Operational axioms for diagonalizing states

Giulio Chiribella
gchiribella@mail.tsinghua.edu.cn

Carlo Maria Scandolo
carlomaria.scandolo@gmail.com

Center for Quantum Information, Institute for Interdisciplinary Information Sciences,
Tsinghua University, Beijing 100084, China

In quantum theory every state can be diagonalized, i.e. decomposed as a convex combination of perfectly distinguishable pure states. This feature is crucial for quantum statistical mechanics, as 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 that occurs with unit probability on some state. Using these axioms, we present a constructive method to decompose every given state into perfectly distinguishable pure states.

1 Introduction

One of the possible routes 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 [48, 49, 50, 47, 24, 32, 28, 10]. The viability of this approach heavily relies 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 [29, 21, 8, 3, 13, 14, 7, 31, 30, 12]. The first step in this direction is to prove an operational version of the spectral theorem from basic principles, showing that every state can be decomposed as a mixture of perfectly distinguishable pure states. This will enable one to define the “eigenvalues” of states, which are at the root of the notion of majorization. A diagonalization result was proven in the context of the axiomatization in Ref. [14], 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, potentially, other generalizations of quantum theory [22, 23]. Our list includes three of the six axioms of Ref. [14] (Causality, Purity Preservation, and Purification) plus one 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 [40]. In these terms, Pure Sharpness requires that for every system there exists at least one property, and at least one state that has such a property. Note that none of our axioms assumes that there exist perfectly distinguishable states. 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 can 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 states
is a non-trivial consequence 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 found-
ation of quantum thermodynamics. From this points of view, Purification is almost a mandatory choice:
a standard thermodynamic procedure is to consider an open system in interaction with its environment,
in such a way that the two together form an isolated system. Purification guarantees that one can asso-
ciate 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 was adopted in a number of works aimed at deriving the microcanonical and canonical
ensembles [34, 26, 41, 27, 25, 36, 9], an idea that has been recently explored also in hypothetical
post-quantum scenarios [35, 37].

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

2 Framework

The present analysis is carried out in the framework of general probabilistic theories, adopting the specific
variant of Refs. [13, 14, 12], known as the framework of operational-probabilistic theories (OPTs). OPTs
arise from the marriage of the graphical language of symmetric monoidal categories [1, 2, 19, 20, 43]
with the toolbox of probability theory.

Physical processes can be combined in sequence or in parallel, giving rise to circuits like the following

\[
\begin{array}{cccc}
\rho & \mathcal{A} & \mathcal{A}' & \mathcal{A}'' \\
\mathcal{B} & \mathcal{B}' & a & b
\end{array}
\]

Here, A, A’, A”, B, B’ are systems, \(\rho\) is a bipartite state, \(\mathcal{A}\), \(\mathcal{A}'\) and \(\mathcal{B}\) 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

- \(\text{St}(A)\) the set of states of system A
- \(\text{Eff}(A)\) the set of effects on A
- \(\text{Transf}(A,B)\) the set of transformations from A to B
- \(A \otimes B\) the composition of systems A and B.
- \(\mathcal{A} \otimes \mathcal{B}\) the parallel composition of the transformations \(\mathcal{A}\) and \(\mathcal{B}\).

A particular system is the trivial system I (mathematically, the unit of the tensor product); 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 mean

\[
(a|\rho) := \begin{array}{c}
\rho \\
A \\
a
\end{array}.
\]
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 \(\rho\) and the effect \(a\) in a circuit. The fact that scalars are real numbers induces a notion of sum for transformations.

A test from \(A\) to \(B\) is a collection of transformations \(\{\mathcal{M}_i\}_{i \in X}\) from \(A\) to \(B\), which can occur in an experiment with outcomes in \(X\). If \(B\) is the trivial system, we have an observation-test, which is a collection of effects. 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 \([12]\), 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 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). For every pair of systems \(A, B\), and every test \(\{\mathcal{M}_i\}_{i \in X}\) from \(A\) to \(B\), there exist a system \(C\), a transformation \(\mathcal{M} \in \text{Transf}(A, B \otimes C)\), and an observation-test \(\{c_i\}_{i \in X}\) such that

\[
\begin{array}{c}
A \xrightarrow{\mathcal{M}} B \\
\text{c}_i \xrightarrow{\mathcal{M}}
\end{array}
\]

\(\forall i \in X\).

