



UNIVERSIDADE ESTADUAL DE CAMPINAS  
Instituto de Física Gleb Wataghin

GABRIELA RUIZ

**CERTIFICAÇÃO DE ESTADOS EMARANHADOS LOCAIS**

**LOCAL ENTANGLED STATES CERTIFICATION**

CAMPINAS  
2023

GABRIELA RUIZ

**CERTIFICAÇÃO DE ESTADOS EMARANHADOS LOCAIS**

**LOCAL ENTANGLED STATES CERTIFICATION**

Tese apresentada ao Instituto de Física Gleb Wataghin da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de MESTRA EM FÍSICA, na Área de FÍSICA.

Thesis presented to the Gleb Wataghin Institute of Physics of the University of Campinas in partial fulfillment of the requirements for the degree of MASTER OF PHYSICS, in the area of PHYSICS.

Orientador: RAFAEL LUIZ DA SILVA RABELO

ESTE TRABALHO CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELA ALUNA GABRIELA RUIZ E ORIENTADA PELO PROF. DR. RAFAEL LUIZ DA SILVA RABELO.

CAMPINAS

2023

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca do Instituto de Física Gleb Wataghin  
Lucimeire de Oliveira Silva da Rocha - CRB 8/9174

R859L Ruiz, Gabriela, 1999-  
Local entangled states certification / Gabriela Ruiz. – Campinas, SP : [s.n.], 2023.

Orientador: Rafael Luiz da Silva Rabelo.  
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Física Gleb Wataghin.

1. Emaranhamento. 2. Steering. 3. Não-localidade de Bell. I. Rabelo, Rafael Luiz da Silva, 1986-. II. Universidade Estadual de Campinas. Instituto de Física Gleb Wataghin. III. Título.

Informações Complementares

**Título em outro idioma:** Certificação de estados emaranhados locais

**Palavras-chave em inglês:**

Entanglement

Steering

Bell nonlocality

**Área de concentração:** Física

**Titulação:** Mestra em Física

**Banca examinadora:**

Rafael Luiz da Silva Rabelo [Orientador]

Marcelo de Oliveira Terra Cunha

Leonardo Guerini de Souza

**Data de defesa:** 04-08-2023

**Programa de Pós-Graduação:** Física

**Identificação e informações acadêmicas do(a) aluno(a)**

- ORCID do autor: <https://orcid.org/0000-0003-3602-2570>

- Currículo Lattes do autor: <http://lattes.cnpq.br/1102310524120714>



INSTITUTO DE FÍSICA  
GLEB WATAGHIN

MEMBROS DA COMISSÃO EXAMINADORA DA DISSERTAÇÃO DE MESTRADO DA ALUNA GABRIELA RUIZ - RA 220525, APRESENTADA E APROVADA AO INSTITUTO DE FÍSICA GLEB WATAGHIN, DA UNIVERSIDADE ESTADUAL DE CAMPINAS, EM 04/08/2023.

COMISSÃO JULGADORA:

- Prof. Dr. Rafael Luiz da Silva Rabelo – Presidente e orientador (IFGW/UNICAMP)
- Prof. Dr. Marcelo de Oliveira Terra Cunha (IMECC/UNICAMP)
- Prof. Dr. Leonardo Guerini de Souza (Departamento de Matemática da Universidade Federal de Santa Maria)

**OBS.:** Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

CAMPINAS

2023

# AGRADECIMENTOS

Agradeço ao meu orientador, Rafael, por me apresentar a vida como ela é e me permitir escolher como seguir nela. Agradeço pelas frequentes validações e por se esforçar imensamente para que nós tenhamos a erudição e a excelência as quais considera nosso direito.

Agradeço a todos que conheci no Grupo de Fundamentos Matemáticos da Teoria Quântica (MFQ), por construírem um ambiente de intensa pesquisa, priorizando consistentemente a produção de trabalhos de alta qualidade. Especialmente, agradeço ao Prof. Dr. Marcelo Terra Cunha. Sua abordagem exigente, mas encorajadora, foi fundamental para moldar não apenas os projetos que realizamos, mas também a minha compreensão geral da matemática e da física.

Sou grata aos professores com os quais tive aulas, ao IFGW pela infraestrutura oferecida e à UNICAMP pelo incrível ambiente interdisciplinar que ampliou meus horizontes e me permitiu florescer profissionalmente e pessoalmente. Agradeço às agências de fomento, ao Ministério da Ciência, Tecnologia e Inovações e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (processo: 161759/2021-1), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES (processo: 88887.640775/2021-00), e principalmente à Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP (processo: 21/13798-7), pelo financiamento deste projeto.

Agradeço aos amigos que, de longa ou curta data, compartilharam esses anos comigo e foram uma fonte inestimável de apoio, encorajamento e motivação. Em especial Naruna, Jerê, Ciça, Moysés, Pedro (o aranha), Pedro (o da maritaca), Berg, Jacinto, Igor, Zé, Kaleb, Carlos, Arthur.

Agradeço aos meus pais, à minha irmã e à minha vó, sem ter nem como listar aqui tudo pelo que sou grata.

À Tandy e ao Vitor, agradeço antes de tudo por estarem presentes sempre, independente de qualquer coisa. E se isso já não fosse o suficiente, agradeço por enriquecerem minha vida.

Ao Filipe, agradeço por tudo, absolutamente tudo.

# RESUMO

Emaranhamento, não-localidade de Bell e *Steering*, anteriormente considerados sinônimos, são conceitos não clássicos distintos. Embora o emaranhamento seja necessário, não é uma condição suficiente para que as correlações exibidas por um sistema sejam não locais e/ou *steerable* e, portanto, exibam um comportamento mais forte que o clássico. A relação exata entre esses três fenômenos é desconhecida. O problema de verificar se um estado quântico geral pode gerar estatísticas não locais é notoriamente difícil. Há pouca esperança de se obter uma ferramenta analítica sistemática para responder a esta questão. Portanto, métodos numéricos têm sido desenvolvidos para a solução parcial de tais problemas. Neste projeto, a aluna utilizou um método que conta com um subconjunto de otimização convexa, chamado de programação semidefinida, que foi apresentado em um artigo de 2016 que inclui o orientador da aluna como um dos autores. Com este novo método otimizado e generalizado, fomos capazes de investigar e certificar a localidade de uma ampla gama de estados quânticos emaranhados e obtivemos limites aprimorados no volume de estados de dois qubits que são emaranhados e locais. No entanto, este método tem limitações e funciona melhor com dimensões baixas e poucas medições. Com base no sucesso das técnicas de Machine Learning em problemas semelhantes com muitas variáveis, e com vários exemplos já classificados pelo nosso método anterior, discutimos a construção de um modelo usando o método Multi-Layer Perceptron para classificar se os estados são emaranhados e locais ou não.

**Palavras-chave:** Emaranhamento. *Steering*. Não-localidade de Bell.

# ABSTRACT

Entanglement, Bell non-locality, and Steering, previously considered synonymous, are distinct nonclassical concepts. Although entanglement is necessary, it is not a sufficient condition for the correlations exhibited by a system to be nonlocal and/or steerable, and thus display stronger-than-classical behavior. The exact relationship between these three phenomena is unknown. The problem of checking whether a general quantum state can generate non-local statistics is notoriously difficult. There is little hope of obtaining a systematic analytical tool for answering this question. Therefore, numerical methods have been developed for the partial solution of such problems. In this project, the student used a method that relies on a subset of convex optimization, called semi-definite programming, which was presented in a 2016 article that includes the student's advisor as one of the authors. With this new optimized and generalized method, we were able to investigate and certify the locality of a wide range of entangled quantum states, and obtained improved bounds on the volume of two-qubit states that are both entangled and local. However, this method has limitations and works best with low dimensions and few measurements. Based on the success of Machine Learning techniques on similar problems with many variables, and with several examples already classified by our previous method, we discuss building a model using the Multi-Layer Perceptron method to classify whether states are entangled and local or not.

**Keywords:** Entanglement. Steering. Bell non-locality.

# Summary

<b>Preface</b>	<b>10</b>
<b>Part I: Gathering the tools</b>	<b>11</b>
<b>1. CONCEPTS</b>	<b>11</b>
1.1 Quantum states . . . . .	11
1.2 Measurements . . . . .	15
1.3 Qubits and Bloch Sphere . . . . .	16
<b>2. PROPERTIES</b>	<b>20</b>
2.1 Measurement incompatibility . . . . .	20
2.2 Bell non-locality . . . . .	23
2.3 Steering . . . . .	26
<b>3. HIERARCHY OF PROPERTIES</b>	<b>28</b>
3.1 Steering and the other properties . . . . .	28
3.2 Bell non-locality and the other properties . . . . .	30
3.3 Hierarchy diagram . . . . .	31
<b>Part II: Building results</b>	<b>33</b>
<b>4. CERTIFICATION OF LOCAL ENTANGLED STATES</b>	<b>33</b>
4.1 Entangled states certification . . . . .	33
4.2 Local states certification . . . . .	35
<b>5. SAMPLING QUANTUM STATES</b>	<b>38</b>
5.1 Hilbert-Schmidt . . . . .	38
5.2 Bures . . . . .	39
<b>6. CODE DOCUMENTATION AND VALIDATION</b>	<b>41</b>



<b>Part III: Looking at the final work</b>	<b>47</b>
<b>7. THE RESULTS</b>	<b>47</b>
7.1 Werner states . . . . .	47
7.2 Relative volume . . . . .	48
7.3 Machine Learning . . . . .	49
<b>8. DISCUSSIONS, CONCLUSIONS AND NEXT STEPS</b>	<b>53</b>
<b>BIBLIOGRAPHY</b>	<b>54</b>
<b>A. CONVEX OPTIMIZATION</b>	<b>59</b>
<b>B. MACHINE LEARNING</b>	<b>64</b>

# PREFACE

The decision to pursue a master's degree is unique to each individual. Often, even the student herself may not have a clear justification for it. In my case, the opportunity to spend two years working on a theoretical physics project on the foundations of quantum physics was an opportunity I couldn't pass up. It gave me the chance to study and re-study concepts in physics and mathematics that had excited me during my undergraduate course. Moreover, it opened up avenues of study that I hadn't even known existed when I graduated.

After completing the first year of the master's, I wrote a report and presented a qualifying seminar. Throughout this process, I was able to revisit articles that I had already read during my undergraduate studies. It was a satisfying feeling to be able to identify the authors, the importance of the articles in the field, the consequences of their results and to see growth in my understanding of the topics. The justification behind my pursuit of this master's degree is simply to learn and to discover new limits in my own knowledge.

With this thesis, I aim to create a material that is as clear, didactic, complete, and adjusted to the intended audience as possible. I imagined myself talking to a friend and tried to present the concepts and ideas learned over those two years in the best possible way. I am aware of the difficulties that the reader may encounter, and I have tried to be practical in overcoming them. The text is progressive, and I have ordered the subjects and built the appendices in a way that would have helped me when I started. I have provided recommendations for other complementary texts at the beginning of each session, in addition to references throughout the text.

I have added most of the proofs, commenting on each step as much as possible, to avoid gaps and unanswered questions. Where possible, I have added images to aid in visualization. I clarify notation abuses and contraindications in the use of each definition.

Some previous knowledge is necessary: the codes were all built in Python, and I will not focus on the study of programming or a tutorial of the language itself. In the appendix, we cover only the construction of the technique that will be used: Convex Optimization, in particular, Semi-Definite Programming and Machine Learning. Mathematically, the reader should be comfortable with functions and linear algebra, and in physics, knowledge of standard undergraduate curricula is sufficient.

# PART I: GATHERING THE TOOLS

## CHAPTER 1

### CONCEPTS

---

We will begin this project by introducing two fundamental building blocks of quantum theory: quantum states and measurements. For those who are new to these concepts, we recommend consulting references [1,2] for further reading.

#### 1.1 Quantum states

In quantum mechanics, the state is an operator  $\rho$  that acts on the vectors that belong to the Hilbert space  $\mathcal{H}^d$ , of dimension  $d$  (in this work we will restrict ourselves to finite dimensions), associated to the system it describes. A density operator  $\rho$ , as it is called, is defined according to the conditions:

- Normalization:  $\text{tr}(\rho) = 1$ , where  $\text{tr}(\cdot)$  denotes the trace.
- Positivity:  $\rho$  is semi-definite positive, that is,  $\langle \psi | \rho | \psi \rangle \geq 0$  for all  $|\psi\rangle \in \mathcal{H}$ . It is denoted  $\rho \geq 0$ . Since a positive operator is necessarily Hermitian, we gain here a second property: Hermiticity. The density operator  $\rho$  is Hermitian,  $\rho = \rho^\dagger$ , where  $\rho^\dagger$  is the conjugate transposition of  $\rho$ .

Every density operator can be written as a convex combination of one-dimensional projectors,

$$\rho = \sum_i q_i |\psi_i\rangle \langle \psi_i|; \quad \sum_i q_i = 1, \quad q_i \geq 0; \quad (1.1)$$

as long as  $\|\psi_i\| = 1$  (normalization condition). This decomposition, in general, is not unique [2], with the exception of the pure states  $\rho = |\psi\rangle \langle \psi|$ . Two vectors that are equal up to one global phase represent the same quantum state,  $|\psi\rangle \sim e^{i\varphi} |\psi\rangle$ .

Here, we use Dirac notation [3] to represent the vectors of the associated Hilbert space, so we represent the “ket”  $|\psi\rangle$  as a column vector. Its dual, called “bra”, is its Hermitian conjugate, i.e.  $|\psi\rangle^\dagger = \langle \psi|$ . In the matrix representation, the Hermitian conjugate operation reduces to taking the conjugate transpose, thus the “bra” represents a row vector.

We can now define two different operations with bras and kets, giving mathematical meaning to the expressions  $\langle\psi|\phi\rangle$  and  $|\psi\rangle\langle\phi|$ . The first represents the inner product (it's a map  $\langle\cdot|\cdot\rangle : \mathcal{H}^d \times \mathcal{H}^d \rightarrow \mathbb{C}$ ) and the second the outer product (is a map  $|\cdot\rangle\langle\cdot| : \mathcal{H}^d \times \mathcal{H}^q \rightarrow \mathbb{C}^{d \times q}$ ).

The Hilbert space associated with a system composed of two or more parts is formed by the tensor product of the spaces associated with the subsystems. A bipartite system, composed of subsystems whose Hilbert spaces are  $\mathcal{H}_A$  and  $\mathcal{H}_B$ , has the space  $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$  associated with it.

Let  $\rho^{AB}$  be a composite state associated with  $\mathcal{H}_A \otimes \mathcal{H}_B$ , where  $\{u_i\}$  and  $\{v_\mu\}$  are bases for  $\mathcal{H}_A$  and  $\mathcal{H}_B$ , respectively. Then  $\{|u\rangle \otimes |v\rangle\}$ , or in a more simplified notation  $\{|u_i v_\mu\rangle\}$ , will be an orthonormal basis for  $\mathcal{H}_A \otimes \mathcal{H}_B$ . We can decompose the composite state into the vector tensor products of the basis,

$$\rho^{AB} = \sum_{i,j,\mu,\nu} \rho_{i\mu,j\nu} |u_i v_\mu\rangle \langle u_j v_\nu|.$$

The reduced density operator for subsystems  $A$  and  $B$  are respectively defined by

$$\begin{aligned} \rho^A &\equiv \text{tr}_B(\rho^{AB}) = \sum_{i,j,\mu} \rho_{i\mu,j\mu} |u_i\rangle \langle u_j|, \\ \rho^B &\equiv \text{tr}_A(\rho^{AB}) = \sum_{i,\mu,\nu} \rho_{i\mu,i\nu} |v_\mu\rangle \langle v_\nu|, \end{aligned}$$

where the operations  $\text{tr}_A$  and  $\text{tr}_B$  represent the partial trace operations on the subsystems  $A$  and  $B$ , respectively.

A product, or uncorrelated, state of this system is a state of the form  $\rho^{AB} = \rho^A \otimes \rho^B$ . However, a crucial point in understanding these reduced operators is the fact that a state is not always a product state. We can also have states said to be separable, or classically correlated, in the form

$$\rho = \sum_{r=1}^n q_r \rho_r^A \otimes \rho_r^B, \quad q_r \geq 0, \quad \sum_{r=1}^n q_r = 1.$$

States that cannot be prepared like this, and are incompatible with the form presented above, are said to be entangled. All construction of these composite systems were presented with bipartite systems, but the generalization is straightforward.

The entanglement of bipartite pure states can be described using the Schmidt decomposition, which constructs a pair of preferred orthonormal bases to emphasize the tight correlations between two quantum subsystems.

It is interesting to note that Erhard Schmidt's original work [4] was published in 1906, before the concept of entanglement appeared in Schrödinger's famous paper on the "Cat Paradox" [5]. Schmidt focused on investigating kernels of integral equations. Hugh

Everett III [6, 7] provided the decomposition for finite-dimensional systems, which is mostly applied in quantum physics nowadays.

**Theorem 1 (Schmidt decomposition)** *Any vector  $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$  can be expressed in the form*

$$|\Psi\rangle = \sum_i \lambda_i |\psi_i\rangle_A |\phi_i\rangle_B,$$

where  $\{|\psi_i\rangle_A\}$  is a basis of  $\mathcal{H}_A$ ,  $\{|\phi_i\rangle_B\}$  is a basis of  $\mathcal{H}_B$  and the coefficients  $\lambda_i$  are real, non-negative and  $\sum_j \lambda_i^2 = 1$ .

