Universidade Estadual de Campinas
Instituto de Física *Gleb Wataghin*

Gabriel Pereira Alves

Causal relaxations and quantum nonlocality on multipartite systems

Relaxações causais e não-localidade quântica em sistemas multipartidos

Campinas, 2020

Gabriel Pereira Alves

Causal relaxations and quantum nonlocality on multipartite systems

Relaxações causais e não-localidade quântica em sistemas multipartidos

Dissertação apresentada ao Instituto de Física *Gleb Wataghin* da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Física, na área de Física.

Dissertation presented to the *Gleb Wataghin* Institute of Physics of the State University of Campinas in partial fulfillment of the requirements for the degree of Master in Physics, in the area of Physics.

Supervisor/orientador: Prof. Dr. Rafael Luiz da Silva Rabelo

O arquivo digital corresponde à versão final da dissertação defendida pelo aluno Gabriel Pereira Alves e orientada pelo Prof. Dr. Rafael Luiz da Silva Rabelo.

Campinas, 2020

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

AL87c

Alves, Gabriel Pereira, 1994-
Causal relaxations and quantum nonlocality on multipartite systems / Gabriel Pereira Alves. – Campinas, SP : [s.n.], 2020.

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

1. Teorema de Bell. 2. Teoria quântica. 3. Correlações quânticas. 4. Causalidade (Física). I. Rabelo, Rafael Luiz da Silva, 1986-. II. Universidade Estadual de Campinas. Instituto de Física Gleb Wataghin. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Relaxações causais e não-localidade quântica em sistemas multipartidos
Palavras-chave em inglês:
Bell's theorem
Quantum theory
Quantum correlations
Causality (Physics)
Área de concentração: Física
Titulação: Mestre em Física
Banca examinadora:
Rafael Luiz da Silva Rabelo [Orientador]
Marcelo de Oliveira Terra Cunha
Diogo de Oliveira Soares Pinto
Data de defesa: 21-05-2020
Programa de Pós-Graduação: Física

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-5439-9665
- Currículo Lattes do autor: http://lattes.cnpq.br/9268890038270187

MEMBROS DA COMISSÃO JULGADORA DA DISSERTAÇÃO DE MESTRADO DE GABRIEL PEREIRA ALVES – RA 135806 APRESENTADA E APROVADA AO INSTITUTO DE FÍSICA “GLEB WATAGHIN”, DA UNIVERSIDADE ESTADUAL DE CAMPINAS, EM 21 / 05 / 2020.

COMISSÃO JULGADORA:

- Prof. Dr. Rafael Luiz da Silva Rabelo – Orientador – DFMCIFGW/UNICAMP

- Prof. Dr. Marcelo de Oliveira Terra Cunha - IMECC/UNICAMP

- Prof. Dr. Diogo de Oliveira Soares Pinto - IFSC/USP

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

CAMPINAS

2020

Esta tese ainda não é o supra sumo do meu trabalho, porém nada mais justo que dedicar o primeiro de muitos que virão a você, mãe!

Acknowledgements

As this is not a fundamental part of the thesis and the people mentioned here are all brazilians, this section will be written in brazilian portuguese.

Essa me parece uma seção muito difícil de ser escrita, já que são muitas pessoas e poucas palavras que devem fazer parte desse texto. Assim, por uma dificuldade de arranjar tudo em uma meia dúzia de frases, mencionarei aqui apenas aqueles que considero mais fundamentais para o desenvolvimento desse trabalho. É claro que, dada a natureza caótica da vida, se algumas das pessoas com as quais me encontrei durante essa jornada não tivessem cruzado meu caminho, é possível que nada disso fosse escrito e você não estaria lendo essa tese.

Como não é este o caso, a primeira dessas pessoas fundamentais é aquela para a qual eu também dedico esta tese. Conta essa pessoa que foi dela de quem eu herdei a inteligência, assim é ela a quem devo o meu primeiro agradecimento. Eu sei que você confia em mim para tudo, da mesma maneira que eu confio em ti. Obrigado pelas noites mal dormidas que você condicionou à minha falta a perda do seu sono. Obrigado pelo tempo despendido na minha educação e por muitas outras coisas a mais que não retornarão nas nossas vidas. Eu sei que, se eu estivesse no seu lugar, teria feito de bom grado o mesmo que você fez por mim.

