

Convexity and uncertainty in  
operational quantum foundations

Ryo Takakura

# Convexity and uncertainty in operational quantum foundations

A thesis presented by

Ryo Takakura

in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy (Engineering) in the

Department of Nuclear Engineering  
Kyoto University



January 2022

# Abstract

To find the essential nature of quantum theory has been an important problem for not only theoretical interest but also applications to quantum technologies. In those studies on quantum foundations, the notion of uncertainty, which appears in many situations, plays a primary role among several stunning features of quantum theory. The purpose of this thesis is to investigate fundamental aspects of uncertainty. In particular, we address this problem focusing on convexity, which has an operational origin.

We first try to reveal why in quantum theory similar bounds are often obtained for two types of uncertainty relations, namely, preparation and measurement uncertainty relations. In order to do this, we consider uncertainty relations in the most general framework of physics called generalized probabilistic theories (GPTs). It is proven that some geometric structures of states connect those two types of uncertainty relations in GPTs in terms of several expressions such as entropic one. From this result, we can find what is essential for the close relation between those uncertainty relations.

Then, we consider a broader expression of uncertainty in quantum theory called quantum incompatibility. Motivated by an operational intuition, we propose and investigate new quantifications of incompatibility which are related directly to the convexity of states. It is also demonstrated that there can be observed a notable phenomenon for those quantities even in the simplest incompatibility, i.e., incompatibility for a pair of mutually unbiased qubit observables.

Finally, we study thermodynamical entropy of mixing in quantum theory, which also can be seen as a quantification of uncertainty. Similarly to the previous approach, we consider its operationally natural extension to GPTs, and then try to characterize how specific the entropy in quantum theory is. It is shown that the operationally natural entropy is allowed to exist only in classical and quantum-like theories among a class of GPTs called regular polygon theories.

# List of papers

This thesis is based on the following papers:

1. (Reproduced from [1], with the permission of AIP Publishing)  
Ryo Takakura, Takayuki Miyadera, “Preparation Uncertainty Implies Measurement Uncertainty in a Class of Generalized Probabilistic Theories”, *Journal of Mathematical Physics*, **61**, 082203 (2020);
2. ([2])  
Ryo Takakura, Takayuki Miyadera, “Entropic uncertainty relations in a class of generalized probabilistic theories”, *Journal of Physics A: Mathematical and Theoretical*, **54**, 315302 (2021);
3. ([3])  
Teiko Heinosaari, Takayuki Miyadera, Ryo Takakura, “Testing incompatibility of quantum devices with few states”, *Physical Review A*, **104**, 032228 (2021);
4. ([4])  
Ryo Takakura, “Entropy of mixing exists only for classical and quantum-like theories among the regular polygon theories”, *Journal of Physics A: Mathematical and Theoretical*, **52**, 465302 (2019).

# Contents

<b>Abstract</b>	<b>2</b>
<b>List of papers</b>	<b>3</b>
<b>1 Introduction</b>	<b>6</b>
<b>2 Generalized Probabilistic Theories</b>	<b>9</b>
2.1 States and effects . . . . .	10
2.1.1 Axiomatic description . . . . .	11
2.1.2 Convex structures and embedding theorems . . . . .	14
2.1.3 Ordered Banach spaces . . . . .	21
2.1.4 Standard formulations of GPTs . . . . .	29
2.2 Composite systems . . . . .	35
2.3 Transformations . . . . .	41
2.3.1 Channels in GPTs . . . . .	41
2.3.2 Compatibility and incompatibility for channels . . . . .	44
2.4 Additional notions . . . . .	45
2.4.1 Physical equivalence of pure states . . . . .	46
2.4.2 Self-duality . . . . .	48
2.5 Examples of GPTs . . . . .	49
2.5.1 Classical theories with finite levels . . . . .	50
2.5.2 Quantum theories with finite levels . . . . .	50
2.5.3 Regular polygon theories . . . . .	52
<b>3 Preparation uncertainty implies measurement uncertainty in a class of GPTs</b>	<b>54</b>
3.1 Preparation uncertainty and measurement uncertainty in GPTs	55
3.1.1 Widths of probability distributions . . . . .	56
3.1.2 Measurement error . . . . .	57
3.1.3 Relations between preparation uncertainty and mea- surement uncertainty in a class of GPTs . . . . .	61
3.2 Entropic uncertainty relations in a class of GPTs . . . . .	69
3.2.1 Entropic PURs . . . . .	69

3.2.2	Entropic MURs . . . . .	72
3.3	Uncertainty relations in regular polygon theories . . . . .	76
3.3.1	Extensions of previous theorems . . . . .	76
3.3.2	Concrete values for Landau-Pollak-type bounds . . . . .	79
<b>4</b>	<b>Testing incompatibility of quantum devices with few states</b>	<b>84</b>
4.1	(In)compatibility on a subset of states . . . . .	85
4.1.1	(In)compatibility for quantum devices . . . . .	86
4.1.2	(In)compatibility dimension of devices . . . . .	88
4.1.3	Remarks on other formulations of incompatibility dimension . . . . .	93
4.2	Incompatibility dimension and incompatibility witness . . . . .	94
4.2.1	Relation between incompatibility dimension and incompatibility witness for observables . . . . .	95
4.2.2	An upper bound on the incompatibility dimension of observables via incompatibility witness . . . . .	98
4.3	(In)compatibility dimension for mutually unbiased qubit observables . . . . .	101
4.3.1	Proof of Proposition 4.14 : Part 1 . . . . .	103
4.3.2	Proof of Proposition 4.14 : Part 2 . . . . .	105
4.3.3	Proof of Proposition 4.14 : Part 3 . . . . .	110
4.3.4	Proof of Proposition 4.14 : Part 4 . . . . .	119
<b>5</b>	<b>Thermodynamical entropy of mixing in regular polygon theories</b>	<b>126</b>
5.1	Entropy of mixing in GPTs . . . . .	127
5.1.1	Perfect distinguishability for regular polygon theories . . . . .	127
5.1.2	Entropy of mixing in GPTs . . . . .	128
5.2	Main result . . . . .	129
<b>6</b>	<b>Summary</b>	<b>135</b>
	<b>Acknowledgments</b>	<b>137</b>
	<b>Appendix</b>	<b>138</b>
A	Proof of Proposition 2.50 . . . . .	138
B	Proof of Proposition 2.52 . . . . .	141
	<b>Bibliography</b>	<b>149</b>

# Chapter 1

## Introduction

Since its birth about a hundred years ago, quantum theory has been crucial in modern physics because of its more accurate description of nature than classical theory; in addition, it was particularly revealed that there are many differences between the mathematical formulations of classical and quantum theories [5]. Then, it is natural to ask the following questions. What is physically the most significant difference between them? Why is nature described by quantum theory? Since the dawn of quantum theory, they have remained central questions, and much effort has been devoted to finding an answer to form the frontier of physics called quantum foundations [6, 7]. Many significant results have been obtained in that field, and for results of particular importance such as uncertainty relations [8] and the violation of Bell inequality [9, 10], active studies are still ongoing. While studies on quantum foundations are motivated by the theoretical interest of exploring the root of nature, it should be emphasized that pursuing fundamental aspects of quantum theory also contributes to the development of its applications, i.e., quantum technologies. For example, the original ideas of quantum cryptography (quantum key distribution) were derived using uncertainty relations and Bell nonlocality [11, 12]. Quantum foundations are valuable research objects from both theoretical and practical perspectives.

In this thesis, we are engaged in further developing of quantum foundations. To elucidate how “special” quantum theory is, we focus on its convexity. In quantum theory, convexity is one of the most fundamental ingredients, and appears in many situations. A basic example that exhibits convexity is the set of all states (the state space) for some quantum system, which is in fact closed under operationally natural convex combinations [5]. There is one noteworthy approach to quantum foundations concentrating on this primitive convexity, which we call the convexity approach [13]. The main aim of the convexity approach is to find what is needed to derive quantum theory besides the convexity, i.e., to distinguish quantum theory from other convex theories. Its mathematical formulation and physical motiva-

tion are today succeeded to the framework called *generalized probabilistic theories* (GPTs). As was seen above or will be seen in detail in subsequent chapters, GPTs are operationally the broadest framework to describe nature, and have been studied actively in recent years in the context of quantum foundations, followed by the intuition that seeing quantum theory from a broader perspective will contribute to elucidating its essence. While this primitive convexity for states is focused in the study of GPTs, there are studies about quantum foundations based on other types of convexity such as convexity for separable states [14, 15] or compatibility [16, 17]. Considering the above facts, in this thesis we regard convexity as a significant concept for the research on quantum foundations, and demonstrate the results of several attempts to capture the essential nature of quantum theory via convexity. In particular, in this study, we focus on “uncertainty”, which is one of the most critical features in quantum theory, and try to reveal its essence. We have to mention that all results were obtained for *operational convexity*, which means that every type of convexity considered in this thesis has an operational origin. By means of the operational descriptions, our results are easier to understand physically, and thus may contribute more to the theoretical insights of quantum theory and technological applications.

In Chapter 2, we review the mathematical foundations of GPTs. In recent studies, GPTs are usually introduced in a mathematically refined manner such as “the state space is a compact convex set in a finite-dimensional Euclidean space.” There, we try to give a detailed explanation of how those expressions are derived from physically abstract notions. We demonstrate how the operational convexity associated with probability mixtures of states or effects (observables) is expressed in terms of ordered Banach spaces. There are also introduced additional topics for GPTs with physical or mathematical motivations such as the descriptions of composite system and transformations or the notions of transitivity and self-duality.

Based on the mathematical foundations of GPTs, in Chapter 3 we extend the concept of uncertainty relations, which is one of the most astonishing consequences in quantum theory, to GPTs, and investigate how specific the quantum uncertainty is. It is explained that two types of uncertainty, *preparation uncertainty* and *measurement uncertainty*, can also be naturally considered in GPTs, and how they are related is examined under various expressions such as entropic uncertainty relations. Following the quantum results [18, 19], we prove that there is a quantitatively close connection between the two types of uncertainty in GPTs with the assumptions of transitivity and self-duality. We also present numerical evaluations of uncertainty for GPTs called *regular polygon theories* from which we can observe how quantum uncertainty for a single qubit system is specific in regular polygon theories.



In Chapter 4, we focus on another fundamental concept for quantum foundations called *quantum incompatibility*. It is known that many astonishing results in quantum theory, such as the no-cloning theorem [20] and uncertainty relations, are examples of quantum incompatibility [21]. In this way, quantum incompatibility provides such a unified framework to describe what is impossible or what becomes uncertain in quantum theory that it plays an essential role in the field of quantum foundations. Further, we consider the operational convexity of quantum incompatibility, which is derived from that of states and effects. There are introduced new quantifications of incompatibility called *compatibility dimension* and *incompatibility dimension* from a very operational perspective, and properties of those quantities are examined for several cases. In particular, for a pair of incompatible qubit observables, we demonstrate that there is a difference of interest between these quantities. We note that similar quantities can also be defined in GPTs because they are introduced based on the convexity for states and effects, but we only concentrate on quantum incompatibility.

Finally, in Chapter 5, we revisit GPTs, and consider thermodynamical entropy there. We introduce operationally natural entropy which can be defined in every theory of GPTs but is required to satisfy some operational convexity for families of perfectly distinguishable states. Then, it is proven that the only theories that admit the existence of the natural entropy are classical and a quantum-like theories among regular polygon theories.

## Chapter 2

# Generalized Probabilistic Theories

Quantum theory is the most successful theory which describes nature: it does explain phenomena that cannot be recognized if we live in the classical world. The existence of superposition or entanglement is an instance of those remarkable phenomena, but probably the most drastic one is that nature is probabilistic: even if we conduct a “perfect” preparation of a physical system and measurement, we do not always obtain one determined outcome. Generalized probabilistic theories (GPTs) are the framework that focuses on those probabilistic behaviors of nature. The only requirement for GPTs is the convexity for primitive notions of states and effects, and there are in general not assumed any Hilbert space structures or operator algebraic properties. In this sense, GPTs are a more general framework than quantum theory and classical theory, and play an active role in the study of quantum foundations [22, 23, 24, 25, 26, 27, 28, 29, 30, 31] after their initial proposition and development in the 1960s and 1970s [32, 33, 34, 35, 36, 37].<sup>1</sup> In this chapter, we explore the mathematical foundations of GPTs in detail to show how they give the most intuitive and fundamental description of nature.

This chapter is organized as follows. In Section 2.1, we give the two most fundamental notions of GPTs, namely, states and effects. They are introduced in a conceptual and operational way, and mathematically embedded into a vector space and its dual (more generally, a Banach space and its Banach dual) respectively. These embeddings form the mathematical foundations of GPTs. In fact, thanks to this embedding theorem, studies on GPTs usually begin with the assumption that the set of all states called the state space is a compact convex set of a finite-dimensional vector space (more generally, a base of a base norm Banach space). After giving the de-

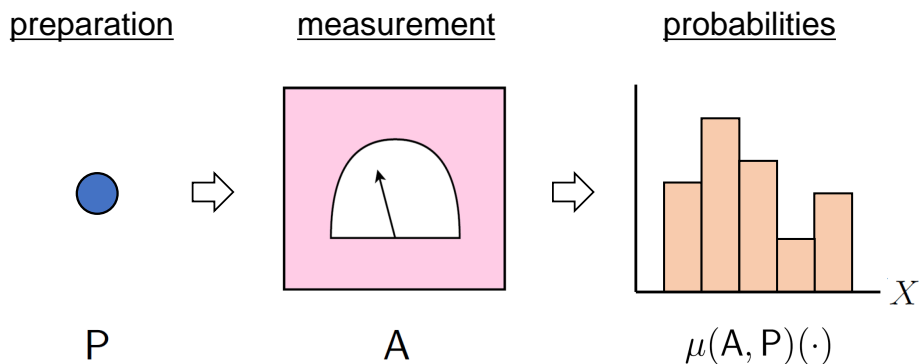
---

<sup>1</sup>For historical review of GPTs, we recommend [30, 38].

descriptions of states and effects, we explain other basic but somewhat more advanced topics, composite systems and transformations in GPTs, in Section 2.2 and Section 2.3. It will be found that the previously introduced embeddings into vector spaces make it mathematically convenient to discuss those concepts. In Section 2.4, we introduce the notions of transitivity and self-duality. These additional notions often appear in the field of GPTs, and our main results in the following chapter are also obtained based on them. In Section 2.5, we illustrate some examples of GPTs which include classical and quantum theories with finite levels, and other important theories often considered in the study of quantum foundations. Throughout this chapter, explicit proofs of mathematical matters are given in principle, but some of them are omitted when they are too technical or lengthy.

## 2.1 States and effects

A physical experiment is described by three procedures: to prepare an object system, to perform a measurement, and to obtain an outcome. However, in general, even if the same preparations are conducted and the same measurements are performed, each outcome obtained is different, and we can only predict from the preparation and measurement how frequently each outcome is obtained, i.e., the probabilities [39, 40, 41, 42].<sup>2</sup> Let us give a concrete description. For a preparation procedure  $P$ , measurement apparatus  $A$ , and a measurable set  $(X, \mathcal{A})$ , where  $X$  is the nonempty set of outcomes associated with  $A$  and  $\mathcal{A}$  is a  $\sigma$ -algebra of subsets of  $X$ , we denote by  $\mu(A, P)(U)$  the probability of obtaining an outcome in  $U \in \mathcal{A}$  when measuring  $A$  on  $P$ . Then, each pair  $(A, U)$  reflects whether a measurement of  $A$  yields a result in the set  $U$  or not. We regard such “yes-no measurements” as a more fundamental notion than the original measurement apparatuses because the latter is an assemblage of the former.



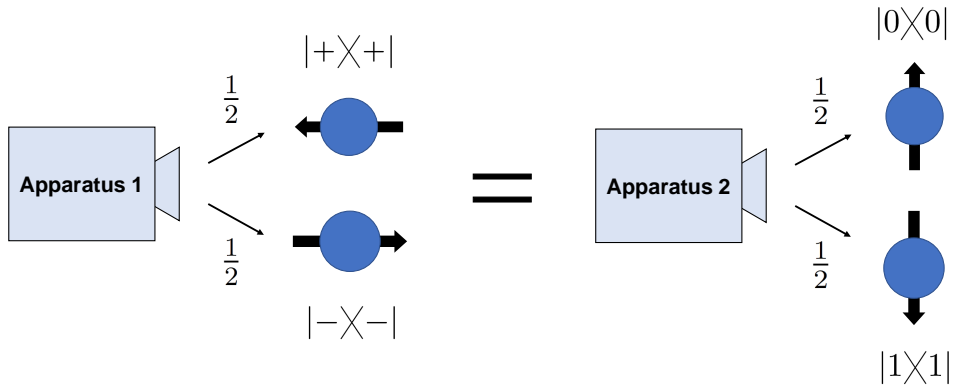
**Figure 2.1:** Description of physical experiments.

<sup>2</sup>In [38], this primitive assumption of physics is called the *statistical causality*.

In this section, we shall demonstrate how to describe two fundamental concepts of physics, preparations and measurements, in mathematical language. As explained above, we focus mainly on yes-no measurements, and write a yes-no measurement and the probability  $\mu(\mathbf{A}, \mathbf{P})(U)$  simply as  $\mathbf{M}$  and  $\mu(\mathbf{M}, \mathbf{P})$  respectively. It will be shown that they are reduced to the notions of *states* and *effects*, and are embedded naturally into some vector space and its dual space respectively. The embedding theorem enables us to treat abstract concepts of preparations and measurements as mathematically well-defined objects, which is the very starting point for GPTs. After their investigations, we will go back to descriptions of general measurement apparatuses to obtain the notion of *observables*. This section is mainly in accord with [30, 31, 41, 43, 44, 45].

### 2.1.1 Axiomatic description

Let  $\text{Prep}$  and  $\text{Meas}$  be the set of all procedures of preparations and yes-no measurements for some physical experiment respectively. For example, in the experiment of detecting the spin of an electron, each element of  $\text{Prep}$  represents an apparatus which emits an electron, and each element of  $\text{Meas}$  represents a value of the meter of some measurement apparatus or the corresponding yes-no apparatus itself. What is specific to this description is that apparatuses with different physical implementations are distinguished. In the previous example, an apparatus that emits randomly (i.e., with probabilities  $\frac{1}{2}$  and  $\frac{1}{2}$ ) electrons with  $x+$  spin and  $x-$  spin, and apparatus which emits randomly electrons with  $z+$  spin and  $z-$  spin are different elements of  $\text{Prep}$ , even though they describe the same quantum state  $\frac{\mathbb{1}}{2}$ . In the field of GPTs, we do not pay attention to those differences of “context” [46] for both preparations and measurements, but only focus on the statistics: if we have two apparatuses that are different but output the same statistics, then we identify those two apparatuses in our framework (see Figure 2.2).



**Figure 2.2:** We identify apparatuses that have different “contexts” but generate the same statistics.

Let us present its mathematical expression. Preparation procedures  $P_1, P_2 \in \text{Prep}$  are called *operationally equivalent* (denoted by  $P_1 \sim P_2$ ) if  $\mu(M, P_1) = \mu(M, P_2)$  holds for all  $M \in \text{Meas}$ . In a similar way, measurement procedures  $M_1, M_2 \in \text{Meas}$  are called *operationally equivalent* (denoted by  $M_1 \sim M_2$ ) if  $\mu(M_1, P) = \mu(M_2, P)$  for all  $P \in \text{Prep}$ . The binary relation  $\sim$  defines an equivalence relation, and thus we can introduce the corresponding quotient sets  $\tilde{\Omega} := \text{Prep}/\sim$  and  $\tilde{\mathcal{E}} := \text{Meas}/\sim$ . These two sets  $\tilde{\Omega}$  and  $\tilde{\mathcal{E}}$  are called the *state space* and *effect space* respectively, and each element of  $\tilde{\Omega}$  and  $\tilde{\mathcal{E}}$  are called a *state* and an *effect* respectively [33, 34, 39, 43]. Here, we express those descriptions above as an axiom.

**Axiom 1 (Separation principle)**

*States and effects separate each other. That is, for any distinct  $\tilde{\omega}_1, \tilde{\omega}_2 \in \tilde{\Omega}$ , there exists an effect  $\tilde{e} \in \tilde{\mathcal{E}}$  such that  $\mu(\tilde{e}, \tilde{\omega}_1) \neq \mu(\tilde{e}, \tilde{\omega}_2)$ , and also, for any distinct  $\tilde{e}_1, \tilde{e}_2 \in \tilde{\mathcal{E}}$ , there exists a state  $\tilde{\omega} \in \tilde{\Omega}$  such that  $\mu(\tilde{e}_1, \tilde{\omega}) \neq \mu(\tilde{e}_2, \tilde{\omega})$ .*

We note that in the statement above we regard the function  $\mu(\cdot, \cdot)$  on  $\text{Meas} \times \text{Prep}$  as on  $\tilde{\mathcal{E}} \times \tilde{\Omega}$  in a well-defined way. States and effects are two primitive notions in GPTs.

Next, we focus on another fundamental concept, *probabilistic mixtures*. It is operationally natural to assume that if we can prepare states  $\tilde{\omega}_1, \tilde{\omega}_2, \dots, \tilde{\omega}_n$ , then we can also prepare a state through the probabilistic mixture of  $\tilde{\omega}_1, \tilde{\omega}_2, \dots, \tilde{\omega}_n$  with respective probabilities  $\lambda_1, \lambda_2, \dots, \lambda_n$ , where  $\lambda_i \geq 0$  and  $\sum_{i=1}^n \lambda_i = 1$ .<sup>3</sup> We denote the newly introduced state by  $\langle \lambda_1, \lambda_2, \dots, \lambda_n; \tilde{\omega}_1, \tilde{\omega}_2, \dots, \tilde{\omega}_n \rangle_{\tilde{\Omega}}$ . The notion of probabilistic mixtures should be considered also for effects, and we denote the effect obtained through the mixture of effects  $\{\tilde{e}_j\}_{j=1}^m \subset \tilde{\mathcal{E}}$  with a probability weight  $\{\sigma_j\}_{j=1}^m$  by  $\langle \sigma_1, \sigma_2, \dots, \sigma_m; \tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_m \rangle_{\tilde{\mathcal{E}}}$ . Then, the nature of probabilistic mixtures motivates us to give the following axiom.

**Axiom 2 (Probabilistic mixtures)**

*For any finite set of states  $\{\tilde{\omega}_i\}_{i=1}^n \subset \tilde{\Omega}$  and probability weight  $\{\lambda_i\}_{i=1}^n$  (i.e.,  $\lambda_i \geq 0$  and  $\sum_i \lambda_i = 1$ ), there exists a state  $\langle \lambda_1, \lambda_2, \dots, \lambda_n; \tilde{\omega}_1, \tilde{\omega}_2, \dots, \tilde{\omega}_n \rangle_{\tilde{\Omega}} \in \tilde{\Omega}$  satisfying*

$$\mu(\tilde{e}, \langle \lambda_1, \lambda_2, \dots, \lambda_n; \tilde{\omega}_1, \tilde{\omega}_2, \dots, \tilde{\omega}_n \rangle_{\tilde{\Omega}}) = \sum_{i=1}^n \lambda_i \mu(\tilde{e}, \tilde{\omega}_i) \quad (2.1)$$

*for all  $\tilde{e} \in \tilde{\mathcal{E}}$ . Similarly, for any finite set of effects  $\{\tilde{e}_j\}_{j=1}^m \subset \tilde{\mathcal{E}}$  and probability weight  $\{\sigma_j\}_{j=1}^m$ , there exists an effect  $\langle \sigma_1, \sigma_2, \dots, \sigma_m; \tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_m \rangle_{\tilde{\mathcal{E}}} \in \tilde{\mathcal{E}}$*

---

<sup>3</sup>From an operational viewpoint, it seems unnatural to consider mixtures with irrational ratios because we can only conduct a finite number of experiments. However, in this thesis, we focus on theories with the completeness assumption (see Mathematical assumption 1), so at this point assume implicitly that mixtures with irrational ratios are admitted for simplicity.

satisfying

$$\mu(\langle\sigma_1, \sigma_2, \dots, \sigma_m; \tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_m\rangle_{\tilde{\mathcal{E}}}, \tilde{\omega}) = \sum_{j=1}^m \sigma_j \mu(\tilde{e}_j, \tilde{\omega}) \quad (2.2)$$

for all  $\tilde{\omega} \in \tilde{\Omega}$ . From Axiom 1, they are uniquely determined.

Axiom 1 and Axiom 2 ensure that, in addition to (2.1), several properties that probabilistic mixtures should satisfy hold successfully for the state  $\langle\lambda_1, \lambda_2, \dots, \lambda_n; \tilde{\omega}_1, \tilde{\omega}_2, \dots, \tilde{\omega}_n\rangle_{\tilde{\Omega}}$ . For example, we can derive easily that

$$\langle\lambda_1, \lambda_2, \dots, \lambda_n; \tilde{\omega}_1, \tilde{\omega}_2, \dots, \tilde{\omega}_n\rangle_{\tilde{\Omega}} = \langle\lambda_2, \lambda_1, \dots, \lambda_n; \tilde{\omega}_2, \tilde{\omega}_1, \dots, \tilde{\omega}_n\rangle_{\tilde{\Omega}}$$

holds, i.e., the mixture does not depend on the “order” of the states and probabilities (similar observations also can be obtained for effects).

We require additional conditions for  $\tilde{\mathcal{E}}$  according to [30, 47, 48]. The first requirement is that  $\tilde{\mathcal{E}}$  includes the *unit effect*  $\tilde{u}$  satisfying  $\mu(\tilde{u}, \tilde{\omega}) = 1$  for all  $\tilde{\omega} \in \tilde{\Omega}$ . In other words, we suppose the existence of a yes-no measurement apparatus that always outputs “yes”, and this seems to be an operationally natural condition. We note that such  $\tilde{u}$  is unique due to Axiom 1. The second one is that if  $\tilde{e}$  is an element of  $\tilde{\mathcal{E}}$ , then the complement effect  $\tilde{e}^\perp$  such that  $\mu(\tilde{e}^\perp, \tilde{\omega}) = 1 - \mu(\tilde{e}, \tilde{\omega})$  for all  $\tilde{\omega} \in \tilde{\Omega}$  is also an element of  $\tilde{\mathcal{E}}$ . This condition comes from the operationally natural intuition that if we admit a certain yes-no measurement apparatus, then we should also admit an apparatus constituted by exchanging the “yes” and “no” of the original one. We note similarly that such  $\tilde{e}^\perp$  is unique. For the complement of the unit effect  $\tilde{u}$ , we sometimes denote it by  $\tilde{0}$  in this thesis. These conditions are summarized as follows.

**Axiom 3 (Existence of unit and complement effects)**

- (i) There exists the unit effect  $\tilde{u}$  in  $\tilde{\mathcal{E}}$  such that  $\mu(\tilde{u}, \tilde{\omega}) = 1$  for all  $\tilde{\omega} \in \tilde{\Omega}$ .
- (ii) If  $\tilde{e} \in \tilde{\mathcal{E}}$ , then its complement  $\tilde{e}^\perp \in \tilde{\mathcal{E}}$  such that  $\mu(\tilde{e}^\perp, \tilde{\omega}) = 1 - \mu(\tilde{e}, \tilde{\omega})$  for all  $\tilde{\omega} \in \tilde{\Omega}$ .

We remark that those effects  $\tilde{u}$  and  $\tilde{e}^\perp$  in Axiom 3 are consistent with Axiom 2.

Now we can give the definition of a GPT.

**Definition 2.1 (Generalized probabilistic theories)**

A triple  $(\tilde{\Omega}, \tilde{\mathcal{E}}, \mu)$  of two sets  $\tilde{\Omega}$  and  $\tilde{\mathcal{E}}$ , and a function  $\mu: \tilde{\Omega} \times \tilde{\mathcal{E}} \rightarrow [0, 1]$  satisfying Axiom 1, 2, 3 is called a *generalized probabilistic theory* (a *GPT* for short). The set  $\tilde{\Omega}$  and its element are called the *state space* and a *state* of the theory, and  $\tilde{\mathcal{E}}$  and its element are called the *effect space* and an *effect* of the theory respectively.

Let us consider infinite countable mixtures for states.<sup>4</sup> In the following, we denote mixtures of two states  $\langle \lambda, 1 - \lambda; \tilde{\omega}_1, \tilde{\omega}_2 \rangle_{\tilde{\Omega}}$  simply by  $\langle \lambda; \tilde{\omega}_1, \tilde{\omega}_2 \rangle$ . In order to treat infinite limits, some topological structure should be introduced into  $\tilde{\Omega}$ . Here we define a topology on  $\tilde{\Omega}$  in line with Gudder [43]. We suppose that if states  $\tilde{\omega}_1$  and  $\tilde{\omega}_2$  are “close”, then

$$\langle \lambda; \tilde{\omega}'_1, \tilde{\omega}_1 \rangle = \langle \lambda; \tilde{\omega}'_2, \tilde{\omega}_2 \rangle$$

with small  $\lambda$  holds for some  $\tilde{\omega}'_1, \tilde{\omega}'_2 \in \tilde{\Omega}$ . That is, the closeness between  $\tilde{\omega}_1$  and  $\tilde{\omega}_2$  should be evaluated by

$$\tilde{d}(\tilde{\omega}_1, \tilde{\omega}_2) := \inf\{0 < \lambda \leq 1 \mid \langle \lambda; \tilde{\omega}'_1, \tilde{\omega}_1 \rangle = \langle \lambda; \tilde{\omega}'_2, \tilde{\omega}_2 \rangle \text{ for some } \tilde{\omega}'_1, \tilde{\omega}'_2 \in \tilde{\Omega}\}. \quad (2.3)$$

We note that (2.3) always can be defined since  $\langle \frac{1}{2}; \tilde{\omega}_2, \tilde{\omega}_1 \rangle = \langle \frac{1}{2}; \tilde{\omega}_1, \tilde{\omega}_2 \rangle$  holds due to Axiom 1. We assume that infinite countable mixtures are allowed in our framework. It is described in the following form.

**Mathematical assumption 1 (Completeness)**

*If  $\tilde{d}$  defined in (2.3) satisfies  $\lim_{n,m \rightarrow \infty} \tilde{d}(\tilde{\omega}_n, \tilde{\omega}_m) = 0$  for a family of states  $\{\tilde{\omega}_n\}_{n=1}^{\infty} \subset \tilde{\Omega}$ , then there exists a unique  $\tilde{\omega} \in \tilde{\Omega}$  such that  $\lim_{n \rightarrow \infty} \tilde{d}(\tilde{\omega}_n, \tilde{\omega}) = 0$ .*

There are two things to remark on Mathematical assumption 1. The first one is about the notion of completeness. In fact, we can prove that the function  $\tilde{d}$  is a metric function on  $\tilde{\Omega}$  (see Subsection 2.1.2), and thus Mathematical assumption 1 is equivalent to the requirement that  $(\tilde{\Omega}, \tilde{d})$  is a complete metric space, which especially admits infinite countable mixtures. The other remark is about the terminology “Mathematical assumption”. In the field of GPTs, the assumption of closedness or completeness for a state space with respect to some physically natural topology is a common one [31]. Namely, if we can prepare states which are very “close” to some fixed state, then it is usually assumed that the fixed state can also be prepared. This seems to be a natural, but at the same time more artificial assumption than the previous ones, so in this thesis, we regard it as a mathematical assumption rather than an axiom.

## 2.1.2 Convex structures and embedding theorems

In the previous section, we presented primitive descriptions of states and effects from a physical perspective. We can rephrase them via the mathematical notion of *convex structures* [36, 43].

---

<sup>4</sup>For infinite countable mixtures of effects, see footnote 17.

**Definition 2.2**

(i) A set  $S$  with a map  $\langle \cdot; \cdot \rangle$  such that

1.  $\langle \lambda_1, \lambda_2, \dots, \lambda_n; s_1, s_2, \dots, s_n \rangle$  defines a unique element of  $S$  for any finite  $s_1, s_2, \dots, s_n \in S$  and probability weight  $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$  (i.e., each  $\lambda_i \geq 0$  and  $\sum_{i=1}^n \lambda_i = 1$ );
2.  $\langle \lambda_1, \lambda_2, \dots, \lambda_n; s, s, \dots, s \rangle = s$

is called a *convex (pre-)structure*. Elements of the form  $\langle \lambda, 1 - \lambda; s, t \rangle$  are denoted simply by  $\langle \lambda; s, t \rangle$ .

(ii) Let  $S$  and  $T$  be convex structures. A map  $F: S \rightarrow T$  is called *affine* if

$$\begin{aligned} F(\langle \lambda_1, \lambda_2, \dots, \lambda_n; s_1, s_2, \dots, s_n \rangle) \\ = \langle \lambda_1, \lambda_2, \dots, \lambda_n; F(s_1), F(s_2), \dots, F(s_n) \rangle, \end{aligned} \quad (2.4)$$

and the set of all affine maps from  $S$  to  $T$  is denoted by  $Aff(S, T)$ . If there exists an affine bijection  $J: S \rightarrow T$ , then  $S$  and  $T$  are called *affinely isomorphic*, and  $J$  is called an *affine isomorphism*.

(iii) Because a convex subset of a vector space is naturally a convex structure with usual convex combinations<sup>5</sup>:  $\langle \lambda_1, \lambda_2, \dots, \lambda_n; s_1, s_2, \dots, s_n \rangle = \sum_{i=1}^n \lambda_i s_i$ , we can define successfully the set  $Aff(S, \mathbb{R})$  for a convex structure  $S$ , and call its element an *affine functional* on  $S$ . In particular, the set of all  $f \in Aff(S, \mathbb{R})$  such that  $f(s) \in [0, 1]$  for all  $s \in S$  is denoted by  $\tilde{\mathcal{E}}_S$ . We regard  $Aff(S, \mathbb{R})$  as a real vector space in a natural way.

(iv) A convex structure  $(S, \langle \cdot; \cdot \rangle)$  is called a *total convex structure* if

1.  $S$  is equipped with a function  $\tilde{d}$  defined in the same way as (2.3), and if  $\lim_{n,m \rightarrow \infty} \tilde{d}(s_n, s_m) = 0$  for a family of elements  $\{s_n\}_n \subset S$ , then there exists a unique  $s \in S$  such that  $\lim_{n \rightarrow \infty} \tilde{d}(s_n, s) = 0$ ;
2.  $f(s) = f(t)$  for every  $f \in \tilde{\mathcal{E}}_S$  implies  $s = t$ .

Let us consider a GPT with a state space  $\tilde{\Omega}$  and effect space  $\tilde{\mathcal{E}}$ . Clearly,  $\tilde{\Omega}$  satisfies conditions (i)-1 and (i)-2 in Definition 2.2, and thus it is a convex structure. On the other hand, it is easy to see that the functional  $\tilde{e}^\circ$  defined for  $\tilde{e} \in \tilde{\mathcal{E}}$  as  $\tilde{e}^\circ: \tilde{\omega} \mapsto \mu(\tilde{e}, \tilde{\omega})$  is an affine functional on  $\tilde{\Omega}$  due to Axiom 2. Because we are interested only in probabilities, it is not problematic to identify the effect  $\tilde{e}$  representing the associated yes-no apparatus with the affine functional  $\tilde{e}^\circ$ , and we also call the latter an *effect*.<sup>6</sup> In other words, if

<sup>5</sup>A subset  $A$  of a vector space  $L$  is called *convex* if  $\lambda x + (1 - \lambda)y \in A$  whenever  $x, y \in A$  and  $\lambda \in (0, 1)$ , and a vector sum  $\sum_{i=1}^n \lambda_i x_i$  for  $x_1, \dots, x_n \in A$  is called a *convex combination* if  $\{\lambda_1, \dots, \lambda_n\}$  is a probability weight. For a more detailed description of convex sets, see [49, 50]

<sup>6</sup>In [38],  $\tilde{e}$  is called an *experimental proposition*, while the term “effect” (also called *experimental function*) is used for the induced affine functional  $\tilde{e}^\circ$ .



we define the map  $\circ: \tilde{e} \mapsto \tilde{e}^\circ$ , then it is an injection from  $\tilde{\mathcal{E}}$  to  $\tilde{\mathcal{E}}_\Omega$  because of Axiom 1, and thus  $\tilde{\mathcal{E}}$  and  $\tilde{\mathcal{E}}^\circ \subset \tilde{\mathcal{E}}_\Omega$  can be identified with each other. Moreover, we can observe from Axiom 2 that the notion of mixtures is represented mathematically as

$$\langle \sigma_1, \sigma_2, \dots, \sigma_m; \tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_m \rangle_{\tilde{\mathcal{E}}}^\circ = \sum_{j=1}^m \sigma_j \tilde{e}_j^\circ, \quad (2.5)$$

and from Axiom 3 that  $\tilde{\mathcal{E}}^\circ$  includes a special effect  $\tilde{u}^\circ$  such that  $\tilde{u}^\circ(\tilde{\omega}) = 1$  for all  $\tilde{\omega} \in \tilde{\Omega}$  and  $\tilde{e}^{\perp \circ} = \tilde{u}^\circ - \tilde{e}^\circ \in \tilde{\mathcal{E}}^\circ$  holds whenever  $\tilde{e}^\circ \in \tilde{\mathcal{E}}^\circ$ . We note that  $\tilde{\mathcal{E}}^\circ$  is a convex subset of the vector space  $\text{Aff}(S, \mathbb{R})$  due to (2.5). In this way, we regard the effect space  $\tilde{\mathcal{E}}^\circ$  as a convex subset of  $\tilde{\mathcal{E}}_\Omega$ :  $\tilde{\mathcal{E}}^\circ \subset \tilde{\mathcal{E}}_\Omega$ . In this thesis, we require that the converse inclusion also holds, which is called the *no-restriction hypothesis* [26].

**Mathematical assumption 2 (No-restriction hypothesis)**

*Any affine functional  $\tilde{e}^\circ$  on  $\tilde{\Omega}$  with  $\tilde{e}^\circ(\tilde{\omega}) \in [0, 1]$  for all  $\tilde{\omega} \in \tilde{\Omega}$  is an effect. That is,  $\tilde{\mathcal{E}}^\circ = \tilde{\mathcal{E}}_\Omega$ .*

The no-restriction hypothesis means that any mathematically valid affine functional is also physically valid. There is no physical background for this assumption, and GPTs without assuming it were investigated for example in [47, 48, 51, 52]. However, in this thesis, we suppose that all theories satisfy the no-restriction hypothesis based on the fact that it is satisfied both in classical and quantum theory. Now we can conclude the following.

**Proposition 2.3**

*A GPT is identified with  $(\tilde{\Omega}, \tilde{\mathcal{E}}_\Omega)$ , where  $\tilde{\Omega}$  is a total convex structure and  $\tilde{\mathcal{E}}_\Omega$  is the set of all affine functionals on it whose values lie in  $[0, 1]$ .*

It is known that a total convex structure can be embedded into a certain Banach space. In order to show this, we need the following lemma.

**Lemma 2.4**

*Let  $(S, \langle \cdot; \cdot \rangle)$  be a total convex structure with a “metric”  $\tilde{d}$  defined in (2.3).*

*(i) If  $\lim_{n \rightarrow \infty} \tilde{d}(s_n, s) = 0$ , then  $\lim_{n \rightarrow \infty} f(s_n) = f(s)$  holds for all  $f \in \tilde{\mathcal{E}}_S$ .*

*(ii) Let  $(T, \langle \cdot; \cdot \rangle_T)$  be another total convex structure equipped with a similar “metric”  $\tilde{d}_T$ . For all  $s_1, s_2 \in S$  and  $F \in \text{Aff}(S, T)$ ,  $\tilde{d}_T(F(s_1), F(s_2)) \leq \tilde{d}(s_1, s_2)$  holds. If  $F$  is bijective, then  $\tilde{d}_T(F(s_1), F(s_2)) = \tilde{d}(s_1, s_2)$ .*

**Proof**

(i) Because  $\lim_{n \rightarrow \infty} \tilde{d}(s_n, s) = 0$  holds, there exists  $N \in \mathbb{N}$  for any  $\varepsilon > 0$  such that  $\tilde{d}(s_n, s) < \frac{\varepsilon}{\varepsilon+2}$  holds whenever  $n > N$ . It implies that there are  $\lambda \in (0, \frac{\varepsilon}{\varepsilon+2})$  and  $t_1, t_2 \in S$  satisfying  $\langle \lambda; t_1, s_n \rangle = \langle \lambda; t_2, s \rangle$ , which results in

$$\lambda f(t_1) + (1 - \lambda)f(s_n) = \lambda f(t_2) + (1 - \lambda)f(s)$$

for  $f \in \tilde{\mathcal{E}}_S$ . It follows that

$$|f(s_n) - f(s)| = \frac{\lambda}{1-\lambda} |f(t_2) - f(t_1)| \leq \frac{2\lambda}{1-\lambda},$$

and thus  $|f(s_n) - f(s)| < \varepsilon$  holds because

$$\frac{2\lambda}{1-\lambda} < \frac{2\lambda}{1-\lambda} \Big|_{\lambda \rightarrow \frac{\varepsilon}{\varepsilon+2}} = \varepsilon.$$

(ii) It holds from the definition of  $\tilde{d}$  that

$$\begin{aligned} \tilde{d}_T(F(s_1), F(s_2)) &= \inf\{0 < \lambda \leq 1 \mid \\ &\quad \langle \lambda; t_1, F(s_1) \rangle_T = \langle \lambda; t_2, F(s_2) \rangle_T, \ t_1, t_2 \in T\} \\ &\leq \inf\{0 < \lambda \leq 1 \mid \\ &\quad \langle \lambda; F(s), F(s_1) \rangle_T = \langle \lambda; F(s'), F(s_2) \rangle_T, \ s, s' \in S\} \\ &= \inf\{0 < \lambda \leq 1 \mid F(\langle \lambda; s, s_1 \rangle) = F(\langle \lambda; s', s_2 \rangle), \ s, s' \in S\} \\ &\leq \inf\{0 < \lambda \leq 1 \mid \langle \lambda; s, s_1 \rangle = \langle \lambda; s', s_2 \rangle, \ s, s' \in S\} \\ &= \tilde{d}(s_1, s_2). \end{aligned}$$

If  $F$  is bijective, then the two “ $\leq$ ” in the above consideration become “ $=$ ”, and thus  $\tilde{d}_T(F(s_1), F(s_2)) = \tilde{d}(s_1, s_2)$  holds.  $\square$

We remember that  $\text{Aff}(S, \mathbb{R})$  is a real vector space for a convex structure  $S$ . The set  $\text{Aff}(S, \mathbb{R})' := \{\alpha \mid \alpha: \text{Aff}(S, \mathbb{R}) \rightarrow \mathbb{R}, \text{ linear}\}$  is naturally a vector space called the *algebraic dual* of  $\text{Aff}(S, \mathbb{R})$ . Then, there is a canonical embedding  $J$  of  $S$  into  $\text{Aff}(S, \mathbb{R})'$  defined as  $[J(s)](f) = f(s)$  for all  $f \in \text{Aff}(S, \mathbb{R})$ . We can prove the following proposition.

**Proposition 2.5**

Let  $(S, \langle \cdot; \cdot \rangle)$  be a total convex structure with a “metric”  $\tilde{d}$  defined in (2.3).

(i) The canonical embedding  $J: S \rightarrow \text{Aff}(S, \mathbb{R})'$  is an affine isomorphism between  $S$  and the convex subset  $J(S)$  of  $\text{Aff}(S, \mathbb{R})'$ .

(ii) If there is an affine isomorphism  $\eta$  between  $S$  and a convex subset  $S_0$  of some real vector space  $V_0$  such that  $\text{aff}(S_0)$  does not include the origin  $0$  of  $V_0$ , then there is a linear bijection  $\Phi: \text{span}(S_0) \rightarrow \text{span}(J(S))$  satisfying  $\Phi(S_0) = J(S)$ .<sup>7</sup>

(iii)  $(S, \tilde{d})$  is a complete metric space.

The claims (i) and (ii) demonstrate that the total convex structure  $S$  can be identified with a convex set in some vector space in an essentially unique way

---

<sup>7</sup>For a subset  $A$  of a vector space  $W$ , its *affine hull*  $\text{aff}(A)$  and *linear span*  $\text{span}(A)$  are defined as  $\text{aff}(A) := \{\sum_{i=1}^n \lambda_i a_i \mid a_i \in A, \lambda_i \in \mathbb{R}, \sum_i \lambda_i = 1, n: \text{finite}\}$  and  $\text{span}(A) := \{\sum_{i=1}^n \lambda_i a_i \mid a_i \in A, \lambda_i \in \mathbb{R}, n: \text{finite}\}$  respectively.

via the canonical embedding  $J$ . We note that the functional  $0 \in \text{Aff}(S, \mathbb{R})'$  defined as  $0(f) = 0$  for all  $f \in \text{Aff}(S, \mathbb{R})$ , which is the origin of the vector space  $\text{Aff}(S, \mathbb{R})'$ , does not belong to  $J(S)$  because  $0 \in J(S)$  contradicts to the existence of the unit effect. On the other hand, the claim (iii) shows that  $\tilde{d}$  is indeed a metric (see Mathematical assumption 1).

**Proof (proof of Proposition 2.5)**

(i) It is easy to see that  $J$  is an affine map from  $S$  to  $\text{Aff}(S, \mathbb{R})'$  and  $J(S)$  is a convex set in  $\text{Aff}(S, \mathbb{R})'$ . Since  $S$  is total (see (iv)-2 in Definition 2.2), for  $s, t \in S$  with  $s \neq t$ , there exists an affine functional  $f \in \text{Aff}(S, \mathbb{R})$  such that  $f(s) \neq f(t)$ , i.e.,  $[J(s)](f) \neq [J(t)](f)$ . This implies  $J(s) \neq J(t)$ .

(ii) Let us introduce a subset  $K := \{\sum_{i=1}^n \lambda_i x_i \mid \lambda_i \geq 0, x_i \in S_0, n : \text{finite}\}$ , i.e., the *conic hull* of  $S_0$  (see Definition 2.6). Then, any  $y \in K \setminus \{0\}$  can be represented as  $y = \lambda x$  with  $\lambda > 0$  and  $x \in S_0$  in a unique way. To see this, assume that  $y \in K \setminus \{0\}$  satisfies  $y = \lambda x = \lambda' x'$  with  $\lambda, \lambda' > 0$  and  $x, x' \in J(S)$ . If  $\lambda \neq \lambda'$ , then it holds that

$$0 = \lambda x - \lambda' x' = (\lambda - \lambda') \left( \frac{\lambda}{\lambda - \lambda'} x - \frac{\lambda'}{\lambda - \lambda'} x' \right).$$

Because  $0 \notin \text{aff}(S_0)$ , it implies  $\lambda = \lambda'$ , which is a contradiction. Thus, we can conclude  $\lambda = \lambda'$  and  $x = x'$ . Now let us construct the linear bijection  $\Phi$  from the affine isomorphism  $\eta$ . First, we define an affine bijection  $\phi_0: S_0 \rightarrow J(S)$  by  $\phi = J \circ \eta^{-1}$  (note that  $J$  is a bijection between  $S$  and  $J(S)$ ). From the above consideration, we can extend this  $\phi_0$  successfully to a bijection  $\phi$  from  $K$  to the conic hull of  $J(S)$ :  $\phi(y) = \lambda \phi_0(x)$  for  $y = \lambda x$  with  $y \in K, x \in S_0$ , and  $\lambda \geq 0$ . It is easy to verify that  $\phi(\lambda y + \mu z) = \lambda \phi(y) + \mu \phi(z)$  holds for  $y, z \in K$  and  $\lambda, \mu \geq 0$ . Since any  $y \in \text{span}(S_0)$  such that  $y = \sum_{i=1}^n \lambda_i x_i$  with  $x_i \in S_0, \lambda_i \in \mathbb{R}$ , and a finite  $n$  can be expressed as  $y = u - v$ , where  $u, v \in K$ , we can consider the extension of  $\phi$  to a map  $\Phi$  from  $\text{span}(S_0)$  to  $\text{span}(J(S))$  by  $\Phi(y) = \phi(u) - \phi(v)$  for  $y = u - v$  with  $y \in \text{span}(S_0)$  and  $u, v \in K$ . We note that this  $\Phi$  is well-defined: if  $y = u_1 - v_1 = u_2 - v_2$  with  $u_1, u_2, v_1, v_2 \in K$  holds, then  $u_1 + v_2 = u_2 + v_1$  holds, and thus  $\Phi(u_1 + v_2) = \Phi(u_2 + v_1)$ , i.e.,  $\Phi(u_1) + \Phi(v_2) = \Phi(u_2) + \Phi(v_1)$  follows, which implies  $\Phi(u_1) - \Phi(u_2) = \Phi(v_1) - \Phi(v_2)$ . It is easy to confirm that  $\Phi: \text{span}(S_0) \rightarrow \text{span}(J(S))$  is linear and bijective.

(iii) It is trivial that  $\tilde{d}(s, t) \geq 0$  and  $\tilde{d}(s, t) = \tilde{d}(t, s)$  holds for all  $s, t \in S$ . Let  $\tilde{d}(s, t) = 0$ . Then, there exist a family of positive numbers  $\{\lambda_i\}_i$  with  $\lim_{i \rightarrow \infty} \lambda_i = 0$  and families  $\{s_i\}_i$  and  $\{t_i\}_i$  of elements of  $S$  such that  $\langle \lambda_i; s_i, s \rangle = \langle \lambda_i; t_i, t \rangle$ . It follows that

$$\lambda_i f(s_i) + (1 - \lambda_i) f(s) = \lambda_i f(t_i) + (1 - \lambda_i) f(t)$$

holds for all  $f \in \tilde{\mathcal{E}}_S$ . Because  $0 \leq f(s_i), f(t_i) \leq 1$  holds, taking  $i \rightarrow \infty$  in the

above equation, we obtain  $f(s) = f(t)$  for all  $f \in \tilde{\mathcal{E}}_S$ . By the assumption of totality, we can conclude  $s = t$ . To verify the triangle inequality for  $\tilde{d}$ , it is enough to prove that  $\tilde{d}' : J(S) \times J(S) \rightarrow \mathbb{R}$  defined on  $J(S)$  in a similar way to  $\tilde{d}$  satisfies it. This is because  $\tilde{d}'(J(s), J(t)) = \tilde{d}(s, t)$  holds for all  $s, t \in S$  as we have seen in Lemma 2.4. For the evaluation of  $\tilde{d}'(p, r) + \tilde{d}'(r, q)$  with  $p, q, r \in J(S)$ , let us assume that  $\lambda_1, \lambda_2 \in (0, 1)$  satisfy

$$\begin{aligned}\lambda_1 p_1 + (1 - \lambda_1)p &= \lambda_1 r_1 + (1 - \lambda_1)r, \\ \lambda_2 r_2 + (1 - \lambda_2)r &= \lambda_2 q_1 + (1 - \lambda_2)q\end{aligned}$$

for  $p_1, q_1, r_1, r_2 \in J(S)$ . We obtain from these equations

$$\begin{aligned}\lambda_1(1 - \lambda_2)p_1 + \lambda_2(1 - \lambda_1)r_2 + (1 - \lambda_1)(1 - \lambda_2)p \\ = \lambda_2(1 - \lambda_1)q_1 + \lambda_1(1 - \lambda_2)r_1 + (1 - \lambda_1)(1 - \lambda_2)q.\end{aligned}$$

It can be rewritten as

$$\lambda_0 p_2 + (1 - \lambda_0)p = \lambda_0 q_2 + (1 - \lambda_0)q, \quad (2.6)$$

where

$$\lambda_0 = \frac{\lambda_1(1 - \lambda_2) + \lambda_2(1 - \lambda_1)}{\lambda_1(1 - \lambda_2) + \lambda_2(1 - \lambda_1) + (1 - \lambda_1)(1 - \lambda_2)} = \frac{\lambda_1 + \lambda_2 - 2\lambda_1\lambda_2}{1 - \lambda_1\lambda_2}$$

and

$$\begin{aligned}p_2 &= \frac{\lambda_1(1 - \lambda_2)}{\lambda_1(1 - \lambda_2) + \lambda_2(1 - \lambda_1)}p_1 + \frac{\lambda_2(1 - \lambda_1)}{\lambda_1(1 - \lambda_2) + \lambda_2(1 - \lambda_1)}r_2, \\ q_2 &= \frac{\lambda_2(1 - \lambda_1)}{\lambda_1(1 - \lambda_2) + \lambda_2(1 - \lambda_1)}q_1 + \frac{\lambda_1(1 - \lambda_2)}{\lambda_1(1 - \lambda_2) + \lambda_2(1 - \lambda_1)}r_1.\end{aligned}$$

Because  $\lambda_0 \leq \lambda_1 + \lambda_2$ , we can see from (2.6) that

$$\tilde{d}'(p, q) \leq \tilde{d}'(p, r) + \tilde{d}'(r, q)$$

holds, and thus we can conclude that  $(S, \tilde{d})$  is a metric space. The completeness clearly holds due to (iv)-1 in Definition 2.2.  $\square$

We note that we can prove the same claim as (iii) also for the function  $\tilde{d}_0$  defined as

$$\tilde{d}_0(s, t) = \frac{\tilde{d}(s, t)}{1 - \tilde{d}(s, t)}. \quad (2.7)$$

It was shown in [36] that this  $\tilde{d}_0$  is a metric on  $S$ , and the completeness holds similarly. Before proceeding to the main theorem of this section, we

introduce the notion of convex cones [50, 53, 54].

**Definition 2.6**

Let  $L$  be a vector space, and  $0 \in L$  be its origin.

(i) A subset  $C$  of  $L$  is called a *cone* of vertex  $0$  if  $\lambda C \subset C$  for all  $\lambda > 0$ . A cone of vertex  $x_0$  is a set of the form  $x_0 + C$ , where  $C$  is a cone of vertex  $0$ . In this thesis, the vertex of a cone is always assumed to be  $0$ .

(ii) A cone  $C \subset L$  is called

1. *convex* if it is convex, i.e., satisfies  $C + C \subset C$ ;
2. *pointed* if  $C \cap -C = \{0\}$ ;
3. *generating* (or *spanning*) if  $\text{span}(C) = L$ , i.e.,  $C - C = L$ .

(iii) The *conic hull* of a subset  $A$  of  $L$  is defined as  $\text{cone}(A) := \{\sum_{i=1}^n \lambda_i a_i \mid \lambda_i \geq 0, a_i \in A, n : \text{finite}\}$ . It is easy to see that  $\text{cone}(A)$  is a convex cone.

Let us write  $\text{cone}(J(S))$  and  $\text{span}(J(S))$  generated by  $J(S)$  simply as  $K$  and  $V$  respectively. It is easy to see that  $K$  is a convex, pointed, and generating cone for  $V$ , and thus any  $v \in V$  is written in the form  $v = k_+ - k_- = \alpha p - \beta q$ , where  $k_{\pm} \in K$ ,  $p, q \in J(S)$ , and  $\alpha, \beta \geq 0$ . It follows that we can introduce the following quantity for  $v \in V$ :

$$\|v\| = \inf\{\alpha + \beta \mid v = \alpha p - \beta q, \alpha, \beta \geq 0, p, q \in J(S)\}. \quad (2.8)$$

Now we can present the embedding theorem for a total convex structure as follows. We shall omit the proof, but it is given in [43] (see the proofs of Theorem 4.11 and Theorem 4.12 there).

**Theorem 2.7**

*Let  $S$  be a total convex structure, and  $K$  and  $V$  be the cone and the real vector space generated by the canonical embedding  $J(S)$  of  $S$  into  $\text{Aff}(S, \mathbb{R})'$  respectively.*

*(i) The function  $\|\cdot\|$  on  $V$  defined in (2.8) is a norm on  $V$  satisfying  $\|J(s) - J(t)\| = 2\tilde{d}_0(s, t)$  for all  $s, t \in S$  and  $\|J(s)\| = 1$  for all  $s \in S$ . Moreover,  $(V, \|\cdot\|)$  is a real Banach space, and  $K$  is closed.*

*(ii) Let  $f \in \text{Aff}(S, \mathbb{R})$ . Then, the affine functional  $f \circ J^{-1}: J(S) \rightarrow \mathbb{R}$  on  $J(S)$  has a unique linear extension  $\check{f}: V \rightarrow \mathbb{R}$ .*

*(iii) If we let  $e: V \rightarrow \mathbb{R}$  be the unique linear extension of  $\tilde{e}^\circ \in \tilde{\mathcal{E}}_S \subset \text{Aff}(S, \mathbb{R})$  described in (ii) above, then  $e$  is continuous, and thus belongs to the Banach dual  $V^* := \{f \mid f: V \rightarrow \mathbb{R}, \text{ linear, bounded (continuous)}\}$  of  $V$ . In particular, the linear extension  $u$  of the unit effect  $\tilde{u}^\circ$  such that  $u(J(s)) = 1$  for all  $J(s) \in J(S)$  satisfies  $u \in V^*$ .*

According to this theorem, together with Proposition 2.3, we can consider the state space of a GPT  $\tilde{\Omega}$  as a closed convex set  $\Omega := J(\tilde{\Omega})$  in a Banach space  $V = \text{span}(\Omega)$  equipped with the norm  $\|\cdot\|$  in (2.8) called the *base norm*, and the effect space  $\tilde{\mathcal{E}}_{\tilde{\Omega}}$  as a subset  $\mathcal{E}_{\Omega} := \{e \mid e(\omega) \in [0, 1] \text{ for all } \omega \in \Omega, \text{ linear}\}$  of the Banach dual  $V^*$ . We also call  $\Omega$  and  $\mathcal{E}_{\Omega}$  the *state space* and the *effect space* of the GPT respectively. In the next part, we give further explanations about the Banach space  $V$  and its Banach dual  $V^*$ .

### 2.1.3 Ordered Banach spaces

The vector spaces  $V$  and  $V^*$  introduced in the previous part are equipped with both order and Banach space structures, that is, they are ordered Banach spaces. In this subsection, we make a brief review of ordered Banach spaces. Mathematical terms shown in this subsection are according mainly to [30, 31, 44, 50, 53, 55, 56]. Also, there can be found the technical proofs of some theorems which we omit. We begin with the definition of an ordered vector space.

#### Definition 2.8

A real vector space  $L$  equipped with a partial ordering<sup>8</sup>  $\leq$  is called an *ordered vector space* if it satisfies

- (i)  $x \leq y$  implies  $x + z \leq y + z$  for all  $x, y, z \in L$ ;
- (ii)  $x \leq y$  implies  $\lambda x \leq \lambda y$  for all  $x, y \in L$  and  $\lambda \geq 0$ .

We can prove easily the following (recall Definition 2.6).

#### Proposition 2.9

- (i)  $L_+ := \{x \in L \mid x \geq 0\} \subset L$  is a convex and pointed cone.
- (ii) If  $(L, \leq)$  is directed, i.e., for every  $x, y \in L$  there is  $z \in L$  such that  $x \leq z, y \leq z$ , then  $L_+$  in (i) is also generating.

#### Proof

(i) For  $x \geq 0, \lambda x \geq 0$  ( $\lambda \geq 0$ ) clearly holds, and thus  $L_+$  is a cone. Because, for  $x, y \geq 0, px \geq 0$  and  $(1-p)y \geq 0$  ( $0 \leq p \leq 1$ ) hold,  $px + (1-p)y \geq 0$  follows, which implies  $L_+$  is convex. The claim that  $L_+$  is pointed follows from the observation that  $x \geq 0$  and  $x \leq 0$  implies  $x = 0$ .

(ii) Because  $L$  is directed, for any  $x \in L$ , there exists  $z \in L$  such that  $x \leq z$  and  $-x \leq z$ , equivalently,  $z-x \geq 0$  and  $z+x \geq 0$  hold. Because  $\frac{1}{2}(z-x) \geq 0$  and  $\frac{1}{2}(z+x) \geq 0, x = \frac{1}{2}(z+x) - \frac{1}{2}(z-x)$  implies that  $L_+$  is generating.  $\square$

---

<sup>8</sup>A binary relation  $\leq$  on a set  $X$  is called a *preorder* if it is reflexive, i.e.,  $x \leq x$  ( $x \in X$ ), and transitive, i.e.,  $x \leq y$  and  $y \leq z$  implies  $x \leq z$  ( $x, y, z \in X$ ). A preorder  $\leq$  is called a *partial order* if it is antisymmetric, i.e.,  $x \leq y$  and  $y \leq x$  implies  $x = y$  ( $x, y \in X$ ). We remark that some authors use the term “partial order” to represent a preorder here [57].

**Definition 2.10**

Let  $L$  be an ordered vector space.

- (i) The cone  $L_+ := \{x \in L \mid x \geq 0\}$  is called the *positive cone* of  $L$ .
- (ii) For the positive cone  $L_+$  of  $L$ , its *order dual cone*  $L_+^\diamond$  is defined as the set of all “positive” functionals on  $L_+$ , i.e.,  $L_+^\diamond := \{f \in L' \mid f(x) \geq 0 \text{ for all } x \in L_+\}$ . It is clear that  $L_+^\diamond$  is a convex cone in the algebraic dual  $L'$  of  $L$  and in the subspace  $L^\diamond := L_+^\diamond - L_+^\diamond = \text{span}(L_+^\diamond)$  called the *order dual* of  $L$ . Moreover, we can find that  $L_+^\diamond$  is pointed in  $L'$  and  $L^\diamond$  if  $L_+$  is generating.

We have proven in Proposition 2.9 that a positive cone can be introduced through an order vector space. Conversely, we can construct an order structure for a vector space when there is a convex cone.

**Proposition 2.11**

Let  $C$  be a convex and pointed cone in a real vector space  $L$ .

(i) If we define a binary relation  $\leq$  as  $x \leq y \iff y - x \in C$  for  $x, y \in V$ , then the relation  $\leq$  is a partial ordering, and  $(L, \leq)$  is an ordered vector space.

(ii) The positive cone  $L_+$  for  $L$  defined via the order  $\leq$  in (i) is identical to  $C$ , i.e.,  $L_+ = C$ .

(iii) If  $C$  is in addition generating, then  $(L, \leq)$  is directed.

**Proof**

(i) Because  $C$  is pointed,  $x - x = 0 \in C$ , and  $y - x \in C$  and  $x - y \in C$  imply  $y - x = 0$ , i.e.,  $x = y$  for  $x, y \in L$ . Moreover, if  $y - x \in C$  and  $z - y \in C$  ( $z, y, z \in L$ ), then  $z - x = (z - y) + (y - x) \in C$ . Therefore, we can conclude that  $\leq$  is a partial ordering. On the other hand, because  $y - x = (y + z) - (x + z)$  ( $z, y, z \in L$ ),  $x + z \leq y + z$  holds when  $x \leq y$ . Since  $C$  is a pointed cone,  $y - x \in C$  ( $x, y \in L$ ) implies  $\lambda(y - x) = \lambda y - \lambda x \in C$  ( $\lambda \geq 0$ ), i.e.,  $\lambda x \leq \lambda y$  when  $x \leq y$ .

(ii)  $L_+ = C$  is trivial since  $x \geq 0$  is equivalent to  $x \in C$ .

(iii) For  $x, y \in L$ , because  $C$  is generating, there exist  $x_1, x_2, y_1, y_2 \in C$  such that  $x = x_1 - x_2$  and  $y = y_1 - y_2$ . Defining  $z = x_1 + y_1$ , we have  $z - x = y_1 + x_2 \in C$  and  $z - y = x_1 + y_2 \in C$ , which means that  $(L, \leq)$  is directed.  $\square$

It follows from these propositions that a positive cone and a convex and pointed cone can be identified naturally with each other.

