Operational axioms for diagonalizing states

Giulio Chiribella

`giulio@cs.hku.hk`

Department of Computer Science, University of Hong Kong, Hong Kong

Carlo Maria Scandolo

`carlomaria.scandolo@st-annes.ox.ac.uk`

Department of Computer Science, University of Oxford, Oxford, UK

In quantum theory every state can be diagonalized, i.e. decomposed as a convex combination of perfectly distinguishable pure states. This elementary structure plays an ubiquitous role in quantum mechanics, quantum information theory, and quantum statistical mechanics, where it provides the foundation for the notions of majorization and entropy. A natural question then arises: can we reconstruct these notions from purely operational axioms? We address this question in the framework of general probabilistic theories, presenting a set of axioms that guarantee that every state can be diagonalized. The first axiom is Causality, which ensures that the marginal of a bipartite state is well defined. Then, Purity Preservation states that the set of pure transformations is closed under composition. The third axiom is Purification, which allows to assign a pure state to the composition of a system with its environment. Finally, we introduce the axiom of Pure Sharpness, stating that for every system there exists at least one pure effect occurring with unit probability on some state. For theories satisfying our four axioms, we show a constructive algorithm for diagonalizing every given state. The diagonalization result allows us to formulate a majorization criterion that captures the convertibility of states in the operational resource theory of purity, where random reversible transformations are regarded as free operations.

1 Introduction

A canonical route to the foundations of quantum thermodynamics is provided by the theory of majorization, used to define an ordering among states according to their degree of mixedness [59, 60, 61, 58]. In recent years, the applications of majorization have seen remarkable developments in the study of quantum and nano thermodynamics [29, 39, 33, 13]. The viability of this approach relies heavily on the Hilbert space framework, for it is based on the fact that density operators can be diagonalized. Ideally, however, it would be desirable to have an axiomatic foundation of quantum thermodynamics based on purely operational axioms.

The problem can be addressed in the framework of general probabilistic theories [35, 26, 9, 4, 15, 16, 8, 37, 36, 14]. The first step in this direction is to consider probabilistic theories that satisfy an operational version of the spectral theorem, according to which every state can be "diagonalized", i.e. decomposed as a mixture of perfectly distinguishable pure states. At this point there are two options: one option is to demand the diagonalizability of states as an axiom. This approach has been adopted in Refs. [7, 41, 3], also in relation to the issue of defining majorization in general probabilistic theories. The other option is to reduce diagonalization to other operational axioms, which may provide deeper insights on the conceptual foundations of quantum thermodynamics. This approach will be the subject of the present paper.

A diagonalization result from operational principles was proved by D'Ariano, Perinotti, and one of the authors in the context of the axiomatization of quantum theory in Ref. [16] (hereafter referred to as CDP), although the proof therein used the full set of axioms implying quantum theory. In this paper we derive the diagonalizability of states from a strictly weaker set of axioms, which is compatible with quantum theory on real Hilbert spaces and, with other potential generalizations of quantum theory, such as the fermionic theory recently proposed by D'Ariano *et al* in Refs. [27, 28]. Our list of axioms consists of:

- two of the six CDP axioms (Causality and Purification);

- one axiom (Purity Preservation) that is close to the CDP axiom Atomicity of Composition, although not exactly equivalent to it;

- a new axiom, which we name *Pure Sharpness*.

Pure Sharpness stipulates that every physical system has at least one pure effect occurring with unit probability on some state. Such a pure effect can be seen as part of a yes-no test designed to check an elementary property, in the sense of Piron [52]. In these terms, Pure Sharpness requires that for every system there exist at least one property, and at least one state possessing such a property. Note that none of our axioms assumes that perfectly distinguishable states exist. A priori, the general probabilistic theories considered here may not contain *any* pair of perfectly distinguishable states—operationally, this would mean that no system described by the theory could be used to transmit a classical bit with zero error. The existence of perfectly distinguishable states, and the fact that every state can be broken down into a mixture of perfectly distinguishable pure states are non-trivial consequences of the axioms.

Note that the presence of Purification among the axioms excludes from the start the case of classical probability theory. Indeed, the aim of our work is *not* to provide the most general conditions for the diagonalization of states, but rather to derive diagonalization as a first step towards an axiomatic foundation of quantum thermodynamics. In particular, we are searching for axioms that capture the characteristic traits of quantum thermodynamics, such as the link with the resource theory of entanglement [21]. From this point of view, Purification is an almost mandatory choice, in that it sets up a fundamental relation between mixed states and pure entangled states. More importantly, Purification is deeply related to the thermodynamic procedure that consists in considering the system in interaction with its environment in such a way that the composite system is isolated. In this scenario, Purification guarantees that one can always associate a pure state with the composite system and that the overall evolution of system and environment can be treated as reversible. In this way, thermodynamics is reconciled with the paradigm of reversible dynamics at the fundamental level. In the concrete Hilbert space setting, the purified view of quantum thermodynamics has been adopted in a number of works aimed at deriving the microcanonical and canonical ensembles [11, 42, 43, 31, 32, 53, 30, 46, 12], an idea that has been recently explored also in general probabilistic theories [45, 48].

After deriving the diagonalizability of states, we discuss the implications of the result. In particular, we discuss the relation of majorization, defined in terms of the probability distributions arising from diagonalization. Combining our axioms with an additional axiom, known as Strong Symmetry [7], we then show that majorization completely determines the convertibility of states in the operational resource theory of purity [21], where random reversible transformations are viewed as free operations. It remains as an open question whether in the context of our axioms Strong Symmetry can be replaced with a weaker requirement [19].

The paper is structured as follows: in section 2 we introduce the basic framework. The four axioms for diagonalization are presented in section 3, and their consequences are examined in section 4. Section 5 contains the main result, namely the diagonalization theorem. In section 6 we discuss a number

of results that arise from the combination of diagonalization with the Strong Symmetry axiom. Using these results, section 7 analyses majorization and its applications to the resource theory of purity. The conclusions are drawn in section 8.

2 Framework

The present analysis is carried out in the framework of general probabilistic theories, adopting the specific variant of Refs. [15, 16, 14], known as the framework of *operational-probabilistic theories* (*OPTs*). OPTs arise from the marriage of the graphical language of symmetric monoidal categories [1, 2, 22, 24, 55] with the toolbox of probability theory. Here we give a quick summary of the framework, referring the reader to the original papers and to the related work by Hardy [37, 38] for a more in-depth presentation. A comprehensive review of the OPT framework is presented in the book chapter [18].

Physical processes can be combined in sequence or in parallel, giving rise to circuits like the following
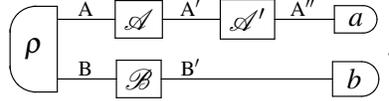
Here, A, A$'$, A$''$, B, B$'$ are *systems*, ρ is a bipartite *state*, $\mathscr{A}$, $\mathscr{A}'$ and $\mathscr{B}$ are *transformations*, a and b are *effects*. Circuits with no external wires, like the one in the above example, are associated with probabilities. We denote by

- $\mathsf{St}\,(\mathrm{A})$ the set of states of system A

- $\mathsf{Eff}\,(\mathrm{A})$ the set of effects on A

- $\mathsf{Transf}\,(\mathrm{A},\mathrm{B})$ the set of transformations from A to B

- $\mathrm{A} \otimes \mathrm{B}$ the composition of systems A and B.

- $\mathscr{A} \otimes \mathscr{B}$ the parallel composition of the transformations $\mathscr{A}$ and $\mathscr{B}$.

A particular system is the trivial system I (mathematically, the unit of the tensor product), corresponding to the degrees of freedom ignored by the theory. States (resp. effects) are transformations with the trivial system as input (resp. output). We will often make use of the short-hand notation $(a|\rho)$ to denote the scalar

$$(a|\rho) \; := \; \boxed{\rho}\!\!-\!\!\overset{\mathrm{A}}{}\!\!-\!\boxed{a}\,,$$

and of the notation $(a|\mathscr{C}|\rho)$ to mean

$$(a|\mathscr{C}|\rho) \; := \; \boxed{\rho}\!\!-\!\!\overset{\mathrm{A}}{}\!\!-\!\boxed{\mathscr{C}}\!\!-\!\!\overset{\mathrm{B}}{}\!\!-\!\boxed{a}\,.$$

We identify the scalar $(a|\rho)$ with a real number in the interval $[0,1]$, representing the probability of a joint occurrence of the state ρ and the effect a in a circuit where suitable non-deterministic elements are put in place. The fact that scalars are real numbers induces a notion of sum for transformations, whereby the sets $\mathsf{St}\,(\mathrm{A})$, $\mathsf{Transf}\,(\mathrm{A},\mathrm{B})$, and $\mathsf{Eff}\,(\mathrm{A})$ become spanning sets of suitable vector spaces over the real numbers, denoted by $\mathsf{St}_{\mathbb{R}}\,(\mathrm{A})$, $\mathsf{Transf}_{\mathbb{R}}\,(\mathrm{A},\mathrm{B})$, and $\mathsf{Eff}_{\mathbb{R}}\,(\mathrm{A})$ respectively. In this paper we will restrict our attention to finite systems, i.e. systems A for which the vector spaces $\mathsf{St}_{\mathbb{R}}\,(\mathrm{A})$ and $\mathsf{Eff}_{\mathbb{R}}\,(\mathrm{A})$ are finite-dimensional. Also, it will be assumed as a default that the sets $\mathsf{St}\,(\mathrm{A})$, $\mathsf{Transf}\,(\mathrm{A},\mathrm{B})$, and $\mathsf{Eff}\,(\mathrm{A})$ are compact in the topology induced by probabilities, by which one has $\lim_{n \to +\infty} \mathscr{C}_n = \mathscr{C}$, where $\mathscr{C}_n, \mathscr{C} \in \mathsf{Transf}\,(\mathrm{A},\mathrm{B})$, if and only if

$$\lim_{n \to +\infty} (E|\mathscr{C}_n \otimes \mathscr{I}_{\mathrm{R}}|\rho) = (E|\mathscr{C} \otimes \mathscr{I}_{\mathrm{R}}|\rho) \quad \forall \mathrm{R}, \forall \rho \in \mathsf{St}\,(\mathrm{A} \otimes \mathrm{R}), \forall E \in \mathsf{Eff}\,(\mathrm{B} \otimes \mathrm{R})\,.$$

A *test* from A to B is a collection of transformations $\{\mathscr{C}_i\}_{i\in\mathsf{X}}$ from A to B, which can occur in an experiment with outcomes in X. If A (resp. B) is the trivial system, the test is called a *preparation-test* (resp. *observation-test*). We stress that not all the collections of transformations are tests: the specification of the collections that are to be regarded as tests is part of the theory, the only requirement being that the set of test is closed under parallel and sequential composition.

