sum_{l \in L} \lambda_{kl} \mathbf{M}_l = \mathbf{N}_k \right\}$$

In view of $\mathscr{N} \to \mathscr{M}$, the set Λ can be seen as a property of the $\mathscr{M}$-meter characterizing its functioning as a nonideal $\mathscr{N}$-meter.

Define also

$$\tilde{I}(\Lambda; (p_l)) := \inf_{\{\lambda_{kl}\} \in \Lambda_{\mathscr{N} \to \mathscr{M}}} \{I(\{\lambda_{kl}\}; (p_l))\}$$

These are used in:

Definition 10. $\tilde{C}_{\mathscr{N} \to \mathscr{M}} := \sup_{(p_l)} \{\tilde{I}(\Lambda_{\mathscr{M} \to \mathscr{N}}; (p_l))\}$.

This modified capacity is indeed a consistent quantification of nonideality, as can be seen from the following theorem:

Theorem 20. Suppose $\mathscr{N} \to \mathscr{M} \to \mathscr{O} = \{\mathbf{O}_m\}_{m \in M}$. Then

$$\tilde{C}_{\mathscr{N} \to \mathscr{O}} \leqslant \tilde{C}_{\mathscr{N} \to \mathscr{M}}$$
$$\tilde{C}_{\mathscr{N} \to \mathscr{O}} \leqslant \tilde{C}_{\mathscr{M} \to \mathscr{O}}$$

This theorem is a consequence of Definition 10 and the following lemma (Ref. 16, p. 27):

Lemma 3. Suppose two nonideality matrices $\{\lambda_{kl}\}$ and $\{\mu_{lm}\}$ $(k \in K, l \in L, m \in M)$ are given. Then

$$I\left(\left\{\sum_{l \in L} \lambda_{kl} \mu_{lm}\right\}; (p_m)\right) \leqslant I\left(\{\lambda_{kl}\}; \left(\sum_{m \in M} \mu_{lm} p_m\right)\right)$$
$$I\left(\left\{\sum_{l \in L} \lambda_{kl} \mu_{lm}\right\}; (p_m)\right) \leqslant I(\{\mu_{lm}\}; (p_m))$$

for all probability distributions $(p_m)_{m \in M}$.

The channel capacity $\tilde{C}$ is not always easy to evaluate. In these cases the following measure can be convenient:

Definition 11. $\hat{I}_{\mathcal{N} \to \mathcal{M}} := \tilde{I}(\Lambda_{\mathcal{N} \to \mathcal{M}}; (\mathrm{Tr}(\mathbf{N}_l)/\dim(\mathcal{H})))$.

Because of Lemma 3, $\hat{I}$ satisfies an analog of Theorem 21. In Ref. 15 we shall use this measure to derive an inaccuracy relation.

ACKNOWLEDGMENTS

The authors would like to thank J. de Graaf for his careful reading of the manuscript, and a referee for his valuable suggestions for improvement. One of us (HM) was supported by the Foundation for Philosophical Research (SWON), which is subsidized by the Netherlands Organization for Scientific Research (NWO).

REFERENCES

1. G. Allcock, *Ann. Phys. (N.Y.)* **53**, 311 (1969).
2. K. L. Chung, *Markov Chains with Stationary Transition Probabilities*, 2nd edn. Springer, Berlin 1967).
3. E. Davies, *J. Funct. Anal.* **6**, 318 (1970).
4. E. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).
5. E. Davies and J. Lewis, *Commun. Math. Phys.* **17**, 239 (1969).
6. P. Dirac, *The Principles of Quantum Mechanics* (Clarendon, Oxford, 1930).
7. C. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976).
8. A. Holevo, *Trans. Mosc. Math. Soc.* **26**, 133 (1972).
9. A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
10. G. Jameson, *Ordered Linear Spaces* (Lecture Notes in Mathematics, Vol. 141) (Springer, New York, 1970).
11. P. Kelly and M. Weiss, *Geometry and Convexity* (Wiley, New York, 1979).
12. P. Kruszynski and W. de Muynck, *J. Math. Phys.* **28**, 1761 (1987).
13. R. Loudon, *Quantum Theory of Light*, 2nd edn. (Clarendon, Oxford, 1983).
14. G. Ludwig, *Foundations of Quantum Mechanics*, Vol. I (Springer, Berlin, 1983).
15. H. Martens and W. de Muynck, "The inaccuracy principle," *Found. Phys.* **20**, 357 (1990).
16. R. McEliece, *The Theory of Information and Coding* (Addison-Wesley, London, 1977).
17. W. de Muynck and J. Koelman, *Phys. Lett. A* **98**, 1 (1983).
18. J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer, New York, 1932, 1982).
19. J. Ortega, *Matrix theory* (Plenum, New York, 1987).
20. E. Prugovečki, *J. Phys. A* **10**, 543 (1977).
21. E. Prugovečki, *Stochastic Quantum Mechanics and Quantum Spacetime.* (Reidel, Dordrecht, 1984).
22. J. Schwinger, *Proc. Natl. Acad. Sci. USA* **46**, 570 (1960).
23. F. Schroeck, *Int. J. Theor. Phys.* **28**, 247 (1989).
24. C. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948).
25. C. Shannon, *Inform. Control* **1**, 390 (1958).
26. J. Uffink and J. Hilgevoord, *Physica B* **151**, 309 (1988).
27. W. Wootters, *Phys. Rev. D* **19**, 473 (1979).