Next, we give descriptions of ordered Banach spaces. An ordered vector space  $L$  is called an ordered Banach space if  $L$  is also a Banach space (see [58] for a review of Banach space). There are two important types of ordered Banach space in the field of GPTs: *base norm Banach spaces* and *order unit Banach spaces*, which are related with state spaces and effect spaces respectively. Let us first introduce base norm Banach spaces.

**Definition 2.12**

Let  $L$  be an ordered vector space with its positive cone  $L_+$ . A convex subset  $B \subset L_+$  is called a *base* of  $L_+$  if for any  $x \in L_+$  there exists a unique  $\lambda \geq 0$  such that  $x \in \lambda B$ .

The following lemma is important.

**Lemma 2.13**

Let  $L$  be an ordered vector space with its positive cone  $L_+$ , and  $B$  be its base. Then,  $\text{aff}(B)$  does not contain the origin  $0$  of  $L$ .

**Proof**

Suppose  $0 \in \text{aff}(B)$ . Then, there exist real numbers  $\{\lambda_i\}_{i=1}^n$  with  $\sum_{i=1}^n \lambda_i = 1$  and elements  $\{x_i\}_{i=1}^n$  of  $B$  such that  $\sum_{i=1}^n \lambda_i x_i = 0$ . Dividing  $\{\lambda_i\}_{i=1}^n$  into positive and negative parts, we obtain

$$\sum_j \lambda_j^+ x_j^+ = \sum_k \lambda_k^- x_k^-,$$

where  $\{x_j^+\}_j$  and  $\{x_k^-\}_k$  are subsets of  $\{x_i\}_{i=1}^n$ , and  $\{\lambda_j^+\}_j$  and  $\{\lambda_k^-\}_k$  are positive numbers satisfying  $\sum_j \lambda_j^+ - \sum_k \lambda_k^- = 1$ . If we assume  $K := \sum_k \lambda_k^- \neq 0$ , then we can rewrite the above equation as

$$\frac{K+1}{K} \cdot \frac{1}{K+1} \sum_j \lambda_j^+ x_j^+ = \frac{1}{K} \sum_k \lambda_k^- x_k^-.$$

Because  $y := \frac{1}{K+1} \sum_j \lambda_j^+ x_j^+$  and  $y' := \frac{1}{K} \sum_k \lambda_k^- x_k^-$  are convex combinations of elements of  $B$ , they belong to  $B$ . Then, the above equation  $\frac{K+1}{K}y = y'$  contradicts to the uniqueness condition in the definition of the base  $B$ , and thus we obtain  $K = 0$ . It implies  $0 \in B$ , but this also contradicts to the uniqueness condition because any positive number  $\lambda$  satisfy  $\lambda 0 = 0$ .  $\square$

By means of this lemma, we can associate a base of a positive cone with a linear functional in the following way [30, 56].

**Proposition 2.14**

Let  $L$  and  $L_+$  be an ordered vector space and its positive cone respectively.  $L_+$  has a base  $B$  if and only if there exists a strictly positive functional  $e_B$  (i.e.,  $e_B \in L^\diamond$  and satisfies  $e_B(x) > 0$  for all nonzero  $x \in L_+$ ) such that

$$B = \{x \in L_+ \mid e_B(x) = 1\}. \quad (2.9)$$

**Proof**

The if part is easy, so we prove the only if part. Let  $B$  be a base of  $L_+$ . Applying Zorn's lemma to the set  $A$  of all affine sets which include  $\text{aff}(B)$  but not  $\{0\}$ , we obtain the maximal affine set  $H$  in  $A$ . It can be shown [59]



that this  $H$  is a hyperplane in  $L$ , and thus there exists a linear functional  $e_B$  such that  $e_B(x) = 1$  for all  $x \in H$ . This functional  $e_B$  is easily found to be strictly positive because  $B$  is a base.  $\square$

We call the functional  $e_B$  the *intensity functional* for the base  $B$  [38].

**Lemma 2.15**

Let  $L$  be an ordered vector space, and  $L_+$  be its positive cone which is generating. For a base  $B \subset L_+$  of  $L_+$ ,  $D := \text{conv}(B \cup -B)$  is a radial, circled, and convex subset of  $L$ .<sup>9</sup>

**Proof**

The convexity is clear. It is easy to see  $0 \in D$ , and thus  $D$  is circled. Because  $L_+$  is generating, any  $x \in L$  can be written as  $x = \lambda_+x_+ + \lambda_-x_-$  with  $\lambda_{\pm} \geq 0$  and  $x_+ \in B, x_- \in -B$ . Let  $\lambda_0 = \lambda_+ + \lambda_-$ . For  $\lambda \geq \lambda_0$ ,  $x$  can be rewritten as

$$x = \lambda \cdot \frac{\lambda_+ + \lambda_-}{\lambda} \left( \frac{\lambda_+}{\lambda_+ + \lambda_-}x_+ + \frac{\lambda_-}{\lambda_+ + \lambda_-}x_- \right).$$

Because  $D$  is circled,  $\frac{\lambda_+ + \lambda_-}{\lambda} \left( \frac{\lambda_+}{\lambda_+ + \lambda_-}x_+ + \frac{\lambda_-}{\lambda_+ + \lambda_-}x_- \right) \in D$  can be obtained. It implies  $x \in \lambda D$ , and thus  $D$  is radial.  $\square$

According to Lemma 2.15, if  $L_+$  is generating, then the Minkowski functional of  $D = \text{conv}(B \cup -B)$  defined as

$$p_D(x) := \inf\{\lambda > 0 \mid x \in \lambda D\} \quad (x \in L) \quad (2.10)$$

is a seminorm on  $L$  [50]. It is not difficult to see that  $p_D$  satisfies

$$p_D(x) = \inf\{e_B(x_+) + e_B(x_-) \mid x = x_+ - x_-, x_{\pm} \in L_+\} \quad (x \in L) \quad (2.11)$$

with  $e_B$  introduced in Proposition 2.14, or

$$p_D(x) = \inf\{\alpha + \beta \mid x = \alpha b_+ - \beta b_-, \alpha, \beta \geq 0, b_{\pm} \in B\} \quad (x \in L) \quad (2.12)$$

since it holds that  $p_D(x_+) = e_B(x_+)$  for all  $x_+ \in L_+$ . Now we can give the definition of a base norm space.

**Definition 2.16**

Let  $L$  be an ordered vector space with its positive cone  $L_+$  generating. For a base  $B$  of  $L_+$ ,  $(L, B)$  is called a *base norm space* if the function  $p_D$  defined in (2.10)-(2.12) through  $B$  is a norm on  $L$ .<sup>10</sup> In this case, we write  $p_D(\cdot)$  as

<sup>9</sup>A subset  $U$  of a vector space  $L$  (assumed to be on the field  $F = \mathbb{R}$  or  $\mathbb{C}$ ) is *radial* if for any  $x \in L$  there exists  $\lambda_0 \in F$  such that  $|\lambda| \geq |\lambda_0|$  implies  $x \in \lambda U$ , and is *circled* if  $\lambda U \subset U$  for any  $\lambda$  with  $|\lambda| \leq 1$  [50].

<sup>10</sup>It is not difficult to see that  $p_D$  is a norm if and only if  $D = \text{conv}(B \cup -B)$  is linearly bounded, i.e.,  $M \cap D$  is a bounded subset of  $L$  whenever  $M$  is a one-dimensional subspace [56].

$\|\cdot\|_B$  and call it the *base norm*. A base norm space  $(L, B)$  is called a *base norm Banach space* if  $L$  is complete with respect to the base norm  $\|\cdot\|_B$ .

In this thesis, for a Banach space  $X$ , we denote its *Banach dual* by  $X^* = \{f \mid f: X \rightarrow \mathbb{R}, \text{ linear, bounded}\}$ . When  $X$  is in addition an ordered vector space (i.e., an *ordered Banach space*) and  $X_+$  is its positive cone, we define a set  $X_+^*$  as  $X_+^* := \{f \in X^* \mid f(x) \geq 0 \text{ for all } x \in X_+\}$ , and call it the *Banach dual cone* for  $X_+$ . It is verified easily that  $L_+^*$  is a convex and closed (in the weak<sup>\*</sup><sup>11</sup> and norm topologies) cone in  $L^{*12}$ , and is in addition pointed if  $L_+$  is generating.

We present miscellaneous facts about base norm Banach spaces.

### Proposition 2.17

Let  $(L, B)$  be a base norm Banach space, and  $L_+$  be the positive cone of  $L$ . For a subset  $A$  of  $L$ , we denote its norm closure by  $\overline{A}$ .

- (i) The intensity functional  $e_B$  in (2.9) is continuous, i.e.,  $e_B \in L^*$ .
- (ii)  $B$  is closed if and only if  $L_+$  is closed.
- (iii) The unit ball of  $L$  is given by  $\overline{D} = \overline{\text{conv}(B \cup -B)}$ .
- (iv) The dual norm  $\|\cdot\|_*$  on the Banach dual  $L^*$  defined as  $\|f\|_* := \sup\{|f(x)| \mid \|x\|_B \leq 1\}$  satisfies  $\|f\|_* := \sup\{|f(x)| \mid x \in B\}$ .
- (v)  $\overline{L_+}$  is a convex, pointed, and generating cone in  $L$ , and  $\overline{B}$  is a base of  $\overline{L_+}$  with its intensity functional identical with that of the original base  $B$ :  $\overline{B} = \overline{L_+} \cap e_B^{-1}(1)$ . Moreover, the base norm induced by  $\overline{B}$  coincides with the original one by  $B$ .
- (vi) If  $L_+$  is closed, then the Banach dual and order dual coincide with each other:  $L^* = L^\diamond$ .

### Proof

- (i) Representing  $x \in L$  as  $x = x_+ - x_-$  ( $x_\pm \in L_+$ ), we have

$$|e_B(x)| = |e_B(x_+) - e_B(x_-)| \leq e_B(x_+) + e_B(x_-).$$

It implies  $|e_B(x)| \leq \|x\|_B$ , i.e.,  $e_B$  is bounded.

- (ii) When  $L_+$  is closed, its base  $B = L_+ \cap \{x \in L \mid e_B(x) = 1\}$  is also closed. Assume conversely that  $B$  is closed. Since  $L$  is complete, for a Cauchy sequence  $\{\alpha_i x_i\}_i$  in  $L_+$  such that  $\alpha_i \geq 0$  and  $x_i \in B$ , there exists

<sup>11</sup>For a Banach space  $X$  and its Banach dual  $X^*$ , the *weak topology* of  $X$  often denoted by  $\sigma(X, X^*)$  is the weakest topology on  $X$  which makes all  $f \in X^*$  continuous, and the *weak\* topology* of  $X^*$  often denoted by  $\sigma(X^*, X)$  is the weakest topology on  $X^*$  which makes all  $x \in X \subset X^{**}$  continuous [50, 58].

<sup>12</sup>Clearly,  $L_+^*$  satisfies  $L_+^* = \bigcap_{x \in L_+} \{f \in L^* \mid f(x) \geq 0\}$ , and thus is weakly\* closed. On the other hand, let  $\{f_n\}$  be a Cauchy sequence in  $L_+^*$  which converges to  $f \in L^*$  in the norm topology, and suppose that there is  $x_0 \in L_+$  such that  $f(x_0) < 0$ . If we set  $\varepsilon = \frac{|f(x_0)|}{\|x_0\|}$ , then for large  $n$  we have  $\sup_{x \in L} \frac{|f(x) - f_n(x)|}{\|x\|} < \varepsilon$ , and thus  $\frac{|f(x_0) - f_n(x_0)|}{\|x_0\|} < \varepsilon$ . It implies  $f_n(x_0) < 0$ , which is a contradiction. Thus,  $f(x) \geq 0$  for all  $x \in L_+$ .

$v_* \in L$  to which  $\{\alpha_i x_i\}_i$  converges. From the continuity of  $e_B$ , we obtain  $\alpha_i = e_B(\alpha_i x_i) \rightarrow e_B(v_*)$ . If  $e_B(v_*) = 0$ , then  $\alpha_i \rightarrow 0$ . It implies  $\|\alpha_i x_i\|_B \rightarrow 0$ , i.e.,  $v_* = \lim_i \alpha_i x_i = 0$ , and thus  $v_* \in L_+$  because  $L_+$  is pointed (see Proposition 2.9). If  $e_B(v_*) \neq 0$ , then

$$\begin{aligned} e_B(v_*)\|x_i - x_j\|_B &\leq \|e_B(v_*)x_i - v_*\|_B + \|v_* - e_B(v_*)x_j\|_B \\ &\leq \|e_B(v_*)x_i - \alpha_i x_i\|_B + \|\alpha_i x_i - v_*\|_B \\ &\quad + \|v_* - \alpha_j x_j\|_B + \|\alpha_j x_j - e_B(v_*)x_j\|_B \\ &= |e_B(v_*) - \alpha_i| + \|\alpha_i x_i - v_*\|_B \\ &\quad + \|v_* - \alpha_j x_j\|_B + |\alpha_j - e_B(v_*)|. \end{aligned}$$

The last equation converges to 0 as  $i, j \rightarrow \infty$ , and thus  $\{x_i\}_i$  is a Cauchy sequence in  $L$ . Because  $B$  is closed,  $\{x_i\}_i$  converges to  $x_* \in B$ . Therefore, we obtain  $\lim_i \alpha_i x_i = e_B(v_*)x_* \in L_+$ .

(iii) This claim follows directly from the definition of  $\|\cdot\|_B$  as the Minkowski functional of  $D$ .

(iv) It can be found that

$$\begin{aligned} \|f\|_* &= \sup\{|f(x)| \mid \|x\|_B \leq 1\} \\ &= \sup\{|f(x)| \mid x \in \overline{D}\} \\ &= \sup\{|f(x)| \mid x \in D\} \\ &= \sup\{|f(x)| \mid x \in B\}. \end{aligned}$$

For the proofs of (v) and (vi), see Proposition 1.40 in [30].  $\square$

Roughly speaking, the base norm and the intensity functional correspond to the trace norm and the identity operator in the usual formulation of quantum theory respectively. In fact, if we let  $L$  be the set  $\mathcal{L}_S(\mathcal{H})$  of all self-adjoint operators on a finite-dimensional Hilbert space  $\mathcal{H}$ , then any  $x \in L$  is decomposed as  $x = x_+ - x_-$  with  $x_{\pm} \geq 0$  in the usual ordering for self-adjoint operators, and thus the trace norm of  $x$  is given via the identity operator  $\mathbb{1}$  by  $\|x\|_{\text{Tr}} = \text{Tr}[x_+] + \text{Tr}[x_-] = \text{Tr}[\mathbb{1}x_+] + \text{Tr}[\mathbb{1}x_-]$ , which corresponds to (2.11).

Let us move to the introduction of order unit Banach spaces.

**Definition 2.18**

Let  $L$  be an ordered vector space equipped with an ordering  $\leq$ .

(i)  $L$  is called *Archimedean* if  $x \leq 0$  whenever there exists  $y \in L$  such that  $nx \leq y$  for all  $n \in \mathbb{N}$ .

(ii)  $L$  is called *almost Archimedean* if  $x = 0$  whenever there exists  $y \in L$  such that  $-y \leq nx \leq y$  for all  $n \in \mathbb{N}$ .

(iii) A positive element  $u$  of  $L$  is called an *order unit* if for any  $x \in L$  there exists some  $n \in \mathbb{N}$  such that  $-nu \leq x \leq nu$ .

It is clear that if  $L$  is Archimedean, then it is almost Archimedean. For  $a, b \in L$ , we define the order interval  $[a, b]$  as  $[a, b] := \{x \in L \mid a \leq x \leq b\}$ . The following lemma is important.

**Lemma 2.19**

Let  $L$  be an ordered vector space with an ordering  $\leq$  and an order unit  $u$ .

(i) The order interval  $\Delta := [-u, u]$  is a radial, circled, and convex subset of  $L$ .

(ii) The Minkowski functional of  $\Delta$  of the form

$$p_\Delta = \inf\{\lambda > 0 \mid x \in \lambda\Delta\} \tag{2.13}$$

is a norm on  $L$  if and only if  $L$  is almost Archimedean.

**Proof**

It is easy to see that (i) holds due to the definition of the order unit  $u$ , and thus the Minkowski functional  $p_\Delta$  is a seminorm on  $L$ . Assume that  $p_\Delta$  is a norm and  $x \in L$  satisfies  $-y \leq nx \leq y$  for all  $n \in \mathbb{N}$  and some  $y \in L$ . Since there exists  $m \in \mathbb{N}$  such that  $-mu \leq y \leq mu$ , it holds that  $-mu \leq nx \leq mu$ , or  $-\frac{m}{n}u \leq x \leq \frac{m}{n}u$  for all  $n \in \mathbb{N}$ . Thus,  $\inf\{\lambda > 0 \mid -\lambda u \leq x \leq \lambda u\} = 0$ , and we can conclude  $x = 0$  because  $\|\cdot\|_u$  is a norm. Conversely, assume that  $L$  is almost Archimedean and  $x \in L$  satisfies  $p_\Delta(x) = 0$ . Then,  $-u \leq \frac{1}{\lambda}x \leq u$  holds for arbitrary small  $\lambda$ , and thus  $x = 0$  follows from the assumption that  $L$  is almost Archimedean, which concludes (ii).  $\square$

We can give the definition of an order unit Banach space.

**Definition 2.20**

Let  $L$  be an ordered vector space with an order unit  $u \in L$  associated with the ordering of  $L$ .  $(L, u)$  is called an *order unit Banach space* if  $L$  is Archimedean and complete with respect to the norm  $p_\Delta$ . In this case, we write  $p_\Delta(\cdot)$  as  $\|\cdot\|_u$ , and call it the *order unit norm*.

We present miscellaneous facts about base norm Banach spaces according mainly to [30, 55].

**Proposition 2.21**

Let  $(L, u)$  be an order unit Banach space, and let  $\leq$  be the ordering of  $L$ .

(i) The positive cone  $L_+$  of  $L$  is generating and closed.

(ii) The unit ball of  $L$  is given by  $\Delta = [-u, u] = \{x \in L \mid -u \leq x \leq u\}$ .

(iii) If  $f$  is a positive functional on  $L$ , then  $f$  is bounded, and its dual norm on the Banach dual  $L^*$  is given by  $\|f\|_* = f(u)$ . Conversely, if a linear functional  $f: L \rightarrow \mathbb{R}$  satisfies  $\|f\|_* = f(u)$ , then  $f$  is positive.

(iv) If we define  $B_u := \{f \in L_+^* \mid f(u) = 1\}$ , then  $B_u$  is a base for the Banach dual cone  $L_+^*$ .

(v) The Banach dual and order dual coincide with each other:  $L^* = L^\diamond$ .

**Proof**

(i) For  $x \in L$ , there exists  $n \in \mathbb{N}$  such that  $-nu \leq x \leq nu$ . Then,  $x = nu + (x - nu)$  shows  $x \in L_+ - L_+$ , i.e.,  $L_+$  is generating. Let  $\{x_i\}_i$  be a Cauchy sequence in  $L_+$  and converge to  $x_\star \in L$ . For any  $n \in \mathbb{N}$ , we have  $\|x_\star - x_i\|_u \leq \frac{1}{n}$  for sufficiently large  $i$ . It implies  $-\frac{1}{n}u \leq x_\star - x_i \leq \frac{1}{n}u$ , and thus  $-nx_\star \leq u$ . Since this holds for all  $n \in \mathbb{N}$  and  $L$  is Archimedean, we obtain  $-x_\star \leq 0$ , i.e.,  $x \in L_+$ .

(ii) It follows easily from (i) that  $\Delta$  is closed. Then, the definition of  $\|\cdot\|_u$  as the Minkowski functional of  $\Delta$  proves the claim.

(iii) Let  $x \in \Delta$ . From the positivity of  $f$ ,  $-f(u) \leq f(x) \leq f(u)$ , i.e.,  $\|f\|_* \leq f(u)$  holds. The equality clearly holds for  $x = u$ , and thus we obtain  $\|f\|_* = f(u)$  (in particular,  $f$  is bounded). Assume conversely that  $\|f\|_* = f(u)$ . For  $x \in L_+$  with  $\|x\|_u = 1$ , we have  $-u \leq x \leq u$ , or  $0 \leq u - x \leq u$ . It follows that  $\|u - x\|_u \leq 1$ , and because  $\|f\|_* = f(u)$ , we obtain  $|f(u) - f(x)| \leq f(u)$ , which implies  $f(x) \geq 0$ .

(iv) It can be seen from (iii) that every  $f \in L_+^*$  satisfies  $\|f\|_* = f(u)$ , and thus when considered as an element of  $L^{**}$ ,  $u$  is a strictly positive functional for  $L_+^*$ . Then, applying Proposition 2.14, we obtain the claim.

(v) See Proposition 1.29 in [30]. □

It can be verified easily that the order unit norm corresponds to the usual operator norm in the formulation of quantum theory.

Now we can give the most general description of GPTs in terms of base norm Banach spaces and order unit Banach spaces. We present first of all a fundamental theorem for our description on a close relationship between base norm Banach spaces and order unit Banach spaces (see [30, 55, 56] for the proof).

**Theorem 2.22**

(i) Let  $(L, B)$  be a base norm Banach space. Then, with  $e_B$  the intensity functional of  $B$  such that  $B = \{x \in L_+ \mid e_B(x) = 1\}$ ,  $(L^*, e_B)$  is an order unit Banach space, and  $L_+^* := \{f \in L^* \mid f(x) \geq 0 \text{ for all } x \in L_+\}$  is its positive cone. Moreover, the order unit norm on  $L^*$  coincides with the usual Banach dual norm.

(ii) Let  $(L, u)$  be an order unit Banach space. Then, with  $B_u := \{f \in L_+^* \mid f(u) = 1\}$ ,  $(L^*, B_u)$  is a base norm Banach space, and  $L_+^* := \{f \in L^* \mid f(x) \geq 0 \text{ for all } x \in L_+\}$  is its positive cone. Moreover, the base norm in  $L^*$  coincides with the usual Banach dual norm, and  $B_u$  is a weakly\* compact subset of  $L^*$ .

In the next subsection, we interpret this theorem in the language of GPTs and present the most standard formulation of GPTs based on it.

### 2.1.4 Standard formulations of GPTs

We adopt Theorem 2.22 (i) to our expression of GPTs. To do this, we recall that in Subsection 2.1.2 (Theorem 2.7) the state space of a GPT was shown to be represented as a convex subset  $\Omega$  of some Banach space  $V$  (note that  $0 \notin \Omega$  by its construction). We presented that the embedding vector space  $V$  is constructed by  $V = \text{span}(\Omega)$ , and there is a convex, pointed, and generating cone  $K$  in  $V$  given by  $K = \text{cone}(\Omega)$ . Moreover, we defined a norm in  $V$  such that

$$\|v\| = \inf\{\alpha + \beta \mid v = \alpha p - \beta q, \alpha, \beta \geq 0, p, q \in \Omega\}$$

(see (2.8)), and found that  $V$  is a Banach space and  $\Omega$  and  $K$  are closed with respect to the norm. These observations can be interpreted in the language of ordered Banach spaces. That is,  $V$  is a base norm Banach space whose positive cone  $V_+$  is given by  $V_+ = K = \text{cone}(\Omega)$  with its closed base  $\Omega$  and is closed and generating. On the other hand, it follows from Proposition 2.14 that there exists a strictly positive functional  $e_\Omega$  such that  $e_\Omega(\omega) = 1$  for all  $\omega \in \Omega$ . Then, Proposition 2.17 (i) and Theorem 2.22 (i) result in that this  $e_\Omega$  is an element of the Banach dual  $V^*$ , and in fact is an order unit of  $V^*$  ordered via the cone  $V_+^*$ . Since  $V = \text{span}(\Omega)$ , we can find that the order unit  $e_B$  coincides with the unit effect  $u \in V^*$  (see Theorem 2.7 (iii)). Overall, we have obtained the following observation.

#### Theorem 2.23

A GPT is given by  $(\Omega, \mathcal{E}_\Omega)$ , where

1. the state space  $\Omega$  is a closed base of the closed positive cone  $V_+$  in some base norm Banach space  $V$  such that  $V_+ = \text{cone}(\Omega)$  and  $V = \text{span}(\Omega)$ ;
2. the effect space  $\mathcal{E}_\Omega$  is a subset  $[0, u] = \{e \in V^* \mid 0 \leq e \leq u\}$  of the order unit Banach space  $V^*$  such that  $V^*$  is the Banach dual of  $V$  ordered via  $V_+^* := \{f \in V^* \mid f(x) \geq 0 \text{ for all } x \in L_+\}$  and  $u$  is the order unit of  $V^*$  determined by  $u(\omega) = 1$  for all  $\omega \in \Omega$ .

The contents of Theorem 2.23 are the most general formulation of GPTs. We remark that the positive cone  $V_+$  represents the set of all “unnormalized” states, which are not necessarily mapped to 1 by the unit effect  $u$ , and that  $\mathcal{E}_\Omega$  spans  $V^*$  because  $V_+^*$  is generating. We define another primitive notion of *observables* based on this representation.<sup>13</sup>

---

<sup>13</sup>Observables can be introduced also in terms of the abstract description of convex structures [43], but in this thesis, we present the definition of observables after embedding them into vector spaces for simplicity.

**Definition 2.24**

Let  $(\Omega, \mathcal{E}_\Omega)$  be a GPT. An *observable* whose outcome space is given by a measurable space  $(X, \mathcal{A})$  is defined as a normalized effect-valued measure  $E$  on  $(X, \mathcal{A})$ , i.e.,  $E: \mathcal{A} \rightarrow \mathcal{E}_\Omega$  such that

- (i)  $E(X) = u$ ;
- (ii)  $E(\bigcup_i U_i) = \sum_i E(U_i)$  for any countable family  $\{U_i\}$  of pairwise disjoint sets in  $\mathcal{A}$  (the sum converges in the weak\* topology on  $V^*$ ).

When the outcome set  $X$  of an observable  $E$  is finite, we often describe it as  $E = \{e_x\}_{x \in X}$  with  $e_x = E(\{x\})$  representing the yes-no measurement corresponding to the outcome  $x \in X$ . We also use the notation  $E = \{e_i\}_{i=1}^l$  when  $|X| = l$  ( $l < \infty$ ), where  $e_i$  represents the  $i$ th yes-no measurement. We note that  $\sum_{x \in X} e_x = u$  and  $\sum_{i=1}^l e_i = u$  hold. In this thesis, we assume that observables are composed of a finite number of nonzero effects, and the trivial observable  $\{u\}$  is not considered.

Although those descriptions above are of the most general form including theories with  $\dim V = \infty$ , we are interested only in finite-dimensional cases in this thesis, and present it explicitly as a mathematical assumption.

**Mathematical assumption 3 (Finite dimensionality)**

*Any GPT  $(\Omega, \mathcal{E}_\Omega)$  is associated with a finite-dimensional Euclidean space  $\mathbb{R}^d$  ( $d < \infty$ ).*

We note that any Hausdorff topological vector space of finite dimension is isomorphic linearly and topologically to the Euclidean space with the same dimension, and the norm, weak, and weak\* topologies on a Banach space and its dual are Hausdorff (thus these topologies coincide with each other to be Euclidean in finite-dimensional cases) [50, 58]. It should be also noted that a finite-dimensional vector space is isomorphic to its dual. If a GPT satisfies Mathematical assumption 3, then we call it a finite-dimensional GPT. Let us develop how we can simplify the formulation of GPTs shown in Theorem 2.23 when dealing with finite-dimensional theories. The following facts derived for the standard Euclidean topology are useful [30, 50, 60].

**Proposition 2.25**

*Let  $L = \mathbb{R}^d$  be a finite-dimensional ordered vector space (in particular, an ordered Banach space with respect to the Euclidean norm) whose positive cone  $L_+$  is generating.*

- (i) *The condition that  $L_+$  is generating is equivalent to the condition that  $L_+$  has an interior point.*
- (ii)  *$L_+$  is closed if and only if  $L$  is Archimedean.*
- (iii) *If  $L_+$  is closed, then the following statements for  $e \in L_+^*$  are equivalent (note that the Banach dual cone  $L_+^*$  is a positive cone for  $L^*$  because  $L_+$  is generating, and thus makes  $L^*$  an ordered vector space):*

1.  $e$  is strictly positive, i.e.,  $e(x) > 0$  for all  $x \in L_+ \setminus \{0\}$ ;

2.  $e$  is an interior point of  $L_+^*$ ;

3.  $e$  is an order unit in  $L^*$ .

(iv) If  $L_+$  is closed and  $B$  is a base of  $L_+$ , then  $B$  is bounded.

(v) If  $L_+$  is closed, then  $L_+$  admits a bounded base, i.e., there exists a bounded base for  $L_+$ .

(vi) If  $L_+$  is closed, then all types of dual  $L'$ ,  $L^\diamond$ , and  $L^*$  coincide with each other.

### Proof

In this proof, we denote the ordering of  $L$  by  $\leq$  (thus,  $x \geq 0$  if and only if  $x \in L_+$ ).

(i) Let  $u$  be an interior point of  $L_+$ . Then, there exists an open ball  $C$  in  $L$  such that  $u + C \subset L_+$ . For  $v \in C$ , because  $C$  is a ball and thus  $-v \in C$ , we have  $u \pm v \geq 0$ , i.e.,  $-u \leq v \leq u$ . Therefore, we obtain  $C \subset [-u, u]$ , i.e.,  $u$  is an order unit, which implies that  $L_+$  is generating (see the proof of Proposition 2.21 (i)). Assume conversely that  $L_+$  is generating. It is not difficult to see that the maximal set  $\{v_i\}_{i=1}^k$  of linearly independent elements in  $L_+$  is a basis of  $L$  (and thus  $k = d$ ). Let us consider a subset  $U := \{v \in L \mid v = \sum_{i=1}^d \lambda_i v_i \text{ with } \sum_{i=1}^d |\lambda_i| < 1\}$  of  $L$ . Because a map  $\|\cdot\|'$  on  $L$  given by  $\|\sum_{i=1}^d \lambda_i v_i\|' = \sum_{i=1}^d |\lambda_i|$  defines a norm on  $L$ , the above  $U$  is an open subset in  $L$  (remember that all norm topologies are equivalent to each other in finite-dimensional cases). Defining  $v_\star := \sum_{i=1}^d v_i \in L_+$ , we can see that for any  $v = \sum_{i=1}^d \lambda_i v_i \in U$ ,  $v_\star + v = \sum_{i=1}^d (1 + \lambda_i) v_i \in L_+$  holds because  $v_i \in L_+$  and  $1 + \lambda_i > 0$ . This implies  $v_\star + U \subset L_+$ , and thus  $v_\star$  is an interior point of  $L_+$ .

(ii) Suppose that  $L_+$  is closed. If  $x, y \in L$  satisfy  $nx \leq y$  for all  $n \in \mathbb{N}$ , then a sequence  $\{\frac{1}{n}y - x\}_n$  in  $L_+$  converges to  $-x \in L_+$ , and thus we have  $x \leq 0$ . Conversely, suppose that  $L$  is Archimedean, and consider  $x \in \overline{L_+}$ . With  $y \in \text{int}(L_+)$ , where  $\text{int}(L_+)$  is the nonempty interior of  $L_+$  (see (i)), we can see that  $\frac{1}{n+1}y + (1 - \frac{1}{n+1})x \in \text{int}(L_+)$  holds for any  $n \in \mathbb{N}$ . It follows that  $y + nx \in \text{int}(L_+) \subset L_+$ , and thus  $-nx \leq y$  for all  $n \in \mathbb{N}$ . Since  $L$  is Archimedean, we obtain  $x \geq 0$ , which means  $\overline{L_+} \subset L_+$ .

(iii) (1 $\rightarrow$ 2) Let  $e \in L^*$  be strictly positive, and consider a unit ball  $C := \{x \in L \mid \|x\| \leq 1\}$  and a unit sphere  $D := \{f \in L \mid \|x\| = 1\}$  in  $L$ , where  $\|\cdot\|$  is the Euclidean norm. Because  $L_+$  is closed and  $L = \mathbb{R}^d$  is finite-dimensional,  $S := L_+ \cap D$  is a compact subset of  $L$ . It implies that there exists a minimum value  $M > 0$  for the strictly positive and continuous functional  $e$  on  $S$ . On the other hand, if we define a unit ball  $C_\star := \{f \in L^* \mid \|f\|_\star \leq 1\}$  in  $L^*$  with the Banach dual norm  $\|\cdot\|_\star$  (which is equivalent to Euclidean norm in this finite-dimensional case), then, for  $f \in C_\star$ , we have  $\|f\|_\star = \sup_{y \in D} |f(y)|$



[58], and thus  $-1 \leq f(y) \leq 1$  holds for all  $y \in S$ . It follows that if we take  $0 < \varepsilon < M$ , then the functional  $e + \varepsilon f$  satisfies  $(e + \varepsilon f)(y) > 0$  for all  $y \in S$ . Since this holds for every  $f \in C_*$  and any  $x \in L_+$  can be represented as  $x = \lambda y$  with  $\lambda = \|x\| \geq 0$  and  $y \in S$ , we can conclude that  $e + \varepsilon C_* \subset L_+^*$ , i.e.,  $e$  is an internal point of  $L_+^*$ .

(2 $\rightarrow$ 3) Because  $e$  is an interior point of  $L_+^*$ , for every  $f \in L^*$ , there exist  $\alpha, \beta > 0$  such that  $e + \alpha f \in L_+^*$  and  $e + \beta(-f) \in L_+^*$ . It can be rewritten as  $-\frac{1}{\alpha}e \leq f \leq \frac{1}{\beta}e$ , and thus we can conclude that  $e$  is an order unit in  $L^*$ .

(3 $\rightarrow$ 1) Suppose that there exists  $x_0 \in L_+ \setminus \{0\}$  such that  $e(x_0) = 0$ . Since  $e$  is an order unit, for  $f \in L^*$ , there exists  $n \in \mathbb{N}$  such that  $-ne \leq f \leq ne$ , i.e.,  $f(x_0) = 0$ . Because this holds for all  $f \in L^*$ , we obtain  $x_0 = 0$ , which is a contradiction.

(iv) Let  $e_B$  be the intensity functional for  $B$ , which is strictly positive according to Proposition 2.14. Since any linear functional is continuous in a finite dimensional topological vector space (see Theorem 3.4 in [50]), we obtain  $e_B \in L_+^*$ . It follows from (iii) that  $e_B$  is an order unit in  $L^*$ , and thus, for  $f \in L^*$ , there exists  $n \in \mathbb{N}$  such that  $-ne_B \leq f \leq ne_B$ . We obtain  $|f(x)| \leq n$  for all  $x \in B$ , and because  $f \in L^*$  is arbitrary, we can conclude that  $B$  is bounded.

(v) For the unit sphere  $D$  in  $L$  introduced above, consider  $T := L_+ \cap D$  and its convex hull  $T' := \text{conv}(T)$ . Clearly,  $T'$  does not include 0, and we can find that  $T'$  is compact because  $T$  is compact (see Theorem 10.2 in [50]). Thus, there exists  $x_0 \in T'$  such that the continuous norm function  $\|\cdot\|$  takes its minimum in  $T'$ . It follows that any  $x' \in T'$  satisfies  $\|x_0\| \leq \|x_0 - t(x_0 - x')\|$  for  $0 \leq t \leq 1$  because  $x_0 - t(x_0 - x') = (1-t)x_0 + tx' \in T'$ . It can be rewritten as  $t^2\|x_0 - x'\|^2 - 2t(x_0, x_0 - x')_E \geq 0$ , where  $(\cdot, \cdot)_E$  is the Euclidean inner product in  $L = \mathbb{R}^d$ . Since this holds for all  $0 \leq t \leq 1$ , it must hold that  $(x_0, x_0 - x')_E \leq 0$ , that is, any  $x' \in T'$  satisfies  $(x_0, x')_E \geq (x_0, x_0)_E > 0$ . On the other hand, any  $x \in L_+$  can be written as  $x = \|x\|y$  with  $y \in D$  (in particular,  $y \in T'$ ). Therefore, we obtain  $(x_0, x)_E > 0$  for all  $x \in L_+ \setminus \{0\}$ . By means of the Riesz representation theorem [58], we can identify the inner product  $(x_0, \cdot)_E$  as an element  $f_0 \in L^*$  such that  $f_0(x) = (x_0, x)_E$ . This  $f_0$  is a strictly positive functional for  $L_+$ , and thus defines a base, which is bounded as shown in (iv).

(vi) As we have seen in (iv) above, any linear functional on  $L = \mathbb{R}^d$  is continuous, and thus we obtain  $L^* = L'$  (and  $L_+^* = L_+^\diamond$ ). On the other hand, it follows from (v) above that there are a base  $B$  in  $L$  and a strictly positive functional  $e_B \in L_+^*$  associated with  $B$ . Then, (iii) and (i) implies that the Banach dual cone  $L_+^*$  generates the Banach dual  $L^*$ , and because  $L_+^* = L_+^\diamond$ , we can conclude the claim (remember that the order dual  $L^\diamond$  is given by  $L^\diamond = \text{span}(L_+^\diamond)$ ).  $\square$

*Remark 2.26*

Proposition (iii)-(vi) do not necessarily hold when  $L_+$  is not closed. To confirm this, let us consider the case when  $L = \mathbb{R}^2$  and  $L_+ = \{(x, y) \in \mathbb{R}^2 \mid y > 0\} \cup (0, 0)$ . It is easy to see that  $L_+$  defines a convex, pointed, and generating cone, but we cannot find a bounded base for this  $L_+$  or verify  $L^* = L^\diamond$ .

Theorem 2.23 now can be rewritten as follows.

**Corollary 2.27**

A GPT is given by  $(\Omega, \mathcal{E}_\Omega)$ , where

1. the state space  $\Omega$  is a compact convex set of some finite-dimensional Euclidean space  $V = \mathbb{R}^{N+1}$  ( $N < \infty$ ) such that  $\text{span}(\Omega) = V$  and  $0 \notin \text{aff}(\Omega)$  (in particular,  $\dim \text{aff}(\Omega) = N$  holds<sup>14</sup>);
2. the effect space  $\mathcal{E}_\Omega$  is a subset  $[0, u] = \{e \in V^* \mid 0 \leq e \leq u\}$  of the dual space  $V^*$  of  $V$  such that  $u(\omega) = 1$  for all  $\omega \in \Omega$ .<sup>15</sup>

The mathematical expression given in Corollary 2.27 is the standard formulation of GPTs in this thesis, and all observations on GPTs are based on this description. We note that order structures similar to the ones described in Theorem 2.23 can be introduced for these finite-dimensional  $V$  and  $V^*$ . In fact, in Corollary 2.27, we can verify easily that an order structure can be introduced for  $V$  by a generating cone  $V_+ := \text{cone}(\Omega)$ , and  $\Omega$  is a compact (thus closed) base for  $V_+$  with which  $V$  is a base norm Banach space. There we can also find that  $V^*$  can be ordered via a generating cone  $V_+^* := \{f \in V^* \mid f(x) \geq 0 \text{ for all } x \in L_+\}$ , and the functional  $u$ , which is the intensity functional for the base  $\Omega$ , is an order unit with which  $V^*$  is an order unit Banach space.

Let us further introduce several notions about finite-dimensional GPTs. For a state space  $\Omega$ , we can consider its extreme points<sup>16</sup>, and denote the set of all extreme points of  $\Omega$  by  $\Omega^{\text{ext}} = \{\omega_i^{\text{ext}}\}_{i \in \mathcal{I}}$ , where  $\mathcal{I}$  is an index set. Because  $\Omega$  is a compact convex set in  $\mathbb{R}^{N+1}$ , thanks to the Krein-Milman theorem,  $\Omega^{\text{ext}}$  is not empty and  $\Omega = \text{conv}(\Omega^{\text{ext}})$  [49, 50, 58]. Similar arguments also hold for the corresponding effect space  $\mathcal{E}_\Omega$  since  $\mathcal{E}_\Omega = V_+^* \cap (u - V_+^*)$  and Proposition 2.21 (ii) imply  $\mathcal{E}_\Omega$  is closed and bounded, i.e., compact.<sup>17</sup>

<sup>14</sup>For an affine set  $A$  of a finite-dimensional vector space  $L$ , its dimension  $\dim A$  is defined as the dimension of the set  $A - a_0$  ( $a_0 \in A$ ) as a vector subspace of  $L$ .

<sup>15</sup>Although the dual space  $V^*$  of  $V$  is isomorphic to  $\mathbb{R}^{N+1}$ , we do not identify them here (see Subsection 2.4.2).

<sup>16</sup>For a convex subset  $C$  in a vector space,  $x \in C$  is called an *extreme point* of  $C$  if  $x = \lambda y + (1 - \lambda)z$  with  $y, z \in C$  and  $0 < \lambda < 1$  implies  $y = z = x$ .

<sup>17</sup>Therefore, the effect space  $\mathcal{E}_\Omega$  is closed under infinite countable mixtures.

**Definition 2.28**

- (i) An extreme point of  $\Omega$  is called a *pure state*, and a state which is not pure is called a *mixed state*.
- (ii) An extreme point of  $\mathcal{E}_\Omega$  is called a *pure effect*, and an effect which is not pure is called a *mixed effect*.
- (iii) An effect  $e$  is called *indecomposable* if  $e \neq 0$  and a decomposition  $e = e_1 + e_2$ , where  $e_1, e_2 \in \mathcal{E}_\Omega$ , implies that both  $e_1$  and  $e_2$  are scalar multiples of  $e$ . We denote the set of all pure and indecomposable effects (shown to be nonempty [61]) by  $\mathcal{E}_\Omega^{\text{ext}} = \{e_j^{\text{ext}}\}_{j \in \mathcal{J}}$ , where  $\mathcal{J}$  is an index set.

It is easy to see that the unit effect  $u$  is pure and  $e^\perp := u - e \in \mathcal{E}_\Omega$  is pure whenever  $e \in \mathcal{E}_\Omega$  is pure. It can be also observed that pure and indecomposable effects correspond to rank-1 projections in quantum theory (see Subsection 2.5.2), and that an effect  $e \in \mathcal{E}_\Omega$  is indecomposable if and only if  $e$  is on an extremal ray of  $V_+^*$ .<sup>18</sup> We call two GPTs  $(\Omega_1, \mathcal{E}_{\Omega_1})$  and  $(\Omega_2, \mathcal{E}_{\Omega_2})$  *equivalent* if there exists an affine bijection (affine isomorphism)  $\psi$  such that  $\psi(\Omega_1) = \Omega_2$ . In this case, we can find easily that  $\mathcal{E}_{\Omega_2} = \mathcal{E}_{\Omega_1} \circ \psi^{-1}$ , and thus physical predictions are covariant (equivalent), which can be regarded as a physical expression of Proposition 2.5 (ii). We remark that the affine isomorphism  $\psi$  is indeed a linear isomorphism on the underlying vector space  $V_1 = \text{span}(\Omega_1)$  and  $V_2 = \text{span}(\Omega_2)$  (see the proof of Proposition 2.5 (ii)). A set of  $m$  states  $\{\omega_1, \omega_2, \dots, \omega_m\}$  is called *perfectly distinguishable* if there exists an observable  $\{e_1, e_2, \dots, e_m\}$  such that  $e_i(\omega_j) = \delta_{ij}$ ,  $i, j = 1, 2, \dots, m$ . In general, we can not identify the state of a system by a single measurement. However, for perfectly distinguishable states, there exists a measurement by which we can detect perfectly in which state the system is prepared.

*Remark 2.29*

There is a physical interpretation for the mathematical assumption of finite dimensionality. In [22], Hardy assumed that any state is determined by a finite set of effects named *fiducial measurements*. If we denote the finite set of fiducial measurements by  $\{e_i^{\text{fid}}\}_{i=0}^N$  ( $N < \infty$ ), then it is equivalent to regarding a state  $\omega$  as a vector

$$\omega = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_N \end{pmatrix}, \tag{2.14}$$

where the  $i$ th row  $a_i$  represents the probability  $e_i^{\text{fid}}(\omega)$ . It is easy to see that Hardy's formulation is consistent with ours: the state space  $\Omega$  composed

---

<sup>18</sup>A ray  $P \subset V_+^*$  is called an *extremal ray* of  $V_+^*$  if  $x \in P$  and  $x = y + z$  with  $y, z \in V_+^*$  imply  $y, z \in P$ .

by  $\omega$  of the form (2.14) is a compact (or closed and bounded) convex set in  $\mathbb{R}^{N+1}$  (by requiring completeness), and the normalization  $u(\omega) = 1$  for the unit effect  $u$  yields the condition  $\dim \text{aff}(\Omega) = N$ . We note that similar formulations for infinite-dimensional cases are given in [40]. That is, a state  $\omega$  is regarded as an element of the product set  $[0, 1]^\mathcal{E}$  with a set of effects  $\mathcal{E}$  similarly to (2.14), and the state space is a subset of  $[0, 1]^\mathcal{E}$  which is compact with respect to the pointwise convergence topology (i.e., the weak\* topology) because  $[0, 1]^\mathcal{E}$  is compact with respect to this topology due to Tychonoff's theorem [62] (see also Theorem 2.22 (ii)).

*Remark 2.30*

In our formulation, effects are constructed from states in the way how a state space is given first as a closed base of a base norm Banach space and then effects are given in its dual (see Theorem 2.23 and Corollary 2.27). On the other hand, as in the operator algebraic formulation of quantum theory [40, 63], it should be allowed to construct theories starting with effects. In fact, for a finite-dimensional GPT  $(\Omega, \mathcal{E}_\Omega)$ , if we consider the set  $\Theta := \{x \in V_+^{**} \mid x(u) = 1\}$ , where  $V^{**}$  is the double Banach dual of  $V$  or the Banach dual of  $V^*$ , i.e.,  $V^{**} = (V^*)^*$ , then by means of the canonical identification of  $V$  with  $V^{**}$  it holds that  $\Theta = \Omega$ .<sup>19</sup> This can be proven in a similar way to Proposition 2.5 (i) by just regarding  $\Omega$  as  $S$  (an explicit proof is given in [31]). For general cases, as was proven that states represented by a total convex structure can be embedded into a base norm Banach space, one can show that an abstract expression of effects called a *convex effect algebra* (with some completeness) can be embedded into an order unit Banach space [64, 65, 66]. Then, due to Theorem 2.23, we can obtain successfully the corresponding state space in a base norm Banach space. Moreover, if  $\Omega$  is weakly compact, which is identical to the reflexivity of the base norm Banach space  $V$  ( $V = V^{**}$ ) [67], then  $\Theta = \Omega$  holds also in an infinite-dimensional case.

## 2.2 Composite systems

In the previous section, we have presented the mathematical formulation of single systems in GPTs. Then, it is natural to ask how a system composed of several single systems, a *composite system*, is described mathematically in GPTs. This is also motivated by another physical reason that it is in general difficult to isolate perfectly a system from environments: a composite system

---

<sup>19</sup>A triple  $(V, V_+, u)$ , where  $V$  is a finite-dimensional ordered vector space with a closed positive cone  $V_+$  and  $u \in V^*$  is a strictly positive functional on  $V$ , is sometimes called an *abstract state space* [29]. The subset  $V_+ \cap u^{-1}(1)$  in this formulation corresponds to a state space in our formulation.

of the target system and its environments emerges naturally [41]. In this part, we establish the mathematical formulation of composite systems in GPTs based on that of single systems. We note that we only study theories for bipartite systems in this thesis. Our description may seem to be only for limited cases and not general, but it is in fact an essential one also for multipartite cases<sup>20</sup>, and we can develop sufficiently interesting observations for this simplest scenario.

Let us consider a composite system composed of two single systems characterized by GPTs  $(\Omega_A, \mathcal{E}_{\Omega_A})$  and  $(\Omega_B, \mathcal{E}_{\Omega_B})$ . By convention, we suppose that the two subsystems are controlled by Alice and Bob respectively. A fundamental assumption that is usually assumed implicitly is that the total system is also expressed by a GPT. In the following, we follow this assumption, and denote the GPT for the total system by  $(\Omega_{AB}, \mathcal{E}_{\Omega_{AB}})$ . Similarly to the previous section, we write the vector spaces that embed canonically  $\Omega_A$ ,  $\Omega_B$ , and  $\Omega_{AB}$  as  $V_A$ ,  $V_B$ , and  $V_{AB}$  respectively (thus the dual vector space  $V_A^*$  embeds canonically  $\mathcal{E}_{\Omega_A}$ , for example). For the joint system, it is natural to require that every individual and independent preparation or measurement by Alice and Bob is a valid preparation or measurement in the bipartite system respectively. It is also reasonable to assume that if such an independent preparation by Alice or Bob is probabilistic with some probability weight, then the total preparation is also probabilistic with the same probability weight (similarly for independent measurements). Its mathematical expression is given as follows [30].

**Axiom 4 (Validity of individual preparations and measurements)**

*There exist bi-affine maps  $\phi: \Omega_A \times \Omega_B \rightarrow \Omega_{AB}$  and  $\psi: \mathcal{E}_{\Omega_A} \times \mathcal{E}_{\Omega_B} \rightarrow \mathcal{E}_{\Omega_{AB}}$  such that*

$$[\psi(e_A, e_B)](\phi(\omega_A, \omega_B)) = e_A(\omega_A) \cdot e_B(\omega_B) \quad (2.15)$$

*for all  $\omega_A \in \Omega_A$ ,  $\omega_B \in \Omega_B$  and  $e_A \in \mathcal{E}_{\Omega_A}$ ,  $e_B \in \mathcal{E}_{\Omega_B}$ . Each  $\phi(\omega_A, \omega_B)$  and  $\psi(e_A, e_B)$  are called a product state and a product effect respectively.*

In the assumption, each product state  $\phi(\omega_A, \omega_B)$  represents the individual preparation of  $\omega_A$  and  $\omega_B$  by Alice and Bob, and the individual convexity is reflected via the notion of bi-affinity of the map  $\phi$  (similarly for each product effect  $\psi(e_A, e_B)$  and the map  $\psi$ ). We also require that if Alice and Bob measure their respective unit effects  $u_A$  and  $u_B$  individually on any joint state (not necessarily a product state), then the observed probability is 1. In other words, the unit effect of the total system is  $\psi(u_A, u_B)$ .

---

<sup>20</sup>For the description of multipartite systems, see [29, 31].

**Axiom 5 (Unit effect of the total system)**

The unit effect  $u_{AB}$  of the joint system is given by the product effect  $\psi(u_A, u_B)$  of each unit effect  $u_A$  and  $u_B$  of Alice and Bob respectively.

Let us give an easy consequence of these axioms according mainly to [30].

**Lemma 2.31**

Assume Axiom 4 and Axiom 5. There are linear injections  $\Phi: V_A \otimes V_B \rightarrow V_{AB}$  and  $\Psi: V_A^* \otimes V_B^* \rightarrow V_{AB}^*$  such that

- (i)  $\Phi(\omega_A \otimes \omega_B) = \phi(\omega_A, \omega_B)$  for all  $\omega_A \in \Omega_A$  and  $\omega_B \in \Omega_B$ ;
- (ii)  $\Psi(e_A \otimes e_B) = \psi(e_A, e_B)$  for all  $e_A \in \mathcal{E}_{\Omega_A}$  and  $e_B \in \mathcal{E}_{\Omega_B}$ ;
- (iii)  $u_{AB} = \Psi(u_A \otimes u_B)$ .

**Proof**

Let us first construct a bi-linear extension  $\Phi'$  on  $V_A \times V_B$  of the bi-affine map  $\phi$  on  $\Omega_A \times \Omega_B$ . Due to the assumption of the bi-affinity,  $\phi(\omega_A, \cdot)$  defines an affine map from  $\Omega_B$  to  $\Omega_{AB}$  for a fixed  $\omega_A \in \Omega_A$ , and it can be extended (uniquely) to a linear map  $[\phi'(\omega_A)](\cdot)$  from  $span(\Omega_B) = V_B$  to  $span(\Omega_{AB}) = V_{AB}$  such that  $[\phi'(\omega_A)](\omega_B) = \phi(\omega_A, \omega_B)$  for all  $\omega_B \in \Omega_B$  (see the proof of Proposition 2.5 (ii)). In this way, we obtain a map  $P: \Omega_A \rightarrow \mathcal{L}(V_B, V_{AB})$ , where  $\mathcal{L}(V_B, V_{AB})$  is the set of all linear operators from  $V_B$  to  $V_{AB}$ . It is easy to see that  $P$  is affine, and thus, similarly to the above argument, it has a unique linear extension  $P: V_A \rightarrow \mathcal{L}(V_B, V_{AB})$  such that  $[P(\omega_A)](\cdot) = [\phi'(\omega_A)](\cdot)$  for all  $\omega_A \in \Omega_A$ . The bi-linear extension  $\Phi'$  of  $\phi$  is now obtained by  $\Phi'(v_A, v_B) = [P(v_A)](v_B)$  for  $v_A \in V_A, v_B \in V_B$ . Then, the existence of the linear map  $\Phi: V_A \otimes V_B \rightarrow V_{AB}$  satisfying  $\Phi(v_A \otimes v_B) = \Phi'(v_A, v_B)$  for all  $v_A \in V_A, v_B \in V_B$  (in particular (i)) follows immediately from the universal property of tensor product [68]. The existence of  $\Psi$  satisfying (ii) is proved similarly, and (iii) is an easy consequence of Axiom 5.

The remaining problem is to show the injectivity of  $\Phi$ . Because  $\Omega_A$  and  $\Omega_B$  span  $V_A$  and  $V_B$  respectively, any  $v_A \otimes v_B \in V_A \otimes V_B$  with  $v_A \in V_A, v_B \in V_B$  is expressed as  $v_A \otimes v_B \in V_{AB} = \sum_{i,j} a_{ij} \omega_A^i \otimes \omega_B^j$  with  $a_{ij} \in \mathbb{R}$  and  $\omega_A^i \in \Omega_A, \omega_B^j \in \Omega_B$ . Similarly, any  $w_A \otimes w_B \in V_A^* \otimes V_B^*$  with  $w_A \in V_A^*, w_B \in V_B^*$  is expressed as  $w_A \otimes w_B = \sum_{k,l} b_{kl} e_A^k \otimes e_B^l$  with  $b_{kl} \in \mathbb{R}$  and  $e_A^k \in \mathcal{E}_{\Omega_A}, e_B^l \in \mathcal{E}_{\Omega_B}$ . Thus, we can observe from the linearity of  $\Phi$  and  $\Psi$  that

$$[\Psi(w_A \otimes w_B)](\Phi(v_A \otimes v_B)) = w_A(v_A) \cdot w_B(v_B).$$

Let  $v_{AB} \in V_{AB}$  satisfy  $\Phi(v_{AB}) = 0$ . Then, since  $v_{AB}$  is expressed by  $v_{AB} = \sum_i v_A^i \otimes v_B^i$  with  $v_A^i \in V_A, v_B^i \in V_B$ , it holds for all  $w_A \in V_A^*, w_B \in V_B^*$  that

$$[w_A \otimes w_B](v_{AB}) = \sum_i w_A(v_A^i) \cdot w_B(v_B^i) = [\Psi(w_A \otimes w_B)](\Phi(v_{AB})) = 0.$$

Because  $\{w_A \otimes w_B \mid w_A \in V_A^*, w_B \in V_B^*\}$  spans  $V_{AB}^*$ , we can conclude  $v_{AB} = 0$ , which means that  $\Phi$  is injective. The injectivity of  $\Psi$  can be proved similarly.  $\square$

*Remark 2.32*

It seems to be assumed implicitly in Axiom 4 and Axiom 5 that Alice's actions do not influence Bob, and vice versa. For example, there we require that Alice and Bob can prepare individually their states and effects without influencing each other, or we can see from the bi-affinity (bi-linearity) of  $\psi$  that the statistics observed by Alice alone are independent of Bob's measurements: for any joint state  $\omega_{AB}$ , the probability of Alice observing  $e_A \in \mathcal{E}_{\Omega_A}$  does not depend on Bob's measurement  $\{f_B^i\}_i$  because it holds that

$$\sum_i [\psi(e_A, f_B^i)](\omega_{AB}) = [\psi(e_A, u_B)](\omega_{AB}).$$

In fact, Axiom 4 and Axiom 5 can be rephrased in terms of the so-called *no-signaling principle* [24, 69, 70]<sup>21</sup>, or the requirement of *causality* [26, 27].

There is another important requirement for bipartite systems. We require that every joint state can be determined by local measurements. This claim called the *tomographic locality for states* [22, 24, 71] is described mathematically as follows.

**Axiom 6 (Tomographic locality for states)**

If  $\omega_{AB}, \omega'_{AB} \in \Omega_{AB}$  satisfy  $[\psi(e_A, e_B)](\omega_{AB}) = [\psi(e_A, e_B)](\omega'_{AB})$  for all  $e_A \in \mathcal{E}_{\Omega_A}$  and  $e_B \in \mathcal{E}_{\Omega_B}$ , then  $\omega_{AB} = \omega'_{AB}$ .

**Lemma 2.33**

Assume Axiom 4, Axiom 5, and Axiom 6. The linear injections  $\Phi$  and  $\Psi$  in Lemma 2.31 are also surjective, that is,  $\Phi$  is a linear bijection between  $V_A \otimes V_B$  and  $V_{AB}$ , and  $\Psi$  between  $V_A^* \otimes V_B^*$  and  $V_{AB}^*$ .

**Proof**

Assume that  $V_{AB}^* \setminus \Psi(V_A^* \otimes V_B^*)$  is nonempty, and  $w'_{AB} \in V_{AB}^* \setminus \Psi(V_A^* \otimes V_B^*)$ . Then, because  $w'_{AB}$  and a basis of  $\Psi(V_A^* \otimes V_B^*)$  are linearly independent, we can construct an element  $v'_{AB}$  of  $V_{AB}^{**}$  such that  $v'_{AB}(w'_{AB}) = 1$  and  $v'_{AB}(w_{AB}) = 0$  for all  $w_{AB} \in \Psi(V_A^* \otimes V_B^*)$ . We note that  $V_{AB}^{**} = V_{AB}$  holds due to the assumption of finite dimensionality, and thus  $v'_{AB}$  above can be regarded as an element of  $V_{AB}$ . It follows that if we define  $M := \{v_{AB} \in V_{AB} \mid w_{AB}(v_{AB}) = 0 \text{ for all } w_{AB} \in \Psi(V_A^* \otimes V_B^*)\}$ , then  $M \setminus \{0\}$  is nonempty. In the following, we prove that  $M = \{0\}$ , which implies  $V_{AB}^* = \Psi(V_A^* \otimes V_B^*)$ . Let  $v_{AB}^* \in M$ . For a state  $\omega_{AB} \in \text{int}(V_{AB+})$ , where  $\text{int}(V_{AB+})$  is the interior of the positive cone  $V_{AB+}$  of  $V_{AB}$  generated by  $\Omega_{AB}$  (see Proposition 2.25),

<sup>21</sup>How the no-signaling principle is formulated in GPTs is explained in detail in [70].

we can make  $\omega_{AB}^* := \omega_{AB} + \varepsilon v_{AB}^*$  belong to  $V_{AB+}$  if we take sufficiently small  $\varepsilon > 0$ . Because  $u_{AB} = \Psi(u_A \otimes u_B)$ , it holds from the definition of  $M$  that  $u_{AB}(\omega_{AB}^*) = u_{AB}(\omega_{AB}) = 1$ , i.e.,  $\omega_{AB}^* \in \Omega_{AB}$ . Moreover, we can find in a similar way that  $[\Psi(e_A \otimes e_B)](\omega_{AB}^*) = [\Psi(e_A \otimes e_B)](\omega_{AB})$  holds for all  $e_A \in \mathcal{E}_{\Omega_A}, e_B \in \mathcal{E}_{\Omega_B}$ , and thus, from Axiom 6,  $\omega_{AB}^* = \omega_{AB}$  holds. This implies  $v_{AB}^* = 0$ , which means  $M = \{0\}$  and  $V_{AB}^* = \Psi(V_A^* \otimes V_B^*)$ . Therefore, we can conclude  $\Psi$  is surjective (i.e., bijective). Then, it is easy to derive  $\dim V_{AB} = \dim V_A \otimes V_B = \dim V_A \cdot \dim V_B$ , and the surjectivity (bijectivity) of  $\Phi$  follows from this observation.  $\square$

We assume Axiom 4, Axiom 5, and Axiom 6 (thus Lemma 2.33) in this thesis. Then, it does not cause any problem to identify the subsets  $\Phi^{-1}(\Omega_{AB})$  and  $\Psi^{-1}(\mathcal{E}_{\Omega_{AB}})$  of  $V_A \otimes V_B$  and  $V_A^* \otimes V_B^* = (V_A \otimes V_B)^*$  with the state space and effect space of the joint system respectively (see the argument above Remark 2.29). We hereafter write  $\Phi^{-1}(\Omega_{AB})$  simply as  $\Omega_{AB}$ , and  $\Psi^{-1}(\mathcal{E}_{\Omega_{AB}})$  as  $\mathcal{E}_{\Omega_{AB}}$ , and work with these expressions of states and effects where product states and effects are represented as  $\omega_A \otimes \omega_B$  and  $e_A \otimes e_B$  ( $\omega_A \in \Omega_A, \omega_B \in \Omega_B$  and  $e_A \in \mathcal{E}_{\Omega_A}, e_B \in \mathcal{E}_{\Omega_B}$ ) respectively.

*Remark 2.34*

One may consider Axiom 5 to be more artificial when compared to the other axioms. In [30], it was explained that  $u_{AB} = u_A \otimes u_B$  holds if the *tomographic locality for effects* is imposed together with Axiom 4 and Axiom 6.

Let us give more detailed specifications of bipartite systems. For GPTs  $(\Omega_A, \mathcal{E}_{\Omega_A})$  and  $(\Omega_B, \mathcal{E}_{\Omega_B})$  of local systems, we define the following classes of convex sets [72, 73].

**Definition 2.35**

Let  $(\Omega_A, \mathcal{E}_{\Omega_A})$  and  $(\Omega_B, \mathcal{E}_{\Omega_B})$  be GPTs.

(i) The convex subset  $\Omega_A \otimes_{\min} \Omega_B$  of  $V_A \otimes V_B$  defined as

$$\Omega_A \otimes_{\min} \Omega_B := \left\{ \sum_i p_i \omega_A^i \otimes \omega_B^i \mid p_i \geq 0, \sum_i p_i = 1, \omega_A^i \in \Omega_A, \omega_B^i \in \Omega_B \right\}$$

is called the *minimal tensor product* of  $\Omega_A$  and  $\Omega_B$ . The minimal tensor product  $\mathcal{E}_{\Omega_A} \otimes_{\min} \mathcal{E}_{\Omega_B}$  of the effect spaces  $\mathcal{E}_{\Omega_A}$  and  $\mathcal{E}_{\Omega_B}$  is defined in the same way.

(ii) The convex subset  $\Omega_A \otimes_{\max} \Omega_B$  of  $V_A \otimes V_B$  defined as

$$\Omega_A \otimes_{\max} \Omega_B := \{ \omega_{AB} \in V_A \otimes V_B \mid (e_A \otimes e_B)(\omega_{AB}) \in [0, 1], \\ e_A \in \mathcal{E}_{\Omega_A}, e_B \in \mathcal{E}_{\Omega_B}, (u_A \otimes u_B)(\omega_{AB}) = 1 \}$$



is called the *maximal tensor product* of  $\Omega_A$  and  $\Omega_B$ . The maximal tensor product  $\mathcal{E}_{\Omega_A} \otimes_{max} \mathcal{E}_{\Omega_B}$  of the effect spaces  $\mathcal{E}_{\Omega_A}$  and  $\mathcal{E}_{\Omega_B}$  is defined in the same way.