For the proof of this theorem, it follows that: Let  $\rho_A$  be the reduced state of the system  $A$  and its spectral decomposition

$$\rho_A = \sum_i \alpha_i |\psi_i\rangle_A \langle \psi_i|_A,$$

given the orthonormal basis  $\{|\psi_i\rangle_A\}$  for  $\mathcal{H}_A$  and let  $\{|\omega_j\rangle_B\}$  be any orthonormal basis for  $\mathcal{H}_B$ , we see that  $\{|\psi_i\rangle_A |\omega_j\rangle_B\}$  will be an orthonormal base for  $\mathcal{H}_A \otimes \mathcal{H}_B$ . Therefore, we can write

$$|\Psi\rangle = \sum_{i,j} c_{ij} |\psi_i\rangle_A |\omega_j\rangle_B.$$

The coefficients  $c_{ij}$  can be calculated through  $c_{ij} = \langle \Psi | (|\psi_i\rangle_A |\omega_j\rangle_B) \rangle$ . If we define  $\lambda_i = \sqrt{\alpha_i}$ , we can rewrite

$$|\Psi\rangle = \sum_i \lambda_i |\psi_i\rangle_A \left( \sum_j \frac{c_{ij}}{\lambda_i} |\omega_j\rangle_B \right). \quad (1.2)$$

We can define  $|\phi_i\rangle_B = \sum_j \frac{c_{ij}}{\lambda_i} |\omega_j\rangle_B$  and thus we have

$$|\Psi\rangle = \sum_i \lambda_i |\psi_i\rangle_A |\phi_i\rangle_B.$$

We have to ensure that this new set  $\{|\phi_i\rangle_B\}$  can be extended to an orthogonal basis of  $\mathcal{H}_B$ . For that, we have to show that  $\langle \phi_i | \phi_{i'} \rangle = 0$  when  $i \neq i'$  and 1 when  $i = i'$ . Calculating, we have

$$\langle \phi_i | \phi_{i'} \rangle = \sum_{j,j'} \frac{c_{ij}^* c_{i'j'}}{\lambda_i \lambda_{i'}} \langle \omega_j | \omega_{j'} \rangle,$$

since  $\{|\omega_j\rangle_B\}$  was an orthonormal basis and using the definition of  $c_{ij}$ , we have

$$\begin{aligned}
 \langle\phi_i|\phi_{i'}\rangle &= \sum_j \frac{c_{ij}^* c_{i'j}}{\lambda_i \lambda_{i'}} \\
 &= \sum_j \frac{\langle\P|(|\psi_{i'}\rangle_A |\omega_j\rangle_B)(\langle\psi_i|_A \langle\omega_j|_B)|\Psi\rangle}{\lambda_i \lambda_{i'}} \\
 &= \frac{\langle\P|(|\psi_{i'}\rangle_A \langle\psi_i|_A \otimes \sum_j |\omega_j\rangle_B \langle\omega_j|_B)|\Psi\rangle}{\lambda_i \lambda_{i'}} \\
 &= \frac{\langle\P|(|\psi_{i'}\rangle_A \langle\psi_i|_A \otimes \mathbb{I})|\Psi\rangle}{\lambda_i \lambda_{i'}}.
 \end{aligned} \tag{1.3}$$

Equivalent to Equation 1.3 is to perform the trace operation

$$\begin{aligned}
 \langle\phi_i|\phi_{i'}\rangle &= \frac{\text{tr}(|\Psi\rangle\langle\P|(|\psi_{i'}\rangle_A \langle\psi_i|_A \otimes \mathbb{I}))}{\lambda_i \lambda_{i'}} \\
 &= \frac{\text{tr}(\text{tr}_B[|\Psi\rangle\langle\P|(|\psi_{i'}\rangle_A \langle\psi_i|_A \otimes \mathbb{I})])}{\lambda_i \lambda_{i'}} \\
 &= \frac{\text{tr}(\rho_A |\psi_{i'}\rangle_A \langle\psi_i|_A)}{\lambda_i \lambda_{i'}} \\
 &= \frac{\langle\psi_i|_A \rho_A |\psi_{i'}\rangle_A}{\lambda_i \lambda_{i'}}.
 \end{aligned}$$

As we know that  $\{|\psi_i\rangle_A\}$  is the orthonormal base for  $\mathcal{H}_A$ , it follows

$$\begin{aligned}
 \langle\phi_i|\phi_{i'}\rangle &= \frac{\alpha_i \langle\psi_i|\psi_{i'}\rangle}{\lambda_i \lambda_{i'}} \\
 &= \frac{\alpha_i}{\lambda_i \lambda_{i'}} \delta_{ii'}.
 \end{aligned}$$

Therefore, and remembering that  $\lambda_i = \sqrt{\alpha_i}$ , we have that  $\langle\phi_i|\phi_{i'}\rangle$  will be 0 when  $i \neq i'$  and 1 when  $i = i'$ , that is,  $\{|\phi_i\rangle_B\}$  is an orthonormal basis. Thus ending the proof of Theorem 1.

Through the Schmidt decomposition theorem, we have to choose suitable bases  $\{|\psi_i\rangle_A\}$  and  $\{|\phi_i\rangle_B\}$ , called Schmidt bases for  $|\Psi\rangle$ . With this choice of basis, we have that we don't need the cross terms of  $i$  and  $j$  in Equation 1.2. The coefficients  $\lambda_i$  are called the Schmidt coefficients for  $|\Psi\rangle$  and the number of non-zero coefficients  $\lambda_i \neq 0$  is called the Schmidt number of  $|\Psi\rangle$ . Some important consequences of this theorem are

- The reduced states  $\rho_A$  and  $\rho_B$  will have the same eigenvalues, and therefore the same rank (Schmidt number):

$$\begin{aligned}
 \rho_A &= \sum_i \lambda_i^2 |\psi_i\rangle_A \langle\psi_i|_A, \\
 \rho_B &= \sum_i \lambda_i^2 |\phi_i\rangle_B \langle\phi_i|_B;
 \end{aligned}$$

- The vector  $|\Psi\rangle$  generates a state product if and only if the Schmidt number equals 1.

Therefore, we have a direct relationship between the Schmidt number and the entanglement of a state. All development here was done with pure states, but there is a generalization for mixed states.

## 1.2 Measurements

Complementary to the states are the so-called measurement operators. A measurement is described by a operator  $M$ , called observable. Let  $a_n$  be the value associated with the measurement  $M$  on the element  $|n\rangle$  of the basis of eigenvectors of  $M$ , so  $M|n\rangle = a_n|n\rangle$ . Constraining the observable to be Hermitian ensures that all measurement results will be real numbers. The matrix representation of the operator  $M$  has its terms given by

$$M_{mn} = \langle m|M|n\rangle = a_n\langle m|n\rangle = a_n\delta_{mn}.$$

Based on the equation above, we can then define a projective measurement as one in which each outcome  $a_n$  is associated with a projector  $\Pi_n$ , and each projector associated with a distinct outcome corresponds to an orthogonal subspace to others [1]. Whence it follows that

$$M = \sum_m \sum_n |m\rangle M_{mn} \langle n| = \sum_m \sum_n |m\rangle \langle m|M|n\rangle \langle n| = \sum_n a_n |n\rangle \langle n| = \sum_n a_n \Pi_n.$$

This result tells us that we can decompose any observable in terms of projective measurements, represented by projectors  $\Pi_n$ .

When applying a projective measurement on any  $\rho$  state, the probability of obtaining the result  $a_n$  is given by  $p(a_n) = \text{tr}(\Pi_n \rho)$  (also known as *Born's rule*). If we want the measurements to be reproducible, that is, that when we subsequently carry out the same operation we obtain the same result, it is necessary that after the first one the state of the system becomes  $\rho_n = (\Pi_n) \rho (\Pi_n) / p(a_n)$ .

By relaxing the condition that we need to know the final state of the system, we can discuss POVMs (*positive-operator valued measure*), where we are only interested in the probabilities of obtaining each outcome. Projective measurements become just a special case of POVMs [1].

If  $x$  is any POVM, we associate the effects  $M_{a|x}$  to the possible results  $a$ . An effect is necessarily a positive operator,  $M_{a|x} \geq 0$ , and also satisfies the completeness relation  $\sum_a M_{a|x} = \mathbb{I}$ . The probability of obtaining the value  $a$  when applying the effect  $x$  to the system  $\rho$  is still  $p_A(a|x) = \text{tr}(M_{a|x} \rho)$ . However, unlike projective measurements, effects associated with distinct  $a$  results do not necessarily correspond to orthogonal subspaces. It is in this sense that the condition of knowing the post-measurement state was relaxed.

In a composite state  $\rho^{AB}$ , measurements of the type  $M_A \otimes \mathbb{I}_B$  acting on  $\rho^{AB}$  present

the same statistics as for all  $M_A$  acting on  $\rho^A$ , and it is in this sense that we say that the operator represents its respective subsystem.

### 1.3 Qubits and Bloch Sphere

Now that we have fairly straightforward and operational descriptions of what our states and measurements are, it's interesting to get an intuition of how they behave and what we can do with them through the geometric visualization of the cases of systems associated with the Hilbert space  $\mathcal{H}^2 = \mathbb{C}^2$ . Here we will be able to use the representations of qubits and the Bloch sphere. Qubits are mathematical abstractions that refer to simpler, non-trivial quantum systems.

In  $\mathcal{H}^2$ , the canonical orthonormal basis, commonly called the computational basis in the literature, is defined as

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

A ket vector can be written as a linear combination of this basis and, given the normalization and global phase conditions, can be conveniently parameterized through two angles:

$$\begin{aligned} |\psi\rangle &= \alpha|0\rangle + \beta|1\rangle = |\alpha|e^{i\varphi_\alpha}|0\rangle + |\beta|e^{i\varphi_\beta}|1\rangle \\ &= \cos\theta e^{i\varphi_\alpha}|0\rangle + \sin\theta e^{i\varphi_\beta}|1\rangle && \text{(Normalization)} \\ &= \cos\left(\frac{\theta}{2}\right)|0\rangle + \sin\left(\frac{\theta}{2}\right)e^{i(\varphi_\beta - \varphi_\alpha)}|1\rangle && \text{(Global phase)} \\ &= \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\varphi} \sin\left(\frac{\theta}{2}\right)|1\rangle, \end{aligned}$$

where  $\theta \in [0, \pi]$  and  $\varphi \in [0, 2\pi]$ . In [Figure 1.1](#) we can do the visualization of this parametrization.

When we use this parametrization in the definition of the density operators [1.1](#) and  $e^{i\psi} = \cos\psi + i\sin\psi$ , we get

$$\begin{aligned} \rho &= \sum_i q_i |\psi_i\rangle \langle \psi_i| \\ &= \sum_i q_i \left( \cos(\theta_i/2)|0\rangle + e^{i\varphi_i} \sin(\theta_i/2)|1\rangle \right) \left( \cos(\theta_i/2)\langle 0| + e^{-i\varphi_i} \sin(\theta_i/2)\langle 1| \right) \\ &= \sum_i q_i \begin{pmatrix} \cos^2(\theta_i/2) & \cos(\theta_i/2) \sin(\theta_i/2) (\cos\psi - i\sin\psi) \\ \cos(\theta_i/2) \sin(\theta_i/2) (\cos\psi + i\sin\psi) & \sin^2(\theta_i/2) \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 1 + v_z & v_x - i v_y \\ v_x + i v_y & 1 - v_z \end{pmatrix}, \end{aligned}$$



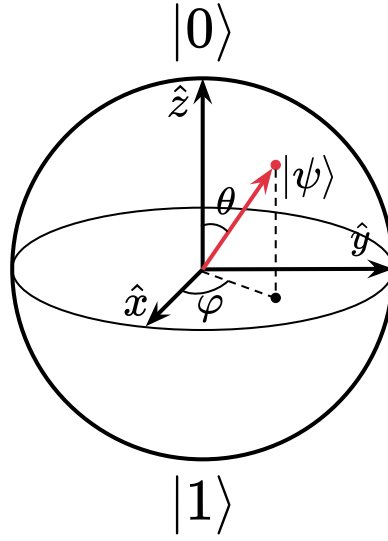


Figure 1.1: **Bloch sphere.** An arbitrary ket  $|\psi\rangle$  on the Bloch sphere is representable as a unique linear combination of the two basis elements, here chosen as  $|0\rangle$  and  $|1\rangle$ , with coefficients defined by the two angles  $\theta$  and  $\varphi$ .

where

$$\begin{aligned} v_z &= \sum_i q_i (\cos^2(\theta_i/2) - \sin^2(\theta_i/2)), \\ v_x &= \sum_i q_i \cos(\theta_i/2) \sin(\theta_i/2) \cos \psi, \\ v_y &= \sum_i q_i \cos(\theta_i/2) \sin(\theta_i/2) \sin \psi. \end{aligned}$$

Defined the identity matrix (matrix representation of the operator with the same name)

$$\mathbb{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and the three Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

any quantum state of a qubit can be written as

$$\rho = \frac{\mathbb{I} + \vec{v} \cdot \vec{\sigma}}{2},$$

where  $\vec{v}$  is called the Bloch vector, with  $\vec{v} \in \mathbb{R}^3$  and  $|\vec{v}| \leq 1$  (positivity condition), and  $\vec{\sigma}$  is the vector formed by the Pauli matrices.

One of the great advantages of this parameterization is the direct geometric interpretation of the state space of a qubit. The correspondence between density operators and vectors  $\vec{v}$  is one-to-one, and this allows identifying the state space of  $\mathcal{H}^2$  with the three-

dimensional ball of unit radius, inserted in  $\mathbb{R}^3$ , formed by the real vectors  $\vec{v}$  with norm less than or equal to 1. We can do the visualization of this in [Figure 1.2](#).

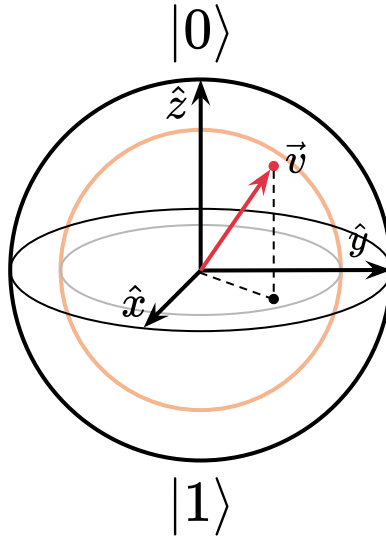


Figure 1.2: **Bloch vector.** Bloch vectors correspond to density operators, allowing the geometric visualization in  $\mathbb{R}^3$  of the space of  $\mathbb{C}^2$ .

Analogously, we can also represent dichotomous projective measurement operators (two results) in Bloch's sphere, given by

$$M_{\pm|x} = \frac{\mathbb{I} \pm \hat{v}_x \cdot \vec{\sigma}}{2}, \quad (1.4)$$

where  $\hat{v}_x$  is a unitary vector that represents the measurement direction and the direction will be associated by the sign, representing the measurement result (+1 or -1). With this, we see that a dichotomous projective measurement can be represented as two antipodal points on the surface of the Bloch sphere. Therefore, we can build polytopes associated with sets of measurements, given that each point representing an effect will be a vertex of the polytope. We can visualize this idea in [Figure 1.3](#).

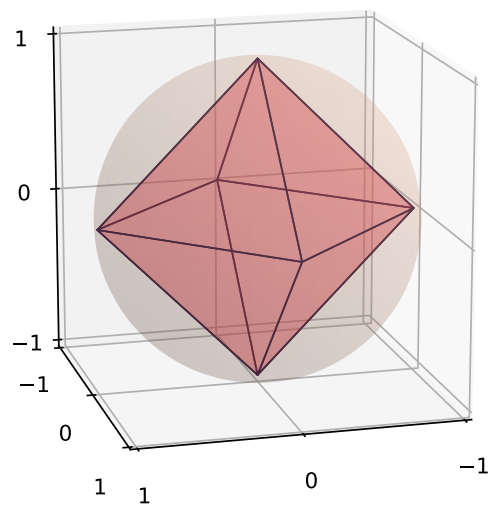


Figure 1.3: **Polytope of measurements.** A 6-vertex polytope, the regular octahedron, representing 3 dichotomous projective measurements with vertices  $\pm(1, 0, 0)$ ,  $\pm(0, 1, 0)$  and  $\pm(0, 0, 1)$ .

# CHAPTER 2

## PROPERTIES

---

After presenting our basic concepts and explaining how to use them, we can begin to analyze the properties that they exhibit. First, let's consider the possibility of two measurements being incompatible with each other. Then, we'll explore scenarios where measurements are performed on states, and non-locality is observed.

### 2.1 Measurement incompatibility

There are a variety of references that can be consulted in order to further explore the topic of measurement incompatibility. I strongly suggest the reading of [8–11].

There are different notions characterizing compatibility relations between quantum observables; such as commutativity, non-disturbance, joint measurability, coexistence, etc. For projective measurements, all four relations are equivalent, however, for generalized observables, all are non-equivalent. There is, however, a hierarchical relationship between some of them [12].

We will begin by defining the more general concept of measurement compatibility. Two observables  $A$  and  $B$ , with outcome sets  $\{m\}$  and  $\{n\}$ , are compatible if there exists a third observable  $C$ , with an outcome set  $\{k\}$ , and functions  $f : \{k\} \rightarrow \{m\}$  and  $g : \{k\} \rightarrow \{n\}$  such that

$$A(x) = \sum_{z: f(z)=x} C(z), \quad B(y) = \sum_{z: g(z)=y} C(z). \quad (2.1)$$

The functions  $f$  and  $g$  are simply relabeling the outcomes of  $C$ . If  $A$  and  $B$  are not compatible, then they are incompatible.

Another compatibility relation is the commutativity criterion. Probably the best known property of measurement incompatibility. We say that two observables commute if  $[O_1, O_2] \equiv O_1 O_2 - O_2 O_1 = 0$ . Due to the relative ease of this calculation it makes sense that it is the best known, especially knowing that it is enough to show the incompatibility of projective measurements.

The second incompatibility relation we will study is the joint measurability. A set of  $m$  POVMs  $\{M_{a|x}\}$  is called jointly measurable if there is a measurement  $M_{\vec{a}}$  with result

$\vec{a} = [a_1, a_2, \dots, a_m]$ , where  $a_x$  gives the result of the measurement  $x$ , i.e.

$$M_{\vec{a}} \geq 0, \quad \sum_{\vec{a}} M_{\vec{a}} = \mathbb{I}, \quad \sum_{\vec{a}/a_x} M_{\vec{a}} = M_{a|x},$$

where  $\vec{a}/a_x$  represents the elements of  $\vec{a}$  except  $a_x$ . Thus, all POVM elements  $M_{a|x}$  are retrieved as marginals from the joint observable  $M_{\vec{a}}$ . We can visualize the concept of joint observable in [Figure 2.1](#).

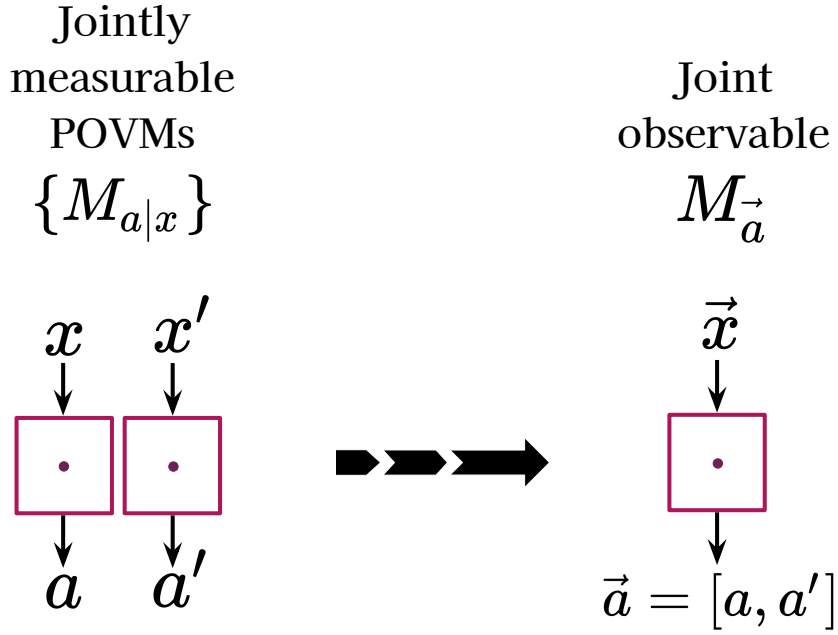


Figure 2.1: **Joint observable.** The elements of the set of POVMs  $\{M_{a|x}\}$  jointly measurable are retrieved as marginals from the joint observable  $M_{\vec{a}}$ .

We can show that commuting observables is a particular case of their being jointly measurable. We can define  $M_{\vec{a}} = M_{a|x} M_{a'|x'}$ , where we have  $\sum_{\vec{a}} M_{\vec{a}} = \mathbb{I}$  and  $\sum_{\vec{a}/a_x} M_{\vec{a}} = M_{a|x}$ , directly, and using the commutativity of  $M_{a|x}$  and  $M_{a'|x'}$ , we will have  $M_{\vec{a}} \geq 0$ . However, two observables can be jointly measurable even without commuting [\[10\]](#).

Joint measurability is a strictly weaker notion of compatibility for POVMs, but if we are considering a finite number of observables, then they are equivalent definitions [\[9\]](#). We can conclude this by seeing that if  $A$  and  $B$  are compatible and  $C, f, g$  are such that [Equation 2.1](#) holds, then we can define a new observable  $G$  on the product outcome set  $\{m\} \times \{n\}$  as

$$G(x, y) = \sum_{z: f(z)=x \wedge g(z)=y} C(z)$$

and then [Equation 2.1](#) implies that

$$A(x) = \sum_y G(x, y), \quad B(y) = \sum_x G(x, y). \quad (2.2)$$

This is a special case of [Equation 2.1](#), where the relabeling functions are taken to be the projections  $f(x, y) = x$  and  $g(x, y) = y$ . Any observable  $G$  satisfying [Equation 2.2](#) is called a joint observable of  $A$  and  $B$ . Therefore, we can visualize a summary of this hierarchy in [Figure 2.2](#).