A segunda dessas pessoas não existe na minha vida de maneira tão eterna quanto a primeira, mas, se Vinicius me permite, me proporciona um sentimento tão eterno enquanto juntos nós permanecemos. Eu não tenho dimensão alguma dessa eternidade e, se um dia o destino decidir que esta terá um fim, eu quero que você saiba que a nossa união transpassa os limites do companheirismo conjugal. Não foi por acaso que eu te encontrei, apesar de parecer que se eu tivesse tropeçado na calçada naquele dia não teria lhe conhecido. Eu agradeço muito que não tenha tropeçado e mais ainda por você ter surgido na minha vida, Kay.

Ainda existe uma terceira e uma quarta pessoa, pessoas que eu admiro muito, e já que eu não consigo sobrepor os agradecimentos e agradecer as duas ao mesmo tempo, vamos

a terceira. Foram muitas as vezes que eu precisei de ti com único propósito de que você me ouvisse. No fundo do meu coração, tudo que eu mais queria é que igualmente você me dissesse aquilo que eu desejava ouvir, mas você nunca cumpriu esse papel; você sempre me disse o que eu precisava ouvir. E assim eu ressignifiquei boa parte das coisas que me cercavam, de tal maneira que eu sempre vou associar *Baby can I hold you* da Tracy Chapman àquele dia que conversávamos atrás da BIF. Muito obrigado!

Por fim, só me resta agardecer ao professor. É possível que um dia eu também me torne um e pare de lhe chamar dessa maneira porque eu sei que lhe incomoda. Mas de que outra maneira eu poderia lhe demonstrar algum respeito e admiração? Nos conhecemos num instante no qual eu estava muito perdido e você precisava com urgência de um aluno. De certa forma nos complementamos nesse momento. Porém, eu nunca vou poder lhe retribuir no mesmo nível todas as coisas que você me ensinou nos últimos meses. Assim, me resta apenas o agradecimento.

0 presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Brasil.

This study was financed in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Brazil.

Seek knowledge, but not too much otherwise it makes you want to cry

- An internet meme

Abstract

Can we get a criteria to obtain stronger quantum correlations? In this thesis, we reproduce the formalism of searching for a Bell inequality. However, instead of using the common notion of nonlocality, we replaced it with a scenario in which the parts are able to communicate with each other. This way, we have defined a new concept of "nonlocality", called super nonlocality, since now not all non-local behaviors in a Bell scenario without communication will be strongly correlated to be also considered non-local in this new definition. To exemplify this idea, we have found a Bell-type inequality that separates super non-local behaviors from others. More interestingly, a violation of this inequality was observed, reaffirming the advantage of quantum correlations upon classical ones. Also, here the reader will find an exhaustive attempt to still use the facet enumeration method for achieving Bell inequalities for high dimensional Bell scenarios.

Resumo

Podemos encontrar um critério para obter correlações quânticas mais fortes? Nesta tese, reproduzimos o formalismo de procura por uma desigualdade de Bell. No entanto, ao invés de usar a noção comum de não-localidade, substituímo-a por um cenário em que as partes são capazes de comunicar-se entre si. Dessa maneira, definimos um novo conceito de "não-localidade", denominado super não-localidade, já que agora nem todos os comportamentos não-locais em um cenário de Bell sem comunicação estarão fortemente correlacionados para serem considerados não-locais nesta nova definição. Para exemplificar esta ideia, encontramos uma desigualdade do tipo Bell que separa comportamentos super não-locais dos demais. Mais interessante ainda, descobrimos a violação dessa desigualdade, reafirmando a vantagem das correlações quânticas sobre as clássicas. Além disso, aqui o leitor encontrará uma tentativa exaustiva de uso do método de enumeração de facetas para obter desigualdades de Bell em cenários de alta dimensão.

Contents

Preface

Among all the advice that I was given before writing this document one of them was that I should write It as the text I wish I had read before joining my master's studies. I may have taken this advice too seriously than I should, but now I definitely have the thesis I wish I had read before all.