A channel \(\mathcal{W}\) from \(A\) to \(B\) is called reversible if there exists another channel \(\mathcal{W}^{-1}\) from \(B\) to \(A\) such that \(\mathcal{W}^{-1} \mathcal{W} = \mathcal{I}_A\) and \(\mathcal{W} \mathcal{W}^{-1} = \mathcal{I}_B\), where \(\mathcal{I}_S\) is the identity channel on a generic system \(S\). A state \(\chi \in \text{St}(A)\) is called invariant if \(\mathcal{W}\chi = \chi\), for every reversible channel \(\mathcal{W}\). 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 \(A \otimes B \simeq B \otimes A\).

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 \(\{\mathcal{E}_i\}_{i \in X}\) is a coarse-graining of the test \(\{\mathcal{D}_j\}_{j \in Y}\) if there is a partition \(\{Y_i\}_{i \in X}\) of \(Y\) such that \(\mathcal{E}_i = \sum_{j \in Y_i} \mathcal{D}_j\) for every \(i \in X\). In this case, we say that \(\{\mathcal{D}_j\}_{j \in Y}\) is a refinement of \(\{\mathcal{E}_i\}_{i \in X}\).

A transformation \(\mathcal{E} \in \text{Transf}(A, B)\) is called pure if it has only trivial refinements, namely for every refinement \(\{\mathcal{D}_j\}\) one has \(\mathcal{D}_j = p_j \mathcal{E}\), 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 \(\text{PurTransf}(A, B)\). In the special case of states (resp. effects) of system \(A\) we use the notation \(\text{PurSt}(A)\) (resp. \(\text{PurEff}(A)\)). The set of normalized pure states (resp. effects) of \(A\) will be denoted by \(\text{Eff}_1(A)\).

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

It is clear that no states are compatible with a pure state, except the pure state itself.

**Definition 2.** We say that two transformations \(\mathcal{A}, \mathcal{A}' \in \text{Transf}(A, B)\) are equal upon input of the state \(\rho \in \text{St}_1(A)\) if \(\mathcal{A}\sigma = \mathcal{A}'\sigma\) for every state \(\sigma\) compatible with \(\rho\). In this case we will write \(\mathcal{A} \equiv_{\rho} \mathcal{A}'\).
Our analysis will be in finite dimension, meaning that the vector space spanned by states is finite-dimensional.

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 [13, 14]).** 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 $\text{Tr}_A$ on $A$. Thanks to that, it is possible to define the marginal state $\rho_{AB}$ on system $A$ as $\rho_A := \text{Tr}_B \rho_{AB}$. In a theory satisfying Causality, the norm of a state $\rho$ is simply given by $\|\rho\| = \text{Tr} \rho$.

Moreover, observation-tests are normalized in the following way (see corollary 3 of Ref. [13]):

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

The second axiom in our list is Purity Preservation:

**Axiom 2 (Purity Preservation [21, 14, 16]).** Sequential and parallel compositions of pure transformations are pure transformations.

We consider Purity Preservation as a most 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 [16].

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 [15, 16]. Specifically, we say that a state $\rho \in \text{St}_1(A)$ can be purified if there exists a pure state $\Psi \in \text{PurSt}_1(A \otimes B)$ that has $\rho$ as its marginal on system $A$. In this case, we call $\Psi$ a purification of $\rho$. The axiom is as follows:

**Axiom 3 (Purification [13]).** 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 $\Psi, \Psi' \in \text{PurSt}_1(A \otimes B)$ are such that $\text{Tr}_B \Psi_{AB} = \text{Tr}_B \Psi'_{AB}$, then $\Psi'_{AB} = (\mathcal{F}_A \otimes \%_B) \Psi_{AB}$, where $\%_B$ is a reversible channel on $B$.

In quantum theory, the validity of Purification lies at the foundation of all dilation theorems, such as Stinespring’s [46], Naimark’s [39], and Ozawa’s [38]. In the finite dimensional setting, these theorems have been reconstructed axiomatically in [13].

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 $a \in \text{PurEff}(A)$ that occurs with probability 1 on some state.

Pure Sharpness is reminiscent of the Sharpness axiom used in Hardy’s 2011 axiomatization [31], which requires a one to one correspondence between pure states and effects that distinguish maximal sets of states.

---

1The name and the formulation of the axiom adopted here are the same as in Ref. [16]. The original axiom was called Atomicity of Composition [21] 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. [14] this property was a consequence of the Local Tomography axiom, which, instead, is not assumed here.
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 first follows from Purification:

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

\[
p \sigma = \Psi A B b_\sigma.
\]

**Proof.** The proof follows the same lines of theorem 6 and corollary 9 in Ref. [13], with the only difference that here we do not assume the existence of 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.

