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*

To appear in EPTCS.

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.

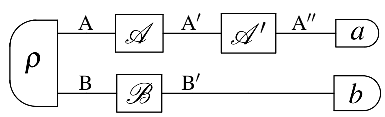
Note that the presence of Purification among the axioms excludes from the start the case of classical probability theory. Indeed, the aim of our work is *not* to provide the most general conditions for the diagonalization of states, but rather to derive diagonalization as a first step towards an axiomatic foundation 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 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 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

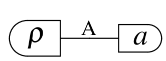


Here, A, A′, A″, B, B′ are *systems*, ρ is a bipartite *state*, $\mathscr{A}$, $\mathscr{A}'$ and $\mathscr{B}$ are *transformations*, a and b are *effects*. Circuits with no external wires, like the one in the above example, are associated with probabilities. We denote by

- $\mathsf{St}(\mathrm{A})$ 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); 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

$$(a|\rho) \;:=$$

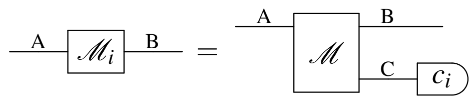


We identify the scalar $(a|\rho)$ with a real number in the interval $[0,1]$, representing the probability of a joint occurrence of the state ρ and the effect a in a circuit. The fact that scalars are real numbers induces a notion of sum for transformations.

A *test* from A to B is a collection of transformations $\{\mathscr{M}_i\}_{i\in\mathsf{X}}$ 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* $\{\mathscr{M}_i\}_{i\in\mathsf{X}}$ *from* A *to* B*, there exist a system* C*, a transformation* $\mathscr{M}\in\mathsf{Transf}\,(\mathrm{A},\mathrm{B}\otimes\mathrm{C})$*, and an observation-test* $\{c_i\}_{i\in\mathsf{X}}$ *such that*



$\forall i\in\mathsf{X}.$

A channel $\mathscr{U}$ from A to B is called *reversible* if there exists another channel $\mathscr{U}^{-1}$ from B to A such that $\mathscr{U}^{-1}\mathscr{U}=\mathscr{I}_{\mathrm{A}}$ and $\mathscr{U}\mathscr{U}^{-1}=\mathscr{I}_{\mathrm{B}}$, where $\mathscr{I}_{\mathrm{S}}$ is the identity channel on a generic system S. A state $\chi\in\mathsf{St}\,(\mathrm{A})$ is called *invariant* if $\mathscr{U}\chi=\chi$, for every reversible channel $\mathscr{U}$. If there exists a reversible channel transforming A into B, we say that A and B are *operationally equivalent*, denoted by $\mathrm{A}\simeq\mathrm{B}$. The composition of systems is required to be *symmetric*, meaning that $\mathrm{A}\otimes\mathrm{B}\simeq\mathrm{B}\otimes\mathrm{A}$.

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

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 $\left\{\mathscr{D}_j\right\}_{j\in\mathsf{Y}}$ 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 $\left\{\mathscr{D}_j\right\}_{j\in\mathsf{Y}}$ 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 $\left\{\mathscr{D}_j\right\}_{j\in\mathsf{Y}}$ is a refinement of $\{\mathscr{C}_i\}_{i\in\mathsf{X}}$, then the transformations $\left\{\mathscr{D}_j\right\}_{j\in\mathsf{Y}_i}$ are a refinement of the transformation $\mathscr{C}_i$.

A transformation $\mathscr{C}\in\mathsf{Transf}(\mathrm{A},\mathrm{B})$ is called *pure* if it has only trivial refinements, namely for every refinement $\left\{\mathscr{D}_j\right\}$ one has $\mathscr{D}_j=p_j\mathscr{C}$, where $\left\{p_j\right\}$ 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 $\mathsf{PurTransf}\,(\mathrm{A},\mathrm{B})$. In the special case of states (resp. effects) of system A we use the notation $\mathsf{PurSt}\,(\mathrm{A})$ (resp. $\mathsf{PurEff}\,(\mathrm{A})$). The set of normalized pure states (resp. effects) of A will be denoted by $\mathsf{PurSt}_1\,(\mathrm{A})$ (resp. $\mathsf{PurEff}_1\,(\mathrm{A})$). As usual, non-pure states are called *mixed*.