Firstly, I tried my best to write this text in a manner that any student in the last year of its undergraduate studies could be able to read it. Unfortunately, this is not always possible and in the moments in which you face this impossibility, then you should look for the references or the textbooks. This way, the first chapter is entirely introductory, so if you are familiar with the most basics concepts of quantum mechanics, then you can, without a doubt, jump to the next chapter.

Chapter two is also introductory, but, instead of commenting on the basics of quantum mechanics, it introduces the great area of study in which this thesis is based on: Bell nonlocality. In this chapter I tried to be as short as I could and the things I have considered not as immeadiate as the fundamentals of nonlocality, I have put in the appendices, as, in a moment or another, these will be necessary concepts.

Finally, chapters three and four are dedicated to present my specific area of study and the problem I aimed to approach, finishing the text presenting the results we got. If you intend to read this text I appreciate it: it is difficult times for sharing knowledge in this current world.

1 The mathematical foundations of quantum mechanics

Like all revolutions, there is not a specific day when quantum mechanics began. Quantum mechanics is an area that is still producing knowledge and is far from being considered as solid as classical mechanics, for instance, as its foundations are still debated by the community today [1, 2]. Thereby, it is impossible up to date to define quantum mechanics in few words and in a manner that pleases every physicist.

In the late of 1920's, quantum mechanics was associated as the field of physics whose aim was to study the world of the tiny things, whose phenomena could not anymore be explained by classical physics. However, this association is no longer true since there are macroscopic phenomena that can only be described by quantum mechanics [3, 4]. Thus, it is convenient on this work to stand up in a position in which quantum mechanics is the set of mathematical tools used to develop physical concepts, among which we intend to explore some.

This first chapter is dedicated to providing this presentation of concepts in the form of postulates. At first, we will not take on any particular experiment, so that it will look just like a collection of axioms. The introduction of a connection to physical reality will be gradual as this chapter develops.

1.1 The postulates of quantum mechanics

As just presented, this is a sensible area, as there is no accordance on the foundations of quantum mechanics. The number of postulates itself is not fixed and varies according to your preferable bibliography. On Cohen-Tannoudji, Diu and Lalöe textbook, this number corresponds to six [5]. The high number of postulates reinforces the idea of a collection of

axioms as, in comparison to other fundamental areas of physics, their axiomatic structure is not simple.

A reformulation of the axiomatic structure of quantum mechanics, although necessary, is not the focus of this thesis. Here, we will move only with the postulates needed to the rest of the work.

The canonical first postulate is commonly dedicated to present the state of the system and the space in which these states belongs. We will call this the quantum realm postulate as its intention is to present the stage in which the quantum phenomena happens.

Postulate 1.1 (quantum realm)

Every quantum system is associated to a complex vector space $\mathcal{H}$ endowed with an inner product and denominated **Hilbert space**. The state of the system is represented by a normalized vector $|\psi\rangle$ (reads as ψ ket) on this space.

In addition, for all $|\psi\rangle$ we can associate a dual element, named **bra** and denoted by $\langle\psi|$. If done so, the normalization rule defined by the inner product then imposes that $\langle\psi|\psi\rangle = 1$. This condition will be again explored on postulate 1.3.

1.1.1 Operators

Before introducing the second postulate, some brief definitions are needed. On classical mechanics, the importance of the state space is not as great as in quantum mechanics. While here we need to define it on a postulate, in Newtonian mechanics this postulate does not even exist. This is related to our common sense which tends to relate the parameters of the physical theory to be real quantities.

Otherwise on quantum mechanics, some quantities like momentum and position are represented by operators which act on the state. The definitions involving these operators will be multiplied according to the development of this work in such a way it is convenient to quickly present some of these definitions on this subsection.