It is verified easily that the minimal and maximal tensor products are dual to each other in the sense that  $\mathcal{E}_{\Omega_A \otimes_{min} \Omega_B} = \mathcal{E}_{\Omega_A} \otimes_{max} \mathcal{E}_{\Omega_B}$  and  $\mathcal{E}_{\Omega_A \otimes_{max} \Omega_B} = \mathcal{E}_{\Omega_A} \otimes_{min} \mathcal{E}_{\Omega_B}$  hold. A similar observation can be obtained if we start from effects (see Remark 2.30). We also note that  $\Omega_A \otimes_{min} \Omega_B \subset \Omega_A \otimes_{max} \Omega_B$  clearly holds.

By means of the axioms introduced so far, we can specify the joint state space  $\Omega_{AB}$  in the following way. First, it can be found that  $\Omega_{AB}$  must include  $\Omega_A \otimes_{min} \Omega_B$  because product states and probabilistic mixtures are required to exist. Similarly, the existence of product effects are imposed, and it follows that  $\Omega_{AB}$  is included in  $\Omega_A \otimes_{max} \Omega_B$ . We have now obtained the following description for bipartite systems.

**Theorem 2.36**

Let  $(\Omega_{AB}, \mathcal{E}_{\Omega_{AB}})$  be a GPT describing a bipartite system composed of two subsystems  $(\Omega_A, \mathcal{E}_{\Omega_A})$  and  $(\Omega_B, \mathcal{E}_{\Omega_B})$ . Then,

$$\Omega_A \otimes_{min} \Omega_B \subset \Omega_{AB} \subset \Omega_A \otimes_{max} \Omega_B \quad (2.16)$$

holds. Dually,

$$\mathcal{E}_{\Omega_A} \otimes_{min} \mathcal{E}_{\Omega_B} \subset \mathcal{E}_{\Omega_{AB}} \subset \mathcal{E}_{\Omega_A} \otimes_{max} \mathcal{E}_{\Omega_B} \quad (2.17)$$

holds.

It can be found that when a bipartite system  $(\Omega_{AB}, \mathcal{E}_{\Omega_{AB}})$  composed of  $(\Omega_A, \mathcal{E}_{\Omega_A})$  and  $(\Omega_B, \mathcal{E}_{\Omega_B})$  satisfies both (2.16) and (2.17), then Axiom 4, Axiom 5, and Axiom 6 hold. In fact, Axiom 4 and Axiom 5 clearly hold, and because any element of  $V_{AB}^*$  can be written as a linear combination of effects of the form  $e_A \otimes e_B$  (remember that  $\mathcal{E}_{\Omega_A}$  and  $\mathcal{E}_{\Omega_B}$  span  $V_A^*$  and  $V_B^*$  respectively), Axiom 6 also can be verified.

**Definition 2.37**

Each element of  $\Omega_A \otimes_{min} \Omega_B$  is called a *separable state*, and an element of the form  $\omega_A \otimes \omega_B$  is particularly called a *product state*. Each element of  $\Omega_A \otimes_{max} \Omega_B \setminus \Omega_A \otimes_{min} \Omega_B$  is called an *entangled state*. Separable effects, product effects, and entangled effects are defined in the same way.

It should be noted that entangled states exist unless either theory is classical. More precisely, it was shown in [74] that  $\Omega_A \otimes_{min} \Omega_B = \Omega_A \otimes_{max} \Omega_B$  holds if and only if either  $\Omega_A$  or  $\Omega_B$  is a simplex (i.e., a classical theory).

*Example 2.38 (Quantum theory over a real Hilbert space)*

Let  $\mathcal{K} = \mathbb{R}^d$  ( $d < \infty$ ) be a finite-dimensional real Hilbert space. We can consider a GPT whose state space is given by  $\Omega_{\text{rQT}}(\mathcal{K}) = \{\rho \in \mathcal{L}_S(\mathcal{K}) \mid \rho \geq 0, \text{Tr}[\rho] = 1\}$  with  $\mathcal{L}_S(\mathcal{K})$  the set of all self-adjoint operators on  $\mathcal{K}$ . The real quantum theory described by  $\Omega_{\text{rQT}}(\mathcal{K})$  often appears in the field of GPTs when deriving the standard quantum theory (i.e., complex quantum theory) from physical principles [22, 51]. It is easy to see that  $\text{aff}(\Omega_{\text{rQT}}(\mathcal{K}))$  and the standard embedding vector space  $V(\mathcal{K})$  are given by  $\text{aff}(\Omega_{\text{rQT}}(\mathcal{K})) = \{\rho \in \mathcal{L}_S(\mathcal{K}) \mid \text{Tr}[\rho] = 1\}$  and  $V(\mathcal{K}) = \mathcal{L}_S(\mathcal{K})$  respectively. We can also observe that  $\dim \text{aff}(\Omega_{\text{rQT}}(\mathcal{K})) = \frac{1}{2}(d^2 + d) - 1$  and  $\dim V(\mathcal{K}) = \frac{1}{2}(d^2 + d)$  hold (in particular,  $\dim V(\mathcal{K}) = \dim \text{aff}(\Omega_{\text{rQT}}(\mathcal{K})) + 1$  holds). Suppose in analogy with the formulation of a finite-dimensional quantum theory over a complex Hilbert space that the state space of the composite system composed of two identical state spaces  $\Omega_{\text{rQT}}(\mathcal{K})$  is given by  $\Omega_{\text{rQT}}(\mathcal{K} \otimes \mathcal{K}) = \{\rho \in \mathcal{L}_S(\mathcal{K} \otimes \mathcal{K}) \mid \rho \geq 0, \text{Tr}[\rho] = 1\}$ . Then, we can derive

$$\dim V_A \cdot \dim V_B = \left[\frac{1}{2}(d^2 + d)\right]^2, \quad \dim V_{AB} = \frac{1}{2}(d^4 + d^2),$$

where  $V_A = V_B = V(\mathcal{K})$  and  $V_{AB} = V(\mathcal{K} \otimes \mathcal{K})$  are the respective embedding vector spaces of the individual and the total state spaces. It implies  $\dim V_A \cdot \dim V_B < \dim V_{AB}$ , i.e.,  $V_A \otimes V_B = V_{AB}$  does not hold. Thus, we can conclude that the tomographic locality is not satisfied in a finite-dimensional quantum theory over a real Hilbert space (it is not difficult to see that Axiom 4 and Axiom 5 hold in this case).<sup>22</sup>

## 2.3 Transformations

In this section, we explain how transformations between systems are formulated in GPTs, which completes our review for basic notions on GPTs. It will be found that not only state changes such as time evolutions but also measurements can be described in terms of transformations or *channels*, which are more refined forms of transformations. We also introduce the notions of compatibility and incompatibility for channels, which play a key role in the following chapters.

### 2.3.1 Channels in GPTs

In quantum theory, transformations of systems is described via the notion of channels [41, 76]. In this part, we explain how channels are generalized in GPTs according mainly to [31, 77].

---

<sup>22</sup>We can also eliminate a finite-dimensional quantum theory over a quaternionic Hilbert space by a similar observation [75].

**Definition 2.39**

Let  $(\Omega_1, \mathcal{E}_{\Omega_1})$  and  $(\Omega_2, \mathcal{E}_{\Omega_2})$  be GPTs. An affine map  $T: \Omega_1 \rightarrow \Omega_2$  is called a *channel* from  $\Omega_1$  to  $\Omega_2$ . A linear map  $T: V_1 \rightarrow V_2$ , where  $V_1$  and  $V_2$  are the embedding vector spaces of  $\Omega_1$  and  $\Omega_2$  respectively, is equivalently called a channel from  $\Omega_1$  to  $\Omega_2$  if  $T(\Omega_1) \subset \Omega_2$  (thus it is positive in the sense that  $T((V_1)_+) \subset T((V_2)_+)$ <sup>23</sup>). We denote the set of all channels from  $\Omega_1$  to  $\Omega_2$  by  $\mathcal{C}(\Omega_1, \Omega_2)$ , and denote the set  $\mathcal{C}(\Omega_1, \Omega_1)$  simply by  $\mathcal{C}(\Omega_1)$

A channel  $T: V_1 \rightarrow V_2$  in the above definition induces a map  $T': V_2^* \rightarrow V_1^*$  such that  $[T'e](\omega) = e(T\omega)$  for all  $e \in \mathcal{E}_{\Omega_2}$  and  $\omega \in \Omega_1$ . In this way, we can focus on transformations between effects instead of transformations between states. However, in this thesis, when channels are considered, they always represent transformations between states, that is, the *Schrodinger picture* is adopted although similar arguments can be developed with channels considered as transformations between effects (the *Heisenberg picture*).

It is easy to obtain the following observations.

**Proposition 2.40**

Let  $(\Omega_1, \mathcal{E}_{\Omega_1})$ ,  $(\Omega_2, \mathcal{E}_{\Omega_2})$ , and  $(\Omega_3, \mathcal{E}_{\Omega_3})$  be GPTs.

- (i) For  $S, T \in \mathcal{C}(\Omega_1, \Omega_2)$ , if we define  $\lambda T + (1 - \lambda)T'$  as  $[\lambda T + (1 - \lambda)T'](\omega_1) = \lambda T(\omega_1) + (1 - \lambda)T'(\omega_1)$  ( $0 \leq \lambda \leq 1$ ), then  $\lambda T + (1 - \lambda)T' \in \mathcal{C}(\Omega_1, \Omega_2)$
- (ii) If  $S \in \mathcal{C}(\Omega_1, \Omega_2)$  and  $T \in \mathcal{C}(\Omega_2, \Omega_3)$ , then  $T \circ S \in \mathcal{C}(\Omega_1, \Omega_3)$ .

Let us give several examples of channels.

*Example 2.41 (Basic examples of channels)*

Let  $(\Omega_1, \mathcal{E}_{\Omega_1})$ ,  $(\Omega_2, \mathcal{E}_{\Omega_2})$  be GPTs, and  $V_1$  and  $V_2$  be the canonical embedding vector spaces of  $\Omega_1$  and  $\Omega_2$  respectively.

- (i) If we define a map  $id_{\Omega_1}: \Omega_1 \rightarrow \Omega_1$  by  $id_{\Omega_1}(\omega_1) = \omega_1$  for all  $\omega_1 \in \Omega_1$ , then  $id_{\Omega_1} \in \mathcal{C}(\Omega_1)$ . We call  $id_{\Omega_1}$  the *identity channel* on  $\Omega_1$ .
- (ii) If we define a map  $T_{\omega^*}: \Omega_1 \rightarrow \Omega_2$  by  $T_{\omega^*}(\omega_1) = \omega^*$  ( $\omega^* \in \Omega_2$ ) for all  $\omega_1 \in \Omega_1$ , then  $T_{\omega^*} \in \mathcal{C}(\Omega_1, \Omega_2)$ .
- (iii) Consider a bipartite system  $(\Omega_{12}, \mathcal{E}_{\Omega_{12}})$  composed of  $(\Omega_1, \mathcal{E}_{\Omega_1})$ ,  $(\Omega_2, \mathcal{E}_{\Omega_2})$ . For the linear maps  $id_{\Omega_1}: V_1 \rightarrow V_1$  and  $u_2: V_2 \rightarrow \mathbb{R}$ , where  $id_{\Omega_1}$  is the identity channel on  $\Omega_1$  and  $u_2$  is the unit effect on  $\Omega_2$ , we define their tensor product  $id_{\Omega_1} \otimes u_2$ . Then,  $id_{\Omega_1} \otimes u_2$  as a linear map from  $V_1 \otimes V_2$  to  $V_1$  is a channel from  $\Omega_{12}$  to  $\Omega_1$ , and called the *partial trace*.

We can demonstrate that even the fundamental notions of states and observables can be represented in terms of channels. To show this, we need to define the following convex sets.

<sup>23</sup>It is sometimes more convenient to consider a linear map  $T: V_1 \rightarrow V_2$  satisfying  $T((V_1)_+) \subset T((V_2)_+)$  and  $u_2(T(\omega)) \leq 1$  ( $\omega \in \Omega_1$ ), where  $u_2$  is the unit effect for  $\Omega_2$ , as representing a transformation of states. Such positive and normalization-nonincreasing maps in GPTs correspond to the notion of *operations* in quantum theory [78], although operations in quantum theory are sometimes assumed also to be completely positive [76].

**Definition 2.42**

Let  $\{x_i\}_{i=1}^{n+1}$  be a set of affinely independent<sup>24</sup> vectors in  $\mathbb{R}^d$  ( $d \geq n$ ). The convex set  $\text{conv}(\{x_i\}_{i=1}^{n+1})$  is called an  $n$ -dimensional simplex [49]. In particular, we denote the simplex generated by orthonormal vectors  $\{p_i\}_{i=1}^{n+1}$  with  $p_1 = (1, 0, \dots, 0)$ ,  $p_2 = (0, 1, 0, \dots, 0), \dots$  simply by  $\Delta_n$ , and call it the  $n$ -dimensional standard simplex. It is trivial that any  $n$ -dimensional simplex is isomorphic to  $\Delta_n$ .

*Example 2.43 (States, observables, and instruments as channels)*

Let  $(\Omega_1, \mathcal{E}_{\Omega_1}), (\Omega_2, \mathcal{E}_{\Omega_2})$  be GPTs, and let us follow the similar notations in Definition 2.42 above.

(i) A state  $\omega \in \Omega_1$  is equivalent to a channel from  $\Delta_1$  to  $\Omega_1$  by the identification of  $\omega$  with a channel  $P_\omega: \Delta_1 \rightarrow \Omega_1$  defined as  $P_\omega(p_1) = \omega$ . Similarly, we can introduce a conditional preparation channel  $P_{\{\omega_i\}_{i=1}^{n+1}} \in \mathcal{C}(\Delta_n, \Omega_1)$  by  $P_{\{\omega_i\}_{i=1}^{n+1}}(v) = \sum_{i=1}^{n+1} v_i \omega_i$ , where  $v_i$  is the  $i$ th element of the vector  $v \in \mathbb{R}^{n+1}$ . The channel  $P_{\{\omega_i\}_{i=1}^{n+1}}$  represents an apparatus which outputs the states  $\{\omega_i\}_{i=1}^{n+1}$  according to the proportion determined by a classical input  $v = (v_1, \dots, v_{n+1})$ .

(ii) An observable  $E = \{e_i\}_{i=1}^{n+1}$  on  $\Omega_1$  with  $(n + 1)$  outcomes is equivalent to a channel from  $\Omega_1$  to  $\Delta_n$  by the identification of  $E$  with a channel  $M_E: \Omega_1 \rightarrow \Delta_n$  defined as  $M_E(\omega) = (e_1(\omega), \dots, e_{n+1}(\omega)) = \sum_{i=1}^{n+1} e_i(\omega) p_i$ .

(iii) For a conditional preparation channel  $P_{\{\omega_i\}_{i=1}^{n+1}} \in \mathcal{C}(\Delta_n, \Omega_2)$  and a measurement channel  $M_E: \Omega_1 \rightarrow \Delta_n$ , the composition  $P_{\{\omega_i\}_{i=1}^{n+1}} \circ M_E \in \mathcal{C}(\Omega_1, \Omega_2)$  is called a *measure-and-prepare channel*. Preparation channels or measurement channels in (i) or (ii) above respectively are examples of measure-and-prepare channels (see [31] for other examples).

(iv) A channel from  $\Omega_1$  to  $\Omega_1 \otimes_{\min} \Delta_n$  is called an *instrument*. It outputs the measurement outcomes of an observable and the ensemble of the post measurement states.

*Remark 2.44*

In this part, we introduce channels in GPTs as positive and normalization-preserving maps, while in quantum theory channels are defined as trace-preserving (normalization-preserving) and completely positive maps [41, 76, 79]. The notion of complete positivity can be introduced also in GPTs based on the above formulation of bipartite systems [31]. However, completely positive maps do not always correspond to physical processes in GPTs. This is because, while in quantum theory all completely positive maps are physically valid transformations in the sense that their physical implementations exist via the Steinspring's theorem [80], there is in general not ensured the existence of such physical implementations in GPTs.

---

<sup>24</sup>Vectors  $v_0, v_1, \dots, v_n$  in a vector space  $V$  are called *affinely independent* if the vectors  $v_1 - v_0, \dots, v_n - v_0$  are linearly independent.

### 2.3.2 Compatibility and incompatibility for channels

In quantum theory, we cannot always obtain simultaneously statistics for a pair of observables such as position and momentum, or cannot always duplicate a family of states [20]. These impossibilities are essential ingredients of quantum theory: for example, without them, the violation of Bell inequality or the security of quantum cryptography never occurs. Those impossibilities can be described by the notion of *incompatibility* in a unified way [21]. In this part, we demonstrate that the notion of incompatibility can be introduced successfully also in GPTs.

#### Definition 2.45

Let  $(\Omega_1, \mathcal{E}_{\Omega_1})$ ,  $(\Omega_2, \mathcal{E}_{\Omega_2})$ , and  $(\Omega_3, \mathcal{E}_{\Omega_3})$  be GPTs, and  $(\Omega_{23}, \mathcal{E}_{\Omega_{23}})$  be a GPT which describe a joint system of  $(\Omega_2, \mathcal{E}_{\Omega_2})$  and  $(\Omega_3, \mathcal{E}_{\Omega_3})$ . Channels  $S \in \mathcal{C}(\Omega_1, \Omega_2)$  and  $T \in \mathcal{C}(\Omega_1, \Omega_3)$  are called *compatible* if there exists a channel  $R \in \mathcal{C}(\Omega_1, \Omega_{23})$  called the *joint channel* of  $S$  and  $T$  such that the marginal actions of  $R$  reproduce each action of  $S$  and  $T$ , that is,

$$\begin{aligned} (id_{\Omega_2} \otimes u_3) \circ R &= S, \\ (u_2 \otimes id_{\Omega_3}) \circ R &= T, \end{aligned}$$

where  $id_{\Omega_2} \otimes u_3$  and  $u_2 \otimes id_{\Omega_3}$  are the partial traces in  $\Omega_{23}$  (see Example 2.41). If  $S$  and  $T$  are not compatible, then they are called *incompatible*.

This definition of incompatibility applies to cases when three or more channels are considered. For incompatibility of observables, we can derive a simpler expression.

#### Proposition 2.46

Let  $(\Omega, \mathcal{E}_{\Omega})$  be a GPT, and  $M_E$  and  $M_F$  be the measurement channels associated with observables  $E = \{e_i\}_{i=1}^l$  and  $F = \{f_j\}_{j=1}^m$  on  $\Omega$  respectively (see Example 2.43 (ii)). Then,  $M_E$  and  $M_F$  are compatible if and only if there exists an observable (called a joint observable)  $G = \{g_{ij}\}_{i=1, j=1}^{l, m}$  on  $\Omega$  such that

$$\sum_{j=1}^m g_{ij} = e_i, \quad \sum_{i=1}^l g_{ij} = f_j.$$

#### Proof

If there exists an observable  $G = \{g_{ij}\}_{i=1, j=1}^{l, m}$  on  $\Omega$  such that

$$\sum_{j=1}^m g_{ij} = e_i, \quad \sum_{i=1}^l g_{ij} = f_j,$$

then it is easy to see that the measurement channel  $M_G \in \mathcal{C}(\Omega, \Delta_{l-1} \otimes_{\min} \Delta_{m-1})$  defined as  $M_G(\omega) = (m_{11}(\omega), \dots, m_{lm}(\omega)) = \sum_{i,j} m_{ij}(\omega) p_i \otimes p_j$ , where

$p_1 = (1, 0, 0, \dots)$ ,  $p_2 = (0, 1, 0, \dots)$ , (see Definition 2.42), is a joint channel of  $M_E$  and  $M_F$ . We note that the composite of two simplices is always given by their minimal tensor product. Conversely, if there exists a joint channel  $M \in \mathcal{C}(\Omega, \Delta_{l-1} \otimes_{\min} \Delta_{m-1})$  of  $M_E$  and  $M_F$ , then, representing  $M(\omega) \in \Delta_{l-1} \otimes_{\min} \Delta_{m-1}$  as  $M(\omega) = \sum M(\omega)_{ij} p_i \otimes p_j$  ( $M(\omega)_{ij} \in [0, 1]$ ), we obtain  $\sum_j M(\omega)_{ij} = e_i(\omega)$  and  $\sum_i M(\omega)_{ij} = f_j(\omega)$ . We can naturally introduce effects  $m_{ij}: \Omega \rightarrow [0, 1]$  by  $m_{ij}(\omega) = M(\omega)_{ij}$ , and it is easy to verify that  $\sum_j m_{ij} = e_i$  and  $\sum_i m_{ij} = f_j$  (and thus  $\sum_{i,j} m_{ij} = u$ , i.e.,  $\{m_{ij}\}_{i,j}$  is an observable).  $\square$

In [81], it was shown that there exists an incompatible pair of observables in every finite-dimensional GPT unless it is classical. We can present the existence of another type of incompatibility.

*Example 2.47 (Generalized no-broadcasting theorem)*

Let  $(\Omega, \mathcal{E}_\Omega)$  be a GPT, and let  $(\Omega_{12}, \mathcal{E}_{\Omega_{12}})$  be a GPT describing a composite system of  $(\Omega_1, \mathcal{E}_{\Omega_1})$  and  $(\Omega_2, \mathcal{E}_{\Omega_2})$ , where  $\Omega_1 = \Omega_2 = \Omega$ . A set of states  $\{\omega_i\}_i \subset \Omega$  is called *broadcastable* if there exists a channel  $T \in \mathcal{C}(\Omega, \Omega_{12})$  such that  $(id_{\Omega_1} \otimes u_2)(T(\omega_i)) = \omega_i$  and  $(u_1 \otimes id_{\Omega_2})(T(\omega_i)) = \omega_i$  hold for all  $i$ . It was shown in [23, 25] (see also [31]) that  $\{\omega_i\}_i \subset \Omega$  is broadcastable if and only if it lies in a simplex. In other words, the identity channels  $id_{\Omega_1}$  and  $id_{\Omega_2}$  are compatible if and only if  $\Omega_1 = \Omega_2 = \Omega$  is a simplex (i.e., the theory is classical).

These results on GPTs manifest interesting facts that properties once thought to be specific to quantum theory are in fact more universal ones.

## 2.4 Additional notions

So far we have reviewed fundamental notions in GPTs especially focusing on states and effects. It was shown that states and effects are represented in terms of ordered Banach spaces, and under the assumption of finite dimensionality, they are reduced to elements of Euclidean spaces. In this part, based on those descriptions, we develop additional notions on states and effects which will play significant roles in demonstrating several results of this thesis. To do this, we follow the notations that have been used so far. That is, a GPT is given by a pair  $(\Omega, \mathcal{E}_\Omega)$  of a state space and the corresponding effect space where  $\Omega \subset V = \mathbb{R}^{N+1}$  with  $\text{span}(\Omega) = V$  and  $0 \notin \Omega$  and  $\mathcal{E}_\Omega \subset V^*$ . We should also recall that the set of all pure states is denoted by  $\Omega^{\text{ext}}$ , and the set of all pure and indecomposable effects by  $\mathcal{E}_\Omega^{\text{ext}}$ .

### 2.4.1 Physical equivalence of pure states

It is known that in quantum theory all pure states are physically equivalent via unitary (and antiunitary) transformations [41]. A similar notion to this physical equivalence of pure states can be introduced also in GPTs.

Let  $\Omega$  be a state space. A map  $T: \Omega \rightarrow \Omega$  is called a *state automorphism* on  $\Omega$  if  $T$  is an affine bijection. We denote the set of all state automorphisms on  $\Omega$  by  $GL(\Omega)$ , and say that a state  $\omega_1 \in \Omega$  is *physically equivalent* to a state  $\omega_2 \in \Omega$  if there exists a  $T \in GL(\Omega)$  such that  $T\omega_1 = \omega_2$ . It was shown in [45] that the physical equivalence of  $\omega_1, \omega_2 \in \Omega$  is equal to the existence of some unit-preserving affine bijection  $T': \mathcal{E}_\Omega \rightarrow \mathcal{E}_\Omega$  satisfying  $e(\omega_1) = T'(e)(\omega_2)$  for all  $e \in \mathcal{E}_\Omega$ , which means  $\omega_1$  and  $\omega_2$  have the same physical contents on measurements. Because any affine map on  $\Omega$  can be extended uniquely to a linear map on  $V$ , it holds that  $GL(\Omega) = \{T: V \rightarrow V \mid T: \text{linear, bijective, } T(\Omega) = \Omega\}$ . It is clear that  $GL(\Omega)$  forms a group, and we can represent the notion of physical equivalence of pure states by means of the transitive action of  $GL(\Omega)$  on  $\Omega^{\text{ext}}$ .

#### Definition 2.48 (Transitive state space)

A state space  $\Omega$  is called *transitive* if  $GL(\Omega)$  acts transitively on  $\Omega^{\text{ext}}$ , that is, for any pair of pure states  $\omega_i^{\text{ext}}, \omega_j^{\text{ext}} \in \Omega^{\text{ext}}$  there exists an affine bijection  $T_{ji} \in GL(\Omega)$  such that  $\omega_j^{\text{ext}} = T_{ji}\omega_i^{\text{ext}}$ .

We remark that the equivalence of pure states does not depend on how the theory is expressed. In fact, when  $\Omega$  is a transitive state space and  $\Omega' := \psi(\Omega)$  is equivalent to  $\Omega$  with a linear bijection  $\psi$ , it is easy to check that  $GL(\Omega') = \psi \circ GL(\Omega) \circ \psi^{-1}$  and  $\Omega'$  is also transitive.

In the remaining of this subsection, we let  $\Omega$  be a transitive state space. In a transitive state space, we can introduce successfully the maximally mixed state as a unique invariant state with respect to every state automorphism.

#### Proposition 2.49 ([82])

*For a transitive state space  $\Omega$ , there exists a unique state  $\omega_M \in \Omega$  (which we call the maximally mixed state) such that  $T\omega_M = \omega_M$  for all  $T \in GL(\Omega)$ . The unique maximally mixed state  $\omega_M$  is given by*

$$\omega_M = \int_{GL(\Omega)} T\omega^{\text{ext}} d\mu(T),$$

*where  $\omega^{\text{ext}}$  is an arbitrary pure state and  $\mu$  is the normalized two-sided invariant Haar measure on  $GL(\Omega)$ .*

Note in Proposition 2.49 that the transitivity of  $\Omega$  guarantees the independence of  $\omega_M$  on the choice of  $\omega^{\text{ext}}$ . When  $\Omega^{\text{ext}}$  is finite and  $\Omega^{\text{ext}} = \{\omega_i^{\text{ext}}\}_{i=1}^n$ ,

$\omega_M$  has a simpler form

$$\omega_M = \frac{1}{n} \sum_{i=1}^n \omega_i^{\text{ext}}.$$

We should recall that the action of the linear bijection  $\eta := \frac{1}{\|\omega_M\|_E} \mathbb{1}_V$  on  $\Omega$  does not change the theory, where  $\|\omega_M\|_E = (\omega_M, \omega_M)_E^{1/2}$  with the standard Euclidean inner product  $(\cdot, \cdot)_E$  and  $\mathbb{1}_V$  is the identity map on  $V$ . Since  $\eta T \eta^{-1} = T$  holds for all  $T \in GL(\Omega)$ ,  $GL(\Omega)$  is invariant under the rescaling by  $\eta$ , i.e.  $GL(\eta(\Omega)) = GL(\Omega)$ . It follows that the unique maximally mixed state of the rescaled state space  $\eta(\Omega)$  is  $\frac{1}{\|\omega_M\|_E} \omega_M$ . In the remaining of this thesis, when a transitive state space is discussed, we apply this rescaling and assume that  $\|\omega_M\|_E = 1$  holds. This assumption makes it easy to prove our main theorems in Chapter 3 via Proposition 2.50 introduced in the following.

The Haar measure  $\mu$  on  $GL(\Omega)$  makes it possible for us to construct a convenient representation of the theory. First of all, we define an inner product  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$  on  $V$  as

$$\langle x, y \rangle_{GL(\Omega)} := \int_{GL(\Omega)} (Tx, Ty)_E d\mu(T) \quad (x, y \in V).$$

Remark that in this thesis we adopt  $(\cdot, \cdot)_E$  as the reference inner product of  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$  although the following discussion still holds even if it is not  $(\cdot, \cdot)_E$ . Thanks to the properties of the Haar measure  $\mu$ , it holds that

$$\langle Tx, Ty \rangle_{GL(\Omega)} = \langle x, y \rangle_{GL(\Omega)} \quad \forall T \in GL(\Omega),$$

which proves that any  $T \in GL(\Omega)$  to be an orthogonal transformation on  $V$  with respect to the inner product  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$ . Therefore, together with the transitivity of  $\Omega$ , we can see that all pure states of  $\Omega$  are of equal norm, that is,

$$\begin{aligned} \|\omega_i^{\text{ext}}\|_{GL(\Omega)} &= \langle \omega_i^{\text{ext}}, \omega_i^{\text{ext}} \rangle_{GL(\Omega)}^{1/2} \\ &= \langle T_{i0} \omega_0^{\text{ext}}, T_{i0} \omega_0^{\text{ext}} \rangle_{GL(\Omega)}^{1/2} \\ &= \langle \omega_0^{\text{ext}}, \omega_0^{\text{ext}} \rangle_{GL(\Omega)}^{1/2} \\ &= \|\omega_0^{\text{ext}}\|_{GL(\Omega)} \end{aligned} \tag{2.18}$$

holds for all  $\omega_i^{\text{ext}} \in \Omega^{\text{ext}}$ , where  $\omega_0^{\text{ext}}$  is an arbitrary reference pure state. We remark that when  $\|\omega_M\|_E = 1$ , we can obtain from the invariance of  $\omega_M$  for



$GL(\Omega)$

$$\begin{aligned}
\|\omega_M\|_{GL(\Omega)}^2 &= \int_{GL(\Omega)} (T\omega_M, T\omega_M)_E d\mu(T) \\
&= \int_{GL(\Omega)} (\omega_M, \omega_M)_E d\mu(T) \\
&= \|\omega_M\|_E^2 \int_{GL(\Omega)} d\mu(T) \\
&= \|\omega_M\|_E^2,
\end{aligned}$$

and thus  $\|\omega_M\|_{GL(\Omega)} = 1$ . The next proposition allows us to give a useful representation of the theory (the proof is given in Appendix A).

**Proposition 2.50**

For a transitive state space  $\Omega$ , there exists a basis  $\{v_l\}_{l=1}^{N+1}$  of  $V$  orthonormal with respect to the inner product  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$  such that  $v_{N+1} = \omega_M$  and

$$x \in \text{aff}(\Omega) \iff x = \sum_{l=1}^N a_l v_l + v_{N+1} = \sum_{l=1}^N a_l v_l + \omega_M \quad (a_1, \dots, a_N \in \mathbb{R}).$$

By employing the representation shown in Proposition 2.50, an arbitrary  $x \in \text{aff}(\Omega)$  can be written as a vector form that

$$x = \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} \quad \text{with} \quad \omega_M = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}, \quad (2.19)$$

where the vector  $\mathbf{x}$  is sometimes called the *Bloch vector* [83, 84] corresponding to  $x$ .

### 2.4.2 Self-duality

In this part, we introduce the notion of self-duality, which also plays an important role in our work.

Let  $V_+$  be the positive cone generated by a state space  $\Omega$ . We define the *internal dual cone* of  $V_+$  relative to an inner product  $(\cdot, \cdot)$  on  $V$  as  $V_{+(\cdot, \cdot)}^{*int} := \{y \in V \mid (x, y) \geq 0, \forall x \in V_+\}$ , which is isomorphic to the dual cone  $V_+^*$  because of the Riesz representation theorem [58].<sup>25</sup> The self-duality of  $V_+$  can be defined as follows.

**Definition 2.51 (Self-duality)**

$V_+$  is called *self-dual* if there exists an inner product  $(\cdot, \cdot)$  on  $V$  such that  $V_+ = V_{+(\cdot, \cdot)}^{*int}$ .

---

<sup>25</sup>In the field of GPTs, effects are often defined as elements of  $V = \mathbb{R}^{N+1}$  through the identification  $V^* = V_{+(\cdot, \cdot)}^{*int}$ , and the action of effects on states is represented via the inner product  $(\cdot, \cdot)$ .

We remark similarly to Definition 2.48 that if  $V_+$  generated by a state space  $\Omega$  is self-dual, then the cone  $V'_+$  generated by  $\Omega' := \psi(\Omega)$  with a linear bijection  $\psi$  (i.e.  $V'_+ = \psi(V_+)$ ) is also self-dual. In fact, we can confirm that if  $V_+ = V_{+(\cdot, \cdot)}^{*int}$  holds for some inner product  $(\cdot, \cdot)$ , then  $V'_+ = V_{+(\cdot, \cdot)'}^{*int}$  holds, where the inner product  $(\cdot, \cdot)'$  is defined as  $(x, y)' = (\psi^{-1}x, \psi^{-1}y)$  ( $x, y \in V$ ).

Let us consider the case when  $\Omega$  is transitive and  $V_+$  is self-dual with respect to the inner product  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$ . Since  $V_+ = V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int}$ , we can regard  $V_+$  also as the set of unnormalized effects. In particular, every pure state  $\omega_i^{\text{ext}} \in \Omega^{\text{ext}}$  can be considered as an unnormalized effect, and if we define

$$e_i := \frac{\omega_i^{\text{ext}}}{\|\omega_i^{\text{ext}}\|_{GL(\Omega)}^2} = \frac{\omega_i^{\text{ext}}}{\|\omega_0^{\text{ext}}\|_{GL(\Omega)}^2}, \quad (2.20)$$

then from Cauchy-Schwarz inequality

$$\langle e_i, \omega_k^{\text{ext}} \rangle_{GL(\Omega)} \leq \|e_i\|_{GL(\Omega)} \|\omega_k^{\text{ext}}\|_{GL(\Omega)} = 1$$

holds for any pure state  $\omega_k^{\text{ext}} \in \Omega^{\text{ext}}$  (thus  $e_i$  is indeed an effect). The equality holds if and only if  $\omega_k^{\text{ext}}$  is parallel to  $e_i$ , i.e.  $\omega_k^{\text{ext}} = \omega_i^{\text{ext}}$ , and we can also conclude that an effect is pure and indecomposable if and only if it is of the form defined as (2.20) together with the fact that effects on the extremal rays of  $V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int} = V_+$  are indecomposable (for more details see [61]):

$$e_i = \frac{\omega_i^{\text{ext}}}{\|\omega_i^{\text{ext}}\|_{GL(\Omega)}^2} = \frac{\omega_i^{\text{ext}}}{\|\omega_0^{\text{ext}}\|_{GL(\Omega)}^2} \equiv e_i^{\text{ext}} \in \mathcal{E}^{\text{ext}}(\Omega). \quad (2.21)$$

When  $|\Omega^{\text{ext}}| < \infty$ , it is sufficient for the discussion above that  $\Omega$  is transitive and self-dual with respect to an arbitrary inner product.

### Proposition 2.52

*Let  $\Omega$  be transitive with  $|\Omega^{\text{ext}}| < \infty$  and  $V_+$  be self-dual with respect to some inner product. There exists a linear bijection  $\Xi: V \rightarrow V$  such that  $\Omega' := \Xi\Omega$  is transitive and the generating positive cone  $V'_+$  is self-dual with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega')}$ , i.e.  $V'_+ = V_{+\langle \cdot, \cdot \rangle_{GL(\Omega')}}^{*int}$ .*

The proof is given in Appendix B. Proposition 2.52 reveals that if a theory with finite pure states is transitive and self-dual, then the theory can be expressed in the way it is self-dual with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$ .

## 2.5 Examples of GPTs

In this section, we present some examples of GPTs with relevant structures to transitivity or self-duality.

### 2.5.1 Classical theories with finite levels

Let us denote by  $\Omega_{\text{CT}}$  the state space of a classical system with a finite level.  $\Omega_{\text{CT}}$  can be represented by means of some finite  $N \in \mathbb{N}$  as the set of all probability distributions (probability vectors)  $\{\mathbf{p} = (p_1, \dots, p_{N+1})\} \subset V = \mathbb{R}^{N+1}$  on some sample space  $\{a_1, \dots, a_{N+1}\}$ , i.e.,  $\Omega_{\text{CT}}$  is the  $N$ -dimensional standard simplex  $\Delta_N$ . It is easy to justify that the set of all pure states  $\Omega_{\text{CT}}^{\text{ext}}$  is given by  $\Omega_{\text{CT}}^{\text{ext}} = \{\mathbf{p}_i^{\text{ext}}\}_{i=1}^{N+1}$ , where  $\mathbf{p}_i^{\text{ext}}$  is the probability distribution satisfying  $(\mathbf{p}_i^{\text{ext}})_j = \delta_{ij}$ , and the positive cone  $V_+$  by  $V_+ = \{\sigma = (\sigma_1, \dots, \sigma_{N+1}) \in V \mid \sigma_i \geq 0, \forall i\}$ . Remark that the set

$$\{\mathbf{p}_i^{\text{ext}}\}_{i=1}^{N+1} = \{(1, 0, \dots, 0), (0, 1, \dots, 0), \dots, (0, 0, \dots, 1)\}$$

forms a standard orthonormal basis of  $V$ . Since any state automorphism maps pure states to pure states, it can be seen that the set  $GL(\Omega_{\text{CT}})$  of all state automorphisms on  $\Omega_{\text{CT}}$  is exactly the set of all permutation matrices with respect to the orthonormal basis  $\{\mathbf{p}_i^{\text{ext}}\}_{i=1}^{N+1}$  of  $V$ . Therefore,  $\Omega_{\text{CT}}$  is a transitive state space, and any  $T \in GL(\Omega_{\text{CT}})$  is orthogonal, which results in

$$\begin{aligned} \langle x, y \rangle_{GL(\Omega_{\text{CT}})} &= \int_{GL(\Omega_{\text{CT}})} (Tx, Ty)_E d\mu(T) \\ &= \int_{GL(\Omega_{\text{CT}})} (x, y)_E d\mu(T) \\ &= (x, y)_E \int_{GL(\Omega_{\text{CT}})} d\mu(T) \\ &= (x, y)_E. \end{aligned} \tag{2.22}$$

The set of all positive linear functionals on  $\Omega_{\text{CT}}$  can be identified with the internal dual cone  $V_{+(\cdot, \cdot)_E}^{*int}$ , and every  $h \in V_{+(\cdot, \cdot)_E}^{*int}$  can be identified with  $h = (h(\mathbf{p}_1^{\text{ext}}), \dots, h(\mathbf{p}_{N+1}^{\text{ext}}))$  with all entries nonnegative since

$$h(\mathbf{p}_i^{\text{ext}}) = (h, \mathbf{p}_i^{\text{ext}})_E = (h)_i \geq 0$$

holds for all  $i$ . Therefore, we can conclude together with (2.22)  $V_+ = V_{+(\cdot, \cdot)_E}^{*int} = V_{+\langle \cdot, \cdot \rangle_{GL(\Omega_{\text{CT}})}}^{*int}$ . Note that we can find the representation (2.19) to be valid for this situation by taking a proper basis of  $V = \mathbb{R}^{N+1}$  and normalization.

### 2.5.2 Quantum theories with finite levels

The state space of a quantum system with a finite level denoted by  $\Omega_{\text{QT}}$  is the set of all density operators on  $N$ -dimensional Hilbert space  $\mathcal{H}$  ( $N < \infty$ ), that is,  $\Omega_{\text{QT}} := \{\rho \in \mathcal{L}_S(\mathcal{H}) \mid \rho \geq 0, \text{Tr}[\rho] = 1\}$ , where  $\mathcal{L}_S(\mathcal{H})$  is the set

of all self-adjoint operators on  $\mathcal{H}$ . The set of all pure states  $\Omega_{\text{QT}}^{\text{ext}}$  is given by the rank-1 projections:  $\Omega_{\text{QT}}^{\text{ext}} = \{|\psi\rangle\langle\psi| \mid |\psi\rangle \in \mathcal{H}, \langle\psi|\psi\rangle = 1\}$ . It has been demonstrated in [85] that with the identity operator  $\mathbb{1}_N$  on  $\mathcal{H}$  and the generators  $\{\sigma_i\}_{i=1}^{N^2-1}$  of  $SU(N)$  satisfying

$$\sigma_i \in \mathcal{L}_S(\mathcal{H}), \quad \text{Tr}[\sigma_i] = 0, \quad \text{Tr}[\sigma_i\sigma_j] = 2\delta_{ij}, \quad (2.23)$$

any  $A \in \mathcal{L}_S(\mathcal{H})$  can be represented as

$$A = c_0\mathbb{1}_N + \sum_{i=1}^{N^2-1} c_i\sigma_i \quad (c_0, c_1, \dots, c_{N^2-1} \in \mathbb{R}) \quad (2.24)$$

and any  $B \in \text{aff}(\Omega_{\text{QT}})$  as

$$B = \frac{1}{N}\mathbb{1}_N + \sum_{i=1}^{N^2-1} c_i\sigma_i \quad (c_1, \dots, c_{N^2-1} \in \mathbb{R}). \quad (2.25)$$

Since (2.23) implies that  $\{\mathbb{1}_N, \sigma_1, \dots, \sigma_{N^2-1}\}$  forms an orthogonal basis of  $\mathcal{L}_S(\mathcal{H})$  with respect to the Hilbert-Schmidt inner product  $(\cdot, \cdot)_{HS}$  defined by

$$(X, Y)_{HS} = \text{Tr}[X^\dagger Y],$$

and (2.24) and (2.25) prove  $\dim(\mathcal{L}_S(\mathcal{H})) = \dim(\text{aff}(\Omega_{\text{QT}})) + 1$ , it seems natural to consider  $\Omega_{\text{QT}}$  to be embedded in  $V = \mathcal{L}_S(\mathcal{H})$  equipped with  $(\cdot, \cdot)_{HS}$ . Because it holds that

$$\begin{aligned} \mathcal{E}(\Omega_{\text{QT}}) &= \{E \in \mathcal{L}_S(\mathcal{H}) \mid 0 \leq \text{Tr}[E\rho] \leq 1, \forall \rho \in \Omega_{\text{QT}}\} \\ &= \{E \in \mathcal{L}_S(\mathcal{H}) \mid 0 \leq E \leq \mathbb{1}_N\}, \end{aligned}$$

we can see  $V_+ = V_{+(\cdot, \cdot)_{HS}}^{*int} = \{A \in \mathcal{L}_S(\mathcal{H}) \mid A \geq 0\}$ , and rank-1 projections are pure and indecomposable effects in quantum theories. We note that while higher dimensional classical theories are represented by simplices as shown in the previous example, higher dimensional quantum theories have more complicated structures [85, 86]: we cannot represent them with higher dimensional balls just generalizing the three dimensional ball for the qubit case (the Bloch ball).

On the other hand, it is known that in quantum theory any state automorphism is either a unitary or antiunitary transformation [41], and for any pair of pure states one can find a unitary operator which links them. Thus,  $\Omega_{\text{QT}}$  is transitive, and any state automorphism is of the form

$$\rho \mapsto U\rho U^\dagger \quad \forall \rho \in \Omega_{\text{QT}},$$

where  $U$  is unitary or antiunitary. Considering that

$$\begin{aligned} (UXU^\dagger, UYU^\dagger)_{HS} &= \text{Tr} [UX^\dagger U^\dagger UYU^\dagger] \\ &= \text{Tr}[X^\dagger Y] \\ &= (X, Y)_{HS} \end{aligned}$$

holds for any unitary or antiunitary operator  $U$ , we can obtain in a similar way to (2.22)

$$\langle X, Y \rangle_{GL(\Omega_{QT})} = (X, Y)_{HS}. \quad (2.26)$$

Therefore, we can conclude  $V_+ = V_{+(\cdot, \cdot)_{HS}}^{*int} = V_{+\langle \cdot, \cdot \rangle_{GL(\Omega_{QT})}}^{*int}$ . We remark similarly to the classical cases that we may rewrite (2.25) as (2.19) by taking a suitable normalization and considering that  $\omega_M = \mathbb{1}_N/N$ .

### 2.5.3 Regular polygon theories

If the state space of a GPT is in the shape of a regular polygon with  $n(\geq 3)$  sides, then we call it a *regular polygon theory* and denote the state space by  $\Omega_n$ . We set  $V = \mathbb{R}^3$  when considering regular polygon theories, and it can be seen in [87] that the pure states of  $\Omega_n$  are described as

$$\Omega_n^{\text{ext}} = \{\omega_i^n\}_{i=0}^{n-1}$$

with

$$\omega_i^n = \begin{pmatrix} r_n \cos(\frac{2\pi i}{n}) \\ r_n \sin(\frac{2\pi i}{n}) \\ 1 \end{pmatrix}, \quad r_n = \sqrt{\frac{1}{\cos(\frac{\pi}{n})}} \quad (2.27)$$

when  $n$  is finite, and when  $n = \infty$  (the state space  $\Omega_\infty$  is a disc),

$$\Omega_\infty^{\text{ext}} = \{\omega_\theta^\infty\}_{\theta \in [0, 2\pi)}$$

with

$$\omega_\theta^\infty = \begin{pmatrix} \cos \theta \\ \sin \theta \\ 1 \end{pmatrix}. \quad (2.28)$$

The state space  $\Omega_3$  represents a classical trit system (the 2-dimensional standard simplex), while  $\Omega_\infty$  represents a qubit system with real coefficients (the unit disc can be considered as an equatorial plane of the Bloch ball). Regular polygon theories can be regarded as intermediate theories of those theories.

The state space of the regular polygon theory with  $n$  sides (including  $n = \infty$ ) defines its positive cone  $V_+$ , and it is also shown in [87] that the corresponding internal dual cone  $V_{+(\cdot, \cdot)_E}^{*int} \subset \mathbb{R}^3$  is given by the conic hull of the following extreme effects (in fact, those effects are also indecomposable)

$$\begin{aligned}
e_i^n &= \frac{1}{2} \begin{pmatrix} r_n \cos(\frac{(2i-1)\pi}{n}) \\ r_n \sin(\frac{(2i-1)\pi}{n}) \\ 1 \end{pmatrix}, \quad i = 0, 1, \dots, n-1 \quad (n : \text{even}) ; \\
e_i^n &= \frac{1}{1+r_n^2} \begin{pmatrix} r_n \cos(\frac{2i\pi}{n}) \\ r_n \sin(\frac{2i\pi}{n}) \\ 1 \end{pmatrix}, \quad i = 0, 1, \dots, n-1 \quad (n : \text{odd}) ; \\
e_\theta^\infty &= \frac{1}{2} \begin{pmatrix} \cos \theta \\ \sin \theta \\ 1 \end{pmatrix}, \quad \theta \in [0, 2\pi) \quad (n = \infty).
\end{aligned} \tag{2.29}$$

Moreover, for finite  $n$ , we can see that the group  $GL(\Omega_n)$  (named the *dihedral group*) is composed of orthogonal transformations with respect to  $(\cdot, \cdot)_E$  [88], which also holds for  $n = \infty$ . Similar calculations to (2.22) or (2.26) demonstrate  $(\cdot, \cdot)_E = \langle \cdot, \cdot \rangle_{GL(\Omega_n)}$  for  $n = 3, 4, \dots, \infty$ . Therefore, from (2.27) - (2.29), we can conclude that  $V_+$  is self-dual, i.e.  $V_+ = V_{+(\cdot, \cdot)_E}^{*int} = V_{+\langle \cdot, \cdot \rangle_{GL(\Omega_n)}}^{*int}$ , when  $n$  is odd or  $\infty$ , while  $V_+$  is not identical but only isomorphic to  $V_{+\langle \cdot, \cdot \rangle_{GL(\Omega_n)}}^{*int}$  when  $n$  is even (in that case,  $V_+$  is called *weakly self-dual* [29, 87]).

Among regular polygon theories, the square theory described by the state space  $\Omega_4$  is physically of particular importance, and is often called a *gbit* (generalized bit) system [24]. It can be observed that the so-called PR-box [69] is represented by a pure entangled state of the composite system  $\Omega_4 \otimes_{max} \Omega_4$  [24], and thus can violate the CHSH inequality maximally in the sense that it attains the value 4 for that entangled state [87]. The square theory is also known for its interesting behavior on incompatibility. It was demonstrated in [89] that a pair of two-outcome observables for  $\Omega_4$  exhibits maximal incompatibility, which means that we need maximal noise to make them compatible (see also Example 3.8).

## Chapter 3

# Preparation uncertainty implies measurement uncertainty in a class of GPTs

Since it was propounded by Heisenberg [8], the existence of uncertainty relations, which is not observed in classical theory, has been regarded as one of the most significant features of quantum theory. The importance of uncertainty relations lies not only in their conceptual aspects but also in practical use such as the security proof of quantum key distribution [11, 90]. There have been researches to capture and formulate the notion of “uncertainty” in several ways. One of the most outstanding works was given by Robertson [91]. There was shown an uncertainty relation in terms of standard derivation which stated that the probability distributions obtained by the measurements of a pair of noncommutative observables cannot be simultaneously sharp. While this type of uncertainty (called *preparation uncertainty*) has been studied also in a more direct way [92, 93, 94] or the entropic way [95, 96, 97, 98, 99, 100], another type of uncertainty called *measurement uncertainty* is known to exist in quantum theory [41]. It describes that when we consider measuring jointly a pair of noncommutative observables, there must exist *measurement error* for the joint measurement, that is, we can only conduct their *approximate joint measurement*. There have been researches on measurement uncertainty with measurement error formulated in terms of standard derivation [101, 102, 103] or entropy [19]. Their measurement uncertainty relations were proven by using preparation uncertainty relations. It implies that there may be a close connection between those two kinds of uncertainty. From this perspective, in [18], simple inequalities were proven which demonstrate in a more explicit way than other previous studies that preparation uncertainty indicates measurement uncertainty and the bound derived from the former also bounds the latter. The main results of [18] were

obtained with preparation uncertainty quantified by overall widths and minimum localization error, and measurement uncertainty by error bar widths, Werner’s measure, and  $l_\infty$  distance [104, 105, 106, 107]. Concerning about uncertainty, both preparation and measurement uncertainty can be introduced naturally also in GPTs. For example, both types of uncertainty for GPTs analogical with a qubit system were investigated in [108], and there are also researches on joint measurability of observables [81, 109, 110, 111, 112], which are related with measurement uncertainty, in GPTs. It is of interest to give further research on how two types of generalized uncertainty are related with each other.

In this part, we study the relations between two kinds of uncertainty in GPTs. We focus on a class of GPTs that are transitive and self-dual including finite-dimensional classical and quantum theories, and demonstrate similar results to [18] in the GPTs: preparation uncertainty relations indicate measurement uncertainty relations. More precisely, it is proven in a certain class of GPTs that if a preparation uncertainty relation gives some bound, then it is also a bound on the corresponding measurement uncertainty relation with the quantifications of uncertainty in [18] generalized to GPTs. We also prove its entropic expression by generalizing the quantum results in [19] to those GPTs. Our results manifest that the close connections between two kinds of uncertainty exhibited in quantum theory are more universal ones. We also present, as an illustration, concrete expressions of our uncertainty relations in regular polygon theories.

This part is organized as follows. In Section 3.1, we introduce measures that quantify the width of a probability distribution. These measures are used for considering whether it is possible to localize jointly two probability distributions obtained by two kinds of measurement on one certain state, that is, they are used for describing preparation uncertainty. We also introduce measures quantifying measurement error by means of which we can formulate measurement uncertainty resulting from approximate joint measurements of two incompatible observables. After the introductions of those quantifications, we present the main theorems and their proofs. In Section 3.2, we demonstrate that similar contents of those theorems can be also expressed in an entropic way. In Section 3.3, we investigate uncertainty relations in regular polygon theories.

### **3.1 Preparation uncertainty and measurement uncertainty in GPTs**

In this section, our main results on the relations between preparation uncertainty and measurement uncertainty are given in GPTs with transitivity and



self-duality with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$  (see Section 2.4). Measures quantifying the width of a probability distribution or measurement error are also given to describe those results. Throughout this section, we consider observables whose sample spaces are finite metric spaces.

### 3.1.1 Widths of probability distributions

In this subsection, we give two kinds of measure to quantify how concentrated a probability distribution is.

Let  $A$  be a finite metric space equipped with a metric function  $d_A$ , and  $O_{d_A}(a; w)$  be the ball defined by  $O_{d_A}(a; w) := \{x \in A \mid d_A(x, a) \leq w/2\}$ . For  $\epsilon \in [0, 1]$  and a probability distribution  $\mathbf{p}$  on  $A$ , we define the *overall width* (at confidence level  $1 - \epsilon$ ) [18, 104] as

$$W_\epsilon(\mathbf{p}) := \inf\{w > 0 \mid \exists a \in A : \mathbf{p}(O_{d_A}(a; w)) \geq 1 - \epsilon\}. \quad (3.1)$$

We can give another formulation for the width of  $\mathbf{p}$ . We define the *minimum localization error* [18] of  $\mathbf{p}$  as

$$LE(\mathbf{p}) := 1 - \max_{a \in A} p(a). \quad (3.2)$$

Both (3.1) and (3.2) can be applied to probability distributions observed in physical experiments. Let us consider a GPT with  $\Omega$  its state space. For a state  $\omega \in \Omega$  and an observable  $F = \{f_a\}_{a \in A}$  on  $A$ , we denote by  $\omega^F$  the probability distribution obtained by the measurements of  $F$  on  $\omega$ , i.e.

$$\omega^F := \{f_a(\omega)\}_{a \in A}.$$

The overall width and minimum localization error for  $\omega^F$  can be defined as

$$W_\epsilon(\omega^F) := \inf\{w > 0 \mid \exists a \in A : \sum_{a' \in O_{d_A}(a; w)} f_{a'}(\omega) \geq 1 - \epsilon\} \quad (3.3)$$

and

$$LE(\omega^F) := 1 - \max_{a \in A} f_a(\omega) \quad (3.4)$$

respectively. Note that as in [18, 104], overall widths can be defined properly even if the sample spaces of probability distributions are infinite. For example, overall widths are considered in [104] for probability measures on  $\mathbb{R}$  derived from the measurement of position or momentum of a particle.

Those two measures above are used for the mathematical description of *preparation uncertainty relations (PURs)*. As a simple example, we consider a qubit system with Hilbert space  $\mathcal{H} = \mathbb{C}^2$ . For two projection-valued measures (PVMs)  $Z = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}$  and  $X = \{|+\rangle\langle +|, |-\rangle\langle -|\}$ , where

$\{|0\rangle, |1\rangle\}$  and  $\{|+\rangle, |-\rangle\} = \{\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\}$  are the  $z$ -basis and  $x$ -basis of  $\mathcal{H}$  respectively, it holds from [93, 99] that

$$LE(\rho^Z) + LE(\rho^X) \geq 1 - \frac{1}{\sqrt{2}} > 0 \quad (3.5)$$

for any state  $\rho$  (see also (3.34)). The inequality (3.5) shows that there is no state  $\rho$  which makes both  $LE(\rho^Z)$  and  $LE(\rho^X)$  zero, that is,  $\rho^Z$  and  $\rho^X$  cannot be localized simultaneously even if the observables are ideal ones (PVMs). PURs in terms of overall widths were also discussed in [104] for the position and momentum observables.

### 3.1.2 Measurement error

In this part, we introduce the concept of measurement error in GPTs, which derives from joint measurement problems, and describe how to quantify it.

Let us consider a GPT with its state space  $\Omega$ , and two observables  $F = \{f_a\}_{a \in A}$  and  $G = \{g_b\}_{b \in B}$  on  $\Omega$ . Although general descriptions of (in)compatibility was already given in Subsection 2.3.2, here we show the definition again. We call  $F$  and  $G$  are *compatible* or *jointly measurable* if there exists a *joint observable*  $M^{FG} = \{m_{ab}^{FG}\}_{(a,b) \in A \times B}$  of  $F$  and  $G$  satisfying

$$\begin{aligned} \sum_{b \in B} m_{ab}^{FG} &= f_a \quad \text{for all } a \in A, \\ \sum_{a \in A} m_{ab}^{FG} &= g_b \quad \text{for all } b \in B, \end{aligned}$$

and if  $F$  and  $G$  are not jointly measurable, then they are called *incompatible* [21, 110]. As was mentioned in Subsection 2.3.2, there exist pairs of observables that are incompatible in all non-classical GPTs, but we can nevertheless conduct their *approximate joint measurements* allowing *measurement error*. Assume that  $F$  and  $G$  are incompatible. It is known that one way to compose their approximate joint measurement is adding some trivial noise to them. To see this, we consider as a simple example the incompatible pair of observables  $Z = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}$  and  $X = \{|+\rangle\langle +|, |-\rangle\langle -|\}$  in a qubit system described in the last subsection. It was demonstrated in [113] that the observables

$$\begin{aligned} \tilde{Z}^\lambda &:= \lambda Z + (1 - \lambda)I \\ &= \left\{ \lambda |0\rangle\langle 0| + \frac{1 - \lambda}{2} \mathbb{1}_2, \lambda |1\rangle\langle 1| + \frac{1 - \lambda}{2} \mathbb{1}_2 \right\}, \\ \tilde{X}^\lambda &:= \lambda X + (1 - \lambda)I, \\ &= \left\{ \lambda |+\rangle\langle +| + \frac{1 - \lambda}{2} \mathbb{1}_2, \lambda |-\rangle\langle -| + \frac{1 - \lambda}{2} \mathbb{1}_2 \right\} \end{aligned} \quad (3.6)$$

are jointly measurable for  $0 \leq \lambda \leq \frac{1}{\sqrt{2}}$ , where  $I := \{\mathbb{1}_2/2, \mathbb{1}_2/2\}$  with  $\mathbb{1}_2$  the identity operator on  $\mathcal{H} = \mathbb{C}^2$  is a trivial observable. The joint measurability of (3.6) implies that the addition of trivial noise described by a trivial observable makes incompatible observables compatible in an approximate way. In fact, it is observed also in GPTs that adding trivial noise results in approximate joint measurements of incompatible observables [110, 111, 113].

Because the notion of measurement error derives from the difference between ideal and approximate observables as discussed above, we have to define ideal observables in GPTs to quantify measurement error. In this chapter, they are defined in an analogical way with the ones in finite-dimensional quantum theories, where PVMs are considered to be ideal [41]. If we denote a PVM by  $E = \{P_a\}_a$ , then each effect is of the form

$$P_a = \sum_{i(a)} |\psi_{i(a)}\rangle\langle\psi_{i(a)}|.$$

In particular, every effect is a sum of pure and indecomposable effects, and we call in a similar way an observable  $F = \{f_a\}_{a \in A}$  on  $\Omega$  *ideal* if each effect  $f_a$  satisfies

$$f_a = \sum_{i(a)} e_{i(a)}^{\text{ext}}, \quad \text{or} \quad f_a = u - \sum_{i(a)} e_{i(a)}^{\text{ext}}, \quad (3.7)$$

where we should recall that the set of all pure and indecomposable effects is denoted by  $\{e_i^{\text{ext}}\}_i$  and we do not consider the trivial observable  $F = \{u\}$ . It is easy to see that observables defined as (3.7) result in PVMs in finite-dimensional quantum theories. This type of observable was considered also in [51].

The introduction of ideal observables makes it possible for us to quantify measurement error. Consider an ideal observable  $F = \{f_a\}_a$  and a general observable  $\tilde{F} = \{\tilde{f}_a\}_a$ , and suppose similarly to the previous subsection that  $A$  is a finite metric space with a metric  $d_A$ .  $F$  may be understood as the measurement intended to be measured, while  $\tilde{F}$  as a measurement conducted actually. Taking into consideration the fact that for each nonzero pure effect there exists at least one state which is mapped to 1 (an “eigenstate” [61]), we can define for  $\epsilon \in [0, 1]$  the *error bar width* of  $\tilde{F}$  relative to  $F$  [18, 104] as

$$\mathcal{W}_\epsilon(\tilde{F}, F) = \inf\{w > 0 \mid \forall a \in A, \forall \omega \in \Omega : \\ f_a(\omega) = 1 \Rightarrow \sum_{a' \in O_{d_A}(a; w)} \tilde{f}_{a'}(\omega) \geq 1 - \epsilon\}. \quad (3.8)$$

$\mathcal{W}_\epsilon(\tilde{F}, F)$  represents the spread of probabilities around the “eigenvalues” of  $F$  observed when the corresponding “eigenstates” of  $F$  are measured by  $\tilde{F}$ ,

and thus it can be thought to be one of the quantifications of measurement error. Note that although error bar widths in general (not necessarily finite) metric spaces were defined in [104], we consider only finite metric spaces in this chapter, so we employ their convenient forms (3.8) in finite metric spaces shown in [18]. Another measure is the one given by Werner [107] as the difference of expectation values of “slowly varying functions” on the probability distributions obtained when  $F$  and  $\tilde{F}$  are measured. It is defined as

$$D_W(\tilde{F}, F) := \sup_{\omega \in \Omega} \sup_{h \in \Lambda} \left| (\tilde{F}[h])(\omega) - (F[h])(\omega) \right|, \quad (3.9)$$

where

$$\Lambda := \{h: A \rightarrow \mathbb{R} \mid |h(a_1) - h(a_2)| \leq d_A(a_1, a_2), \forall a_1, a_2 \in A\}$$

is the set of all “slowly varying functions” (called the *Lipshitz ball* of  $(A, d_A)$ ) and

$$F[h] := \sum_{a \in A} h(a) f_a$$

is a map which gives the expectation value of  $h \in \Lambda$  when  $F$  is measured on a state  $\omega$  (similarly for  $\tilde{F}[h]$ ). There is known a simple relation between (3.8) and (3.9).

**Proposition 3.1** ([18, 104])

Let  $(A, d_A)$  be a finite metric space, and  $F = \{f_a\}_{a \in A}$  and  $\tilde{F} = \{\tilde{f}_a\}_{a \in A}$  be an ideal and general observable respectively. Then,

$$\mathcal{W}_\epsilon(\tilde{F}, F) \leq \frac{2}{\epsilon} D_W(\tilde{F}, F)$$

holds for  $\epsilon \in (0, 1]$ .