If X contains a single outcome, we say that the test is *deterministic*. We will refer to deterministic transformations as *channels*. Following the most recent version of the formalism [14], we assume as part of the framework that every test arises from an observation-test performed on one of the outputs of a channel. The motivation for such an assumption is the idea that the readout of the outcome could be interpreted physically as a measurement allowed by the theory. Precisely, the assumption is the following.

Assumption 1 (Physicalization of readout [14])**.** *For every pair of systems* A, B, *and every test* $\{\mathscr{M}_i\}_{i\in\mathsf{X}}$ *from* A *to* B, *there exist a system* C, *a channel* $\mathscr{M}\in\mathsf{Transf}(A, B\otimes C)$, *and an observation-test* $\{c_i\}_{i\in\mathsf{X}}\subset \mathsf{Eff}(C)$ *such that*

$$\begin{array}{c}\underline{\quad A\quad}\boxed{\mathscr{M}_i}\underline{\quad B\quad}\end{array} = \begin{array}{c}\xrightarrow{A}\boxed{\quad}\xrightarrow{B}\\ \boxed{\mathscr{M}}\\ \xrightarrow{C}\!\!\overline{c_i}\end{array}\qquad \forall i\in\mathsf{X}.$$

A channel $\mathscr{U}$ from A to B is called *reversible* if there exists a channel $\mathscr{U}^{-1}$ from B to A such that $\mathscr{U}^{-1}\mathscr{U}=\mathscr{I}_A$ and $\mathscr{U}\mathscr{U}^{-1}=\mathscr{I}_B$, where $\mathscr{I}_S$ is the identity channel on a generic system S. If there exists a reversible channel transforming A into B, we say that A and B are *operationally equivalent*, denoted by $A\simeq B$. The composition of systems is required to be *symmetric*, meaning that $A\otimes B\simeq B\otimes A$.

A state $\chi\in\mathsf{St}(A)$ is called *invariant* if $\mathscr{U}\chi=\chi$, for every reversible channel $\mathscr{U}$. Note that, in general, invariant states may not exist. In this paper their existence will be a consequence of the axioms and of a standing assumption of finite-dimensionality.

The pairing between states and effects leads naturally to a notion of norm. We define the norm of a state ρ as $\|\rho\|:=\sup_{a\in\mathsf{Eff}(A)}(a|\rho)$. The set of normalized (i.e. with unit norm) states of A will be denoted by $\mathsf{St}_1(A)$. Similarly, the norm of an effect a is defined as $\|a\|:=\sup_{\rho\in\mathsf{St}(A)}(a|\rho)$. The set of normalized effects of system A will be denoted by $\mathsf{Eff}_1(A)$.

The probabilistic structure also offers an easy way to define pure transformations. The definition is based on the notion of *coarse-graining*, i.e. the operation of joining two or more outcomes of a test into a single outcome. More precisely, a test $\{\mathscr{C}_i\}_{i\in\mathsf{X}}$ is a *coarse-graining* of the test $\{\mathscr{D}_j\}_{j\in\mathsf{Y}}$ if there is a partition $\{\mathsf{Y}_i\}_{i\in\mathsf{X}}$ of Y such that $\mathscr{C}_i=\sum_{j\in\mathsf{Y}_i}\mathscr{D}_j$ for every $i\in\mathsf{X}$. In this case, we say that $\{\mathscr{D}_j\}_{j\in\mathsf{Y}}$ is a *refinement* of $\{\mathscr{C}_i\}_{i\in\mathsf{X}}$. The refinement of a given transformation is defined via the refinement of a test: if $\{\mathscr{D}_j\}_{j\in\mathsf{Y}}$ is a refinement of $\{\mathscr{C}_i\}_{i\in\mathsf{X}}$, then the transformations $\{\mathscr{D}_j\}_{j\in\mathsf{Y}_i}$ are a refinement of the transformation $\mathscr{C}_i$.

A transformation $\mathscr{C}\in\mathsf{Transf}(A,B)$ is called *pure* if it has only trivial refinements, namely for every refinement $\{\mathscr{D}_j\}$ one has $\mathscr{D}_j=p_j\mathscr{C}$, where $\{p_j\}$ is a probability distribution. Pure transformations are those for which the experimenter has maximal information about the evolution of the system. We denote the set of pure transformations from A to B as $\mathsf{PurTransf}(A,B)$. In the special case of states (resp. effects) of system A we use the notation $\mathsf{PurSt}(A)$ (resp. $\mathsf{PurEff}(A)$). The set of normalized pure states (resp. effects) of A will be denoted by $\mathsf{PurSt}_1(A)$ (resp. $\mathsf{PurEff}_1(A)$). As usual, non-pure states are called *mixed*.

Definition 1. Let ρ be a normalized state. We say that a state σ is *contained* in ρ if we can write $\rho=p\sigma+(1-p)\tau$, where $p\in(0,1]$ and τ is another state.

It is clear that no states are contained in a pure state, except the pure state itself. At the opposite side there are *completely mixed* states [16], such that every state is contained in them.

Definition 2. We say that two transformations $\mathscr{A}, \mathscr{A}' \in \mathsf{Transf}(A,B)$ are *equal upon input* of the state $\rho \in \mathsf{St}_1(A)$ if $\mathscr{A}\sigma = \mathscr{A}'\sigma$ for every state σ contained in ρ. In this case we will write $\mathscr{A} =_\rho \mathscr{A}'$.

3   Axioms

Here we present our four axioms for diagonalizing states. As a first axiom, we assume Causality, which forbids signalling from the future to the past:

Axiom 1 (Causality [15, 16]). *The outcome probabilities of a test do not depend on the choice of other tests performed later in the circuit.*

Causality is equivalent to the requirement that, for every system A, there exists a unique deterministic effect u_A on A (or simply u, when no ambiguity can arise). Thanks to that, it is possible to define the *marginal state* of a bipartite state ρ_{AB} on system A as

$$
\boxed{\rho_A}\!-\!\!\!\!\!\!^A\ = \ \left(\rho_{AB}\ \genfrac{}{}{0pt}{}{\ -\!\!^A}{\ -\!\!_B\!-\!\boxed{u}}\right).
$$

In this case we will also write $\rho_A := \mathrm{Tr}_B \rho_{AB}$, calling u_B as Tr_B, to remind that the deterministic effect acts as the partial trace in quantum theory. We will tend to keep the notation Tr in formulas where the deterministic effect is directly applied to a state, e.g. $\mathrm{Tr}\,\rho := (u|\rho)$.

In a causal theory (i.e. satisfying Causality), the norm of a state ρ is simply given by $\|\rho\| = \mathrm{Tr}\,\rho$. Moreover, observation-tests are normalized in the following way (see corollary 3 of Ref. [15]):

Proposition 1. *In a causal theory, if $\{a\}_{i \in \mathsf{X}}$ is an observation-test, then $\sum_{i \in \mathsf{X}} a_i = u$.*

Causality guarantees that it is consistent to assume that the choice of a test can depend on the outcomes of previous tests—namely that it is possible to perform *conditional tests* [15]. Combined with the assumption of compactness, the ability to perform conditional tests implies that every state is proportional to a normalized state [18]. Another consequence is that all the sets $\mathsf{St}(A)$, $\mathsf{Transf}(A,B)$, and $\mathsf{Eff}(A)$ are *convex*. In the following we will take for granted the ability to perform conditional tests, the fact that every state is proportional to a normalized state, and the convexity of all the sets of transformations.

The second axiom in our list is Purity Preservation.

Axiom 2 (Purity Preservation[1] [26, 16, 20, 21]). *Sequential and parallel compositions of pure transformations are pure transformations.*

We consider Purity Preservation as a fundamental requirement. Considering the theory as an algorithm to make deductions about physical processes, Purity Preservation ensures that, when presented with maximal information about two processes, the algorithm outputs maximal information about their composition [20].