Firstly, let us assume that $|\psi\rangle = \sum_i c_i |\psi_i\rangle$, where $c_i \in \mathbb{C}$, is a finite discrete vector which can be spanned in the $\{|\psi_i\rangle\}$ basis. Similarly, let $|\psi'\rangle = \sum_i c'_i |\psi_i\rangle$ be another finite discrete vector characterized by coefficients $c'_i \in \mathbb{C}$. If A is an operator acting on $|\psi\rangle$ and mapping it onto $|\psi'\rangle$, then A can be represented by a matrix whose elements are given by

$A_{ij} = \langle\psi_i|\, A\, |\psi_j\rangle$. So, the action of A implies

$$\begin{aligned} c_i' &= \langle\psi_i \,|\, \psi'\rangle = \langle\psi_i \,|\, A \,|\, \psi\rangle \\ &= \langle\psi_i|\, A \left(\sum_j c_j \,|\psi_j\rangle\right) = \sum_j \langle\psi_i \,|\, A \,|\, \psi_j\rangle\, c_j \\ &= \sum_j A_{ij} c_j. \end{aligned}$$

The first interesting property to be explored is the unitarity. An unitary operator A is defined as the operator acting on $|\psi\rangle$ such that $A : \mathcal{H}_\psi \to \mathcal{H}_\psi$, where $|\psi\rangle \in \mathcal{H}_\psi$. If $A : |\psi\rangle \mapsto |\psi'\rangle$, its inverse operator is given by the action $A^{-1} : |\psi'\rangle \mapsto |\psi\rangle$ such that $AA^{-1} = \mathbb{1}$. Some desirable properties of the unitaries worth to be mentioned:

1. Its inverse operator equals its complex and adjoint (also named Hermitian) conjugation: $A^{-1} = A^* = A^\dagger$;

2. Unitaries are inner product preserving; if $A : |\psi\rangle \mapsto |\psi'\rangle$ and $A : |\varphi\rangle \mapsto |\varphi'\rangle$, then $\langle\psi'|\varphi'\rangle = \langle\psi|\varphi\rangle$;

3. Their eigenvalues $\{a_i\}$ are such that $a_i = e^{-i\theta_i}$, for some $\theta_i \in \mathbb{R}$, and their eigenvectors $\{|a_i\rangle\}$ are orthonormal;

4. The unitaries admit a **spectral decomposition**: $A = \sum_i a_i \,|a_i\rangle\langle a_i|$.

As just mentioned on item 1 above, we can make a conjugation of A named Hermitian. It is obtained just by transposing and taking the complex conjugate of the operator. The notation is as defined earlier: $A^\dagger$ is the Hermitian conjugate of A. Also, if $A = A^\dagger$, the operator is said Hermitian. Like the unitaries, the Hermitians also admit a spectral decomposition, but, since A is Hermitian, its eigenvalues $\{a_i\}$ are real.

It is also interesting to define the projector $P_i = |a_i\rangle\langle a_i|$ as the outer product with itself. This way, we can write the spectral decomposition of A as $A = \sum_i a_i P_i$. Here, the definition of a projector is the same as in linear algebra. P_i is an operator whose action projects the state $|\psi\rangle$ onto a subspace of $\mathcal{H}_\psi$, then $P_i\,|\psi\rangle = |\psi_i\rangle$. Any operator is easily checked to be a projector if it satisfies $P_i^2 = P_i$.

Finally, we can make a combination of projectors in such a way that this combination

returns an operator which we will call **observable**. Let

$$O = \sum_i a_i P_i,$$

where $a_i \in \mathbb{R}$. Then, O is an observable if the above combination is such that O is Hermitian.

Now, we can comfortably return to the postulates. The next immediate step is to define the dynamics of the theory. By convenience, it will be assumed to act on a closed physical system. A closed physical system is a system which cannot interact with others. However, this definition lacks on real world applications as we cannot study any fully isolated system. So, in the postulate below, it will be considered that a closed physical system is the best approximation of a fully isolated system which is able to be studied.

Postulate 1.2 (dynamics)

A closed physical system can evolve subject to the action of a unitary transformation, represented by the unitary operator U, such that if $|\psi\rangle$ is the state immediately before the evolution and $|\psi'\rangle$ is the state immeaditely after, then $|\psi'\rangle = U|\psi\rangle$.

Note that the unitary operator does not specify that the evolution of the physical system happens on time. However, we can extend this definition by introducing the Hamiltonian operator H of $|\psi(t)\rangle$ if t is a parameter.

$$i\hbar \frac{\partial |\psi(t)\rangle}{\partial t} \equiv H|\psi(t)\rangle, \tag{1.1}$$

which correspond to the known Schrödinger's equation.