Definition 1. Let ρ be a normalized state. We say that a state σ is *compatible* with 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 compatible with a pure state, except the pure state itself.

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

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 Tr_{A} on A. Thanks to that, it is possible to define the *marginal state* of a bipartite state ρ_{AB} on system A as $\rho_{\mathrm{A}} := \mathrm{Tr}_{\mathrm{B}}\rho_{\mathrm{AB}}$. In a theory 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. [13]):

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

The second axiom in our list is Purity Preservation:

Axiom 2 (Purity Preservation[1] [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 $\rho \in \mathsf{St}_1(\mathrm{A})$ can be purified if there exists a pure state $\Psi \in \mathsf{PurSt}(\mathrm{A}\otimes\mathrm{B})$ that has ρ as its marginal on system A. In this case, we call Ψ a *purification* of ρ. 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 $\Psi, \Psi' \in \mathsf{PurSt}_1(\mathrm{A}\otimes\mathrm{B})$ are such that $\mathrm{Tr}_{\mathrm{B}}\Psi_{\mathrm{AB}} = \mathrm{Tr}_{\mathrm{B}}\Psi'_{\mathrm{AB}}$, then $\Psi'_{\mathrm{AB}} = (\mathscr{I}_{\mathrm{A}}\otimes\mathscr{U}_{\mathrm{B}})\Psi_{\mathrm{AB}}$, where $\mathscr{U}_{\mathrm{B}}$ is a reversible channel on B.

In quantum theory, the validity of Purification lies at the foundation of all dilation theorems, such 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* $a \in \mathsf{PurEff}(\mathrm{A})$ *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.

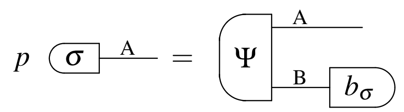
[1] The name and the formulation of the axiom adopted here are the same as in Ref. [16]. The original axiom was called *Atomicity of Composition* [21] and involved only sequential composition. Extending the axiom to parallel composition is important for our purposes, because it guarantees that the product 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 $\rho \in \mathsf{St}_1(\mathrm{A})$ and let $\Psi \in \mathsf{PurSt}_1(\mathrm{A}\otimes\mathrm{B})$ be a purification of ρ. Then σ is compatible with ρ if and only if there exist an effect b_σ on the purifying system* B *and a non-zero probability p such that*



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 $\psi,\psi' \in \mathsf{PurSt}_1(\mathrm{A})$ there exists a reversible channel $\mathscr{U}$ on* A *such that $\psi' = \mathscr{U}\psi$.*

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 χ_{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 $\{a_i\}_{i\in\mathsf{X}}$ on* A*, there is a system* B *and a test $\{\mathscr{A}_i\}_{i\in\mathsf{X}} \subset \mathsf{Transf}(\mathrm{A},\mathrm{B})$ such that every $\mathscr{A}_i$ is pure and $a_i = \mathrm{Tr}_{\mathrm{B}}\mathscr{A}_i$.*

Proposition 5. *Let a be an effect such that $(a|\rho) = 1$, for some $\rho \in \mathsf{St}_1(\mathrm{A})$. Then there exists a transformation $\mathscr{T}$ on* A *such that $a = \mathrm{Tr}\mathscr{T}$ and $\mathscr{T} =_\rho \mathscr{I}$, where $\mathscr{I}$ 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 ρ must be pure (see lemma 26 and theorem 7 of Ref. [14] for the proof idea):