The third axiom is Purification. This axiom characterizes the physical theories admitting a description where all deterministic processes are pure and reversible at a fundamental level. Essentially, Purification expresses a strengthened version of the principle of conservation of information [17, 20]. In its simplest form, Purification is phrased as a requirement about *causal* theories, where the marginal of a bipartite state is defined in a canonical way. Specifically, we say that a state $\rho \in \mathsf{St}_1(A)$ can be purified if there exists a pure state $\Psi \in \mathsf{PurSt}(A \otimes B)$ that has ρ as its marginal on system A. In this case, we call Ψ a *purification* of ρ, and B a *purifying system*. The axiom is as follows.

---

[1] The name and the formulation of the axiom adopted here are the same as in Ref. [20]. The original axiom was called *Atomicity of Composition* [26] and involved *only* sequential composition. Extending the axiom to parallel composition is important for our purposes, because it guarantees that the product of two pure states is pure. In the axiomatization of Ref. [16] this property was a consequence of the Local Tomography axiom, which, instead, is not assumed here.

Axiom 3 (Purification [15, 16]). *Every state can be purified and two purifications with the same purifying system differ by a reversible channel on the purifying system.*

Technically, the second part of the axiom states that, if $\Psi, \Psi' \in \mathsf{PurSt}_1(A \otimes B)$ are such that $\mathrm{Tr}_B \Psi_{AB} = \mathrm{Tr}_B \Psi'_{AB}$, then $\Psi'_{AB} = (\mathscr{I}_A \otimes \mathscr{U}_B) \Psi_{AB}$, where $\mathscr{U}_B$ is a reversible channel on B. In diagrams,

$$\begin{array}{c} \Psi' \begin{array}{l} \text{A} \\ \text{B} \end{array} \end{array} = \begin{array}{c} \Psi \begin{array}{l} \text{A} \\ \text{B} \boxed{\mathscr{U}} \text{B} \end{array} \end{array},$$

In quantum theory, the validity of Purification lies at the foundation of all dilation theorems, such as Stinespring's [57], Naimark's [51], and Ozawa's [50]. In the finite-dimensional setting, these theorems (or at least some aspects thereof) were reconstructed axiomatically in [15].

Finally, we introduce a new axiom, which we name *Pure Sharpness*. This axiom ensures that there exists at least one elementary property associated with every system:

Axiom 4 (Pure Sharpness). *For every system* A*, there exists at least one pure effect $a \in \mathsf{PurEff}(A)$ occurring with probability 1 on some state.*

Pure Sharpness is reminiscent of the Sharpness axiom used in Hardy's 2011 axiomatization [37], which requires a one-to-one correspondence between pure states and effects that distinguish maximal sets of states.

4   Consequences of the axioms

4.1   Consequences of Causality, Purity Preservation, and Purification

Here we list a few consequences of the first three axioms, which will become useful later.

The easiest consequence of Purification is that reversible channels act transitively on the set of pure states (see lemma 20 in Ref. [15]):

Proposition 2. *For any pair of pure states $\psi, \psi' \in \mathsf{PurSt}_1(A)$ there exists a reversible channel $\mathscr{U}$ on* A *such that $\psi' = \mathscr{U}\psi$.*

As a consequence, every finite-dimensional system possesses one invariant state (see corollary 34 of Ref. [15]):

Proposition 3. *For every system* A*, there exists a unique invariant state χ_A, which is also a completely mixed state.*

Also, transitivity implies that the set of pure states is compact for every system (see corollary 32 of Ref. [15]). This property is generally a non-trivial property—cf. Ref. [5] for a counterexample of a state space with a non-closed set of pure states.

A crucial consequence of Purification is the *steering property*:

Theorem 1 (Steering property). *Let $\rho \in \mathsf{St}_1(A)$ and let $\Psi \in \mathsf{PurSt}_1(A \otimes B)$ be a purification of ρ. Then σ is contained in ρ if and only if there exist an effect b_σ on the purifying system* B *and a non-zero probability p such that*

$$p \begin{array}{c} \boxed{\sigma} \text{A} \end{array} = \begin{array}{c} \Psi \begin{array}{l} \text{A} \\ \text{B} \boxed{b_\sigma} \end{array} \end{array}.$$

Proof. The proof follows the same lines of theorem 6 and corollary 9 in Ref. [15], with the only difference that here we do not assume the existence of perfectly distinguishable states. In its place, we use the framework assumption 1, which guarantees that the outcome of every test can be read out from a physical system. $\square$

Now we introduce a definition and a proposition which will be used later.

Definition 3. We say that a state $\rho \in \mathsf{St}_1(A \otimes B)$ is *faithful for effects* of system A if, for any $a, a' \in \mathsf{Eff}(A)$, we have $a = a'$ if

$$
\begin{array}{c}
\rho \!\!-\!\!\!\begin{array}{c} A \!-\! a \\ B \end{array}
\end{array}
=
\begin{array}{c}
\rho \!\!-\!\!\!\begin{array}{c} A \!-\! a' \\ B \end{array}
\end{array}
$$

Proposition 4. *A pure state* Ψ_{AB} *is faithful for effects of system A if and only if its marginal* ω_A *on A is completely mixed.*

See theorems 8 and 9 of Ref. [15] for the proof.

Combining Purification with Purity Preservation one obtains the following properties:

Proposition 5. *For every observation-test* $\{a_i\}_{i \in X}$ *on A, there is a system B and a test* $\{\mathscr{A}_i\}_{i \in X} \subset \mathsf{Transf}(A, B)$ *such that every* $\mathscr{A}_i$ *is pure and* $a_i = u_B \mathscr{A}_i$.

Proposition 6. *Let a be an effect such that* $(a|\rho) = 1$*, for some* $\rho \in \mathsf{St}_1(A)$*. Then there exists a transformation* $\mathscr{T}$ *on A such that* $a = u\mathscr{T}$ *and* $\mathscr{T} =_\rho \mathscr{I}$*, where* $\mathscr{I}$ *is the identity.*

The proofs of the above propositions can be found in lemma 18 and corollary 9 of Ref. [16].

Finally, thanks to Purification, proposition 1 becomes also a sufficient conditions for a set of effects to be an observation-test (cf. theorem 18 of Ref. [15]).

Proposition 7. *A set of effects* $\{a_i\}_{i=1}^n$ *is an observation-test if and only if* $\sum_{i=1}^n a_i = u$.

4.2 Consequences of all the axioms

In quantum theory, diagonalizing a state means decomposing it as a convex combination of orthogonal pure states, i.e. pure states that can be perfectly distinguished by a measurement.

In a general theory, perfectly distinguishable states are defined as follows:

Definition 4. The normalized states $\{\rho_i\}$ are *perfectly distinguishable* if there exists an observation-test $\{a_j\}$ such that $(a_j|\rho_i) = \delta_{ij}$. $\{a_j\}$ is called *perfectly distinguishing test*.

Suppose we know that $(a|\rho) = 1$, where a is a pure effect. Then, we can conclude that the state ρ must be pure:

Proposition 8. *Let* $a \in \mathsf{PurEff}_1(A)$*. Then, there exists a* pure *state* $\alpha \in \mathsf{PurSt}(A)$ *such that* $(a|\alpha) = 1$*. Furthermore, for every* $\rho \in \mathsf{St}(A)$*, if* $(a|\rho) = 1$*, then* $\rho = \alpha$*.*

See lemma 26 and theorem 7 of Ref. [16] for the proof idea.

Combining the above result with our Pure Sharpness axiom, we derive the following

Proposition 9. *For every pure state* $\alpha \in \mathsf{PurSt}(A)$*, there exists at least one pure effect* $a \in \mathsf{PurEff}(A)$ *such that* $(a|\alpha) = 1$*.*

Proof. By Pure Sharpness, there exists at least one pure effect a_0 such that $(a_0|\alpha_0) = 1$ for some state α_0. By proposition 8, α_0 is pure. Now, for a generic pure state α, by transitivity, there is a reversible channel $\mathscr{U}$ such that $\alpha = \mathscr{U}\alpha_0$. Hence, the effect $a := a_0 \mathscr{U}^{-1}$ is pure and $(a|\alpha) = 1$. $\square$

The above result will turn out to be useful for the construction of our diagonalization procedure. A crucial ingredient in the derivation of the diagonalization theorem is the following

Theorem 2. *Let ρ be a normalized state of system* A *and let p_* be the probability defined as*[2]

$$p_* = \max_{\alpha \in \mathsf{PurSt}_1(A)} \{p \in [0,1] : \rho = p\alpha + (1-p)\sigma, \sigma \in \mathsf{St}_1(A)\}.$$

Let $\Psi \in \mathsf{PurSt}_1(A \otimes B)$ be a purification of ρ and let $\widetilde{\rho} \in \mathsf{St}_1(B)$ be the complementary state of ρ, namely $\widetilde{\rho} := \mathrm{Tr}_A \Psi$. Then, there exists a pure state $\beta \in \mathsf{PurSt}_1(B)$ such that $\widetilde{\rho} = p_*\beta + (1 - p_*)\tau$ for some state $\tau \in \mathsf{St}_1(B)$.