Moreover, Purification implies that reversible channels act transitively on the set of pure states (see lemma 20 in Ref. [13]):

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

As a consequence, every system possesses one invariant state (see corollary 34 of Ref. [13]):

**Proposition 3.** For every system \( A \), there exists a unique invariant state \( \chi_A \).

Also, transitivity implies that the set of pure states is compact for every system (see corollary 32 of Ref. [13]). This property is generally a non-trivial property—cf. Ref. [4] for a counterexample of a state space with a non-closed set of pure states.

Finally, combining Purification with Purity Preservation one obtains the following properties [14]:

**Proposition 4.** For every observation-test \( \{a_i\}_{i \in X} \) on \( A \), there is a system \( B \) and a test \( \{\mathcal{A}_i\}_{i \in X} \subset \text{Transf}(A, B) \) such that every \( \mathcal{A}_i \) is pure and \( a_i = \text{Tr}_B \mathcal{A}_i \).

**Proposition 5.** Let \( a \) be an effect such that \( (a|\rho) = 1 \), for some \( \rho \in \text{St}_1(A) \). Then there exists a transformation \( \mathcal{F} \) on \( A \) such that \( a = \text{Tr} \mathcal{F} \) and \( \mathcal{F} = \rho \mathcal{I} \), where \( \mathcal{I} \) is the identity.

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

Suppose we know that \( (a|\rho) = 1 \), where \( a \) is a pure effect. Then, we can conclude that the state \( \rho \) must be pure (see lemma 26 and theorem 7 of Ref. [14] for the proof idea):

**Proposition 6.** Let \( a \in \text{PurEff}_1(A) \). Then, there exists a pure state \( \alpha \in \text{PurSt}(A) \) such that \( (a|\alpha) = 1 \). Furthermore, for every \( \rho \in \text{St}(A) \), if \( (a|\rho) = 1 \), then \( \rho = \alpha \).
Combining the above result with our Pure Sharpness axiom, we derive the following

**Proposition 7.** For every pure state \( \alpha \in \text{PurSt}(A) \), there exists at least one pure effect \( a \in \text{PurEff}(A) \) 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 \( \alpha_0 \). By proposition \([6]\), \( \alpha_0 \) is pure. Now, for a generic pure state \( \alpha \), by transitivity, there is a reversible channel \( \mathcal{U} \) such that \( \alpha = \mathcal{U} \alpha_0 \). Hence, the effect \( a := a_0 \mathcal{U}^{-1} \) is pure and \( (a|\alpha) = 1 \). \( \square \)

Now we prove an important theorem, which lies at the core of our procedure for diagonalizing states.

**Theorem 2.** Let \( \rho_A \) be a normalized state of system \( A \) and let \( p_* \) be the probability defined \( p_* := \max_{\alpha \in \text{PurSt}_1(A)} \{ p \in [0, 1] : \rho_A = p \alpha + (1 - p) \sigma, \sigma \in \text{St}_1(A) \} \).

Let \( \Psi_{AB} \) be a purification of \( \rho_A \) and let \( \tilde{\rho}_B \in \text{St}_1(B) \) be the complementary state of \( \rho_A \), namely \( \tilde{\rho}_B := \text{Tr}_A \Psi_{AB} \). Then, there exists a pure state \( \beta \in \text{PurSt}(B) \) such that \( \tilde{\rho}_B = p_* \beta + (1 - p_*) \tau \) for some state \( \tau \in \text{St}(B) \).

**Proof.** By hypothesis, one can write \( \rho_A = p_* \alpha + (1 - p_*) \sigma \), where \( \alpha \) is a pure state and \( \sigma \) is possibly mixed. Let us purify \( \rho_A \), and let \( \Psi_{AB} \) be one of its purifications. According to the steering property, there exists an effect \( b \) that prepares \( \alpha \) with probability \( p_* \), namely

\[
\Psi_{AB} b = p_* \alpha_A.
\]

Let \( a \) be a pure effect such that \( (a|\alpha) = 1 \) (such an effect exists by proposition \([7]\]). Applying \( a \) on both sides of Eq. (1), we get

\[
\Psi_{AB} a = p_* \alpha_A.
\]

On the other hand, applying \( a \) to the state \( \Psi_{AB} \) we obtain

\[
\Psi a = q \beta_B,
\]

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

\[
\Psi_{AB} a b = p_* = q \beta_B b.
\]