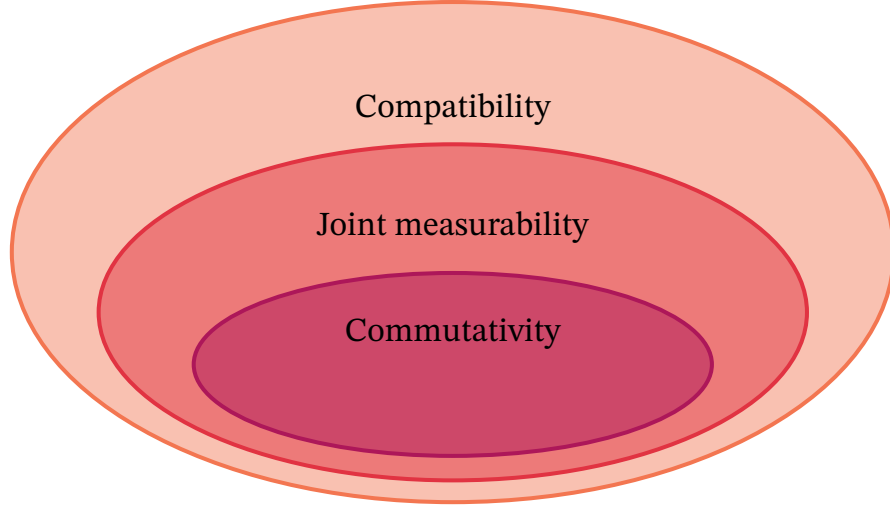


Figure 2.2: **Hierarchy of measurement compatibility concepts.** Commutativity implies joint measurability and compatibility, but a set of measurements can be jointly measurable and do not commute, or even be compatible and neither jointly measurable nor commute.

We will define a last notion of incompatibility called coexistence. Two POVMs,  $A$  and  $B$ , are said to coexist if there exists a POVM  $M$  such that each effect of  $A$  or  $B$  can be simulated by  $M$ , and all binary measurements that can be formed from  $A$  and  $B$  can be measured simultaneously. However, this does not provide a direct way to measure the entire observables  $A$  and  $B$  simultaneously. The relationship between coexistence and joint measurability was studied in article [\[13\]](#), where it was shown that coexistence does not imply joint measurability. Furthermore, article [\[14\]](#) demonstrated that there are even stronger relationships than compatibility, such as broadcastability, one-side broadcastability, mutual nondisturbance, and nondisturbance.

Another way to certify measurement incompatibility is through incompatibility witnesses. This method, however, must be used on a case-by-case basis, as we will have to find the specific witness that will allow us to certify the incompatibility of the measurements in question, as we will see below.

In general, a witness is any experimentally assessable linear function whose values are greater than or equal to zero whenever the measured object does not have the investigated property, but results in a negative value for at least one object that does have the property.

The general framework of witnesses is formally defined from  $\mathcal{V}$  a real linear space and  $S_0 \subset \mathcal{V}$  a compact convex subset that mathematically describes the objects of interest. This set is then divided into two disjoint subsets (empty intersection)  $S$  and  $\bar{S}$ , with  $S$  being

closed and convex. We can think of subsets as properties, either the element  $x \in S_0$  is in  $S$  or in  $\bar{S}$ . A witness to the property  $\bar{S}$ , or  $\bar{S}$ -witness, is a map  $\xi : S \rightarrow \mathbb{R}$  such that

1.  $\xi(x) \geq 0$  for all  $x \in S$  and  $\xi < 0$  at least for some  $x \in \bar{S}$ ;
2.  $\xi(tx + (1-t)y) = t\xi(x) + (1-t)\xi(y)$  for all  $x, y \in S_0$  and  $t \in [0, 1]$ .

By the second condition, each witness generates a hyperplane separating  $\mathcal{V}$  into two half-spaces. The first condition ensures that one of the halves completely contains  $S$ , but not all of  $S_0$ .

Witnesses are associated with hyperplanes and are equivalent in detection if they produce the same set separation. In Figure 2.3, we can visualize the representation of two equivalent witnesses, which detect the red dot, but not the black one.

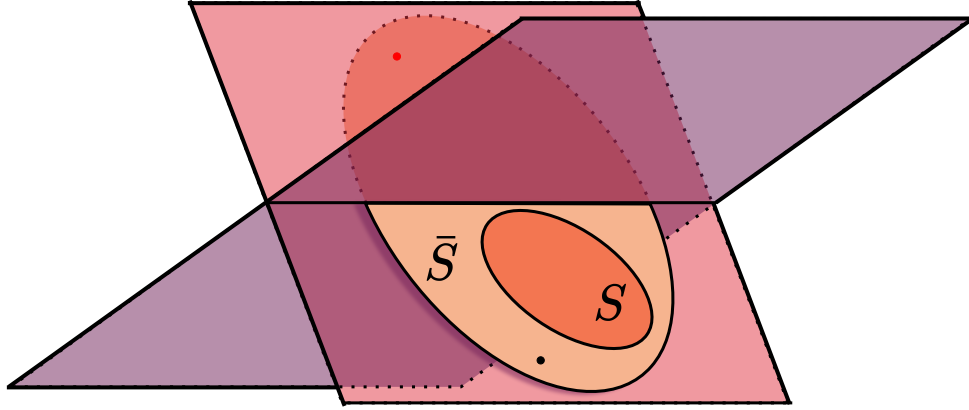


Figure 2.3: **Witnesses.** Two equivalent witnesses, each one associated with one hyperplane. They both detect the red point, but not the black one. Image based on figure 1 of article [15].

## 2.2 Bell non-locality

The EPR paradox is a thought experiment by Einstein, Podolsky, and Rosen arguing that quantum mechanics is incomplete. In a 1935 paper titled “Can Quantum-Mechanical Description of Physical Reality be Considered Complete?” [16] they proposed the existence of “elements of reality” that were not part of quantum theory, and that it should be possible to construct a theory containing the so called hidden variables. Later, in 1964, John Bell’s paper “On the Einstein Podolsky Rosen paradox” [17] presented a theorem that demonstrates the incompatibility between quantum mechanics and local hidden-variable theories. In this context, “local” refers to the principle of locality, the idea that a particle can only be influenced by its immediate surroundings. The interactions mediated by physical fields cannot propagate faster than the speed of light.

The scenario we will consider is that of two distant observers, Alice and Bob, making local measurements on a shared entangled quantum state,  $\rho$ . The measurements are described by the POVMs  $\{A_{a|x}\}$  and  $\{B_{b|y}\}$  (where  $A_{a|x} \geq 0$  and  $\sum_a A_{a|x} = \mathbb{I}$ , likewise for  $B_{b|y}$ ), where  $x$  and  $y$  denote the choice of measurements and  $a$  and  $b$  denote the corresponding results. We say that each observer has what we can call a measurement box, which gets clearer if we visualize this scenario in [Figure 2.4](#).

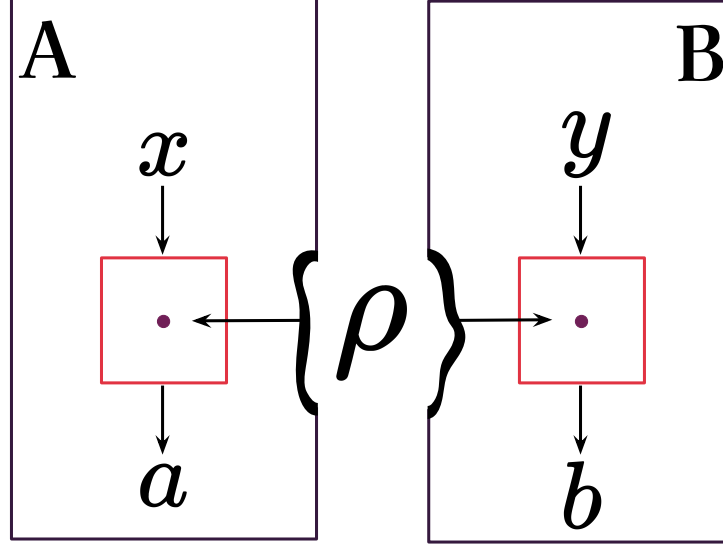


Figure 2.4: **Bell non-locality scenario.** Alice and Bob each receive a part of the shared entangled quantum state and perform the measurement  $x$  and  $y$ , respectively. The results obtained are  $a$  and  $b$ , respectively.

The corresponding probability distributions are given by

$$p(ab|xy) = \text{tr}([A_{a|x} \otimes B_{b|y}]\rho). \quad (2.3)$$

If two measurement events are spatially separated, as represented in [Figure 2.5](#), there is no direct causal connection between them. However, the observed correlations could have arisen from the causal influence of an unknown third event in the common past of the first two. Let us investigate how a classical mechanism based on such unknown common event would explain the observed correlations.

Let  $\Lambda$  be a random variable that assumes values in the set  $\{\lambda\}$ , with probabilities  $\{p(\lambda)\}$ . A measurement of  $\Lambda$  is performed in the common past of the spatially separated measurement events  $a|x$  and  $b|y$ , and the value  $\lambda$  obtained influences the operation of both measurement boxes. This way:

$$p(ab|xy, \lambda) = p_A(a|x, \lambda) p_B(b|y, \lambda).$$



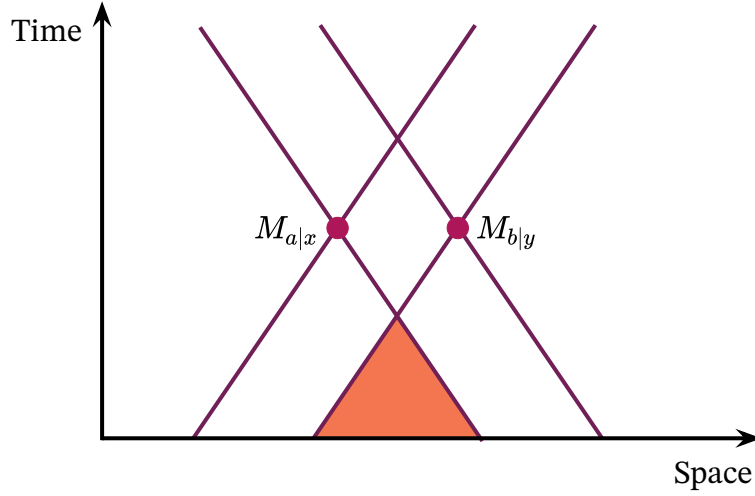


Figure 2.5: **Measurement events spatially separated.** Two measurements events  $M_{a|x}$  and  $M_{b|y}$  separated by a space-like interval represented by points. The light cones are delimited and we can see where they meet. The painted area represents the common past of the events.

The observed correlations come from the marginalization on  $\lambda$ :

$$\begin{aligned}
 p(ab|xy) &= \int_L p(ab, \lambda|xy) d\lambda \\
 &= \int_L p(ab|xy, \lambda) p(\lambda|xy) d\lambda \\
 &= \int_L p_A(a|x, \lambda) p_B(b|y, \lambda) p(\lambda) d\lambda,
 \end{aligned} \tag{2.4}$$

with  $\int p(\lambda) d\lambda = 1$ . Here we are using the assumption that  $\lambda$  does not depend on the choices of measurements  $x$  and  $y$  and therefore  $p(\lambda|xy) = p(\lambda)$ . A quantum state  $\rho$  is said to be local, or equivalently admits a local hidden variable model (LHV), when the statistics of arbitrary local measurements can be reproduced by a distribution of the form Equation 2.4. In contrast, if such a decomposition does not exist, the state is non-local and violates a Bell inequality for appropriately chosen local measurements [2].

In Bell scenario with finite numbers of parts, measurements per part, and possible outcomes per measurement, we can discretize Equation 2.4. For this, we can assume that the marginal probabilities  $p_A(a|x, \lambda)$  and  $p_B(b|y, \lambda)$  assume only the values 0 or 1, for any measurements and results. The randomness in the marginal probabilities is incorporated to the randomness of the variable  $\lambda$  and we can build the local behaviors from deterministic strategies, given by the set of predetermined results of each of the possible measurements of Alice and Bob. In the case of  $m_A$  measurements of  $k_A$  results each for Alice and  $m_B$  measurements of  $k_B$  results each for Bob, we will have the formation of  $N = (k_A)^{m_A} (k_B)^{m_B}$  deterministic strategies of the form

$$D = [D(a_1 b_1 | x_1 y_1), D(a_2 b_1 | x_1 y_1), \dots, D(a_k b_1 | x_1 y_1), D(a_1 b_1 | x_2 y_1), \dots, D(a_k b_k | x_{m_A} y_{m_B})],$$

where

- $D(ab|xy) = D(a|x)D(b|y)$ , since Bob's choice of measurements cannot influence Alice's and vice versa;
- $D(a|x)$  and  $D(b|y)$  are either 0 or 1 (locally deterministic);
- $\sum_{i=1}^{k_A} \sum_{j=1}^{k_B} D(a_i b_j | xy) = 1$  for all  $x$  and  $y$ , since probabilities must add up to 1.

Therefore, we can write the discretized local distribution of the state as

$$p(ab|xy) = \sum_{\lambda} p(\lambda) D(ab|xy).$$

## 2.3 Steering

We continue with the scenario of two distant observers. However, now only Alice will perform the measurements in the  $\rho$  state, Bob will perform his measurements in the unnormalized state

$$\sigma_{a|x} = \text{tr}_A[(A_{a|x} \otimes \mathbb{I}_B)\rho], \quad (2.5)$$

where  $\mathbb{I}_B$  is the density operator in  $B$ , conditioned on Alice having observed the result  $a$  in the measurement  $x$ , with  $\text{tr}(\sigma_{a|x}) = p_A(a|x)$ . For each possible outcome/measurement Alice has, Bob will receive a different state, we call the set of all possible states an assemblage  $\{\sigma_{a|x}\}$ . We can visualize this scenario in [Figure 2.6](#).

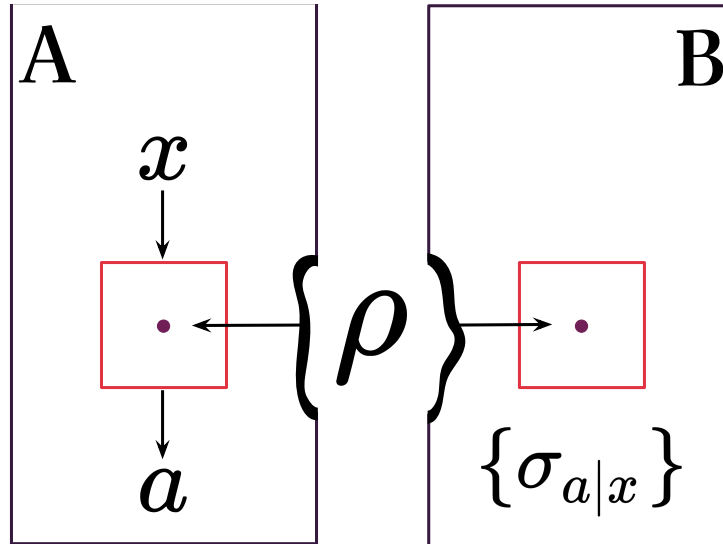


Figure 2.6: **Steering scenario.** Alice and Bob each receive a part of the shared entangled quantum state. Only Alice performs the measurement  $x$  in her subsystem and obtains the result  $a$ . Bob can perform measurements in his conditioned state to reconstruct his assemblage.

Analogously to the Bell locality scenario, let  $\Lambda$  be a random variable in the common past of Alice and Bob's measurement events, that assumes values in the set  $\{\lambda\}$  with

probability distribution  $\{p(\lambda)\}$ . The value  $\lambda$  obtained after the measurement of  $\Lambda$ , again in the common past, influences the operation of Alice's measurement box resulting  $\{p_A(a|x, \lambda)\}$  and a set of states  $\{\rho_\lambda(\lambda)\}$  of Bob's quantum system such that

$$\sigma_{a|x} = \int \rho_\lambda(\lambda) p_A(a|x, \lambda) p(\lambda) d\lambda,$$

with

$$\rho_\lambda(\lambda) \geq 0, \quad \text{tr}(\rho_\lambda(\lambda)) = 1 \quad \text{and} \quad \int d\lambda q(\lambda) = 1,$$

for all  $\lambda$ . Then the state  $\rho$  has a local state model (LHS) for these measurements and we say it is unsteerable. Otherwise, it has no local state model and we say it is steerable [18].

As in the case of local models of hidden variables, it is possible to assign all randomness to the choice of variable  $\lambda$ , and take only local deterministic behaviors for Alice

$$\begin{aligned} \sigma_{a|x} &= \sum_{\lambda} q_{\lambda}(\lambda) p_D(a|x, \lambda) \rho_{\lambda} \\ &= \sum_{\lambda} p_D(a|x, \lambda) \sigma(\lambda). \end{aligned} \tag{2.6}$$

## CHAPTER 3

# HIERARCHY OF PROPERTIES

---

Armed with the definition of these properties, we can begin to understand the exact relationships between them and identify possible hierarchies. For this topic, one might consider delving into the works [19–23].

We should start this discussion by establishing that in the case of pure states, the properties of entanglement, steering and Bell non-locality are equivalent. This means that for any pure entangled state there exists a choice of incompatible measurements that produce Bell nonlocal correlations [24]. Bearing this in mind, we can move on to study the hierarchical relationships between properties in relation to mixed states.

### 3.1 Steering and the other properties

#### Steering and entanglement

Entanglement is necessary for steering, and we can show this by first considering a non-entangled state, i.e. one that admits decomposition

$$\rho^{AB} = \sum_{\lambda} q(\lambda) \rho^A(\lambda) \otimes \rho^B(\lambda).$$

By using this state in a steering scenario, we arrive at the post measurement state given by

$$\sigma_{a|x} = \text{tr}_A[(A_{a|x} \otimes \mathbb{I}_B) \rho^{AB}] \quad (3.1)$$

$$= \sum_{\lambda} q(\lambda) \text{tr}[A_{a|x} \rho^A(\lambda)] \rho^B(\lambda) \quad (3.2)$$

$$= \sum_{\lambda} q(\lambda) p_A(a|x, \lambda) \rho^B(\lambda), \quad (3.3)$$

that is, the state is local, being necessarily unsteerable.

In Werner’s 1989 article [25], the author presents a family of states that became known as “Werner States”. A portion of these states were certified as entangled and local, that is, which are not steerable. Thus, we have that entanglement is necessary for steering, but not sufficient. This family of states will be studied later in this work in [section 7.1](#).

## Steering and measurement incompatibility

Analyzing the [Theorem 2](#) presented in [20], we have that

**Theorem 2** *The assemblage  $\{\sigma_{a|x}\}$ , with  $\sigma_{a|x} = \text{tr}_A[(A_{a|x} \otimes \mathbb{I}_B)\rho^{AB}]$ , is unsteerable for any state  $\rho^{AB}$  acting on  $\mathbb{C}^d \otimes \mathbb{C}^d$  if and only if the set of POVMs  $\{A_{a|x}\}$  acting on  $\mathbb{C}^d$  is jointly measurable.*

To prove this theorem, let's first focus on the *if* part. Our goal is to show that  $\{\sigma_{a|x}\}$  admits a decomposition of the form of [Equation 2.6](#), when  $\{A_{a|x}\}$  is jointly measurable, for any state  $\rho^{AB}$ .