Proof. By hypothesis, one can write $\rho = p_*\alpha + (1 - p_*)\sigma$, where α is a pure state and σ is possibly mixed. Let us purify ρ, and let Ψ be one of its purifications, with purifying system B. According to the steering property, there exists an effect b that prepares α with probability p_*, namely

$$\Psi \begin{array}{c} A \\ B \end{array} b = p_* \quad \alpha \underline{} A \ . \tag{1}$$

Let a be a pure effect such that $(a|\alpha) = 1$. Applying a on both sides of Eq. (1), we get

$$\Psi \begin{array}{c} A \end{matrix} a \\ B \end{array} b = p_* \ .$$

On the other hand, applying a to the state Ψ we obtain

$$\Psi \begin{array}{c} A \end{matrix} a \\ B \end{array} = q \quad \beta \underline{} B \ , \tag{2}$$

where $q \in [0,1]$ and β is a pure state (due to Purity Preservation). Now if we apply b, we have

$$\Psi \begin{array}{c} A \end{matrix} a \\ B \end{array} b = p_* = q \quad \beta \underline{} B b \ .$$

Since $(b|\beta) \in [0,1]$, we must have $q \geq p_*$. We now prove that, in fact, equality holds. Let $\widetilde{b}$ be a pure effect such that $\left(\widetilde{b}|\beta\right) = 1$. Applying $\widetilde{b}$ on both sides of Eq. (2), we obtain

$$\Psi \begin{array}{c} A \end{matrix} a \\ B \end{array} \widetilde{b} = q.$$

By Purity Preservation, $\widetilde{b}$ will induce a pure state on system A, namely

$$q = \Psi \begin{array}{c} A \end{matrix} a \\ B \end{array} \widetilde{b} = \widetilde{p} \quad \widetilde{\alpha} \underline{} A a \ ,$$

where $\widetilde{p} \in [0,1]$. From the above equation, we have the inequality $q \leq \widetilde{p}$. Since by definition we have $\widetilde{p} \leq p_*$, we finally get the chain of inequalities $p_* \leq q \leq \widetilde{p} \leq p_*$, whence $p_* = q = \widetilde{p}$. Hence, Eq. (2) implies that the pure state β arises with probability p_* in a convex decomposition of the state $\widetilde{\rho}$. $\square$

---

[2]Note that the maximum is well defined because the set of pure states is compact, thanks to transitivity.

A similar proof was used in lemma 30 of Ref. [16] in the special case where ρ is the invariant state, and with stronger assumptions, i.e. Ideal Compression, which is not assumed here.

The effect b that prepares α with probability p_* can always be taken to be pure. Indeed, $\widetilde{b}$ is a *pure* effect that prepares the pure state $\widetilde{\alpha}$ on A with probability $\widetilde{p}$. But since $\widetilde{p} = p_*$, then $(a|\widetilde{\alpha}) = 1$. Therefore, by proposition 8, $\widetilde{\alpha} = \alpha$. This shows that α can always be prepared with probability p_* by using a pure effect on B.

As a corollary we have the following:

Corollary 1. *Let $\rho \in \mathsf{St}_1(A)$ be a state and let $\widetilde{\rho} \in \mathsf{St}_1(B)$ be a complementary state of ρ. Let $p_*(\rho)$ and $p_*(\widetilde{\rho})$ be defined like in theorem 2, for ρ and $\widetilde{\rho}$ respectively. Then $p_*(\rho) = p_*(\widetilde{\rho})$.*

Proof. By theorem 2, we know that there exists a pure state $\beta \in \mathsf{PurSt}_1(B)$ arising in a convex decomposition of $\widetilde{\rho}$ with probability $p_*(\rho)$:

$$\widetilde{\rho} = p_*(\rho)\beta + (1 - p_*(\rho))\tau,$$

where τ is another state of system B. Therefore $p_*(\widetilde{\rho}) \geq p_*(\rho)$. By theorem 2 applied to $\widetilde{\rho}$, we know that there is a pure state $\alpha' \in \mathsf{PurSt}_1(A)$ arising in a convex decomposition of ρ with probability $p_*(\widetilde{\rho})$:

$$\rho = p_*(\widetilde{\rho})\alpha' + (1 - p_*(\widetilde{\rho}))\sigma',$$

where $\sigma' \in \mathsf{St}(A)$. By definition of $p_*(\rho)$, we have $p_*(\rho) \geq p_*(\widetilde{\rho})$, whence we conclude that $p_*(\rho) = p_*(\widetilde{\rho})$. $\qquad\square$

Now we are ready to prove the uniqueness of the pure effect associated with a pure state. The proof uses the following lemma (see lemma 29 of Ref. [16]).

Lemma 1. *Let χ be the invariant state of system A and let α be a normalized pure state. Then*

$$p_{\max} := p_\alpha = \max\{p : \exists \sigma, \chi = p\alpha + (1 - p)\sigma\}$$

does not depend on α.

Proposition 10. *For every normalized pure state α there exists a unique pure effect a such that $(a|\alpha) = 1$.*

The proof is identical to the that of theorem 8 of Ref. [16], even though we are assuming fewer axioms.

We will denote by $\alpha^\dagger$ the unique pure effect associated with the pure state α, namely such that $(\alpha^\dagger|\alpha) = 1$.

We are able to establish a bijective correspondence between normalized pure states and normalized pure effects. As a result, we obtain the following corollary (cf. corollary 13 of Ref. [16]).

Corollary 2. *For every pair of $a, a' \in \mathsf{PurEff}_1(A)$, there exists a reversible channel $\mathscr{U}$ on A such that $a' = a\mathscr{U}$.*

5 Diagonalization of states

A *diagonalization* of ρ is a convex decomposition of ρ into perfectly distinguishable pure states. The probabilities in such a convex decomposition will be called the *eigenvalues* of ρ. Note that, since we are assuming the vector space $\mathsf{St}_\mathbb{R}(A)$ to be finite-dimensional, diagonalizations of states will have a *finite* number of terms. Here we are not postulating the existence of perfectly distinguishable pure states, but this will be a result of the present set of axioms (see corollary 3).

The starting point for diagonalization is the following

Proposition 11. *Consider $\rho = p_* \alpha + (1 - p_*) \sigma$, where p_* is defined in theorem 2. We have $\left(\alpha^\dagger | \rho\right) = p_*$.*

Proof. Let Ψ_{AB} be a purification of ρ. Then, the proof of theorem 2 yields the following equality

$$
\left(\!\! \Psi \!\! \begin{array}{c} A \!-\! \boxed{\alpha^\dagger} \\ B \end{array} \right) = p_* \left(\beta \right) \!-\! B
$$

By applying the deterministic effect on both sides of the above equation, we obtain

$$
p_* = p_* \left(\beta \right) \!-\! B \!-\! \boxed{u} = \left(\!\! \Psi \!\! \begin{array}{c} A \!-\! \boxed{\alpha^\dagger} \\ B \!-\! \boxed{u} \end{array} \right) = \left(\rho \right) \!-\! A \!-\! \boxed{\alpha^\dagger} .
$$

This shows that $\left(\alpha^\dagger | \rho\right) = p_*$. $\qquad\square$

The following proposition enables us to define p_* in an alternative, and perhaps simpler, way starting from measurements.

Proposition 12. *Let $\rho \in \mathsf{St}_1(A)$. Define $p^* := \max_{a \in \mathsf{PurEff}_1(A)} (a | \rho)$. Then $p^* = p_*$.*

Proof. By proposition 11, clearly one has $p^* \geq p_*$. Since p^* is the maximum, it is achieved by some $a^* \in \mathsf{PurEff}_1(A)$. Therefore,

$$
p^* = \left(\rho \right) \!-\! A \!-\! \boxed{a^*} = \left(\!\! \Psi \!\! \begin{array}{c} A \!-\! \boxed{a^*} \\ B \!-\! \boxed{u} \end{array} \right) ,
$$

where Ψ_{AB} is a purification of ρ. Now, a^* prepares a pure state β^* on B with probability $q \leq p_*$ (cf. corollary 1).

$$
p^* = \left(\!\! \Psi \!\! \begin{array}{c} A \!-\! \boxed{a^*} \\ B \!-\! \boxed{u} \end{array} \right) = q \left(\beta^* \right) \!-\! B \!-\! \boxed{u} = q
$$

We then obtain $p^* = q \leq p_*$, whence, in fact, $p^* = p_*$. $\qquad\square$

The result expressed in proposition 11 has important consequences about diagonalization. Since $\left(\alpha^\dagger | \rho\right) = p_*$, if $\rho = p_* \alpha + (1 - p_*) \sigma$, then $\left(\alpha^\dagger | \sigma\right) = 0$, provided[3] $p_* \neq 1$. Besides, if $\left(\alpha^\dagger | \sigma\right) = 0$, then $\left(\alpha^\dagger | \tau\right) = 0$ for any state τ contained in σ. As a consequence, we have the following important corollary, which guarantees the existence of perfectly distinguishable pure states.

Corollary 3. *Every pure state is perfectly distinguishable from some other pure state.*

Proof. Let us consider the invariant state χ. For every normalized pure state α, we have $\chi = p_{\max} \alpha + (1 - p_{\max}) \sigma$ (see lemma 1), where σ is another normalized state. By proposition 11, $\left(\alpha^\dagger | \sigma\right) = 0$. If σ is pure, then α is perfectly distinguishable from σ by means of the observation-test $\left\{ \alpha^\dagger, u - \alpha^\dagger \right\}$. If σ is mixed, than $\left(\alpha^\dagger | \psi\right) = 0$ for every pure state ψ contained in σ. Therefore α is perfectly distinguishable from ψ again via the observation-test $\left\{ \alpha^\dagger, u - \alpha^\dagger \right\}$. $\qquad\square$

It is quite remarkable that the existence of perfectly distinguishable (pure) states pops out from the axioms, without being assumed from the start. In principle, the general theories considered in our framework might not have had any perfectly distinguishable states at all!