Postulates 1 and 2 are very mathematical rules. It does not say much about the connection of the state with the real physical system and their measurable properties. Maybe the most important postulate is the one which makes this remaining link. A fundamental feature of quantum mechanics is that equation (1.1) was able to predict the energies of certain physical system. The way we can understand it is seeing the Hamiltonian operator as an observable, whose eigenvalues E corresponds to the energies

of the physical system. If $|\psi\rangle$ is an eigenvector of H, so

$$H\,|\psi\rangle = E\,|\psi\rangle\,.$$

Whereas H is an observable, we can assume it has an spectral decomposition in terms of projectors. Let us say $H = \sum_i E_i\,|\psi_i\rangle\langle\psi_i|$ or $H = \sum_i E_i P_i$, where $|\psi_i\rangle$ is the i-th eigenvector of H.

When quantum mechanics was established in the early of 1920's there was a clear interpretation of the Hamiltonian's eigenvalues, but the interpretation associated to the eigenvector was suggested only in 1926 by Max Born [6] when he was working on the problem of an electron scattering through a potential in which he aimed to calculate and interpret the asymptotic state after the scattering. This interpretation in most known nowadays as one of the standard postulates of quantum mechanics. It is a consensus and you will see it whatever the textbook you prefer. It also has a special name which is attributed to Max Born.

Postulate 1.3 (Born's rule)

Associated to every observable eigenvector there is a probability, with magnitude of

$$p_i = |\,\langle\psi_i|\psi\rangle\,|^2, \tag{1.2}$$

to observe the i-th observable eigenvalue in a measurement of $|\psi\rangle$.

So far, we have not mentioned anything about probabilities, but this specific postulate turns quantum mechanics into a probabilistic theory. It is because of this result that we impose that the state must be normalized, otherwise (1.2) would not represent a probability.

If the eigenvectors of an observable H constitute a basis of $\mathcal{H}$, we can write the state as a linear combination of its basis,

$$|\psi\rangle = \sum_i c_i\,|\psi_i\rangle\,, \tag{1.3}$$

that is, $|\psi\rangle$ can be written in a **superposition** of the eigenvectors of H. On this ordinary situation, it is easy to see that $p_i = |c_i|^2\,,\ \forall\, i$, where the physical $|\psi\rangle$ is normalized.

Note that postulate 1.3 does not mention which of the eigenvalues of the observable will be obtained once you perform a measurement of it. This is, obviously, expected as the theory can only give probabilities; then the result of a measurement only can be predicted with certainty if one of the p_i's equals to one. Postulate 1.3 is what gives quantum mechanics the characteristic of being nondeterministic, differently of most of the physical theories.

There is, still, a subtlety on postulate 1.3 which is so fundamental that deserves another postulate for itself. Consider, for instance, the aforementioned observable $H = \sum_i E_i P_i$ which can be written as a linear combination of projective operators. Once you perform a measurement, one of the observable's eigenvalues is picked at random and revealed. If, just after the first measurement, you perform a second measurement equal to the first one, then it is natural to expect that the result of this second projective measurement is the same as before. This criterion is denominated as repeatability and its validity enables us to consider that there is a well-defined post-measurement state. Its form is given on the postulate below.

Postulate 1.4 (post-measurement state)

Given that the i-th eigenvalue of the observable being measured occurred, the state of the quantum system immediately after the measurement is

$$\frac{P_i \left|\psi\right\rangle}{\sqrt{p_i}}, \tag{1.4}$$

where the P_i projects $\left|\psi\right\rangle$ onto the subspace spanned by the i-th eigenvector.

The whole set of operators $\{P_i\}$ and the process depicted on postulate 1.4 is also known as a projective measurement or sharp measurement. This is not the only case to be considered, but the most intuitive and simple to illustrate this postulate. However, there can be other cases which depend on the experimental setup you are considering. For instance, if $\left|\psi\right\rangle$ represents a photon, the post-measurement state does not even exist - as the measurement process destroys the photon - and the repeatability criterion can no longer be assumed valid.