Consider  $M_{\vec{a}}$ , the joint observable for  $\{A_{a|x}\}$ , and define Alice's local variable to be  $\lambda = \vec{a}$ , distributed according to  $\Pi(\vec{a}) = \text{tr}(M_{\vec{a}}\rho^A)$ , where  $\rho^A = \text{tr}_B(\rho^{AB})$ . Next Alice sends the local state  $\sigma_{\vec{a}} = \text{tr}_A[(M_{\vec{a}} \otimes \mathbb{I})\rho^{AB}]/\Pi(\vec{a})$ . When asked by Bob to perform measurement  $x$ , Alice announces an outcome  $a$  according to  $p_A(a|x, \vec{a}) = \delta_{a, a_x}$ . In this way, we can write

$$\sigma_{a|x} = \sum_{\vec{a}} p_A(a|x, \vec{a}) \sigma(\vec{a}),$$

therefore admitting a local decomposition.

We now move to the *only if* part. Our goal is now to show that if  $\sigma_{a|x}$  is unsteerable then  $\{A_{a|x}\}$  is jointly measurable. Consider an arbitrary pure state  $\rho^{AB} = |\Psi\rangle\langle\Psi|$  with Schmidt number  $d$ . We can always write  $|\Psi\rangle = (D \otimes \mathbb{I})|\Phi\rangle$ , where  $|\Phi\rangle = \sum_i |ii\rangle$  is an (unnormalized) maximally entangled vector in  $\mathbb{C}^d \otimes \mathbb{C}^d$ , and  $D$  is a diagonal matrix that contains only strictly positive numbers. The assemblage resulting from a set of POVMs  $\{A_{a|x}\}$  on  $\rho^{AB}$  is given by

$$\sigma_{a|x} = \text{tr}_A[(A_{a|x} \otimes \mathbb{I}_B)\rho^{AB}] = DM_{a|x}^T D$$

where  $M_{a|x}^T$  is the transpose of  $A_{a|x}$ . As  $\sigma_{a|x}$  is unsteerable, we have that

$$\sigma_{a|x} = \sum_{\lambda} p_A(a|x, \lambda) \sigma_{\lambda}$$

which allows us to define the positive definite operator

$$\sigma_{\vec{a}} = \sum_{\lambda} \sigma_{\lambda} \prod_x p(a_x|x, \lambda)$$

from which we can recover the assemblage  $\{\sigma_{a|x}\}$  as marginals, i.e.,  $\sigma_{a|x} = \sum_{\vec{a} \setminus a_x} \sigma_{\vec{a}}$ . Since the diagonal matrix  $D$  is invertible, we can define  $M_{\vec{a}} := D^{-1} \sigma_{\vec{a}}^T D^{-1}$ .

It is straightforward to check that  $M_{\vec{a}}$  is a joint observable for  $\{A_{a|x}\}$ , since

$$M_{\vec{a}} \geq 0, \quad \sum_{\vec{a}} M_{\vec{a}} = \mathbb{I}, \quad \sum_{\vec{a}/a_x} M_{\vec{a}} = A_{a|x}.$$

Hence,  $\{A_{a|x}\}$  is jointly measurable, which concludes the proof. Note, finally, an interesting point that follows from the above. Considering a set of incompatible measurements acting

on  $\mathbb{C}^d$ , any pure entangled state of the Schmidt number  $d$  can be used to demonstrate steering. Specifically, for any set of POVMs that are incompatible (i.e., not jointly measurable), one can find an entangled state, such that the resulting statistics violates a steering inequality.

### One-way steering

Due to the asymmetry of the concept of steering, there are entangled states that can only be steered in one direction. The article [26] provides examples of families of states that can be steered in a single direction.

## 3.2 Bell non-locality and the other properties

### Bell non-locality and steering

If a state is Bell non-local, the state is necessarily steerable. To prove this, we start with an unsteerable state  $\rho$ , where we have

$$\text{tr}_A[(A_{a|x} \otimes \mathbb{I}_B)\rho] = \int \rho_\lambda(\lambda) p_A(a|x, \lambda) p(\lambda) d\lambda.$$

When performing the locality Bell scenario, we will obtain the probability distribution given by

$$p(ab|xy) = \int p_A(a|x, \lambda) \text{tr}(B_{b|y} \rho_\lambda) p(\lambda) d\lambda,$$

which will necessarily be a local distribution.

However, a steerable state can be Bell local, as presented in [21]. The authors begin by presenting the class of states of the form

$$\rho_G = \frac{1}{9} \left[ \alpha |\psi^-\rangle \langle \psi^-| + (3 - \alpha) \frac{\mathbb{I}_2}{2} \otimes |2\rangle \langle 2| + 2\alpha |2\rangle \langle 2| \otimes \frac{\mathbb{I}_2}{2} + (6 - 2\alpha) |22\rangle \langle 22| \right]$$

where  $|\psi^-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$  and  $\mathbb{I}_2$  denotes the identity with dimension 2. For  $0 < \alpha \leq 1/2$ , these states are proven to be local for POVMs, but they are steerable (in both directions) for  $0 < \alpha \leq 1$ . Therefore, the state  $\rho_G$  with  $0 < \alpha \leq 1/2$  is two-way steerable but local for all POVMs.

### Bell non-locality and entanglement

Entanglement is necessary for Bell non-locality, and we can show this by first considering a non-entangled state, i.e. one that admits decomposition

$$\rho^{AB} = \sum_{\lambda} q(\lambda) \rho^A(\lambda) \otimes \rho^B(\lambda),$$

then the generated probabilities will be local

$$\begin{aligned}
 p(ab|xy) &= \text{tr}[(A_{a|x} \otimes B_{b|y})\rho^{AB}] \\
 &= \sum_{\lambda} q(\lambda) \text{tr}[A_{a|x}\rho^A(\lambda) \otimes B_{b|y}\rho^B(\lambda)] \\
 &= \sum_{\lambda} q(\lambda) \text{tr}[A_{a|x}\rho^A(\lambda)] \text{tr}[B_{b|y}\rho^B(\lambda)] \\
 &= \sum_{\lambda} q(\lambda) p_A(a|x, \lambda) p_B(b|y, \lambda).
 \end{aligned}$$

Just as entanglement is not enough for steering, it will not be for Bell non-locality.

### Bell non-locality and measurement incompatibility

We investigate now how joint measurability relates to Bell nonlocality. First, the [Theorem 2](#) implies that if the set of POVMs  $\{A_{a|x}\}$  used by Alice is jointly measurable, then the statistics  $p(ab|xy)$  can always be reproduced by a local model, for any state  $\rho^{AB}$  and measurements of Bob  $\{B_{b|y}\}$ . The converse problem is much more interesting. The question is whether for any set of POVMs  $\{A_{a|x}\}$  that is not jointly measurable, there exists a state  $\rho^{AB}$  and a set of measurements  $\{B_{b|y}\}$  such that the resulting statistics  $p(ab|xy)$  violates a Bell inequality. This was shown to hold true for the case of sets of two POVMs with binary outcomes [\[27\]](#). But in [\[28\]](#), they present explicitly a given set of non jointly measurable POVMs  $\{A_{a|x}\}$  from which it is not possible to construct a scenario that violates a Bell inequality. Considering a bipartite Bell test where Alice uses  $\{A_{a|x}\}$ , then for any possible shared entangled state  $\rho^{AB}$  and any set of (possibly infinitely many) POVMs  $\{B_{b|y}\}$  performed by Bob, the resulting statistics admits a local model. Consequently, these statistics do not violate any Bell inequality. This shows that quantum measurement incompatibility does not imply Bell non-locality in general.

### 3.3 Hierarchy diagram

The hierarchical relationships presented here can be conveniently represented through the diagram shown in [Figure 3.1](#), based on article [\[21\]](#).

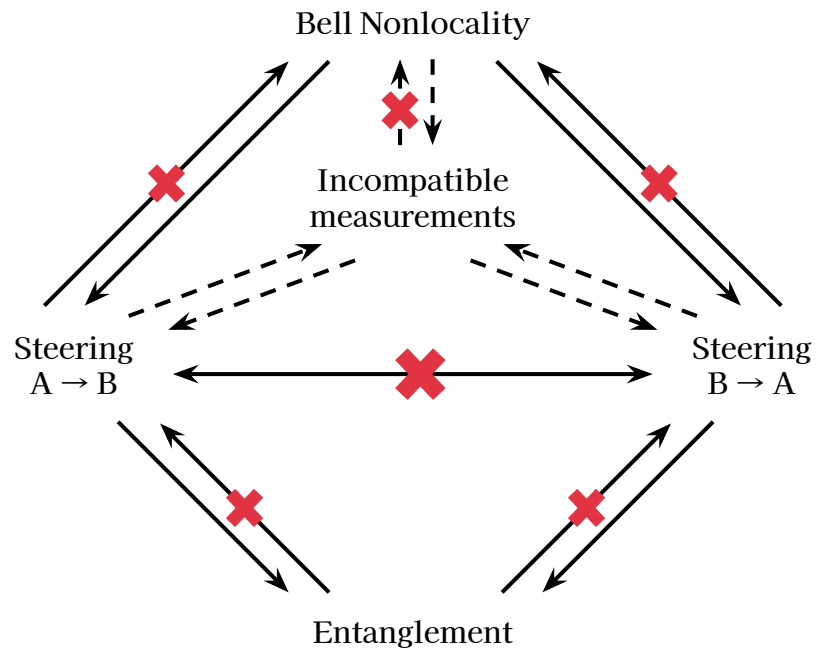


Figure 3.1: **Hierarchical relationships between the properties.** Diagram of hierarchical relationships between the properties of Entanglement, Steering, Bell Non-locality and Incompatible measurements.



## PART II: BUILDING RESULTS

### CHAPTER 4

## CERTIFICATION OF LOCAL ENTANGLED STATES

---

Now that we've developed the necessary tools for our study, we can proceed to establish the methods required to certify entanglement and state locality. This entails conducting two distinct certification processes: one focused on entanglement and another centered on locality. Subsequently, we can merge the outcomes of these certifications to certify states that exhibit both entanglement and locality.

### 4.1 Entangled states certification

An interesting criterion to study in order to certify entangled states is the Positive Partial Transpose (PPT) criterion, presented in [Theorem 3](#) [29].

**Theorem 3 (PPT criterion)** *Let  $\rho_{AB}$  be a separable state of a bipartite quantum system, acting on  $\mathcal{H}_A \otimes \mathcal{H}_B$ . Then the transposed partial in subsystem B (or analogously for subsystem A) is positive semi-definite  $\rho^{T_B} = (\mathbb{I}_{d_A} \otimes T_B)[\rho] \geq 0$ . Therefore if  $\rho^{T_B} \leq 0$ , the state  $\rho_{AB}$  is entangled.*

The proof of this theorem is quite straightforward. If  $\rho$  is separable, it can be written as

$$\rho = \sum p_i \rho_i^A \otimes \rho_i^B.$$

In this case, the effect of the partial transposition is trivial, given by

$$\rho^{T_B} = (I \otimes T)(\rho) = \sum p_i \rho_i^A \otimes (\rho_i^B)^T.$$

As the transposition map preserves eigenvalues, the spectrum of  $(\rho_i^B)^T$  is the same as the spectrum of  $\rho_i^B$ , and in particular  $(\rho_i^B)^T$  must still be positive semidefinite. Thus  $\rho^{T_B}$  must also be positive semidefinite. This proves the necessity of the PPT criterion.

In arbitrary dimensions of the two subsystems, the PPT criterion only provides a necessary but not sufficient condition for separability, that is, in “high” dimensions there are

states that remain positive under partial transposition (PPT states) even if they are entangled. If, instead, the bipartite system has dimension  $d_A \cdot d_B \leq 6$ , that is, the bipartite system is of type  $2 \otimes 2$ ,  $2 \otimes 3$  or  $3 \otimes 2$ , then the criterion PPT provides a separability condition [30], given as follows.

**Theorem 4 (PPT criterion for  $d_A \cdot d_B \leq 6$ )** *Let  $\rho_{AB}$  be a state of a bipartite quantum system with dimension  $d_A \cdot d_B \leq 6$ , acting on  $\mathcal{H}_A \otimes \mathcal{H}_B$ . It will be separable if and only if the transposed partial in subsystem B is positive semi-definite  $\rho^{T_B} = (\mathbb{I}_{d_A} \otimes T_B)[\rho] \geq 0$ .*

To do the proof of this theorem and show that being PPT is also sufficient for the  $d_A \cdot d_B \leq 6$  cases is more involved. First, it was shown that for every entangled state there exists an entanglement witness [30]. This is a result of geometric nature and invokes the Hahn-Banack theorem.

From the existence of entanglement witnesses, one can show that  $I \otimes \Lambda(\rho)$  being positive for all positive maps  $\Lambda$  is a necessary and sufficient condition for the separability of  $\rho$ , where  $\Lambda$  maps the space of density matrices  $\mathcal{L}(\mathcal{H}_B)$  to  $\mathcal{L}(\mathcal{H}_A)$ .

Furthermore, every positive map from  $\mathcal{L}(\mathcal{H}_B)$  to  $\mathcal{L}(\mathcal{H}_A)$  can be decomposed into a sum of completely positive and completely copositive maps, when  $d_B = 2$  and  $d_A = 2$  or 3. In other words, every such map  $\Lambda$  can be written as

$$\Lambda = \Lambda_1 + \Lambda_2 \circ T,$$

where  $\Lambda_1$  and  $\Lambda_2$  are completely positive and  $T$  is the transposition map. This follows from the Størmer-Woronowicz theorem.

Loosely speaking, the transposition map is therefore the only one that can generate negative eigenvalues in these dimensions. So if  $\rho^{T_B}$  is positive,  $I \otimes \Lambda(\rho)$  is positive for any  $\Lambda$ . Thus we conclude that the PPT criterion is also sufficient for separability when  $d_A \cdot d_B \leq 6$ .

In higher dimensions, however, there exist maps that can't be decomposed in this fashion, and the criterion is no longer sufficient. Consequently, there are entangled states which have a positive partial transpose. For the higher dimension cases, we can study the DPS hierarchy, also known as k-symmetric PPT extension.

Let  $\rho_{AB}$  be the state of a bipartite quantum system, acting on  $\mathcal{H}_A \otimes \mathcal{H}_B$ . The state  $\rho_{AB}$  is said to admit a  $k$ -symmetric extension if there is a state  $\rho_{AB_1 \dots B_k}$  acting on  $\mathcal{H}_A \otimes \mathcal{H}_B^{\otimes k}$  such that the partial trace over any  $(k-1)$  copies of subsystem B is equal to  $\rho_{AB}$ :

$$\text{tr}_{B_{(1, \dots, k)/i}}(\rho_{AB_1 \dots B_k}) = \rho_{AB_i},$$

for all  $i \in \{1, \dots, k\}$ , where the notation  $B_{(1, \dots, k)/i}$  denotes all  $k$  parts  $B$ , except the part  $B_i$ .

If in addition to these conditions we have

$$\rho_{AB_1 \dots B_k}^{T_A} \geq 0 \quad \text{and} \quad \rho_{AB_1 \dots B_k}^{T_{B_1}} \geq 0,$$

$\rho_{AB}$  is said to admit a  $k$ -symmetrical PPT extension. Finally, we can present the [Theorem 5 \[31\]](#).

**Theorem 5 (Doherty-Parrilo-Spedalieri (DPS) Criterion)** *Let  $\rho_{AB}$  be the state of a bipartite quantum system, acting on  $\mathcal{H}_A \otimes \mathcal{H}_B$ . The state  $\rho_{AB}$  is separable if and only if it admits a  $k$ -symmetric PPT extension for all  $k$ .*

This theorem presents an entanglement detection criterion that allows the construction of a numerical algorithm of arbitrary precision using convex optimization methods [\[32\]](#).

## 4.2 Local states certification

To perform the certification of local states, we start by analyzing one of the theorems presented in [\[18\]](#).

**Theorem 6** *Let  $\mathcal{M}$  be a finite collection of projective measurements in  $\mathbb{C}^{d_A}$ . A state  $\rho_{AB}$  acting on  $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$  admits an LHS model for all projective measurements if there is a unit trace operator  $O_{AB}$  acting on the same Hilbert space, so that  $O_{AB}$  admits an LHS model for measurements in  $\mathcal{M}$ , and*

$$\rho_{AB} = r O_{AB} + (1 - r) \xi_A \otimes O_B, \quad (4.1)$$

where  $r$  is the radius of the inner sphere, the largest centric sphere contained in the polytope generated by  $\mathcal{M}$ , and  $\xi_A$  is an arbitrary density matrix (of dimension  $d_A \times d_A$ ).

To prove this theorem we have to show that

$$\text{tr}_A[(M_a^r \otimes \mathbb{I}_{d_B}) O_{AB}] = \text{tr}_A[(M_a \otimes \mathbb{I}_{d_B}) \rho_{AB}]. \quad (4.2)$$

Given that we are constructing a continuous set of shrunk measurements  $M_a^r = r M_a + (1 - r) \text{tr}[\xi_A M_a] \mathbb{I}_{d_A}$  for any  $M_a \in \mathcal{M}$ . Then  $r$  is exactly the largest number such that all  $M_a^r$  can be written as a convex combination of the elements of  $\{A_{a|x}\}$ , i.e.  $M_a^r = \sum_x p_x A_{a|x}$  with  $\sum_x p_x = 1$  and  $p_x \geq 0$ . The left side of [Equation 4.2](#), can be written

$$\begin{aligned} \text{tr}_A[(M_a^r \otimes \mathbb{I}_{d_B}) O_{AB}] &= \text{tr}_A[(r M_a + (1 - r) \text{tr}[\xi_A M_a] \mathbb{I}_{d_A}) \otimes \mathbb{I}_{d_B} O_{AB}] \\ &= \text{tr}_A[r(M_a \otimes \mathbb{I}_{d_B}) O_{AB} + ((1 - r) \text{tr}[\xi_A M_a] \mathbb{I}_{d_A}) \otimes \mathbb{I}_{d_B} O_{AB}] \\ &= \text{tr}_A[r(M_a \otimes \mathbb{I}_{d_B}) O_{AB}] + \text{tr}_A[((1 - r) \text{tr}[\xi_A M_a]) \mathbb{I}_{d_A} \otimes \mathbb{I}_{d_B} O_{AB}] \\ &= \text{tr}_A[r(M_a \otimes \mathbb{I}_{d_B}) O_{AB}] + \text{tr}_A[((1 - r) \text{tr}[\xi_A M_a]) \mathbb{I}_{d_A \times d_B} O_{AB}] \\ &= r(\text{tr}_A[(M_a \otimes \mathbb{I}_{d_B}) O_{AB}]) + ((1 - r) \text{tr}[\xi_A M_a]) \text{tr}_A[O_{AB}] \\ &= r(\text{tr}_A[(M_a \otimes \mathbb{I}_{d_B}) O_{AB}]) + ((1 - r) \text{tr}[\xi_A M_a]) O_B, \end{aligned} \quad (4.3)$$

and the right side can be written

$$\begin{aligned}
\text{tr}_A[(M_a \otimes \mathbb{I}_{d_B})\rho_{AB}] &= \text{tr}_A[(M_a \otimes \mathbb{I}_{d_B})(rO_{AB} + (1-r)\xi_A \otimes O_B)] \\
&= \text{tr}_A[r(M_a \otimes \mathbb{I}_{d_B})O_{AB} + (1-r)(M_a \otimes \mathbb{I}_{d_B})(\xi_A \otimes O_B)] \\
&= \text{tr}_A[r(M_a \otimes \mathbb{I}_{d_B})O_{AB}] + \text{tr}_A[(1-r)(M_a \xi_A) \otimes (\mathbb{I}_{d_B} O_B)] \\
&= r(\text{tr}_A[(M_a \otimes \mathbb{I}_{d_B})O_{AB}]) + ((1-r)\text{tr}[M_a \xi_A])O_B,
\end{aligned} \tag{4.4}$$

where we made use of the property  $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ , and also that  $\text{tr}_A[(1-r)(M_a \xi_A) \otimes O_B] = ((1-r)\text{tr}[M_a \xi_A])O_B$ , given that the partial trace operator can be defined invariantly. Combining Equation 4.3 and Equation 4.4 into Equation 4.2, we have

$$\begin{aligned}
r(\text{tr}_A[(M_a \otimes \mathbb{I}_{d_B})O_{AB}]) + ((1-r)\text{tr}[\xi_A M_a])O_B &= \\
r(\text{tr}_A[(M_a \otimes \mathbb{I}_{d_B})O_{AB}]) + ((1-r)\text{tr}[M_a \xi_A])O_B,
\end{aligned}$$

which is true, since  $\text{tr}[\xi_A M_a] = \text{tr}[M_a \xi_A]$ , which concludes the proof.