---

[3]If $p_* = 1$, then ρ is pure, and we are done. Therefore, without loss of generality we can assume $p_* \neq 1$.

5.1 The diagonalization theorem

Theorem 3. *In a theory satisfying Causality, Purity Preservation, Purification, and Pure Sharpness every state of every system can be diagonalized.*

The proof uses the following lemma, which provides a condition for the perfect distinguishability of a set of pure states:

Lemma 2. *If the pure states $\{\alpha_i\}_{i=1}^n$ satisfy the condition $\left(\alpha_i^\dagger | \alpha_j\right) = 0$ for every $j > i$, they are perfectly distinguishable.*

Proof. By hypothesis, the observation-test $\left\{\alpha_i^\dagger, u - \alpha_i^\dagger\right\}$ distinguishes perfectly between α_i and all the other pure states α_j with $j > i$. Equivalently, the test distinguishes perfectly between α_i and the mixed state $\rho_i := \frac{1}{n-i}\sum_{j>i}\alpha_j$. As a result, we have the condition $\left(u - \alpha_i^\dagger | \rho_i\right) = 1$. Applying proposition 6, we can construct a transformation $\mathscr{A}_i^\perp$, which occurs with the same probability as $u - \alpha_i^\dagger$, such that $\mathscr{A}_i^\perp =_{\rho_i} \mathscr{I}$, and, specifically,

$$\mathscr{A}_i^\perp \alpha_j = \alpha_j \qquad \forall j > i.$$

Moreover, the transformation $\mathscr{A}_i^\perp$ never occurs on the state α_i. Let $\left\{\mathscr{A}_i, \mathscr{A}_i^\perp\right\}$ be a binary test containing the transformation $\mathscr{A}_i^\perp$. By construction, this test distinguishes without error between the state α_i and all the states α_j with $j > i$, in such a way that the latter are not disturbed. Using the tests $\left\{\mathscr{A}_i, \mathscr{A}_i^\perp\right\}$ it is easy to construct a protocol that distinguishes perfectly between the states $\{\alpha_i\}_{i=1}^n$. The protocol works as follows: for i going from 1 to $n-1$, perform the test $\left\{\mathscr{A}_i, \mathscr{A}_i^\perp\right\}$. If the transformation $\mathscr{A}_i$ takes place, then the state is α_i. If the transformation $\mathscr{A}_i^\perp$ takes place, then perform the test $\left\{\mathscr{A}_{i+1}, \mathscr{A}_{i+1}^\perp\right\}$, and so on. $\square$

Proof of theorem 3. The proof consists of a constructive procedure for diagonalizing arbitrary states. In order to diagonalize the state ρ, it is enough to proceed along the following steps:

1. Set $\rho_1 = \rho$ and $p_{*,0} = 0$

2. Starting from $i = 1$, decompose ρ_i as $\rho_i = p_{*,i}\alpha_i + (1 - p_{*,i})\sigma_i$ as in theorem 2, and set $\rho_{i+1} = \sigma_i$, $p_i = p_{*,i}\prod_{j=0}^{i-1}(1 - p_{*,j})$. If $p_{*,i} = 1$, then stop, otherwise continue to the step $i+1$.

Recall that, at every step of the procedure, proposition 11 guarantees the condition $\left(\alpha_i^\dagger | \sigma_i\right) = 0$. Since by construction every state α_j with $j > i$ is contained in the convex decomposition of σ_i, we also have $\left(\alpha_k^\dagger | \alpha_l\right) = 0$ for $l > k$. Hence, lemma 2 implies that the states $\{\alpha_k\}_{k=1}^i$, generated by the first i iterations of the protocol, are perfectly distinguishable, for any i. For a finite dimensional system, this means that the procedure has to terminate in a finite number of iterations. Once the procedure has been completed, the state ρ is diagonalized as $\rho = \sum_i p_i \alpha_i$. $\square$

Note that the diagonalization procedure in the above proof returns a diagonalization of ρ where the eigenvalues are naturally listed in decreasing order, namely $p_i \geq p_{i+1}$ for every i. Such an ordering will become useful when dealing with majorization.

5.2   Unique vs non-unique diagonalization and the majorization criterion

In quantum theory the diagonalization of every state is unique, up to different choices of bases for degenerate eigenspaces. Is this property satisfied by the operational diagonalization? In general, it is conceivable that different diagonalization procedures may yield different sets of eigenvalues for the same state. On top of that, even our algorithm for diagonalizing states may not yield a single, canonical diagonalization. It *does* when the eigenvalues are all distinct, but the situation may be different when two eigenvalues coincide.

The uniqueness of the eigenvalues of a state is particularly important. In a theory where the diagonalization is not unique any attempt to define entropies from the eigenvalues is in serious danger of failure: indeed, the resulting entropies would not be functions of the state, but rather of its diagonalization. At this stage it is not clear whether the present set of axioms (Causality, Purity Preservation, Purification, and Pure Sharpness) implies that all the diagonalizations of a given state have the same eigenvalues. We conjecture that the answer is affirmative and plan to provide a rigorous proof in a forthcoming paper [19]. For the moment, in this paper we will prove an intermediate result, showing that the eigenvalues are unique if one assumes the Strong Symmetry axiom by Barnum, Müller, and Ududec [7] in addition to our axioms.

6   Combining diagonalization with Strong Symmetry

Strong Symmetry is a requirement on the ability to transform maximal sets of perfectly distinguishable pure states using reversible channels. In general, a maximal set is defined as follows:

Definition 5. Let $\{\rho_i\}_{i=1}^n$ be a set of perfectly distinguishable states. We say that $\{\rho_i\}_{i=1}^n$ is *maximal* if there is *no* state ρ_{n+1} such that the states $\{\rho_i\}_{i=1}^{n+1}$ are perfectly distinguishable.

When the maximal set is made of pure states, this definition gives an operational characterization of the orthonormal bases of a finite-dimensional Hilbert space. Another operational of characterization of them was given in Ref. [25] in terms of commutative †-Frobenius monoids.

With this definition, Strong Symmetry reads

Axiom 5 (Strong Symmetry [7])**.** *The group of reversible channels acts transitively on maximal sets of perfectly distinguishable* pure *states.*

Strong Symmetry implies that all maximal sets of perfectly distinguishable pure states have the same cardinality, sometimes referred to as the *dimension* of the system. We call a system of dimension d a *d-level system*. Note that in a d-level system, the diagonalizations of a state have at most d terms.

In the following we present a number of results arising from the combination of diagonalization with Strong Symmetry. These results were preliminarily discussed in the master's thesis of one of the authors [54] and, more recently, they have appeared independently in Refs. [41, 3]. The first result is that the eigenvalues of the invariant state are uniquely defined:

Proposition 13. *Every diagonalization of the invariant state $\chi = \sum_{i=1}^d p_i \alpha_i$ has $p_i = \frac{1}{d}$, for every i.*

Proof. Let $\{a_i\}_{i=1}^d$ be the test that perfectly distinguishes between the states $\{\alpha_i\}_{i=1}^d$. Then, one has $p_i = (a_i|\chi)$, for every i. Let us consider all the possible permutations of the pure states $\{\alpha_i\}_{i=1}^d$. For instance, if $\pi \in S_d$, where S_d is the symmetric group over d elements, we can consider the permuted

states $\{\alpha_{\pi(i)}\}_{i=1}^d$, which are obviously still perfectly distinguishable. By Strong Symmetry, there is a reversible channel $\mathscr{U}_\pi$ that implements this permutation, namely $\alpha_{\pi(i)} = \mathscr{U}_\pi \alpha_i$. Let us apply $\mathscr{U}_\pi$ to χ.

$$\chi = \sum_{j=1}^d p_j \mathscr{U}_\pi \alpha_j = \sum_{j=1}^d p_j \alpha_{\pi(j)}$$

Now let us apply a_i to χ. We have

$$p_i = (a_i|\chi) = \sum_{j=1}^d p_j \delta_{i,\pi(j)} = p_{\pi^{-1}(i)}.$$

Since this holds for every $\pi \in S_d$, one has $p_i = p_j$ for every j. This implies that the eigenvalues are equal, therefore $p_i = \frac{1}{d}$. □

Proposition 13 implies that the pure states arising in every diagonalization of the invariant state χ form a maximal set of perfectly distinguishable pure states. One can wonder about the converse: is it true that every maximal set of perfectly distinguishable pure states, combined with equal weights, yields the invariant state? In this case, the answer is immediate from Strong Symmetry:

Proposition 14. *Let $\{\psi_i\}_{i=1}^d$ be a maximal set of perfectly distinguishable pure states. Then one has $\chi = \frac{1}{d}\sum_{i=1}^d \psi_i$.*

Proof. Let us consider a diagonalization of χ, say $\chi = \frac{1}{d}\sum_{i=1}^d \varphi_i$. By Strong Symmetry, there is a reversible channel $\mathscr{U}$ such that $\mathscr{U}\varphi_i = \psi_i$ for every i. Then we have

$$\chi = \mathscr{U}\chi = \frac{1}{d}\sum_{i=1}^d \mathscr{U}\varphi_i = \frac{1}{d}\sum_{i=1}^d \psi_i.$$

□

So far we have used the diagonalization theorem as a "black box", without referring to the axioms used to prove it. Using the full power of the axioms allows us to prove stronger results. For example, we are able to prove that every pure maximal set admits a pure, perfectly distinguishing test:

Lemma 3. *For every pure maximal set $\{\alpha_i\}_{i=1}^d$, the pure effects $\left\{\alpha_i^\dagger\right\}_{i=1}^d$ form an observation-test, which distinguishes perfectly between the states $\{\alpha_i\}_{i=1}^d$.*

Proof. Let us consider the pure maximal set $\{\alpha_i\}_{i=1}^d$. By proposition 14, we know that $\chi = \frac{1}{d}\sum_{i=1}^d \alpha_i$. Let us prove that the perfectly distinguishing test for $\{\alpha_i\}_{i=1}^d$ is pure, namely made of the pure effects $\left\{\alpha_i^\dagger\right\}_{i=1}^d$. Recalling lemma 1, each α_i arises in the the diagonalization of χ with weight $p_* = \frac{1}{d}$, and one has that $\left(\alpha_i^\dagger|\alpha_j\right) = 0$ for $j \neq i$. Therefore $\left(\alpha_i^\dagger|\alpha_j\right) = \delta_{ij}$. We want now to prove that $\left\{\alpha_i^\dagger\right\}_{i=1}^d$ is an observation-test. Thanks to Purification, it is sufficient to show that $\sum_{i=1}^d \alpha_i^\dagger = u$ (see proposition 7). Let us consider a purification Φ_{AB} of the invariant state χ_A. By theorem 2, one has

for some set of pure states $\{\beta_i\}_{i=1}^d$. By theorem 2, we know that a diagonalization of the complementary state is $\widetilde{\rho} = \frac{1}{d}\sum_{i=1}^d \beta_i$. Hence, we have the equality

where the last equality follows from the definition of the complementary state $\widetilde{\rho}$. Since χ_A is completely mixed, Φ_{AB} is faithful for effects of system A (by proposition 4). Therefore we conclude that $\sum_{i=1}^d \alpha_i^\dagger = u$, thus proving that $\left\{\alpha_i^\dagger\right\}_{i=1}^d$ is an observation-test.   □

The above lemma allows us to prove an important result, which will be essential for the theory of majorization discussed in the next section. The result is the following:

Lemma 4. *Let $\{\psi_i\}_{i=1}^d$ and $\{\varphi_i\}_{i=1}^d$ be two maximal sets of perfectly distinguishable pure states. The matrix with entries $\left(\psi_i^\dagger|\varphi_j\right)$ is doubly stochastic*[4].

Proof. Clearly $\left(\psi_i^\dagger|\varphi_j\right) \geq 0$ because $\left(\psi_i^\dagger|\varphi_j\right)$ is a probability. Let us calculate $\sum_{i=1}^d \left(\psi_i^\dagger|\varphi_j\right)$. By lemma 3 we know that $\left\{\psi_i^\dagger\right\}_{i=1}^d$ is an observation-test and therefore

$$\sum_{i=1}^d \left(\psi_i^\dagger|\varphi_j\right) = (u|\varphi_j) = 1,$$

because the φ_j's are normalized. On the other hand, we know that the invariant state can be decomposed as

$$\chi = \frac{1}{d}\sum_{j=1}^d \varphi_j = \frac{1}{d}\sum_{i=1}^d \psi_i$$

(cf. proposition 14). Hence, we have

$$\sum_{j=1}^d \left(\psi_i^\dagger|\varphi_j\right) = d\left(\psi_i^\dagger|\chi\right) = d\cdot\frac{1}{d} = 1.$$

This proves that the matrix with entries $\left(\psi_i^\dagger|\varphi_j\right)$ is doubly stochastic.   □

Double stochasticity will be the key ingredient for the results of the following section.

7   Majorization and the resource theory of purity

Majorization is traditionally used as a criterion to compare the degree of mixedness of probability distributions. Here we extend this approach to general probabilistic theories satisfying our axioms and, provisionally, Strong Symmetry. In order to define the degree of mixedness operationally, we adopt the resource theory of purity defined in our earlier work [21], which considered the situation where an experimenter has limited control on the dynamics of a closed system. In this scenario, the set of free operations are the *Random Reversible (RaRe) channels*, defined as random mixtures of reversible transformations:

---

[4]See chapter 2, A.1 of Ref. [44] for the definition of doubly stochastic matrix.

Definition 6. A channel $\mathscr{R}$ is *RaRe* if there exist a probability distribution $\{p_i\}_{i\in\mathsf{X}}$ and a set of reversible channels $\{\mathscr{U}_i\}_{i\in\mathsf{X}}$ such that $\mathscr{R} = \sum_{i\in\mathsf{X}} p_i \mathscr{U}_i$.

By definition, RaRe channels cannot increase the purity of a state. If $\rho = \mathscr{R}\sigma$, where $\mathscr{R}$ is a RaRe channel, we say that ρ is *more mixed*[5] than σ [21]. If ρ is more mixed than σ and σ is more mixed than ρ we say that ρ and σ are *equally mixed*.

Like in all resource theories, it is important to devise some methods capable of detecting the convertibility of states under free operations [23], which gives the (pre)ordering of states. We will now show that, under the assumptions made so far in our paper, the ordering of states according to their mixedness is completely determined by majorization, just as it happens in quantum theory [49]. Let us start by recalling the definition of majorization:

Definition 7. Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in $\mathbb{R}^d$, with the components arranged in *decreasing* order. Then, $\mathbf{x}$ is *majorized* by $\mathbf{y}$ (or $\mathbf{y}$ *majorizes* $\mathbf{x}$), and we write $\mathbf{x} \preceq \mathbf{y}$, if

- $\sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i$, for every $k = 1, \ldots, d-1$
- $\sum_{i=1}^{d} x_i = \sum_{i=1}^{d} y_i$.

It is known that $\mathbf{x} \preceq \mathbf{y}$ if and only if $\mathbf{x} = P\mathbf{y}$, where P is a doubly stochastic matrix [34, 44].

Thanks to the results proved in the previous section, we are now in the position to show that majorization of the eigenvalues is a necessary condition for the mixedness ordering of two states:

Theorem 4. *In a theory satisfying Causality, Purity Preservation, Purification, Pure Sharpness, and Strong Symmetry, let ρ and σ be two states of a generic system and let $\mathbf{p}$ and $\mathbf{q}$ be the vectors of the eigenvalues in the diagonalizations of ρ and σ. If ρ is more mixed than σ, then $\mathbf{p} \preceq \mathbf{q}$.*

Proof. If ρ is more mixed than σ, by definition, we have $\rho = \sum_k \lambda_k \mathscr{U}_k \sigma$, where $\{\lambda_k\}$ is a probability distribution and $\mathscr{U}_k$ is a reversible channel, for every k. Suppose $\rho = \sum_{j=1}^{d} p_j \psi_j$ and $\sigma = \sum_{j=1}^{d} q_j \varphi_j$ are diagonalizations of ρ and σ. Then, $\rho = \sum_k \lambda_k \mathscr{U}_k \sigma$ becomes

$$\sum_{j=1}^{d} p_j \psi_j = \sum_k \lambda_k \sum_{j=1}^{d} q_j \mathscr{U}_k \varphi_j.$$

By applying $\psi_i^\dagger$ we get

$$p_i = \sum_{j=1}^{d} q_j \sum_k \lambda_k \left(\psi_i^\dagger \middle| \mathscr{U}_k \middle| \varphi_j \right).$$

This expression can be rewritten as $p_i = \sum_{j=1}^{d} P_{ij} q_j$, where

$$P_{ij} := \sum_k \lambda_k \left(\psi_i^\dagger \middle| \mathscr{U}_k \middle| \varphi_j \right)$$

Now, $\left(\psi_i^\dagger \middle| \mathscr{U}_k \middle| \varphi_j \right)$ is a doubly stochastic matrix because $\{ \mathscr{U}_k \varphi_j \}_{j=1}^{d}$ is a maximal set of perfectly distinguishable pure states. Since the set of doubly stochastic matrices is convex [44], P_{ij} is a doubly stochastic matrix, whence the thesis. ◻

As a corollary, we prove the desired result about the uniqueness of the eigenvalues.

Corollary 4. *In a theory satisfying Causality, Purity Preservation, Purification, Pure Sharpness, and Strong Symmetry, all the diagonalizations of a given state have the same eigenvalues.*

Proof. Let $\rho = \sum_{i=1}^{d} p_i \psi_i$ and $\rho = \sum_{j=1}^{d} q_j \varphi_j$ be two diagonalizations of a generic state ρ, and let $\mathbf{p}$ and $\mathbf{q}$ be the corresponding vectors of eigenvalues. Trivially, ρ is more mixed than ρ, which implies $\mathbf{p} \preceq \mathbf{q}$, but also $\mathbf{q} \preceq \mathbf{p}$, therefore $\mathbf{p} = \Pi \mathbf{q}$, for some permutation matrix Π [44]. This means that $\mathbf{p}$ and $\mathbf{q}$ differ only by a rearrangement of their entries, whence the eigenvalues of ρ are uniquely defined. $\qquad\square$

In Ref. [47] Müller and Masanes proved that two states that are equally mixed (in our terminology) differ by a reversible channel. For theories satisfying the axioms adopted in this paper, majorization provides an alternative proof:

Proposition 15. *In a theory satisfying Causality, Purity Preservation, Purification, Pure Sharpness, and Strong Symmetry, two states ρ and σ are equally mixed under RaRe channels if and only if $\rho = \mathscr{U} \sigma$, for some reversible channel $\mathscr{U}$. In particular, two equally mixed states must have the same eigenvalues.*