1.1.2 Positive Operator-Valued Measure

For those cases where we cannot repeat a measurement, consider an operator slightly different than the projectors P_i. We will name it an **effect** and represent the set of effects by $\{E_i\}$. For each effect we will consider that it is a positive semi-definite operator, which means that

$$\langle\psi|E_i|\psi\rangle \geq 0,$$

for all $|\psi\rangle \in \mathcal{H}$. The above property is also a projector property, otherwise the probabilities given by equation (1.2) could be negative. We also want this property for the effects in such a way we can define a Born's rule for the effects[1].

Note that equation (1.2) can be written as

$$p_i = |\langle\psi_i|\psi\rangle|^2 = \langle\psi|\psi_i\rangle\langle\psi_i|\psi\rangle = \langle\psi|P_i|\psi\rangle .$$

So, if the i-th output of the measurement is observed, the same Born's rule can be rewritten as $p_i = \langle\psi|E_i|\psi\rangle$. On this case, the post-measurement state corresponds to

$$\frac{E_i^{1/2}|\psi\rangle}{\sqrt{p_i}}. \tag{1.5}$$

Because E_i is no longer a projector, not necessarily $E_i^{1/2}$ equal E_i and if a second measurement is taken, the post-measurement state is not necessarily equals to (1.5), thus this measurement is not repeatable.

For the $\{E_i\}$ measurement, we call by the old fashioned name Positive Operator-Valued Measure. This name is almost unused anymore in detriment of its acronym, POVM. Like the projective measurement case, the effects of a POVM also satisfy a completeness relation, $\sum_i E_i = \mathbb{1}$. In a projective measurement, however, if all of the projectors are one-dimensional operators, then this sum has d_ψ terms, where d_ψ is the dimension of $\mathcal{H}_\psi$. For the POVMs, the number of possible outputs is no longer limited by the dimension of the Hilbert space and the completeness relation can be a sum of an arbitrary number of terms.

Until now, we have been discussing about a single physical system we described by the

[1]For an interesting discussion about the difference between measurements see [7].

state $|\psi\rangle$. Still, we need, for the sake of completeness, consider a rule for what to do when dealing with more than one physical system. The last postulate is no more than a recipe for these cases (which are the huge majority in nature).

Postulate 1.5 (composite systems)

The Hilbert space $\mathcal{H}$ of a system composed by n parts, where $\mathcal{H}_i$ is the Hilbert space of the i-th part, for $i = 1, ..., n$, is obtained by taking the tensor product between the parts, such that $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes ... \otimes \mathcal{H}_n$.

In summary, for a composite physical system, where each of the parts are *individually* prepared in a state $|\psi\rangle_n$, for $i = 1, ..., n$, the joint state can be obtained by taking the tensor product of the individual states. Altough we can affirm that $|\psi\rangle = |\psi\rangle_1 \otimes ... \otimes |\psi\rangle_n$ is the composite state for this case[2], it is not true for an arbitrary state, let say, $|\varphi\rangle \in \mathcal{H}$, that $|\varphi\rangle = |\varphi\rangle_1 \otimes ... \otimes |\varphi\rangle_n$, for $|\varphi\rangle_i \in \mathcal{H}_i$, $i = 1, ..., n$. In the first case $|\psi\rangle$ is denominated as a **product state**, whilst $|\varphi\rangle$ is **entangled**. This last definition will be presented in more detail in section 1.4.

1.2 The qubit

This section is no more than an introduction to a generic physical system that will be useful to illustrate the further concepts we will explore during this thesis. This system is known as qubit - the abbreviation of quantum bit - and makes reference to the bit, the basic unit of information in information theory.

A qubit can be seen as any two-level quantum system. So, suppose there is a generic observable G which have two eigenvalues: the first one will be called g, as ground, and the second one e, as excited. For each of these eigenvalues there will be an associated orthonormal eigenvector, $|g\rangle$ and $|e\rangle$, respectively. A qubit is then any state $|\psi\rangle$ which can be written as a superposition of $|g\rangle$ and $|e\rangle$, $|\psi\rangle = \alpha |g\rangle + \beta |e\rangle$, with $|\alpha|^2 + |\beta|^2 = 1$.