**Proof**

Let us define  $n := \frac{D_W(\tilde{F}, F)}{\epsilon}$  for  $\epsilon \in (0, 1]$ , and consider for  $a \in A$  a state  $\omega \in \Omega$  satisfying  $f_a(\omega) = 1$ . Remember that such state does exist for every  $a \in A$  because  $F$  is ideal. We also define a function  $h_n$  on  $A$  as

$$h_n(x) := \begin{cases} n - d_A(x, a) & (d(x, a) \leq n) \\ 0 & (d(x, a) > n). \end{cases}$$

It can be seen that

$$|h_n(x_1) - h_n(x_2)| \leq d_A(x_1, x_2)$$

holds for  $x_1, x_2 \in A$ , and thus we can obtain from the definition of  $D_W(\tilde{F}, F)$  (3.9)

$$\left| (\tilde{F}[h_n])(\omega) - (F[h_n])(\omega) \right| \leq D_W(\tilde{F}, F).$$

It results in

$$\left| (\tilde{F}[g_n])(\omega) - (F[g_n])(\omega) \right| \leq \frac{D_W(\tilde{F}, F)}{n} = \epsilon, \quad (3.10)$$

where we set  $g_n := h_n/n$ . Since it holds that  $g_n(x) \leq \chi_{O_{d_A}(a; 2n)}(x) \leq 1$  for all  $x \in A$ , where  $\chi_{O_{d_A}(a; 2n)}$  is the indicator function of the ball  $O_{d_A}(a; 2n) = \{x \in A \mid d_A(x, a) \leq n\}$ , and

$$(F[g_n])(\omega) = \sum_{x \in A} g_n(x) f_x(\omega) = g_n(a) f_a(\omega) = 1$$

because  $f_a(\omega) = 1$ , (3.10) can be rewritten as

$$1 - (\tilde{F}[\chi_{O_{d_A}(a; 2n)}])(\omega) \leq \epsilon,$$

that is,

$$\sum_{x \in O_{d_A}(a; 2n)} \tilde{f}_x(\omega) \geq 1 - \epsilon. \quad (3.11)$$

(3.11) holds for all  $a \in A$  and all  $\omega \in \Omega$  such that  $f_a(\omega) = 1$ , and thus

$$2n = \frac{2}{\epsilon} D_W(\tilde{F}, F) \geq \mathcal{W}_\epsilon(\tilde{F}, F)$$

is concluded (see the definition of  $\mathcal{W}_\epsilon(\tilde{F}, F)$  (3.8)).  $\square$

On the other hand, there can be introduced a more intuitive quantification of measurement error called *l<sub>∞</sub> distance* [106]:

$$D_\infty(\tilde{F}, F) := \sup_{\omega \in \Omega} \max_{a \in A} \left| \tilde{f}_a(\omega) - f_a(\omega) \right|. \quad (3.12)$$

By means of those quantifications of measurement error above, we can formulate *measurement uncertainty relations (MURs)*. As an illustration, we again consider the joint measurement problem of incompatible observables  $Z$  and  $X$  in a qubit system. Suppose that  $\tilde{M}^{ZX}$  is an approximate joint observable of  $Z$  and  $X$ , and  $\tilde{M}^Z$  and  $\tilde{M}^X$  are its marginal observables corresponding to  $Z$  and  $X$  respectively. It was proven in [106] that

$$D_\infty(\tilde{M}^Z, Z) + D_\infty(\tilde{M}^X, X) \geq 1 - \frac{1}{\sqrt{2}} > 0. \quad (3.13)$$

(3.13) gives a quantitative representation of the incompatibility of  $Z$  and  $X$  that  $D_\infty(\tilde{M}^Z, Z)$  and  $D_\infty(\tilde{M}^X, X)$  cannot be simultaneously zero, that is, measurement error must occur when conducting any approximate joint measurement of  $Z$  and  $X$  (see [105] for another inequality). MURs for the position and momentum observables were given in [104] and [107] in terms

of (3.8) and (3.9) respectively.

### 3.1.3 Relations between preparation uncertainty and measurement uncertainty in a class of GPTs

In the previous subsections, we have introduced several measures to review two kinds of uncertainty, preparation uncertainty and measurement uncertainty. In this part, we shall manifest as our main results how they are related with each other in GPTs, which is a generalization of the quantum ones in [18].

Before demonstrating our main theorems, we confirm the physical settings and mathematical assumptions to state them. In the following, we focus on a GPT with a state space  $\Omega$ , and suppose that  $\Omega$  is transitive and the positive cone  $V_+$  is self-dual with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$  (see Section 2.4). While our assumptions may seem curious, it can be observed in [84] that those two conditions are satisfied simultaneously if the state space is *bit-symmetric*. There are also researches where they are derived from certain conditions possible to be interpreted physically [51, 114]. In addition, we consider ideal observables  $F = \{f_a\}_{a \in A}$  and  $G = \{g_b\}_{b \in B}$  on  $\Omega$ , whose sample spaces are finite metric spaces  $(A, d_A)$  and  $(B, d_B)$  respectively, and consider an observable  $\widetilde{M}^{FG} := \{\widetilde{m}_{ab}^{FG}\}_{(a,b) \in A \times B}$  as an approximate joint observable of  $F$  and  $G$ , whose marginal observables are given by

$$\begin{aligned} \widetilde{M}^F &:= \{\widetilde{m}_a^F\}_a, & \widetilde{m}_a^F &:= \sum_{b \in B} \widetilde{m}_{ab}^{FG}; \\ \widetilde{M}^G &:= \{\widetilde{m}_b^G\}_b, & \widetilde{m}_b^G &:= \sum_{a \in A} \widetilde{m}_{ab}^{FG}. \end{aligned}$$

Remember that, as shown in Subsection 3.1.2, the ideal observable  $F = \{f_a\}_a$  satisfies

$$f_a = \sum_{i(a)} e_{i(a)}^{\text{ext}}, \quad \text{or} \quad f_a = u - \sum_{i(a)} e_{i(a)}^{\text{ext}} \quad (3.14)$$

in terms of the pure and indecomposable effects  $\{e_i^{\text{ext}}\}_i$  shown in (2.21) (similarly for  $G = \{g_b\}_b$ ). The following lemmas are needed to prove our main results.

#### Lemma 3.2

If  $\Omega$  is transitive, then the unit effect  $u \in V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*\text{int}} \subset V$  is identical to the maximally mixed state  $\omega_M$ , i.e.  $u = \omega_M$ .

#### Proof

It is an easy consequence of Proposition 2.50. In fact, (2.19) gives

$$u = \omega_M = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}.$$

□

**Lemma 3.3**

If  $\Omega$  is a transitive state space and its positive cone  $V_+$  is self-dual with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$ , then for any effect  $e \in \mathcal{E}(\Omega)$  on  $\Omega$  it holds that

$$\frac{e}{\langle u, e \rangle_{GL(\Omega)}} \in \Omega, \quad (3.15)$$

and for any ideal observable  $F = \{f_a\}_{a \in A}$  on  $\Omega$  it holds that

$$\left\langle f_a, \frac{f_a}{\langle u, f_a \rangle_{GL(\Omega)}} \right\rangle_{GL(\Omega)} = 1 \quad (3.16)$$

for all  $a \in A$ . In particular, each  $f_a / \langle u, f_a \rangle_{GL(\Omega)}$  is an “eigenstate” of  $F$ .

**Proof**

In this proof, we denote the inner product  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$  and the norm  $\| \cdot \|_{GL(\Omega)}$  simply by  $\langle \cdot, \cdot \rangle$  and  $\| \cdot \|$  respectively.

For any element  $e \in V_{+\langle \cdot, \cdot \rangle}^{*\text{int}}$ ,  $e / \langle u, e \rangle$  defines a state because  $\langle u, e / \langle u, e \rangle \rangle = 1$  and  $e \in V_+$  due to the self-duality:  $V_+ = V_{+\langle \cdot, \cdot \rangle}^{*\text{int}}$ , which proves (3.15). To prove (3.16), we focus on the fact that  $f_a$  in (3.14) is an effect (thus  $u - f_a$  is also an effect), that is,  $\sum_{i(a)} e_{i(a)}^{\text{ext}}$  is an effect and it satisfies  $0 \leq \langle \sum_{i(a)} e_{i(a)}^{\text{ext}}, \omega \rangle \leq 1$  for any state  $\omega \in \Omega$ . However, if we act  $\sum_{i(a)} e_{i(a)}^{\text{ext}}$  on the pure state  $\omega_{j(a)}^{\text{ext}}$ , then (2.21) shows that  $\langle e_{j(a)}^{\text{ext}}, \omega_{j(a)}^{\text{ext}} \rangle = 1$ , and thus we have

$$\langle e_{i(a)}^{\text{ext}}, \omega_{j(a)}^{\text{ext}} \rangle = 0 \quad \text{for } i(a) \neq j(a),$$

that is,

$$\langle e_{i(a)}^{\text{ext}}, e_{j(a)}^{\text{ext}} \rangle = 0 \quad \text{for } i(a) \neq j(a). \quad (3.17)$$

Because

$$\langle e_{i(a)}^{\text{ext}}, e_{i(a)}^{\text{ext}} \rangle = \frac{1}{\|\omega_0^{\text{ext}}\|^2} \quad \text{and} \quad \langle u, e_{i(a)}^{\text{ext}} \rangle = \frac{1}{\|\omega_0^{\text{ext}}\|^2}$$

hold from (2.21), we obtain together with (3.17)

$$\begin{aligned} \left\langle \sum_{i(a)} e_{i(a)}^{\text{ext}}, \sum_{i(a)} e_{i(a)}^{\text{ext}} \right\rangle &= \frac{(\#i(a))}{\|\omega_0^{\text{ext}}\|^2}, & \langle u, \sum_{i(a)} e_{i(a)}^{\text{ext}} \rangle &= \frac{(\#i(a))}{\|\omega_0^{\text{ext}}\|^2}, \\ \left\langle u - \sum_{i(a)} e_{i(a)}^{\text{ext}}, u - \sum_{i(a)} e_{i(a)}^{\text{ext}} \right\rangle &= 1 - \frac{(\#i(a))}{\|\omega_0^{\text{ext}}\|^2}, & \langle u, u - \sum_{i(a)} e_{i(a)}^{\text{ext}} \rangle &= 1 - \frac{(\#i(a))}{\|\omega_0^{\text{ext}}\|^2}, \end{aligned} \quad (3.18)$$

where  $(\#i(a))$  is the number of elements of the index set  $\{i(a)\}$  and we use  $\langle u, u \rangle = \langle u, \omega_M \rangle = 1$  (Lemma 3.2). Therefore, we can conclude that every

effect  $f_a = \sum_{i(a)} e_{i(a)}^{\text{ext}}$  or  $u - \sum_{i(a)} e_{i(a)}^{\text{ext}}$  composing  $F$  satisfies

$$\left\langle f_a, \frac{f_a}{\langle u, f_a \rangle} \right\rangle = 1.$$

□

Now, we can state our main theorems connecting PURs and MURs. Similar results to ours were proven [18] for finite-dimensional quantum theories. Because GPTs shown above include those theories, our theorems can be considered to demonstrate that the relations between PURs and MURs introduced in [18] are more general ones.

### Theorem 3.4

Let  $\Omega$  be a transitive state space and its positive cone  $V_+$  be self-dual with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$ , and let  $(F, G)$  be a pair of ideal observables on  $\Omega$ . For an arbitrary approximate joint observable  $\widetilde{M}^{FG}$  of  $(F, G)$  and  $\epsilon_1, \epsilon_2 \in [0, 1]$  satisfying  $\epsilon_1 + \epsilon_2 \leq 1$ , there exists a state  $\omega \in \Omega$  such that

$$\begin{aligned} \mathcal{W}_{\epsilon_1}(\widetilde{M}^F, F) &\geq W_{\epsilon_1 + \epsilon_2}(\omega^F), \\ \mathcal{W}_{\epsilon_2}(\widetilde{M}^G, G) &\geq W_{\epsilon_1 + \epsilon_2}(\omega^G). \end{aligned}$$

Theorem 3.4 manifests that if one cannot make both  $W_{\epsilon_1 + \epsilon_2}(\omega^F)$  and  $W_{\epsilon_1 + \epsilon_2}(\omega^G)$  vanish, then one also cannot make both  $\mathcal{W}_{\epsilon_1}(\widetilde{M}^F, F)$  and  $\mathcal{W}_{\epsilon_2}(\widetilde{M}^G, G)$  vanish. That is, if there exists a PUR, then there also exists a MUR. Moreover, Theorem 3.4 also demonstrates that bounds for MURs in terms of error bar widths can be given by ones for PURs described by overall widths.

### Proof (Proof of Theorem 3.4)

In this proof, we denote again the inner product  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$  and the norm  $\|\cdot\|_{GL(\Omega)}$  simply by  $\langle \cdot, \cdot \rangle$  and  $\|\cdot\|$  respectively.

From Lemma 3.3 and the definition of  $\mathcal{W}_{\epsilon_1}(\widetilde{M}^F, F)$  (3.8), for any  $w_1 \geq \mathcal{W}_{\epsilon_1}(\widetilde{M}^F, F)$  we have

$$\sum_{a' \in O_{d_A}(a; w_1)} \left\langle \widetilde{m}_{a'}^F, \frac{f_a}{\langle u, f_a \rangle} \right\rangle \geq 1 - \epsilon_1,$$

equivalently,

$$\sum_{b' \in B} \sum_{a' \in O_{d_A}(a; w_1)} \left\langle \widetilde{m}_{a'b'}^{FG}, \frac{f_a}{\langle u, f_a \rangle} \right\rangle \geq 1 - \epsilon_1$$

for all  $a \in A$ . Multiplying both sides by  $\langle u, f_a \rangle = \langle \omega_M, f_a \rangle (> 0)$  (Lemma 3.2) and taking the summation over  $a$  yield

$$\sum_{a \in A} \sum_{b' \in B} \sum_{a' \in O_{d_A}(a; w_1)} \langle \widetilde{m}_{a'b'}^{FG}, f_a \rangle \geq 1 - \epsilon_1, \quad (3.19)$$



where we use the relation  $\sum_{a \in A} \langle u, f_a \rangle = \langle u, u \rangle = \langle u, \omega_M \rangle = 1$ . Defining a function  $\chi_{[d_A, w_1]}$  on  $A \times A$  such that

$$\chi_{[d_A, w_1]}(a, a') = \begin{cases} 1 & (d_A(a, a') \leq \frac{w_1}{2}) \\ 0 & (d_A(a, a') > \frac{w_1}{2}), \end{cases}$$

it holds that

$$\begin{aligned} \sum_{a \in A} \sum_{a' \in O_{d_A}(a; w_1)} \langle \tilde{m}_{a'b'}^{FG}, f_a \rangle &= \sum_{(a, a') \in A \times A} \chi_{[d_A, w_1]}(a, a') \langle \tilde{m}_{a'b'}^{FG}, f_a \rangle \\ &= \sum_{a' \in A} \sum_{a \in O_{d_A}(a'; w_1)} \langle \tilde{m}_{a'b'}^{FG}, f_a \rangle \end{aligned}$$

because of the symmetric action of  $\chi_{[d_A, w_1]}$  on  $a$  and  $a'$ . Therefore, (3.19) can be rewritten as

$$\sum_{a' \in A} \sum_{b' \in B} \sum_{a \in O_{d_A}(a'; w_1)} \langle \tilde{m}_{a'b'}^{FG}, f_a \rangle \geq 1 - \epsilon_1.$$

Overall, we obtain

$$\sum_{a' \in A} \sum_{b' \in B} \sum_{a \in O_{d_A}(a'; w_1)} \langle u, \tilde{m}_{a'b'}^{FG} \rangle \left\langle f_a, \frac{\tilde{m}_{a'b'}^{FG}}{\langle u, \tilde{m}_{a'b'}^{FG} \rangle} \right\rangle \geq 1 - \epsilon_1. \quad (3.20)$$

Similar calculations show that for any  $w_2 \geq \mathcal{W}_{\epsilon_2}(\tilde{M}^G, G)$

$$\sum_{a' \in A} \sum_{b' \in B} \sum_{b \in O_{d_B}(b'; w_2)} \langle u, \tilde{m}_{a'b'}^{FG} \rangle \left\langle g_b, \frac{\tilde{m}_{a'b'}^{FG}}{\langle u, \tilde{m}_{a'b'}^{FG} \rangle} \right\rangle \geq 1 - \epsilon_2 \quad (3.21)$$

holds. We obtain from (3.20) and (3.21)

$$\begin{aligned} \sum_{a' \in A} \sum_{b' \in B} \langle u, \tilde{m}_{a'b'}^{FG} \rangle &\left[ \left( \sum_{a \in O_{d_A}(a'; w_1)} \left\langle f_a, \frac{\tilde{m}_{a'b'}^{FG}}{\langle u, \tilde{m}_{a'b'}^{FG} \rangle} \right\rangle \right) \right. \\ &\left. + \left( \sum_{b \in O_{d_B}(b'; w_2)} \left\langle g_b, \frac{\tilde{m}_{a'b'}^{FG}}{\langle u, \tilde{m}_{a'b'}^{FG} \rangle} \right\rangle \right) \right] \geq 2 - \epsilon_1 - \epsilon_2, \end{aligned}$$

which implies that there exists a  $(a'_0, b'_0) \in A \times B$  such that

$$\begin{aligned} & \left( \sum_{a \in O_{d_A}(a'_0; w_1)} \left\langle f_a, \frac{\tilde{m}_{a'_0 b'_0}^{FG}}{\langle u, \tilde{m}_{a'_0 b'_0}^{FG} \rangle} \right\rangle \right) \\ & + \left( \sum_{b \in O_{d_B}(b'_0; w_2)} \left\langle g_b, \frac{\tilde{m}_{a'_0 b'_0}^{FG}}{\langle u, \tilde{m}_{a'_0 b'_0}^{FG} \rangle} \right\rangle \right) \geq 2 - \epsilon_1 - \epsilon_2 \end{aligned} \quad (3.22)$$

since  $\sum_{a' \in A} \sum_{b' \in B} \langle u, \tilde{m}_{a' b'}^{FG} \rangle = \langle u, u \rangle = 1$  and  $0 \leq \langle u, \tilde{m}_{a' b'}^{FG} \rangle \leq 1$  for all  $(a', b') \in A \times B$ . We can see from (3.22) that

$$\begin{aligned} & \sum_{a \in O_{d_A}(a'_0; w_1)} \left\langle f_a, \frac{\tilde{m}_{a'_0 b'_0}^{FG}}{\langle u, \tilde{m}_{a'_0 b'_0}^{FG} \rangle} \right\rangle \geq 1 - \epsilon_1 - \epsilon_2 \\ & + \left( 1 - \sum_{b \in O_{d_B}(b'_0; w_2)} \left\langle g_b, \frac{\tilde{m}_{a'_0 b'_0}^{FG}}{\langle u, \tilde{m}_{a'_0 b'_0}^{FG} \rangle} \right\rangle \right) \\ & \geq 1 - \epsilon_1 - \epsilon_2 \end{aligned} \quad (3.23)$$

holds for an arbitrary  $w_1 \geq \mathcal{W}_{\epsilon_1}(\tilde{M}^F, F)$ , where we use

$$\sum_{b \in O_{d_B}(b'_0; w_2)} \left\langle g_b, \frac{\tilde{m}_{a'_0 b'_0}^{FG}}{\langle u, \tilde{m}_{a'_0 b'_0}^{FG} \rangle} \right\rangle \leq \sum_{b \in B} \left\langle g_b, \frac{\tilde{m}_{a'_0 b'_0}^{FG}}{\langle u, \tilde{m}_{a'_0 b'_0}^{FG} \rangle} \right\rangle = 1,$$

and similarly

$$\sum_{b \in O_{d_B}(b'_0; w_2)} \left\langle g_b, \frac{\tilde{m}_{a'_0 b'_0}^{FG}}{\langle u, \tilde{m}_{a'_0 b'_0}^{FG} \rangle} \right\rangle \geq 1 - \epsilon_1 - \epsilon_2 \quad (3.24)$$

holds for an arbitrary  $w_2 \geq \mathcal{W}_{\epsilon_2}(\tilde{M}^G, G)$ . Because

$$\omega'_0 := \frac{\tilde{m}_{a'_0 b'_0}^{FG}}{\langle u, \tilde{m}_{a'_0 b'_0}^{FG} \rangle}$$

defines a state ((3.15) in Lemma 3.3), (3.23) and (3.24) together with the definition of the overall width (3.3) result in

$$\begin{aligned} w_1 & \geq W_{\epsilon_1 + \epsilon_2}(\omega'_0{}^F), \\ w_2 & \geq W_{\epsilon_1 + \epsilon_2}(\omega'_0{}^G). \end{aligned}$$

These equations hold for any  $w_1 \geq \mathcal{W}_{\epsilon_1}(\tilde{M}^F, F)$  and  $w_2 \geq \mathcal{W}_{\epsilon_2}(\tilde{M}^G, G)$ , so

we finally obtain

$$\begin{aligned}\mathcal{W}_{\epsilon_1}(\widetilde{M}^F, F) &\geq W_{\epsilon_1+\epsilon_2}(\omega_0^F) \\ \mathcal{W}_{\epsilon_2}(\widetilde{M}^G, G) &\geq W_{\epsilon_1+\epsilon_2}(\omega_0^G).\end{aligned}$$

□

The next corollary results immediately from Proposition 3.1. It describes a similar content to Theorem 3.4 in terms of another measure.

**Corollary 3.5**

Let  $\Omega$  be a transitive state space and its positive cone  $V_+$  be self-dual with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$ , and let  $(F, G)$  be a pair of ideal observables on  $\Omega$ . For an arbitrary approximate joint observable  $\widetilde{M}^{FG}$  of  $(F, G)$  and  $\epsilon_1, \epsilon_2 \in (0, 1]$  satisfying  $\epsilon_1 + \epsilon_2 \leq 1$ , there exists a state  $\omega \in \Omega$  such that

$$\begin{aligned}D_W(\widetilde{M}^F, F) &\geq \frac{\epsilon_1}{2} W_{\epsilon_1+\epsilon_2}(\omega^F), \\ D_W(\widetilde{M}^G, G) &\geq \frac{\epsilon_2}{2} W_{\epsilon_1+\epsilon_2}(\omega^G).\end{aligned}$$

There is also another formulation by means of minimum localization error and  $l_\infty$  distance.

**Theorem 3.6**

Let  $\Omega$  be a transitive state space and its positive cone  $V_+$  be self-dual with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$ , and let  $(F, G)$  be a pair of ideal observables on  $\Omega$ . For an arbitrary approximate joint observable  $\widetilde{M}^{FG}$  of  $(F, G)$ , there exists a state  $\omega \in \Omega$  such that

$$D_\infty(\widetilde{M}^F, F) + D_\infty(\widetilde{M}^G, G) \geq LE(\omega^F) + LE(\omega^G).$$

**Proof**

We can see from (3.16) in Lemma 3.3 and the definition of the  $l_\infty$  distance (3.12) that

$$\left| \left\langle f_a, \frac{f_a}{\langle u, f_a \rangle} \right\rangle - \left\langle \widetilde{m}_a^F, \frac{f_a}{\langle u, f_a \rangle} \right\rangle \right| \leq D_\infty(\widetilde{M}^F, F)$$

holds for all  $a \in A$ , which can be rewritten as

$$1 - \sum_{b \in B} \left\langle \widetilde{m}_{ab}^{FG}, \frac{f_a}{\langle u, f_a \rangle} \right\rangle \leq D_\infty(\widetilde{M}^F, F),$$

for all  $a \in A$ . Multiplying both sides by  $\langle u, f_a \rangle$  and taking the summation over  $a$ , we have

$$1 - \sum_{a \in A} \sum_{b \in B} \langle \widetilde{m}_{ab}^{FG}, f_a \rangle \leq D_\infty(\widetilde{M}^F, F),$$

namely

$$1 - \sum_{a' \in A} \sum_{b' \in B} \langle u, \tilde{m}_{a'b'}^{FG} \rangle \left\langle f_{a'}, \frac{\tilde{m}_{a'b'}^{FG}}{\langle u, \tilde{m}_{a'b'}^{FG} \rangle} \right\rangle \leq D_\infty(\tilde{M}^F, F) \quad (3.25)$$

In a similar way, we also have

$$1 - \sum_{a' \in A} \sum_{b' \in B} \langle u, \tilde{m}_{a'b'}^{FG} \rangle \left\langle g_{b'}, \frac{\tilde{m}_{a'b'}^{FG}}{\langle u, \tilde{m}_{a'b'}^{FG} \rangle} \right\rangle \leq D_\infty(\tilde{M}^G, G). \quad (3.26)$$

Since  $\sum_{a' \in A} \sum_{b' \in B} \langle u, \tilde{m}_{a'b'}^{FG} \rangle = 1$ , (3.25) and (3.26) give

$$\begin{aligned} \sum_{a' \in A} \sum_{b' \in B} \langle u, \tilde{m}_{a'b'}^{FG} \rangle \left[ \left( 1 - \left\langle f_{a'}, \frac{\tilde{m}_{a'b'}^{FG}}{\langle u, \tilde{m}_{a'b'}^{FG} \rangle} \right\rangle \right) + \left( 1 - \left\langle g_{b'}, \frac{\tilde{m}_{a'b'}^{FG}}{\langle u, \tilde{m}_{a'b'}^{FG} \rangle} \right\rangle \right) \right] \\ \leq D_\infty(\tilde{M}^F, F) + D_\infty(\tilde{M}^G, G), \end{aligned}$$

which indicates that there exists a  $(a'_0, b'_0) \in A \times B$  satisfying

$$\begin{aligned} \left( 1 - \left\langle f_{a'_0}, \frac{\tilde{m}_{a'_0 b'_0}^{FG}}{\langle u, \tilde{m}_{a'_0 b'_0}^{FG} \rangle} \right\rangle \right) + \left( 1 - \left\langle g_{b'_0}, \frac{\tilde{m}_{a'_0 b'_0}^{FG}}{\langle u, \tilde{m}_{a'_0 b'_0}^{FG} \rangle} \right\rangle \right) \\ \leq D_\infty(\tilde{M}^F, F) + D_\infty(\tilde{M}^G, G). \quad (3.27) \end{aligned}$$

Because

$$\omega'_0 := \frac{\tilde{m}_{a'_0 b'_0}^{FG}}{\langle u, \tilde{m}_{a'_0 b'_0}^{FG} \rangle}$$

is a state ((3.15) in Lemma 3.3), we can conclude from (3.27) and the definition of the minimum localization error (3.4) that

$$LE(\omega'^F_0) + LE(\omega'^G_0) \leq D_\infty(\tilde{M}^F, F) + D_\infty(\tilde{M}^G, G),$$

which proves the theorem.  $\square$

It is easy to see from the proofs that our theorems can be generalized to the case when three or more observables are considered.

*Remark 3.7*

It was claimed in [108] similarly to our theorems that PURs imply MURs in GPTs. However, the result in [108] was obtained for a pair of binary (i.e. two-outcome), extreme, sharp, and *postprocessing clean* [115] observables. It is known that any effect of a sharp and postprocessing clean observable is pure and indecomposable, and such observables do not always exist for a GPT [115, 61]. The only finite-dimensional quantum theory admitting those observables is a qubit system (remember that pure and indecomposable effects correspond to rank-1 projections in finite-dimensional quantum

theories). On the other hand, although our GPTs are assumed to be transitive and self-dual, or regular polygon theories, our theorems are obtained for more general forms of observables (3.7) always possible to be defined.

Theorem 3.6 (and Theorem 3.12) has an application to evaluate the *degree of incompatibility* [110, 111, 113] of a GPT.

*Example 3.8 (Evaluation of degree of incompatibility)*

Suppose that  $\Omega$  is an arbitrary state space, and  $F$  and  $G$  are two-outcome observables on  $\Omega$ , namely  $F = \{f_0, f_1\}$  and  $G = \{g_0, g_1\}$ , and consider similarly to (3.6) their “fuzzy” versions

$$\begin{aligned}\tilde{F}^\lambda &:= \lambda F + (1 - \lambda) \left\{ \frac{u}{2}, \frac{u}{2} \right\} = \left\{ \lambda f_0 + \frac{1 - \lambda}{2} u, \lambda f_1 + \frac{1 - \lambda}{2} u \right\}, \\ \tilde{G}^\lambda &:= \lambda G + (1 - \lambda) \left\{ \frac{u}{2}, \frac{u}{2} \right\} = \left\{ \lambda g_0 + \frac{1 - \lambda}{2} u, \lambda g_1 + \frac{1 - \lambda}{2} u \right\}\end{aligned}\quad (3.28)$$

for  $\lambda \in [0, 1]$ . It is known that we can find a  $\lambda_{F,G} \geq \frac{1}{2}$  such that the distorted observables  $\tilde{F}^\lambda$  and  $\tilde{G}^\lambda$  in (3.28) are jointly measurable for any  $\lambda \in [0, \lambda_{F,G}]$ , and  $\lambda_{\text{opt}} := \inf_{F,G} \lambda_{F,G}$  can be thought describing the degree of incompatibility of the theory.  $\lambda_{\text{opt}}$  has been calculated in various theories: for example,  $\lambda_{\text{opt}} = \frac{1}{\sqrt{2}}$  in finite-dimensional quantum theories [113], and  $\lambda_{\text{opt}} = \frac{1}{2}$  in the square theory (a regular polygon theory with  $n = 4$ ) [89].

To see how Theorem 3.6 contributes to the degree of incompatibility, we consider the situations in Theorem 3.6 (and Theorem 3.12) with the marginals  $\tilde{M}^F$  and  $\tilde{M}^G$  of the approximate joint observable being  $\tilde{F}^\lambda$  and  $\tilde{G}^\lambda$  in (3.28) for  $\lambda \in [0, \lambda_{F,G}]$  respectively. In this case, we can represent the measurement error  $D_\infty(\tilde{F}^\lambda, F)$  in a more explicit way:

$$\begin{aligned}D_\infty(\tilde{F}^\lambda, F) &= \sup_{\omega \in \Omega} \max_{i \in \{0,1\}} \left| \left( \lambda f_i + \frac{1 - \lambda}{2} u \right) (\omega) - f_i(\omega) \right| \\ &= (1 - \lambda) \sup_{\omega \in \Omega} \max_{i \in \{0,1\}} \left| f_i(\omega) - \frac{1}{2} \right| \\ &= \frac{1 - \lambda}{2},\end{aligned}\quad (3.29)$$

where we use the relation

$$\left| f_0(\omega) - \frac{1}{2} \right| = \left| (u - f_1)(\omega) - \frac{1}{2} \right| = \left| f_1(\omega) - \frac{1}{2} \right|$$

and the fact that there is an “eigenstate”  $\omega_i$  for each ideal effect  $f_i$  satisfying  $f_i(\omega_i) = 1$  as we have seen in (3.16) or (3.57). Therefore, we can conclude from Theorem 3.6 (and Theorem 3.12) that for any  $\lambda \in [0, \lambda_{F,G}]$  and for

some state  $\omega_0$

$$1 - \lambda \geq \left(1 - \max_{i \in \{0,1\}} f_i(\omega_0)\right) + \left(1 - \max_{j \in \{0,1\}} g_j(\omega_0)\right)$$

holds, that is,

$$\lambda_{F,G} \leq \max_{\omega \in \Omega} \left( \max_{i \in \{0,1\}} f_i(\omega) + \max_{j \in \{0,1\}} g_j(\omega) \right) - 1 \quad (3.30)$$

holds, and  $\lambda_{\text{opt}}$  can be evaluated by taking the infimum of both sides of (3.30) over all two-outcome observables. We remark that the maximum value in the right hand side of (3.30) does exist due to the compactness of  $\Omega$ . The concrete value of the right hand side of (3.30) for regular polygon theories will be given in Subsection 3.3.2.

## 3.2 Entropic uncertainty relations in a class of GPTs

Entropic uncertainty relations have the advantages of their compatibility with information theory and independence from the structure of the sample spaces. They indeed have been applied to the field of quantum information in various ways [116]. In this section, we present our main results on two types of entropic uncertainty in a certain class of GPTs. While our results reproduce entropic uncertainty relations obtained in finite-dimensional quantum theories, they indicate that similar relations hold also in a broader class of physical theories.

### 3.2.1 Entropic PURs

We continue following the notations in the previous section. Let us consider a GPT with its state space  $\Omega$ , and two ideal observables (see (3.7))  $F = \{f_a\}_{a \in A}$  and  $G = \{g_b\}_{b \in B}$  on  $\Omega$ . Here we do not assume that  $A$  and  $B$  are metric spaces but assume that they are finite sets. For the probability distribution  $\omega^F = \{f_a(\omega)\}_a$  obtained in the measurement of  $F$  on a state  $\omega \in \Omega$  (and similarly for  $\{g_b(\omega)\}_b$ ), its Shannon entropy is defined as

$$H(\omega^F) = - \sum_{a \in A} f_a(\omega) \log f_a(\omega). \quad (3.31)$$

Note that  $H(\omega^F) \geq 0$  and  $H(\omega^F) = 0$  if and only if  $\omega^F$  is definite, i.e.  $f_{a^*}(\omega) = 1$  for some  $a^*$  and  $f_a(\omega) = 0$  for  $a \neq a^*$ . If there exists a relation

such as

$$H(\omega^F) + H(\omega^G) \geq \Gamma_{F,G} \quad \forall \omega \in \Omega$$

with a constant  $\Gamma_{F,G} > 0$ , then it is called an entropic PUR because it demonstrates that we cannot prepare a state which makes simultaneously  $H(\omega^F)$  and  $H(\omega^G)$  vanish, or  $\omega^F$  and  $\omega^G$  definite. One way to obtain an entropic PUR is to consider the Landau-Pollak-type relation [92, 93, 94]:

$$\max_{a \in A} f_a(\omega) + \max_{b \in B} g_b(\omega) \leq \gamma_{F,G} \quad \forall \omega \in \Omega \quad (3.32)$$

with a constant  $\gamma_{F,G} \in (0, 2]$ . Remark that relations of the form (3.32) always can be found for any pair of observables. It is known [99, 117] that  $\max_{a \in A} f_a(\omega)$  is related with  $H(\omega^F)$  by

$$\exp[-H(\omega^F)] \leq \max_{a \in A} f_a(\omega),$$

and thus we can observe from (3.32)

$$\exp[-H(\omega^F)] + \exp[-H(\omega^G)] \leq \gamma_{F,G}.$$

Considering that

$$\exp[-H(\omega^F)] + \exp[-H(\omega^G)] \geq 2 \exp\left[\frac{-H(\omega^F) - H(\omega^G)}{2}\right]$$

holds, we can finally obtain an entropic relation

$$H(\omega^F) + H(\omega^G) \geq -2 \log \frac{\gamma_{F,G}}{2} \quad \forall \omega \in \Omega. \quad (3.33)$$

If  $\gamma_{F,G} < 2$ , then (3.33) gives an entropic PUR because it indicates that it is impossible to prepare a state which makes both  $H(\omega^F)$  and  $H(\omega^G)$  zero, that is, there is no state preparation on which  $F$  and  $G$  take simultaneously definite values (note that (3.32) also gives a PUR if  $\gamma_{F,G} < 2$ ). In a finite-dimensional quantum theory with its state space  $\Omega_{\text{QT}}$ , it can be shown that

$$\max_a f_a(\omega) + \max_b g_b(\omega) \leq 1 + \max_{a,b} |\langle f_a | g_b \rangle| \quad \forall \omega \in \Omega_{\text{QT}}, \quad (3.34)$$

where  $F = \{|f_a\rangle\langle f_a|\}_a$  and  $B = \{|g_b\rangle\langle g_b|\}_b$  are rank-1 PVMs. In that case, (3.33) can be rewritten as

$$H(\omega^F) + H(\omega^G) \geq 2 \log \frac{2}{1 + \max_{a,b} |\langle f_a | g_b \rangle|} \quad \forall \omega \in \Omega_{\text{QT}}, \quad (3.35)$$

which is the entropic PUR proven by Deutsch [98]. There have been studies to find a better bound [99] or generalization [100] of (3.35).

*Remark 3.9*

Entropic PURs in quantum theory can be derived also by means of majorization [118, 119, 120, 121, 122, 123]. This method of majorization can be also applied to GPTs. To see this, let us introduce probability vectors  $\mathbf{f}(\omega)$  and  $\mathbf{g}(\omega)$  defined simply through  $\omega^F = \{f_a(\omega)\}_a$  and  $\omega^G = \{g_b(\omega)\}_b$  respectively. By adding outcomes to either  $A$  or  $B$ , we can assume without loss of generality that their cardinalities are equal:  $|A| = |B| = d$ , and  $\mathbf{f}(\omega)$  and  $\mathbf{g}(\omega)$  are  $d$ -dimensional vectors. If  $d$ -dimensional probability vectors  $\mathbf{p} = (p_i)_i$  and  $\mathbf{q} = (q_i)_i$  satisfy

$$\sum_{j=1}^k p_j^\downarrow \leq \sum_{j=1}^k q_j^\downarrow \quad \forall k = 1, 2, \dots, d,$$

where  $p_j^\downarrow$ 's are obtained through ordering the components of  $\mathbf{p}$  in decreasing order:  $\{p_j^\downarrow\}_j = \{p_i\}_i$  and  $p_1^\downarrow \geq p_2^\downarrow \geq p_3^\downarrow \geq \dots$  (similarly for  $q_j^\downarrow$ 's), then  $\mathbf{p}$  is called *majorized* by  $\mathbf{q}$  and we write  $\mathbf{p} < \mathbf{q}$ . For  $\mathbf{f}(\omega)$  and  $\mathbf{g}(\omega)$ , a relation of the form

$$\mathbf{f}(\omega) \otimes \mathbf{g}(\omega) < \mathbf{r} \quad \forall \omega \in \Omega, \quad (3.36)$$

where  $\mathbf{r} = (r_i)_i$  is a  $d^2$ -dimensional probability vector defined below, was proven in [119]. The vector  $\mathbf{r}$  was given by

$$\mathbf{r} = (R_1, R_2 - R_1, \dots, R_d - R_{d-1}, 0, 0, \dots, 0)$$

with

$$\begin{cases} R_k = \max_{\mathcal{I}_k} \max_{\omega \in \Omega} \sum_{(x,y) \in \mathcal{I}_k} f_a(\omega) g_b(\omega) \\ \mathcal{I}_k = \{(a_1, b_1), \dots, (a_k, b_k) \mid (a_i, b_i) \in A \times B, (a_i, b_i) \neq (a_j, b_j) \text{ for } i \neq j\} \end{cases}$$

(thus we can see  $R_k = 1$  for  $d \leq k \leq d^2$  because  $F$  and  $G$  are ideal). From (3.36), we can derive [118]

$$H(\omega^F) + H(\omega^G) \geq H(\{r_i\}_i) \quad \forall \omega \in \Omega, \quad (3.37)$$

which gives a similar entropic relation to (3.33). Note that when  $F$  and  $G$  are binary, the vector  $\mathbf{r}$  is completely determined by

$$R_1 = \max_{(a,b)} f_a(\omega) g_b(\omega).$$



In [119],  $R_1$  was evaluated as

$$R_1 = \max_{(a,b)} f_a(\omega)g_b(\omega) \leq \frac{\gamma^2}{4}$$

with

$$\gamma = \max_{(a,b)} (f_a + g_b)(\omega),$$

and it was shown that in quantum theory the equality holds:

$$R_1 = \max_{(a,b)} f_a(\omega)g_b(\omega) = \frac{\gamma^2}{4}.$$

We will consider in Subsection 3.3.2 similar cases when  $R_1 = \frac{\gamma^2}{4}$  holds, and give concrete value of  $\gamma$ .

### 3.2.2 Entropic MURs

Let  $\Omega$  be a state space which is transitive and its positive cone  $V_+$  satisfy  $V_+ = V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}^{*int}}$ , and we hereafter denote the inner product  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$  simply by  $\langle \cdot, \cdot \rangle$  as in the previous section. There can be defined measurement error in terms of entropy in the identical way with the quantum one by Buscemi et al. [19]. Let in the GPT  $E = \{e_x\}_{x \in X}$  be an ideal observable and  $M = \{m_{\hat{x}}\}_{\hat{x} \in \hat{X}}$  be an observable with finite outcome sets  $X, \hat{X}$ . Since

$$\left\langle e_{x'}, \frac{e_x}{\langle u, e_x \rangle} \right\rangle = \delta_{x'x} \quad (3.38)$$

holds for all  $x, x' \in X$ , and

$$\begin{aligned} \omega_M = u &= \sum_x e_x \\ &= \sum_x \langle u, e_x \rangle \frac{e_x}{\langle u, e_x \rangle} \end{aligned} \quad (3.39)$$

holds from Lemma 3.2 and Lemma 3.3, the joint probability distribution

$$\{p(x, \hat{x})\}_{x, \hat{x}} = \{\langle e_x, m_{\hat{x}} \rangle\}_{x, \hat{x}} = \left\{ \langle u, e_x \rangle \left\langle \frac{e_x}{\langle u, e_x \rangle}, m_{\hat{x}} \right\rangle \right\}_{x, \hat{x}} \quad (3.40)$$

is considered to be obtained in the measurement of  $M$  on the “eigenstates”  $\{e_x / \langle u, e_x \rangle\}_x$  of  $E$  (see (3.38)) with the initial distribution

$$\{p(x)\}_x = \{\langle u, e_x \rangle\}_x. \quad (3.41)$$

According to [19], the conditional entropy

$$\begin{aligned}
\mathbf{N}(M; E) &:= H(E|M) \\
&= \sum_{\hat{x}} p(\hat{x}) H(\{p(x|\hat{x})\}_x) \\
&= \sum_{\hat{x}} \langle u, m_{\hat{x}} \rangle H\left(\left\{\left\langle e_x, \frac{m_{\hat{x}}}{\langle u, m_{\hat{x}} \rangle} \right\rangle\right\}_x\right)
\end{aligned} \tag{3.42}$$

calculated via (3.40) describes how inaccurately the actual observable  $\mathcal{M}$  can estimate the input eigenstates of the ideal observable  $E$ . In fact, if we consider measuring  $\mathcal{M}$  on  $e_x/\langle u, e_x \rangle$  and estimating the input state from the output probability distribution

$$\{p(\hat{x}|x)\}_{\hat{x}} = \left\{ \left\langle m_{\hat{x}}, \frac{e_x}{\langle u, e_x \rangle} \right\rangle \right\}_{\hat{x}}$$

by means of a guessing function  $f : \hat{X} \rightarrow X$ , then the error probability  $p_{\text{error}}^f(x)$  is given by

$$p_{\text{error}}^f(x) = 1 - \sum_{\hat{x}: f(\hat{x})=x} p(\hat{x}|x) = \sum_{\hat{x}: f(\hat{x}) \neq x} p(\hat{x}|x).$$

When similar procedures are conducted for all  $x \in X$  with the probability distribution  $\{p(x)\}_x$  in (3.41), the total error probability  $p_{\text{error}}^f$  is

$$p_{\text{error}}^f = \sum_x p(x) p_{\text{error}}^f(x) = \sum_{x \in X} \sum_{\hat{x}: f(\hat{x}) \neq x} p(x, \hat{x}), \tag{3.43}$$

and it was shown in [19] that

$$\min_f p_{\text{error}}^f \rightarrow 0 \iff \mathbf{N}(M; E) = H(E|M) \rightarrow 0.$$

We can conclude from the consideration above that the entropic quantity (3.42) represents the difference between  $E$  to be measured ideally and  $\mathcal{M}$  measured actually, and thus we can define their entropic measurement error as (3.42).

We are now in the position to derive a similar entropic relation to [19] with the generalized entropic measurement error (3.42). We continue focusing on a GPT with its state space  $\Omega$  being transitive and  $V_+$  being self-dual with respect to the inner product  $\langle \cdot, \cdot \rangle_{GL(\Omega)} \equiv \langle \cdot, \cdot \rangle$ , that is,  $V_+ = V_{+\langle \cdot, \cdot \rangle}^{*int}$ . Let  $F = \{f_a\}_{a \in A}$  and  $G = \{g_b\}_{b \in B}$  be a pair of ideal observables defined in (3.7), and consider their approximate joint observable  $\tilde{M}^{FG} := \{\tilde{m}_{ab}^{FG}\}_{(a,b) \in A \times B}$  and

its marginals

$$\begin{aligned}\widetilde{M}^F &:= \{\widetilde{m}_a^F\}_a, & \widetilde{m}_a^F &:= \sum_{b \in B} \widetilde{m}_{ab}^{FG}; \\ \widetilde{M}^G &:= \{\widetilde{m}_b^G\}_b, & \widetilde{m}_b^G &:= \sum_{a \in A} \widetilde{m}_{ab}^{FG}.\end{aligned}$$

as in the previous section. We can prove the following theorem.

**Theorem 3.10**

Suppose that  $\Omega$  is a transitive state space with its positive cone  $V_+$  being self-dual with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega)} \equiv \langle \cdot, \cdot \rangle$ ,  $F = \{f_a\}_a$  and  $G = \{g_b\}_b$  are ideal observables on  $\Omega$ , and  $\widetilde{M}^{FG}$  is an arbitrary approximate joint observable of  $(F, G)$  with its marginals  $\widetilde{M}^F$  and  $\widetilde{M}^G$ . If there exists a relation

$$H(\omega^F) + H(\omega^G) \geq \Gamma_{F,G} \quad \forall \omega \in \Omega$$

with a constant  $\Gamma_{F,G}$ , then it also holds that

$$\mathbf{N}(\widetilde{M}^F; F) + \mathbf{N}(\widetilde{M}^G; G) \geq \Gamma_{F,G}.$$

**Proof**

Since for every  $\hat{a} \in A$  and  $\hat{b} \in B$   $\omega_{\hat{a}\hat{b}} := \widetilde{m}_{\hat{a}\hat{b}}^{FG} / \langle u, \widetilde{m}_{\hat{a}\hat{b}} \rangle$  is a state due to the self-duality, it holds that

$$H(\omega_{\hat{a}\hat{b}}^F) + H(\omega_{\hat{a}\hat{b}}^G) \geq \Gamma_{F,G}$$

for all  $\hat{a} \in A$  and  $\hat{b} \in B$ . Therefore, taking into consideration that  $\langle u, \widetilde{m}_{\hat{a}\hat{b}}^{FG} \rangle \geq 0$  for all  $\hat{a}, \hat{b}$  and  $\sum_{\hat{a}\hat{b}} \langle u, \widetilde{m}_{\hat{a}\hat{b}}^{FG} \rangle = \langle u, u \rangle = \langle u, \omega_M \rangle = 1$ , we have

$$\sum_{\hat{a} \in A} \sum_{\hat{b} \in B} \langle u, \widetilde{m}_{\hat{a}\hat{b}}^{FG} \rangle [H(\omega_{\hat{a}\hat{b}}^F) + H(\omega_{\hat{a}\hat{b}}^G)] \geq \Gamma_{A,B},$$

or equivalently (see (3.42))

$$H(A | \widetilde{M}^{FG}) + H(B | \widetilde{M}^{FG}) \geq \Gamma_{F,G}. \quad (3.44)$$

Note that the conditional entropy  $H(A | \widetilde{M}^{FG})$  is obtained through a joint probability distribution  $\{p(a, \hat{a}, \hat{b})\}_{a, \hat{a}, \hat{b}} := \{\langle f_a, \widetilde{m}_{\hat{a}\hat{b}}^{FG} \rangle\}$ , and we can also obtain  $H(A | \widetilde{M}^F)$  from its marginal distribution  $\{p(a, \hat{a})\}_{a, \hat{a}} = \{\langle f_a, \widetilde{m}_{\hat{a}}^F \rangle\}$ . The quantity

$$H(A | \widetilde{M}^F) - H(A | \widetilde{M}^{FG})$$

defined from those two conditional entropies is called the (classical) conditional mutual information, and it is known [124] to be nonnegative:

$$H(A | \widetilde{M}^F) - H(A | \widetilde{M}^{FG}) \geq 0.$$

A similar relation holds also for  $H(G \mid \widetilde{M}^{FG})$  and  $H(G \mid \widetilde{M}^G)$ , and thus, together with (3.44), we can conclude that

$$H(F \mid \widetilde{M}^F) + H(G \mid \widetilde{M}^G) \geq \Gamma_{F,G}$$

holds, which proves the theorem.  $\square$

Theorem 3.10 is a generalization of the quantum result [19] to a class of GPTs. In fact, when we consider a finite-dimensional quantum theory and a pair of rank-1 PVMs  $F = \{|f_a\rangle\langle f_a|\}_a$  and  $G = \{|g_b\rangle\langle g_b|\}_b$ , our theorem results in the one in [19] with the quantum bound  $\Gamma_{F,G} = -2 \log \max_{a,b} |\langle f_a | g_b \rangle|$  by Maassen and Uffink [99]. Theorem 3.10 demonstrates that if there is an entropic PUR, i.e.  $\Gamma_{F,G} > 0$ , then there is also an entropic MUR which shows that we cannot make both  $\mathbf{N}(\widetilde{M}^F; F)$  and  $\mathbf{N}(\widetilde{M}^G; G)$  vanish. It is again easy to prove that this theorem holds for three or more observables.

*Remark 3.11*

There is another type of entropic uncertainty relation on successive measurements in quantum theory [125, 126, 127, 128]. With a suitable introduction of transformations associated with ideal observables, we can derive similar entropic relations also in GPTs considered above. For an ideal observable  $E = \{e_x\}_{x \in X}$ , we define the corresponding (Schrödinger) channel  $\Phi_E$ , which gives the post-measurement states as

$$\Phi_E: \Omega \rightarrow \Omega: \omega \mapsto \sum_x \langle e_x, \omega \rangle \frac{e_x}{\langle u, e_x \rangle} \quad (3.45)$$

in analogy with the channel associated with a rank-1 projective measurement (Lüders measurement [41] for a rank-1 PVM) in quantum theory (remember (3.38)). Note that this channel is found easily to be a measure-and-prepare channel (see Example 2.43). In the Heisenberg picture, it becomes

$$\Phi_E^*: \mathcal{E}_\Omega \rightarrow \mathcal{E}_\Omega: e \mapsto \sum_x \left\langle e, \frac{e_x}{\langle u, e_x \rangle} \right\rangle e_x. \quad (3.46)$$

Let  $F = \{f_a\}_a$  and  $G = \{g_b\}_b$  be ideal observables associated with the channel defined in (3.45) (or (3.46)). It is easy to see that

$$H(\omega^F) + H(\omega^G) \geq \Gamma_{F,G} \quad \forall \omega \in \Omega$$

with

$$\Gamma_{F,G} := \inf_{\omega} [H(\omega^F) + H(\omega^G)]$$

holds. We consider measuring successively  $F$  and  $G$  on a state  $\omega$ : measuring  $F$  first, and then  $G$ . The observed statistics are  $\omega^F = \{f_a(\omega)\}_a$  and

$\Phi_F(\omega)^G = \{g_b(\Phi_F(\omega))\}_b$ , and we can derive

$$H(\omega^F) + H(\Phi_F(\omega)^G) \geq \Gamma'_{F,G} \quad (3.47)$$

with

$$\Gamma'_{F,G} := \inf_{\Phi_F(\omega)} [H(\omega^F) + H(\Phi_F(\omega)^G)] \quad (3.48)$$

$$= \inf_{\Phi_A(\omega)} [H(\Phi_F(\omega)^F) + H(\Phi_F(\omega)^G)] \quad (3.49)$$

because  $f_a(\omega) = f_a(\Phi_F(\omega))$ . We can see that  $\Gamma'_{F,G} \geq \Gamma_{F,G}$  holds, and thus there is more uncertainty in the successive measurement than the individual measurements of  $F$  and  $G$ . The entropic relation(3.47) together with (3.48) can be considered as a generalization of the quantum result [125]. Note that similarly to [125] we can present another bound for (3.47) in terms of the joint entropy. In fact, considering that  $\{f_a\}_a$  and  $\{\Phi_F^*(g_b)\}$  are jointly measurable ( $\left\langle g_b, \frac{f_a}{\langle u, f_a \rangle} \right\rangle f_a$  is the joint observable), that is, the probability distributions  $\{f_a(\omega)\}_a$  and  $\{g_b(\Phi_F(\omega))\}_b$  are obtained from the joint distribution  $\left\langle g_b, \frac{f_a}{\langle u, f_a \rangle} \right\rangle \langle f_a, \omega \rangle$ , it can be shown [124] that

$$H(\omega^F) + H(\Phi_F(\omega)^G) \geq H\left(\left\langle g_b, \frac{f_a}{\langle u, f_a \rangle} \right\rangle \langle f_a, \omega \rangle\right).$$

It is easy to see that the right hand side is also greater than or equal to  $\Gamma_{F,G}$ .

### 3.3 Uncertainty relations in regular polygon theories

In this section, we restrict ourselves to regular polygon theories, and consider similar situations to the previous sections.

#### 3.3.1 Extensions of previous theorems

Our theorems in Section 3.1 and Section 3.2 have been proven only for a class of theories such as finite-dimensional classical and quantum theories, and regular polygon theories with odd sides (see Section 2.5). What is essential to the proofs of the theorems is that we can see effects as states (the self-duality), and that every effect of an ideal observable is an ‘‘eigenstate’’ of itself (Lemma 3.3). In fact, taking those points into consideration, although it may be a minor generalization, we can demonstrate similar theorems for even-sided regular polygon theories.

**Theorem 3.12**

*Theorem 3.4, Corollary 3.5, Theorem 3.6, and Theorem 3.10 hold for every regular polygon theory.*

**Proof**

We only need to prove the claim for even-sided regular polygon theories. The proof is done by confirming that the claim of Lemma 3.3 holds for even-sided regular polygon theories with modified parametrizations. We again denote the inner product  $\langle \cdot, \cdot \rangle_{GL(\Omega_n)}$  by  $\langle \cdot, \cdot \rangle$  in this proof.

In the  $n$ -sided regular polygon theory with even  $n$ , if  $F = \{f_a\}_a$  is an ideal observable, then it is of the form

$$F = \{f_0, f_1\} \quad (3.50)$$

with

$$f_0 = e_i^n \quad \text{and} \quad f_1 = u - e_i^n = e_{i+\frac{n}{2}}^n \quad (3.51)$$

for some  $i$  (remember that we do not consider the trivial observable  $F = \{u\}$ ). Let us introduce an affine bijection

$$\psi := \begin{pmatrix} r_n & 0 & 0 \\ 0 & r_n & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.52)$$

on  $\mathbb{R}^3$ . Because  $(e, \omega)_E = (\psi^{-1}(e), \psi(\omega))_E$  holds for any  $\omega \in \Omega_n$  and  $e \in \mathcal{E}(\Omega_n)$ , we can consider an equivalent expression of the theory with  $\psi(\Omega_n) =: \widehat{\Omega}_n$  and  $\psi^{-1}(\mathcal{E}(\Omega_n))$  being its state and effect space respectively (remember that  $(\cdot, \cdot)_E$  is the standard Euclidean inner product). The pure states (2.27) and the extreme effects (2.29) shown in Subsection 2.5.3 are modified as

$$\omega_i^n \rightarrow \widehat{\omega}_i^n := \psi(\omega_i^n) = \begin{pmatrix} r_n^2 \cos(\frac{2\pi i}{n}) \\ r_n^2 \sin(\frac{2\pi i}{n}) \\ 1 \end{pmatrix}; \quad (3.53)$$

$$e_i^n \rightarrow \check{e}_i^n := \psi^{-1}(e_i^n) = \frac{1}{2} \begin{pmatrix} \cos(\frac{(2i-1)\pi}{n}) \\ \sin(\frac{(2i-1)\pi}{n}) \\ 1 \end{pmatrix} \quad (3.54)$$

respectively, and their conic hull (the positive cone and the internal dual cone) as

$$\begin{aligned} V_+ &\rightarrow \widehat{V}_+ := \psi(V_+); \\ V_{+\langle \cdot, \cdot \rangle}^{*int} &\rightarrow \check{V}_{+\langle \cdot, \cdot \rangle}^{*int} := \psi^{-1}(V_{+\langle \cdot, \cdot \rangle}^{*int}), \end{aligned}$$

respectively. Note in the equations above that  $GL(\Omega_n) = GL(\widehat{\Omega}_n)$  and

$(\cdot, \cdot)_E = \langle \cdot, \cdot \rangle_{GL(\Omega_n)} = \langle \cdot, \cdot \rangle_{GL(\hat{\Omega}_n)} = \langle \cdot, \cdot \rangle$  hold, and  $\omega_M = u = {}^t(0, 0, 1)$  is invariant for  $\psi$  (and  $\psi^{-1}$ ). We can also find that an observable  $E = \{e_a\}_a$  in the original expression is rewritten as  $\check{E} := \{\check{e}_a\}_a$  with  $\check{e}_a := \psi^{-1}(e_a)$ , and that an ideal observable  $F$  in (3.50) and (3.51) gives

$$\check{F} = \{\check{f}_0, \check{f}_1\} \quad (3.55)$$

with

$$\check{f}_0 = \check{e}_i^n \quad \text{and} \quad \check{f}_1 = u - \check{e}_i^n = \check{e}_{i+\frac{n}{2}}^n \quad (3.56)$$

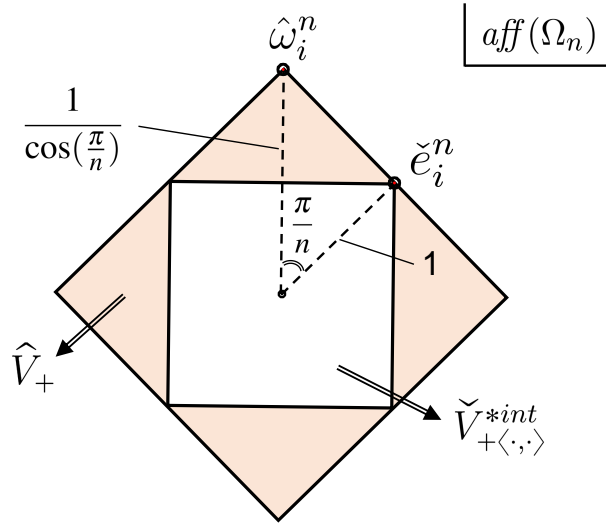
which is also ideal in the rewritten theory. Since

$$\left\langle \check{e}_i^n, \frac{\check{e}_i^n}{\langle u, \check{e}_i^n \rangle} \right\rangle = 1 \quad (3.57)$$

holds for any  $i$  (see (3.54)), we can conclude together with (3.55) and (3.56) that any ideal observables  $\check{F} = \{\check{f}_k\}_{k=0,1}$  satisfies

$$\left\langle \check{f}_k, \frac{\check{f}_k}{\langle u, \check{f}_k \rangle} \right\rangle = 1. \quad (3.58)$$

On the other hand, it can be seen from (3.53) and (3.54) that  $\hat{V}_+$  generated by (3.53) includes  $\check{V}_{+\langle \cdot, \cdot \rangle}^{*int}$  generated by (3.54), i.e.  $\check{V}_{+\langle \cdot, \cdot \rangle}^{*int} \subset \hat{V}_+$  (see FIG 3.1).



**Figure 3.1:** Illustration of  $(\text{aff}(\Omega_n) \cap \hat{V}_+) = \hat{\Omega}_n$  generated by  $\{\hat{\omega}_i^n\}_{i=1}^n$  (3.53) and  $(\text{aff}(\Omega_n) \cap \check{V}_{+\langle \cdot, \cdot \rangle}^{*int})$  generated by  $\{2\check{e}_i^n\}_{i=1}^n$  (3.54) for  $n = 4$ . It is observed that  $\check{V}_{+\langle \cdot, \cdot \rangle}^{*int} \subset \hat{V}_+$ , which holds also for every even  $n$ .

Therefore,

$$\frac{\check{e}}{\langle u, \check{e} \rangle} \in \hat{\Omega}_n \quad (3.59)$$

holds for any effect  $\check{e} \in \check{V}_{+\langle \cdot, \cdot \rangle}^{*int}$ . It follows from (3.58) and (3.59) that the claim of Lemma 3.3 holds also for even-sided regular polygon theories in a rewritten expression (3.53) and (3.54).

We also need to confirm that all of our measures (3.3), (3.4), (3.8), (3.9), (3.12), (3.31), and (3.42) depend only on probabilities, and thus they are invariant for the modification above. For example, for a pair of observables  $M = \{m_a\}_a$  and  $F = \{f_a\}_a$  on the original state space  $\Omega_n$ , we can see easily from (3.4) and (3.12) that

$$\begin{aligned} LE(\omega^F) &= 1 - \max_{a \in A} f_a(\omega) \\ &= 1 - \max_{a \in A} \check{f}_a(\hat{\omega}) \\ &= LE(\hat{\omega}^{\check{F}}) \end{aligned}$$

and

$$\begin{aligned} D_\infty(M, F) &= \sup_{\omega \in \Omega_n} \max_{a \in A} |m_a(\omega) - f_a(\omega)| \\ &= \sup_{\hat{\omega} \in \hat{\Omega}_n} \max_{a \in A} |\check{m}_a(\hat{\omega}) - \check{f}_a(\hat{\omega})| \\ &= D_\infty(\check{M}, \check{F}) \end{aligned}$$

respectively. It results in that if Theorem 3.6 holds in the modified theory, then it holds also in the original theory. In fact, by virtue of (3.58) and (3.59) (the ‘‘generalized version of Lemma 3.3’’), we can repeat the same calculations as in Theorem 3.6, and obtain a similar result to it in the modified theory. Similar considerations can be adapted also for the other measures, and it proves Theorem 3.12.  $\square$

### 3.3.2 Concrete values for Landau-Pollak-type bounds

In this part, we shall concentrate on the Landau-Pollak-type relation (see (3.34)) for the  $n$ -sided regular polygon theory of the form

$$\max_a f_a(\omega) + \max_b g_b(\omega) \leq \Gamma_{F,G}(n) \quad \forall \omega \in \Omega_n, \quad (3.60)$$

where  $F = \{f_a\}_a$  and  $G = \{g_b\}_b$  are ideal observables as usual, and show a concrete calculation for the bound  $\Gamma_{F,G}(n)$  of uncertainty.