Proposition 6. *Let $a \in \mathsf{PurEff}_1(\mathrm{A})$. Then, there exists a pure state $\alpha \in \mathsf{PurSt}(\mathrm{A})$ such that $(a|\alpha) = 1$. Furthermore, for every $\rho \in \mathsf{St}(\mathrm{A})$, 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 $\alpha \in \mathsf{PurSt}(\mathrm{A})$, there exists at least one pure effect $a \in \mathsf{PurEff}(\mathrm{A})$ 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 6, 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 $(a|\alpha) = 1$. □

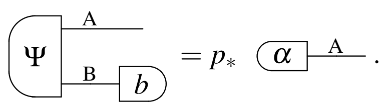
Now we prove an important theorem, which lies at the core of our procedure for diagonalizing states.

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

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

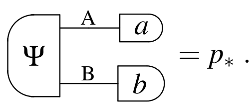
Let Ψ_AB be a purification of ρ_A and let $\widetilde{\rho} \in \mathsf{St}_1(\mathrm{B})$ be the complementary state *of ρ_A, namely $\widetilde{\rho}_\mathrm{B} := \mathrm{Tr}_\mathrm{A}\Psi_\mathrm{AB}$. Then, there exists a pure state $\beta \in \mathsf{PurSt}(\mathrm{B})$ such that $\widetilde{\rho}_\mathrm{B}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

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

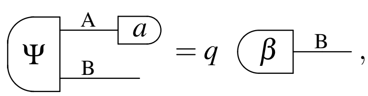


(1)

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

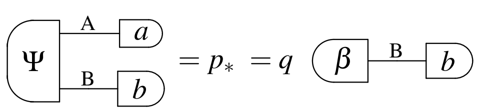


On the other hand, applying a to the state Ψ_AB we obtain

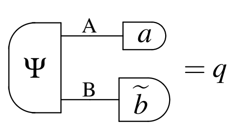


(2)

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

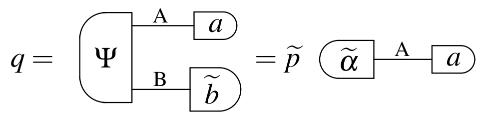


Since $(b|\beta) \in [0,1]$, we must have $q \geq p_*$. We now prove that, in fact, equality holds. Let $\widetilde{b}$ be a pure effect such that $\left(\widetilde{b}|\beta\right) = 1$. Applying $\widetilde{b}$ on both sides of Eq. (2), we obtain



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

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



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 can arise with probability p_* in a convex decomposition of the state $\widetilde{\rho}_{\mathrm{B}}$. □

A similar proof was used in lemma 30 of Ref. [14] 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 6, $\widetilde{\alpha} = \alpha$. This shows that α can always be prepared with probability p_* 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 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 8. *For every normalized pure state α 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 $a, a' \in \mathsf{PurEff}_1(\mathrm{A})$, there exists a reversible channel $\mathscr{U}$ on* A *such that $a' = a\mathscr{U}$.*

5 Diagonalization of states

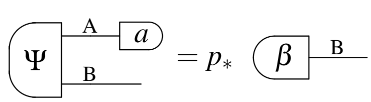
A *diagonalization* of ρ is a convex decomposition of ρ into perfectly distinguishable pure states. .The probabilities in such a convex decomposition will be called the *eigenvalues* of ρ. By Carathéodory's theorem [11, 45], we know that every diagonalization of $\rho \in \mathsf{St}_1(\mathrm{A})$ can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

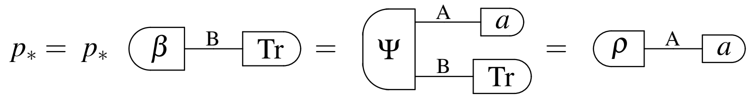
The starting point for diagonalization is the following