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

\[
\Psi_{AB} \tilde{b} = q
\]

\(2\)Note that the maximum is well defined because the set of pure states is compact, thanks to transitivity.
By Purity Preservation, \( \tilde{b} \) will induce a pure state on system A, namely

\[
q = \begin{pmatrix} \Psi \\ \alpha \end{pmatrix} = p \begin{pmatrix} \alpha \\ \tilde{\beta} \end{pmatrix},
\]

where \( p \in [0, 1] \). From the above equation, we have the inequality \( q \leq p \). Since by definition we have \( p \leq p_s \), we finally get the chain of inequalities \( p_s \leq q \leq p \), whence \( p_s = q = \tilde{p} \). Hence, Eq. (2) implies that the pure state \( \tilde{\beta} \) can arise with probability \( p_s \) in a convex decomposition of the state \( \tilde{\rho}_B \).

A similar proof was used in lemma 30 of Ref. [14] in the special case where \( \rho \) is the invariant state, and with stronger assumptions, i.e. Ideal Compression, which is not assumed here.

The effect \( b \) that prepares \( \alpha \) with probability \( p_s \) can always be taken to be pure. Indeed, \( \tilde{b} \) is a pure effect that prepares the pure state \( \tilde{\alpha} \) on A with probability \( \tilde{p} \). But since \( \tilde{p} = p_s \), then \( (a|\tilde{\alpha}) = 1 \). Therefore, by proposition 6 \( \tilde{\alpha} = \alpha \). This shows that \( \alpha \) can always be prepared with probability \( p_s \) by using a pure effect on B.

Now we are ready to prove the uniqueness of the pure effect associated with a pure state. The proof is identical to the that of theorem 8 of Ref. [14], even though we are assuming fewer axioms. In the proof the following lemma is necessary (see lemma 29 of Ref. [14]).

**Lemma 1.** Let \( \chi \) be the invariant state of system A and let \( \alpha \) be a normalized pure state. Then

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

does not depend on \( \alpha \).

**Proposition 8.** For every normalized pure state \( \alpha \) there exists a unique pure effect a such that \( (a|\alpha) = 1 \).

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. [14]).

**Corollary 1.** For every pair of \( a, a' \in \text{PurEff}_1(A) \), there exists a reversible channel \( \mathcal{U} \) on A such that \( a' = aU \).

## 5 Diagonalization of states

A diagonalization of \( \rho \) is a convex decomposition of \( \rho \) into perfectly distinguishable pure states. The probabilities in such a convex decomposition will be called the eigenvalues of \( \rho \). By Carathéodory’s theorem [11, 45], we know that every diagonalization of \( \rho \in \text{St}_1(A) \) can have at most \( D_A + 1 \) terms, where \( D_A \) is the dimension of the vector space spanned by the state space \( \text{St}(A) \) (see Ref. [13]).

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 2).

The starting point for diagonalization is the following

**Proposition 9.** Consider \( \rho = p_s \alpha + (1 - p_s)\sigma \), where \( p_s \) is defined in theorem 2 and let a be the pure effect associated with \( \alpha \). Then \( (a|\rho) = p_s \).

**Proof.** Let \( \Psi_{AB} \) be a purification of \( \rho \). Then, the proof of theorem 2 yields the following equality

\[
\begin{pmatrix} \Psi \\ \alpha \end{pmatrix} = p_s \begin{pmatrix} \beta \\ \tilde{\beta} \end{pmatrix}.
\]
Applying the deterministic effect on both sides of the above equation, we obtain

\[ p_\ast = p_\ast \bigg( \begin{array}{c} \beta \\ B \bigg) \bigg( \begin{array}{c} \Psi \\ A \\ B \end{array} \bigg) = \bigg( \begin{array}{c} \Psi \\ A \\ B \bigg) \bigg( \begin{array}{c} \alpha \\ B \end{array} \bigg) = \bigg( \begin{array}{c} \rho \\ A \\ B \end{array} \bigg) a. \]

This shows that \((a|\rho) = p_\ast\).

The following proposition enables us to define \(p_\ast\) in an alternative way starting from measurements.

**Proposition 10.** Let \(\rho \in \text{St}_1(A)\). Define \(p_\ast := \max_{a \in \text{PurEff}_1(A)} (a|\rho)\). Then \(p_\ast = p_\ast\).