Using Theorem 6, we can define a method that provides a sequence of tests (sufficient conditions) that converge to the brute force test, in terms of the set  $\mathcal{M}$ , so that a state has a model LHS in the limit where  $\mathcal{M}$  is the infinite set of all projective measurements [18].

This method is based on a convex optimization program, with the semi-definite programming (SDP) technique, defined and discussed in Appendix A. Initially, we choose a finite set of measurements  $\mathcal{M}$  and calculate  $r$ , given by the distance between the closest facet of the polytope generated by  $\mathcal{M}$  and the origin (calculated by enumeration algorithms of default vertices). Since  $\mathcal{M}$  is finite, we can restrict it to a finite set of hidden variables by imposing an LHS model on the  $O_{AB}$  operator.

Without loss of generality, we consider  $\lambda = \lambda_1 \cdots \lambda_{m_A}$  a bit string of length  $m_A$ , which specifies a (deterministic) result for each of Alice's  $m_A$  measurements. When the measurement along the  $\hat{u}_x$  direction is performed, we get  $a = \lambda_x$ . If each measurement has  $k$  possible outcomes, there are  $d_{m_A} = k^{m_A}$  distinct deterministic specifications. Thus, according to Theorem 6, the following SDP tests an LHS model for projective measurements in state  $\rho_{AB}$ :

$$\begin{aligned}
&\text{given } \rho_{AB}, \mathcal{M}, r, \xi \\
&\text{find } O_{AB}, \{\rho_\lambda\}_\lambda \\
&\text{subject to } \text{tr}_A[(\Pi_{a|\hat{u}_x} \otimes \mathbb{I}_B)O_{AB}] = \sum_\lambda D_\lambda(a|x)\rho_\lambda, \quad \forall a, x \\
&\quad \rho_\lambda \geq 0, \quad \forall \lambda \\
&\quad rO_{AB} + (1-r)\xi \otimes O_B = \rho_{AB},
\end{aligned}$$

where  $D_\lambda(a|x) = \delta_{a,\lambda_x}$  are the deterministic response functions.

If we change the format of our state of interest to

$$\rho_{AB} = \rho_q = q\rho + (1 - q)\rho_{\text{sep}},$$

with  $0 \leq q \leq 1$ , and where state  $\rho$  can be called target state and  $\rho_{\text{sep}}$  is a separable state (thus unsteerable), we can change the SDP to find the maximum value of  $q$  such that  $\rho_q$  admits an LHS model. The changed SDP is as follows

$$\begin{aligned} & \text{maximize} && q \\ & \text{subject to} && \text{tr}_A[(\Pi_{a|\hat{v}_x} \otimes \mathbb{I}_B) O_{AB}] = \sum_{\lambda} D_{\lambda}(a|x) \rho_{\lambda}, \quad \forall a, x \\ & && \rho_{\lambda} \geq 0, \quad \forall \lambda \\ & && r O_{AB} + (1 - r) \xi \otimes O_B = \rho_q. \end{aligned}$$

When the  $q$  found is equal to 1, it means that it was not necessary to add noise to find the LHS model, that is, the state is unsteerable (local).

# CHAPTER 5

## SAMPLING QUANTUM STATES

---

The structure of the abstract space of theories is characterized in terms of its metrics. In order to understand what are the possible metrics that define the mathematical structure of quantum theory, we must understand what would be a natural measure of distance for it. In particular, two metrics are interesting: the Hilbert-Schmidt distance and the Bures metric. Some interesting references on this subject are [33–35].

### 5.1 Hilbert-Schmidt

Quantum tomography, also known as quantum state tomography, is a process used to reconstruct a quantum state by performing measurements on a collection of identical quantum states. Consider two quantum systems associated with a  $d$ -dimensional Hilbert space. To determine the proximity between two quantum states,  $\rho_1$  and  $\rho_2$ , a tomographically complete set of mutually complementary measurements  $\mathcal{M} = \{M_x\}$  is employed. These measurements are non-degenerate and orthogonal, allowing for the unique identification of the state. It is crucial for a proximity measure to be invariant under the specific choice of a complete set of mutually complementary measurements.

The measurement is performed independently and equivalently for each quantum system and its probability vector is denoted by  $\vec{p}_x(\rho)$  for the system  $\rho$ . The distance of the two probability vectors,  $\vec{p}_x(\rho_1)$  and  $\vec{p}_x(\rho_2)$ , is defined as,

$$D_x(\rho_1, \rho_2) = |\vec{p}_x(\rho_1) - \vec{p}_x(\rho_2)|^2,$$

where  $|\cdot|$  is a vector norm. The distance  $D_x$  is called the single operational distance for measurement  $M_x$  belonging to the complete set of mutually complementary measurements. The total operating distance can be defined by summing the individual operating distances over the complete set of complementary measurements:

$$D_{\text{total}}(\rho_1, \rho_2) = \sum_x D_x(\rho_1, \rho_2).$$

The present geometric approach leads to the result that the total operating dis-

tance is equivalent to the Hilbert-Schmidt distance, i.e.

$$D_{\text{total}}(\rho_1, \rho_2) = \|\rho_1 - \rho_2\|^2,$$

where  $\|\cdot\|$  is a Hilbert-Schmidt norm.

To generate a Hilbert-Schmidt random density matrix of dimension  $n$ , we take a square complex random matrix  $G$  of size  $n$  pertaining to Ginibre's ensemble, which is built with real and imaginary parts of each element being a random variable sampled from the normal distribution. Therefore, we can write the density matrix as

$$\rho_{HS} = \frac{GG^\dagger}{\text{tr}[GG^\dagger]}.$$

## 5.2 Bures

The fidelity serves as another measure of the proximity between two quantum states. It can be defined as the transition probability between two pure states  $\rho_1 = |\psi_1\rangle\langle\psi_1|$  and  $\rho_2 = |\psi_2\rangle\langle\psi_2|$ , as

$$F(\rho_1, \rho_2) = |\langle\psi_1|\psi_2\rangle|,$$

where  $F = 0$  when the states are orthogonal (perfectly distinguishable) and  $F = 1$  when  $|\psi_1\rangle = |\psi_2\rangle$ .

This definition can be generalized to the case where one of the states is pure  $\rho_1 = |\psi_1\rangle\langle\psi_1|$  and the other one is mixed  $\rho_2$ ,

$$F(\rho_1, \rho_2) = \sqrt{\langle\psi_1|\rho_2|\psi_1\rangle}.$$

In this case, the fidelity can be understood as the average value obtained from considering both pure states over an ensemble of pure states represented by a density matrix  $\rho_2$ .

When fidelity is extended to incorporate mixed states, we end up with

$$F(\rho_1, \rho_2) = \left( \text{tr} \sqrt{\sqrt{\rho_1} \rho_2 \sqrt{\rho_1}} \right)^2,$$

and its interpretation becomes vague from an operational perspective. Rather, fidelity can be indirectly interpreted in terms of the statistical distance or “statistical distinguishability” of finding the measure that optimally resolves neighboring density operators.

This measure can be used to define the Bures metric on the space of density matrices, by defining the Bures distance, given by

$$D_B(\rho_1, \rho_2) = \sqrt{2(1 - \sqrt{F(\rho_1, \rho_2)})}.$$

To generate a Bures random density matrix of dimension  $n$ , we take a square complex random matrix  $G$  of size  $n$  pertaining to the Ginibre ensemble, a random unitary matrix  $U$  distributed according to the Haar measure on  $U(n)$  and write the random density matrix

$$\rho_B = \frac{(\mathbb{I} + U)GG^\dagger(\mathbb{I} + U^\dagger)}{\text{tr}[(\mathbb{I} + U)GG^\dagger(\mathbb{I} + U^\dagger)]}.$$



# CHAPTER 6

## CODE DOCUMENTATION AND VALIDATION

---

All results referring to this project were built numerically, using codes written in the Python language and stored in a GitHub digital repository with public access [36]. Next, I present the documentation and validation of the code, as well as their limitations and associated precisions.

### Entanglement certification

As we will be interested in states of dimensions  $d_A \cdot d_B \leq 6$ , we use the positive partial transpose criterion (PPT) to certify the entanglement of the states. Given the target state  $\rho$  as input, we are able to calculate the partial transpose and determine if the state is entangled or not. We can also find the entanglement threshold  $\alpha_{\text{ent}}$  for each state  $\rho_\alpha = \alpha\rho + (1 - \alpha)\rho_{\text{sep}}$  to become separable. The values given by this method have a precision of  $10^{-16}$ .

### Generating the quantum states

We perform the generation of entangled quantum states according to Bures and Hilbert-Schmidt [37] metrics. For code validation, we first perform the certification of the mixed-state density matrix generation function by checking the normalization properties ( $\text{tr}(\rho) = 1$ ) and hermiticity ( $\rho = \rho^\dagger$ ) with an accuracy of up to  $10^{-15}$ . Next, we recovered the percentage of generation of separable states expected by geometrical arguments. We uniformly sampled  $10^5$  two-qubit states first with the Bures metric, which our method generated  $\approx 7.4\%$  (expected value =  $7.3\%$  [38]) and next with the Hilbert-Schmidt metric, which our method generates  $\approx 24.2\%$  (value expected =  $24.2\%$  [38]). We certified the separability of the states generated with the aforementioned code. On an Intel i5-7200U CPU, it took 37 seconds to create and certify the states generated according to the Bures metric, and 11 seconds for Hilbert-Schmidt metric.

## Generating the measurements

As we are interested in studying dichotomous projective measurements performed on qubits, we can use the format presented in [Equation 1.4](#) to generate our measurements. Thus, when assembling the set of measurements, we are determining a polytope in which each vertex will represent a measurement effect  $M_{\pm|x}$ . We will see that it will be interesting that this polytope is constructed in such a way as to maximize the value of the radius of the inscribed sphere. It is important to notice here that the inradius is not necessarily proportional to the number of vertices, for example, if the vertices are all concentrated on the poles, the inradius will be smaller than if they are well distributed.

The polytopes that will be used in this work can be seen in [Figure 6.1](#), the first has 6 vertices and an inradius of 0.577, the second has 18 vertices and an inradius of 0.816 and the third one has 26 vertices and an inradius of 0.863. The polytopes constructed here share vertices, meaning that the 18-vertex polytope contains the 6 vertices of the 3-measurements polytope, and the 26-vertex polytope contains the 18 vertices of the 9-measurements polytope.

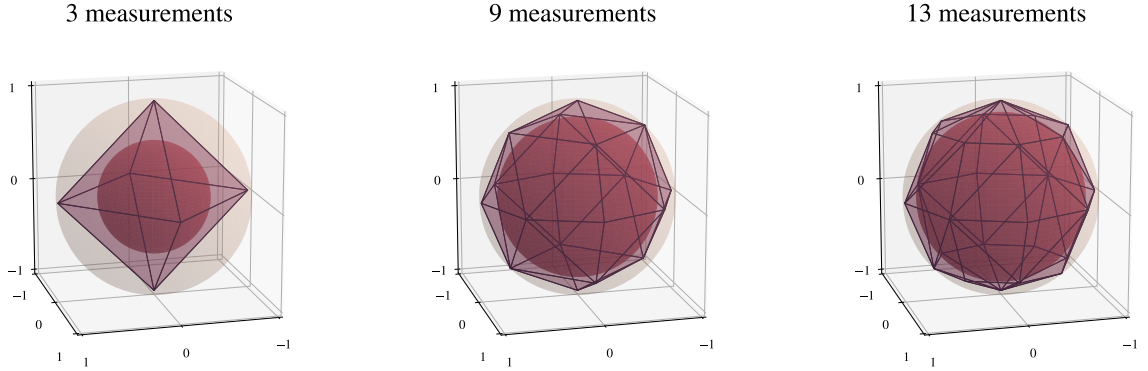


Figure 6.1: **Polytopes of dichotomic projective measurements.** In the left, we have the polytope that represent 3 measurements. In the middle, the polytope that represent 9 measurements. In the right, the polytope that represent 13 measurements.

## Locality certification

To do the locality certification, we constructed a code that follows the structure presented in [Figure 6.2](#).

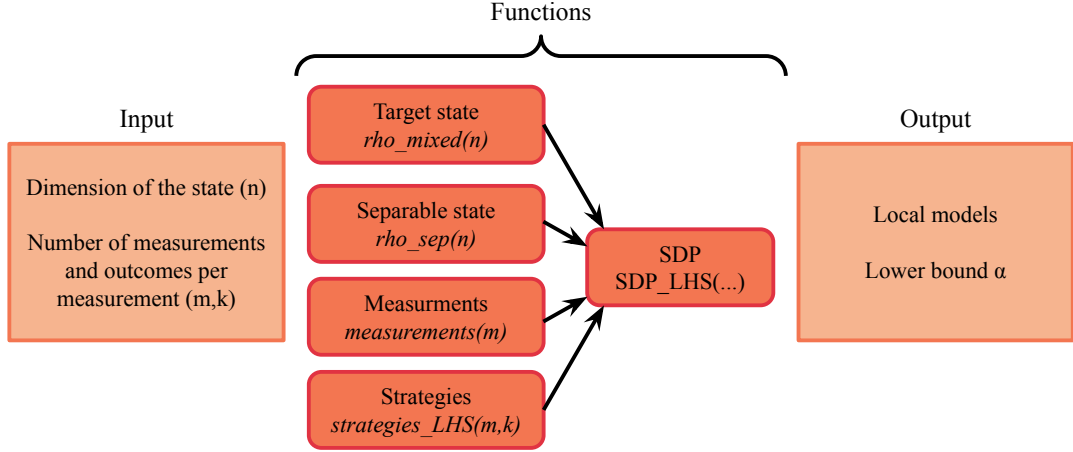


Figure 6.2: **Program structure.** We input the dimension of the state, 4 in the case of two qubits, the number of measurements per part and the number of results per measurement, 2 in the dichotomous case. The program will then provide the target and separable state, the measurements and the deterministic strategies of the scenario to feed the SDP which will return as output the lower bound of locality for the target state and the respective local model.

Looking back at the SDP presented in [section 4.2](#),

$$\begin{aligned}
 & \text{maximize} && q \\
 & \text{subject to} && \text{tr}_A[(\Pi_{a|\hat{v}_x} \otimes \mathbb{I}_B) O_{AB}] = \sum_{\lambda} D_{\lambda}(a|x) \rho_{\lambda}, \quad \forall a, x \\
 & && \rho_{\lambda} \geq 0, \quad \forall \lambda \\
 & && r O_{AB} + (1-r) \xi \otimes O_B = q \rho + (1-q) \rho_{\text{sep}},
 \end{aligned}$$

we see that it has input variables and optimization variables that are not predefined. Knowing that  $d_A$  is the dimension of Alice's part,  $d_B$  is the dimension of Bob's part,  $k$  is the number of results per measurement,  $m$  is the number of measurements, the dimensions of each variable is presented in [Table 6.1](#).

Table 6.1: Dimension of variables

Input variables						
Variable	$\rho, \rho_{\text{sep}}$	$\mathbb{I}_B$	$\xi$	$\{\Pi_{a \hat{v}_x}\}_{a,x}$	$r$	$\{D_{\lambda}(a x)\}_{\lambda,a,x}$
Dimension	$(d_A \cdot d_B, d_A \cdot d_B)$	$(d_B, d_B)$	$(d_A, d_A)$	$(k \cdot m, d_A \cdot d_A)$	1	$(k^m, k \cdot m)$
Optimization variables						
Variable	$q$	$O_{AB}$			$\{\rho_{\lambda}\}_{\lambda}$	
Dimension	1	$(d_A \cdot d_B, d_A \cdot d_B)$			$(k^m, d_B, d_B)$	

Analyzing variable by variable, we see that  $\{D_{\lambda}(a|x)\}_{\lambda,a,x}$  will be the most computationally costly of the input variables. This is because to build the strategies, we need to take in consideration that they are deterministic, so, given a result  $a$  of a measurement  $\hat{v}_x$ ,

the associated probability will be either 0 or 1. For each round of the experiment it is possible to determine all possible deterministic strategies, so we create  $k^m$  ( $k$  being the number of results per measurement and  $m$  the number of measurements) in the form

$$D_\lambda = (D_\lambda(a_1|\hat{v}_1), D_\lambda(a_2|\hat{v}_1), \dots, D_\lambda(a_k|\hat{v}_1), \dots, D_\lambda(b_k|\hat{v}_m)),$$

with  $\sum_{i=1}^k D_\lambda(a_i, \hat{v}_j) = 1$  for all  $1 \leq j \leq m$ .

For the optimization variables,  $\{\rho_\lambda\}_\lambda$  will be the most computationally costly. This happens because for each  $\lambda$ , we create a local state of Bob's state dimension, that is,  $k^m$  states of dimension  $d_B$ .

By fixing the number of measurements and the number of results per measurement, we find that the spatial complexity increases as a quadratic function of the dimension of the state, i.e.,  $\mathcal{O}(n^2)$ . If we set the number of measurements to  $m$  and the size of the state, the spatial complexity increases as a function of the number of results per measurement by a factor of  $\mathcal{O}(n^m)$ . Conversely, if we fix the number of results per measurement  $k$  and the dimension of the state, we see that the spatial complexity increases as a function of the number of results per measurement by a factor of  $\mathcal{O}(n \cdot k^n)$ .

## Machine Learning

To improve my results using the method presented above, I need to increase the number of measurements, which is computationally expensive. Therefore, it becomes interesting to study alternative techniques. Given the success of Machine Learning techniques on similar problems [39] we propose building a model for the certification of local entangled states.

For our purposes, we will employ the Multi-Layer Perceptron Neural Network, which is readily available in the popular Python library called Scikit-learn. Scikit-learn leverages the extensive capabilities of Python to offer cutting-edge implementations of numerous renowned machine learning algorithms. It ensures a user-friendly interface that seamlessly integrates with the Python language, facilitating a smooth and intuitive experience [40].

Through this method, we can effectively categorize the input quantum state into one of two potential output classes: entangled and local or not. To interpret the outcomes accurately, our analysis will primarily rely on the confusion matrix, presented in Table 6.2, where we use acronyms such as TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative). TP signifies cases correctly predicted as positive when they are indeed positive, while TN indicates cases correctly predicted as negative when they are indeed negative. Conversely, FP points to cases incorrectly predicted as positive when they are actually negative, and FN denotes cases incorrectly predicted as negative when they are actually positive.

Table 6.2: Confusion matrix

		Actual values	
		Positive (1)	Negative (0)
Predicted value	Positive (1)	TP	FP
	Negative (0)	FN	TN

In an ideal scenario, we anticipate that the data points will predominantly fall along the main diagonal of the matrix, indicating good performance. We can derive the following additional quantitative performance metrics from these values, based on the questions

**Precision** “How many of the selected items are relevant?”. Given by  $\frac{TP}{TP+FP}$ .

**Recall** “How many of the relevant items are selected?”. Given by  $\frac{TP}{TP+FN}$ .

**Accuracy** “What percentage of predictions were correct?”. Given by  $\frac{TP+TN}{TP+FP+TN+FN}$ .

This method is elaborated upon in [Appendix B](#), where you can also find a more precise definition of the performance metrics.