The standard notation is to consider the observable $G = \sigma_z$, where σ_z is the z Pauli matrix. On this case, $g = 1$ and $e = -1$ and the eigenvectors are $|0\rangle = [\ 1\ \ 0\ ]^{\mathrm{T}}$ and $|1\rangle = [\ 0\ \ 1\ ]^{\mathrm{T}}$. It is also useful to consider the Bloch sphere representation. Any qubit

[2]For discrete Hilbert spaces, the tensor product reduces to the Kronecker product. Worth mentioning that the usual notation of composite systems omits the "$\otimes$" symbol. If $|\psi\rangle$ and $|\varphi\rangle$ are two physical systems, and the joint state is a product state, then $|\psi\rangle\,|\varphi\rangle$ is the notation used to evidence the composition. Also, $|\psi,\, \varphi\rangle$ implicitly denotes the composite state.

state can be parametrized as

$$|\psi\rangle = \cos(\theta/2)\,|0\rangle + \sin(\theta/2)\,e^{i\varphi}\,|1\rangle \tag{1.6}$$

where $0 \leq \theta \leq \pi$ and $0 \leq \varphi \leq 2\pi$. This leads us to define a representation known as the Bloch sphere. The Bloch sphere is an unit radius sphere and the geometric location of all qubit states, that is, for each possible qubit state there are an associated θ and φ.

A well known physical system which can be satisfactorily represented by a qubit is the spin-$1/2$. The spin observable is written as a combination of the Pauli matrices,

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

in the compact form, $\mathbf{S} = \hbar/2\,\boldsymbol{\sigma}$, where $\boldsymbol{\sigma} = [\,\sigma_x\,\sigma_y\,\sigma_z\,]$, a vector whose elements are matrices. If one intends to measure the spin of a spin-$1/2$ particle so one needs to choose an arbitrary direction in which the spin vector will be projected. By convention this direction is named z. One could rotate the system and measure it on the x direction, for instance. Now, the results of the measurement are the eigenvalues of the S_x observable. As the choice of the direction of such a measurement is always arbitrary, the diagonalization of S_x provides the same eigenvalues as for S_z which is an important property of the Pauli matrices: their eigenvalues are ± 1. Therefore, the spin state $|s\rangle$ can either be written in these two forms

$$\begin{aligned} |s\rangle &= \alpha_z\,|\uparrow\rangle_z + \beta_z\,|\downarrow\rangle_z \\ &= \alpha_x\,|\uparrow\rangle_x + \beta_x\,|\downarrow\rangle_x\,, \end{aligned}$$

where $|\uparrow\rangle_x$ and $|\downarrow\rangle_x$ are eigenvectors of S_x and $|\uparrow\rangle_z$ and $|\downarrow\rangle_z$ are eigenvectors of S_z. Moreover, $\alpha_{x,z}$, $\beta_{x,z} \in \mathbb{C}$.

It is always possible to change the representation of a state as long as there are two distinct observables. By its spectral decomposition, you can find two different basis and rewrite the state. This brief example illustrate this changing for a spin-$1/2$ system. For instance, in a composite two level system of two qubits, any state can be written as

$$|\psi\rangle = \alpha\,|00\rangle + \beta\,|01\rangle + \gamma\,|10\rangle + \delta\,|11\rangle\,,$$

where $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1$. Or, alternatively, we can write it in the **Bell basis**.

$$|\psi\rangle = \alpha' |\Psi^+\rangle + \beta' |\Psi^-\rangle + \gamma' |\Phi^+\rangle + \delta' |\Phi^-\rangle ,$$

where the normalization is also valid for the prime indices and

$$\begin{aligned} |\Phi^+\rangle &= \tfrac{1}{\sqrt{2}} (|00\rangle + |11\rangle) \\ |\Phi^-\rangle &= \tfrac{1}{\sqrt{2}} (|00\rangle - |11\rangle) \\ |\Psi^+\rangle &= \tfrac{1}{\sqrt{2}} (|01\rangle + |10\rangle) \\ |\Psi^-\rangle &= \tfrac{1}{\sqrt{2}} (|01\rangle - |10\rangle) . \end{aligned} \tag{1.7}$$

For each of the elements in the Bell basis, the two qubits are configured in a way to present **entanglement**. Entanglement is a quantum resource which can be used for quantum information protocols. We will also address this quantum feature at this chapter, but, before we go deeper into this amazing concept, let us introduce another approach to the description of quantum systems that slightly differs from that made in terms of complex vectors.