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

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



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

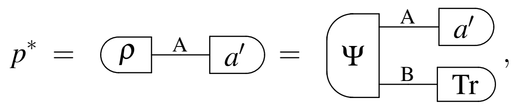


This shows that $(a|\rho) = p_*$. □

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

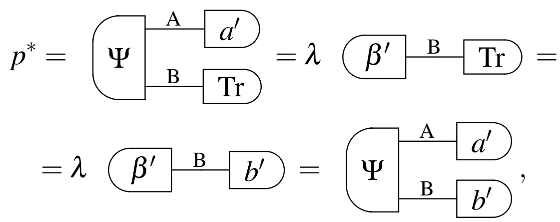
Proposition 10. *Let $\rho \in \mathsf{St}_1(\mathrm{A})$. Define $p^* := \max_{a \in \mathsf{PurEff}_1(\mathrm{A})} (a|\rho)$. Then $p^* = p_*$.*

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

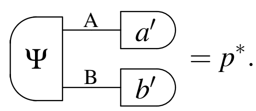


(3)

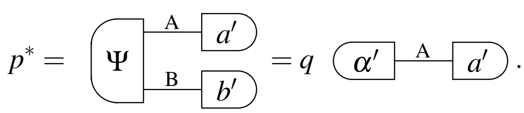
where Ψ_{AB} is a purification of ρ. Now, a' prepares a pure state β' on B with probability λ. If we take the pure effect b' associated with β', we have, recalling Eq. (3),



which means



Now, b' prepares a pure state α' on A with probability q such that



This implies $q \geq p^* > p_*$. Therefore $q > p_*$, and the pure state α' arises in a convex decomposition of ρ with a weight q strictly greater than p_*. This contradicts the fact that p_* is the maximum weight for a pure state, whence in fact $p^* = p_*$. □

The result expressed in proposition 9 has important consequences about diagonalization. Since $(a|\rho) = p_*$, if $\rho = p_*\alpha + (1 - p_*)\sigma$, then $(a|\sigma) = 0$, provided[3] $p_* \neq 1$. Besides, if $(a|\sigma) = 0$, then $(a|\tau) = 0$ for any state τ compatible with σ. 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 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 9, if a is the pure effect associated with α, then $(a|\sigma) = 0$. If σ is pure, then α is perfectly distinguishable from σ by means of the observation-test $\{a, \mathrm{Tr} - a\}$. If σ is mixed, than $(a|\psi) = 0$ for every pure state ψ compatible with σ. Therefore α is perfectly distinguishable from ψ again via the observation-test $\{a, \mathrm{Tr} - a\}$. □

[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

We devise the following procedure for diagonalizing any state ρ.

1. Once we have determined $p_* =: q_1$ 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 σ_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 α_2 is a pure state and σ_2 is another state; and so on.

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

At the end of this procedure, we can write ρ as $\rho = \sum_i p_i\alpha_i$, where $p_1 := q_1$ and $p_i := q_i \prod_{j<i}(1-q_j)$ 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 α_i.

We claim that all the α_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 α_i, $(a_i|\alpha_j) = 0$ for every $j > i$, where a_i is the pure effect associated with α_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 $\{a_1, \mathrm{Tr} - a_1\}$, which perfectly distinguishes α_1 from the other pure states, and we apply the associated test $\left\{\mathscr{A}_1, \mathscr{A}_1^\perp\right\}$, according to proposition 4. If $\mathscr{A}_1$ occurs, we conclude that the state is α_1. If not, the state is one of the others. Consider $\rho_1 = \frac{1}{n-1}\sum_{i=2}^n \alpha_i$. Since $(\mathrm{Tr} - a_1|\rho_1) = 1$, because $(a_1|\alpha_1) = 1$, $\mathscr{A}_1^\perp$ 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 $\{a_2, \mathrm{Tr} - a_2\}$, which perfectly distinguishes α_2 from the remaining states $\{\alpha_i\}_{i=3}^n$, and we apply the associated test $\left\{\mathscr{A}_2, \mathscr{A}_2^\perp\right\}$. If $\mathscr{A}_2$ happens, the state is α_2. If not, we consider $\rho_2 = \frac{1}{n-2}\sum_{i=3}^n \alpha_i$. Since $(\mathrm{Tr} - a_2|\rho_2) = 1$, then $\mathscr{A}_2^\perp$ 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

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

[1] S. Abramsky and B. Coecke. A categorical semantics of quantum protocols. In *Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science*, pages 415–425, July 2004.