## Additional certification methods

In addition to the techniques presented in this document for locality certification, there are others already discussed in the literature. There is also Bell non-locality certification. Each technique has its own unique strengths, optimal target problems, and inherent limitations. In the following, we provide a brief overview of some of the notable state-of-the-art works in this particular research domain.

In a comprehensive review published in 2017 [41], the authors explored the characterization of quantum steering using semidefinite programming. This review presents efficient numerical methods for addressing various problems related to steering, including detection, quantification, and practical applications.

Another article from 2017 [42] combines two recently developed numerical methods to derive a result. By utilizing these methods, an analytical construction of the local hidden variable (LHV) model is achieved, employing a larger set of measurements. However, it should be noted that this result is specific to a particular class of entangled states and lacks general applicability. Generalizing this method to construct LHV models for other classes of entangled states, especially in higher-dimensional Hilbert spaces, has yet to be explored. Additionally, the adaptation of this technique to construct local hidden state (LHS) models specifically for studying steering was not discussed in the article.

In a 2023 article [43], researchers examined the limiting value for the non-locality of two-qubit Werner states under projective measurements. This approach is not limited to

the specific case considered but can be applied to scenarios involving any number of parties and inputs (uniform for all parties) with only two outputs. The lower bound derived in this work is entirely analytic and provides a tight constraint on the value of the Grothendieck constant of order three. The authors employed a convex optimization algorithm called Frank-Wolfe, which seeks to find the vertex of the polytope closest to the point of interest at each iteration.

In a recent 2023 paper [44], the authors introduced a confidence polytope. This polytope, taking the form of a hyperoctahedron, serves as a valuable tool for determining if a quantum state can be described by a local hidden state model. By verifying the absence of steering and establishing Bell-locality, the authors provide a computationally efficient method for this verification process. Remarkably, the authors demonstrate that assessing the locality of the polytope's vertices alone is sufficient to establish the overall locality of the confidence polytope.

# PART III: LOOKING AT THE FINAL WORK

## CHAPTER 7

### THE RESULTS

---

With the concepts and methods at our disposal, we can now embark on the certification processes and engage in a comprehensive discussion of the insights derived from the results. Our primary focus will be on the examination of two-qubit states and dichotomous projective measurements.

#### 7.1 Werner states

The first result we sought was to characterize the entanglement and locality limits for a class of states already analytically defined, in order to perform code verification. We use the 2-qubit Werner state class

$$\rho_W(\alpha) = \alpha|\psi^-\rangle\langle\psi^-| + (1-\alpha)\mathbb{I}/4, \quad 0 \leq \alpha \leq 1, \quad (7.1)$$

with  $|\psi^-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$ .

After the entanglement certification, we recovered  $\alpha_{\text{ent}} = 1/3$  (expected value =  $1/3$  [25]), with a precision of  $10^{-16}$ . As for the locality certification, we were able to reach a numerical limit of  $\alpha_{\text{local}} = 0.44$  with 13 dichotomic measurements (insphere radius of 0.86). When comparing to the analytical limit of  $1/2$  [25], our method already recaptures  $\approx 88\%$  of local Werner states.

Moreover, we run the code with the pentakis dodecahedron polytope with 32 vertices (representing 16 dichotomous measurements), with an internal radius of 0.92. In this case, the vertices do not include all the vertices of the 13-measurement polytope. We were able to reach a numerical limit of  $\alpha_{\text{local}} = 0.47$ , recapturing  $\approx 94\%$  of local Werner states.

When running the code, the choice of measurements itself will be important to the performance of the code. Therefore, when we use the polytopes that contain the vertices of the smaller polytope, we guarantee that the improvement of the values found was given by the increase in the inscribed radius and we have a more reliable visualization of the inscribed

radius and lower bound relationship.

When analyzing the graph presented in Figure 7.1, we see that the radius of the inscribed sphere is crucial for approaching the expected value. However, the number of vertices in the polytope also defines the time it takes to run the code. That is exactly why we had to perform the optimization to find the polytopes with the smallest number of vertices and the largest associated inscribed radius.

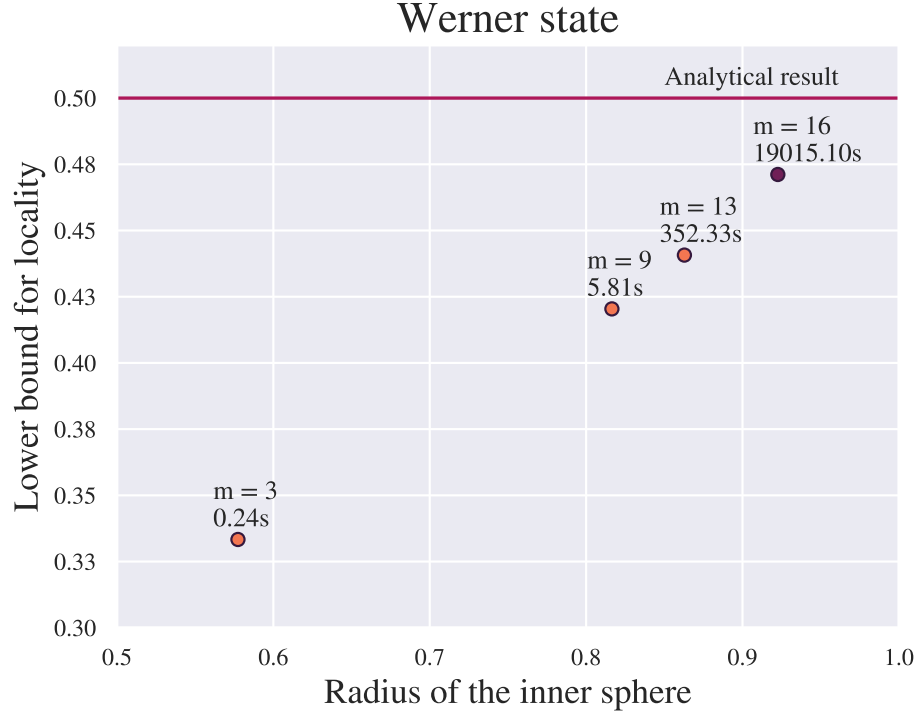


Figure 7.1: **Werner State's lower bounds for locality as a function of insphere radius.** We used the locality certification method by SDP to calculate the lower bounds for the Werner state, referring to the different measurement polytopes used (3, 9, 13, and 16 measurements). The polytope of 16 measurements is presented in a different color, as it does not possess the vertices of the smaller polytopes. In the graph, the continuous line at 0.5 represents the known analytical result of the locality threshold for the Werner state. Code runtimes accounted for on an Intel Xeon CPU E5-2630 v4.

This initial result serves as validation for the selected methods employed in our certification process. It is crucial to bear in mind that, despite being able to analytically characterize this specific family, the challenge of determining whether a general quantum state can produce local statistics is an exceptionally challenging problem. The prospect of obtaining a systematic analytical tool to answer this question is exceedingly slim. Therefore, our numerical method provides a partial yet significant solution to this complex set of issues.

## 7.2 Relative volume

Using this method, we can determine a lower bound for the volume of the set of local entangled states. We use a set of 9 dichotomous measures (radius 0.82) to certify which states



are entangled and Bell local. We selected this particular polytope due to its favorable characteristics, as demonstrated in our previous result. It exhibits an optimal balance between the number of vertices, inscribed radius, and the computational time required.

We started by generating  $10^4$  two-qubit states uniformly according to the Hilbert-Schmidt metric. We find that approximately 35.3% of the entangled states are local. Then, we generated  $10^4$  two-qubit states uniformly according to the Bures metric. We find that approximately 11.6% of the entangled states are local. The charts shown in Figure 7.2 summarize the relative volumes found for each metric.

In a related study [18], the authors use this same technique but with the icosahedron polytope (radius 0.79). They find that approximately 25% of the  $2 \times 10^4$  entangled states sampled according to the Hilbert-Schmidt measure are local, while using the Bures measure there was 7%. We have already observed an increase for the lower bound of the volume of local entangled states, given by the increase in the inscribed radius in this project.

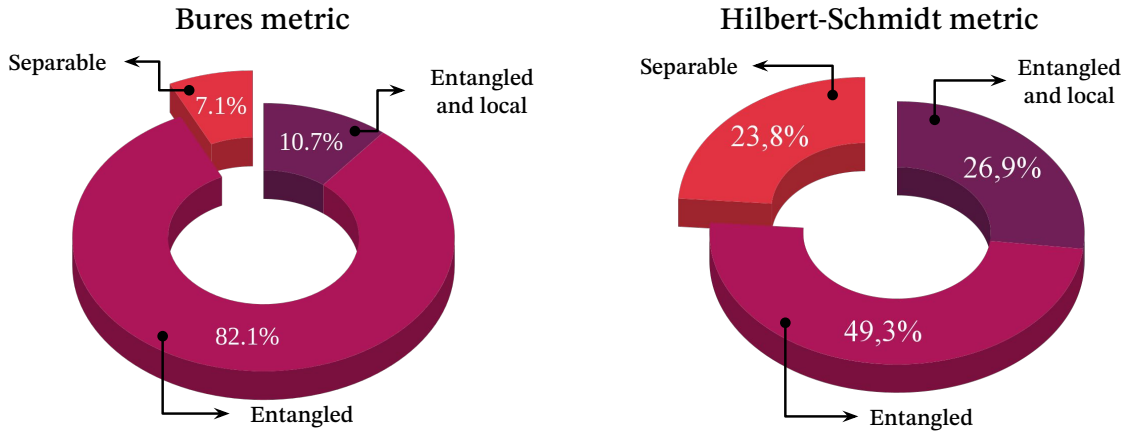


Figure 7.2: **Relative Volumes.** In the left, we have the states generated by Bures metric. We have successfully certified that 7.2% states are separable. Additionally, we can confirm that at least 10.8% of the states (or 11.6% of the entangled states) exhibit entanglement and satisfy the condition of locality. In the right, we have the states generated by Hilbert-Schmidt metric. We have achieved successful certification indicating that 23.8% of the states are separable. Furthermore, we can confidently affirm that a minimum of 26.9% of the states (or 35.3% of the entangled states) demonstrate entanglement while satisfying the condition of locality.

### 7.3 Machine Learning

In the previous result, we generated a dataset of quantum states and employed entanglement and locality certifications to classify them. This dataset is well-suited for serving as input to train and evaluate a Multilayer Perceptron model. Initially, we generated 5000 states using the Bures metric. However, in order to maintain a balanced dataset with an equal number of states for each classification, we reduced this number to 1034.

For the dataset partition, 70% of the states were utilized for training, while the remaining 30% were used for testing. The confusion matrix and performance metrics (pre-

sented in [Appendix B](#)) are presented in [Table 7.1](#) and [Table 7.2](#), respectively. After performing cross-validation with a 5-fold process, we were able to recover a mean accuracy of  $0.81 \pm 0.03$ .

Table 7.1: Confusion matrix (Bures metric)

	Positive (1)	Negative (0)
Positive (1)	135	29
Negative (0)	32	115

Table 7.2: Performance metrics (Bures metric)

Precision	Recall	Accuracy
0.81	0.82	0.80

The second dataset initially consisted of 5000 states generated using the Hilbert-Schmidt metric. After balancing the dataset, the total number of states decreased to 2664. Among these, 70% were allocated for training the model, while the remaining 30% were reserved for testing. The evaluation of the model's performance is presented through a confusion matrix in [Table 7.3](#), illustrating the classification results. Additionally, [Table 7.4](#) provides the performance metrics, offering a comprehensive analysis of the model's accuracy, precision and recall. After conducting a 5-fold cross-validation process, we achieved a mean accuracy score of  $0.66 \pm 0.03$ .

Table 7.3: Confusion matrix (Hilbert-Schmidt metric)

	Positive (1)	Negative (0)
Positive (1)	274	128
Negative (0)	143	255

Table 7.4: Performance metrics (Hilbert-Schmidt metric)

Precision	Recall	Accuracy
0.66	0.68	0.66

Upon comparing the results of the different datasets, an interesting observation emerges: the model exhibits better performance when applied to the states generated by the Bures metric. This finding may appear counterintuitive, considering that the dataset based on the Hilbert-Schmidt metric was larger, and one might expect the neural network to learn more effectively with a larger number of examples.

In addition to the individual dataset analyses, we also generated a third dataset by combining the previous two. The classification outcomes are depicted in the confusion

matrix shown in Table 7.5, while the comprehensive results of the performance metrics are presented in Table 7.6. The mean accuracy recovered after cross-validation with a 5-fold process was  $0.707 \pm 0.009$ .

Table 7.5: Confusion matrix (Both metrics)

	Positive (1)	Negative (0)
Positive (1)	384	169
Negative (0)	151	406

Table 7.6: Performance metrics (Both metric)

Precision	Recall	Accuracy
0.72	0.69	0.71

Across all three datasets, the low standard deviation of the accuracy after cross-validation provides strong evidence that the model is capable of accurately classifying new data. This indicates that the model has not suffered from overfitting or selection bias, and it is expected to generalize well to unseen data.

In this study, we opted to use a fixed model, specifically the Multilayer Perceptron, due to its simplicity and ease of implementation. However, it is worth noting that there are automated methods available in the field of machine learning that can help identify the best model for a given dataset. One such method is the Tree-based Pipeline Optimization Tool (TPOT) [45], a Python Automated Machine Learning tool that leverages genetic programming to optimize machine learning models. Using the TPOT, we found that there are models that lead our data to higher accuracy, as shown in Table 7.7. Interestingly, in the case of the Hilbert-Schmidt dataset, our model outperformed the one suggested by the automated machine learning (AutoML) process.

Table 7.7: Accuracy of the TPOT-discovered models

Bures metric	Hilbert-Schmidt metric	Both metrics
0.894	0.655	0.823

Using the Multilayer Perceptron (MLP) model and training it with  $10^4$  states generated by Bures, achieving an accuracy of 0.88, we successfully estimated a new relative volume. The estimation was performed using  $10^5$  input states, among which 42% were classified as entangled and local. The certification was solely based on the machine learning model without the need for additional certification methods. The prediction process, carried out on an Intel Xeon CPU 2.20GHz, took 0.17 seconds, highlighting the efficiency of our approach.

We observed a significant increase in the relative volume compared to the result obtained through convex optimization, which was previously recorded at 10.8%. It is important to approach this result with caution, considering the following aspects. Firstly, the previous technique provided certification of locality and lower bound values for the relative volume, whereas our machine learning approach estimates state classifications through data analysis. Additionally, our model is still in its initial stages and currently exhibits relatively low accuracy. For instance, there are opportunities to enhance the model by expanding the training dataset, implementing the model recommended by auto ML, increasing the number of hidden layers, neurons, and iterations of backpropagation, among other potential modifications. These adjustments hold the potential to further enhance the performance and accuracy of the method.

## CHAPTER 8

# DISCUSSIONS, CONCLUSIONS AND NEXT STEPS

---

Following extensive theoretical investigations into the fundamental concepts of quantum states and measurements, including their interrelationships and inherent properties, we embarked on the challenging task of devising a framework for entanglement and locality certification applicable to arbitrary two-qubit states subjected to projective dichotomous measurements. Subsequently, we constructed two distinct numerical methodologies: one grounded in convex optimization and the other rooted in machine learning techniques.

Our analysis uncovered distinct advantages and limitations associated with each method. Notably, the SDP method necessitates the initial definition of a set of measurements used to construct the polytope, an inherently critical aspect of the program. Conversely, the machine learning method does not rely on the definition of measurements but heavily depends on the database, composed exclusively of certified states. The accuracy of the machine learning model is significantly influenced by the composition of this dataset.

Upon examining the results for the lower bound of locality in the Werner state, we observed that while the SDP method delivers high accuracy for individual two-qubit states, it incurs substantial computational costs in terms of memory usage and processing time. Consequently, for a large number of states, the SDP method becomes impractical quickly. In contrast, the machine learning method offers computational cost advantages in both memory and time, serving as a viable alternative for handling numerous states when provided with an appropriate dataset. Nevertheless, with the SDP method, we successfully investigated and certified the locality of a broad spectrum of entangled quantum states, leading to improved bounds on the volume of two-qubit states featuring both entanglement and locality under the Bures and Hilbert-Schmidt metrics employed for uniform state generation.

It is essential to acknowledge that the SDP method faces scalability limitations concerning the dimension of the states under consideration. Consequently, it is pertinent to explore avenues for enhancing the methodology's scalability. This exploration might involve investigating alternative optimization techniques capable of handling higher-dimensional states or potentially devising hybrid methods that strike a balance between accuracy and computational efficiency. By pursuing such avenues, we could significantly bolster the scalability of our approach, rendering it applicable to a broader array of quantum problems and scenarios.

# BIBLIOGRAPHY

- [1] Nielsen, M. & Chuang, I. *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010). URL <https://books.google.com.br/books?id=-s4DEy7o-a0C>.
- [2] Rabelo, R. Não-localidade quântica: matemática e fundamentos. *Universidade Federal de Minas Gerais* (2010).
- [3] Dirac, P. A. M. A new notation for quantum mechanics. *Mathematical Proceedings of the Cambridge Philosophical Society* **35**, 416–418 (1939).
- [4] Schmidt, E. Zur theorie der linearen und nichtlinearen integralgleichungen. i. teil: Entwicklung willkürlicher funktionen nach systemen vorgeschriebener. *Mathematische Annalen* **63**, 433–476 (1907). URL <http://eudml.org/doc/158296>. [English translation: G.W. Stewart, “FREDHOLM, HILBERT, SCHMIDT: Three Fundamental Papers on Integral Equations Translated with commentary by G. W. Stewart”, <http://www.cs.umd.edu/~stewart>.].
- [5] Schrödinger, E. Die gegenwärtige situation in der quantenmechanik. *Naturwissenschaften* **23**, 807–812 (1935). [English translation: J.D. Trimmer, “The present situation in quantum mechanics: a translation of Schrödinger’s “cat paradox” paper”, *Proc. of American Philosophical Society* **124**, 323-338 (1980).].
- [6] Everett, H. Relative state” formulation of quantum mechanics. *Rev. Mod. Phys.* **29**, 454–462 (1957). URL <https://link.aps.org/doi/10.1103/RevModPhys.29.454>.
- [7] Ekert, A. & Knight, P. L. Entangled quantum systems and the Schmidt decomposition. *American Journal of Physics* **63**, 415–423 (1995). URL <https://doi.org/10.1119/1.17904>. [https://pubs.aip.org/aapt/ajp/article-pdf/63/5/415/11806926/415\\_1\\_online.pdf](https://pubs.aip.org/aapt/ajp/article-pdf/63/5/415/11806926/415_1_online.pdf).
- [8] Heinosaari, T., Miyadera, T. & Ziman, M. An invitation to quantum incompatibility. *Journal of Physics A: Mathematical and Theoretical* **49**, 123001 (2016). URL <http://dx.doi.org/10.1088/1751-8113/49/12/123001>.
- [9] Heinosaari, T. Quantum incompatibility from the viewpoint of entanglement theory. *Journal of Physics: Conference Series* **1638**, 012002 (2020). URL <http://dx.doi.org/10.1088/1742-6596/1638/1/012002>.
- [10] Heinosaari, T. & Wolf, M. M. Nondisturbing quantum measurements. *Journal of Mathematical Physics* **51**, 092201 (2010). URL <https://doi.org/10.1063/1.3480658>. <https://doi.org/10.1063/1.3480658>.