Let us focus on the state space  $\Omega_n$ . Any nontrivial ideal observable is of the form  $\{e_i^n, u - e_i^n\}$  (see (2.29)). Note that although  $\{e_i^n\}_{i=0,1,2}$  is also



an ideal observable when  $n = 3$  (a classical trit system), we focus only on ideal observables with two outcomes in this subsection. Thus, if we consider a pair of ideal observables  $F$  and  $G$ , then we can suppose that they are binary:  $F = F_i \equiv \{f_i^0, f_i^1\}$  and  $G = G_j \equiv \{g_j^0, g_j^1\}$  with  $f_i^0 = e_i^n$  and  $g_j^0 = e_j^n$  for  $i, j \in \{0, 1, \dots, n-1\}$  (or  $i, j \in [0, 2\pi)$  when  $n = \infty$ ). On the other hand, it holds that

$$\begin{aligned} \max_{x=0,1} f_i^x(\omega) + \max_{y=0,1} g_j^y(\omega) &\leq \sup_{\omega \in \Omega_n} \max_{(x,y) \in \{0,1\}^2} [(f_i^x + g_j^y)(\omega)] \\ &= \max_{\omega \in \Omega_n^{\text{ext}}} \max_{(x,y) \in \{0,1\}^2} [(f_i^x + g_j^y)(\omega)] \end{aligned} \quad (3.61)$$

because  $\Omega_n$  is a compact set and any state can be represented as a convex combination of pure states. Therefore, if we let  $\omega_k^n$  be a pure state ((2.27) and (2.28)), then the value

$$\gamma_{F_i, G_j}^n := \max_k \max_{(x,y) \in \{0,1\}^2} [(f_i^x + g_j^y)(\omega_k^n)] \quad (3.62)$$

gives a Landau-Pollak-type relation

$$\max_{x=0,1} f_i^x(\omega) + \max_{y=0,1} g_j^y(\omega) \leq \gamma_{A^i, B^j}^n \quad \forall \omega \in \Omega_n. \quad (3.63)$$

From this inequality, we can derive, for example, entropic relations

$$H(\omega^F) + H(\omega^G) \geq -2 \log \frac{\gamma_{F_i, G_j}^n}{2} \quad \forall \omega \in \Omega_n \quad (3.64)$$

and

$$\mathbf{N}(\widetilde{M}^F; F) + \mathbf{N}(\widetilde{M}^G; G) \geq -2 \log \frac{\gamma_{F_i, G_j}^n}{2}. \quad (3.65)$$

**Table 3.1:** The value  $(f_i^x + g_j^y)(\omega_k^n)$  when  $n$  is even.

$x = 0, y = 0$	$1 + r_n^2 \cos \left[ \frac{\theta_i + \theta_j}{2} - \phi_k \right] \cos \left[ \frac{\theta_i - \theta_j}{2} \right]$
$x = 1, y = 0$	$1 + r_n^2 \sin \left[ \frac{\theta_i + \theta_j}{2} - \phi_k \right] \sin \left[ \frac{\theta_i - \theta_j}{2} \right]$
$x = 0, y = 1$	$(i \longleftrightarrow j \text{ in the case of } x = 1, y = 0)$
$x = 1, y = 1$	$1 - r_n^2 \cos \left[ \frac{\theta_i + \theta_j}{2} - \phi_k \right] \cos \left[ \frac{\theta_i - \theta_j}{2} \right]$
$\theta_i = \frac{2i-1}{n}\pi, \theta_j = \frac{2j-1}{n}\pi, \phi_k = \frac{2k}{n}\pi \quad (i, j, k = 0, 1, \dots, n-1)$	

**Table 3.2:** The value  $(f_i^x + g_j^y)(\omega_k^n)$  when  $n$  is odd.

$x = 0, y = 0$	$\frac{2}{1+r_n^2} + \frac{2r_n^2}{1+r_n^2} \cos \left[ \frac{\theta_i + \theta_j}{2} - \phi_k \right] \cos \left[ \frac{\theta_i - \theta_j}{2} \right]$
$x = 1, y = 0$	$1 + \frac{2r_n^2}{1+r_n^2} \sin \left[ \frac{\theta_i + \theta_j}{2} - \phi_k \right] \sin \left[ \frac{\theta_i - \theta_j}{2} \right]$
$x = 0, y = 1$	$(i \longleftrightarrow j \text{ in the case of } x = 1, y = 0)$
$x = 1, y = 1$	$\frac{2r_n^2}{1+r_n^2} - \frac{2r_n^2}{1+r_n^2} \cos \left[ \frac{\theta_i + \theta_j}{2} - \phi_k \right] \cos \left[ \frac{\theta_i - \theta_j}{2} \right]$
$\theta_i = \frac{2i}{n}\pi, \theta_j = \frac{2j}{n}\pi, \phi_k = \frac{2k}{n}\pi \quad (i, j, k = 0, 1, \dots, n-1)$	

**Table 3.3:** The value  $(f_i^x + g_j^y)(\omega_k^n)$  when  $n$  is  $\infty$ .

$x = 0, y = 0$	$1 + \cos \left[ \frac{\theta_i + \theta_j}{2} - \phi_k \right] \cos \left[ \frac{\theta_i - \theta_j}{2} \right]$
$x = 1, y = 0$	$1 + \sin \left[ \frac{\theta_i + \theta_j}{2} - \phi_k \right] \sin \left[ \frac{\theta_i - \theta_j}{2} \right]$
$x = 0, y = 1$	$(i \longleftrightarrow j \text{ in the case of } x = 1, y = 0)$
$x = 1, y = 1$	$1 - \cos \left[ \frac{\theta_i + \theta_j}{2} - \phi_k \right] \cos \left[ \frac{\theta_i - \theta_j}{2} \right]$
$\theta_i = i, \theta_j = j, \phi_k = k \quad (0 \leq i, j, k < 2\pi)$	

Table 3.1 - Table 3.3 show the value of  $(f_i^x + g_j^y)(\omega_k^n)$  in terms of the angles  $\theta_i, \theta_j$ , and  $\phi_k$  between the  $x$ -axis and the effects  $f_i^0 = e_i^n, g_j^0 = e_j^n$ , and the state  $\omega_k^n$  respectively when viewed from the  $z$ -axis (see (2.27) - (2.29) in Subsection 2.5.3). Maximizing the values in those tables over all pure states, we can obtain the optimal bound  $\gamma_{F_i, G_j}^n$  in (3.62) for each regular polygon theory. Note that focusing only on the case when  $j = 0$  and  $0 < i < \frac{n}{2}$  ( $0 < i < \pi$  when  $n = \infty$ ) is sufficient for the universal description of  $\gamma_{F_i, G_j}^n$  due to the geometric symmetry of the regular polygon theories.  $\gamma_{F_i, G_0}^n$  for the regular polygon theory with  $n (< \infty)$  sides is exhibited in Table 3.4 and Table 3.5, and  $\gamma_{F_i, G_0}^n$  for the disc theory (the regular polygon theory with  $n = \infty$  sides) can be calculated from Table 3.3 as

$$\gamma_{F_i, G_0}^n = \max \left\{ 1 + \cos \frac{\theta'_i}{2}, 1 + \sin \frac{\theta'_i}{2} \right\}, \quad (3.66)$$

where  $\theta'_i = \theta_i - \theta_0 = \theta_i$  similarly to Table 3.4 and Table 3.5. (3.66) can be regarded as giving the quantum bound in (3.34) for a qubit system in terms of the usual Bloch representation. Note that when  $n$  is even or  $\infty$ , due to

the geometric symmetry,  $(f_i^x + g_j^y)(\omega_k^n)$  takes its maximum where  $\omega_k^n$  lies just “halfway” between the effects  $f_i^x$  and  $g_j^y$ , that is,  $f_i^x(\omega) = g_j^y(\omega)$  and thus  $f_i^x(\omega)g_j^y(\omega) = \frac{1}{4}(f_i^x(\omega) + g_j^y(\omega))^2$  holds (see Remark 3.9), while this does not hold generally when  $n$  is odd. From Table 3.4, Table 3.5 and (3.66), we can obtain the corresponding entropic inequalities (3.64) (also (3.37)) and (3.65) for an arbitrary regular polygon theory. We should recall that the value  $\gamma_{F_i, G_0}^n$  can be used also to evaluate the nonlocality of the theory via its degree of incompatibility (see Example 3.8).

**Table 3.4:** The value  $\gamma_{F_i, G_0}^n$  when  $n$  is even.

$n \equiv 0 \pmod{4}, i: \text{ even}$	$\max \left\{ 1 + \cos \frac{\theta'_i}{2}, 1 + \sin \frac{\theta'_i}{2} \right\}$
$n \equiv 0 \pmod{4}, i: \text{ odd}$	$\max \left\{ 1 + r_n^2 \cos \frac{\theta'_i}{2}, 1 + r_n^2 \sin \frac{\theta'_i}{2} \right\}$
$n \equiv 2 \pmod{4}, i: \text{ even}$	$\max \left\{ 1 + \cos \frac{\theta'_i}{2}, 1 + r_n^2 \sin \frac{\theta'_i}{2} \right\}$
$n \equiv 2 \pmod{4}, i: \text{ odd}$	$\max \left\{ 1 + r_n^2 \cos \frac{\theta'_i}{2}, 1 + \sin \frac{\theta'_i}{2} \right\}$
$\theta'_i = \frac{2i}{n}\pi = \theta_i - \theta_0$	

**Table 3.5:** The value  $\gamma_{F_i, G_0}^n$  when  $n$  is odd.

$i: \text{ even}$	$\max \left\{ \frac{2r_n^2}{1+r_n^2} + \frac{2}{1+r_n^2} \cos \frac{\theta'_i}{2}, 1 + \frac{1}{\cos \frac{\pi}{2n}} \sin \frac{\theta'_i}{2} \right\}$
$i: \text{ odd}$	$\max \left\{ \frac{2r_n^2}{1+r_n^2} + \frac{2r_n^2}{1+r_n^2} \cos \frac{\theta'_i}{2}, 1 + \frac{1}{\cos \frac{\pi}{2n}} \sin \frac{\theta'_i}{2} \right\}$
$\theta'_i = \frac{2i}{n}\pi = \theta_i - \theta_0 = \theta_i$	

*Remark 3.13*

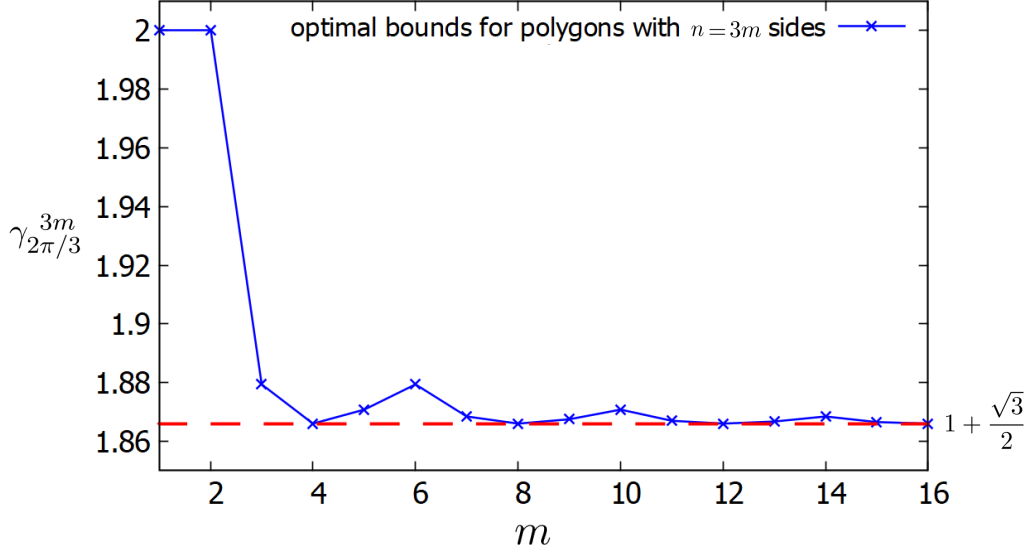
With the angle  $\theta'_i$  fixed, we can see from Table 3.4, Table 3.5, and (3.66) that  $\gamma_{F_i, G_0}^n \geq \gamma_{F_i, G_0}^\infty$  holds for all  $n$ . In fact, if we assume, for example,  $n$  is odd and  $i$  is even, then

$$\gamma_{F_i, G_0}^n = \max \left\{ \frac{2r_n^2}{1+r_n^2} + \frac{2}{1+r_n^2} \cos \frac{\theta'_i}{2}, 1 + \frac{1}{\cos \frac{\pi}{2n}} \sin \frac{\theta'_i}{2} \right\}$$

(see Table 3.5), and it can be easily shown that

$$\begin{aligned} \frac{2r_n^2}{1+r_n^2} + \frac{2}{1+r_n^2} \cos \frac{\theta'_i}{2} &\geq 1 + \cos \frac{\theta'_i}{2}, \\ 1 + \frac{1}{\cos \frac{\pi}{2n}} \sin \frac{\theta'_i}{2} &\geq 1 + \sin \frac{\theta'_i}{2} \end{aligned}$$

hold for  $0 < i < \frac{n}{2}$  (or  $0 < \theta'_i < \frac{\pi}{2}$ ). Thus, we can conclude  $\gamma_{F_i, G_0}^n \geq \gamma_{F_i, G_0}^\infty$ .



**Figure 3.2:** The optimal bound  $\gamma_{2\pi/3}^{3m}$  for the Landau-Pollak-type inequality on a pair of observables  $(F_m, G_0)$  in the regular polygon theory with  $n = 3m$ .

To see this in a more explicit way, let us consider, as an illustration, regular polygon theories with  $n = 3m$  ( $m = 1, 2, \dots$ ), and let the angle  $\theta'_i$  be  $\theta'_i = \frac{2\pi}{3}$  (i.e.  $i = m$ ). We can calculate the corresponding optimal bound  $\gamma_{2\pi/3}^n \equiv \gamma_{2\pi/3}^{3m}$  for any  $m$  from Table 3.4, Table 3.5 and (3.66), and describe its behavior as a function of  $m$  in Figure 3.2. There can be observed that theories with  $m = 1, 2$  ( $n = 3, 6$ ) admit  $\gamma_{2\pi/3}^{3m} = 2$ , that is, there is a state on which both  $F_i = F_m$  and  $G_0$  take simultaneously exact values when  $m = 1, 2$ . It exhibits that when  $m \geq 3$ , there exists preparation uncertainty for this  $(F_i, G_0)$ . Hence, it follows from our theorems that there also exists measurement uncertainty for  $(F_i, G_0)$ , and their entropic representations (entropic PUR and MUR) are given by similar inequalities with the same bound. Also, it can be observed that  $\gamma_{2\pi/3}^{3m} \geq \gamma_{2\pi/3}^\infty = 1 + \frac{\sqrt{3}}{2}$  holds for all  $m$ , which has been shown in the argument above. Note that we can derive easily an observable-independent relation

$$\min_i \gamma_{F_i, G_0}^n \geq \min_i \gamma_{F_i, G_0}^\infty.$$

In other words, the disc theory shows the “maximum uncertainty” in terms of the Landau-Pollak-type formulation.

## Chapter 4

# Testing incompatibility of quantum devices with few states

Quantum information processing, including the exciting fields of quantum communication and quantum computation, is ultimately based on the fact that there are new types of resources that can be utilized in carefully designed information processing protocols. The best-known feature of quantum information is that quantum systems can be in superposition and entangled states, and these resources lead to applications such as superdense coding and quantum teleportation. While superposition and entanglement are attributes of quantum states, quantum measurements have also features that can power a new type of applications. The best known and most studied property is the incompatibility of pairs (or collections) of quantum measurements [21]. It is crucial e.g. in the BB84 quantum key distribution protocol [11] that the used measurements are incompatible.

From the resource perspective, it is important to quantify the incompatibility. There have been several studies on incompatibility robustness, i.e., how incompatibility is affected by noise. This is motivated by the fact that noise is unavoidable in any actual implementation of quantum devices and similar to other quantum properties (e.g. entanglement), large amount of noise destroys incompatibility. Earlier studies have mostly focused on quantifying noise [129] and finding those pairs or collections of measurements that are most robust to certain types of noise [130], or to find conditions under which all incompatibility is completely erased [131]. In this work, we introduce quantifications of incompatibility which are motivated by an operational aspect of testing whether a collection of devices is incompatible or not. We focus on two integer valued quantifications of incompatibility, called *compatibility dimension* and *incompatibility dimension*. We formulate these concepts for arbitrary collections of devices. Roughly speaking, the first one quantifies how many states we minimally need to use to detect incompatibility if we choose the test states carefully, whereas the second

one quantifies how many (affinely independent) states we may have to use if we cannot control their choice. We study some of the basic properties of these quantifications of incompatibility and we present several examples to demonstrate their behaviour.

This part is organized as follows. In Section 4.1, we introduce the notion of compatibility and incompatibility dimension, which reflects operationally how easy it is to detect the incompatibility of quantum devices considered. We also give brief reviews on related studies recently reported in [132, 133, 134, 135] for the case of quantum observables, and explain the interconnections of these studies to ours. In Section 4.2, we show that incompatibility dimension is related with the concept of *incompatibility witness* [16, 17, 112]. We also derive a useful bound for incompatibility dimension by means from the relation between them. In Section 4.3, we give a particular analysis for compatibility and incompatibility dimensions of a pair of mutually unbiased qubit observables. We show that, remarkably, even for the standard example of noisy orthogonal qubit observables the incompatibility dimension has a jump in a point where all noise robustness measures are continuous and indicate nothing special to happen. More precisely, the noise parameter has a threshold value where the number of needed test states to reveal incompatibility shifts from 2 to 3. This means that even in this simple class of incompatible pairs of qubit observables there is a qualitative difference in the incompatibility of less noisy and more noisy pairs of observables. An interesting additional fact is that the compatibility dimension of these pairs of observables does not depend on the noise parameter.

For simplicity and clarity, we will restrict to finite-dimensional Hilbert spaces and observables with a finite number of outcomes. Our definitions apply not only to quantum theory but also to any GPT. However, for the sake of concreteness, we keep the discussion in the realm of quantum theory. The main definitions work in any GPT without any changes. We expect that similar findings as the aforementioned result on noisy orthogonal qubit observables can be made in subsequent studies on other collections of devices.

## 4.1 (In)compatibility on a subset of states

In this section, we introduce the notion of incompatibility dimension and compatibility dimension as quantifications of incompatibility. We again mention that we focus on compatibility and incompatibility in quantum theory in this chapter, but those quantities can be defined naturally also in GPTs.

### 4.1.1 (In)compatibility for quantum devices

We start with presenting explicit descriptions of compatibility and incompatibility for quantum observables, although we have already given their definitions in the general framework of GPTs (see Definition 2.45 and Proposition 2.46). A quantum observable is mathematically described as a positive operator valued measure (POVM) [76]. A quantum observable with finite number of outcomes is hence a map  $x \mapsto \mathbf{A}(x)$  from the outcome set to the set of linear operators on a Hilbert space. The compatibility of quantum observables  $\mathbf{A}_1, \dots, \mathbf{A}_n$  with outcome sets  $X_1, \dots, X_n$  means that there exists an observable  $\mathbf{G}$ , called *joint observable*, defined on the product outcome set  $X_1 \times \dots \times X_n$  such that from an outcome  $(x_1, \dots, x_n)$  of  $\mathbf{G}$ , one can infer outcomes for every  $\mathbf{A}_1, \dots, \mathbf{A}_n$  by ignoring the other outcomes. More precisely, the requirement is that

$$\begin{aligned} \mathbf{A}_1(x_1) &= \sum_{x_2, \dots, x_n} \mathbf{G}(x_1, x_2, \dots, x_n), \\ \mathbf{A}_2(x_2) &= \sum_{x_1, x_3, \dots, x_n} \mathbf{G}(x_1, x_2, \dots, x_n), \\ &\vdots \\ \mathbf{A}_n(x_n) &= \sum_{x_1, \dots, x_{n-1}} \mathbf{G}(x_1, x_2, \dots, x_n). \end{aligned} \tag{4.1}$$

If  $\mathbf{A}_1, \dots, \mathbf{A}_n$  are not compatible, then they are called *incompatible*.

#### Example 4.1

(*Unbiased qubit observables*) We recall a standard example to fix the notation that we will use in later examples. An unbiased qubit observable is a dichotomic observable with outcomes  $\pm$  and determined by a vector  $\mathbf{a} \in \mathbb{R}^3$ ,  $|\mathbf{a}| \leq 1$  via

$$\mathbf{A}^{\mathbf{a}}(\pm) = \frac{1}{2}(\mathbb{1} \pm \mathbf{a} \cdot \boldsymbol{\sigma}),$$

where  $\mathbf{a} \cdot \boldsymbol{\sigma} = a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3$  and  $\sigma_i$ ,  $i = 1, 2, 3$ , are the Pauli matrices. The Euclidean norm  $|\mathbf{a}|$  of  $\mathbf{a}$  reflects the noise in  $\mathbf{A}^{\mathbf{a}}$ ; in the extreme case of  $|\mathbf{a}| = 1$  the operators  $\mathbf{A}^{\mathbf{a}}(\pm)$  are projections and the observable is called sharp. As shown in [136], two unbiased qubit observables  $\mathbf{A}^{\mathbf{a}}$  and  $\mathbf{A}^{\mathbf{b}}$  are compatible if and only if

$$|\mathbf{a} + \mathbf{b}| + |\mathbf{a} - \mathbf{b}| \leq 2. \tag{4.2}$$

There are two extreme cases. Firstly, if  $\mathbf{A}^{\mathbf{a}}$  is sharp then it is compatible with some  $\mathbf{A}^{\mathbf{b}}$  if and only if  $\mathbf{b} = r\mathbf{a}$  for some  $-1 \leq r \leq 1$ . Secondly, if  $|\mathbf{a}| = 0$ , then  $\mathbf{A}^{\mathbf{a}}(\pm) = \frac{1}{2}\mathbb{1}$  and it is called a trivial qubit observable, in which case it is compatible with all other qubit observables.

How can we test if a given family of observables is compatible or incompatible? From the operational point of view, the existence of an observable  $\mathbf{G}$  satisfying (4.1) is equivalent to the existence of  $\mathbf{G}$  such that for any state  $\varrho$  the equation

$$\mathrm{Tr}[\varrho \mathbf{A}_1(x_1)] = \sum_{x_2, \dots, x_n} \mathrm{Tr}[\varrho \mathbf{G}(x_1, x_2, \dots, x_n)] \quad (4.3)$$

holds. Before contemplating into these questions, we recall that analogous definitions of quantum compatibility and incompatibility make sense for other types of quantum devices, in particular, for instruments and channels [21, 137, 138, 139, 140, 141, 142]. We denote by  $\mathcal{S}(\mathcal{H})$  the set of all density operators on a Hilbert space  $\mathcal{H}$ . The input space of all types of devices must be  $\mathcal{S}(\mathcal{H}_{in})$  on the same Hilbert space  $\mathcal{H}_{in}$  as the devices operate on a same system. We denote  $\mathcal{S}(\mathcal{H}_{in})$  simply by  $\mathcal{S}$ . A device is a completely positive map and the “type” of the device is characterized by its output space. Output spaces for the three basic types of devices are:

- observable:  $P(X) := \{p = \{p(x)\}_{x \in X} \mid 0 \leq p(x) \leq 1, \sum_x p(x) = 1\}$ ,
- channel:  $\mathcal{S}(\mathcal{H}_{out})$ ,
- instrument:  $\mathcal{S}(\mathcal{H}_{out}) \otimes P(X)$ .

In this classification, an observable  $\mathbf{A}$  is identified with a map  $\varrho \mapsto \mathrm{Tr}[\varrho \mathbf{A}(x)]$  from  $\mathcal{S}(\mathcal{H}_{in})$  to  $P(X)$ . We limit our investigation to the cases where the number of outcomes in  $X$  is finite and the output Hilbert space  $\mathcal{H}_{out}$  is finite-dimensional. Regarding  $P(X) \subset \mathcal{S}(\mathbb{C}^{|X|})$  as the set of all diagonal density operators, we can summarize that quantum devices are normalized completely positive maps to different type of output spaces.

Devices  $\mathbf{D}_1, \dots, \mathbf{D}_n$  are *compatible* if there exists a device  $\mathbf{D}$  that can simulate  $\mathbf{D}_1, \dots, \mathbf{D}_n$  simultaneously, meaning that by ignoring disjoint parts of the output of  $\mathbf{D}$  we get the same actions as  $\mathbf{D}_1, \dots, \mathbf{D}_n$  (see [21]). This kind of device is called a *joint device* of  $\mathbf{D}_1, \dots, \mathbf{D}_n$ . The input space of  $\mathbf{D}$  is the same as for  $\mathbf{D}_1, \dots, \mathbf{D}_n$ , but the output space is the tensor product of their output spaces. As an illustration, let  $\mathbf{D}_j: \mathcal{S}(\mathcal{H}_{in}) \rightarrow \mathcal{S}(\mathcal{H}_j)$  ( $j = 1, \dots, n$ ) be quantum channels. They are compatible iff there exists a channel  $\mathbf{D}: \mathcal{S}(\mathcal{H}_{in}) \rightarrow \mathcal{S}(\otimes_{j=1}^n \mathcal{H}_j)$  satisfying

$$\begin{aligned} \mathbf{D}_1(\varrho) &= \mathrm{tr}_{\mathcal{H}_2, \dots, \mathcal{H}_n} \mathbf{D}(\varrho), \\ \mathbf{D}_2(\varrho) &= \mathrm{tr}_{\mathcal{H}_1, \mathcal{H}_3, \dots, \mathcal{H}_n} \mathbf{D}(\varrho), \\ &\vdots \\ \mathbf{D}_n(\varrho) &= \mathrm{tr}_{\mathcal{H}_1, \dots, \mathcal{H}_{n-1}} \mathbf{D}(\varrho) \end{aligned}$$



for all  $\varrho \in \mathcal{S}(\mathcal{H}_{in})$  (see (4.1)). If  $D_1, \dots, D_n$  are not compatible, then they are *incompatible*. We recall a qubit example to exemplify the general definition.

*Example 4.2*

(*Unbiased qubit observable and partially depolarizing noise*) A measurement of an unbiased qubit observable  $A^a$  necessarily disturbs the system. This trade-off is mathematically described by the compatibility relation between observables and channels. Let us consider partially depolarizing qubit channels, which have the form

$$\Gamma_p(\varrho) = p\varrho + (1-p)\frac{1}{2}\mathbb{1} \quad (4.4)$$

for  $0 \leq p \leq 1$ . A joint device for a channel and observable is an instrument. Hence,  $A^a$  and  $\Gamma_p$  are compatible if there exists an instrument  $x \mapsto \Phi_x$  such that

$$\sum_x \Phi_x(\varrho) = \Gamma_p(\varrho) \quad \text{and} \quad \text{Tr}[\Phi_x(\varrho)] = \text{Tr}[\varrho A^a(x)]$$

for all states  $\varrho$  and outcomes  $x$ . It has been proven in [140] that  $A^a$  and  $\Gamma_p$  are compatible if and only if

$$|a| \leq \frac{1}{2} \left( 1 - p + \sqrt{(1-p)(1+3p)} \right). \quad (4.5)$$

This shows that higher is the norm  $|a|$ , smaller must  $p$  be.

### 4.1.2 (In)compatibility dimension of devices

To test the incompatibility we should hence check the validity of (4.3) in a subset of states that spans the whole state space. An obvious question is then if we really need all those states, or if a smaller number of test states is enough. Further, does the number of needed test states depend on the given family of observables? How does noise affect the number of needed test states? The earlier discussion motivates the following definition, which is central to our investigation.

#### Definition 4.3

Let  $\mathcal{S}_0 \subset \mathcal{S}$ . Devices  $D_1, \dots, D_n$  are  $\mathcal{S}_0$ -*compatible* if there exist compatible devices  $D'_1, \dots, D'_n$  of the same type such that

$$D'_j(\varrho) = D_j(\varrho) \quad (4.6)$$

for all  $j = 1, \dots, n$  and states  $\varrho \in \mathcal{S}_0$ . Otherwise,  $D_1, \dots, D_n$  are  $\mathcal{S}_0$ -*incompatible*.

The definition is obviously interesting only when  $D_1, \dots, D_n$  are incompatible in the usual sense, i.e., with respect to the full state space. In that

case the definition means that if devices  $D_1, \dots, D_n$  are  $\mathcal{S}_0$ -compatible, their incompatibility cannot be verified by taking test states from  $\mathcal{S}_0$  only, and vice versa, if devices  $D_1, \dots, D_n$  are  $\mathcal{S}_0$ -incompatible, their actions on  $\mathcal{S}_0$  cannot be simulated by any collection of compatible devices and therefore their incompatibility should be able to be observed in some way.

The  $\mathcal{S}_0$ -compatibility depends not only on the size of  $\mathcal{S}_0$  but also on its structure. We start with a simple example showing that there exist sets  $\mathcal{S}_0$  such that an arbitrary family of devices is  $\mathcal{S}_0$ -compatible.

*Example 4.4*

Any set of devices  $D_1, \dots, D_n$  is  $\mathcal{S}_0$ -compatible if  $\mathcal{S}_0 = \{\varrho_1, \dots, \varrho_k\}$  consists of perfectly distinguishable states. In fact, one may construct a device  $D'_k$  which outputs  $D_k(\varrho_j)$  after confirming an input state is  $\varrho_j$  by measuring an observable that distinguishes the states in  $\mathcal{S}_0$ . It is easy to see that the devices  $D'_1, \dots, D'_n$  are compatible. The same argument works for devices in general probabilistic theories and one can use the same reasoning for a subset  $\mathcal{S}_0$  that is broadcastable [25]. (We recall that a subset  $\mathcal{S}_0$  is broadcastable if there exists a channel  $B : \mathcal{S} \rightarrow \mathcal{S} \otimes \mathcal{S}$  such that the bipartite state  $B(\varrho)$  has marginals equal to  $\varrho$  for all  $\varrho \in \mathcal{S}_0$ .) For instance, two qubit states  $\mathbb{1}/2$  and  $|0\rangle\langle 0|$  are broadcastable even though not distinguishable. Any pair of qubit channels  $\Lambda_1$  and  $\Lambda_2$  is  $\mathcal{S}_0$ -compatible for  $\mathcal{S}_0 = \{\mathbb{1}/2, |0\rangle\langle 0|\}$  as we can define  $\Lambda'_j(\varrho) = \sum_{i=0}^1 \langle i|\varrho|i\rangle \Lambda_j(|i\rangle\langle i|)$  for  $j = 1, 2$ . The channel  $\Lambda'_j$  has clearly the same action as  $\Lambda_j$  on  $\mathcal{S}_0$ . A joint channel  $\Lambda$  for  $\Lambda'_1$  and  $\Lambda'_2$  is given as

$$\Lambda(\varrho) = \sum_{i=0}^1 \langle i|\varrho|i\rangle \Lambda_1(|i\rangle\langle i|) \otimes \Lambda_2(|i\rangle\langle i|),$$

and it is clear that, in fact,  $\text{Tr}_2[\Lambda(\varrho)] = \Lambda_1(\varrho)$  and  $\text{Tr}_1[\Lambda(\varrho)] = \Lambda_2(\varrho)$ .

For a subset  $\mathcal{S}_0 \subset \mathcal{S}$ , we denote by  $\bar{\mathcal{S}}_0$  the intersection of the linear hull of  $\mathcal{S}_0$  with  $\mathcal{S}$ , i.e.,

$$\bar{\mathcal{S}}_0 = \{\varrho \in \mathcal{S} \mid \varrho = \sum_{i=1}^l c_i \varrho_i \text{ for some } c_i \in \mathbb{C} \text{ and } \varrho_i \in \mathcal{S}_0\}$$

In this definition we can assume without restriction that  $c_i \in \mathbb{R}$  and  $\sum_i c_i = 1$  as they follow from the positivity and unit-trace of states. Since the condition (4.6) is linear in  $\varrho$ , we conclude that devices  $D_1, \dots, D_n$  are  $\mathcal{S}_0$ -compatible if and only if they are  $\bar{\mathcal{S}}_0$ -compatible. This makes sense: if we can simulate the action of devices for states in  $\mathcal{S}_0$ , we can simply calculate the action for all states that are linear combinations of those states. This observation also shows that a reasonable way to quantify the size of a subset  $\mathcal{S}_0$  for the task in question is the number of affinely independent states. We

consider the following questions. Given a collection of incompatible devices  $D_1, \dots, D_n$ ,

- (a) what is the smallest subset  $\mathcal{S}_0$  such that  $D_1, \dots, D_n$  are  $\mathcal{S}_0$ -incompatible?
- (b) what is the largest subset  $\mathcal{S}_0$  such that  $D_1, \dots, D_n$  are  $\mathcal{S}_0$ -compatible?

Smallest and largest here mean the number of affinely independent states in  $\bar{\mathcal{S}}_0$ . It agrees with the linear dimension of the linear hull of  $\mathcal{S}_0$ , or  $\dim \text{aff} \mathcal{S}_0 + 1$ , where  $\dim \text{aff} \mathcal{S}_0$  is the affine dimension of the affine hull  $\text{aff} \mathcal{S}_0$  of  $\mathcal{S}_0$  [49, 54]. The answer to (a) quantifies how many states we need to use to detect incompatibility if we choose them carefully, whereas the answer to (b) quantifies how many (affinely independent) states we may have to use if we cannot control their choice. Hence for both of these quantities lower number means more incompatibility in the sense of easier detection. The precise mathematical definitions read as follows.

**Definition 4.5**

For a collection of incompatible devices  $D_1, \dots, D_n$ , we denote

$$\chi_{incomp}(D_1, \dots, D_n) = \min_{\mathcal{S}_0 \subset \mathcal{S}} \{ \dim \text{aff} \mathcal{S}_0 + 1 \mid D_1, \dots, D_n: \mathcal{S}_0\text{-incompatible} \}$$

and

$$\chi_{comp}(D_1, \dots, D_n) = \max_{\mathcal{S}_0 \subset \mathcal{S}} \{ \dim \text{aff} \mathcal{S}_0 + 1 \mid D_1, \dots, D_n: \mathcal{S}_0\text{-compatible} \}.$$

We call these numbers the *incompatibility dimension* and *compatibility dimension* of  $D_1, \dots, D_n$  respectively.

From Example 4.4 and the fact that the linear dimension of the linear hull of  $\mathcal{S}$  is  $d^2$  we conclude that

$$2 \leq \chi_{incomp}(D_1, \dots, D_n) \leq d^2 \tag{4.7}$$

and

$$d \leq \chi_{comp}(D_1, \dots, D_n) \leq d^2 - 1. \tag{4.8}$$

Further, from the definitions of these quantities it directly follows that

$$\chi_{incomp}(D_1, \dots, D_n) \leq \chi_{comp}(D_1, \dots, D_n) + 1. \tag{4.9}$$

We note that based on their definitions, both  $\chi_{incomp}$  and  $\chi_{comp}$  are expected to be smaller for collections of devices that are more incompatible. The following monotonicity property of  $\chi_{incomp}$  and  $\chi_{comp}$  under pre-processing is a basic property that any quantification of incompatibility is expected to satisfy.

**Proposition 4.6**

Let  $\Lambda : \mathcal{S} \rightarrow \mathcal{S}$  be a quantum channel and let  $\tilde{D}_j$  be a pre-processing of  $D_j$  with  $\Lambda$  for each  $j = 1, \dots, n$ , i.e.,  $\tilde{D}_j(\varrho) = D_j(\Lambda(\varrho))$ . If  $\tilde{D}_j$ 's are incompatible, then also  $D_j$ 's are incompatible and

$$\chi_{incomp}(\tilde{D}_1, \dots, \tilde{D}_n) \geq \chi_{incomp}(D_1, \dots, D_n) \quad (4.10)$$

and

$$\chi_{comp}(\tilde{D}_1, \dots, \tilde{D}_n) \geq \chi_{comp}(D_1, \dots, D_n). \quad (4.11)$$

**Proof**

Suppose that  $D_1, \dots, D_n$  are  $\mathcal{S}_0$ -compatible for some subset  $\mathcal{S}_0$ . Let  $D'$  be a device that gives devices  $D'_1, \dots, D'_n$  as marginals and these marginals satisfy (4.6) in  $\mathcal{S}_0$ . Then the pre-processing of  $D'$  with  $\Lambda$  gives  $\tilde{D}_1, \dots, \tilde{D}_n$  as marginals in  $\mathcal{S}_0$ . The claimed inequalities then follow.  $\square$

The post-processing map of a device  $D$  depends on type of the device. For instance, the output set of an observable is  $P(X)$  and post-processing is then described as a stochastic matrix [143]. We formulate and prove the following monotonicity property of  $\chi_{incomp}$  and  $\chi_{comp}$  under post-processing only for observables. The formulation is analogous for other types of devices.

**Proposition 4.7**

Let  $\tilde{A}_j$  be a post-processing of  $A_j$  (i.e.  $\tilde{A}_j(x') = \sum_x \nu_j(x', x)A_j(x)$  for some stochastic matrix  $\nu_j$ ) for each  $j = 1, \dots, n$ . If  $\tilde{A}_j$ 's are  $\mathcal{S}_0$ -incompatible, then also  $A_j$ 's are  $\mathcal{S}_0$ -incompatible and

$$\chi_{incomp}(\tilde{A}_1, \dots, \tilde{A}_n) \geq \chi_{incomp}(A_1, \dots, A_n) \quad (4.12)$$

and

$$\chi_{comp}(\tilde{A}_1, \dots, \tilde{A}_n) \geq \chi_{comp}(A_1, \dots, A_n). \quad (4.13)$$

**Proof**

Suppose that  $A_1, \dots, A_n$  are  $\mathcal{S}_0$ -compatible for some subset  $\mathcal{S}_0$ . This means that there exists an observable  $G$  satisfying for all  $\varrho \in \mathcal{S}_0$ , any  $j$  and  $x_j$ ,

$$\text{Tr}[\varrho A_j(x_j)] = \sum_{l \neq j} \sum_{x_l} \text{Tr}[\varrho G(x_1, \dots, x_n)]. \quad (4.14)$$

We define an observable  $\tilde{G}$

$$\tilde{G}(x'_1, \dots, x'_n) = \sum_{x_1, \dots, x_n} \nu(x'_1|x_1) \cdots \nu(x'_n|x_n) G(x_1, \dots, x_n),$$

and it then satisfies

$$\text{Tr}[\varrho \tilde{A}_j(x'_j)] = \sum_{l \neq j} \sum_{x'_l} \text{Tr}[\varrho \tilde{G}(x'_1, \dots, x'_n)] \quad (4.15)$$

for all  $\varrho \in \mathcal{S}_0$ , any  $j$  and  $x'_j$ . This shows that  $\tilde{\mathbf{A}}_1, \dots, \tilde{\mathbf{A}}_n$  are  $\mathcal{S}_0$ -compatible. The claimed inequalities then follow.  $\square$

We will now have some examples to demonstrate the values of  $\chi_{incomp}$  and  $\chi_{comp}$  in some standard cases.

*Example 4.8*

Let us consider the identity channel  $\text{id} : \mathcal{S}(\mathbb{C}^d) \rightarrow \mathcal{S}(\mathbb{C}^d)$ . It follows from the definitions that two identity channels are  $\mathcal{S}_0$ -compatible if and only if  $\mathcal{S}_0$  is a broadcastable set. It is known that a subset of states is broadcastable only if the states commute with each other [144], and for this reason the pair of two identity channels is  $\mathcal{S}_0$ -incompatible whenever  $\mathcal{S}_0$  contains two noncommuting states. Therefore, we have  $\chi_{incomp}(\text{id}, \text{id}) = 2$ . On the other hand,  $\mathcal{S}_0$  consisting of distinguishable states makes the identity channels  $\mathcal{S}_0$ -compatible. As  $\mathcal{S}_0$  consisting of commutative states has at most  $d$  affinely independent states, we conclude that  $\chi_{comp}(\text{id}, \text{id}) = d$ .

A comparison of the results of Example 4.8 to the bounds (4.7) and (4.8) shows that the pair of identity channels has the smallest possible incompatibility and compatibility dimensions. This is quite expectable as that pair is consider to be the most incompatible pair - any device can be post-processed from the identity channel. Perhaps surprisingly, the lower bound of  $\chi_{incomp}$  can be attained already with a pair of dichotomic observables; this is shown in the next example.

*Example 4.9*

Let  $P$  and  $Q$  be two noncommuting one-dimensional projections in a  $d$ -dimensional Hilbert space  $\mathcal{H}$ . We define two dichotomic observables  $\mathbf{A}$  and  $\mathbf{B}$  as

$$\mathbf{A}(1) = P, \mathbf{A}(0) = \mathbb{1} - P, \quad \mathbf{B}(1) = Q, \mathbf{B}(0) = \mathbb{1} - Q.$$

Let us then consider a subset consisting of two states,

$$\mathcal{S}_0 = \{\varrho^P, \varrho^Q\} := \left\{ \frac{1}{d-1}(\mathbb{1} - P), \frac{1}{d-1}(\mathbb{1} - Q) \right\}.$$

We find that the dichotomic observables  $\mathbf{A}$  and  $\mathbf{B}$  are  $\mathcal{S}_0$ -incompatible. To see this, let us make a counter assumption that  $\mathbf{A}$  and  $\mathbf{B}$  are  $\mathcal{S}_0$ -compatible, in which case there exists  $\mathbf{G}$  such that the marginal condition (4.3) holds for both observables and for all  $\varrho \in \mathcal{S}_0$ . We have  $\text{Tr}[\varrho^P \mathbf{A}(1)] = 0$  and therefore

$$0 = \text{Tr}[(\mathbb{1} - P)\mathbf{G}(1, 1)] = \text{Tr}[(\mathbb{1} - P)\mathbf{G}(1, 0)].$$

It follows that  $\mathbf{G}(1, 1) = \alpha P$  and  $\mathbf{G}(1, 0) = \beta P$ . Further,  $\text{Tr}[P\mathbf{A}(1)] = 1$  and hence  $\alpha + \beta = 1$ . In a similar way we obtain  $\mathbf{G}(1, 1) = \gamma Q$  and

$G(0, 1) = \delta Q$  with  $\gamma + \delta = 1$ . It follows that  $\alpha = \gamma = 0$  and  $\beta = \delta = 1$ . But  $G(1, 0) + G(0, 1) = P + Q$  contradicts  $G(1, 0) + G(0, 1) \leq \mathbb{1}$ . Thus we conclude  $\chi_{incomp}(\mathbf{A}, \mathbf{B}) = 2$ .

For two incompatible sharp qubit observables (Example 4.1) the previous example gives a concrete subset of two states such that the observables are incompatible and proves that  $\chi_{incomp}(\mathbf{A}^{\mathbf{a}}, \mathbf{A}^{\mathbf{b}}) = 2$  for such a pair. The incompatibility dimension for unsharp qubit observables is more complicated and will be treated in Section 4.3.

*Example 4.10*

Let us consider two observables  $\mathbf{A}$  and  $\mathbf{B}$ . Fix a state  $\varrho_0 \in \mathcal{S}$  and define

$$\mathcal{S}_0 = \{\varrho \in \mathcal{S} : \text{Tr}[\varrho \mathbf{A}(x)] = \text{Tr}[\varrho_0 \mathbf{A}(x)] \forall x\}.$$

Then  $\mathbf{A}$  and  $\mathbf{B}$  are  $\mathcal{S}_0$ -compatible. To see this, we define an observable  $\mathbf{G}$  as

$$\mathbf{G}(x, y) = \text{Tr}[\varrho_0 \mathbf{A}(x)] \mathbf{B}(y).$$

It is then straightforward to verify that (4.3) holds for all  $\varrho \in \mathcal{S}_0$ . As a special instance of this construction, let  $\mathbf{A}^{\mathbf{a}}$  be a qubit observable and  $\mathbf{a} \neq 0$  (see Example 4.1). We choose  $\mathcal{S}_0 = \{\varrho \in \mathcal{S} \mid \text{Tr}[\varrho \mathbf{A}^{\mathbf{a}}(+)] = \frac{1}{2}\}$ . We then have  $\mathcal{S}_0 = \{\frac{1}{2}(\mathbb{1} + \mathbf{r} \cdot \sigma) \mid \mathbf{r} \cdot \mathbf{a} = 0\}$  and hence  $\dim_{\text{aff}} \mathcal{S}_0 = 2$ . Based on the previous argument,  $\mathbf{A}^{\mathbf{a}}$  is  $\mathcal{S}_0$ -compatible with any  $\mathbf{A}^{\mathbf{b}}$ . Therefore,  $\chi_{comp}(\mathbf{A}^{\mathbf{a}}, \mathbf{A}^{\mathbf{b}}) = 3$  for all incompatible qubit observables  $\mathbf{A}^{\mathbf{a}}$  and  $\mathbf{A}^{\mathbf{b}}$ .

### 4.1.3 Remarks on other formulations of incompatibility dimension

The notion of  $\mathcal{S}_0$ -compatibility for quantum observables has been introduced in [132] and in that particular case (i.e. quantum observables) it is equivalent to Definition 4.3. In the current investigation, our focus is on the largest or smallest  $\mathcal{S}_0$  on which devices  $\mathbf{D}_1, \dots, \mathbf{D}_n$  are compatible or incompatible, and this has some differences to the earlier approaches. In [134], the term ‘‘compatibility dimension’’ was introduced and for observables  $\mathbf{A}_1, \dots, \mathbf{A}_n$  on a  $d$ -dimensional Hilbert space  $\mathcal{H} = \mathbb{C}^d$ : it is given by

$$R(\mathbf{A}_1, \dots, \mathbf{A}_n) = \max\{r \leq d \mid \exists V: \mathbb{C}^r \rightarrow \mathbb{C}^d \text{ isometry} \\ \text{s.t. } V^* \mathbf{A}_1 V, \dots, V^* \mathbf{A}_n V \text{ are compatible}\},$$

Evaluations of  $R(\mathbf{A}_1, \dots, \mathbf{A}_n)$  in various cases such as  $n = 2$  and  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are rank-1 were presented in [134]. To describe it in our notions, let us denote  $\mathbb{C}^r$  by  $\mathcal{K}$ , and define  $\mathcal{S}_{\mathcal{H}}$  and  $\mathcal{S}_{\mathcal{K}}$  as the set of all density operator on

$\mathcal{H}$  and  $\mathcal{K}$  respectively. We also introduce  $\mathcal{S}_{V\mathcal{K}}$  as

$$\mathcal{S}_{V\mathcal{K}} := \{\varrho \in \mathcal{S} \mid \text{supp}\varrho \subset V\mathcal{K}\} = V\mathcal{S}_{\mathcal{K}}V^* \subset \mathcal{S}_{\mathcal{H}}.$$

Then, we can see that the  $\mathcal{S}_{\mathcal{K}}$ -compatibility of  $V^*A_1V, \dots, V^*A_nV$  is equivalent to the  $\mathcal{S}_{V\mathcal{K}}$ -compatibility of  $A_1, \dots, A_n$ . Therefore, if we focus only on sets of states such as  $\mathcal{S}_{V\mathcal{K}}$  (i.e. states with fixed support), then there is no essential difference between our compatibility dimension and the previous one:  $R(A_1, \dots, A_n) = r$  iff  $\chi_{comp}(A_1, \dots, A_n) = r^2$ . In [134] also the concept of “strong compatibility dimension” was defined as

$$\begin{aligned} \bar{R}(A_1, \dots, A_n) = \max\{r \leq d \mid \forall V: \mathbb{C}^r \rightarrow \mathbb{C}^d \text{ isometry} \\ \text{s.t. } V^*A_1V, \dots, V^*A_nV \text{ are compatible}\}. \end{aligned}$$

It is related to our notion of incompatibility dimension. In fact, if we only admit sets of states such as  $\mathcal{S}_{V\mathcal{K}}$ , then  $\bar{R}(A_1, \dots, A_n)$  and  $\chi_{incomp}(A_1, \dots, A_n)$  are essentially the same:  $\bar{R}(A_1, \dots, A_n) = r$  iff  $\chi_{incomp}(A_1, \dots, A_n) = (r+1)^2$ .

Similar notions have been introduced and investigated also in [133, 135]. As in [134], these works focus on quantum observables and on subsets of states that are lower dimensional subspaces of the original state space. Therefore, the notions are not directly applicable in GPTs. In [135] incompatibility is classified into three types. They are explained exactly in terms of [134] as

- (i) incompressive incompatibility:  $(A_1, \dots, A_n)$  are  $\mathcal{S}_{V\mathcal{K}}$ -compatible for all  $\mathcal{K}$  and  $V$
- (ii) fully compressive incompatibility:  $(A_1, \dots, A_n)$  are  $\mathcal{S}_{V\mathcal{K}}$ -incompatible for all nontrivial  $\mathcal{K}$  and  $V$
- (iii) partly compressive incompatibility: there is a  $V$  and  $\mathcal{K}$  such that  $(A_1, \dots, A_n)$  are  $\mathcal{S}_{V\mathcal{K}}$ -compatible, and some  $V'$  and  $\mathcal{K}'$  such that  $(A_1, \dots, A_n)$  are  $\mathcal{S}_{V'\mathcal{K}'}$ -incompatible.

In [135] concrete constructions of these three types of incompatible observables were given.

## 4.2 Incompatibility dimension and incompatibility witness

In this section we show how the notion of incompatibility dimension is related to the notion of incompatibility witness.

### 4.2.1 Relation between incompatibility dimension and incompatibility witness for observables

An *incompatibility witness* is an affine functional  $\xi$  defined on  $n$ -tuples of observables such that  $\xi$  takes non-negative values on all compatible  $n$ -tuples and a negative value at least for some incompatible  $n$ -tuple [16, 17, 112]. Every incompatibility witness  $\xi$  is of the form

$$\xi(\oplus_{j=1}^n \mathbf{A}_j) = \delta - f(\oplus_{j=1}^n \mathbf{A}_j), \quad (4.16)$$

where  $\delta \in \mathbb{R}$  and  $f$  is a linear functional on  $\oplus_{j=1}^n \mathcal{L}_s(\mathcal{H})^{m_j}$  with  $\mathcal{L}_s(\mathcal{H})$  being the set of all self-adjoint operators on  $\mathcal{H}$  and  $m_j$  the number of outcomes of  $\mathbf{A}_j$ . It can be written also in the form

$$\xi(\mathbf{A}_1, \dots, \mathbf{A}_n) = \delta - \sum_{j=1}^n \sum_{x_j=1}^{m_j} c_{j,x_j} \text{Tr}[\varrho_{j,x_j} \mathbf{A}_j(x_j)], \quad (4.17)$$

where  $c_{j,x_j}$ 's are real numbers, and  $\varrho_{j,x_j}$ 's are states. This result has been proven in [17] for incompatibility witnesses acting on pairs of observables and the generalization to  $n$ -tuples is straightforward. A witness  $\xi$  *detects* the incompatibility of observables  $\mathbf{A}_1, \dots, \mathbf{A}_n$  if  $\xi(\mathbf{A}_1, \dots, \mathbf{A}_n) < 0$ . The following proposition gives a simple relation between incompatibility dimension and incompatibility witness.

#### Proposition 4.11

*Assume that an incompatibility witness  $\xi$  has the form (4.17) and it detects the incompatibility of observables  $\mathbf{A}_1, \dots, \mathbf{A}_n$ . Then  $\mathbf{A}_1, \dots, \mathbf{A}_n$  are  $\mathcal{S}_0$ -incompatible for  $\mathcal{S}_0 = \{\varrho_{j,x_j} \mid j = 1, \dots, n, x_j = 1, \dots, m_j\}$ .*

#### Proof

Let  $\mathbf{A}_1, \dots, \mathbf{A}_n$  be  $\mathcal{S}_0$ -compatible. Then we have compatible observables  $\tilde{\mathbf{A}}_1, \dots, \tilde{\mathbf{A}}_n$  such that  $\text{Tr}[\varrho \mathbf{A}_j(x_j)] = \text{Tr}[\varrho \tilde{\mathbf{A}}_j(x_j)]$  for all  $\varrho \in \mathcal{S}_0$ . This implies that

$$\xi(\mathbf{A}_1, \dots, \mathbf{A}_n) = \xi(\tilde{\mathbf{A}}_1, \dots, \tilde{\mathbf{A}}_n) \geq 0,$$

which contradicts to the assumption that  $\xi$  detects the incompatibility of observables  $\mathbf{A}_1, \dots, \mathbf{A}_n$ .  $\square$

It has been shown in [17] that any incompatible pair of observables is detected by some incompatibility witness of the form (4.17). The proof is straightforward to generalize to  $n$ -tuples of observables, and thus, together with Proposition 4.11, we can obtain

$$\chi_{incomp}(\mathbf{A}_1, \dots, \mathbf{A}_n) \leq m_1 + \dots + m_n. \quad (4.18)$$



That is, the incompatibility dimension of  $\mathbf{A}_1, \dots, \mathbf{A}_n$  can be evaluated via their incompatibility witness (we will derive a better upper bound later in this section). We can further prove the following proposition.

**Proposition 4.12**

The statements (i) and (ii) for a set of incompatible observables  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  are equivalent:

(i)  $\chi_{incomp}(\mathbf{A}_1, \dots, \mathbf{A}_n) \leq N$

(ii) There exist a family of linearly independent states  $\{\varrho_1, \dots, \varrho_N\}$  and real numbers  $\delta$  and  $\{c_{l,j,x_j}\}_{l,j,x_j}$  ( $l = 1, \dots, N, j = 1, \dots, n, x_j = 1, \dots, m_j$ ) such that the incompatibility witness  $\xi$  defined by

$$\xi(\mathbf{B}_1, \dots, \mathbf{B}_n) = \delta - \sum_{l=1}^N \sum_{j=1}^n \sum_{x_j=1}^{m_j} c_{l,j,x_j} \text{tr}[\varrho_l \mathbf{B}_j(x_j)]$$

detects the incompatibility of  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$ .

The claim (i)  $\Rightarrow$  (ii) may be regarded as the converse of the previous argument to obtain (4.18). It manifests that we can find an incompatibility witness detecting the incompatibility of  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  reflecting their incompatibility dimension.

**Proof**

(ii)  $\Rightarrow$  (i) can be proven in the same way as Proposition 4.11. Thus, we focus on proving (i)  $\Rightarrow$  (ii). Suppose that a family of observables  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  satisfies  $\chi_{incomp}(\mathbf{A}_1, \dots, \mathbf{A}_n) = N$ . Then there exists a family of linearly independent states  $\{\varrho_1, \varrho_2, \dots, \varrho_N\}$  in  $\mathcal{L}_s(\mathcal{H})$  on which  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  are incompatible. We can regard the family  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  as an element of a vector space  $\mathcal{L}$  defined as  $\mathcal{L} := \bigoplus_{j=1}^n \mathcal{L}_s(\mathcal{H})^{m_j}$ , that is,  $\mathbf{A} := \bigoplus_{j=1}^n \mathbf{A}_j \in \mathcal{L}$ . For each  $l = 1, \dots, N, j = 1, \dots, n$ , and  $x_j = 1, \dots, m_j$ , let us define a subset  $K(\mathbf{A}, \varrho_l, j, x_j)$  of  $\mathcal{L}$  as

$$K(\mathbf{A}, \varrho_l, j, x_j) := \{\mathbf{B} \in \mathcal{L} \mid \langle \varrho_l | \mathbf{B}_j(x_j) \rangle_{HS} = \langle \varrho_l | \mathbf{A}_j(x_j) \rangle_{HS}\}, \quad (4.19)$$

where  $\langle \varrho_l | \mathbf{A}_j(x_j) \rangle_{HS} := \text{tr}[\varrho_l \mathbf{A}_j(x_j)]$  is the Hilbert-Schmidt inner product on  $\mathcal{L}_s(\mathcal{H})$ . Note that this inner product can be naturally extended to an inner product  $\langle\langle \cdot | \cdot \rangle\rangle$  on  $\mathcal{L}$ :

$$\langle\langle \mathbf{A} | \mathbf{B} \rangle\rangle = \sum_{j=1}^n \sum_{x_j=1}^{m_j} \langle \mathbf{A}_j(x_j) | \mathbf{B}_j(x_j) \rangle_{HS}.$$

Embedding  $\varrho_l$  into  $\mathcal{L}$  by  $\hat{\varrho}_l^{j,x_j} = \bigoplus_{i=1}^n \bigoplus_{y=1}^{m_i} \delta_{ij} \delta_{yx_j} \varrho_l$  for each  $j, x_j$  and  $l$ , we

obtain another representation of (4.19) as

$$K(\mathbf{A}, \varrho_l, j, x_j) = \{\mathbf{B} \mid \langle\langle \hat{\varrho}_l^{j,x} | \mathbf{B} \rangle\rangle = \langle\langle \hat{\varrho}_l^{j,x_j} | \mathbf{A} \rangle\rangle\}. \quad (4.20)$$

Thus this set is a hyperplane in  $\mathcal{L}$ . Note that  $\{\hat{\varrho}_l^{j,x}\}_{l,j,x_j}$  is a linearly independent set in  $\mathcal{L}$ . Consider an affine set  $K := \cap_{l=1}^N \cap_{j=1}^n \cap_{x_j=1}^{m_j} K(\mathbf{A}, \varrho_l, j, x_j)$ . Because  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  is incompatible in  $\{\varrho_1, \dots, \varrho_N\}$ , it satisfies

$$K \cap C = \emptyset, \quad (4.21)$$

where  $C := \{\mathbf{C} \in \mathcal{L} \mid \{\mathbf{C}_1, \dots, \mathbf{C}_n\} \text{ is compatible}\}$ . Thus, by the separating hyperplane theorem [49], there exists a hyperplane in  $\mathcal{L}$  which separates strongly the (closed) convex sets  $K$  and  $C$ . In the following, we will show that one of those separating hyperplanes can be constructed from  $\{\hat{\varrho}_l^{j,x}\}_{l,j,x_j}$ .

Let us extend a family of linearly independent vectors  $\{\hat{\varrho}_l^{j,x_j}\}_{l,j,x_j}$  to form a basis of  $\mathcal{L}$ . That is, we introduce a basis  $\{w_b\}_{b=1, \dots, \dim \mathcal{L}}$  of  $\mathcal{L}$  satisfying  $\{v_a\}_{a=1, \dots, N(\sum_j m_j)} = \{\hat{\varrho}_l^{j,x_j}\}_{l,j,x_j}$ . We introduce its dual basis  $\{w_b\}_{b=1, 2, \dots, \dim \mathcal{L}}$  satisfying  $\langle\langle v_a | w_b \rangle\rangle = \delta_{ab}$ . Because  $K$  can be written as

$$K = \{\mathbf{B} \mid \langle\langle \hat{\varrho}_l^{j,x_j} | (\mathbf{B} - \mathbf{A}) \rangle\rangle = 0, \forall l, j, x_j\},$$

it is represented in terms this (dual) basis as

$$K = \mathbf{A} + K_0,$$

where  $K_0$  is an affine set defined by

$$K_0 := \left\{ \sum_{a=N(\sum_j m_j)+1}^{\dim \mathcal{L}} c_a w_a \mid c_a \in \mathbb{R} \right\} \quad (4.22)$$

Now we can construct a hyperplane separating  $K$  and  $C$ . To do this, let us focus on the convex sets  $K_0$  and  $C' := C - \mathbf{A}$  instead of  $K$  and  $C$ , which satisfy  $K_0 \cap C' = \emptyset$  because of (4.21). We can apply the separating hyperplane theorem (Theorem 11.2 in [49]) for the affine set  $K_0$  and convex set  $C'$ . There exists a hyperplane  $H_0$  in  $\mathcal{L}$  such that  $K_0$  and  $C'$  are contained by  $H_0$  and one of its associating open half-spaces respectively. That is, there exists  $h \in \mathcal{L}$  satisfying

$$H_0 = \{\mathbf{B} \in \mathcal{L} \mid \langle\langle \mathbf{B} | h \rangle\rangle = 0\}$$

with  $K_0 \subset H_0$ , and  $\langle\langle C' | h \rangle\rangle < 0$  for all  $C' \in C'$ . Let us examine the vector  $h$ . It satisfies

$$\langle\langle w_a | h \rangle\rangle = 0 \quad \text{for all } a = N(\sum_j m_j) + 1, \dots, \dim \mathcal{L}$$

because  $K_0 \subset H_0$  (see (4.22)). Thus, if we write  $h$  as  $h = \sum_{a=1}^{\dim \mathcal{L}} c_a v_a$ , then we can find that  $c_a = 0$  holds for all  $a = N(\sum_j m_j) + 1, \dots, \dim \mathcal{L}$ . It follows that

$$h = \sum_{a=1}^{N(\sum_j m_j)} c_a v_a = \sum_l \sum_j \sum_{x_j} c_{l,j,x_j} \hat{\varrho}_l^{j,x_j}$$

holds, and the hyperplane  $H_0$  can be written as

$$H_0 = \{\mathbf{B} \in \mathcal{L} \mid \sum_l \sum_j \sum_{x_j} c_{l,j,x_j} \text{Tr}[\varrho_l \mathbf{B}_j(x_j)] = 0\}.$$

Then, the hyperplane  $H' := \mathbf{A} + H_0$ , a translation of  $H_0$ , of the form

$$H' = \{\mathbf{B} \in \mathcal{L} \mid \sum_l \sum_j \sum_{x_j} c_{l,j,x_j} \text{Tr}[\varrho_l \mathbf{B}_j(x_j)] = \delta'\}$$

contains the original sets  $K$ , and satisfy

$$\sum_l \sum_j \sum_{x_j} c_{l,j,x_j} \text{Tr}[\varrho_l \mathbf{C}_j(x_j)] < \delta'$$

for all  $\mathbf{C} \in C$ . We can displace  $H'$  slightly in the direction of  $C$  to obtain a hyperplane  $H$  defined as

$$H = \{\mathbf{B} \in \mathcal{L} \mid \sum_l \sum_j \sum_{x_j} c_{l,j,x_j} \text{Tr}[\rho_l \mathbf{B}_j(x_j)] = \delta\},$$

which (strongly) separates  $H'$  (in particular  $K$ ) and  $C$  because  $H'$  is closed and  $C$  is compact (see Corollary 11.4.2 in [49]). The claim now follows as  $\mathbf{A} \in K$ .  $\square$

## 4.2.2 An upper bound on the incompatibility dimension of observables via incompatibility witness

We can give a better upper bound than (4.18) for the incompatibility dimension by slightly modifying the previous argument in [17] on incompatibility witness.

### Proposition 4.13

Let  $\mathbf{A}_1, \dots, \mathbf{A}_n$  be incompatible observables with  $m_1, \dots, m_n$  outcomes, respectively. Then

$$\chi_{\text{incomp}}(\mathbf{A}_1, \dots, \mathbf{A}_n) \leq \sum_{j=1}^n m_j - n + 1.$$

**Proof**

We continue following the same notations as the proof of Proposition 4.12. Let us assume that the incompatibility of  $\mathbf{A}_1, \dots, \mathbf{A}_n$  is detected by an incompatibility witness  $\xi$ . The functional  $\xi$  is of the form

$$\xi(\mathbf{A}) = \delta - f(\mathbf{A})$$

with a real number  $\delta$  and a functional  $f$  on  $\mathcal{L}$  (see (4.16)). Then, Riesz representation theorem shows that the functional  $f$  can be represented as

$$f(\mathbf{A}) = \sum_{j=1}^n \sum_{x_j}^{m_j} \langle F_j(x_j) | \mathbf{A}_j(x_j) \rangle_{HS}$$

with some  $F_j(x_j) \in \mathcal{L}_s(\mathcal{H})$  ( $j = 1, \dots, n$ ,  $x_j = 1, \dots, m_j$ ). If we define  $F'_j(x_j) = F_j(x_j) + \epsilon_j \mathbb{1}$ , then we find

$$\xi(\mathbf{A}) = \delta + d \sum_j \epsilon_j - \sum_{j=1}^n \sum_{x_j=1}^{m_j} \langle F'_j(x_j) | \mathbf{A}_j(x_j) \rangle_{HS}.$$

We choose  $\epsilon_j$  so that

$$\sum_{x_j} \text{tr}[F'_j(x_j)] = \sum_{x_j} \langle F'_j(x_j) | \mathbb{1} \rangle_{HS} = 0$$

holds. The choice of  $\{F'_j(x_j)\}_{j,x_j}$  has still some freedom. Each  $F'_j(x_j)$  can be replaced with  $F''_j(x_j) = F'_j(x_j) + T_j$ , where  $T_j \in \mathcal{L}_s(\mathcal{H})$  satisfies  $\text{tr}[T_j] = \langle T_j | \mathbb{1} \rangle_{HS} = 0$ . In fact, it holds that

$$\begin{aligned} \sum_{x_j} \langle F''_j(x_j) | \mathbf{A}_j(x_j) \rangle_{HS} &= \sum_{x_j} \langle F'_j(x_j) | \mathbf{A}_j(x_j) \rangle_{HS} + \sum_{x_j} \langle T_j | \mathbf{A}_j(x_j) \rangle_{HS} \\ &= \sum_{x_j} \langle F'_j(x_j) | \mathbf{A}_j(x_j) \rangle_{HS} + \langle T_j | \mathbb{1} \rangle_{HS} \\ &= \sum_{x_j} \langle F'_j(x_j) | \mathbf{A}_j(x_j) \rangle_{HS}. \end{aligned}$$

We choose  $T_j$  as  $m_j T_j = -\sum_{x_j=1}^{m_j} F'_j(x_j)$  which indeed satisfies

$$m_j \langle T_j | \mathbb{1} \rangle_{HS} = -\sum_{x_j=1}^{m_j} \langle F'_j(x_j) | \mathbb{1} \rangle_{HS} = 0,$$

i.e.,  $\text{Tr}[T_j] = 0$ , to obtain

$$\sum_{x_j} F''_j(x_j) = 0.$$

We further choose large numbers  $\alpha_j \geq 0$  so that  $G_j(x_j) := F_j''(x_j) + \alpha_j \mathbb{1} \geq 0$  for all  $j$  and  $x_j$ . Now we obtain a representation of the witness which is equivalent to  $\xi$  for  $n$ -tuples of observables as

$$\xi^*(\mathbf{A}) = \delta + d \sum_j (\epsilon_j + \alpha_j) - \sum_j \sum_{x_j} \langle G_j(x_j) | \mathbf{A}_j(x_j) \rangle_{HS},$$

where positive operators  $G_j(x_j)$ 's satisfy  $\sum_{x_j} G_j(x_j) = m_j \alpha_j \mathbb{1}$ . Defining density operators  $\varrho_j(x_j)$  by  $\varrho_j(x_j) = \frac{G_j(x_j)}{\text{tr}[G_j(x_j)]}$ , we obtain yet another representation

$$\xi^*(\mathbf{A}) = \delta + d \sum_j (\epsilon_j + \alpha_j) - \sum_j \sum_{x_j} \text{tr}[G_j(x_j)] \text{tr}[\varrho_j(x_j) \mathbf{A}_j(x_j)]$$

with  $\varrho_j(x_j)$ 's satisfying constraints

$$\sum_{x_j} \text{tr}[G_j(x_j)] \varrho_j(x_j) = m_j \alpha_j \mathbb{1}. \quad (4.23)$$

Thus, according to Proposition 4.11,  $\mathbf{A}_1, \dots, \mathbf{A}_n$  are  $\mathcal{S}_0$ -incompatible with  $\mathcal{S}_0 = \{\varrho_j(x_j)\}_{j,x_j}$ . To evaluate  $\text{dim aff } \mathcal{S}_0$ , we focus on the condition (4.23). Introducing parameters  $p_j(x_j) := \text{tr}[G_j(x_j)]/dm_j \alpha_j$  such that  $\sum_{x_j} p_j(x_j) = 1$ , we obtain

$$\sum_{x_j} p_j(x_j) \varrho_j(x_j) = \frac{1}{d} \mathbb{1},$$

or

$$\sum_{x_j} p_j(x_j) \tilde{\varrho}_j(x_j) = 0,$$

where  $\tilde{\varrho}_j(x_j) := \varrho_j(x_j) - \frac{1}{d} \mathbb{1}$ . It follows that  $\{\tilde{\varrho}_j(x_j)\}_{x_j}$  are linearly dependent, and thus

$$\text{dimspan}\{\tilde{\varrho}_j(x_j)\}_{x_j} \leq m_j - 1.$$

Similar arguments for the other  $j$ 's result in

$$\text{dimspan}\{\tilde{\varrho}_j(x_j)\}_{j,x_j} \leq \sum_j (m_j - 1) = \sum_j m_j - n.$$

Considering that

$$\text{dimspan}\{\tilde{\varrho}_j(x_j)\}_{j,x_j} = \text{dim aff}\{\varrho_j(x_j)\}_{j,x_j}$$

holds, we can obtain the claim of the proposition.  $\square$

The bound in Proposition 4.13 is not tight in general since the right-hand side of the inequality can exceed the bound obtained in (4.7). However, for small  $n$  and  $m_j$ 's, the bound can be tight. In fact, while for  $n = 2$  and

$m_1 = m_2 = 2$  it gives  $\chi_{incomp}(\mathbf{A}_1, \mathbf{A}_2) \leq 3$ , we will construct an example which attains this upper bound in the next section.