[2] S. Abramsky and B. Coecke. Categorical quantum mechanics. In K. Engesser, D. M. Gabbay, and D. Lehmann, editors, *Handbook of Quantum Logic and Quantum Structures: Quantum Logic*, pages 261–324. Elsevier, 2008.

[3] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. Generalized no-broadcasting theorem. *Phys. Rev. Lett.*, 99:240501, Dec 2007.

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

[5] H. Barnum, C. P. Gaebler, and A. Wilce. Ensemble steering, weak self-duality, and the structure of probabilistic theories. *Foundations of Physics*, 43(12):1411–1427, 2013.

[6] H. Barnum, M. P. Müller, and C. Ududec. Higher-order interference and single-system postulates characterizing quantum theory. *New Journal of Physics*, 16(12):123029, 2014.

[7] H. Barnum and A. Wilce. Information processing in convex operational theories. *Electronic Notes in Theoretical Computer Science*, 270(1):3 – 15, 2011. Proceedings of the Joint 5th International Workshop on Quantum Physics and Logic and 4th Workshop on Developments in Computational Models (QPL/DCM 2008).

[8] J. Barrett. Information processing in generalized probabilistic theories. *Phys. Rev. A*, 75:032304, Mar 2007.

[9] F. G. S. L. Brandão and M. Cramer. Equivalence of statistical mechanical ensembles for non-critical quantum systems. *arXiv preprint arXiv:1502.03263*, 2015.

[10] F. G. S. L. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner. The second laws of quantum thermodynamics. *Proceedings of the National Academy of Sciences*, 112(11):3275–3279, 2015.

[11] C. Carathéodory. Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen. *Rendiconti del Circolo Matematico di Palermo*, 32(1):193–217, 1911.

[12] G. Chiribella. Dilation of states and processes in operational-probabilistic theories. In B. Coecke, I. Hasuo, and P. Panangaden, editors, Proceedings 11th workshop on *Quantum Physics and Logic,* Kyoto, Japan, 4-6th June 2014, volume 172 of *Electronic Proceedings in Theoretical Computer Science*, pages 1–14. Open Publishing Association, 2014.

[13] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Probabilistic theories with purification. *Phys. Rev. A*, 81:062348, Jun 2010.

[14] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Informational derivation of quantum theory. *Phys. Rev. A*, 84:012311, Jul 2011.

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

[16] G. Chiribella and C. M. Scandolo. Conservation of information and the foundations of quantum mechanics. *EPJ Web of Conferences*, 95:03003, 2015.

[17] G. Chiribella and C. M. Scandolo. Entanglement and thermodynamics in general probabilistic theories. *arXiv preprint arXiv:1504.07045*, 2015.

[18] G. Chiribella and C. M. Scandolo. Towards an axiomatic foundation of (quantum) thermodynamics. in preparation.

[19] B. Coecke. Quantum picturalism. *Contemporary Physics*, 51:59, 2010.

[20] B. Coecke and É. O. Paquette. Categories for the practising physicist. In B. Coecke, editor, *New Structures for Physics*, volume 813 of *Lecture Notes in Physics*, pages 173–286. Springer Berlin Heidelberg, 2011.

[21] G. M. D'Ariano. Probabilistic theories: what is special about quantum mechanics? In A. Bokulich and G. Jaeger, editors, *Philosophy of Quantum Information and Entanglement*, pages 85–126. Cambridge University Press, Cambridge, 2010.

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

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

[24] P. Faist, F. Dupuis, J. Oppenheim, and R. Renner. A quantitative Landauer's principle. *arXiv preprint arXiv:1211.1037*, 2012.