**Proof.** By proposition 9, clearly one has \(p_\ast \geq p_\ast\). Suppose that \(p_\ast > p_\ast\). Since \(p_\ast\) is the maximum, it is achieved by some \(a' \in \text{PurEff}_1(A)\). Therefore,

\[ p_\ast = \bigg( \begin{array}{c} \rho \\ A \\ B \end{array} \bigg) a' = \bigg( \begin{array}{c} \Psi \\ A \\ B \end{array} \bigg) a' = \bigg( \begin{array}{c} \rho \\ A \\ B \end{array} \bigg) a', \]

where \(\Psi_{AB}\) is a purification of \(\rho\). Now, \(a'\) prepares a pure state \(\beta'\) on \(B\) with probability \(\lambda\). If we take the pure effect \(b'\) associated with \(\beta'\), we have, recalling Eq. (3),

\[ p_\ast = \bigg( \begin{array}{c} \Psi \\ A \\ B \end{array} \bigg) a' = \lambda \bigg( \begin{array}{c} \beta' \\ B \bigg) \bigg( \begin{array}{c} \Psi \\ A \\ B \bigg) b' = \bigg( \begin{array}{c} \Psi \\ A \\ B \bigg) b', \]

which means

\[ \bigg( \begin{array}{c} \Psi \\ A \\ B \bigg) a' b' = p_\ast. \]

Now, \(b'\) prepares a pure state \(\alpha'\) on \(A\) with probability \(q\) such that

\[ p_\ast = \bigg( \begin{array}{c} \Psi \\ A \\ B \bigg) a' b' = q \bigg( \begin{array}{c} \alpha' \\ A \bigg) \bigg( \begin{array}{c} \beta' \\ B \bigg) \bigg( \begin{array}{c} \Psi \\ A \\ B \bigg) b' = q \bigg( \begin{array}{c} \alpha' \\ A \bigg) b'. \]

This implies \(q \geq p_\ast > p_\ast\). Therefore \(q > p_\ast\), and the pure state \(\alpha'\) arises in a convex decomposition of \(\rho\) with a weight \(q\) strictly greater than \(p_\ast\). This contradicts the fact that \(p_\ast\) is the maximum weight for a pure state, whence in fact \(p_\ast = p_\ast\). \(\square\)

The result expressed in proposition 9 has important consequences about diagonalization. Since \((a|\rho) = p_\ast\), if \(\rho = p_\ast \alpha + (1 - p_\ast) \sigma\), then \((a|\sigma) = 0\), provided \(p_\ast \neq 1\). Besides, if \((a|\sigma) = 0\), then \((a|\tau) = 0\) for any state \(\tau\) compatible with \(\sigma\). As a consequence, we have the following important corollary, which guarantees the existence of perfectly distinguishable pure states.

**Corollary 2.** Every pure state is perfectly distinguishable from some other pure state.

**Proof.** Let us consider the invariant state \(\chi\). For every normalized pure state \(\alpha\), we have \(\chi = p_{\max} \alpha + (1 - p_{\max}) \sigma\) (see lemma 1), where \(\sigma\) is another normalized state. By proposition 9, if \(a\) is the pure effect associated with \(\alpha\), then \((a|\sigma) = 0\). If \(\sigma\) is pure, then \(\alpha\) is perfectly distinguishable from \(\sigma\) by means of the observation-test \(\{a, Tr - a\}\). If \(\sigma\) is mixed, then \((a|\psi) = 0\) for every pure state \(\psi\) compatible with \(\sigma\). Therefore \(\alpha\) is perfectly distinguishable from \(\psi\) again via the observation-test \(\{a, Tr - a\}\). \(\square\)
5.1 The diagonalization theorem

We devise the following procedure for diagonalizing any state \( \rho \).

1. Once we have determined \( p_1 = q_1 \) and we have found \( \alpha =: \alpha_1 \) such that \( \rho = q_1 \alpha_1 + (1 - q_1) \sigma_1 \), we repeat the same procedure for the state \( \sigma_1 \).

2. We find \( q_2 \), the maximum weight such that we can write \( \sigma_1 = q_2 \alpha_2 + (1 - q_2) \sigma_2 \), where \( \alpha_2 \) is a pure state and \( \sigma_2 \) is another state; and so on.

This process will end sooner or later when we find that the remaining state \( \sigma_i \) is pure. At each step \( i \), \( \alpha_i \) is perfectly distinguishable from \( \sigma_i \) and therefore it is perfectly distinguishable from each of the pure states compatible with \( \sigma_i \).