- [11] Gühne, O., Haapasalo, E., Kraft, T., Pellonpää, J.-P. & Uola, R. Colloquium: Incompatible measurements in quantum information science. *Rev. Mod. Phys.* **95**, 011003 (2023). URL <https://link.aps.org/doi/10.1103/RevModPhys.95.011003>.
- [12] Zhu, H. Information complementarity: A new paradigm for decoding quantum incompatibility. *Scientific Reports* **5** (2015). URL <http://dx.doi.org/10.1038/srep14317>.
- [13] Reeb, D., Reitzner, D. & Wolf, M. M. Coexistence does not imply joint measurability. *Journal of Physics A: Mathematical and Theoretical* **46**, 462002 (2013). URL <https://dx.doi.org/10.1088/1751-8113/46/46/462002>.
- [14] Heinosaari, T. Simultaneous measurement of two quantum observables: Compatibility, broadcasting, and in-between. *Phys. Rev. A* **93**, 042118 (2016). URL <https://link.aps.org/doi/10.1103/PhysRevA.93.042118>.
- [15] Carmeli, C., Heinosaari, T. & Toigo, A. Quantum incompatibility witnesses. *Phys. Rev. Lett.* **122**, 130402 (2019). URL <https://link.aps.org/doi/10.1103/PhysRevLett.122.130402>.
- [16] Einstein, A., Podolsky, B. & Rosen, N. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, 777–780 (1935). URL <https://link.aps.org/doi/10.1103/PhysRev.47.777>.
- [17] Bell, J. S. On the einstein podolsky rosen paradox. *Physics Physique Fizika* **1**, 195–200 (1964). URL <https://link.aps.org/doi/10.1103/PhysicsPhysiqueFizika.1.195>.
- [18] Cavalcanti, D., Guerini, L., Rabelo, R. & Skrzypczyk, P. General method for constructing local hidden variable models for entangled quantum states. *Physical Review Letters* **117** (2016). URL <http://dx.doi.org/10.1103/PhysRevLett.117.190401>.
- [19] Wiseman, H. M., Jones, S. J. & Doherty, A. C. Steering, entanglement, nonlocality, and the einstein-podolsky-rosen paradox. *Phys. Rev. Lett.* **98**, 140402 (2007). URL <https://link.aps.org/doi/10.1103/PhysRevLett.98.140402>.
- [20] Quintino, M. T., Vértesi, T. & Brunner, N. Joint measurability, einstein-podolsky-rosen steering, and bell nonlocality. *Phys. Rev. Lett.* **113**, 160402 (2014). URL <https://link.aps.org/doi/10.1103/PhysRevLett.113.160402>.
- [21] Quintino, M. T. *et al.* Inequivalence of entanglement, steering, and bell nonlocality for general measurements. *Physical Review A* **92** (2015). URL <http://dx.doi.org/10.1103/PhysRevA.92.032107>.

- [22] Uola, R., Budroni, C., Gühne, O. & Pellonpää, J.-P. One-to-one mapping between steering and joint measurability problems. *Phys. Rev. Lett.* **115**, 230402 (2015). URL <https://link.aps.org/doi/10.1103/PhysRevLett.115.230402>.
- [23] Schmid, D. *et al.* Understanding the interplay of entanglement and nonlocality: motivating and developing a new branch of entanglement theory (2021). [2004.09194](https://arxiv.org/abs/2004.09194).
- [24] Brunner, N., Cavalcanti, D., Pironio, S., Scarani, V. & Wehner, S. Bell nonlocality. *Reviews of Modern Physics* **86**, 419–478 (2014). URL <https://doi.org/10.1103/RevModPhys.86.419>.
- [25] Werner, R. F. Quantum states with einstein-podolsky-rosen correlations admitting a hidden-variable model. *Phys. Rev. A* **40**, 4277–4281 (1989). URL <https://link.aps.org/doi/10.1103/PhysRevA.40.4277>.
- [26] Bowles, J., Vértesi, T., Quintino, M. T. & Brunner, N. One-way einstein-podolsky-rosen steering. *Phys. Rev. Lett.* **112**, 200402 (2014). URL <https://link.aps.org/doi/10.1103/PhysRevLett.112.200402>.
- [27] Wolf, M. M., Perez-Garcia, D. & Fernandez, C. Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory. *Phys. Rev. Lett.* **103**, 230402 (2009). URL <https://link.aps.org/doi/10.1103/PhysRevLett.103.230402>.
- [28] Hirsch, F., Quintino, M. T. & Brunner, N. Quantum measurement incompatibility does not imply bell nonlocality. *Physical Review A* **97** (2018). URL <https://doi.org/10.1103/PhysRevA.97.012129>.
- [29] Peres, A. Separability criterion for density matrices. *Phys. Rev. Lett.* **77**, 1413–1415 (1996). URL <https://link.aps.org/doi/10.1103/PhysRevLett.77.1413>.
- [30] Horodecki, M., Horodecki, P. & Horodecki, R. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A* **223**, 1–8 (1996). URL <https://www.sciencedirect.com/science/article/pii/S0375960196007062>.
- [31] Doherty, A., Parrilo, P. & Spedalieri, F. Complete family of separability criteria. *Physical Review A* **69** (2004).
- [32] Navascués, M., Owari, M. & Plenio, M. B. Power of symmetric extensions for entanglement detection. *Phys. Rev. A* **80**, 052306 (2009). URL <https://link.aps.org/doi/10.1103/PhysRevA.80.052306>.
- [33] Bengtsson, I. & Życzkowski, K. *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, 2017). URL <https://books.google.com.br/books?id=sYswDwAAQBAJ>.



- [34] Osipov, V., Sommers, H.-J. & Życzkowski, K. Random bures mixed states and the distribution of their purity. *Journal of Physics A Mathematical and Theoretical* **43** (2009).
- [35] Życzkowski, K., Penson, K. A., Nechita, I. & Collins, B. Generating random density matrices. *Journal of Mathematical Physics* **52**, 062201 (2011). URL <https://doi.org/10.1063/1.3595693>. <https://doi.org/10.1063/1.3595693>.
- [36] Ruiz, G. Github - digital repository. [https://github.com/BiduRuiz/SDP\\_LocalModels](https://github.com/BiduRuiz/SDP_LocalModels) (2023).
- [37] Miszczak, J. Generating and using truly random quantum states in mathematica. *Computer Physics Communications* **183** (2011).
- [38] Slater, P. B. Two-qubit separability probabilities and beta functions. *Phys. Rev. A* **75**, 032326 (2007). URL <https://link.aps.org/doi/10.1103/PhysRevA.75.032326>.
- [39] Canabarro, A., Brito, S. & Chaves, R. Machine learning nonlocal correlations. *Phys. Rev. Lett.* **122**, 200401 (2019). URL <https://link.aps.org/doi/10.1103/PhysRevLett.122.200401>.
- [40] Pedregosa, F. *et al.* Scikit-learn: Machine learning in python (2018). [1201.0490](https://arxiv.org/abs/1201.0490).
- [41] Cavalcanti, D. & Skrzypczyk, P. Quantum steering: a review with focus on semidefinite programming. *Reports on Progress in Physics* **80**, 024001 (2016). URL <https://dx.doi.org/10.1088/1361-6633/80/2/024001>.
- [42] Hirsch, F., Quintino, M. T., Vértesi, T., Navascués, M. & Brunner, N. Better local hidden variable models for two-qubit Werner states and an upper bound on the Grothendieck constant  $K_G(3)$ . *Quantum* **1**, 3 (2017). URL <https://doi.org/10.22331/q-2017-04-25-3>.
- [43] Designolle, S. *et al.* Improved local models and new bell inequalities via frank-wolfe algorithms. *arXiv preprint arXiv:2302.04721* (2023).
- [44] Steinberg, J., Nguyen, H. C. & Kleinmann, M. Certifying activation of quantum correlations with finite data (2023). [2305.03748](https://arxiv.org/abs/2305.03748).
- [45] Le, T. T., Fu, W. & Moore, J. H. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. *Bioinformatics* **36**, 250–256 (2020).
- [46] Boyd, S., Vandenberghe, L. & Press, C. U. *Convex Optimization*. No. pt. 1 in Berichte über verteilte messsysteme (Cambridge University Press, 2004). URL <https://books.google.com.br/books?id=mYm0bLd3fcoC>.

- 
- [47] Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual (2023). URL <https://www.gurobi.com>.
  - [48] ApS, M. *The MOSEK Optimizer API for Python. Version 10.0.* (2019). URL <https://docs.mosek.com/latest/pythonapi/index.html>.
  - [49] Sagnol, G. & Stahlberg, M. PICOS: A Python interface to conic optimization solvers. *Journal of Open Source Software* **7**, 3915 (2022).
  - [50] Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016). <http://www.deeplearningbook.org>.
  - [51] Fushiki, T. Estimation of prediction error by using k-fold cross-validation. *Statistics and Computing* **21**, 137–146 (2011).

# APPENDIX A

## CONVEX OPTIMIZATION

---

Optimization is a field that lies at the intersection of applied mathematics and computer science. Its main aim is to study and apply algorithms to solve optimization problems. Solving an optimization problem entails finding the best possible value for the parameters of a system given a goal (for example, minimizing or maximizing a function).

The convexity of the problems we are interested in has the convenient property that, in many cases, we can use the powerful tools of convex optimization. The main reference used for the construction of this appendix was the book *Convex Optimization*, by Stephen Boyd and Lieven Vandenberghe [46].

To understand this important tool, we must first establish the mathematical definitions that lead us to it.

### Line and line segment

Assuming  $x_1 \neq x_2$  two points in  $\mathbf{R}^n$ . Points of the form

$$y = \theta x_1 + (1 - \theta)x_2,$$

where  $\theta \in \mathbf{R}$ , forms the line passing through  $x_1$  and  $x_2$ . Values of the  $\theta$  parameter between 0 and 1 correspond to the (closed) line segment between  $x_1$  and  $x_2$ . We can visualize both definitions in [Figure A.1](#).

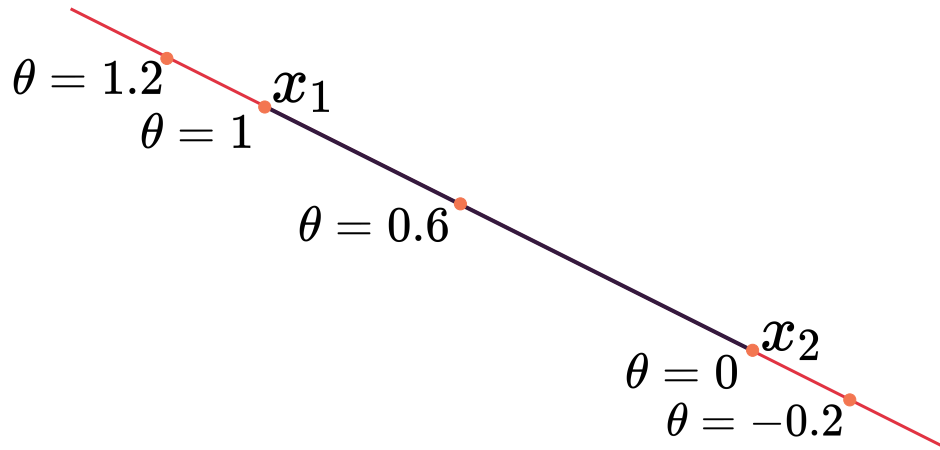


Figure A.1: **Line and line segment.** The line passing through  $x_1$  and  $x_2$  is shown lighter than the line segment between  $x_1$  and  $x_2$ .

## Affine Sets

A set  $C \subset \mathbf{R}^n$  is affine if the line through any two distinct points in  $C$  lies in  $C$ , i.e., if for any  $x_1, x_2 \in C$  and  $\theta \in \mathbf{R}$ , we have  $\theta x_1 + (1 - \theta)x_2 \in C$ . In other words,  $C$  contains the linear combination of any two points in  $C$ , provided the coefficients in the linear combination sum to one. This idea can be generalized to more than two points.

## Convex Sets

A set  $C$  is convex if the line segment between any two points in  $C$  is in  $C$ , i.e., if for any  $x_1, x_2 \in C$  and any  $\theta$  with  $0 \leq \theta \leq 1$ , we have  $\theta x_1 + (1 - \theta)x_2 \in C$ .

Every affine set is also convex, since it contains the entire line between any two distinct points in it, and therefore also the line segment between the points.

We refer to a point of the form  $\theta_1 x_1 + \dots + \theta_k x_k$ , where  $\theta_1 + \dots + \theta_k = 1$  and  $\theta_i \geq 0$ ,  $i = 1, \dots, k$ , as a convex combination of the points  $x_1, \dots, x_k$ . A set can be shown to be convex if and only if it contains all convex combinations of its points. We can visualize examples of convex and nonconvex sets in [Figure A.2](#), given in [46].

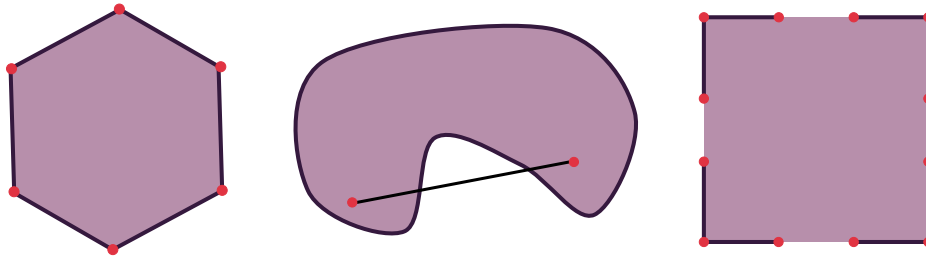


Figure A.2: **Convex and nonconvex sets.** On the left we have a hexagon, which includes its boundary (shown darker), is convex. On the middle we have a kidney shaped set is not convex, since the line segment between the two points in the set shown as dots is not contained in the set. On the right we have a square that contains some boundary points but not others, and, therefore, is not convex. Image based on figure 2.2 of book [46].

## Convex Functions

A function  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  is convex if  $\text{dom } f$  is a convex set and for all  $x, y \in \text{dom } f$ , and  $\theta$  with  $0 \leq \theta \leq 1$ , we have

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y). \quad (\text{A.1})$$

Geometrically, this inequality means that the line segment between  $(x, f(x))$  and  $(y, f(y))$ , which is the chord from  $x$  to  $y$ , is above the graph of  $f$ . A function  $f$  is strictly convex if the strict inequality holds in Equation A.1 whenever  $x \neq y$  and  $0 < \theta < 1$ . We say that  $f$  is concave if  $-f$  is convex, and strictly concave if  $-f$  is strictly convex.

## Optimization problems

Optimization problems follow the format

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t} \quad & f_i(x) \leq b_i, \quad i = 1, \dots, m. \end{aligned} \quad (\text{A.2})$$

Here, the vector  $x = (x_1, \dots, x_n)$  is the optimizable variable of the problem, the function  $f_0 : \mathbf{R}^n \rightarrow \mathbf{R}$  is the objective function, the functions  $f_i : \mathbf{R}^n \rightarrow \mathbf{R}$ ,  $i = 1, \dots, m$ , are the constraint functions (inequalities), and the constants  $b_1, \dots, b_m$  are the limits of restrictions.

A vector  $x^*$  is said to be optimal, or a solution to the problem A.2, if it has the smallest objective value among all vectors satisfying the constraints: for any  $z$  with  $f_1(z) \leq$

$b_1, \dots, f_m(z) \leq b_m$ , we have  $f_0(z) \geq f_0(x^*)$ .

## Convex Optimization

A problem is said to be convex when the objective and constraint functions are convex, which means that they satisfy the inequality

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y)$$

for all  $x, y \in \mathbf{R}^n$  and all  $a, b \in \mathbf{R}$  with  $\alpha + \beta = 1$ ,  $\alpha \geq 0$ ,  $\beta \geq 0$ .

It can be shown that in a convex program, every local minimum is also a global minimum, and the set of all optimal points is convex. If the objective function is strictly convex, we can also say that there is only one optimal point. The first of these properties is closely related to the existence of efficient algorithms for this type of problem.

## Linear programs (LPs) and Semi-definite programs (SDPs)

Linear programs (LPs) and semi-definite programs (SDPs) are two particularly interesting cases of convex optimization problems. In linear programs, the objective function and constraint functions are linear, so the region  $C$  over which the optimization takes place is a convex polytope. Generally speaking, LPs can be written as:

$$\begin{aligned} \min_x \quad & c^T x \\ \text{such that} \quad & Ax \leq b, \\ & x \geq 0, \end{aligned} \tag{A.3}$$

where,  $x, b, c \in \mathbf{R}^n$ ,  $A$  is a matrix  $p \times n$ , where  $p$  is the number of linear constraints, and the inequality in [Equation A.3](#) denotes term-by-term inequality.

Semi-definite programs have a slightly more elaborate definition. Very superficially, the inequalities of the semi-definite program are all half-spaces in the space of Hermitian matrices. Then, its set  $M$  is the intersection of the cone of semi-defined matrices with a polytope. Generally speaking, SDPs can be written as

$$\begin{aligned} \min_{X \in \mathbb{S}^n} \quad & \text{tr}(CX) \\ \text{such that} \quad & \text{tr}(A_i X) \leq b_i, \quad i \in \{1, \dots, p\}, \\ & X \geq 0, \end{aligned} \tag{A.4}$$

where  $\mathbb{S}^n$  denotes the set of symmetric matrices  $n \times n$  (or, more generally, the Hermitian

matrices  $n \times n$ ),  $A_i$  and  $C$  are matrices  $n \times n$ , for all  $i$ ,  $b_i$  is scalar, for all  $i$ , and the inequality in (A.4) denotes that  $X$  is positive semi-definite.

Among the class of convex problems, linear programming is the smallest subset, and semi-definite programming lies between the two. They differ in the restrictions imposed on their respective structures. In convex optimization, we are only interested in minimizing or maximizing convex functions within an equally convex set of possible values for the variables. In linear programming, our restriction is even greater, since our objective function must be linear, and the feasible set is a convex polytope. Finally, in semi-definite programming, we are also restricted to linear functions but within an affine set, of which convex polytopes are a subset.

These types of problems can be solved efficiently, from a computational point of view, with a wide range of tools that have already been developed, such as the solvers Gurobi [47] and Mosek [48]. As we are using the Python language in this project, we make use of the Picos library [49].

## APPENDIX B

# MACHINE LEARNING

---

Machine learning is a powerful data analysis method that offers numerous advantages, including efficient data management, continuous improvement, pattern identification, and the ability to handle multi-dimensional data. However, it also presents certain challenges, such as data acquisition difficulties, time and space requirements, result interpretation complexities, and susceptibility to high error rates. To delve deeper into this subject, I highly recommend reading the book [\[50\]](#) for a comprehensive understanding.

Artificial intelligence encompasses various aspects of intelligence, including the abilities to perceive, synthesize, and infer information. One significant subfield within artificial intelligence is machine learning, which involves automated processes capable of extracting patterns and relationships from data without explicit programming. Deep learning, a subarea of machine learning, utilizes high-level modeling abstractions to analyze data through multiple processing layers. It typically requires large datasets for optimal performance. Neural networks represent another subfield within deep learning, consisting of interconnected artificial neurons. These networks receive signals, process them, and transmit signals to other connected neurons, enabling complex information processing. The relationships between these concepts can be visualized in [Figure B.1](#).