[25] J. Gemmer, M. Michel, and G. Mahler. *Quantum Thermodynamics: Emergence of Thermodynamic Behavior Within Composite Quantum Systems*, volume 784 of *Lecture Notes in Physics*. Springer Verlag, Heidelberg, 2009.

[26] J. Gemmer, A. Otte, and G. Mahler. Quantum approach to a derivation of the second law of thermodynamics. *Phys. Rev. Lett.*, 86:1927–1930, Mar 2001.

[27] S. Goldstein, J. L. Lebowitz, R. Tumulka, and N. Zanghì. Canonical typicality. *Phys. Rev. Lett.*, 96:050403, Feb 2006.

[28] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Yunger Halpern. The resource theory of informational nonequilibrium in thermodynamics. *arXiv preprint arXiv:1309.6586*, 2013.

[29] L. Hardy. Quantum theory from five reasonable axioms. *arXiv preprint quant-ph/0101012*, 2001.

[30] L. Hardy. Foliable operational structures for general probabilistic theories. In H. Halvorson, editor, *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, pages 409–442. Cambridge University Press, Cambridge, 2011.

[31] L. Hardy. Reformulating and reconstructing quantum theory. *arXiv preprint arXiv:1104.2066*, 2011.

[32] M. Horodecki and J. Oppenheim. Fundamental limitations for quantum and nanoscale thermodynamics. *Nature communications*, 4, 2013.

[33] G. Kimura, K. Nuida, and H. Imai. Distinguishability measures and entropies for general probabilistic theories. *Reports on Mathematical Physics*, 66(2):175 – 206, 2010.

[34] E. Lubkin and T. Lubkin. Average quantal behavior and thermodynamic isolation. *International Journal of Theoretical Physics*, 32(6):933–943, 1993.

[35] M. P. Müller, O. C. O. Dahlsten, and V. Vedral. Unifying typical entanglement and coin tossing: on randomization in probabilistic theories. *Communications in Mathematical Physics*, 316(2):441–487, 2012.

[36] M. P. Müller, D. Gross, and J. Eisert. Concentration of measure for quantum states with a fixed expectation value. *Communications in Mathematical Physics*, 303(3):785–824, 2011.

[37] M. P. Müller, J. Oppenheim, and O. C. O. Dahlsten. The black hole information problem beyond quantum theory. *Journal of High Energy Physics*, 2012(9), 2012.

[38] M. Ozawa. Quantum measuring processes of continuous observables. *Journal of Mathematical Physics*, 25(1):79–87, 1984.

[39] V. Paulsen. *Completely bounded maps and operator algebras*. Cambridge University Press, Cambridge, 2002.

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

[41] S. Popescu, A. J. Short, and A. Winter. Entanglement and the foundations of statistical mechanics. *Nature Physics*, 2(11):754–758, 2006.

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

[43] P. Selinger. A survey of graphical languages for monoidal categories. In B. Coecke, editor, *New Structures for Physics*, pages 289–356. Springer, Berlin, 2011.

[44] A. J. Short and S. Wehner. Entropy in general physical theories. *New Journal of Physics*, 12(3):033023, 2010.

[45] E. Steinitz. Bedingt konvergente Reihen und konvexe Systeme. *J. Reine Angew. Math.*, 143(143):128–175, 1913.

[46] W. F. Stinespring. Positive functions on C*-algebras. *Proceedings of the American Mathematical Society*, 6(2):211–216, 1955.

[47] W. Thirring. *Quantum mathematical physics*. Springer Science & Business Media, 2002.

[48] A. Uhlmann. Sätze über Dichtematrizen. *Wiss. Z. Karl-Marx Univ. Leipzig*, 20:633–653, 1971.

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

[50] A. Uhlmann. Endlich-dimensionale Dichtematrizen II. *Wiss. Z. Karl-Marx-Univ. Leipzig*, 22:139, 1973.