### 4.3 (In)compatibility dimension for mutually unbiased qubit observables

In this section we study the incompatibility dimension of pairs of unbiased qubit observables introduced in Example 4.1. We concentrate on pairs that are mutually unbiased, i.e.,  $\text{Tr}[\mathbf{A}^{\mathbf{a}}(\pm)\mathbf{A}^{\mathbf{b}}(\pm)] = 1/2$  (this terminology originates from the fact that if the observables are sharp, then the respective orthonormal bases are mutually unbiased. In the previously written form the definition makes sense also for unsharp observables [145]). The condition of mutual unbiasedness is invariant under a global unitary transformation, hence it is enough to fix the basis  $\mathbf{x} = (1, 0, 0)$ ,  $\mathbf{y} = (0, 1, 0)$ ,  $\mathbf{z} = (0, 0, 1)$  in  $\mathbb{R}^3$  and choose two of these unit vectors. We will study the observables  $\mathbf{A}^{t\mathbf{x}}$  and  $\mathbf{A}^{t\mathbf{y}}$ , where  $0 \leq t \leq 1$ . The observables are written explicitly as

$$\mathbf{A}^{t\mathbf{x}}(\pm) = \frac{1}{2}(\mathbb{1} \pm t\sigma_1), \quad \mathbf{A}^{t\mathbf{y}}(\pm) = \frac{1}{2}(\mathbb{1} \pm t\sigma_2).$$

The condition (4.2) shows that  $\mathbf{A}^{t\mathbf{x}}$  and  $\mathbf{A}^{t\mathbf{y}}$  are incompatible if and only if  $1/\sqrt{2} < t \leq 1$ . The choice of having mutually unbiased observables as well as using a single noise parameter instead of two is to simplify the calculations.

We have seen in Example 4.10 that  $\chi_{comp}(\mathbf{A}^{t\mathbf{x}}, \mathbf{A}^{t\mathbf{y}}) = 3$  for all values  $t$  for which the pair is incompatible. We have further seen (discussion after Example 4.9) that  $\chi_{incomp}(\mathbf{A}^{\mathbf{x}}, \mathbf{A}^{\mathbf{y}}) = 2$ , and from Prop. 4.13 follows that  $\chi_{incomp}(\mathbf{A}^{t\mathbf{x}}, \mathbf{A}^{t\mathbf{y}}) \leq 3$  for all  $1/\sqrt{2} < t \leq 1$ . The remaining question is then about the exact value of  $\chi_{incomp}(\mathbf{A}^{t\mathbf{x}}, \mathbf{A}^{t\mathbf{y}})$ , which can depend on the noise parameter  $t$  and will be in our focus in this section (see Table 4.1).

	$\chi_{incomp}(\mathbf{A}^{t\mathbf{x}}, \mathbf{A}^{t\mathbf{y}})$	$\chi_{comp}(\mathbf{A}^{t\mathbf{x}}, \mathbf{A}^{t\mathbf{y}})$
$t \leq \frac{1}{\sqrt{2}}$	-	-
$\frac{1}{\sqrt{2}} < t < 1$	2 or 3 (Proposition 4.14)	3 (Example 4.10)
$t = 1$	2 (Example 4.9)	

**Table 4.1:**  $\chi_{incomp}$  and  $\chi_{comp}$  for  $(\mathbf{A}^{t\mathbf{x}}, \mathbf{A}^{t\mathbf{y}})$  with  $0 \leq t \leq 1$ . For  $t \leq 1/\sqrt{2}$  the observables  $\mathbf{A}^{t\mathbf{x}}$  and  $\mathbf{A}^{t\mathbf{y}}$  are compatible and  $\chi_{incomp}$  and  $\chi_{comp}$  are not defined.

Let us first make a simple observation that follows from Prop. 4.7. Considering that  $\mathbf{A}^{s\mathbf{x}}$  is obtained as a post-processing of  $\mathbf{A}^{t\mathbf{x}}$  if and only if  $s \leq t$ ,

we conclude that

$$\chi_{incomp}(\mathbf{A}^{sx}, \mathbf{A}^{sy}) = 2 \Rightarrow \chi_{incomp}(\mathbf{A}^{tx}, \mathbf{A}^{ty}) = 2 \quad \text{for } \frac{1}{\sqrt{2}} < s \leq t,$$

and

$$\chi_{incomp}(\mathbf{A}^{s'x}, \mathbf{A}^{s'y}) = 3 \Rightarrow \chi_{incomp}(\mathbf{A}^{t'x}, \mathbf{A}^{t'y}) = 3 \quad \text{for } s' \geq t' > \frac{1}{\sqrt{2}}.$$

Interestingly, there is a threshold value  $t_0$  where the value of  $\chi_{incomp}(\mathbf{A}^{tx}, \mathbf{A}^{ty})$  changes; this is the content of the following proposition.

**Proposition 4.14**

*There exists  $1/\sqrt{2} < t_0 < 1$  such that  $\chi_{incomp}(\mathbf{A}^{tx}, \mathbf{A}^{ty}) = 3$  for  $1/\sqrt{2} < t \leq t_0$  and  $\chi_{incomp}(\mathbf{A}^{tx}, \mathbf{A}^{ty}) = 2$  for  $t_0 < t \leq 1$ .*

The main line of the lengthy proof of Proposition 4.14 is the following. Defining two subsets  $L$  and  $M$  of  $(\frac{1}{\sqrt{2}}, 1]$  as

$$L := \{t \mid \chi_{incomp}(\mathbf{A}^{tx}, \mathbf{A}^{ty}) = 2\}, \quad M := \{t \mid \chi_{incomp}(\mathbf{A}^{tx}, \mathbf{A}^{ty}) = 3\}, \quad (4.24)$$

we see that

$$\inf L = \sup M (=: t'_0) \quad (4.25)$$

holds unless  $L$  and  $M$  are empty. By its definition, the number  $t'_0$  satisfies

$$\chi_{incomp}(\mathbf{A}^{tx}, \mathbf{A}^{ty}) = 2 \quad \text{for } t > t'_0, \quad \chi_{incomp}(\mathbf{A}^{tx}, \mathbf{A}^{ty}) = 3 \quad \text{for } t < t'_0.$$

Based on the considerations above, the proof of Proposition 4.14 proceeds as follows. First, in Part 1 - 3 (Subsection 4.3.1 - 4.3.3), we prove that  $M$  is nonempty while  $L$  has already been shown to be nonempty as  $t = 1 \in L$ . It will be found that  $\chi_{incomp}(\mathbf{A}^{tx}, \mathbf{A}^{ty}) = 3$  for  $t$  sufficiently close to  $\frac{1}{\sqrt{2}}$ , and thus  $t'_0$  introduced above can be defined successfully. Then, we demonstrate in Part 4 (Subsection 4.3.4) that  $\sup M = \max M$ , i.e.  $t'_0$  equals to  $t_0$  in the claim of Prop. 4.14.

*Remark 4.15*

In [132] a similar problem to ours was considered. While in that work the focus was on several affine sets, and a threshold value  $t_0$  was given for each of them by means of their semidefinite programs where observables  $\{\mathbf{A}^{tx}, \mathbf{A}^{ty}, \mathbf{A}^{tz}\}$  become compatible, we are considering *all* affine sets with dimension 2.

### 4.3.1 Proof of Proposition 4.14 : Part 1

In order to prove that  $M$  is nonempty, let us introduce some relevant notions:

$$\begin{aligned} D &:= \{\mathbf{v} \mid |\mathbf{v}| \leq 1, v_z = 0\} \subset B := \{\mathbf{v} \mid |\mathbf{v}| \leq 1\}, \\ \mathcal{S}_D &:= \{\varrho^{\mathbf{v}} \mid \mathbf{v} \in D\} \subset \mathcal{S} = \{\varrho^{\mathbf{v}} \mid \mathbf{v} \in B\}, \end{aligned}$$

where  $\mathbf{v} = v_x \mathbf{x} + v_y \mathbf{y} + v_z \mathbf{z} \in \mathbb{R}^3$ , and  $\varrho^{\mathbf{v}} := \frac{1}{2}(\mathbb{1} + \mathbf{v} \cdot \boldsymbol{\sigma})$ . Since  $\mathcal{S}_D$  is a convex set, we can treat  $\mathcal{S}_D$  almost like a quantum system. In the following, we will do it without giving precise definitions because they are obvious. For an observable  $\mathbf{E}$  on  $\mathcal{S}$  with effects  $\{\mathbf{E}(x)\}_x$ , we write its restriction to  $\mathcal{S}_D$  as  $\mathbf{E}|_D$  with effects  $\{\mathbf{E}(x)|_D\}_x$ , which is an observable on  $\mathcal{S}_D$ . It is easy to obtain the following Lemma.

#### Lemma 4.16

*The followings are equivalent:*

- (i)  $\mathbf{A}^{tx}$  and  $\mathbf{A}^{ty}$  are incompatible (thus  $\frac{1}{\sqrt{2}} < t \leq 1$ ).
- (ii)  $\mathbf{A}^{tx}$  and  $\mathbf{A}^{ty}$  are  $\mathcal{S}_D$ -incompatible.
- (iii)  $\mathbf{A}^{tx}|_D$  and  $\mathbf{A}^{ty}|_D$  are incompatible as observables on  $\mathcal{S}_D$ .

#### Proof

(i)  $\Rightarrow$  (iii). Suppose that  $\mathbf{A}^{tx}|_D$  and  $\mathbf{A}^{ty}|_D$  are compatible in  $\mathcal{S}_D$ . There exists an observable  $\mathbf{M}$  on  $\mathcal{S}_D$  whose marginals coincide with  $\mathbf{A}^{tx}|_D$  and  $\mathbf{A}^{ty}|_D$ . One can extend this  $\mathbf{M}$  to the whole  $\mathcal{S}$  so that it does not depend on  $z$  (for example, one can simply regard its effect  $c_0 \mathbb{1} + c_1 \sigma_1 + c_2 \sigma_2$  as an effect on  $\mathcal{S}$ ). Since both  $\mathbf{A}^{tx}|_D$  and  $\mathbf{A}^{ty}|_D$  also do not depend on  $z$ , the extension of  $\mathbf{M}$  gives a joint observable of  $\mathbf{A}^{tx}$  and  $\mathbf{A}^{ty}$ .

(iii)  $\Rightarrow$  (ii). Suppose that  $\mathbf{A}^{tx}$  and  $\mathbf{A}^{ty}$  are  $\mathcal{S}_D$ -compatible. There exists an observable  $\mathbf{M}$  on  $\mathcal{S}$  whose marginals coincide with  $\mathbf{A}^{tx}$  and  $\mathbf{A}^{ty}$  in  $\mathcal{S}_D$ . The restriction of  $\mathbf{M}$  on  $\mathcal{S}_D$  proves that (iii) is false.

(ii)  $\Rightarrow$  (i). Suppose that  $\mathbf{A}^{tx}$  and  $\mathbf{A}^{ty}$  are compatible, then they are  $\mathcal{S}_D$ -compatible.  $\square$

This lemma demonstrates that the incompatibility of  $\mathbf{A}^{tx}$  and  $\mathbf{A}^{ty}$  means the incompatibility of  $\mathbf{A}^{tx}|_D$  and  $\mathbf{A}^{ty}|_D$ . We can present further observations.

#### Lemma 4.17

*Let us consider two pure states  $\varrho^{\mathbf{r}_1}$  and  $\varrho^{\mathbf{r}_2}$  ( $\mathbf{r}_1, \mathbf{r}_2 \in \partial B$ ,  $\mathbf{r}_1 \neq \mathbf{r}_2$ ), and a convex subset  $\mathcal{S}_0$  of  $\mathcal{S}$  generated by them:  $\mathcal{S}_0 := \{p\varrho^{\mathbf{r}_1} + (1-p)\varrho^{\mathbf{r}_2} \mid 0 \leq p \leq 1\}$ . We also introduce an affine projection  $P$  by  $P\varrho^{\mathbf{v}} = \varrho_{\mathbf{P}\mathbf{v}}$ , where  $\varrho^{\mathbf{v}} \in \mathcal{S}$  with  $\mathbf{v} = v_x \mathbf{x} + v_y \mathbf{y} + v_z \mathbf{z}$  and  $\mathbf{P}\mathbf{v} = v_x \mathbf{x} + v_y \mathbf{y}$ , and extend it affinely. The affine hull of  $\mathcal{S}_0$  is projected to  $\mathcal{S}_D$  as*

$$P\mathcal{S}_0 := \{\lambda P\varrho^{\mathbf{r}_1} + (1-\lambda)P\varrho^{\mathbf{r}_2} \mid \lambda \in \mathbf{R}\} \cap \mathcal{S}_D. \quad (4.26)$$



If  $A^{tx}$  and  $A^{ty}$  are  $\mathcal{S}_0$ -incompatible, then their restrictions  $A^{tx}|_D$  and  $A^{ty}|_D$  are  $P\mathcal{S}_0$ -incompatible.

**Proof**

Suppose that  $A^{tx}$  and  $A^{ty}$  are  $\mathcal{S}_0$ -incompatible. It implies  $\Pr_1 \neq \Pr_2$ , i.e.,  $P\varrho^{r_1} \neq P\varrho^{r_2}$  (see Example 4.10), and thus  $P\mathcal{S}_0$  is a segment in  $\mathcal{S}_D$ . If  $A^{tx}|_D$  and  $A^{ty}|_D$  are  $P\mathcal{S}_0$ -compatible, then there exists a joint observable  $M$  on  $\mathcal{S}_D$  such that its marginals coincide with  $A^{tx}|_D$  and  $A^{ty}|_D$  on  $P\mathcal{S}_0 \subset \mathcal{S}_D$ . This  $M$  can be extended to an observable on  $\mathcal{S}$  so that the extension does not depend on  $z$ . Because

$$\begin{aligned} \text{Tr}[A^{tx}(\pm)P\varrho^{r_1}] &= \text{Tr}[A^{tx}(\pm)\varrho^{r_1}], \\ \text{Tr}[A^{tx}(\pm)P\varrho^{r_2}] &= \text{Tr}[A^{tx}(\pm)\varrho^{r_2}] \end{aligned}$$

(and their  $y$ -counterparts) hold due to the independence of  $A^{tx}(\pm)$  from  $\sigma_3$ , the marginals of  $M$  coincide with  $A^{tx}$  and  $A^{ty}$  on  $\mathcal{S}_0$ . It results in the  $\mathcal{S}_0$ -compatibility of  $A^{tx}$  and  $A^{ty}$ , which is a contradiction.  $\square$

It follows from this lemma that  $\chi_{incomp}(A^{tx}|_D, A^{ty}|_D)$  is two when  $\chi_{incomp}(A^{tx}, A^{ty})$  is two, equivalently  $\chi_{incomp}(A^{tx}, A^{ty})$  is three when  $\chi_{incomp}(A^{tx}|_D, A^{ty}|_D)$  is three (remember that  $\chi_{incomp}(A^{tx}, A^{ty}) \leq 3$ ). In fact, the converse also holds.

**Lemma 4.18**

$\chi_{incomp}(A^{tx}|_D, A^{ty}|_D)$  is three when  $\chi_{incomp}(A^{tx}, A^{ty})$  is three.

**Proof**

Let  $\chi_{incomp}(A^{tx}, A^{ty}) = 3$ . It follows that for any line  $S \subset \mathcal{S}$ ,  $A^{tx}$  and  $A^{ty}$  are  $S$ -compatible. In particular,  $A^{tx}$  and  $A^{ty}$  are  $S'$ -compatible for any line  $S'$  in  $\mathcal{S}_D$ , and thus there is an observable  $M$  such that its marginals coincide with  $A^{tx}$  and  $A^{ty}$  on  $S'$ . It is easy to see that the marginals of  $M|_D$  coincide with  $A^{tx}|_D$  and  $A^{ty}|_D$  on  $S'$ , which results in the  $S'$ -compatibility of  $A^{tx}|_D$  and  $A^{ty}|_D$ . Because  $S'$  is arbitrary, we can conclude  $\chi_{incomp}(A^{tx}|_D, A^{ty}|_D) = 3$ .  $\square$

The lemmas above manifest that if  $A^{tx}$  and  $A^{ty}$  are incompatible, then  $A^{tx}|_D$  and  $A^{ty}|_D$  are also incompatible and

$$\chi_{incomp}(A^{tx}, A^{ty}) = \chi_{incomp}(A^{tx}|_D, A^{ty}|_D).$$

Therefore, in the following, we denote  $A^{tx}|_D$  and  $A^{ty}|_D$  simply by  $A_D^{tx}$  and  $A_D^{ty}$  respectively, and focus on the quantity  $\chi_{incomp}(A_D^{tx}, A_D^{ty})$  instead of the original  $\chi_{incomp}(A^{tx}, A^{ty})$ .

Before proceeding to the next step, let us confirm our strategy in the following parts. In Part 2 (Subsection 4.3.2), we will consider a line (segment)  $\mathcal{S}_1$  in  $\mathcal{S}_D$ , and consider for  $0 < t < 1$  all pairs of observables  $(\tilde{A}_1^t, \tilde{A}_2^t)$

on  $\mathcal{S}_D$  that coincide with  $(\mathbf{A}_D^{tx}, \mathbf{A}_D^{ty})$  on  $\mathcal{S}_1$ . Then, in Part 3 (Subsection 4.3.3), we will investigate the (in)compatibility of those  $\tilde{\mathbf{A}}_1^t$  and  $\tilde{\mathbf{A}}_2^t$  in order to obtain  $\chi_{incomp}(\mathbf{A}_D^{tx}, \mathbf{A}_D^{ty})$ . It will be shown that when  $t$  is sufficiently small, there exists a compatible pair  $(\tilde{\mathbf{A}}_1^t, \tilde{\mathbf{A}}_2^t)$  for any  $\mathcal{S}_1$ , that is,  $\mathbf{A}_D^{tx}$  and  $\mathbf{A}_D^{ty}$  are  $\mathcal{S}_1$ -compatible for any line  $\mathcal{S}_1$ . It results in  $\chi_{incomp}(\mathbf{A}_D^{tx}, \mathbf{A}_D^{ty}) = 3$ , and thus  $M \neq \emptyset$ .

### 4.3.2 Proof of Proposition 4.14 : Part 2

Let us consider two pure states  $\varrho^{\mathbf{r}_1}$  and  $\varrho^{\mathbf{r}_2}$  with  $\mathbf{r}_1, \mathbf{r}_2 \in \partial D$  ( $\mathbf{r}_1 \neq \mathbf{r}_2$ ), and a convex set  $\mathcal{S}_1 := \{p\varrho^{\mathbf{r}_1} + (1-p)\varrho^{\mathbf{r}_2} \mid 0 \leq p \leq 1\}$ . We set parameters  $\varphi_1$  and  $\varphi_2$  as

$$\mathbf{r}_1 = \cos \varphi_1 \mathbf{x} + \sin \varphi_1 \mathbf{y}, \quad (4.27)$$

$$\mathbf{r}_2 = \cos \varphi_2 \mathbf{x} + \sin \varphi_2 \mathbf{y}, \quad (4.28)$$

where  $-\pi \leq \varphi_1 < \varphi_2 < \pi$ . By exchanging  $\pm$  properly, without loss of generality we can assume the line connecting  $\mathbf{r}_1$  and  $\mathbf{r}_2$  passes through above the origin (instead of below). In this case, from geometric consideration, we have

$$\begin{aligned} 0 &< \varphi_2 - \varphi_1 \leq \pi, \\ 0 &\leq \frac{\varphi_1 + \varphi_2}{2} \leq \frac{\pi}{2}. \end{aligned} \quad (4.29)$$

Note that when  $\varphi_2 - \varphi_1 = \pi$ ,  $\varrho^{\mathbf{r}_1}$  and  $\varrho^{\mathbf{r}_2}$  are perfectly distinguishable, which results in the  $\mathcal{S}_1$ -compatibility of  $\mathbf{A}_D^{tx}$  and  $\mathbf{A}_D^{ty}$  (see Example 4.4). On the other hand, when  $\frac{\varphi_1 + \varphi_2}{2} = 0$  or  $\frac{\pi}{2}$ ,  $\text{Tr}[\varrho \mathbf{A}_D^{tx}(+)]$  or  $\text{Tr}[\varrho \mathbf{A}_D^{ty}(+)]$  is constant for  $\varrho \in \mathcal{S}_1$  respectively, so  $\mathbf{A}_D^{tx}$  and  $\mathbf{A}_D^{ty}$  are  $\mathcal{S}_1$ -compatible (see Example 4.10). Thus, instead of (4.29), we hereafter assume

$$\begin{aligned} 0 &< \frac{\varphi_2 - \varphi_1}{2} < \frac{\pi}{2}, \\ 0 &< \frac{\varphi_1 + \varphi_2}{2} < \frac{\pi}{2}. \end{aligned} \quad (4.30)$$

Next, we consider a binary observable  $\tilde{\mathbf{A}}_1^t$  on  $\mathcal{S}_D$  that coincides with  $\mathbf{A}_D^{tx}$  on  $\mathcal{S}_1 \subset \mathcal{S}_D$ . There are many possible  $\tilde{\mathbf{A}}_1^t$ , and each  $\tilde{\mathbf{A}}_1^t$  is determined completely by its effect  $\tilde{\mathbf{A}}_1^t(+)$  corresponding to the outcome ‘+’ because it is binary. The effect  $\tilde{\mathbf{A}}_1^t(+)$  is associated with a vector  $\mathbf{v}_1 \in D$  defined as

$$\mathbf{v}_1 := \text{argmax}_{\mathbf{v} \in D} \text{tr}[\varrho_{\mathbf{v}} \tilde{\mathbf{A}}_1^t(+)]. \quad (4.31)$$

Let us introduce a parameter  $\xi_1 \in [-\pi, \pi)$  by

$$\mathbf{v}_1 = \cos \xi_1 \mathbf{x} + \sin \xi_1 \mathbf{y}, \quad (4.32)$$

and express  $\tilde{\mathbf{A}}_1^t(+)$  as

$$\tilde{\mathbf{A}}_1^t(+)=\frac{1}{2}((1+w(\xi_1))\mathbb{1}+\mathbf{m}_1(\xi_1)\cdot\sigma),\quad(4.33)$$

where we set

$$\mathbf{m}_1(\xi_1)=C_1(\xi_1)\mathbf{v}_1\quad\text{with}\quad 0\leq C_1(\xi_1)\leq 1.\quad(4.34)$$

Because

$$\begin{aligned}\mathrm{Tr}[\varrho^{\mathbf{r}_1}\mathbf{A}_D^{tx}(+)]&=\mathrm{Tr}[\varrho^{\mathbf{r}_1}\tilde{\mathbf{A}}_1^t(+)],\\ \mathrm{Tr}[\varrho^{\mathbf{r}_2}\mathbf{A}_D^{ty}(+)]&=\mathrm{Tr}[\varrho^{\mathbf{r}_2}\tilde{\mathbf{A}}_1^t(+)],\end{aligned}$$

namely

$$\begin{aligned}\frac{1}{2}+\frac{t}{2}\cos\varphi_1&=\frac{1+w_1(\xi_1)}{2}+\frac{C_1(\xi_1)}{2}\cos(\varphi_1-\xi_1),\\ \frac{1}{2}+\frac{t}{2}\cos\varphi_2&=\frac{1+w_1(\xi_1)}{2}+\frac{C_1(\xi_1)}{2}\cos(\varphi_2-\xi_1),\end{aligned}\quad(4.35)$$

hold, we can obtain

$$C_1(\xi_1)=\frac{t(\cos\varphi_1-\cos\varphi_2)}{\cos(\varphi_1-\xi_1)-\cos(\varphi_2-\xi_1)}=\frac{t\sin\varphi_0}{\sin(\varphi_0-\xi_1)},\quad(4.36)$$

$$w_1(\xi_1)=-t\left(\frac{\sin(\varphi_1-\varphi_2)}{2\sin(\frac{\varphi_1-\varphi_2}{2})}\right)\cdot\left(\frac{\sin\xi_1}{\sin(\varphi_0-\xi_1)}\right)=\frac{-t\cos\psi_0\sin\xi_1}{\sin(\varphi_0-\xi_1)},\quad(4.37)$$

where we set  $\varphi_0:=\frac{\varphi_1+\varphi_2}{2}$  and  $\psi_0:=\frac{\varphi_2-\varphi_1}{2}$  ( $0<\varphi_0<\frac{\pi}{2}$ ,  $0<\psi_0<\frac{\pi}{2}$ ). Note that if  $\sin(\varphi_0-\xi_1)=0$  or  $\cos(\varphi_1-\xi_1)-\cos(\varphi_2-\xi_1)=0$  holds, then  $\cos\varphi_1-\cos\varphi_2=0$  holds (see (4.36)). It means  $\varphi_0=0$ , which is a contradiction, and thus  $\sin(\varphi_0-\xi_1)\neq 0$  (that is,  $C_1(\xi_1)$  and  $w_1(\xi_1)$  in (4.36), (4.37) are well-defined). Moreover, because  $C_1(\xi_1)\geq 0$ , we can see from (4.36) that  $\sin(\varphi_0-\xi_1)>0$  holds, which results in

$$0\leq\xi_1<\varphi_0,\quad(4.38)$$

or

$$-\pi+\varphi_0<\xi_1\leq 0.\quad(4.39)$$

In addition,  $\xi_1$  is restricted also by the condition that  $\tilde{\mathbf{A}}_1^t(\pm)$  are positive. Since the eigenvalues of  $\tilde{\mathbf{A}}_1^t(\pm)$  are  $\frac{1}{2}((1+w_1(\xi_1))\pm C_1(\xi_1))$ , the restriction comes from both

$$\begin{aligned}1+w_1(\xi_1)+C_1(\xi_1)&\leq 2,\\ 1+w_1(\xi_1)-C_1(\xi_1)&\geq 0,\end{aligned}\quad(4.40)$$

equivalently

$$1 - w_1(\xi_1) \geq C_1(\xi_1), \quad (4.41)$$

$$1 + w_1(\xi_1) \geq C_1(\xi_1). \quad (4.42)$$

When (4.39) (i.e.  $\sin \xi_1 \leq 0$ ) holds,  $w_1(\xi_1) \geq 0$  holds, and thus (4.41) is sufficient. It is written explicitly as

$$\sin(\varphi_0 - \xi_1) + t \sin \xi_1 \cos \psi_0 \geq t \sin \varphi_0,$$

or

$$\frac{1}{t} \cos \xi_1 + \frac{1}{t \sin \varphi_0} (t \cos \psi_0 - \cos \varphi_0) \sin \xi_1 \geq 1. \quad (4.43)$$

In order to investigate (4.43), we adopt a geometric method here while it can be solved in an analytic way. Let us define

$$h_1(t, \varphi_0, \psi_0) = \frac{1}{t \sin \varphi_0} (t \cos \psi_0 - \cos \varphi_0). \quad (4.44)$$

Then, we can rewrite (4.43) as

$$(\cos \xi_1, \sin \xi_1) \cdot \left[ \left( \frac{1}{t}, h_1 \right) - (\cos \xi_1, \sin \xi_1) \right] \geq 0. \quad (4.45)$$

In fact, it can be verified easily that  $(\frac{1}{t}, h_1)$  is the intersection of the line  $l_1 := \{\lambda \mathbf{r}_1 + (1 - \lambda) \mathbf{r}_2 \mid \lambda \in \mathbb{R}\}$  and the line  $x = \frac{1}{t}$  in  $\mathbb{R}^2$ . Considering this fact, we can find that  $\xi_1$  satisfies (4.45) if and only if

$$\xi_1^{\min}(t, \varphi_0, \psi_0) \leq \xi_1 \leq 0, \quad (4.46)$$

where  $\xi_1^{\min}(t, \varphi_0, \psi_0)$  is determined by the condition

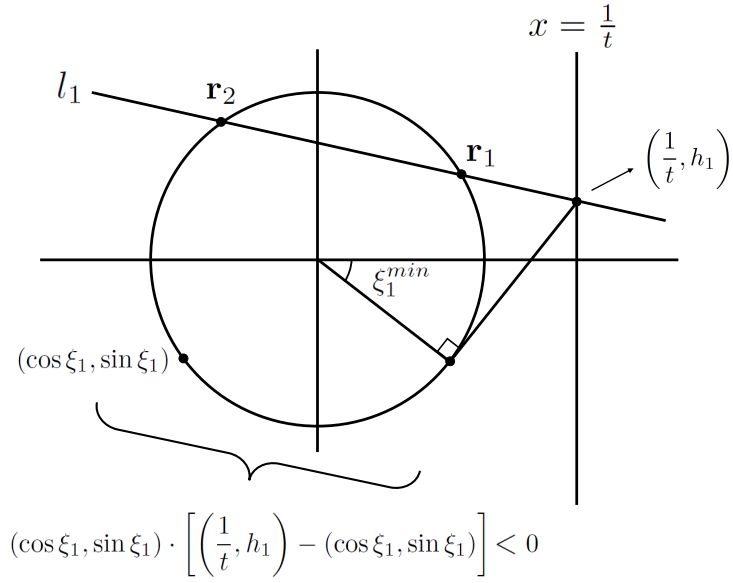
$$\left[ \left( \frac{1}{t}, h_1 \right) - (\cos \xi_1^{\min}, \sin \xi_1^{\min}) \right] \perp (\cos \xi_1^{\min}, \sin \xi_1^{\min}) \quad (4.47)$$

(see FIG. 4.1). Analytically, it corresponds to the case when the equality of (4.43) holds:

$$\frac{1}{t} \cos \xi_1^{\min} + \frac{1}{t \sin \varphi_0} (t \cos \psi_0 - \cos \varphi_0) \sin \xi_1^{\min} = 1, \quad (4.48)$$

or

$$1 - w_1(\xi_1^{\min}) = C_1(\xi_1^{\min}).$$



**Figure 4.1:** Geometric description of determining  $\xi_1^{min}$ .

It can be represented explicitly as

$$\begin{aligned} (t^2 \cos^2 \psi_0 - 2t \cos \varphi_0 \cos \psi_0 + 1) \sin^2 \xi_1^{min} \\ - 2t \sin \varphi_0 (t \cos \psi_0 - \cos \varphi_0) \sin \xi_1^{min} + (t^2 - 1) \sin^2 \varphi_0 = 0, \end{aligned} \quad (4.49)$$

and  $\sin \xi_1^{min}$  is obtained as its negative solution. Note that since the coefficient  $(t^2 \cos^2 \psi_0 - 2t \cos \varphi_0 \cos \psi_0 + 1)$  is strictly positive, the solutions do not show any singular behavior. In summary, we have obtained

$$\xi_1^{min}(t, \varphi_0, \psi_0) \leq \xi_1 \leq 0 \quad (4.50)$$

with  $\xi_1^{min}(t, \varphi_0, \psi_0)$  uniquely determined for  $t$ ,  $\varphi_0$ , and  $\psi_0$  by

$$\begin{cases} -\pi + \varphi_0 < \xi_1^{min}(t, \varphi_0, \psi_0) \leq 0, \\ 1 - w_1(\xi_1^{min}(t, \varphi_0, \psi_0)) = C_1(\xi_1^{min}(t, \varphi_0, \psi_0)). \end{cases} \quad (4.51)$$

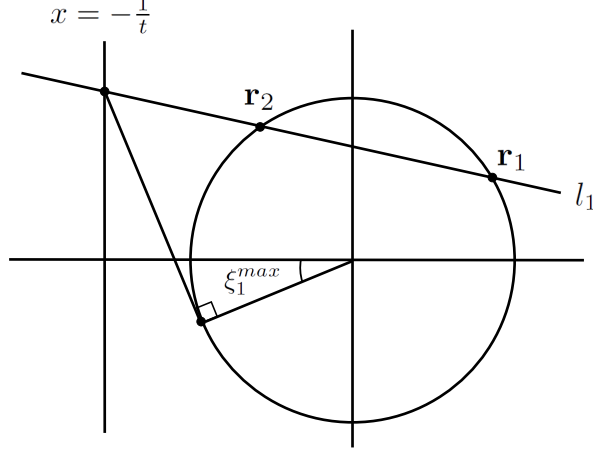
On the other hand, when (4.38) (i.e.  $\sin \xi_1 \geq 0$ ) holds, (4.42) is sufficient. It results in a tight condition for  $\xi_1$ :

$$0 \leq \xi_1 \leq \xi_1^{max}(t, \varphi_0, \psi_0), \quad (4.52)$$

where  $\xi_1^{max}(t, \varphi_0, \psi_0)$  is a constant uniquely determined for  $\varphi_0$  and  $\psi_0$  by

$$\begin{cases} 0 \leq \xi_1^{max}(t, \varphi_0, \psi_0) < \varphi_0 \\ 1 + w_1(\xi_1^{max}(t, \varphi_0, \psi_0)) = C_1(\xi_1^{max}(t, \varphi_0, \psi_0)). \end{cases} \quad (4.53)$$

We remark that this can be obtained by a similar geometric method to the previous case: consider the intersection of the line  $l_1$  and the line  $x = -\frac{1}{t}$  in turn (see FIG. 4.2). Overall, we have demonstrated that  $\xi_1$  for  $\tilde{\mathbf{A}}_1^t$  satisfies



**Figure 4.2:** Geometric description of determining  $\xi_1^{max}$ .

$$\xi_1^{min}(t, \varphi_0, \psi_0) \leq \xi_1 \leq \xi_1^{max}(t, \varphi_0, \psi_0), \quad (4.54)$$

where  $\xi_1^{min}(t, \varphi_0, \psi_0)$  and  $\xi_1^{max}(t, \varphi_0, \psi_0)$  are obtained thorough (4.51) and (4.53) respectively. Note that  $\xi_1^{min}(t, \varphi_0, \psi_0)$  and  $\xi_1^{max}(t, \varphi_0, \psi_0)$  depend continuously on  $t$  (and  $\varphi_1, \varphi_2$  through  $\varphi_0$  and  $\psi_0$ ).

Similarly, we consider a binary observable  $\tilde{\mathbf{A}}_2^t$  on  $\mathcal{S}_D$  which coincides with  $\mathbf{A}_D^{ty}$  in  $\mathcal{S}_1$ , and focus on its effect  $\tilde{\mathbf{A}}_2^t(+)$ . We define parameters  $\mathbf{v}_2 \in D$  and  $\xi_2 \in [-\pi, \pi)$  as

$$\mathbf{v}_2 = \sin \xi_2 \mathbf{x} + \cos \xi_2 \mathbf{y} = \operatorname{argmax}_{\mathbf{v} \in D} \operatorname{tr}[\tilde{\mathbf{A}}_2^t(+) \varrho_{\mathbf{v}}]. \quad (4.55)$$

$\tilde{\mathbf{A}}_2^t(+)$  is represented as

$$\tilde{\mathbf{A}}_2^t(+)= \frac{1}{2} ((1 + w_2(\xi_2)) \mathbb{1} + \mathbf{m}_2(\xi_2)) \quad (4.56)$$

with

$$\mathbf{m}_2(\xi_2) = C_2(\xi_2) \mathbf{v}_2 \quad (0 \leq C_2(\xi_2) \leq 1).$$

(4.35) becomes

$$\begin{aligned} \frac{1}{2} + \frac{t}{2} \cos\left(\frac{\pi}{2} - \varphi_1\right) &= \frac{1 + x_2(\xi_2)}{2} + \frac{C_2(\xi_2)}{2} \cos\left(\frac{\pi}{2} - \varphi_1 - \xi_1\right), \\ \frac{1}{2} + \frac{t}{2} \cos\left(\frac{\pi}{2} - \varphi_2\right) &= \frac{1 + x_2(\xi_2)}{2} + \frac{C_2(\xi_2)}{2} \cos\left(\frac{\pi}{2} - \varphi_2 - \xi_1\right), \end{aligned} \quad (4.57)$$

so defining  $\overline{\varphi}_1 := \frac{\pi}{2} - \varphi_1$  and  $\overline{\varphi}_2 := \frac{\pi}{2} - \varphi_2$ , we can obtain similarly to (4.36) and (4.37)

$$C_2(\xi_2) = \frac{t \sin \overline{\varphi}_0}{\sin(\overline{\varphi}_0 - \xi_2)}, \quad (4.58)$$

$$w_2(\xi_2) = \frac{-t \cos \psi_0 \sin \xi_2}{\sin(\overline{\varphi}_0 - \xi_2)}, \quad (4.59)$$

where  $\overline{\varphi}_0 := \frac{\overline{\varphi}_1 + \overline{\varphi}_2}{2} = \frac{\pi}{2} - \varphi_0$ . It follows that properties of  $\tilde{\mathbf{A}}_2^t$  can be obtained just by replacing  $\xi_1$  and  $\varphi_0$  exhibited in the argument for  $\tilde{\mathbf{A}}_1^t$  by  $\xi_2$  and  $\overline{\varphi}_0$  respectively. Remark that  $0 < \overline{\varphi}_0 < \frac{\pi}{2}$  holds similarly to  $\varphi_0$ , and that the change  $\psi_0 \rightarrow \overline{\psi}_0 := \frac{\overline{\varphi}_2 - \overline{\varphi}_1}{2} = -\psi_0$  does not affect the equations above, so we dismiss it. From (4.58) and (4.59), we have

$$\xi_2^{\min}(t, \varphi_0, \psi_0) \leq \xi_2 \leq \xi_2^{\max}(t, \varphi_0, \psi_0), \quad (4.60)$$

where

$$\xi_2^{\min}(t, \varphi_0, \psi_0) = \xi_1^{\min}(t, \overline{\varphi}_0, \psi_0) = \xi_1^{\min}\left(t, \frac{\pi}{2} - \varphi_0, \psi_0\right), \quad (4.61)$$

and

$$\xi_2^{\max}(t, \varphi_0, \psi_0) = \xi_1^{\max}(t, \overline{\varphi}_0, \psi_0) = \xi_1^{\max}\left(t, \frac{\pi}{2} - \varphi_0, \psi_0\right), \quad (4.62)$$

which satisfy

$$\begin{cases} -\frac{\pi}{2} + \varphi_0 < \xi_2^{\min}(t, \varphi_0, \psi_0) \leq 0 \\ 1 - w_2(\xi_2^{\min}(t, \varphi_0, \psi_0)) = C_2(\xi_2^{\min}(t, \varphi_0, \psi_0)) \end{cases} \quad (4.63)$$

and

$$\begin{cases} 0 \leq \xi_2^{\max}(t, \varphi_0, \psi_0) < \frac{\pi}{2} - \varphi_0 \\ 1 + w_2(\xi_2^{\max}(t, \varphi_0, \psi_0)) = C_2(\xi_2^{\max}(t, \varphi_0, \psi_0)) \end{cases} \quad (4.64)$$

respectively.

### 4.3.3 Proof of Proposition 4.14 : Part 3

In this part, we shall consider the (in)compatibility of the observables  $\tilde{\mathbf{A}}_1^t$  and  $\tilde{\mathbf{A}}_2^t$  defined in (a) for  $t$  close to  $\frac{1}{\sqrt{2}}$  ( $t \sim \frac{1}{\sqrt{2}}$ ). It is related directly with the  $\mathcal{S}_1$ -(in)compatibility of  $\mathbf{A}_D^{tx}$  and  $\mathbf{A}_D^{ty}$  as we have shown in the beginning of this section. Let us examine the behavior of  $\xi_1^{\min}(t, \varphi_0, \psi_0)$  for  $t \sim \frac{1}{\sqrt{2}}$ . We denote  $\xi_1^{\min}(t = \frac{1}{\sqrt{2}}, \varphi_0, \psi_0)$  and  $h_1(t = \frac{1}{\sqrt{2}}, \varphi_0, \psi_0)$  simply by  $\hat{\xi}_1^{\min}(\varphi_0, \psi_0)$

and  $\hat{h}_1(\varphi_0, \psi_0)$  respectively. The following lemma is useful.

**Lemma 4.19**

With  $\varphi_0$  fixed,  $\hat{\xi}_1^{min}$  is a strictly decreasing function of  $\psi_0$ .

**Proof**

The claim can be observed to hold by a geometric consideration in terms of FIG. 4.1. In fact, increasing  $\psi_0$  with  $\varphi_0$  fixed corresponds to moving the line  $l_1$  down with its inclination fixed. The movement makes  $h_1$  (or  $\hat{h}_1$ ) and hence  $\xi_1^{min}$  (or  $\hat{\xi}_1^{min}$ ) smaller, which proves the claim. Here we show an analytic proof of this fact. We can see from (4.44) and (4.48) that

$$\sqrt{2} \cos \hat{\xi}_1^{min} + \hat{h}_1 \sin \hat{\xi}_1^{min} = 1, \quad (4.65)$$

i.e.

$$\hat{h}_1 = \frac{1}{\sin \hat{\xi}_1^{min}} \left( 1 - \sqrt{2} \cos \hat{\xi}_1^{min} \right)$$

holds (note that  $\sin \hat{\xi}_1^{min} \neq 0$  because  $\sin \hat{\xi}_1^{min} = 0$  contradicts (4.65)). Then, the claim follows from the observation that

$$\frac{d\hat{h}_1}{d\hat{\xi}_1^{min}} = \frac{1}{(\sin \hat{\xi}_1^{min})^2} \left( \sqrt{2} - \cos \hat{\xi}_1^{min} \right) > 0,$$

and  $\hat{h}_1 = \frac{1}{\sin \varphi_0} (\cos \psi_0 - \sqrt{2} \cos \varphi_0)$  is a decreasing function of  $\psi_0$ . □

From this lemma, it follows that

$$\hat{\xi}_1^{min}(\varphi_0, \psi_0) < \lim_{\psi_0 \rightarrow +0} \hat{\xi}_1^{min}(\varphi_0, \psi_0) =: \Xi_1^{min}(\varphi_0), \quad (4.66)$$

and

$$\hat{\xi}_2^{min}(\varphi_0, \psi_0) < \Xi_2^{min}(\varphi_0) \quad (4.67)$$

hold for all  $\varphi_0 \in (0, \frac{\pi}{2})$  and  $\psi_0 \in (0, \frac{\pi}{2})$ , where

$$\begin{aligned} \hat{\xi}_2^{min}(\varphi_0, \psi_0) &:= \xi_2^{min} \left( t = \frac{1}{\sqrt{2}}, \varphi_0, \psi_0 \right) \left( = \hat{\xi}_1^{min} \left( \frac{\pi}{2} - \varphi_0, \psi_0 \right) \right), \\ \Xi_2^{min}(\varphi_0) &:= \lim_{\psi_0 \rightarrow +0} \hat{\xi}_2^{min}(\varphi_0, \psi_0) \left( = \Xi_1^{min} \left( \frac{\pi}{2} - \varphi_0 \right) \right). \end{aligned} \quad (4.68)$$

We can prove the following lemma.

**Lemma 4.20**

$$\Xi_1^{min}(\varphi_0) + \Xi_2^{min}(\varphi_0) \leq -\frac{\pi}{2}$$

holds for all  $0 < \varphi_0 < \frac{\pi}{2}$ .



**Proof**

Let us define

$$\begin{aligned} H_1(\varphi_0) &:= \lim_{\psi_0 \rightarrow +0} \widehat{h}_1(\varphi_0, \psi_0) = \lim_{\psi_0 \rightarrow +0} h_1 \left( t = \frac{1}{\sqrt{2}}, \varphi_0, \psi_0 \right) \\ &= \frac{1}{\sin \varphi_0} \left( 1 - \sqrt{2} \cos \varphi_0 \right). \end{aligned}$$

It holds similarly to (4.65) that

$$\sqrt{2} \cos \Xi_1^{min} + H_1 \sin \Xi_1^{min} = 1. \quad (4.69)$$

Hence, together with  $\sin^2 \Xi_1^{min} + \cos^2 \Xi_1^{min} = 1$ , we can obtain

$$\cos \Xi_1^{min} = \frac{1}{\sqrt{2}} \cdot \frac{2 + H_1 \sqrt{2H_1^2 + 2}}{H_1^2 + 2}, \quad (4.70)$$

or its more explicit form

$$\cos \Xi_1^{min} = \frac{1}{\sqrt{2}} \cdot \frac{4 - 3\sqrt{2} \cos \varphi_0}{3 - 2\sqrt{2} \cos \varphi_0}. \quad (4.71)$$

It results in

$$\Xi_1^{min}(\varphi_0) = -\arccos \left( \frac{1}{\sqrt{2}} \cdot \frac{4 - 3\sqrt{2} \cos \varphi_0}{3 - 2\sqrt{2} \cos \varphi_0} \right), \quad (4.72)$$

where we follow the convention that  $\arccos: [-1, 1] \rightarrow [0, \pi]$ , and thus  $\Xi_1^{min} \in (-\pi + \varphi_0, 0]$  is obtained through  $-\arccos: [-1, 1] \rightarrow [-\pi, 0]$ . Because

$$\frac{d}{d\varphi_0} \left( \frac{1}{\sqrt{2}} \cdot \frac{4 - 3\sqrt{2} \cos \varphi_0}{3 - 2\sqrt{2} \cos \varphi_0} \right) = \frac{\sin \varphi_0}{(3 - 2\sqrt{2} \cos \varphi_0)^2},$$

and

$$\sqrt{1 - \left( \frac{1}{\sqrt{2}} \cdot \frac{4 - 3\sqrt{2} \cos \varphi_0}{3 - 2\sqrt{2} \cos \varphi_0} \right)^2} = \sqrt{\left( \frac{\sin \varphi_0}{3 - 2\sqrt{2} \cos \varphi_0} \right)^2} = \frac{\sin \varphi_0}{3 - 2\sqrt{2} \cos \varphi_0},$$

we can observe that

$$\frac{d\Xi_1^{min}}{d\varphi_0} = \left( \frac{\sin \varphi_0}{3 - 2\sqrt{2} \cos \varphi_0} \right)^{-1} \cdot \frac{\sin \varphi_0}{(3 - 2\sqrt{2} \cos \varphi_0)^2} = \frac{1}{3 - 2\sqrt{2} \cos \varphi_0},$$

and

$$\frac{d^2 \Xi_1^{min}}{d\varphi_0^2} = \frac{-2\sqrt{2} \sin \varphi_0}{(3 - 2\sqrt{2} \cos \varphi_0)^2} < 0, \quad (4.73)$$

which means  $\Xi_1^{min}$  is concave. Therefore, for any  $\varphi_0 \in (0, \frac{\pi}{2})$ , the concavity results in

$$\begin{aligned} \frac{1}{2}\Xi_1^{min}(\varphi_0) + \frac{1}{2}\Xi_2^{min}(\varphi_0) &= \frac{1}{2}\Xi_1^{min}(\varphi_0) + \frac{1}{2}\Xi_1^{min}\left(\frac{\pi}{2} - \varphi_0\right) \\ &\leq \Xi_1^{min}\left(\frac{1}{2}\varphi_0 + \frac{1}{2}\left(\frac{\pi}{2} - \varphi_0\right)\right) \\ &= \Xi_1^{min}\left(\frac{\pi}{4}\right). \end{aligned}$$

Since we can see from (4.72) that  $\Xi_1^{min}\left(\frac{\pi}{4}\right) = -\frac{\pi}{4}$ ,

$$\Xi_1^{min}(\varphi_0) + \Xi_2^{min}(\varphi_0) \leq -\frac{\pi}{2}$$

holds for any  $\varphi_0 \in (0, \frac{\pi}{2})$ . □

According to Lemma 4.19 and Lemma 4.20,

$$\widehat{\xi}_1^{min}(\varphi_0, \psi_0) + \widehat{\xi}_2^{min}(\varphi_0, \psi_0) < \Xi_1^{min}(\varphi_0) + \Xi_2^{min}(\varphi_0) \leq -\frac{\pi}{2},$$

that is,

$$\xi_1^{min}\left(t = \frac{1}{\sqrt{2}}, \varphi_0, \psi_0\right) + \xi_2^{min}\left(t = \frac{1}{\sqrt{2}}, \varphi_0, \psi_0\right) < -\frac{\pi}{2}$$

holds for any  $\varphi_0$  and  $\psi_0$  (i.e. for any  $\varphi_1$  and  $\varphi_2$ ). However, we cannot conclude that

$$\xi_1^{min}(t, \varphi_0, \psi_0) + \xi_2^{min}(t, \varphi_0, \psi_0) \leq -\frac{\pi}{2} \quad (4.74)$$

holds for  $t \sim \frac{1}{\sqrt{2}}$ : it may fail when

$$\sup_{\varphi_0, \psi_0} \left[ \xi_1^{min}\left(t = \frac{1}{\sqrt{2}}, \varphi_0, \psi_0\right) + \xi_2^{min}\left(t = \frac{1}{\sqrt{2}}, \varphi_0, \psi_0\right) \right] = -\frac{\pi}{2}.$$

On the other hand, because we can observe similarly to Lemma 4.19 that  $\xi_1^{min}$  is a strictly decreasing function of  $\psi_0$ , it is anticipated that (4.74) holds for  $t \sim \frac{1}{\sqrt{2}}$  and for  $\psi_0$  sufficiently close to  $\frac{\pi}{2}$ . In fact, for  $\psi_0 \in [\frac{\pi}{4}, \frac{\pi}{2})$ , we can prove the following proposition.

**Proposition 4.21**

*There exists a constant  $C < -\frac{\pi}{2}$  such that*

$$\widehat{\xi}_1^{min}(\varphi_0, \psi_0) + \widehat{\xi}_2^{min}(\varphi_0, \psi_0) < C,$$

*i.e.*

$$\xi_1^{\min} \left( t = \frac{1}{\sqrt{2}}, \varphi_0, \psi_0 \right) + \xi_2^{\min} \left( t = \frac{1}{\sqrt{2}}, \varphi_0, \psi_0 \right) < C,$$

holds for all  $\psi_0 \in [\frac{\pi}{4}, \frac{\pi}{2})$  and  $\varphi_0 \in (0, \frac{\pi}{2})$ .

**Proof**

Because

$$\widehat{\xi}_1^{\min}(\varphi_0, \psi_0) + \widehat{\xi}_2^{\min}(\varphi_0, \psi_0) = \widehat{\xi}_1^{\min}(\varphi_0, \psi_0) + \widehat{\xi}_1^{\min} \left( \frac{\pi}{2} - \varphi_0, \psi_0 \right),$$

we can assume without loss of generality that  $0 < \varphi_0 \leq \frac{\pi}{4}$ . Due to Lemma 4.19, it holds for any  $\psi_0 \in [\frac{\pi}{4}, \frac{\pi}{2})$  that

$$\begin{aligned} \widehat{\xi}_1^{\min}(\varphi_0, \psi_0) &\leq \widehat{\xi}_1^{\min} \left( \varphi_0, \psi_0 = \frac{\pi}{4} \right), \\ \widehat{\xi}_1^{\min} \left( \frac{\pi}{2} - \varphi_0, \psi_0 \right) &\leq \widehat{\xi}_1^{\min} \left( \frac{\pi}{2} - \varphi_0, \psi_0 = \frac{\pi}{4} \right). \end{aligned} \quad (4.75)$$

Let us denote  $\widehat{\xi}_1^{\min}(\varphi_0, \psi_0 = \frac{\pi}{4})$  simply by  $\widetilde{\Xi}_1^{\min}(\varphi_0)$ . In order to investigate  $\widetilde{\Xi}_1^{\min}(\varphi_0)$  and  $\widetilde{\Xi}_1^{\min}(\frac{\pi}{2} - \varphi_0)$ , we have to recall (4.65). Similarly to (4.69) and (4.70) in the proof of Lemma 4.20, it results in

$$\cos \widetilde{\Xi}_1^{\min} = \frac{1}{\sqrt{2}} \cdot \frac{2 + \widetilde{H}_1 \sqrt{2\widetilde{H}_1^2 + 2}}{\widetilde{H}_1^2 + 2}, \quad (4.76)$$

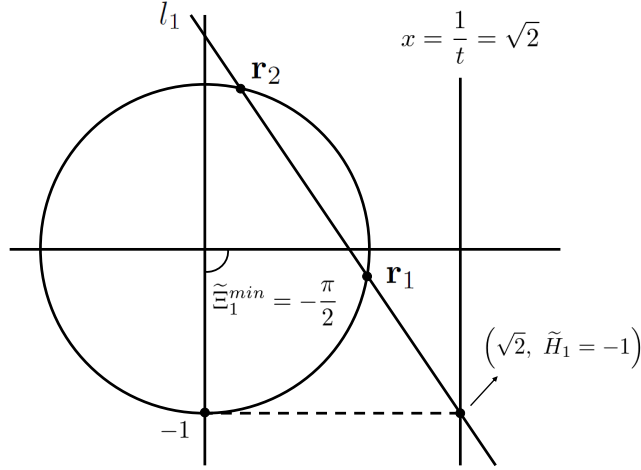
where

$$\widetilde{H}_1(\varphi_0) = \widehat{h}_1 \left( \varphi_0, \psi_0 = \frac{\pi}{4} \right) = \frac{1}{\sin \varphi_0} \left( \frac{1}{\sqrt{2}} - \sqrt{2} \cos \varphi_0 \right). \quad (4.77)$$

Note that in this case we cannot apply a similar method to the one in Lemma 4.20 because  $\widetilde{\Xi}_1^{\min}$  does not have a clear form like (4.72). Alternatively, we focus on the following monotone relations between  $\widetilde{\Xi}_1^{\min}$ ,  $\widetilde{H}_1$ , and  $\varphi_0$  (referring to the proof of Lemma 4.19 may be helpful):

$$\frac{d\widetilde{\Xi}_1^{\min}}{d\widetilde{H}_1} > 0, \quad \frac{d\widetilde{H}_1}{d\varphi_0} > 0 \quad \left( \text{thus } \frac{d\widetilde{\Xi}_1^{\min}}{d\varphi_0} > 0 \right). \quad (4.78)$$

From these relations, it can be seen that our restriction  $0 < \varphi_0 \leq \frac{\pi}{4}$  is equivalent to the condition  $\widetilde{H}_1 \leq 1 - \sqrt{2}$  since  $\widetilde{H}_1(0) = -\infty$  and  $\widetilde{H}_1(\frac{\pi}{4}) = 1 - \sqrt{2}$ . The claim of the proposition can be shown easily when  $\widetilde{H}_1 \leq -1$  (or  $0 < \varphi_0 \leq \varphi^* := \arccos \frac{2+\sqrt{10}}{6}$ , where  $\widetilde{H}_1(\varphi^*) = -1$ ). In fact,



**Figure 4.3:** Geometric description of  $\tilde{\Xi}_1^{min}$ . It can be observed that  $\tilde{\Xi}_1^{min} = -\frac{\pi}{2}$  when  $\tilde{H}_1 = -1$ .

$$\tilde{\Xi}_1^{min}(\varphi_0) \leq \tilde{\Xi}_1^{min}(\varphi^*) = -\frac{\pi}{2}$$

and

$$\tilde{\Xi}_1^{min}\left(\frac{\pi}{2} - \varphi_0\right) < \tilde{\Xi}_1^{min}\left(\frac{\pi}{2}\right) = -\arccos\frac{2\sqrt{2} + \sqrt{3}}{5}$$

hold (see FIG. 4.3 and (4.76)), and thus we can conclude

$$\tilde{\Xi}_1^{min}(\varphi_0) + \tilde{\Xi}_1^{min}\left(\frac{\pi}{2} - \varphi_0\right) < C_1,$$

where

$$C_1 = -\frac{\pi}{2} - \arccos\frac{2\sqrt{2} + \sqrt{3}}{5} \left(< -\frac{\pi}{2}\right).$$

When  $-1 < \tilde{H}_1 \leq 1 - \sqrt{2}$  (or  $\varphi^* < \varphi_0 \leq \frac{\pi}{4}$ ), we need a bit complicated evaluations. It holds similarly to the previous calculations that

$$\begin{aligned} \tilde{\Xi}_1^{min}(\varphi_0) &\leq \tilde{\Xi}_1^{min}\left(\frac{\pi}{4}\right), \\ \tilde{\Xi}_1^{min}\left(\frac{\pi}{2} - \varphi_0\right) &< \tilde{\Xi}_1^{min}\left(\frac{\pi}{2} - \varphi^*\right). \end{aligned}$$

Since

$$\tilde{\Xi}_1^{min} = -\frac{\pi}{4} \iff \tilde{H}_1 = 0 \iff \varphi_0 = \frac{\pi}{3},$$

$\tilde{\Xi}_1^{min}\left(\frac{\pi}{4}\right) < -\frac{\pi}{4} = \tilde{\Xi}_1^{min}\left(\frac{\pi}{3}\right)$  holds due to the monotone relations (4.78). On the other hand, we have

$$\cos\varphi^* - \cos\frac{\pi}{6} = \frac{2 + \sqrt{10}}{6} - \frac{\sqrt{3}}{2} = -0.0056... < 0,$$

that is,

$$\varphi^* > \frac{\pi}{6}.$$

It follows that  $\frac{\pi}{2} - \varphi^* < \frac{\pi}{3}$ , and thus  $\tilde{\Xi}_1^{min}(\frac{\pi}{2} - \varphi_0) < -\frac{\pi}{4}$ . Therefore, we can conclude also in this case

$$\tilde{\Xi}_1^{min}(\varphi_0) + \tilde{\Xi}_1^{min}\left(\frac{\pi}{2} - \varphi_0\right) < C_2,$$

where

$$C_2 = \tilde{\Xi}_1^{min}\left(\frac{\pi}{4}\right) + \tilde{\Xi}_1^{min}\left(\frac{\pi}{2} - \varphi^*\right) \left(< -\frac{\pi}{2}\right).$$

Overall, we have obtained

$$\tilde{\Xi}_1^{min}(\varphi_0, \psi_0) + \tilde{\Xi}_2^{min}(\varphi_0, \psi_0) < \max\{C_1, C_2\} \left(< -\frac{\pi}{2}\right)$$

for all  $\varphi_0 \in (0, \frac{\pi}{2})$  and  $\psi_0 \in [\frac{\pi}{4}, \frac{\pi}{2})$ .  $\square$

By virtue of this proposition, for  $t$  sufficiently close to  $\frac{1}{\sqrt{2}}$ ,

$$\xi_1^{min}(t, \varphi_0, \psi_0) + \xi_2^{min}(t, \varphi_0, \psi_0) \leq -\frac{\pi}{2}$$

follows from the continuity of  $\xi_1^{min}$  and  $\xi_2^{min}$  with respect to  $t$  when  $\frac{\pi}{4} \leq \psi_0 < \frac{\pi}{2}$ . It means that there always exist  $\xi_1^* \geq \xi_1^{min}$  and  $\xi_2^* \geq \xi_2^{min}$  for such  $t$  and for any  $\varphi_1$  and  $\varphi_2$  satisfying  $\xi_1^* + \xi_2^* = -\frac{\pi}{2}$ . For these  $\xi_1^*$  and  $\xi_2^*$ ,  $\mathbf{v}_1 = -\mathbf{v}_2$  holds, and thus  $\tilde{\mathbf{A}}_1^t$  and  $\tilde{\mathbf{A}}_2^t$  are compatible, i.e.  $\mathbf{A}_D^{tx}$  and  $\mathbf{A}_D^{ty}$  are  $\mathcal{S}_1$ -compatible.