At the end of this procedure, we can write \( \rho = \sum_i p_i \alpha_i \), where \( p_i = q_i \prod_{j<i} (1 - q_j) \) for \( i > 1 \). Here, by construction, \( p_i \geq p_{i+1} \) for every \( i \), and \( (a_i|\alpha_j) = 0 \) for every \( j > i \), where \( a_i \) is the pure effect associated with \( \alpha_i \).

We claim that all the \( \alpha_i \)'s are perfectly distinguishable. Suppose somebody prepares a normalized pure state taken from the set \( \{\alpha_i\}_{i=1}^n \), and we know that, for every pure state \( \alpha_i \), \( (a_i|\alpha_j) = 0 \) for every \( j > i \), where \( a_i \) is the pure effect associated with \( \alpha_i \). Are we able to distinguish the pure states perfectly, namely to identify with certainty which state has been prepared?

The answer is affirmative. To construct the distinguishing protocol, the key idea is to switch from pure effects to pure transformations that occur with the same probability as the effects, according to proposition 4. We are interested in transformations rather than in effects because we want a procedure that can be iterated, and using effects would mean destroying the system under concern. We act as follows.

1. Let us consider the observation-test \( \{a_1, \text{Tr} - a_1\} \), which perfectly distinguishes \( \alpha_1 \) from the other pure states, and we apply the associated test \( \{\mathcal{A}_1, \mathcal{A}_1^\perp\} \), according to proposition 4. If \( \mathcal{A}_1 \) occurs, we conclude that the state is \( \alpha_1 \). If not, the state is one of the others. Consider \( \rho_1 = \frac{1}{n-1} \sum_{i=2}^n \alpha_i \).

Since \( (\text{Tr} - a_1|\rho_1) = 1 \), because \( (a_1|\alpha_1) = 1 \), \( \mathcal{A}_1^\perp \) leaves all the pure states \( \{\alpha_i\}_{i=2}^n \) invariant, according to proposition 5. Now we can repeat the test.

2. This time we consider the observation-test \( \{a_2, \text{Tr} - a_2\} \), which perfectly distinguishes \( \alpha_2 \) from the remaining states \( \{\alpha_i\}_{i=3}^n \), and we apply the associated test \( \{\mathcal{A}_2, \mathcal{A}_2^\perp\} \). If \( \mathcal{A}_2 \) happens, the state is \( \alpha_2 \). If not, we consider \( \rho_2 = \frac{1}{n-2} \sum_{i=3}^n \alpha_i \). Since \( (\text{Tr} - a_2|\rho_2) = 1 \), then \( \mathcal{A}_2^\perp \) leaves all the pure states \( \{\alpha_i\}_{i=3}^n \) invariant. Now we repeat the procedure again.

In this way, by iterating the procedure several times, we are able to identify the state with certainty.

Therefore, our procedure is a way to write every state as a convex combination of perfectly distinguishable pure states.

6 Conclusions

In this work we 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. 6, where diagonalization is assumed as Axiom 1. For example, combining our axioms with Hardy’s Permutability axiom we obtain that all set of perfectly distinguishable pure states of the same cardinality can be converted into one another by reversible channels. Then, applying
Operational axioms for diagonalizing states

Proposition 3 of Ref. [6] we obtain that the state space is self-dual—a property that plays an important role in the reconstruction of quantum theory [5].

Another important application of our result is the axiomatic reconstruction of (quantum) thermodynamics. In a previous work [17], we defined a set of thermodynamic transformations in the operational framework, establishing a duality between the resource theory of entanglement and a resource theory of purity defined in terms of random-reversible transformations. A natural application of the diagonalization theorem is the formulation of a majorization criterion capable to detect whether a thermodynamic transition is possible or not, and to establish quantitative measures of mixedness [42, 18]. Such an application contributes also to the difficult problem of finding axioms that imply well-behaved notions of entropy in general probabilistic theories [4, 44, 33] and, therefore, to the development of an axiomatic approach to information theory—specifically including data compression and transmission over noisy channels.

Acknowledgements. We acknowledge P Perinotti for a useful discussion on the fermionic quantum theory of Refs. [22, 23]. This work is supported by the National Basic Research Program of China (973) 2011CBA00300 (2011CBA00301), by the National Natural Science Foundation of China through Grants 11450110096, 11350110207, 61033001, and 61061130540, by the Foundational Questions Institute through the large grant “The fundamental principles of information dynamics”, and by the 1000 Youth Fellowship Program of China. The research by CMS has been supported by a scholarship from “Fondazione Ing. Aldo Gini” and by the Chinese Government Scholarship.

References


Operational axioms for diagonalizing states