Proof. Sufficiency is straightforward. The proof of necessity is close to the proof of corollary 4. If ρ is equivalent to σ, then $\mathbf{p} \preceq \mathbf{q}$ and $\mathbf{q} \preceq \mathbf{p}$, where $\mathbf{p}$ and $\mathbf{q}$ are the vectors of the eigenvalues of ρ and σ respectively. This means that ρ and σ have the same eigenvalues (see above). Thus, $\rho = \sum_{i=1}^{d} p_i \psi_i$ and $\sigma = \sum_{i=1}^{d} p_i \varphi_i$. By Strong Symmetry, there exists a reversible channel $\mathscr{U}$ such that $\mathscr{U} \varphi_i = \psi_i$ for every i. Therefore,

$$\mathscr{U} \sigma = \sum_{i=1}^{d} p_i \mathscr{U} \varphi_i = \sum_{i=1}^{d} p_i \psi_i = \rho.$$

$\square$

We conclude this section by providing a complete equivalence between majorization and the mixedness relation. While in theorem 4 we proved that majorization of the eigenvalues is a necessary condition for the mixedness ordering, we now show that majorization is also sufficient:

Theorem 5. *In a theory satisfying Causality, Purity Preservation, Purification, Pure Sharpness, and Strong Symmetry, let ρ and σ be two states of a generic system and let $\mathbf{p}$ and $\mathbf{q}$ be the vectors of their eigenvalues respectively. If $\mathbf{p} \preceq \mathbf{q}$, then ρ is more mixed than σ.*

Proof. If $\mathbf{p} \preceq \mathbf{q}$, one has $\mathbf{p} = P\mathbf{q}$ for some doubly stochastic matrix P [34, 44]. Now, by Birkhoff's theorem [10, 44], $P = \sum_k \lambda_k \Pi_k$, where the Π_k's are permutation matrices and $\{\lambda_k\}$ is a probability distribution. Therefore $\mathbf{p} = \sum_k \lambda_k \Pi_k \mathbf{q}$; specifically, this means that $p_i = \sum_k \lambda_k \sum_{j=1}^{d} [\Pi_k]_{ij} q_j$. Therefore, we have

$$\rho = \sum_{i=1}^{d} p_i \psi_i = \sum_{i=1}^{d} \sum_k \lambda_k \sum_{j=1}^{d} [\Pi_k]_{ij} q_j \psi_i =$$

$$= \sum_k \lambda_k \sum_{j=1}^{d} q_j \sum_{i=1}^{d} [\Pi_k]_{ij} \psi_i. \tag{3}$$

Now, $\sum_{i=1}^{d} [\Pi_k]_{ij} \psi_i$ is a pure state, given by $\psi_{\pi_k(i)}$ for a suitable permutation $\pi_k \in S_d$. By Strong Symmetry, the permutation π_k is implemented by a reversible channel $\mathscr{V}_k$. Moreover, Strong Symmetry implies that there exists a reversible channel $\mathscr{U}$ such that $\mathscr{U} \varphi_i = \psi_i$ for every $i \in \{1, \dots, d\}$. Defining $\mathscr{U}_k := \mathscr{V}_k \mathscr{U}$, we then have

$$\mathscr{U}_k \varphi_i = \mathscr{V}_k \psi_i = \psi_{\pi_k(i)} = \sum_{i=1}^{d} [\Pi_k]_{ij} \psi_i,$$

which combined with Eq. (3) yields

$$\rho = \sum_k \lambda_k \sum_{j=1}^{d} q_j \mathcal{U}_k \varphi_j = \sum_k \lambda_k \mathcal{U}_k \sigma.$$

Hence, ρ is more mixed than σ.   □

8 Conclusions

In this work we have derived the diagonalization of states from four basic operational axioms: Causality, Purity Preservation, Purification, and Pure Sharpness. Our result has several applications: first of all, it allows one to import all the known consequences of diagonalization in the axiomatic context, such as those presented in Ref. [7], where diagonalization was assumed as Axiom 1. For example, adding Strong Symmetry, we obtain that the state space is self-dual—a property that plays an important role in the reconstruction of quantum theory [6]. The combination of our four axioms with Strong Symmetry leads to important consequences, such as the fact that the eigenvalues in the diagonalization of a state are uniquely determined. While our results use Strong Symmetry, it remains as an open question whether this requirement can be dropped or replaced by other, weaker requirements. We conjecture that this is indeed the case, and we plan to investigate the issue further in a forthcoming paper [19].

Another important application of our results is in the axiomatic reconstruction of (quantum) thermodynamics. In a previous work [21], we defined an operational resource theory of purity—dual to the resource theory of entanglement—in which free operations are random reversible channels. A natural application of the diagonalization theorem is the formulation of a majorization criterion capable of detecting whether a thermodynamic transition is possible or not, and to establish quantitative measures of mixedness [54]. Specifically, when Strong Symmetry is added to our axioms, the ordering of states in the operational resource theory of purity is completely characterized by the majorization criterion. Such an application contributes also to the difficult problem of finding the right requirements that guarantee a well-behaved notion of entropy in general probabilistic theories [5, 56, 40]. To some extent, our results suggest that having a sensible notion of entropy (and therefore having a sensible thermodynamics) is not a generic feature of general probabilistic theories, but rather a quite stringent constraint. In addition to the application to the axiomatization of quantum thermodynamics, it is our hope that this work will contribute to the development of an axiomatic approach to information theory—in particular including data compression and transmission over noisy channels.

References

[1] S. Abramsky & B. Coecke (2004): *A categorical semantics of quantum protocols*. In: *Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science*, pp. 415–425, doi:10.1109/LICS.2004.1319636.

[2] S. Abramsky & B. Coecke (2008): *Categorical Quantum Mechanics*. In K. Engesser, D. M. Gabbay & D. Lehmann, editors: *Handbook of Quantum Logic and Quantum Structures: Quantum Logic*, Elsevier, pp. 261–324.

[3] H. Barnum, J. Barrett, M. Krumm & M. P. Müller (2015): *Entropy, majorization and thermodynamics in general probabilistic theories*. arXiv:1508.03107.

[4] H. Barnum, J. Barrett, M. Leifer & A. Wilce (2007): *Generalized No-Broadcasting Theorem*. Phys. Rev. Lett. 99, p. 240501, doi:10.1103/PhysRevLett.99.240501.

[5] H. Barnum, J. Barrett, L. Orloff Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce & R. Wilke (2010): *Entropy and information causality in general probabilistic theories*. New Journal of Physics 12(3), p. 033024, doi:10.1088/1367-2630/12/3/033024.

[6] H. Barnum, C. P. Gaebler & A. Wilce (2013): *Ensemble Steering, Weak Self-Duality, and the Structure of Probabilistic Theories*. Foundations of Physics 43(12), pp. 1411–1427, doi:10.1007/s10701-013-9752-2.

[7] H. Barnum, M. P. Müller & C. Ududec (2014): *Higher-order interference and single-system postulates characterizing quantum theory*. New Journal of Physics 16(12), p. 123029, doi:10.1088/1367-2630/16/12/123029.

[8] H. Barnum & A. Wilce (2011): *Information Processing in Convex Operational Theories*. Electronic Notes in Theoretical Computer Science 270(1), pp. 3–15, doi:10.1016/j.entcs.2011.01.002. Proceedings of the Joint 5th International Workshop on Quantum Physics and Logic and 4th Workshop on Developments in Computational Models (QPL/DCM 2008).

[9] J. Barrett (2007): *Information processing in generalized probabilistic theories*. Phys. Rev. A 75, p. 032304, doi:10.1103/PhysRevA.75.032304.

[10] G. Birkhoff (1946): *Tres observaciones sobre el algebra lineal*. Univ. Nac. Tucumán Rev. Ser. A 5, pp. 147–151.

[11] P. Bocchieri & A. Loinger (1959): *Ergodic Foundation of Quantum Statistical Mechanics*. Phys. Rev. 114, pp. 948–951, doi:10.1103/PhysRev.114.948.

[12] F. G. S. L. Brandão & M. Cramer (2015): *Equivalence of Statistical Mechanical Ensembles for Non-Critical Quantum Systems*. arXiv:1502.03263.

[13] F. G. S. L. Brandão, M. Horodecki, N. Ng, J. Oppenheim & S. Wehner (2015): *The second laws of quantum thermodynamics*. Proceedings of the National Academy of Sciences 112(11), pp. 3275–3279, doi:10.1073/pnas.1411728112.

[14] G. Chiribella (2014): *Dilation of states and processes in operational-probabilistic theories*. In B. Coecke, I. Hasuo & P. Panangaden, editors: Proceedings 11th workshop on *Quantum Physics and Logic*, Kyoto, Japan, 4-6th June 2014, *Electronic Proceedings in Theoretical Computer Science* 172, Open Publishing Association, pp. 1–14, doi:10.4204/EPTCS.172.1.

[15] G. Chiribella, G. M. D'Ariano & P. Perinotti (2010): *Probabilistic theories with purification*. Phys. Rev. A 81, p. 062348, doi:10.1103/PhysRevA.81.062348.

[16] G. Chiribella, G. M. D'Ariano & P. Perinotti (2011): *Informational derivation of quantum theory*. Phys. Rev. A 84, p. 012311, doi:10.1103/PhysRevA.84.012311.

[17] G. Chiribella, G. M. D'Ariano & P. Perinotti (2012): *Quantum theory, namely the pure and reversible theory of information*. Entropy 14(10), pp. 1877–1893, doi:10.3390/e14101877.