On the other hand, when  $0 < \psi_0 < \frac{\pi}{4}$ , it may not hold for  $t \sim \frac{1}{\sqrt{2}}$  that  $\xi_1^{min}(t, \varphi_0, \psi_0) + \xi_2^{min}(t, \varphi_0, \psi_0) \leq -\frac{\pi}{2}$ , and thus we cannot apply the same argument. Nevertheless, we can demonstrate that there exist  $\xi_1$  and  $\xi_2$  such that  $\tilde{\mathbf{A}}_1^t$  and  $\tilde{\mathbf{A}}_2^t$  are compatible even when  $0 < \psi_0 < \frac{\pi}{4}$ . To see this, let us assume  $0 < \psi_0 < \frac{\pi}{4}$  and apply the necessary and sufficient condition for (in)compatibility. According to the result proven in [146, 147, 148],  $\tilde{\mathbf{A}}_1^t$  and  $\tilde{\mathbf{A}}_2^t$  with (4.33) and (4.56) respectively are compatible if and only if

$$(1 - F_1^2 - F_2^2) \left(1 - \frac{w_1^2}{F_1^2} - \frac{w_2^2}{F_2^2}\right) \leq (\mathbf{m}_1 \cdot \mathbf{m}_2 - w_1 w_2)^2 \quad (4.79)$$

holds, where

$$F_1 := \frac{1}{2} \left( \sqrt{(1 + w_1)^2 - C_1^2} + \sqrt{(1 - w_1)^2 - C_1^2} \right), \quad (4.80)$$

$$F_2 := \frac{1}{2} \left( \sqrt{(1 + w_2)^2 - C_2^2} + \sqrt{(1 - w_2)^2 - C_2^2} \right). \quad (4.81)$$

For  $\xi_1^{min}$  and  $\xi_2^{min}$ , since it holds that

$$1 - w_1(\xi_1^{min}) = C_1(\xi_1^{min}), \quad (4.82)$$

$$1 - w_2(\xi_2^{min}) = C_2(\xi_2^{min}), \quad (4.83)$$

they become

$$F_1 = \sqrt{w_1(\xi_1^{min})}, \quad F_2 = \sqrt{w_2(\xi_2^{min})}. \quad (4.84)$$

Therefore, (4.79) can be rewritten as

$$\begin{aligned} & [(1 - \sin(\xi_1^{min} + \xi_2^{min}))w_1(\xi_1^{min})w_2(\xi_2^{min}) \\ & \quad - (1 + \sin(\xi_1^{min} + \xi_2^{min}))(1 - w_1(\xi_1^{min}) - w_2(\xi_2^{min}))] \\ & \quad \cdot [(1 - w_1(\xi_1^{min}))(1 - w_2(\xi_2^{min}))(1 - \sin(\xi_1^{min} + \xi_2^{min}))] \geq 0. \end{aligned} \quad (4.85)$$

If  $1 - \sin(\xi_1^{min} + \xi_2^{min}) = 0$ , then (4.85) holds, that is,  $\tilde{A}_1^t$  and  $\tilde{A}_2^t$  for  $\xi_1^{min}$  and  $\xi_2^{min}$  respectively are compatible. Therefore, we hereafter assume  $1 - \sin(\xi_1^{min} + \xi_2^{min}) > 0$ , and rewrite (4.85) as (note that  $0 < w_1(\xi_1^{min}) < 1$ ,  $0 < w_2(\xi_2^{min}) < 1$ )

$$\begin{aligned} & (1 + \sin(\xi_1^{min} + \xi_2^{min}))(1 - w_1(\xi_1^{min}) - w_2(\xi_2^{min})) \\ & \leq (1 - \sin(\xi_1^{min} + \xi_2^{min}))w_1(\xi_1^{min})w_2(\xi_2^{min}). \end{aligned} \quad (4.86)$$

In other words,  $\tilde{A}_1^t$  and  $\tilde{A}_2^t$  with respect to  $\xi_1^{min}$  and  $\xi_2^{min}$  are incompatible if and only if

$$\begin{aligned} & (1 + \sin(\xi_1^{min} + \xi_2^{min}))(1 - w_1(\xi_1^{min}) - w_2(\xi_2^{min})) \\ & > (1 - \sin(\xi_1^{min} + \xi_2^{min}))w_1(\xi_1^{min})w_2(\xi_2^{min}) \end{aligned} \quad (4.87)$$

holds. In order to investigate whether (4.87) holds, it is helpful to introduce a function  $Z$  defined as

$$\begin{aligned} Z(t, \varphi_0, \psi_0) := & [1 + \sin(\xi_1^{min}(t, \varphi_0, \psi_0) + \xi_2^{min}(t, \varphi_0, \psi_0))] \\ & [1 + w_1(\xi_1^{min}(t, \varphi_0, \psi_0)) + w_2(\xi_2^{min}(t, \varphi_0, \psi_0))] \\ & - [1 - \sin(\xi_1^{min}(t, \varphi_0, \psi_0) + \xi_2^{min}(t, \varphi_0, \psi_0))] \\ & w_1(\xi_1^{min}(t, \varphi_0, \psi_0))w_2(\xi_2^{min}(t, \varphi_0, \psi_0)). \end{aligned} \quad (4.88)$$

Because

$$\begin{aligned} & (1 + \sin(\xi_1^{min} + \xi_2^{min}))(1 - w_1(\xi_1^{min}) - w_2(\xi_2^{min})) \\ & < (1 + \sin(\xi_1^{min} + \xi_2^{min}))(1 + w_1(\xi_1^{min}) + w_2(\xi_2^{min})), \end{aligned}$$

$$Z(t, \varphi_0, \psi_0) > 0 \quad (4.89)$$

holds if  $\tilde{\mathbf{A}}_1^t$  and  $\tilde{\mathbf{A}}_2^t$  with respect to  $\xi_1^{min}$  and  $\xi_2^{min}$  are incompatible. Let us focus on the case when  $t = \frac{1}{\sqrt{2}}$  (i.e.  $\xi_1^{min} = \hat{\xi}_1^{min}$ ). If a pair  $(\varphi_0, \psi_0)$  satisfies  $\hat{\xi}_1^{min}(\varphi_0, \psi_0) \leq -\frac{\pi}{2}$  or  $\hat{\xi}_2^{min}(\varphi_0, \psi_0) \leq -\frac{\pi}{2}$ , then

$$\hat{\xi}_1^{min}(\varphi_0, \psi_0) + \hat{\xi}_2^{min}(\varphi_0, \psi_0) < C$$

with

$$C = -\frac{\pi}{2} + \lim_{\substack{\varphi_0 \rightarrow \frac{\pi}{2}-0 \\ \psi_0 \rightarrow +0}} \hat{\xi}_1^{min}(\varphi_0, \psi_0) = -\frac{\pi}{2} - \arccos\left(\frac{2\sqrt{2}}{3}\right) < -\frac{\pi}{2}$$

holds due to similar monotone relations to (4.78) between  $\varphi_0, \psi_0$ , and  $\hat{\xi}_1^{min}$  (remember that  $\hat{\xi}_2^{min}(\varphi_0, \psi_0) = \hat{\xi}_1^{min}(\frac{\pi}{2} - \varphi_0, \psi_0)$ ). Therefore, in this case, we can apply the same argument as Proposition 4.21, which results in the compatibility of  $\tilde{\mathbf{A}}_1^t$  and  $\tilde{\mathbf{A}}_2^t$  for  $t \sim \frac{1}{\sqrt{2}}$ . On the other hand, let us examine the case when  $(\varphi_0, \psi_0)$  satisfies  $\psi_0 \in (0, \frac{\pi}{4})$ , and  $\hat{\xi}_1^{min}(\varphi_0, \psi_0) > -\frac{\pi}{2}$  and  $\hat{\xi}_2^{min}(\varphi_0, \psi_0) > -\frac{\pi}{2}$ . Because  $\psi_0 \in (0, \frac{\pi}{4})$ , we obtain for general  $t$  (see (4.37))

$$w_1(\xi_1^{min}) > -\frac{t}{\sqrt{2}} \frac{\sin \xi_1^{min}}{\sin(\varphi_0 - \xi_1^{min})} \geq \frac{t}{\sqrt{2}} (-\sin \xi_1^{min}). \quad (4.90)$$

For  $t = \frac{1}{\sqrt{2}}$ , since

$$-\frac{\pi}{2} < \hat{\xi}_1^{min}(\varphi_0, \psi_0) < \lim_{\substack{\varphi_0 \rightarrow \frac{\pi}{2}-0 \\ \psi_0 \rightarrow +0}} \hat{\xi}_1^{min}(\varphi_0, \psi_0),$$

it gives a bound

$$w_1(\xi_1^{min}) > \frac{1}{2} \sin \hat{\xi}_0, \quad (4.91)$$

where we define

$$\hat{\xi}_0 = -\lim_{\substack{\varphi_0 \rightarrow \frac{\pi}{2}-0 \\ \psi_0 \rightarrow +0}} \hat{\xi}_1^{min}(\varphi_0, \psi_0) = \arccos\left(\frac{2\sqrt{2}}{3}\right).$$

Let  $\varepsilon$  be a positive constant satisfying  $\varepsilon < \frac{1}{16}(\sin \hat{\xi}_0)^2$ . Due to the continuity of sine, there exists a positive constant  $\delta$  such that  $\sin x \in (-1, -1 + \varepsilon)$  whenever  $x \in (-\frac{\pi}{2} - \delta, -\frac{\pi}{2})$ . If  $(\varphi_0, \psi_0)$  satisfies  $\hat{\xi}_1^{min}(\varphi_0, \psi_0) + \hat{\xi}_2^{min}(\varphi_0, \psi_0) \leq -\frac{\pi}{2} - \delta$ , then it again leads to the same argument as Proposition 4.21, and we can see that  $\tilde{\mathbf{A}}_1^t$  and  $\tilde{\mathbf{A}}_2^t$  for this  $(\varphi_0, \psi_0)$  are compatible. Conversely, if

$(\varphi_0, \psi_0)$  satisfies  $-\frac{\pi}{2} - \delta < \widehat{\xi}_1^{min}(\varphi_0, \psi_0) + \widehat{\xi}_2^{min}(\varphi_0, \psi_0) < -\frac{\pi}{2}$  (remember Lemma 4.20), then

$$-1 < \sin(\widehat{\xi}_1^{min}(\varphi_0, \psi_0) + \widehat{\xi}_2^{min}(\varphi_0, \psi_0)) < -1 + \varepsilon$$

follows from the definition of  $\delta$ . Therefore, by virtue of (4.88), we have

$$\begin{aligned} Z\left(t = \frac{1}{\sqrt{2}}, \varphi_0, \psi_0\right) &= \left[1 + \sin(\widehat{\xi}_1^{min}(\varphi_0, \psi_0) + \widehat{\xi}_2^{min}(\varphi_0, \psi_0))\right] \\ &\quad \left[1 + w_1(\widehat{\xi}_1^{min}(\varphi_0, \psi_0)) + w_2(\widehat{\xi}_2^{min}(\varphi_0, \psi_0))\right] \\ &\quad - \left[1 - \sin(\widehat{\xi}_1^{min}(\varphi_0, \psi_0) + \widehat{\xi}_2^{min}(\varphi_0, \psi_0))\right] \\ &\quad w_1(\widehat{\xi}_1^{min}(\varphi_0, \psi_0))w_2(\widehat{\xi}_2^{min}(\varphi_0, \psi_0)) \\ &< \varepsilon \left[1 + w_1(\widehat{\xi}_1^{min}(\varphi_0, \psi_0)) + w_2(\widehat{\xi}_2^{min}(\varphi_0, \psi_0))\right] \\ &\quad - (2 - \varepsilon)w_1(\widehat{\xi}_1^{min}(\varphi_0, \psi_0))w_2(\widehat{\xi}_2^{min}(\varphi_0, \psi_0)) \\ &= \varepsilon \left[1 + w_1(\widehat{\xi}_1^{min}(\varphi_0, \psi_0))\right] \left[1 + w_2(\widehat{\xi}_1^{min}(\varphi_0, \psi_0))\right] \\ &\quad - 2w_1(\widehat{\xi}_1^{min}(\varphi_0, \psi_0))w_2(\widehat{\xi}_2^{min}(\varphi_0, \psi_0)) \\ &< 4\varepsilon - 2w_1(\widehat{\xi}_1^{min}(\varphi_0, \psi_0))w_2(\widehat{\xi}_2^{min}(\varphi_0, \psi_0)). \end{aligned}$$

Because

$$\begin{aligned} 4\varepsilon - 2w_1(\widehat{\xi}_1^{min}(\varphi_0, \psi_0))w_2(\widehat{\xi}_2^{min}(\varphi_0, \psi_0)) &< \frac{1}{4}(\sin \widehat{\xi}_0)^2 - \frac{1}{2}(\sin \widehat{\xi}_0)^2 \\ &= -\frac{1}{4}(\sin \widehat{\xi}_0)^2, \end{aligned}$$

it holds that

$$Z\left(t = \frac{1}{\sqrt{2}}, \varphi_0, \psi_0\right) < -\frac{1}{4}(\sin \widehat{\xi}_0)^2 < 0.$$

Therefore, for  $t \sim \frac{1}{\sqrt{2}}$ ,  $Z(t, \varphi_0, \psi_0) \leq 0$  holds, that is,  $\widetilde{\mathbf{A}}_1^t$  and  $\widetilde{\mathbf{A}}_2^t$  with respect to  $\xi_1^{min}$  and  $\xi_2^{min}$  are compatible. Overall, we have demonstrated that when  $t \sim \frac{1}{\sqrt{2}}$ , there exist compatible observables  $\widetilde{\mathbf{A}}_1^t$  and  $\widetilde{\mathbf{A}}_2^t$  for any line  $\mathcal{S}_1 \subset \mathcal{S}_D$  such that they agree with  $\mathbf{A}_D^{tx}$  and  $\mathbf{A}_D^{ty}$  on  $\mathcal{S}_1$  respectively. That is, when  $t \sim \frac{1}{\sqrt{2}}$ ,  $\mathbf{A}_D^{tx}$  and  $\mathbf{A}_D^{ty}$  are  $\mathcal{S}_1$ -compatible for any line  $\mathcal{S}_1 \subset \mathcal{S}_D$ . Therefore, we can conclude that  $\chi_{incomp}(\mathbf{A}_D^{tx}, \mathbf{A}_D^{ty}) = 3$  for  $t \sim \frac{1}{\sqrt{2}}$ , and thus the set  $M$  in (4.24) is nonempty.

#### 4.3.4 Proof of Proposition 4.14 : Part 4

In this part, we shall show that

$$t'_0 := \inf L = \sup M \in M,$$



where  $L$  and  $M$  are defined in (4.24). In order to prove this, we will see that if  $t \in L$ , then  $t - \delta \in L$  for sufficiently small  $\delta > 0$ , that is,  $t'_0 \notin L$ .

Let us focus again on a system described by a two-dimensional disk state space  $\mathcal{S}_D$ . It is useful to identify this system with the system of a quantum bit with real coefficients by replacing  $\{\sigma_1, \sigma_2\}$  with  $\{\sigma_3, \sigma_1\}$ . Then, defining  $\mathcal{E}_D$  as the set of all effects on  $\mathcal{S}_D$ , we can see that any  $E \in \mathcal{E}_D$  can be expressed as a real-coefficient positive matrix smaller than  $\mathbb{1}$ . We also define  $O_D(2) \subset \mathcal{E}_D \times \mathcal{E}_D$  as the set of all binary observables on  $\mathcal{S}_D$ , which is isomorphic naturally to  $\mathcal{E}_D$  since a binary observable  $\mathbf{A}$  is completely specified by its effect  $\mathbf{A}(+) \in \mathcal{E}_D$ . With introducing a topology (e.g. norm topology) on  $\mathcal{E}_D$ , it also can be observed that  $O_D(2)$  is homeomorphic to  $\mathcal{E}_D$ . Note that because the system is described by finite-dimensional matrices, any (natural) topology (norm topology, weak topology, etc.) coincides with each other. For a pair of states  $\{\varrho^{\mathbf{r}^1}, \varrho^{\mathbf{r}^2}\}$  in  $\mathcal{S}_D$ , and a binary observable  $\mathbf{A} \in O_D(2)$ , we define a set of observables  $C(\mathbf{A} : \varrho^{\mathbf{r}^1}, \varrho^{\mathbf{r}^2})$  as the set of all binary observables  $\tilde{\mathbf{A}} \in O_D(2)$  such that

$$\begin{aligned}\mathrm{Tr}[\varrho^{\mathbf{r}^1} \tilde{\mathbf{A}}(\pm)] &= \mathrm{Tr}[\varrho^{\mathbf{r}^1} \mathbf{A}(\pm)], \\ \mathrm{Tr}[\varrho^{\mathbf{r}^2} \tilde{\mathbf{A}}(\pm)] &= \mathrm{Tr}[\varrho^{\mathbf{r}^2} \mathbf{A}(\pm)].\end{aligned}$$

It can be confirmed easily that  $C(\mathbf{A} : \varrho^{\mathbf{r}^1}, \varrho^{\mathbf{r}^2})$  is closed in  $O_D(2) \simeq \mathcal{E}_D$ . Let us denote by  $O_D(4)$  the set of all observables with four outcomes, which is a compact (i.e. bounded and closed) subset of  $\mathcal{E}_D^4$ . For each  $\mathbf{M} = \{\mathbf{M}(x, y)\} \in O_D(4)$ , we can introduce a pair of binary observables by

$$\pi_1(\mathbf{M}) = \left\{ \sum_y \mathbf{M}(x, y) \right\}_x, \quad \pi_2(\mathbf{M}) = \left\{ \sum_x \mathbf{M}(x, y) \right\}_y.$$

Since  $\pi_j : O_D(4) \rightarrow O_D(2)$  is continuous, the set of all compatible binary observables denoted by

$$JM(2, 2) := \{(\pi_1(\mathbf{M}), \pi_2(\mathbf{M})) \mid \mathbf{M} \in O_D(4)\}$$

is compact in  $O_D(2) \times O_D(2) \simeq \mathcal{E}_D \times \mathcal{E}_D$  as well. As we have seen in the previous part,  $\chi_{incomp}(\mathbf{A}^{t\mathbf{x}}, \mathbf{A}^{t\mathbf{y}}) = 2$  (i.e.  $\chi_{incomp}(\mathbf{A}_D^{t\mathbf{x}}, \mathbf{A}_D^{t\mathbf{y}}) = 2$ ) if and only if there exists a pair of vectors  $\mathbf{r}_1, \mathbf{r}_2 \in \partial D$  such that

$$(C(\mathbf{A}_D^{t\mathbf{x}} : \varrho^{\mathbf{r}^1}, \varrho^{\mathbf{r}^2}) \times C(\mathbf{A}_D^{t\mathbf{y}} : \varrho^{\mathbf{r}^1}, \varrho^{\mathbf{r}^2})) \cap JM(2 : 2) = \emptyset.$$

Let us examine concrete representations of the sets. Each effect  $E \in \mathcal{E}_D$  is written as  $E = \frac{1}{2}(e_0\mathbb{1} + \mathbf{e} \cdot \boldsymbol{\sigma}) = \frac{1}{2}(e_0\mathbb{1} + e_1\sigma_1 + e_2\sigma_2)$  with  $(e_0, \mathbf{e}) = (e_0, e_1, e_2) \in \mathbb{R}^3$  satisfying  $0 \leq e_0 \pm |\mathbf{e}| \leq 2$ .

If we consider another effect  $F = \frac{1}{2}(f_0\mathbb{1} + \mathbf{f} \cdot \sigma)$ , the operator norm of  $E - F$  is calculated as

$$\|E - F\| = \frac{1}{2}(|e_0 - f_0| + |\mathbf{e} - \mathbf{f}|). \quad (4.92)$$

We may employ this norm to define a topology on  $\mathcal{E}_D$  and  $O_D(2) \simeq \mathcal{E}_D$ . On the other hand, each state in  $\mathcal{S}_D$  is parameterized as  $\varrho^{\mathbf{r}_1} = \frac{1}{2}(\mathbb{1} + x_1\sigma_1 + y_1\sigma_2)$ , where  $\mathbf{r}_1 = (x_1, y_1)$  satisfies  $|\mathbf{r}_1| \leq 1$ . For an effect  $E$  and a state  $\varrho^{\mathbf{r}_1}$ , we have  $\text{tr}[\varrho^{\mathbf{r}_1} E] = \frac{1}{2}(e_0 + \mathbf{r}_1 \cdot \mathbf{e})$ . In particular, when considering  $\mathbf{A}^{tx}(\pm) = \frac{1}{2}(\mathbb{1} \pm t\sigma_1)$ , a binary observable  $\mathbf{C}$  determined by the effect  $\mathbf{C}(+) = \frac{1}{2}(c_0\mathbb{1} + \mathbf{c} \cdot \sigma) = \frac{1}{2}(c_0\mathbb{1} + c_1\sigma_1 + c_2\sigma_2)$  satisfies  $\mathbf{C} \in C(\mathbf{A}^{tx} : \varrho^{\mathbf{r}_1}, \varrho^{\mathbf{r}_2})$  if and only if

$$\begin{aligned} \text{Tr}[\varrho^{\mathbf{r}_1} \mathbf{A}^{tx}(+)] &= \text{Tr}[\varrho^{\mathbf{r}_1} \mathbf{C}(+)], \\ \text{Tr}[\varrho^{\mathbf{r}_2} \mathbf{A}^{tx}(+)] &= \text{Tr}[\varrho^{\mathbf{r}_2} \mathbf{C}(+)], \end{aligned}$$

i.e.

$$\begin{aligned} 1 + tx_1 &= c_0 + \mathbf{r}_1 \cdot \mathbf{c} = c_0 + x_1c_1 + y_1c_2, \\ 1 + tx_2 &= c_0 + \mathbf{r}_2 \cdot \mathbf{c} = c_0 + x_2c_1 + y_2c_2. \end{aligned}$$

hold, where we set  $\mathbf{r}_2 = (x_2, y_2)$ . The set of their solutions for  $(c_0, \mathbf{c})$  is represented as

$$(c_0, \mathbf{c}) = (1, t, 0) + \lambda' \left( -\frac{x_1y_2 - y_1x_2}{x_1 - x_2}, -\frac{y_1 - y_2}{x_1 - x_2}, 1 \right)$$

with  $\lambda' \in \mathbb{R}$ . Let us define a vector  $\mathbf{n} \in \mathbb{R}^2$  such that

$$(1, \mathbf{n}) \cdot (1, \mathbf{r}_1) = (1, \mathbf{n}) \cdot (1, \mathbf{r}_2) = 0$$

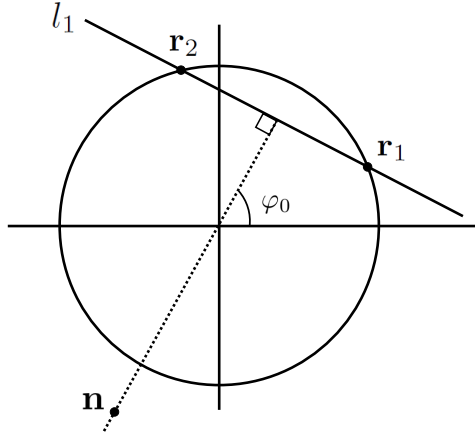
(i.e.  $\mathbf{n} \cdot \mathbf{r}_1 = \mathbf{n} \cdot \mathbf{r}_2 = -1$ ). It is easy to see that

$$\left( -\frac{x_1y_2 - y_1x_2}{x_1 - x_2}, -\frac{y_1 - y_2}{x_1 - x_2}, 1 \right) \propto (1, \mathbf{n}),$$

and thus the set of solutions can be rewritten as

$$(c_0, \mathbf{c}) = (1, t, 0) + \lambda(1, \mathbf{n}) \quad (4.93)$$

with  $\lambda \in \mathbb{R}$ . Note that because we are interested in the case when  $\mathbf{A}_D^{tx}$  and  $\mathbf{A}_D^{ty}$  are  $\mathcal{S}_1$ -incompatible, we do not consider the case when  $\mathbf{r}_1$  and  $\mathbf{r}_2$  are parallel or when  $x_1 = x_2$  corresponding to  $\psi_0 = \frac{\pi}{2}$  or  $\varphi_0 = 0$  in Part 1 respectively. Therefore, the vector  $\mathbf{n} = (n_x, n_y)$  can be defined successfully, and it is easy to verify that  $|\mathbf{n}| = \sqrt{n_x^2 + n_y^2} > 1$ . Moreover, because  $\varphi_0$  is



**Figure 4.4:** Geometric description of  $\mathbf{n}$ : we can observe that it lies in the third quadrant.

supposed to be  $0 < \varphi_0 < \frac{\pi}{2}$  as shown in Part 1, we can assume without loss of generality that its components  $n_x$  and  $n_y$  are negative (see FIG. 4.4). In order for  $\mathbf{C}$  to be an element of  $C(\mathbf{A}^{tx} : \varrho^{r_1}, \varrho^{r_2})$ , (4.93) should also satisfy

$$0 \leq 1 + \lambda \pm |(t, 0) + \lambda \mathbf{n}| \leq 2,$$

i.e.

$$1 + \lambda - |(t, 0) + \lambda \mathbf{n}| \geq 0, \quad 1 + \lambda + |(t, 0) + \lambda \mathbf{n}| \leq 2.$$

It can be reduced to

$$\lambda_1^t \leq \lambda \leq \lambda_2^t \tag{4.94}$$

with

$$\lambda_1^t = \frac{1 - n_x t - \sqrt{(1 - n_x t)^2 + (|\mathbf{n}|^2 - 1)(1 - t^2)}}{|\mathbf{n}|^2 - 1},$$

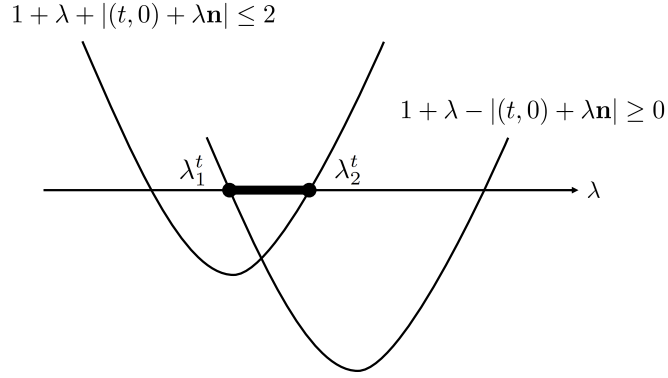
$$\lambda_2^t = \min \left\{ 1, \frac{-1 - n_x t + \sqrt{(1 + n_x t)^2 + (|\mathbf{n}|^2 - 1)(1 - t^2)}}{|\mathbf{n}|^2 - 1} \right\}, \tag{4.95}$$

where we used  $|\mathbf{n}| > 1$  and  $n_x < 0$  (see FIG. 4.5). Overall,  $C(\mathbf{A}_D^{tx} : \varrho^{r_1}, \varrho^{r_2})$  is isomorphic to the set parameterized as

$$\{(1, t, 0) + \lambda(1, \mathbf{n}) \mid \lambda_1^t \leq \lambda \leq \lambda_2^t\}, \tag{4.96}$$

where  $\lambda_1^t$  and  $\lambda_2^t$  are shown in (4.95). Remark that the same argument can be applied for  $C(\mathbf{A}^{ty} : \varrho^{r_1}, \varrho^{r_2})$ .

We shall now prove  $t'_0 = \inf L \notin L$ . Suppose that  $t \in L$ , i.e.  $\chi_{incomp}(\mathbf{A}_D^{tx}, \mathbf{A}_D^{ty}) =$



**Figure 4.5:** Solutions for  $\lambda$ .

2. It follows that there exist  $\mathbf{r}_1$  and  $\mathbf{r}_2$  in  $\partial D$  such that

$$(C(\mathbf{A}_D^{tx} : \varrho^{\mathbf{r}_1}, \varrho^{\mathbf{r}_2}) \times C(\mathbf{A}_D^{ty} : \varrho^{\mathbf{r}_1}, \varrho^{\mathbf{r}_2})) \cap JM(2 : 2) = \emptyset.$$

Denoting  $C(\mathbf{A}_D^{tx} : \varrho^{\mathbf{r}_1}, \varrho^{\mathbf{r}_2})$  and  $C(\mathbf{A}_D^{ty} : \varrho^{\mathbf{r}_1}, \varrho^{\mathbf{r}_2})$  simply by  $X^t$  and  $Y^t$  respectively, we can rewrite it as

$$X^t \times Y^t \cap JM(2 : 2) = \emptyset.$$

We need the following lemma.

**Lemma 4.22**

Let  $\delta > 0$ . There exists  $\Delta > 0$  such that for all  $\tau \in [0, \Delta]$  and for all  $C \in X^{t-\tau}$ , there exists  $\mathbf{A} \in X^t$  satisfying

$$d(C, \mathbf{A}) := \|C(+) - \mathbf{A}(+)\| < \delta$$

where  $d$  is a metric on  $O_D(2)$  defined through the operator norm  $\|\cdot\|$  on  $\mathcal{E}_D \simeq O_D(2)$ .

**Proof**

By its definition,  $X^t$  is a convex set of  $O_D(2)$ , and thus for all  $\mathbf{E} \in O_D(2)$  we can define successfully the distance between  $\mathbf{E}$  and  $X^t$ :

$$d(\mathbf{E}, X^t) = \min_{\mathbf{F} \in X^t} d(\mathbf{E}, \mathbf{F}).$$

In particular, for  $\mathbf{E}' \in X^{t-\Delta'} \subset O_D(2)$  with  $\Delta' > 0$  and  $\mathbf{E}'(+) = \frac{1}{2}(e'_0 \mathbb{1} + \mathbf{e}' \cdot \sigma)$ , it becomes

$$d(\mathbf{E}', X^t) = \min_{\mathbf{F} \in X^t} d(\mathbf{E}', \mathbf{F}) = \min_{\mathbf{F} \in X^t} \frac{1}{2} (|e'_0 - f_0| + |\mathbf{e}' - \mathbf{f}|), \quad (4.97)$$

where  $\mathbf{F}(+) = \frac{1}{2}(f_0 \mathbb{1} + \mathbf{f} \cdot \sigma)$  (see (4.92)). Since, in terms of (4.96),  $\mathbf{E}' \in X^{t-\Delta'}$

and  $\mathbf{F} \in X^t$  imply

$$(e'_0, \mathbf{e}') = (1, t - \Delta', 0) + \lambda'(1, \mathbf{n})$$

with  $\lambda_1^{t-\Delta'} \leq \lambda' \leq \lambda_2^{t-\Delta'}$  and

$$(f_0, \mathbf{f}) = (1, t, 0) + \lambda(1, \mathbf{n})$$

with  $\lambda_1^t \leq \lambda \leq \lambda_2^t$  respectively, (4.97) can be rewritten as

$$2d(\mathbf{E}', X^t) = \min_{\lambda \in [\lambda_1^t, \lambda_2^t]} (|\lambda' - \lambda| + |(-\Delta', 0) + (\lambda' - \lambda)\mathbf{n}|).$$

It follows that

$$2d(\mathbf{E}', X^t) \leq \Delta' + \min_{\lambda \in [\lambda_1^t, \lambda_2^t]} |\lambda' - \lambda|(1 + |\mathbf{n}|). \quad (4.98)$$

Let us evaluate its right hand side. It is easy to see that

$$\min_{\lambda \in [\lambda_1^t, \lambda_2^t]} |\lambda' - \lambda| = \begin{cases} \lambda_1^t - \lambda' & (\lambda' < \lambda_1^t) \\ 0 & (\lambda_1^t \leq \lambda' \leq \lambda_2^t) \\ \lambda' - \lambda_2^t & (\lambda' > \lambda_2^t) \end{cases}.$$

Suppose that  $\lambda' < \lambda_1^t$  holds, for example. In this case, because  $\lambda_1^{t-\Delta'} \leq \lambda'$ , we can obtain

$$\lambda_1^t - \lambda' \leq \lambda_1^t - \lambda_1^{t-\Delta'}.$$

In a similar way, it can be demonstrated that

$$\sup_{\lambda' \in [\lambda_1^{t-\Delta'}, \lambda_2^{t-\Delta'}]} \min_{\lambda \in [\lambda_1^t, \lambda_2^t]} |\lambda' - \lambda| = \max \left\{ \lambda_1^t - \lambda_1^{t-\Delta'}, 0, \lambda_2^{t-\Delta'} - \lambda_2^t \right\}.$$

By virtue of (4.95), the right hand side converges to 0 as  $\Delta' \rightarrow 0$ , and thus we can see from (4.98) that

$$\sup_{\mathbf{E}' \in X^{t-\Delta'}} d(\mathbf{E}', X^t) \xrightarrow{\Delta' \rightarrow 0} 0$$

It results in that there exists  $\Delta > 0$  such that for all  $\tau \in [0, \Delta]$ ,

$$\sup_{\mathbf{E}' \in X^{t-\tau}} d(\mathbf{E}', X^t) < \delta$$

holds, that is,  $d(\mathbf{C}, X^t) < \delta$  holds for any  $\mathbf{C} \in X^{t-\tau}$ . Moreover, because  $X^t$  is convex, there exists  $\mathbf{A} \in X^t$  satisfying  $d(\mathbf{C}, X^t) = d(\mathbf{C}, \mathbf{A})$ , which proves the claim of the lemma.  $\square$

Note that a similar statement also holds for  $Y^t$ : there exists  $\tilde{\Delta} > 0$  such that for all  $\tilde{\tau} \in [0, \tilde{\Delta}]$  and for all  $\mathbf{D} \in Y^{t-\tilde{\tau}}$ , there exists  $\mathbf{B} \in Y^t$  satisfying  $d(\mathbf{D}, \mathbf{B}) < \delta$ . Let  $V := O_D(2) \times O_D(2) (\simeq \mathcal{E}_D \times \mathcal{E}_D)$  and let  $d_V$  be a product metric on  $V$  defined as

$$d_V((\mathbf{A}, \mathbf{B}), (\mathbf{C}, \mathbf{D})) = \max\{d(\mathbf{A}, \mathbf{C}), d(\mathbf{B}, \mathbf{D})\}.$$

According to Lemma 4.22 and its  $Y^t$ -counterpart, if we take  $\Delta_0 = \min\{\Delta, \tilde{\Delta}\} (> 0)$ , then there exists  $(\mathbf{A}, \mathbf{B}) \in X^t \times Y^t$  for all  $(\mathbf{C}, \mathbf{D}) \in X^{t-\Delta_0} \times Y^{t-\Delta_0}$  such that  $d_V((\mathbf{A}, \mathbf{B}), (\mathbf{C}, \mathbf{D})) < \delta$ . On the other hand, as we have seen, it holds that

$$X^t \times Y^t \cap JM(2 : 2) = \emptyset.$$

Since  $X^t \times Y^t$  and  $JM(2 : 2)$  are closed in  $V$ , and  $V$  is a metric space, we can apply Urysohn's Lemma [62]. It follows that there exists a continuous (in fact uniformly continuous since  $V$  is compact) function  $f: V \rightarrow [0, 1]$  satisfying  $f(U) = 0$  for any  $U \in X^t \times Y^t$  and  $f(W) = 1$  for any  $W \in JM(2 : 2)$ . The uniform continuity of  $f$  implies that for some  $\varepsilon \in (0, 1)$ , there is  $\delta > 0$  such that

$$d_V((\mathbf{E}', \mathbf{F}'), (\mathbf{E}, \mathbf{F})) < \delta \Rightarrow |f((\mathbf{E}', \mathbf{F}')) - f((\mathbf{E}, \mathbf{F}))| < \varepsilon \quad (4.99)$$

holds for any  $(\mathbf{E}, \mathbf{F}) \in V$ . For this  $\delta$ , we can apply the argument above: we can take  $\Delta_0 > 0$  such that for any  $(\mathbf{C}, \mathbf{D}) \in X^{t-\Delta_0} \times Y^{t-\Delta_0}$ , there exists  $(\mathbf{A}, \mathbf{B}) \in X^t \times Y^t$  satisfying  $d_V((\mathbf{A}, \mathbf{B}), (\mathbf{C}, \mathbf{D})) < \delta$ . Because  $f((\mathbf{A}, \mathbf{B})) = 0$ , we have  $f((\mathbf{C}, \mathbf{D})) < \varepsilon < 1$  (see (4.99)), and thus  $(\mathbf{C}, \mathbf{D}) \notin JM(2 : 2)$ . It indicates that  $X^{t-\Delta_0} \times Y^{t-\Delta_0} \cap JM(2 : 2) = \emptyset$ , that is, there is  $\Delta_0 > 0$  for any  $t \in L$  satisfying  $t - \Delta_0 \in L$ . Therefore,  $t'_0 = \inf L \notin L$  can be concluded.

## Chapter 5

# Thermodynamical entropy of mixing in regular polygon theories

The concept of entropy plays an important role in thermodynamics [149, 150]. It is possible to calculate the thermodynamical entropy of a mixture of classically different kinds of particles (such as a mixture of nitrogens and oxygens), and similar ideas were applied by von Neumann to the case when the system was composed of particles with different quantum internal states [5]. Similarly to the previous parts, it is expected that generalizing the notion of entropy to GPTs will help us to understand how entropy can affect our world. In fact, there have been researches which aim to introduce and investigate the concept of entropy in GPTs from informational perspectives [61, 151, 152, 153]. In those researches, some kinds of entropy were defined in all theories of GPTs and their information-theoretical properties were investigated. Meanwhile, there have been also researches referring to the thermodynamical entropy in terms of the microcanonical or canonical formulation in GPTs [154, 155], and researches referring to the thermodynamical entropy of mixing in GPTs [114, 156]. However, in those works, the entropy was only defined in or applied to some restricted theories of GPTs with special assumptions such as the existence of a spectral decomposition for any state into perfectly distinguishable pure states. In particular, it can be found that regular polygon theories do not always satisfy those assumptions, and thus entropy in regular polygon has never been investigated although they can be regarded as intermediate theories of a classical trit system and a qubit-like system, where entropy is defined successfully. It seems natural to ask how entropy of mixing behaves in regular polygon theories.

In this part, we consider thermodynamical entropy of mixing in regular

polygon theories. It is proven that the operationally natural thermodynamical entropy of a mixture of ideal particles with different internal states described by a regular polygon theory exists if and only if the state space of theory is triangle-shaped or disc-shaped, i.e., the theory is either classical or quantum-like. More precisely, we demonstrate that the thermodynamical entropy of mixing satisfying conditions imposed in [114], where the concrete operational construction of the entropy was given as von Neumann did under the assumption of the existence of semipermeable membranes, does not exist in all the regular polygon theories except for classical and quantum-like ones.

This part is organized as follows. In Section 5.1, we present a generalization of thermodynamically natural entropy of mixing in GPTs. We will see that the notion of perfect distinguishability plays an important role to define entropy also in GPTs. Then, we demonstrate our main theorem and its proof in Section 5.2.

## 5.1 Entropy of mixing in GPTs

In this section, we introduce the thermodynamically consistent definition of entropy of mixing in GPTs based on the notion of perfect distinguishability.

### 5.1.1 Perfect distinguishability for regular polygon theories

We recall that a family of states  $\{\omega_i\}_i$  is called perfectly distinguishable if there exists an observable  $\{e_i\}_i$  such that  $e_i(\omega_j) = \delta_{ij}$ . Let us characterize perfectly distinguishable states in regular polygon theories. We first consider the regular polygon theory with  $n$  sides, where  $n$  is an even number greater than two. Calculating the Euclidean inner product (denoted by  $(\cdot, \cdot)$  here) of pure effects and pure states in Subsection 2.5.3, we obtain

$$(e_i^n, \omega_i^n) = (e_i^n, \omega_{i-1}^n) = 1, \quad (e_i^n, \omega_{i+\frac{n}{2}-1}^n) = (e_i^n, \omega_{i+\frac{n}{2}}^n) = 0.$$

These equations indicate that any state in  $\Omega_n^{[i-1, i]}$  is perfectly distinguishable from any state in  $\Omega_n^{[i+\frac{n}{2}-1, i+\frac{n}{2}]}$ , where we define

$$\Omega_n^{[k-1, k]} = \{\omega \in \Omega_n \mid \omega = p\omega_{k-1}^n + (1-p)\omega_k^n, 0 \leq p \leq 1\},$$

since the measurement  $\{e_i^n, u - e_i^n\}$  distinguishes perfectly those two states. For odd  $n$  ( $\geq 3$ ), we obtain

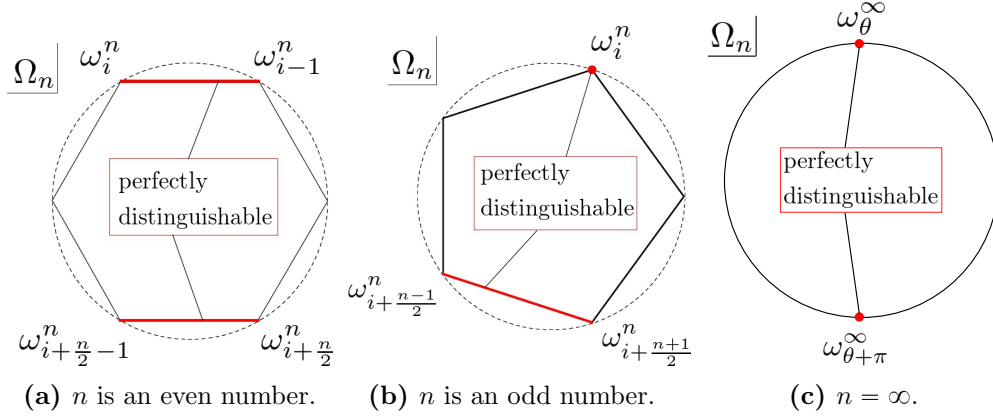
$$(e_i^n, \omega_i^n) = 1, \quad (e_i^n, \omega_{i+\frac{n-1}{2}}^n) = (e_i^n, \omega_{i+\frac{n+1}{2}}^n) = 0.$$



Hence,  $\omega_i^n$  and an arbitrary state in  $\Omega_n^{[i+\frac{n-1}{2}, i+\frac{n+1}{2}]}$  are perfectly distinguishable. Finally, when  $n = \infty$ ,

$$(e_\theta^\infty, \omega_\theta^\infty) = 1, \quad (e_\theta^\infty, \omega_{\theta+\pi}^\infty) = 0.$$

hold, so there is only one perfectly distinguishable state for each pure state (see Figure 5.1).



**Figure 5.1:** Pairs of perfectly distinguishable states in the  $n$ -gon state space.

### 5.1.2 Entropy of mixing in GPTs

In this part, we consider the thermodynamical entropy of mixing in a system composed of ideal gases with different internal degrees of freedom described by a GPT. In thermodynamics, it is well known that a mixture of several classically distinct ideal gases, such like a mixture of ideal hydrogens and nitrogens, causes an increase of entropy. The amount of increase by the mixture can be calculated under the assumption of the existence of semipermeable membranes which distinguish perfectly those particles. We assume in a similar way that if the internal states  $\omega_1, \omega_2, \dots, \omega_l$  described by a GPT are perfectly distinguishable, then there exist semipermeable membranes which can identify completely a state among them without disturbing every  $\omega_j$  ( $j = 1, 2, \dots, l$ ).

We consider ideal gases in thermal equilibrium with its temperature  $T$ , volume  $V$ , and  $N$  particles, and do not focus on the mechanical part of the particles in the following. All of these  $N$  particles are in the same internal state  $\omega = \sum_{i=1}^l p_i \omega_i$ , where  $\{\omega_1, \omega_2, \dots, \omega_l\}$  is a perfectly distinguishable set of states, and  $\forall i, p_i \geq 0$ , and  $\sum_{i=1}^l p_i = 1$ , meaning that this system is composed of the mixture of  $l$  different kinds of particles whose internal states are  $\omega_1, \omega_2, \dots, \omega_l$  with a probability weight  $\{p_1, p_2, \dots, p_l\}$ . We note again that in this chapter, classical species of particles are also regarded as

the internal states of them. In classical thermodynamics, thermodynamical entropy is calculated by constructing concrete thermodynamical operations such as isothermal or adiabatic quasistatic operations. We follow this doctrine of thermodynamics also in GPTs that thermodynamical entropy, especially thermodynamical entropy of mixing, should be operationally-derived quantity. In fact, as shown in [114], our assumption of the existence of semipermeable membranes makes it possible to realize concrete thermodynamical operations to calculate the thermodynamical entropy of mixing of the system mentioned above in the same way as von Neumann did when the internal degrees of freedom were quantum [5]. Strictly speaking, it has been demonstrated operationally in [114] that the thermodynamical entropy of mixing in the system is

$$S(\omega) = \sum_{i=1}^l p_i S(\omega_i) - \sum_{i=1}^l p_i \log p_i, \quad (5.1)$$

where  $S(\sigma)$  means the per-particle thermodynamical entropy of mixing in the system which consists of particles in the same state  $\sigma$ , and we set the Boltzmann constant  $k_B = 1$  (also  $0 \log 0 = 0$ ). In the process of deriving (5.2), the additivity and extensivity of the thermodynamical entropy, and the continuity of  $S$  with respect to states are assumed. The latter one is needed in order to apply (5.2) to arbitrary states with an arbitrary probability weight, while its operational derivation has been given only when each  $p_i N$  is the number of particles in the state  $\omega_i$  and thus each  $p_i$  is rational. We impose additional assumption that the entropy of any pure state equals to zero, that is,  $S(\sigma) = 0$  whenever  $\sigma$  is a pure state.

## 5.2 Main result

Our main result is in the following form.

### Theorem 5.1

*Consider a system in thermal equilibrium composed of ideal gases whose internal states are described an element of the state space  $\Omega_n$  of the regular polygon theory with  $n$  sides ( $n \geq 3$ ). The (per-particle) thermodynamical entropy of mixing  $S : \Omega_n \rightarrow \mathbb{R}$  satisfying*

$$S(\omega) = \sum_{i=1}^l p_i S(\omega_i) - \sum_{i=1}^l p_i \log p_i, \quad (5.2)$$

*where  $\{\omega_i\}_i$  is a family of perfectly distinguishable states, exists if and only if  $n = 3$  or  $\infty$ , that is, the state space is classical or quantum-like.*

**Proof**

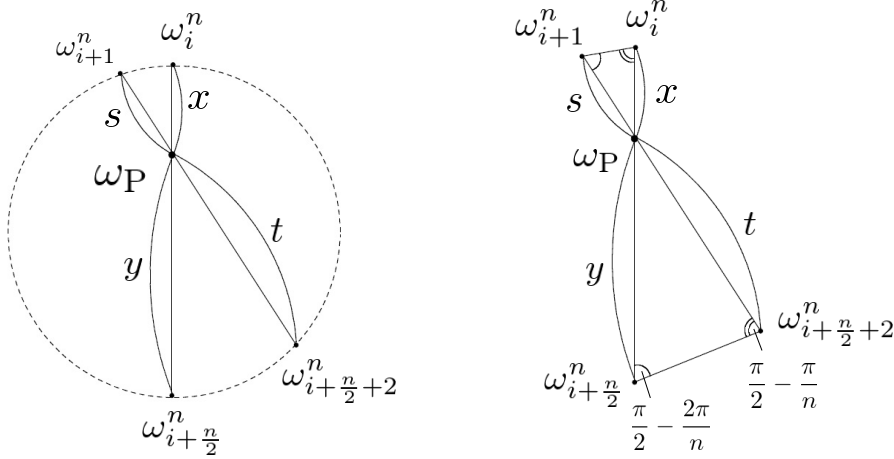
For  $n = 3$ , because it is a classical system, any  $\omega \in \Omega_3$  is decomposed uniquely into perfectly distinguishable pure states as  $\omega = p\omega_0^3 + q\omega_1^3 + (1-p-q)\omega_2^3$ , where  $\omega_i^3$  ( $i = 0, 1, 2$ ) are the three pure states in  $\Omega_3$  and  $\{p, q, 1-p-q\}$  is a probability weight. In this settings, we define  $S$  as

$$S(\omega) = -p \log p - q \log q - (1-p-q) \log(1-p-q).$$

This  $S$  gives the well-defined entropy satisfying (5.2). Similarly, when  $n = \infty$ , any state has only one decomposition into perfectly distinguishable (pure) states except for the central state of  $\Omega_\infty$  (the maximally mixed state). For states that are not maximally mixed, we define  $S$  as

$$S(\omega) = H(p),$$

where we decompose a non-maximally-mixed  $\omega \in \Omega_\infty$  as  $\omega = p\omega_\theta^\infty + (1-p)\omega_{\theta+\pi}^\infty$  ( $0 \leq p \leq 1$ ) and  $H(p) = -p \log p - (1-p) \log(1-p)$  is the *1-bit Shannon entropy*. We can apply this  $S$  to the maximally mixed state, for the probability weight does not depend on the way of decompositions and they are always  $\{\frac{1}{2}, \frac{1}{2}\}$ . Therefore, we can define successfully the thermodynamical entropy  $S$  which meets (5.2) for  $n = 3, \infty$ . In the following, we prove the only if part.



**Figure 5.2:** Illustration of the state  $\omega_P$ .

The case when  $n = 4$  was proven in [114], so we only consider  $n \geq 5$ . At first, we assume  $n$  is an even number, and consider the state  $\omega_P$  represented in Figure 5.2, that is,

$$\omega_P = \frac{y}{x+y} \omega_i^n + \frac{x}{x+y} \omega_{i+\frac{n}{2}}^n = \frac{t}{s+t} \omega_{i+1}^n + \frac{s}{s+t} \omega_{i+\frac{n}{2}+2}^n,$$

where  $x, y, s, t$  are all nonnegative, and  $x \leq y$  and  $s \leq t$  as shown in Figure

5.2. Note that  $\{\omega_i^n, \omega_{i+\frac{n}{2}}^n\}$  and  $\{\omega_{i+1}^n, \omega_{i+\frac{n}{2}+2}^n\}$  are two perfectly distinguishable pairs of pure states. From the observations in the previous section, we obtain two forms of the thermodynamical entropy of mixing:

$$S(\omega_P) = H\left(\frac{x}{x+y}\right) = H\left(\frac{s}{s+t}\right), \quad (5.3)$$

which means

$$\frac{x}{x+y} = \frac{s}{s+t}$$

because  $x \leq y$  and  $s \leq t$ . On the other hand, applying sine theorem to Figure 5.2 we can see that

$$\frac{x}{\sin(\frac{\pi}{2} - \frac{2\pi}{n})} = \frac{s}{\sin(\frac{\pi}{2} - \frac{\pi}{n})}, \quad \frac{y}{\sin(\frac{\pi}{2} - \frac{\pi}{n})} = \frac{t}{\sin(\frac{\pi}{2} - \frac{2\pi}{n})},$$

namely

$$\frac{s}{s+t} = \frac{x}{x + \left(\frac{\cos \frac{2\pi}{n}}{\cos \frac{\pi}{n}}\right)^2 y}$$

holds. It follows from these two equations that

$$\left(\frac{\cos \frac{2\pi}{n}}{\cos \frac{\pi}{n}}\right)^2 = 1,$$

and because for even  $n$ , this equation holds if and only if  $n = \infty$  ( $\cos(\frac{\pi}{n}) = 1$ ), the entropy in (5.3) has been proven to be ill-defined.

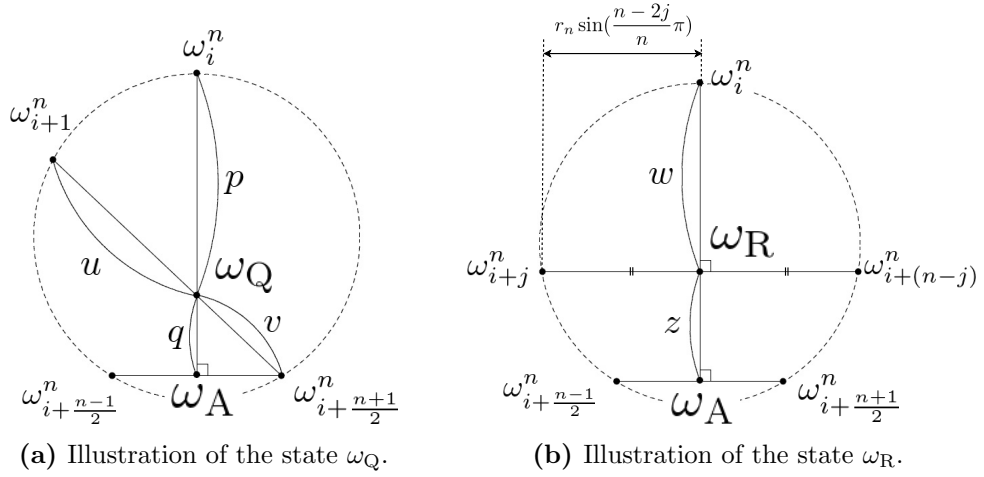
Next, we consider the case when  $n$  is an odd number greater than three. We define the state  $\omega_A$  as  $\omega_A = \frac{1}{2}(\omega_{i+\frac{n-1}{2}}^n + \omega_{i+\frac{n+1}{2}}^n)$ , and consider two states  $\omega_Q$  and  $\omega_R$  shown in Figure 5.3, where  $j = \frac{n+1}{4}$  or  $\frac{n-1}{4}$  corresponding to the case when  $n \equiv 3$  or  $n \equiv 1 \pmod{4}$  respectively. Note that  $\{\omega_i^n, \omega_A\}$  and  $\{\omega_{i+1}^n, \omega_{i+\frac{n+1}{2}}^n\}$  in Figure 5.3a, and  $\{\omega_{i+j}^n, \omega_{i+(n-j)}^n\}$  in Figure 5.3b are perfectly distinguishable pairs of states. Then,

$$S(\omega_Q) = \frac{p}{p+q}S(\omega_A) + H\left(\frac{p}{p+q}\right) = H\left(\frac{u}{u+v}\right),$$

and

$$S(\omega_R) = \frac{w}{w+z}S(\omega_A) + H\left(\frac{w}{w+z}\right) = H\left(\frac{1}{2}\right)$$

hold. We assume that the entropies of the two states  $\omega_Q, \omega_R$  are well-defined



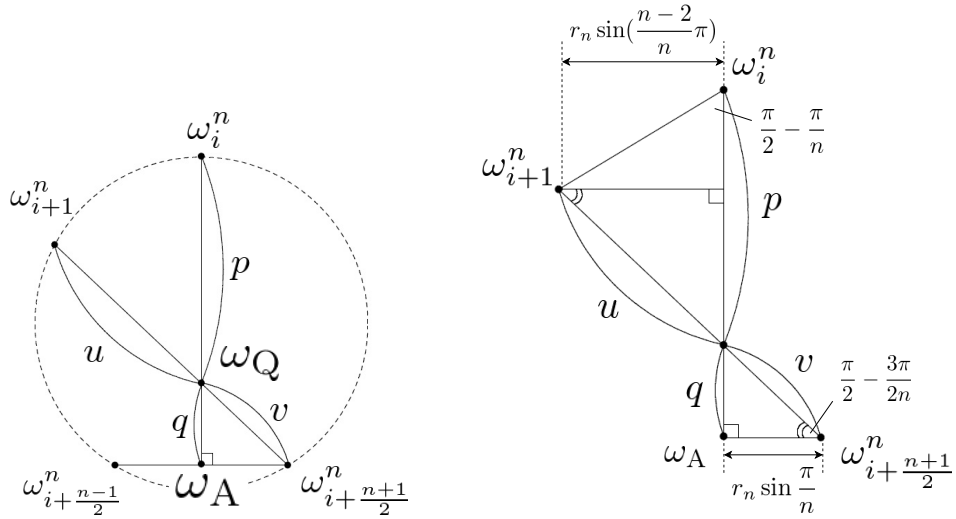
**Figure 5.3:** Illustration of the states  $\omega_Q$  and  $\omega_R$ .

(so is  $\omega_A$ ). Then,

$$S(\omega_A) = \frac{p+q}{p} \left\{ H\left(\frac{u}{u+v}\right) - H\left(\frac{p}{p+q}\right) \right\} \quad (5.4)$$

$$= \frac{w+z}{w} \left\{ H\left(\frac{1}{2}\right) - H\left(\frac{w}{w+z}\right) \right\} \quad (5.5)$$

holds. Let us give the explicit expressions of (5.5). From Figure 5.4, we obtain



**Figure 5.4:** The decomposition for  $\omega_Q$ .

$$\frac{u}{v} = \frac{r_n \sin\left(\frac{n-2}{n}\pi\right)}{r_n \sin\frac{\pi}{n}} = \frac{\sin\frac{2\pi}{n}}{\sin\frac{\pi}{n}} = 2 \cos\frac{\pi}{n},$$

and by sine theorem,

$$\frac{p}{\sin(\frac{\pi}{2} - \frac{\pi}{2n})} = \frac{u}{\sin(\frac{\pi}{2} - \frac{\pi}{n})}, \quad q = v \cos \frac{3\pi}{2n}$$

hold. Therefore,

$$\begin{aligned} \frac{q}{p} &= \frac{v}{u} \cdot \frac{\cos \frac{\pi}{n} \cdot \cos \frac{3\pi}{2n}}{\cos \frac{\pi}{2n}} \\ &= \frac{\cos \frac{3\pi}{2n}}{2 \cos \frac{\pi}{2n}} \\ &= \frac{1}{2} \left( 4 \cos^2 \frac{\pi}{2n} - 3 \right) \\ &= \frac{1}{2} \left( 2 \cos \frac{\pi}{n} - 1 \right). \end{aligned}$$

On the other hand, from Figure 5.3b, we obtain

$$\begin{aligned} \frac{z}{w} &= \frac{2r_n \cos^2 \frac{\pi}{2n} - 2r_n \cos^2(\frac{n-2j}{2n}\pi)}{2r_n \cos^2(\frac{n-2j}{2n}\pi)} \\ &= \frac{\cos^2 \frac{\pi}{2n} - \sin^2 \frac{j\pi}{n}}{\sin^2 \frac{j\pi}{n}} \quad \text{for } j = \frac{n \pm 1}{4}. \end{aligned} \quad (5.6)$$

Since

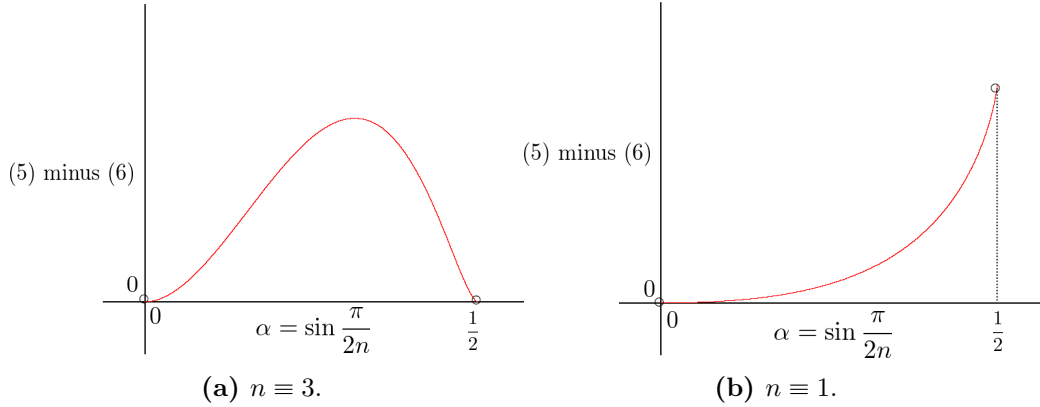
$$\begin{aligned} \sin^2 \frac{j\pi}{n} &= \frac{1}{2} \left( 1 - \cos \frac{2j\pi}{n} \right) \\ &= \frac{1}{2} \left( 1 - \cos \frac{(n \pm 1)\pi}{2n} \right) \\ &= \frac{1}{2} \left( 1 \pm \sin \frac{\pi}{2n} \right), \end{aligned} \quad (5.7)$$

the equation above can be written as

$$\begin{aligned} \frac{z}{w} &= \frac{2(1 + \sin \frac{\pi}{2n})(1 - \sin \frac{\pi}{2n}) - (1 \pm \sin \frac{\pi}{2n})}{1 \pm \sin \frac{\pi}{2n}} \\ &= 1 \mp 2 \sin \frac{\pi}{2n}, \end{aligned}$$

where the double sign corresponds to the ones in (5.6) and (5.7), and the upper and lower sign correspond to the case of  $n \equiv 3$  and  $n \equiv 1 \pmod{4}$  respectively. Substituting these results to (5.4) and (5.5), we obtain

$$\begin{aligned} S(\omega_A) &= \left( \frac{2 \cos \frac{\pi}{n} + 1}{2} \right) \left\{ H \left( \frac{1}{2 \cos \frac{\pi}{n} + 1} \right) - H \left( \frac{2}{2 \cos \frac{\pi}{n} + 1} \right) \right\} \\ &= \left( 2 \mp 2 \sin \frac{\pi}{2n} \right) \left\{ H \left( \frac{1}{2} \right) - H \left( \frac{1}{2 \mp 2 \sin \frac{\pi}{2n}} \right) \right\}. \end{aligned}$$



**Figure 5.5:** The difference between (5.8) and (5.9).

By letting  $\alpha = \sin \frac{\pi}{2n}$ , it can be rewritten as

$$S(\omega_A) = 2\alpha^2 \log 2 + \frac{1 - 4\alpha^2}{2} \log(1 - 4\alpha^2) - (1 - 2\alpha^2) \log(1 - 2\alpha^2) \quad (5.8)$$

$$= (1 \mp 2\alpha) \log(1 \mp 2\alpha) - (2 \mp 2\alpha) \log(1 \mp \alpha), \quad (5.9)$$

where the upper and lower signs correspond to the case of  $n \equiv 3$  and  $n \equiv 1 \pmod{4}$  respectively. The differences between (5.8) and (5.9) in the case of  $n \equiv 3$  and  $n \equiv 1$  are displayed in Figures 5.5a and 5.5b respectively, and we can see that the two forms of  $S(\omega_A)$  shown in (5.8) and (5.9) do not agree with each other. In conclusion, it has been proven that if  $n \neq 3, \infty$ , then there exists some state whose thermodynamical entropy of mixing is ill-defined.  $\square$

We can see the ill-defined values of entropy become well-defined if  $n = 3, \infty$  in our proof. For example, when  $n$  is an odd number,  $\alpha = \sin \frac{\pi}{2n}$  equals to  $\frac{1}{2}$  or 0 if  $n$  equals to three or infinite, respectively, and two values (5.8) and (5.9) coincide with each other in these cases (see Figure 5.5).

*Remark 5.2*

Similar results were obtained in [114], where it was assumed that any state could be represented as a convex combination of perfectly distinguishable pure states. However, a state of a regular polygon theory is not always represented by a convex combination of perfectly distinguishable pure states. For instance, we can see from Figure 5.1 that the state  $\omega_A$  in Figure 5.3a or Figure 5.3b can not be decomposed into perfectly distinguishable pure states. Thus, regular polygon theories generally do not satisfy the assumption in the previous study [114], and our result is the one about the existence of well-defined thermodynamical entropy in such a broader class of theories where “spectral decompositions” of states are not generally possible.

# Chapter 6

## Summary

In this thesis, we have studied the notion of uncertainty in quantum theory via convexity. We focused on three expressions of uncertainty: uncertainty relations, incompatibility, and thermodynamical entropy. Our establishments were based on the idea that to see quantum uncertainty from a broader point of view makes it possible to understand its essence. In fact, in each chapter, we considered uncertainty relations and thermodynamical entropy in a broader class of theories than quantum theory, and investigated quantum incompatibility, which is a broader notion than uncertainty.

In Chapter 2, we introduced the mathematical framework of GPTs. We saw that GPTs are constructed by requiring only primitive convexity originating from probability mixtures, and thus in this sense, they are the most general description of nature (in particular, broader than quantum theory). There were proven that GPTs can be expressed mathematically in terms of ordered Banach spaces, and that it indeed reproduce the usual formulation of quantum theory.

In Chapter 3, several generalizations of uncertainty relations such as entropic uncertainty relations were considered in a class of GPTs which can be considered as generalized theories of quantum theory. It was revealed that similar quantitative relations between preparation and measurement uncertainty to quantum case hold also in GPTs although only theories with transitivity and self-duality with respect to a certain inner product were considered. We also gave concrete calculations of our results for regular polygon theories. What is also specific to theorems is that they were obtained without considering entanglement or even composite systems while the quantum results of the previous studies were based on the “ricochet” property of maximally entangled states. It may be indicated that some of the characteristics of quantum theory can be obtained without entanglement. Future research should reveal the relations between the maximal entanglement and self-duality, which will be a key to generalizing our theorems to infinite-dimensional cases (remember that the maximally entangled



states cannot be defined in infinite-dimensional quantum theories such as  $\mathcal{H} = L^2(\mathbb{R})$ . To find information-theoretic applications of our results is also left for future work.

In Chapter 4, we focused on incompatibility in quantum theory, and introduced the notions of compatibility and incompatibility dimensions for collections of quantum devices. They describe the minimum number of states which are needed to detect incompatibility and the maximum number of states on which incompatibility vanishes, respectively. We have not only presented general properties of those quantities but also examined concrete behaviors of them for a pair of unbiased qubit observables. We have proven that even for this simple pair of incompatible observables there exist two types of incompatibility with different incompatibility dimensions which cannot be observed if we focus only on robustness of incompatibility under noise. We expect that it is possible to apply this difference to some quantum protocols such as quantum cryptography. Future work will be needed to investigate whether similar results can be obtained for observables in higher dimensional Hilbert space or other quantum devices. As the definitions apply to devices in GPTs, an interesting task is further to see how quantum incompatibility dimension differs from incompatibility dimension in general.