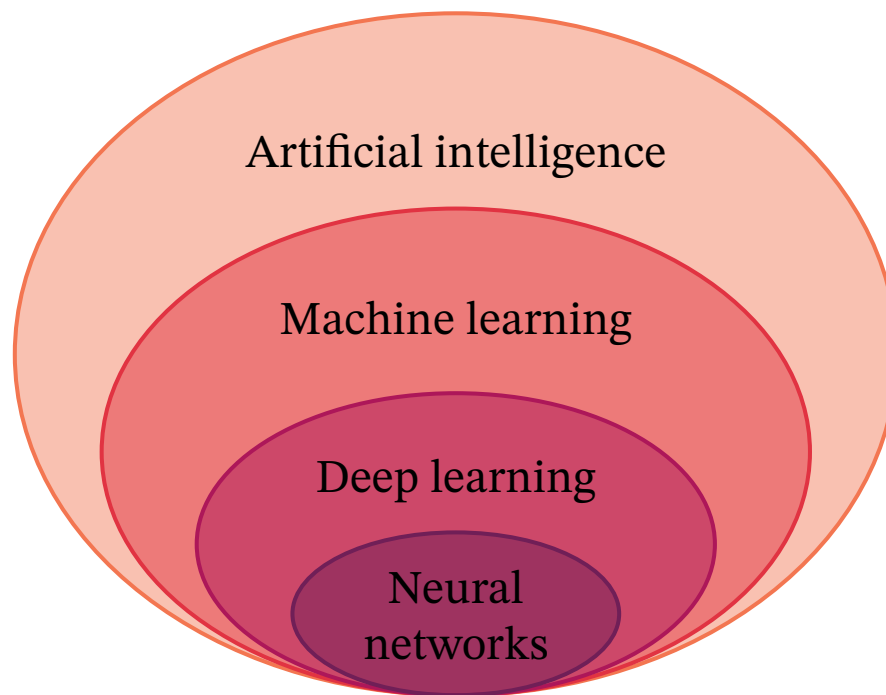


Figure B.1: **Hierarchy of the artificial intelligence subfields.** Within the field of artificial intelligence, there exists the subfield of machine learning. Deep learning is a specialized subfield within machine learning, and further within deep learning, we find the subfield of neural networks.

## Multilayer Perceptron

A Multilayer Perceptron (MLP) is a neural network architecture that consists of an input layer, an output layer, and one or more hidden layers. Each hidden layer is composed of multiple interconnected neurons, creating a stacked structure within the network. The key characteristic of MLPs is their ability to model non-linear mappings between inputs and outputs. MLPs fall under the category of feedforward algorithms, as they receive inputs from the previous layer, which are then combined with initial weights to compute a weighted sum. This result is then passed through an activation function. This process is repeated for each layer, with the output of one layer becoming the input for the next. We can visualize the basic construction of the MLP in the [Figure B.2](#).

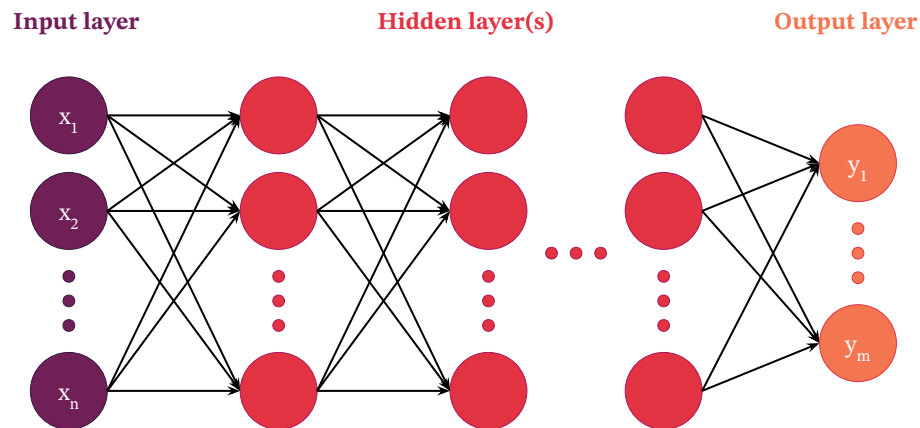


Figure B.2: **Multilayer Perceptron.** A Multilayer Perceptron (MLP) is a type of neural network composed of input and output layers, along with one or more hidden layers containing multiple stacked neurons.

However, there is more to MLPs than just computing weighted sums and propagating results to the output layer. To enable learning and minimize the cost function, MLPs utilize a mechanism called backpropagation. Backpropagation iteratively adjusts the weights in the network during the training process. If the algorithm only computed one iteration or stopped at the output layer, it would not be capable of learning. Backpropagation ensures that the weights are adjusted by computing the gradient of the Mean Squared Error across all input-output pairs.

During each iteration, after the weighted sums are forwarded through all layers, the gradient of the Mean Squared Error is calculated for each input-output pair. Then, this gradient is propagated back by updating the weights of the first hidden layer. This process continues, propagating the weights back through the neural network. The iterations persist until the gradient for each input-output pair converges, meaning the newly computed gradient has not changed significantly beyond a specified convergence threshold compared to the previous iteration. This iterative backpropagation mechanism allows the MLP to adjust its weights and improve its performance over time. We can visualize in the [Figure B.3](#) the neurons of the hidden layers in more detail.

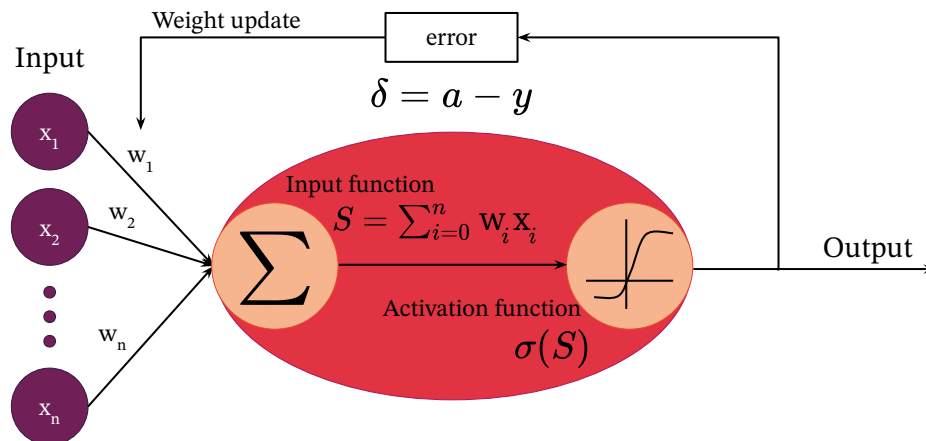


Figure B.3: **Neuron of the hidden layer.** The neurons of the hidden layer receive inputs of the previous layer, combine them with initial weights in a weighted sum, and pass the result through an activation function. Through the process of backpropagation, the network iteratively adjusts the weights by calculating the error of the output.

There are various examples of activation functions commonly used in neural networks. These include:

**Binary step** It outputs 0 if the input is less than zero and 1 if the input is greater than zero.

**Hyperbolic tangent (tanh)** The formula for tanh is  $(1 - \exp(-2x)) / (1 + \exp(-2x))$ . It produces values between -1 and 1, with negative inputs mapping to negative outputs and positive inputs mapping to positive outputs.

**Sigmoid** The sigmoid function is given by  $\text{sig}(x) = 1 / (1 + \exp(-x))$ . It maps inputs to a range between 0 and 1, with smaller negative inputs approaching 0 and larger positive inputs approaching 1.

**Rectified Linear Unit (ReLU)** ReLU returns 0 if the input  $x$  is less than 0 and the input itself if the input is greater than or equal to 0. It is defined as  $f(x) = \max(0, x)$ .

These activation functions serve different purposes and are chosen based on the specific requirements and characteristics of the neural network being used. We can visualize these functions in [Figure B.4](#).

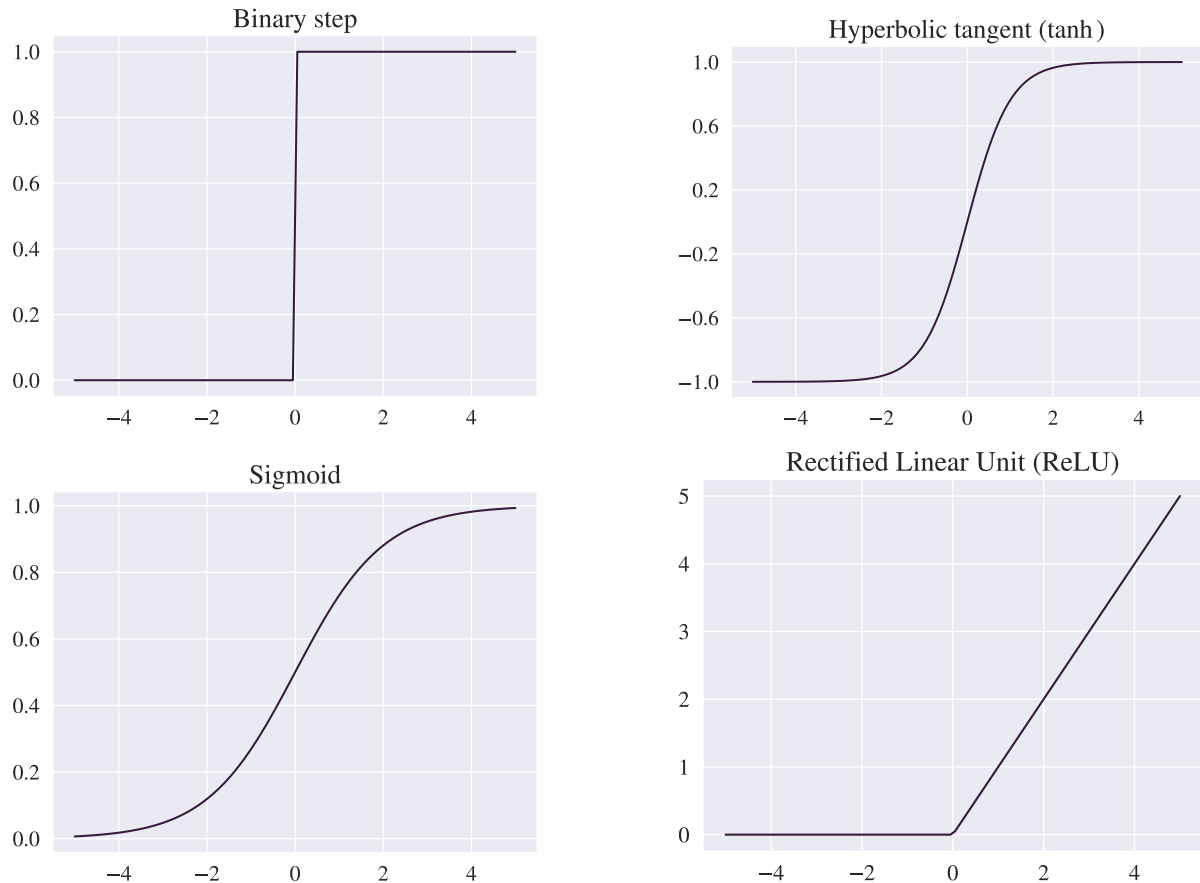


Figure B.4: **Activation functions.** By considering the same range of values, we can observe the behaviors of four activation functions. In the upper left corner, we find the binary step function. Moving to the upper right corner, we encounter the hyperbolic tangent function. In the lower left corner, we have the sigmoid function, and finally, in the lower right corner, we have the ReLU function.

## Performance metrics

We can categorize the classification of inputs into four classes:

**True Positive (TP)** The case was positive, and it was correctly predicted as positive.

**True Negative (TN)** The case was negative, and it was correctly predicted as negative.

**False Positive (FP)** The case was negative, but it was incorrectly predicted as positive.

**False Negative (FN)** The case was positive, but it was incorrectly predicted as negative.

A visualization of this concepts is given in [Figure B.5](#).

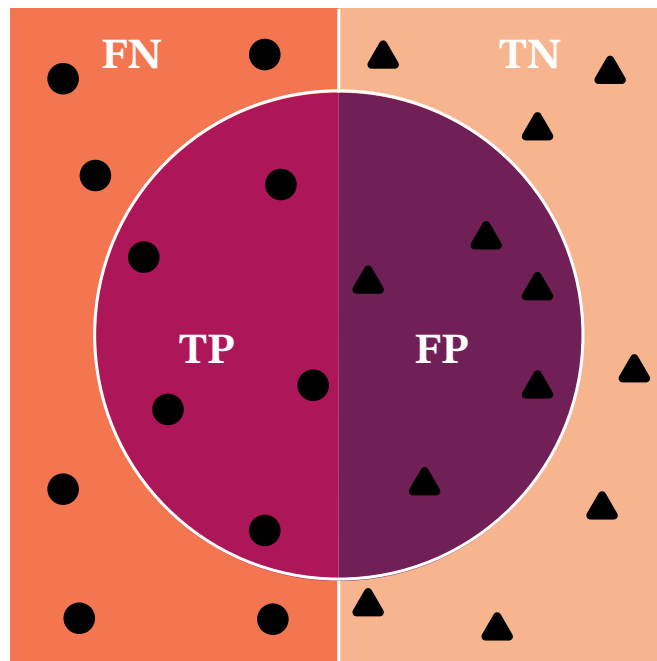


Figure B.5: **Categories of classification.** We designate circles to represent relevant data (positive cases) and triangles to represent irrelevant data (negative cases). Consequently, the left or darker side of the box corresponds to the relevant area. Following the classification by the machine learning model, the selected area, depicted by a circular region, represents the positive cases determined by the model, also called the decision boundary. Within this context, the true positive area of the circumference represents the portion that intersects with the relevant area of the square, while the false positive area refers to the section outside the relevant area. The false negative area refers to the part of the relevant area that was not included in the circumference of selected cases, whereas the true negative area refers to the square region that was neither selected nor relevant.

With this information, we are able to construct our first performance visualization, the confusion matrix. The confusion matrix, presented in [Table B.1](#), provides a clear and intuitive representation of the performance of the classification model used. In an ideal scenario, we anticipate that the data points will predominantly fall along the main diagonal of the matrix, indicating good performance. Any cases that deviate from the main diagonal should ideally be minimal or non-existent. This alignment signifies accurate predictions and a strong correlation between the predicted and actual values.

Table B.1: Confusion matrix

		Actual values	
		Positive (1)	Negative (0)
Predicted value	Positive (1)	TP	FP
	Negative (0)	FN	TN

There are several other performance metrics that provide valuable insights, such as:

**Precision** Precision answers the question, “How many of the selected items are relevant?”

It measures the accuracy of positive predictions, indicating the proportion of correctly identified positive cases among the total selected positives.

**Recall** Recall answers the question, “How many of the relevant items are selected?” It represents the fraction of actual positives that were correctly identified by the model, indicating its ability to capture all relevant cases.

**Accuracy** Accuracy answers the question, “What percentage of predictions were correct?” It measures the overall fraction of predictions made by the model that are correct, providing an assessment of the model’s overall performance.

These performance metrics provide different perspectives on the model’s effectiveness in different aspects of prediction accuracy and can help evaluate and fine-tune the model’s performance. We can visualize each metric in [Figure B.6](#).

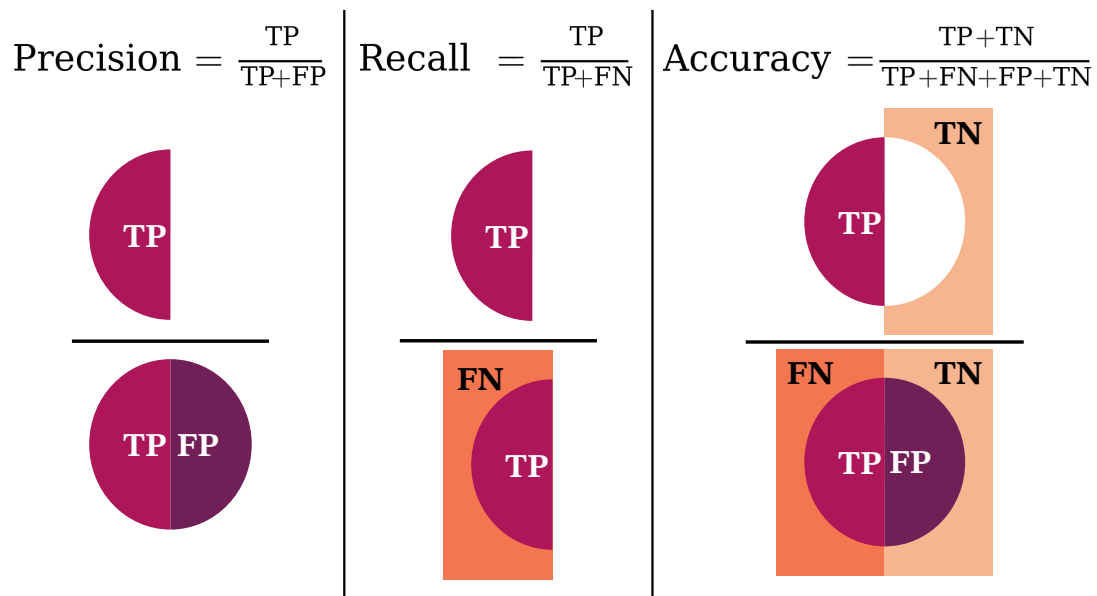


Figure B.6: **Performance metrics.** The formula and the illustrative representation for the precision (in the left), the recall (in the middle) and the accuracy (in the right).

Another metric is support. Different from the others, it remains consistent across models. Support refers to the actual number of occurrences of a class within a given dataset. Imbalanced training data can indicate potential issues with the reported scores of the classifier and may suggest the need for strategies such as stratified sampling or rebalancing techniques. Related to this concept, we have the cross-validation diagnosis. Cross-validation is a resampling method that involves dividing the data into different subsets to train and test the model across multiple iterations, as depicted in [Figure B.7](#). The purpose of cross-validation is to assess the model’s performance in predicting new data that was not used during its estimation. This approach helps identify problems like overfitting or selection bias and provides insights into how the model will generalize to an independent dataset, such as real-world

scenarios. It serves as a means to evaluate the model's ability to perform well on unseen or unknown data [51].

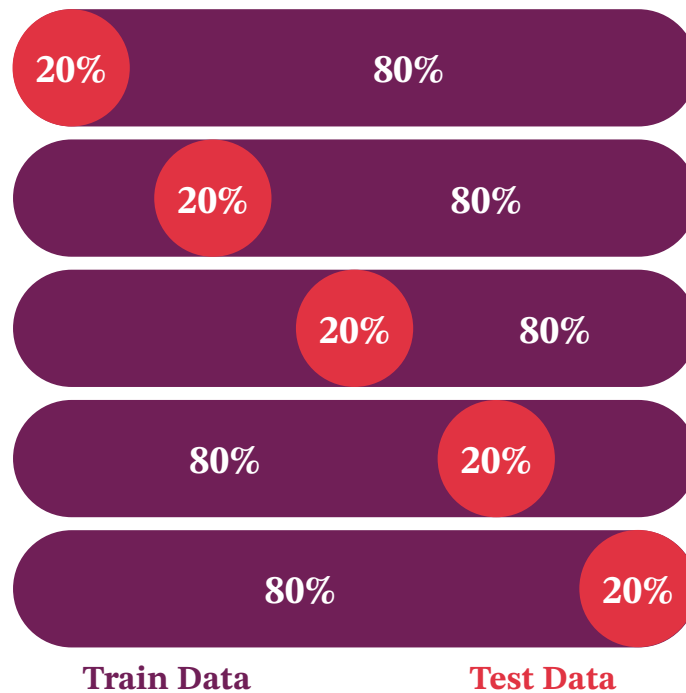


Figure B.7: **Resampling ( $k$ -fold)**. The resampling of  $k$ -fold involves randomly dividing the database into  $k$  subsets, where  $k$  is a predefined value. In this case, we illustrate the resampling using 5 folds, ensuring that each subset contains approximately an equal number of samples.