[18] G. Chiribella, G. M. D'Ariano & P. Perinotti (2016): *Quantum from principles*. In G. Chiribella & R. W. Spekkens, editors: *Quantum Theory: Informational Foundations and Foils*, Springer Netherlands, Dordrecht, pp. 171–222, doi:10.1007/978-94-017-7303-4.

[19] G. Chiribella & C. M. Scandolo: *Towards an axiomatic foundation of (quantum) thermodynamics*. In preparation.

[20] G. Chiribella & C. M. Scandolo (2015): *Conservation of information and the foundations of quantum mechanics*. EPJ Web of Conferences 95, p. 03003, doi:10.1051/epjconf/20149503003.

[21] G. Chiribella & C. M. Scandolo (2015): *Entanglement and thermodynamics in general probabilistic theories*. *arXiv:1504.07045*.

[22] B. Coecke (2010): *Quantum picturalism*. *Contemporary Physics* 51, pp. 59–83, doi:10.1080/00107510903257624.

[23] B. Coecke, T. Fritz & R. W. Spekkens (2014): *A mathematical theory of resources*. *arXiv:1409.5531*.

[24] B. Coecke & É. O. Paquette (2011): *Categories for the Practising Physicist*. In B. Coecke, editor: *New Structures for Physics*, Lecture Notes in Physics 813, Springer, Berlin, Heidelberg, pp. 173–286, doi:10.1007/978-3-642-12821-9_3.

[25] B. Coecke, D. Pavlovic & J. Vicary (2013): *A new description of orthogonal bases*. *Mathematical Structures in Computer Science* 23, pp. 555–567, doi:10.1017/S0960129512000047.

[26] G. M. D'Ariano (2010): *Probabilistic theories: what is special about quantum mechanics?* In A. Bokulich & G. Jaeger, editors: *Philosophy of Quantum Information and Entanglement*, Cambridge University Press, Cambridge, pp. 85–126, doi:10.1017/CBO9780511676550.007.

[27] G. M. D'Ariano, F. Manessi, P. Perinotti & A. Tosini (2014): *Fermionic computation is non-local tomographic and violates monogamy of entanglement*. *EPL (Europhysics Letters)* 107(2), p. 20009, doi:10.1209/0295-5075/107/20009.

[28] G. M. D'Ariano, F. Manessi, P. Perinotti & A. Tosini (2014): *The Feynman problem and fermionic entanglement: Fermionic theory versus qubit theory*. *International Journal of Modern Physics A* 29(17), p. 1430025, doi:10.1142/S0217751X14300257.

[29] P. Faist, F. Dupuis, J. Oppenheim & R. Renner (2015): *The minimal work cost of information processing*. *Nature Communications* 6, doi:10.1038/ncomms8669.

[30] J. Gemmer, M. Michel & G. Mahler (2009): *Quantum Thermodynamics: Emergence of Thermodynamic Behavior Within Composite Quantum Systems*. Lecture Notes in Physics 784, Springer Verlag, Berlin, Heidelberg, doi:10.1007/978-3-540-70510-9.

[31] J. Gemmer, A. Otte & G. Mahler (2001): *Quantum Approach to a Derivation of the Second Law of Thermodynamics*. *Phys. Rev. Lett.* 86, pp. 1927–1930, doi:10.1103/PhysRevLett.86.1927.

[32] S. Goldstein, J. L. Lebowitz, R. Tumulka & N. Zanghì (2006): *Canonical Typicality*. *Phys. Rev. Lett.* 96, p. 050403, doi:10.1103/PhysRevLett.96.050403.

[33] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens & N. Yunger Halpern (2015): *The resource theory of informational nonequilibrium in thermodynamics*. *Physics Reports* 583, pp. 1–58, doi:10.1016/j.physrep.2015.04.003.

[34] G. H. Hardy, J. E. Littlewood & G. Pólya (1929): *Some simple inequalities satisfied by convex functions*. *Messenger Math* 58(145–152), p. 310.

[35] L. Hardy (2001): *Quantum theory from five reasonable axioms*. *arXiv quant-ph/0101012*.

[36] L. Hardy (2011): *Foliable operational structures for general probabilistic theories*. In H. Halvorson, editor: *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, Cambridge University Press, Cambridge, pp. 409–442, doi:10.1017/CBO9780511976971.013.

[37] L. Hardy (2011): *Reformulating and reconstructing quantum theory*. *arXiv:1104.2066*.

[38] L. Hardy (2016): *Reconstructing quantum theory*. In G. Chiribella & R. W. Spekkens, editors: *Quantum Theory: Informational Foundations and Foils*, Springer Netherlands, Dordrecht, pp. 223–248, doi:10.1007/978-94-017-7303-4.

[39] M. Horodecki & J. Oppenheim (2013): *Fundamental limitations for quantum and nanoscale thermodynamics*. *Nature Communications* 4, doi:10.1038/ncomms3059.

[40] G. Kimura, K. Nuida & H. Imai (2010): *Distinguishability measures and entropies for general probabilistic theories*. *Reports on Mathematical Physics* 66(2), pp. 175–206, doi:10.1016/S0034-4877(10)00025-X.

[41] M. Krumm (2015): *Thermodynamics and the Structure of Quantum Theory as a Generalized Probabilistic Theory*. *arXiv:1508.03299*. Master's thesis.

[42] S. Lloyd (1988): *Black Holes, Demons, and the Loss of Coherence*. Ph.D. thesis, Rockfeller University. Available at `http://meche.mit.edu/documents/slloyd_thesis.pdf`.

[43] E. Lubkin & T. Lubkin (1993): *Average quantal behavior and thermodynamic isolation*. International Journal of Theoretical Physics 32(6), pp. 933–943, doi:10.1007/BF01215300.

[44] A. W. Marshall, I. Olkin & B. C. Arnold (2011): *Inequalities: Theory of Majorization and Its Applications*. Springer Series in Statistics, Springer, New York, doi:10.1007/978-0-387-68276-1.

[45] M. P. Müller, O. C. O. Dahlsten & V. Vedral (2012): *Unifying Typical Entanglement and Coin Tossing: on Randomization in Probabilistic Theories*. Communications in Mathematical Physics 316(2), pp. 441–487, doi:10.1007/s00220-012-1605-x.

[46] M. P. Müller, D. Gross & J. Eisert (2011): *Concentration of Measure for Quantum States with a Fixed Expectation Value*. Communications in Mathematical Physics 303(3), pp. 785–824, doi:10.1007/s00220-011-1205-1.

[47] M. P. Müller & L. Masanes (2013): *Three-dimensionality of space and the quantum bit: an information-theoretic approach*. New Journal of Physics 15(5), p. 053040, doi:10.1088/1367-2630/15/5/053040.

[48] M. P. Müller, J. Oppenheim & O. C. O. Dahlsten (2012): *The black hole information problem beyond quantum theory*. Journal of High Energy Physics 2012(9):116, doi:10.1007/JHEP09(2012)116.

[49] M. A. Nielsen & I. L. Chuang (2010): *Quantum computation and quantum information*. Cambridge University Press, Cambridge, doi:10.1017/CBO9780511976667.

[50] M. Ozawa (1984): *Quantum measuring processes of continuous observables*. Journal of Mathematical Physics 25(1), pp. 79–87, doi:10.1063/1.526000.

[51] V. Paulsen (2002): *Completely bounded maps and operator algebras*. Cambridge University Press, Cambridge, doi:10.1017/CBO9780511546631.

[52] C. Piron (1976): *Foundations of quantum physics*. Mathematical Physics Monograph Series, Benjamin-Cummings Publishing Company.

[53] S. Popescu, A. J. Short & A. Winter (2006): *Entanglement and the foundations of statistical mechanics*. Nature Physics 2(11), pp. 754–758, doi:10.1038/nphys444.

[54] C. M. Scandolo (2014): *Entanglement and thermodynamics in general probabilistic theories*. Master's thesis, Università degli Studi di Padova, Italy. Available at `http://tesi.cab.unipd.it/46015/1/Scandolo_carlo_maria.pdf`.

[55] P. Selinger (2011): *A Survey of Graphical Languages for Monoidal Categories*. In B. Coecke, editor: *New Structures for Physics*, Lecture Notes in Physics 813, Springer, Berlin, Heidelberg, pp. 289–356, doi:10.1007/978-3-642-12821-9_4.

[56] A. J. Short & S. Wehner (2010): *Entropy in general physical theories*. New Journal of Physics 12(3), p. 033023, doi:10.1088/1367-2630/12/3/033023.

[57] W. F. Stinespring (1955): *Positive functions on C*-algebras*. Proceedings of the American Mathematical Society 6(2), pp. 211–216, doi:10.1090/S0002-9939-1955-0069403-4.

[58] W. Thirring (2002): *Quantum mathematical physics*. Springer-Verlag, Berlin, Heidelberg, doi:10.1007/978-3-662-05008-8.

[59] A. Uhlmann (1971): *Sätze über Dichtematrizen*. Wiss. Z. Karl-Marx Univ. Leipzig 20, pp. 633–637.

[60] A. Uhlmann (1972): *Endlich-dimensionale Dichtematrizen I*. Wiss. Z. der Karl-Marx Univ. Leipzig. 21, pp. 421–452.

[61] A. Uhlmann (1973): *Endlich-dimensionale Dichtematrizen II*. Wiss. Z. Karl-Marx-Univ. Leipzig 22, pp. 139–177.