In Chapter 5, we returned to GPTs and considered thermodynamical entropy in regular polygon theories. We showed that only classical and quantum-like theories (i.e. the triangle and disc theories respectively) allow the operationally natural entropy to be consistent. Further research is required to reveal if we can obtain the same results in higher dimensional cases. Moreover, the proof of our main theorem indicates that the entropy discussed above is defined successfully in other theories where the probability coefficients obtained when a state is decomposed into perfectly distinguishable states are unique even though the state space is neither classical nor quantum. This means that we need to impose additional conditions on the entropy to remove those “unreasonable” theories, which is also a future problem.

# Acknowledgments

I would like to, first and foremost, thank my supervisor Takayuki Miyadera. Since I became his student in 2016, he has taught me many interesting things on quantum theory to lead me to the study of quantum foundations. Topics such as quantum uncertainty or incompatibility are examples of those things, and they are now the main interest of my research. His insightful guidance, not only as a great researcher but also as a great person, has helped and will help me a lot in my research and daily life. Nothing in my nine-year college life makes me more proud of than being his student and able to learn from him. I also would like to thank my collaborator Teiko Heinosaari. I feel very proud of working with such a great person and exhibiting our results in my PhD thesis. I cannot miss thanking all members of my laboratory, especially Kenzo Ogure, Ikko Hamamura, and Kazuki Yamaga, for their fruitful help in my study. This research has been supported by JSPS KAKENHI Grant Number JP21J10096, and I also acknowledge the JSPS Research Fellowship for Young Scientists.

Finally, I would like to express my thanks to my family and friends. I am grateful to my parents for giving me birth with so good a constitution that I can stay healthy in everyday life. I also thank my friends for their encouragement and support. They are all great persons, and examples for me. I have learned from them what are the most important things in life. Those things are always within my mind, and are the principles that guide my life. My deepest gratitude is expressed to my grandparents. The greatest thing in my life is that they are my grandparents. We live so far that we cannot often see each other, but I always feel their love supporting and encouraging me in my daily life. I hope that my accomplishment of this PhD study will be their pleasure.

# Appendix

## A Proof of Proposition 2.50

In this part, we give a proof of Proposition 2.50. We need the following proposition, which holds without the assumption of the transitivity of  $\Omega$ .

### Proposition A.1

For a state space  $\Omega$ , define a linear map  $P_M: V \rightarrow V$  by

$$P_M x = \int_{GL(\Omega)} T x \, d\mu(T).$$

Then,  $P_M$  is an orthogonal projection with respect to the inner product  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$ , i.e.

$$P_M = P_M^2 \quad \text{and} \quad \langle P_M x, y \rangle_{GL(\Omega)} = \langle x, P_M y \rangle_{GL(\Omega)} \quad \text{for all } x, y \in V.$$

### Proof

We denote the inner product  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$  simply by  $\langle \cdot, \cdot \rangle$  in this proof.

Let  $V_M := \{x \in V \mid Tx = x \text{ for all } T \in GL(\Omega)\}$  be the set of all fixed points with respect to  $GL(\Omega)$ . Then, it is easy to see that  $P_M x_M = x_M$  for any  $x_M \in V_M$  and  $V_M = \text{Im} P_M$  (in particular  $V_M$  is a subspace of  $V$ ). Therefore,

$$P_M^2 x = P_M(P_M x) = P_M x$$

holds for any  $x \in V$ , and thus  $P_M^2 = P_M$ . On the other hand, we can observe

$$\begin{aligned} \langle P_M x, y \rangle &= \int_{GL(\Omega)} d\mu(T) (TP_M x, Ty)_E \\ &= \int_{GL(\Omega)} d\mu(T) (P_M x, Ty)_E \\ &= \int_{GL(\Omega)} d\mu(T) \left( \int_{GL(\Omega)} d\mu(S) Sx, Ty \right)_E. \end{aligned} \quad (\text{A.1})$$

Let us fix an orthonormal basis  $\{w_i\}_{i=1}^{N+1}$  of  $V$  compatible with the standard

Euclidean inner product of  $V$ , i.e.

$$(w_i, w_j)_E = \delta_{ij}.$$

We can consider representing the vector  $\int_{GL(\Omega)} d\mu(S)Sx \in V$  by means of the orthonormal basis  $\{w_i\}_i$  as

$$\int_{GL(\Omega)} d\mu(S)Sx = \sum_i \left( w_i, \int_{GL(\Omega)} d\mu(S)Sx \right)_E w_i.$$

In fact, the “ $i$ th-element”  $\left( w_i, \int_{GL(\Omega)} d\mu(S)Sx \right)_E$  is given by (see [157] for more details)

$$\left( w_i, \int_{GL(\Omega)} d\mu(S)Sx \right)_E = \int_{GL(\Omega)} d\mu(S) (w_i, Sx)_E.$$

It results in

$$\begin{aligned} \left( \int_{GL(\Omega)} d\mu(S)Sx, Ty \right)_E &= \sum_i \left[ \int_{GL(\Omega)} d\mu(S)(w_i, Sx)_E \right] (w_i, Ty)_E \\ &= \int_{GL(\Omega)} d\mu(S) \left[ \sum_i (Sx, w_i)_E (w_i, Ty)_E \right] \\ &= \int_{GL(\Omega)} d\mu(S) (Sx, Ty)_E. \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} \int_{GL(\Omega)} d\mu(T) \left( \int_{GL(\Omega)} d\mu(S)Sx, Ty \right)_E &= \int_{GL(\Omega)} d\mu(T) \left[ \int_{GL(\Omega)} d\mu(S) (Sx, Ty)_E \right] \\ &= \int_{GL(\Omega)} d\mu(S) \left[ \int_{GL(\Omega)} d\mu(T) (Sx, Ty)_E \right] \\ &= \int_{GL(\Omega)} d\mu(S) \left( Sx, \int_{GL(\Omega)} d\mu(T)Ty \right)_E, \end{aligned}$$

where we use Fubini’s theorem for the finite Haar measure  $\mu$  on  $GL(\Omega)$ . We can conclude together with (A.1) that

$$\langle P_M x, y \rangle = \langle x, P_M y \rangle$$

holds. □

Proposition A.1 enables us to give an orthogonal decomposition of a

vector  $x \in V$  such that

$$x = (\mathbb{1} - P_M)x + P_Mx, \quad (\text{A.2})$$

where  $(\mathbb{1} - P_M)x \in V_M^\perp$  and  $P_Mx \in V_M$ . When the transitivity of  $\Omega$  is assumed, (A.2) is reduced to Proposition 2.50.

**Proposition 2.50**

*For a transitive state space  $\Omega$ , there exists a basis  $\{v_l\}_{l=1}^{N+1}$  of  $V$  orthonormal with respect to the inner product  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$  such that  $v_{N+1} = \omega_M$  and*

$$x \in \text{aff}(\Omega) \iff x = \sum_{l=1}^N a_l v_l + v_{N+1} = \sum_{l=1}^N a_l v_l + \omega_M \quad (a_1, \dots, a_N \in \mathbb{R}).$$

**Proof**

Since we set  $\dim \text{aff}(\Omega) = N$ , there exists a set of  $N$  linear independent vectors  $\{v_l\}_{l=1}^N \subset [\text{aff}(\Omega) - \omega_M]$  which forms a basis of the  $N$ -dimensional vector subspace  $[\text{aff}(\Omega) - \omega_M] \subset V$ , and we can assume by taking an orthonormalization that they are orthonormal with respect to the inner product  $\langle \cdot, \cdot \rangle$ . Hence,  $x \in \text{aff}(\Omega)$  if and only if it is represented as

$$x = \sum_{l=1}^N a_l v_l + \omega_M \quad (a_1, \dots, a_N \in \mathbb{R}). \quad (\text{A.3})$$

Moreover, because of the definition of  $\text{aff}(\Omega)$ , for every  $v_l \in [\text{aff}(\Omega) - \omega_M]$  there exist  $k \in \mathbb{N}$ , real numbers  $\{b_i\}_{i=1}^k$  satisfying  $\sum_{i=1}^k b_i = 1$ , and states  $\{\omega_i\}_{i=1}^k$  such that  $v_l = \sum_{i=1}^k b_i \omega_i - \omega_M$ . By means of Proposition 2.49, we obtain for all  $l = 1, 2, \dots, N$

$$\begin{aligned} P_M v_l &= \sum_{i=1}^k b_i P_M \omega_i - P_M \omega_M \\ &= \sum_{i=1}^k b_i \omega_M - \omega_M = 0. \end{aligned} \quad (\text{A.4})$$

Therefore, because of Proposition A.1

$$\begin{aligned} \langle \omega_M, v_l \rangle &= \langle P_M \omega_M, v_l \rangle \\ &= \langle \omega_M, P_M v_l \rangle \\ &= 0 \end{aligned}$$

holds for all  $l = 1, 2, \dots, N$ , and we can conclude together with the unit norm of  $\omega_M$  that  $\{v_1, \dots, v_N, \omega_M\}$  in (A.3) forms an orthonormal basis of the  $(N+1)$ -dimensional vector space  $V$  with respect to  $\langle \cdot, \cdot \rangle$  and Proposition

2.50 is proven (we can also find that (A.3) corresponds to (A.2)).  $\square$

## B Proof of Proposition 2.52

In this part, we prove Proposition 2.52. As we have so far, we let  $\Omega$  be a state space,  $V_+$  be the positive cone generated by  $\Omega$ , and  $GL(\Omega)$  be the set of all state automorphisms on  $\Omega$  in the following.

### Lemma B.1

$V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int}$  is a  $GL(\Omega)$ -invariant set. That is,  $TV_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int} = V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int}$  for all  $T \in GL(\Omega)$ .

### Proof

Let  $w \in V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int}$ . It holds that  $\langle w, v \rangle_{GL(\Omega)} \geq 0$  for all  $v \in V_+$ . Because any  $T \in GL(\Omega)$  is an orthogonal transformation with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$ , we obtain

$$\langle Tw, v \rangle_{GL(\Omega)} = \langle w, T^{-1}v \rangle_{GL(\Omega)} \geq 0$$

for all  $v \in V_+$ . Therefore,  $TV_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int} \subset V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int}$  holds, and a similar argument for  $T^{-1} \in GL(\Omega)$  proves the lemma.  $\square$

### Lemma B.2

Let  $(\cdot, \cdot)$  be an arbitrary inner product on  $V$ .  $V_+$  is self-dual if and only if there exists a linear map  $J: V \rightarrow V$  such that  $J$  is strictly positive with respect to  $(\cdot, \cdot)$ , i.e.  $(x, Jy) = (Jx, y)$  for all  $x, y \in V$  and  $(x, Jx) > 0$  for all  $x \in V$ , and  $J(V_+) = V_{+(\cdot, \cdot)}^{*int}$ .

### Proof

If part: We introduce an inner product  $(\cdot, \cdot)_J = (\cdot, J\cdot)$ .  $V_{+(\cdot, \cdot)_J}^{*int}$  is written as

$$\begin{aligned} V_{+(\cdot, \cdot)_J}^{*int} &= \{v \mid (v, w)_J \geq 0, \forall w \in V_+\} \\ &= \{v \mid (v, Jw) \geq 0, \forall w \in V_+\} \\ &= \{v \mid (Jv, w) \geq 0, \forall w \in V_+\}. \end{aligned}$$

Thus,  $v \in V_{+(\cdot, \cdot)_J}^{*int}$  is equivalent to  $Jv \in V_{+(\cdot, \cdot)}^{*int}$ . It concludes  $V_{+(\cdot, \cdot)_J}^{*int} = J^{-1}(V_{+(\cdot, \cdot)}^{*int}) = V_+$ .

Only if part: Let  $V_+$  be self-dual with respect to an inner product  $\langle \cdot, \cdot \rangle$ . There exists some  $K: V \rightarrow V$  strictly positive with respect to  $(\cdot, \cdot)$  such that  $\langle \cdot, \cdot \rangle = (\cdot, K\cdot)$ . We obtain

$$\begin{aligned} V_+ &= V_{+\langle \cdot, \cdot \rangle}^{*int} = \{v \mid \langle v, w \rangle \geq 0, \forall w \in V_+\} \\ &= \{v \mid (v, Kw) \geq 0, \forall w \in V_+\} \\ &= \{v \mid (Kv, w) \geq 0, \forall w \in V_+\} \end{aligned}$$

Thus,  $v \in V_+ = V_{+\langle \cdot, \cdot \rangle}^{*int}$  is equivalent to  $Kv \in V_{+\langle \cdot, \cdot \rangle}^{*int}$ , i.e.  $KV_+ = V_{+\langle \cdot, \cdot \rangle}^{*int}$ . Define  $J = K$ .  $\square$

In Lemma B.2, we gave a necessary and sufficient condition for  $V_+$  with an inner product  $(\cdot, \cdot)$  to be self-dual. The condition was the existence of a strictly positive map  $J$  satisfying  $J(V_+) = V_{+\langle \cdot, \cdot \rangle}^{*int}$ . This map  $J$  may not be unique. For instance, let us consider a classical system in  $\mathbb{R}^2$  whose extreme points are two points  $(1, 1)$  and  $(1, -1)$ . The positive cone is a ‘‘forward lightcone’’  $V_+ = \{(x_0, x_1) \mid x_0 \geq 0, x_0^2 - x_1^2 \geq 0\}$ . It is easy to see that  $V_+ = V_{+\langle \cdot, \cdot \rangle_E}^{*int}$  with the standard Euclidean inner product  $(\cdot, \cdot)_E$ . However, if we choose an orthogonal basis  $\{v_0, v_1\}$  of  $\mathbb{R}^2$  given by  $v_0 = (1, 1)$  and  $v_1 = (1, -1)$ , then every linear map of the form

$$\begin{pmatrix} v_0 \\ v_1 \end{pmatrix} \mapsto \begin{pmatrix} \lambda_0 v_0 \\ \lambda_1 v_1 \end{pmatrix}$$

for  $\lambda_0, \lambda_1 > 0$  (which contains ‘‘Lorentz transformations’’ in 1 + 1 dimension) is strictly positive and makes  $V_+$  invariant. Nevertheless, when  $|\Omega^{\text{ext}}| < \infty$ , we can demonstrate that such strictly positive maps are ‘‘equivalent’’ to each other .

### Lemma B.3

Let  $|\Omega^{\text{ext}}| < \infty$ . If a linear map  $J : V \rightarrow V$  is strictly positive with respect to an inner product  $(\cdot, \cdot)$ , i.e.  $(x, Jy) = (Jx, y)$  for all  $x, y \in V$  and  $(x, Jx) > 0$  for all  $x \in V$ , and satisfies  $J(V_+) = V_+$ , then for each  $\omega^{\text{ext}} \in \Omega^{\text{ext}}$  there exists  $\mu(\omega^{\text{ext}}) > 0$  such that  $J(\omega^{\text{ext}}) = \mu(\omega^{\text{ext}})\omega^{\text{ext}}$ .

### Proof

Any  $\omega^{\text{ext}} \in \Omega^{\text{ext}}$  is represented as  $\omega^{\text{ext}} = c(\omega^{\text{ext}})w$  with  $c(\omega^{\text{ext}}) := \|\omega^{\text{ext}}\| = (\omega^{\text{ext}}, \omega^{\text{ext}})^{1/2}$  and  $w$  satisfying  $\|w\| = 1$ . Suppose that there exists a family

$$\{\omega_k^{\text{ext}}\}_{k=1}^Z = \{c(\omega_k^{\text{ext}})w_k\}_{k=1}^Z \subset \Omega^{\text{ext}}$$

such that there is no  $\mu(\omega_k^{\text{ext}}) > 0$  for every  $k = 1, 2, \dots, Z$  satisfying  $J(\omega_k^{\text{ext}}) = \mu(\omega_k^{\text{ext}})\omega_k^{\text{ext}}$ , and define  $W := \{w_k\}_{k=1}^Z$ . Since  $J$  maps each extreme ray of  $V_+$  to an extreme ray of  $V_+$ ,  $J(w_k)$  with  $w_k \in W$  is proportional to some  $\omega^{\text{ext}} \in \Omega^{\text{ext}}$  (remember that an extreme ray of  $V_+$  is the set of positive scalar multiples of an extreme point of  $\Omega$ ). We can see that  $J(w_k)$  is proportional to some  $w_p \in W$  with  $p \neq k$  considering that  $J(J(w_k)) = \mu J(w_k)$  holds if and only if  $J(w_k) = \mu w_k$  holds.

We shall show in the following that there is a  $w_q \in W$  such that  $J(w_q) \notin W$  despite of the argument above. To prove the claim, let us diagonalize  $J$ . It is written as  $J = \sum_{n=1}^M \tau_n R_n$ , where  $\tau_1 > \tau_2 > \dots > \tau_M > 0$  and  $\{R_n\}_{n=1}^M$  are orthogonal projections. We choose  $w_1$  so that  $0 \neq (w_1, R_1 w_1) \geq (w_k, R_1 w_k)$

for all  $w_k \in W$ . Although such  $w_1$  may not be unique, the following argument does not depend on the choice. If it happens that  $(w_k, R_1 w_k) = 0$  for all  $w_k \in W$ , we choose  $w_1$  so that  $0 \neq (w_1, R_2 w_1) \geq (w_k, R_2 w_k)$  for all  $w_k \in W$ . If still  $(w_k, R_2 w_k) = 0$  for all  $w_k \in W$ , we repeat the argument for  $R_3, R_4, \dots$ . For simplicity, we assume hereafter that  $(w_1, R_1 w_1) \neq 0$  holds. The general cases can be treated similarly. Let  $r_1 := R_1 w_1 / \|R_1 w_1\| \neq 0$ , then  $J$  is written as

$$J = \tau_1 |r_1 \times r_1| + \tau_1 (R_1 - |r_1 \times r_1|) + \sum_{n \geq 2} \tau_n E_n = \tau_1 \hat{R}_0 + \tau_1 \hat{R}_1 + \sum_{n \geq 2} \tau_n \hat{R}_n,$$

where we define  $\hat{R}_0 := |r_1 \times r_1|$ ,  $\hat{R}_1 := R_1 - |r_1 \times r_1|$  and  $\hat{R}_n := R_n$  for  $n \geq 2$  satisfying  $\hat{R}_a \hat{R}_b = \delta_{ab} \hat{R}_a$  for  $a, b = 0, 1, \dots, M$ . Now we consider a vector

$$\frac{J(w_1)}{\|J(w_1)\|} = \frac{\tau_1 \hat{R}_0 w_1 + \tau_1 \hat{R}_1 w_1 + \sum_{n \geq 2} \tau_n \hat{R}_n w_1}{\left( \tau_1^2 (w_1, \hat{R}_0 w_1) + \tau_1^2 (w_1, \hat{R}_1 w_1) + \sum_{n \geq 2} \tau_n^2 (w_1, \hat{R}_n w_1) \right)^{1/2}},$$

which must coincide with some  $w_p \in W$ . Its “ $\hat{R}_0$ -element” can be calculated as

$$\begin{aligned} & \left( \frac{J(w_1)}{\|J(w_1)\|}, \hat{R}_0 \frac{J(w_1)}{\|J(w_1)\|} \right) \\ &= \frac{\tau_1^2 (w_1, \hat{R}_0 w_1)}{\tau_1^2 (w_1, \hat{R}_0 w_1) + \tau_1^2 (w_1, \hat{R}_1 w_1) + \sum_{n \geq 2} \tau_n^2 (w_1, \hat{R}_n w_1)} \\ &= \frac{(w_1, \hat{R}_0 w_1)}{(w_1, \hat{R}_0 w_1) + (w_1, \hat{R}_1 w_1) + \sum_{n=2}^M \frac{\tau_n^2}{\tau_1^2} (w_1, \hat{R}_n w_1)}. \end{aligned} \quad (\text{B.1})$$

On the other hand, we can obtain that

$$\begin{aligned} & (w_1, \hat{R}_0 w_1) + (w_1, \hat{R}_1 w_1) + \sum_{n=2}^M \frac{\tau_n^2}{\tau_1^2} (w_1, \hat{R}_n w_1) \\ & < (w_1, \hat{R}_0 w_1) + (w_1, \hat{R}_1 w_1) + \sum_{n=2}^M (w_1, \hat{R}_n w_1) = 1 \end{aligned}$$

because there exists a  $n \geq 2$  such that  $(w_1, \hat{R}_n w_1) \neq 0$  (otherwise  $w_1 = (\hat{R}_0 + \hat{R}_1) w_1 = R_1 w_1$  and thus  $J(w_1) = \tau_1 w_1$  hold, which contradicts  $w_1 \in W$ ). Therefore, (B.1) results in

$$\left( \frac{J(w_1)}{\|J(w_1)\|}, \hat{R}_0 \frac{J(w_1)}{\|J(w_1)\|} \right) > (w_1, \hat{R}_0 w_1).$$

This observation concludes a contradiction to  $J(w_1) / \|J(w_1)\| = w_p \in W$



because  $w_1$  satisfies  $(w_1, \hat{R}_0 w_1) \geq (w_k, \hat{R}_0 w_k)$  for all  $w_k \in W$ . Overall, we find that every  $\omega^{\text{ext}} \in \Omega^{\text{ext}}$  has some  $\mu(\omega^{\text{ext}}) > 0$  such that  $J(\omega^{\text{ext}}) = \mu(\omega^{\text{ext}})\omega^{\text{ext}}$ .  $\square$

**Lemma B.4**

Let  $|\Omega^{\text{ext}}| < \infty$ , and suppose that linear maps  $J$  and  $K$  strictly positive with respect to an inner product  $(\cdot, \cdot)$  satisfy  $J(V_+) = K(V_+) = V_+^{*int}_{+(\cdot, \cdot)}$  (in particular,  $V_+$  is self-dual). Then, there exists a  $\mu(\omega^{\text{ext}}) > 0$  for each  $\omega^{\text{ext}} \in \Omega^{\text{ext}}$  such that  $K(\omega^{\text{ext}}) = \mu(\omega^{\text{ext}})J(\omega^{\text{ext}})$  holds.

**Proof**

As was seen in Lemma B.2, the inner products  $(\cdot, \cdot)_J := (\cdot, J\cdot)$  and  $(\cdot, \cdot)_K := (\cdot, K\cdot)$  satisfy  $V_+^{*int}_{+(\cdot, \cdot)_J} = V_+$  and  $V_+^{*int}_{+(\cdot, \cdot)_K} = V_+$  respectively. Because  $(\cdot, \cdot)_K$  is represented as  $(\cdot, \cdot)_K = (\cdot, L\cdot)_J$  with some linear map  $L$  strictly positive with respect to  $(\cdot, \cdot)_J$ , we have for arbitrary  $v, w \in V$

$$(v, w)_K = (v, Kw) = (v, Lw)_J = (v, JLv),$$

and thus  $L = J^{-1} \circ K$  holds. On the other hand,  $L$  satisfies

$$\begin{aligned} V_+^{*int}_{+(\cdot, \cdot)_K} &= \{v \mid (v, w)_K \geq 0, \forall w \in V_+\} \\ &= \{v \mid (v, Lw)_J \geq 0, \forall w \in V_+\} \\ &= \{v \mid (Lv, w)_J \geq 0, \forall w \in V_+\} = L^{-1}(V_+^{*int}_{+(\cdot, \cdot)_J}). \end{aligned}$$

That is,  $L(V_+) = V_+$  holds. Therefore, we can apply Lemma B.3 to  $L$ , and conclude that

$$L(\omega^{\text{ext}}) = \mu(\omega^{\text{ext}})\omega^{\text{ext}} = J^{-1}(K(\omega^{\text{ext}})),$$

i.e.  $K(\omega^{\text{ext}}) = \mu(\omega^{\text{ext}})J(\omega^{\text{ext}})$  holds.  $\square$

**Proposition 2.52**

Let  $\Omega$  be transitive with  $|\Omega^{\text{ext}}| < \infty$  and  $V_+$  be self-dual with respect to some inner product. There exists a linear bijection  $\Xi: V \rightarrow V$  such that  $\Omega' := \Xi\Omega$  is transitive and the generating positive cone  $V'_+$  is self-dual with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega')}$ , i.e.  $V'_+ = V_+^{*int}_{+\langle \cdot, \cdot \rangle_{GL(\Omega')}}$ .

**Proof**

Because of the transitivity of  $\Omega$ , we can adopt the orthogonal coordinate system of  $V$  introduced in Proposition 2.50. Since  $V_+$  is self-dual, there exists a linear map  $J: V \rightarrow V$  strictly positive with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$  such that  $J(V_+) = V_+^{*int}_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}$  (Lemma B.2). We can assume without loss of generality that  $J$  satisfies  $\langle \omega_M, J\omega_M \rangle_{GL(\Omega)} = 1$ . Let us introduce

$$\Omega^* := V_+^{*int}_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}} \cap [z = 1] = \{v \in V_+^{*int}_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}} \mid \langle v, \omega_M \rangle_{GL(\Omega)} = 1\},$$

where we identify the “ $\omega_M$ -coordinate” with “ $z$ -coordinate” in  $V$  and define  $[z = 1] := \{x \in V \mid \langle x, \omega_M \rangle_{GL(\Omega)} = 1\}$  (=  $\text{aff}(\Omega)$ ) (see Proposition 2.50). Note that since both  $V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int}$  and  $[z = 1]$  are  $GL(\Omega)$ -invariant,  $\Omega^*$  is also  $GL(\Omega)$ -invariant. It is easy to demonstrate that  $\Omega^*$  is convex (and compact), and we denote by  $\Omega^{*ext}$  the set of all extreme points of  $\Omega^*$ . We can also see that  $\Omega^{*ext}$  generates the extreme rays of  $V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int}$ . Because  $J$  satisfying  $J(V_+) = V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int}$  is bijective and maps extreme rays of  $V_+$  to extreme rays of  $V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int}$ , it holds that  $|\Omega^{*ext}| = |\Omega^{ext}|$ . Thus, there exists a bijection  $f: \Omega^{ext} \rightarrow \Omega^{*ext}$  and  $\beta(\omega^{ext}) > 0$  for each  $\omega^{ext} \in \Omega^{ext}$  satisfying  $J(\omega^{ext}) = \beta(\omega^{ext})f(\omega^{ext})$ .

For each  $T \in GL(\Omega)$ , we introduce  $J_T := T^{-1} \circ J \circ T$ . It is easy to see that  $J_T$  satisfies  $J_T(V_+) = V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int}$  by virtue of Lemma B.1. Furthermore,  $J_T$  is shown to be strictly positive with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$  because  $T \in GL(\Omega)$  is an orthogonal transformation with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$ . Therefore, applying Lemma B.4 to  $J$  and  $J_T$ , there exists  $\mu_T: \Omega^{ext} \rightarrow \mathbf{R}_{>0}$  such that  $J_T(\omega^{ext}) = \mu_T(\omega^{ext})J(\omega^{ext})$  for  $\omega^{ext} \in \Omega^{ext}$ , that is,

$$\begin{aligned} J_T(\omega^{ext}) &= \mu_T(\omega^{ext})J(\omega^{ext}) \\ &= \mu_T(\omega^{ext})\beta(\omega^{ext})f(\omega^{ext}) \\ &=: \beta_T(\omega^{ext})f(\omega^{ext}), \end{aligned}$$

where we define  $\beta_T(\omega^{ext}) := \mu_T(\omega^{ext})\beta(\omega^{ext})$ . We calculate this  $\beta_T(\omega^{ext})$ . It holds that

$$\begin{aligned} J_T(\omega^{ext}) &= T^{-1} \circ J(T\omega^{ext}) \\ &= T^{-1}(\beta(T\omega^{ext})f(T\omega^{ext})) \\ &= \beta(T\omega^{ext})T^{-1}f(T\omega^{ext}) \\ &= \beta_T(\omega^{ext})f(\omega^{ext}). \end{aligned}$$

This relation shows that  $T^{-1}f(T\omega^{ext})$  is proportional to  $f(\omega^{ext})$ . Considering that the  $z$ -coordinates of  $f(T\omega^{ext})$  and  $f(\omega^{ext})$  are 1 and that  $T^{-1}$  preserves  $z$ -coordinates, we find that  $T^{-1}f(T\omega^{ext}) = f(\omega^{ext})$  (equivalently,  $f(T\omega^{ext}) = Tf(\omega^{ext})$ ) holds. Consequently, we obtain

$$J_T(\omega^{ext}) = \beta(T\omega^{ext})f(\omega^{ext}).$$

Now we introduce

$$J_{av} := \frac{1}{|GL(\Omega)|} \sum_{T \in GL(\Omega)} J_T.$$

We note that  $|GL(\Omega)| < \infty$  when  $|\Omega^{ext}| < \infty$  because  $|GL(\Omega)| \leq |\Omega^{ext}|!$ .  $J_{av}$

acts on  $\omega^{\text{ext}} \in \Omega^{\text{ext}}$  as

$$J_{av}(\omega^{\text{ext}}) = \frac{1}{|GL(\Omega)|} \sum_{T \in GL(\Omega)} \beta(T\omega^{\text{ext}}) \cdot f(\omega^{\text{ext}}) =: Cf(\omega^{\text{ext}}),$$

where  $C := \frac{1}{|GL(\Omega)|} \sum_{T \in GL(\Omega)} \beta(T\omega^{\text{ext}})$  is a positive constant which does not depend on the choice of  $\omega^{\text{ext}} \in \Omega^{\text{ext}}$  because of the transitivity of  $\Omega$ . Thus, the map satisfies  $J_{av}(V_+) = V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{\text{int}}$  since  $J_{av}(\Omega^{\text{ext}}) = C\Omega^{\text{ext}}$ , and is strictly positive with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$  since it is a summation of the strictly positive operators  $\{J_T\}_{T \in GL(\Omega)}$ . Moreover, it satisfies

$$J_{av} \circ T = T \circ J_{av}.$$

for any  $T \in GL(\Omega)$ . We thus find that  $J_{av} \circ P_M = P_M \circ J_{av}$  holds for the orthogonal projection  $P_M$  introduced in Proposition A.1. In fact,

$$\begin{aligned} J_{av}(P_M x) &= \frac{1}{|GL(\Omega)|} J_{av} \left( \sum_{T \in GL(\Omega)} T x \right) \\ &= \frac{1}{|GL(\Omega)|} \sum_{T \in GL(\Omega)} T(J_{av} x) \\ &= P_M(J_{av} x) \end{aligned}$$

holds for all  $x \in V$ . Therefore,  $J_{av}$  is decomposed into two parts as

$$J_{av} = P_M \circ J_{av} \circ P_M + P_M^\perp \circ J_{av} \circ P_M^\perp, \quad (\text{B.2})$$

where  $P_M^\perp = \mathbb{1} - P_M$ . We note that  $V_M^\perp = \text{Im} P_M^\perp = [\text{aff}(\Omega) - \omega_M] = \mathbb{R}^N$  and  $\dim V_M = \dim \text{Im} P_M = 1$  hold by virtue of Proposition 2.50. Therefore, the first part of (B.2) is proportional to  $\mathbb{1}_{V_M} = \mathbb{1}_z = P_M$ , and because we set  $\langle \omega_M, J\omega_M \rangle_{GL(\Omega)} = 1$  and thus

$$\begin{aligned} \langle \omega_M, P_M \circ J_{av} \circ P_M \omega_M \rangle_{GL(\Omega)} &= \langle \omega_M, J_{av} \omega_M \rangle_{GL(\Omega)} \\ &= \langle \omega_M, P_M J \omega_M \rangle_{GL(\Omega)} \\ &= \langle \omega_M, J \omega_M \rangle_{GL(\Omega)} \\ &= 1 \\ &= \langle \omega_M, P_M \omega_M \rangle_{GL(\Omega)} \end{aligned}$$

holds, it is proven that

$$P_M \circ J_{av} \circ P_M = P_M.$$

Let us examine the second part. Suppose that there exists a nonzero  $x \in V_M^\perp$

such that  $Tx = x$  for all  $T \in GL(\Omega)$ . Then,  $P_M x = x \neq 0$  holds, and it contradicts to (A.4). Thus, we can find that  $GL(\Omega)$  acts irreducibly on  $V_M^\perp$ , that is, only  $\{0\}$  and  $V_M^\perp = \mathbb{R}^N$  itself are invariant subspaces. It concludes that  $P_M^\perp J_{av} P_M^\perp$ , which commutes with every element in  $GL(\Omega)$ , is proportional to  $\mathbb{1}_{V_M^\perp} = \mathbb{1}_{\mathbb{R}^N} = P_M^\perp$  due to Schur's lemma. Consequently, we obtain for some  $\xi > 0$

$$J_{av} = P_M + \xi P_M^\perp,$$

and thus

$$J_{av}(V_+) = (P_M + \xi P_M^\perp)(V_+) = V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int}. \quad (\text{B.3})$$

Let us introduce a linear bijection

$$\Xi := \sqrt{J_{av}} = P_M + \sqrt{\xi} P_M^\perp,$$

strictly positive with respect to  $\langle \cdot, \cdot \rangle_{GL(\Omega)}$ , and define  $\Omega' := \Xi \Omega$ . It is easy to check that the positive cone  $V'_+$  generated by  $\Omega'$  is given by  $V'_+ = \Xi V_+$ , and  $GL(\Omega') = \Xi GL(\Omega) \Xi^{-1} = GL(\Omega)$  (moreover, the unique maximally mixed state of  $\Omega'$  is still  $\omega_M$ ). In addition, we can find that

$$\begin{aligned} V_{+\langle \cdot, \cdot \rangle_{GL(\Omega')}}^{*int} &= \{v \mid \langle v, w' \rangle_{GL(\Omega)} \geq 0, \forall w' \in V'_+\} \\ &= \{v \mid \langle v, \Xi w \rangle_{GL(\Omega)} \geq 0, \forall w \in V_+\} \\ &= \Xi^{-1} V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int}. \end{aligned}$$

holds. Since (B.3) can be rewritten as

$$\Xi V_+ = \Xi^{-1} V_{+\langle \cdot, \cdot \rangle_{GL(\Omega)}}^{*int},$$

we can conclude

$$V'_+ = V_{+\langle \cdot, \cdot \rangle_{GL(\Omega')}}^{*int}.$$

□

*Remark*

In the case of  $|\Omega^{\text{ext}}| = \infty$ , there exists a counterexample of Lemma B.3. Let us consider a state space

$$\Omega = \{ {}^t(1, \mathbf{x}) = {}^t(1, x_1, x_2, x_3) \in \mathbb{R}^4 \mid |\mathbf{x}|^2 = x_1^2 + x_2^2 + x_3^2 \leq 1 \}$$

(the Bloch ball).  $\Omega$  defines a corresponding positive cone  $V_+$  as

$$V_+ = \{x \in \mathbb{R}^4 \mid x_0^2 - |\mathbf{x}|^2 \geq 0, x_0 \geq 0\},$$

which can be identified with a forward light cone of a Minkowski spacetime. We examine a pure Lorentz transformation  $\Lambda$  defined for  $\lambda \in \mathbb{R}$  as

$$\Lambda = \begin{bmatrix} \cosh \lambda & \sinh \lambda & 0 & 0 \\ \sinh \lambda & \cosh \lambda & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

It is easy to prove that this  $\Lambda$  is strictly positive. Since the pure Lorentz transformation preserves the Minkowski metric, it satisfies  $\Lambda(V_+) = V_+$ . However,  $\Lambda$  transforms an extreme point  $x = {}^t(1, 0, 1, 0)$  to

$$\Lambda(x) = {}^t(\cosh \lambda, \sinh \lambda, 1, 0),$$

which is not proportional to  $x$ . Investigating whether Proposition 2.52 still holds when  $|\Omega^{\text{ext}}| = \infty$  is a future problem.

# Bibliography

- [1] R. Takakura and T. Miyadera, “Preparation uncertainty implies measurement uncertainty in a class of generalized probabilistic theories,” *Journal of Mathematical Physics*, vol. 61, no. 8, p. 082203, 2020.
- [2] R. Takakura and T. Miyadera, “Entropic uncertainty relations in a class of generalized probabilistic theories,” *Journal of Physics A: Mathematical and Theoretical*, vol. 54, p. 315302, July 2021.
- [3] T. Heinosaari, T. Miyadera, and R. Takakura, “Testing incompatibility of quantum devices with few states,” *Physical Review A*, vol. 104, p. 032228, Sept. 2021.
- [4] R. Takakura, “Entropy of mixing exists only for classical and quantum-like theories among the regular polygon theories,” *Journal of Physics A: Mathematical and Theoretical*, vol. 52, p. 465302, Oct. 2019.
- [5] J. Von Neumann, *Mathematical Foundations of Quantum Mechanics*. Princeton: Princeton University Press, 1955.
- [6] L. Hardy and R. Spekkens, “Why physics needs quantum foundations,” *Physics in Canada*, vol. 66, no. 2, pp. 73–76, 2010.
- [7] G. Chiribella and R. W. Spekkens, eds., *Quantum Theory: Informational Foundations and Foils*. No. 181 in *Fundamental Theories of Physics*, Springer, Dordrecht, 1st ed., 2016.
- [8] W. Heisenberg, “Über den anschaulichen inhalt der quantentheoretischen Kinematik und Mechanik,” *Zeitschrift für Physik*, vol. 43, pp. 172–198, Mar. 1927.
- [9] A. Einstein, B. Podolsky, and N. Rosen, “Can quantum-mechanical description of physical reality be considered complete?,” *Physical Review*, vol. 47, pp. 777–780, May 1935.
- [10] J. S. Bell, “On the einstein podolsky rosen paradox,” *Physics Physique Fizika*, vol. 1, pp. 195–200, Nov. 1964.

- [11] C. H. Bennett and G. Brassard, “Quantum cryptography: Public key distribution and coin tossing,” in *International Conference on Computers, Systems & Signal Processing*, pp. 175–179, 1984.
- [12] A. K. Ekert, “Quantum cryptography based on bell’s theorem,” *Physical Review Letters*, vol. 67, pp. 661–663, Aug. 1991.
- [13] P. J. Lahti and S. Bugajski, “Fundamental principles of quantum theory. ii. from a convexity scheme to the dhb theory,” *International Journal of Theoretical Physics*, vol. 24, pp. 1051–1080, 1985.
- [14] M. Horodecki, P. Horodecki, and R. Horodecki, “Separability of mixed states: necessary and sufficient conditions,” *Physics Letters A*, vol. 223, no. 1, pp. 1–8, 1996.
- [15] B. M. Terhal, “Bell inequalities and the separability criterion,” *Physics Letters A*, vol. 271, no. 5, pp. 319–326, 2000.
- [16] C. Carmeli, T. Heinosaari, T. Miyadera, and A. Toigo, “Witnessing incompatibility of quantum channels,” *Journal of Mathematical Physics*, vol. 60, no. 12, p. 122202, 2019.
- [17] C. Carmeli, T. Heinosaari, and A. Toigo, “Quantum incompatibility witnesses,” *Physical Review Letters*, vol. 122, p. 130402, Apr. 2019.
- [18] T. Miyadera, “Uncertainty relations for joint localizability and joint measurability in finite-dimensional systems,” *Journal of Mathematical Physics*, vol. 52, no. 7, p. 072105, 2011.
- [19] F. Buscemi, M. J. W. Hall, M. Ozawa, and M. M. Wilde, “Noise and disturbance in quantum measurements: An information-theoretic approach,” *Physical Review Letters*, vol. 112, p. 050401, Feb. 2014.
- [20] W. K. Wootters and W. H. Zurek, “A single quantum cannot be cloned,” *Nature*, vol. 299, pp. 802–803, 1982.
- [21] T. Heinosaari, T. Miyadera, and M. Ziman, “An invitation to quantum incompatibility,” *Journal of Physics A: Mathematical and Theoretical*, vol. 49, p. 123001, Feb. 2016.
- [22] L. Hardy, “Quantum theory from five reasonable axioms,” 2001, arXiv:quant-ph/0101012.
- [23] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, “Cloning and broadcasting in generic probabilistic theories,” 2006, arXiv:quant-ph/0101012.

- [24] J. Barrett, “Information processing in generalized probabilistic theories,” *Physical Review A*, vol. 75, p. 032304, Mar. 2007.
- [25] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, “Generalized no-broadcasting theorem,” *Physical Review Letters*, vol. 99, p. 240501, Dec. 2007.
- [26] G. Chiribella, G. M. D’Ariano, and P. Perinotti, “Probabilistic theories with purification,” *Physical Review A*, vol. 81, p. 062348, June 2010.
- [27] G. Chiribella, G. M. D’Ariano, and P. Perinotti, “Informational derivation of quantum theory,” *Physical Review A*, vol. 84, p. 012311, July 2011.
- [28] L. Masanes and M. P. Müller, “A derivation of quantum theory from physical requirements,” *New Journal of Physics*, vol. 13, p. 063001, June 2011.
- [29] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, “Teleportation in general probabilistic theories,” in *Proceedings of Symposia in Applied Mathematics*, vol. 71, pp. 25–48, 2012.
- [30] L. Lami, *Non-classical correlations in quantum mechanics and beyond*. PhD thesis, Universitat Autònoma de Barcelona, 2017.
- [31] M. Plávala, “General probabilistic theories: An introduction,” 2021, arXiv:2103.07469 [quant-ph].
- [32] G. W. Mackey, *The mathematical foundations of quantum mechanics: a lecture-note volume*. New York ; Amsterdam: W.A. Benjamin, Inc., 1963.
- [33] G. Ludwig, “Versuch einer axiomatischen grundlegung der quantenmechanik und allgemeinerer physikalischer theorien,” *Zeitschrift für Physik*, vol. 181, pp. 233–260, 1964.
- [34] G. Ludwig, “Attempt of an axiomatic foundation of quantum mechanics and more general theories, ii,” *Communications in Mathematical Physics*, vol. 4, pp. 331–348, 1967.
- [35] E. B. Davies and J. T. Lewis, “An operational approach to quantum probability,” *Communications in Mathematical Physics*, vol. 17, pp. 239–260, Sept. 1970.
- [36] S. Gudder, “Convex structures and operational quantum mechanics,” *Communications in Mathematical Physics*, vol. 29, no. 3, pp. 249 – 264, 1973.



- [37] A. Hartkämper and H. Neumann, eds., *Foundations of Quantum Mechanics and Ordered Linear Spaces: Advanced Study Institute held in Marburg 1973*, vol. 29 of *Lecture Notes in Physics*. Springer-Verlag Berlin Heidelberg, 1st ed., 1974.
- [38] G. Cassinelli and P. Lahti, “An axiomatic basis for quantum mechanics,” *Foundations of Physics*, vol. 46, pp. 1341–1373, 2016.
- [39] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory*, vol. 190 of *Lecture Notes in Physics*. Springer-Verlag Berlin Heidelberg, 1st ed., 1983.
- [40] H. Araki, *Mathematical Theory of Quantum Fields*. International Series of Monographs on Physics, Oxford: Oxford University Press, 1999.
- [41] P. Busch, P. J. Lahti, J.-P. Pellonpää, and K. Ylänen, *Quantum Measurement*. Theoretical and Mathematical Physics, Springer International Publishing, 2016.
- [42] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, vol. 1 of *North-Holland series in statistics and probability*. Amsterdam ; New York: North-Holland Pub. Co., 1982.
- [43] S. P. Gudder, *Stochastic Methods in Quantum Mechanics*. New York: Dover, 1979.
- [44] Y. Kuramochi, “Compact convex structure of measurements and its applications to simulability, incompatibility, and convex resource theory of continuous-outcome measurements,” 2020, arXiv:2002.03504 [math.FA].
- [45] G. Kimura, K. Nuida, and H. Imai, “Physical equivalence of pure states and derivation of qubit in general probabilistic theories,” 2010, arXiv:1012.5361 [quant-ph].
- [46] R. W. Spekkens, “Contextuality for preparations, transformations, and unsharp measurements,” *Physical Review A*, vol. 71, p. 052108, May 2005.
- [47] S. N. Filippov, S. Gudder, T. Heinosaari, and L. Leppäjärvi, “Operational restrictions in general probabilistic theories,” *Foundations of Physics*, vol. 50, p. 850–876, 2020.
- [48] P. Janotta and R. Lal, “Generalized probabilistic theories without the no-restriction hypothesis,” *Physical Review A*, vol. 87, p. 052131, May 2013.

- [49] R. T. Rockafellar, *Convex Analysis*, vol. 28 of *Princeton Mathematical Series*. Princeton University Press, 1970.
- [50] H. H. Schaefer, *Topological Vector Spaces*, vol. 3 of *Graduate Texts in Mathematics*. Springer-Verlag New York, 2nd ed., 1999.
- [51] H. Barnum, M. P. Müller, and C. Ududec, “Higher-order interference and single-system postulates characterizing quantum theory,” *New Journal of Physics*, vol. 16, p. 123029, Dec. 2014.
- [52] A. B. Sainz, Y. Guryanova, A. Acín, and M. Navascués, “Almost-quantum correlations violate the no-restriction hypothesis,” *Physical Review Letters*, vol. 120, p. 200402, May 2018.
- [53] N. Bourbaki, *Topological Vector Spaces*. Elements of Mathematics, Springer-Verlag Berlin Heidelberg, 2003. Original French edition published by Masson, Paris, France, 1981.
- [54] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [55] E. M. Alfsen, *Compact Convex Sets and Boundary Integrals*, vol. 57 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer-Verlag Berlin Heidelberg, 1971.
- [56] A. J. Ellis, “The Duality of Partially Ordered Normed Linear Spaces,” *Journal of the London Mathematical Society*, vol. s1-39, no. 1, pp. 730–744, 1964.
- [57] J. L. Kelley and I. Namioka, *Linear Topological Spaces*. Springer-Verlag Berlin Heidelberg, 2nd ed., 1963.
- [58] J. B. Conway, *A Course in Functional Analysis*, vol. 96 of *Graduate Texts in Mathematics*. Springer-Verlag New York, 1st ed., 1985.
- [59] O. Güler, *Foundations of Optimization*, vol. 258 of *Graduate Texts in Mathematics*. Springer New York, 1st ed., 2010.
- [60] C. D. Aliprantis and R. Tourky, *Cones and Duality*, vol. 84 of *Graduate Studies in Mathematics*. American Mathematical Society, 2007.
- [61] G. Kimura, K. Nuida, and H. Imai, “Distinguishability measures and entropies for general probabilistic theories,” *Reports on Mathematical Physics*, vol. 66, no. 2, pp. 175 – 206, 2010.
- [62] J. L. Kelley, *General Topology*, vol. 27 of *Graduate Texts in Mathematics*. Springer-Verlag New York, 1st ed., 1975.

- [63] O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics 1: C\*- and W\*-Algebras Symmetry Groups Decomposition of States*. Texts and Monographs in Physics, Springer-Verlag Berlin Heidelberg, 2nd ed., 1987.
- [64] E. G. Beltrametti and S. Bugajski, “Effect algebras and statistical physical theories,” *Journal of Mathematical Physics*, vol. 38, no. 6, pp. 3020–3030, 1997.
- [65] S. Gudder and S. Pulmannová, “Representation theorem for convex effect algebras,” *Commentationes Mathematicae Universitatis Carolinae*, vol. 39, no. 4, pp. 645–659, 1998.
- [66] S. Gudder, “Convex structures and effect algebras,” *International Journal of Theoretical Physics*, vol. 38, pp. 3179–3187, 1999.
- [67] E. M. Alfsen and F. W. Shultz, *Geometry of State Spaces of Operator Algebras*. Mathematics: Theory & Applications, Boston: Birkhäuser, 1st ed., 2003.
- [68] S. Roman, *Advanced Linear Algebra*, vol. 135 of *Graduate Texts in Mathematics*. Springer-Verlag New York, 1st ed., 1992.
- [69] S. Popescu and D. Rohrlich, “Quantum nonlocality as an axiom,” *Foundations of Physics*, vol. 24, pp. 379–385, 1994.
- [70] H. Barnum and A. Wilce, “Post-classical probability theory,” 2012, arXiv:1205.3833 [quant-ph].
- [71] L. Hardy, “Reformulating and reconstructing quantum theory,” 2011, arXiv:1104.2066 [quant-ph].
- [72] A. Hulanicki and R. Phelps, “Some applications of tensor products of partially-ordered linear spaces,” *Journal of Functional Analysis*, vol. 2, no. 2, pp. 177–201, 1968.
- [73] A. L. Peressini and D. R. Sherbert, “Ordered topological tensor products,” *Proceedings of the London Mathematical Society*, vol. s3-19, no. 1, pp. 177–190, 1969,
- [74] G. Aubrun, L. Lami, C. Palazuelos, and M. Plávala, “Entangleability of cones,” *Geometric and Functional Analysis*, vol. 31, pp. 181–205, May 2021.
- [75] H. Araki, “On a characterization of the state space of quantum mechanics,” *Communications in Mathematical Physics*, vol. 75, pp. 1–24, 1980.

- [76] T. Heinosaari and M. Ziman, *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement*. Cambridge: Cambridge University Press, 2011.
- [77] T. Heinosaari, L. Leppäjärvi, and M. Plávala, “No-free-information principle in general probabilistic theories,” *Quantum*, vol. 3, p. 157, July 2019.
- [78] R. Haag and D. Kastler, “An algebraic approach to quantum field theory,” *Journal of Mathematical Physics*, vol. 5, no. 7, pp. 848–861, 1964.
- [79] K. Kraus, “General state changes in quantum theory,” *Annals of Physics*, vol. 64, no. 2, pp. 311–335, 1971.
- [80] A. K. G. K. T. O. Masahito Hayashi, Satoshi Ishizaka, *Introduction to Quantum Information Science*. Graduate Texts in Physics, Springer-Verlag Berlin Heidelberg, 1st ed., 2015.
- [81] M. Plávala, “All measurements in a probabilistic theory are compatible if and only if the state space is a simplex,” *Physical Review A*, vol. 94, p. 042108, Oct. 2016.
- [82] E. B. Davies, “Symmetries of compact convex sets,” *The Quarterly Journal of Mathematics*, vol. 25, pp. 323–328, Jan. 1974.
- [83] 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*, vol. 316, pp. 441–487, 2012.
- [84] M. P. Müller and C. Ududec, “Structure of reversible computation determines the self-duality of quantum theory,” *Physical Review Letters*, vol. 108, p. 130401, Mar. 2012.
- [85] G. Kimura, “The bloch vector for n-level systems,” *Physics Letters A*, vol. 314, no. 5, pp. 339–349, 2003.
- [86] I. Bengtsson, S. Weis, and K. Życzkowski, *Geometry of the Set of Mixed Quantum States: An Apophatic Approach*, pp. 175–197. Basel: Springer Basel, 2013.
- [87] P. Janotta, C. Gogolin, J. Barrett, and N. Brunner, “Limits on non-local correlations from the structure of the local state space,” *New Journal of Physics*, vol. 13, no. 6, p. 063024, 2011.

- [88] D. S. Dummit and R. M. Foote, *Abstract Algebra*. Hoboken, New Jersey: John Wiley & Sons, Inc., 3rd ed., 2003.
- [89] A. Jenčová and M. Plávala, “Conditions on the existence of maximally incompatible two-outcome measurements in general probabilistic theory,” *Physical Review A*, vol. 96, p. 022113, Aug. 2017.
- [90] M. Koashi, “Unconditional security of quantum key distribution and the uncertainty principle,” *Journal of Physics: Conference Series*, vol. 36, pp. 98–102, Apr. 2006.
- [91] H. P. Robertson, “The uncertainty principle,” *Physical Review*, vol. 34, pp. 163–164, July 1929.
- [92] J. B. M. Uffink, *Measures of uncertainty and the uncertainty principle*. PhD thesis, University of Utrecht, Utrecht, 1990.
- [93] J. I. de Vicente and J. Sánchez-Ruiz, “Separability conditions from the Landau-Pollak uncertainty relation,” *Physical Review A*, vol. 71, p. 052325, May 2005.
- [94] T. Miyadera and H. Imai, “Generalized Landau-Pollak uncertainty relation,” *Physical Review A*, vol. 76, p. 062108, Dec. 2007.
- [95] I. I. Hirschman, “A note on entropy,” *American Journal of Mathematics*, vol. 79, no. 1, pp. 152–156, 1957.
- [96] W. Beckner, “Inequalities in fourier analysis,” *Annals of Mathematics*, vol. 102, no. 1, pp. 159–182, 1975.
- [97] I. Białyński-Birula and J. Mycielski, “Uncertainty relations for information entropy in wave mechanics,” *Communications in Mathematical Physics*, vol. 44, pp. 129–132, June 1975.
- [98] D. Deutsch, “Uncertainty in quantum measurements,” *Physical Review Letters*, vol. 50, pp. 631–633, Feb. 1983.
- [99] H. Maassen and J. B. M. Uffink, “Generalized entropic uncertainty relations,” *Physical Review Letters*, vol. 60, pp. 1103–1106, Mar. 1988.
- [100] M. Krishna and K. R. Parthasarathy, “An entropic uncertainty principle for quantum measurements,” *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, vol. 64, no. 3, pp. 842–851, 2002.
- [101] E. Arthurs and M. S. Goodman, “Quantum correlations: A generalized Heisenberg uncertainty relation,” *Physical Review Letters*, vol. 60, pp. 2447–2449, June 1988.

- [102] E. Arthurs and J. L. Kelly Jr., “On the simultaneous measurement of a pair of conjugate observables,” *Bell System Technical Journal*, vol. 44, no. 4, pp. 725–729, 1965,
- [103] M. Ozawa, “Universally valid reformulation of the Heisenberg uncertainty principle on noise and disturbance in measurement,” *Physical Review A*, vol. 67, p. 042105, Apr. 2003.
- [104] P. Busch and D. B. Pearson, “Universal joint-measurement uncertainty relation for error bars,” *Journal of Mathematical Physics*, vol. 48, no. 8, p. 082103, 2007,
- [105] T. Miyadera and H. Imai, “Heisenberg’s uncertainty principle for simultaneous measurement of positive-operator-valued measures,” *Physical Review A*, vol. 78, p. 052119, Nov. 2008.
- [106] P. Busch and T. Heinosaari, “Approximate joint measurements of qubit observables,” *Quantum Information & Computation*, vol. 8, p. 797–818, Sept. 2008.
- [107] R. F. Werner, “The uncertainty relation for joint measurement of position and momentum,” *Quantum Info. Comput.*, vol. 4, p. 546–562, Dec. 2004.
- [108] D. Saha, M. Oszmaniec, L. Czekaj, M. Horodecki, and R. Horodecki, “Operational foundations for complementarity and uncertainty relations,” *Physical Review A*, vol. 101, p. 052104, May 2020.
- [109] M. M. Wolf, D. Perez-Garcia, and C. Fernandez, “Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory,” *Physical Review Letters*, vol. 103, p. 230402, Dec. 2009.
- [110] P. Busch, T. Heinosaari, J. Schultz, and N. Stevens, “Comparing the degrees of incompatibility inherent in probabilistic physical theories,” *Europhysics Letters*, vol. 103, p. 10002, July 2013.
- [111] N. Stevens and P. Busch, “Steering, incompatibility, and Bell-inequality violations in a class of probabilistic theories,” *Physical Review A*, vol. 89, p. 022123, Feb. 2014.
- [112] A. Jenčová, “Incompatible measurements in a class of general probabilistic theories,” *Physical Review A*, vol. 98, p. 012133, July 2018.
- [113] M. Banik, M. R. Gazi, S. Ghosh, and G. Kar, “Degree of complementarity determines the nonlocality in quantum mechanics,” *Physical Review A*, vol. 87, p. 052125, May 2013.

- [114] M. Krumm, H. Barnum, J. Barrett, and M. P. Müller, “Thermodynamics and the structure of quantum theory,” *New Journal of Physics*, vol. 19, no. 4, p. 043025, 2017.
- [115] S. N. Filippov, T. Heinosaari, and L. Leppäjärvi, “Simulability of observables in general probabilistic theories,” *Physical Review A*, vol. 97, p. 062102, June 2018.
- [116] P. J. Coles, M. Berta, M. Tomamichel, and S. Wehner, “Entropic uncertainty relations and their applications,” *Reviews of Modern Physics*, vol. 89, p. 015002, Feb. 2017.
- [117] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*. Cambridge: Cambridge University Press, 2nd ed., 1988.
- [118] M. H. Partovi, “Majorization formulation of uncertainty in quantum mechanics,” *Physical Review A*, vol. 84, p. 052117, Nov. 2011.
- [119] S. Friedland, V. Gheorghiu, and G. Gour, “Universal uncertainty relations,” *Physical Review Letters*, vol. 111, p. 230401, Dec. 2013.
- [120] L. Rudnicki, Z. Puchała, and K. Życzkowski, “Strong majorization entropic uncertainty relations,” *Physical Review A*, vol. 89, p. 052115, May 2014.
- [121] Z. Puchała, L. Rudnicki, and K. Życzkowski, “Majorization entropic uncertainty relations,” *Journal of Physics A: Mathematical and Theoretical*, vol. 46, p. 272002, June 2013.
- [122] Z. Puchała, L. Rudnicki, A. Krawiec, and K. Życzkowski, “Majorization uncertainty relations for mixed quantum states,” *Journal of Physics A: Mathematical and Theoretical*, vol. 51, p. 175306, Apr. 2018.
- [123] K. Baek, H. Nha, and W. Son, “Entropic uncertainty relations via direct-sum majorization relation for generalized measurements,” *Entropy*, vol. 21, no. 3, 2019.
- [124] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2nd ed., 2006.
- [125] M. Srinivas, “Optimal entropic uncertainty relation for successive measurements in quantum information theory,” *Pramana*, vol. 60, pp. 1137–1152, 2003.

- [126] K. Baek, T. Farrow, and W. Son, “Optimized entropic uncertainty for successive projective measurements,” *Physical Review A*, vol. 89, p. 032108, Mar. 2014.
- [127] J. Zhang, Y. Zhang, and C.-s. Yu, “Rényi entropy uncertainty relation for successive projective measurements,” *Quantum Information Processing*, vol. 14, pp. 2239–2253, 2015.
- [128] A. E. Rastegin, “Entropic uncertainty relations for successive measurements of canonically conjugate observables,” *Annalen der Physik*, vol. 528, no. 11-12, pp. 835–844, 2016.
- [129] S. Designolle, M. Farkas, and J. Kaniewski, “Incompatibility robustness of quantum measurements: a unified framework,” *New Journal of Physics*, vol. 21, p. 113053, Nov. 2019.
- [130] T. Heinosaari, J. Schultz, A. Toigo, and M. Ziman, “Maximally incompatible quantum observables,” *Physics Letters A*, vol. 378, no. 24, pp. 1695–1699, 2014.
- [131] T. Heinosaari, J. Kiukas, D. Reitzner, and J. Schultz, “Incompatibility breaking quantum channels,” *Journal of Physics A: Mathematical and Theoretical*, vol. 48, p. 435301, Oct. 2015.
- [132] L. Guerini, M. T. Quintino, and L. Aolita, “Distributed sampling, quantum communication witnesses, and measurement incompatibility,” *Physical Review A*, vol. 100, p. 042308, Oct. 2019.
- [133] J. Kiukas, “Subspace constraints for joint measurability,” *Journal of Physics: Conference Series*, vol. 1638, p. 012003, Oct. 2020.
- [134] F. Loulidi and I. Nechita, “The compatibility dimension of quantum measurements,” *Journal of Mathematical Physics*, vol. 62, no. 4, p. 042205, 2021,
- [135] R. Uola, T. Kraft, S. Designolle, N. Miklin, A. Tavakoli, J.-P. Pellonpää, O. Gühne, and N. Brunner, “Quantum measurement incompatibility in subspaces,” *Physical Review A*, vol. 103, p. 022203, Feb. 2021.
- [136] P. Busch, “Unsharp reality and joint measurements for spin observables,” *Physical Review D*, vol. 33, pp. 2253–2261, Apr. 1986.
- [137] T. Heinosaari, T. Miyadera, and D. Reitzner, “Strongly incompatible quantum devices,” *Foundations of Physics*, vol. 44, pp. 34–57, 2014.



- [138] E. Haapasalo, “Robustness of incompatibility for quantum devices,” *Journal of Physics A: Mathematical and Theoretical*, vol. 48, p. 255303, June 2015.
- [139] T. Heinosaari and T. Miyadera, “Incompatibility of quantum channels,” *Journal of Physics A: Mathematical and Theoretical*, vol. 50, p. 135302, Mar. 2017.
- [140] T. Heinosaari, D. Reitzner, T. c. v. Rybár, and M. Ziman, “Incompatibility of unbiased qubit observables and pauli channels,” *Physical Review A*, vol. 97, p. 022112, Feb. 2018.
- [141] Y. Kuramochi, “Quantum incompatibility of channels with general outcome operator algebras,” *Journal of Mathematical Physics*, vol. 59, no. 4, p. 042203, 2018,
- [142] E. Haapasalo, “Compatibility of covariant quantum channels with emphasis on weyl symmetry,” *Annales Henri Poincaré*, vol. 20, p. 3163–3195, 2019.
- [143] H. Martens and W. M. de Muynck, “Nonideal quantum measurements,” *Foundations of Physics*, vol. 20, pp. 255–281, 1990.
- [144] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, “Noncommuting mixed states cannot be broadcast,” *Physical Review Letters*, vol. 76, pp. 2818–2821, Apr. 1996.
- [145] R. Beneduci, T. J. Bullock, P. Busch, C. Carmeli, T. Heinosaari, and A. Toigo, “Operational link between mutually unbiased bases and symmetric informationally complete positive operator-valued measures,” *Physical Review A*, vol. 88, p. 032312, Sept. 2013.
- [146] P. Stano, D. Reitzner, and T. Heinosaari, “Coexistence of qubit effects,” *Physical Review A*, vol. 78, p. 012315, July 2008.
- [147] P. Busch and H.-J. Schmidt, “Coexistence of qubit effects,” *Quantum Information Processing*, vol. 9, p. 143–169, 2010.
- [148] S. Yu, N.-l. Liu, L. Li, and C. H. Oh, “Joint measurement of two unsharp observables of a qubit,” *Physical Review A*, vol. 81, p. 062116, June 2010.
- [149] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2nd ed., 1985.
- [150] M. W. Zemansky and R. H. Dittman, *Heat and Thermodynamics*. New York: McGraw-Hill Companies, Inc., 7th ed., 1997.

- [151] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke, “Entropy and information causality in general probabilistic theories,” *New Journal of Physics*, vol. 12, no. 3, p. 033024, 2010.
- [152] A. J. Short and S. Wehner, “Entropy in general physical theories,” *New Journal of Physics*, vol. 12, p. 033023, Mar. 2010.
- [153] G. Kimura, J. Ishiguro, and M. Fukui, “Entropies in general probabilistic theories and their application to the holevo bound,” *Physical Review A*, vol. 94, p. 042113, Oct. 2016.
- [154] G. Chiribella and C. M. Scandolo, “Microcanonical thermodynamics in general physical theories,” *New Journal of Physics*, vol. 19, p. 123043, Dec. 2017.
- [155] G. Chiribella and C. M. Scandolo, “Entanglement as an axiomatic foundation for statistical mechanics,” 2016, arxiv:1608.04459 [quant-ph].
- [156] H. Barnum, J. Barrett, M. Krumm, and M. P. Müller, “Entropy, majorization and thermodynamics in general probabilistic theories,” in Proceedings of the 12th International Workshop on *Quantum Physics and Logic*, Oxford, U.K., July 15-17, 2015, vol. 195, pp. 43–58, Open Publishing Association, 2015.
- [157] G. Kimura and K. Nuida, “On affine maps on non-compact convex sets and some characterizations of finite-dimensional solid ellipsoids,” *Journal of Geometry and Physics*, vol. 86, pp. 1–18, 2014.