1.3 Density operator

In a given quantum experiment, it is desirable, for various purposes, to know what is the state of a particular physical system. There are many situations that we could address, but let us consider that either the state is already known by theoretical approaches or that it is entirely unknown. If the situation is the first one, then the problem is solved and we carry on with the experiment. However, if the situation is the second one, then we must consider a source which produces a statistical *ensemble* of identical states. If the measurement operators are wisely chosen, then it is possible to experimentally determine the state of this *ensemble* by a process known as quantum tomography [8], which we will not detail. Note that if there was a single copy of the state, then it would be destroyed in the measurement process and the resultant state would be that given by postulate 1.4, which is different from the original state.

Now, let us assume that there is a source which can produce several copies of

$$|\psi\rangle = \frac{1}{2} (|00\rangle + |01\rangle + |10\rangle + |11\rangle) .$$

The statistics observed at the end of the experiment will allow the experimentalist to conclude that the state is indeed a quantum superposition as in the equation above by observing the probability 1/4 of obtaning any of the kets in the two-qubit basis.

Secondly, let us consider a source which produces, with 50% of probability, the state $|\Psi^+\rangle$, and $|\Phi^+\rangle$, with another 50% of probability. Altough we also obtain 1/4 of probability for each element of the two-qubit basis the resulting state is not a quantum superposition of all of these states, like the first case. Instead, we have a kind of classical mixture between states $|\Psi^+\rangle$ and $|\Phi^+\rangle$. For cases like this, we have a limitation on the notation. Though we can denote the state of the first case as $|\psi\rangle$, in the second case we are forced to emphasize we have a combination of $|\Psi^+\rangle$ and $|\Phi^+\rangle$ with 50% of probability for each.

We can formally differentiate the two cases by introducing an operator we will call **density** and denote by ρ. In the first case, where the state $|\psi\rangle$ is purely a quantum superposition, the ρ operator will be obtained by

$$\rho = |\psi\rangle\langle\psi| \,. \tag{1.8}$$

This state is called **pure**. In the second case we must introduce the classical misture by making a combination between $|\Psi^+\rangle$ and $|\Phi^+\rangle$,

$$\rho = \frac{1}{2}\left(|\Psi^+\rangle\langle\Psi^+| + |\Phi^+\rangle\langle\Phi^+|\right).$$

For a general case, we have

$$\rho = \sum_i p_i \, |\psi\rangle_i\langle\psi|_i \,, \tag{1.9}$$

where all the $|\psi\rangle_i\langle\psi|_i$ are pure states. Equation (1.9) is represented by a convex combination, that is, its weights are constrained by $\sum_i p_i = 1$ and $p_i \geq 0$, for all i. In this case, we have a **mixed** state, where the weights are represented by the respective probabilities p_i.

1.3.1 The postulates for density operators

When considering an *ensemble* of physical systems, the density operator represents a complete description of the systems, working as a proper representative state of this *ensemble*. For the purpose of this thesis, the experiments treated here are represented by

an *ensemble* of equally prepared experiments instead of a single round. Then, for the sake of convenience, it is worth to rewrite the postulates in the density operator representation.

Now, we consider that the state of the physical system is represented by $\rho \in \mathcal{H}^2$, a complex matrix. The normalization condition now imposes

$$\text{tr}\ \rho = 1, \tag{1.10}$$

where 'tr' denotes the sum over all diagonal elements of ρ. The action of any unitary can be seen as a transformation on ρ, $\rho' = U\rho U^\dagger$, where ρ' is the state after the unitary evolution. A generalization of equation (1.1) can also be obtained if we consider

$$\frac{\partial \rho}{\partial t} = -\frac{i}{\hbar}[H,\ \rho], \tag{1.11}$$

where $[A,\ B] \equiv AB - BA$ denotes the commutator between operators A and B, and H is the Hamiltonian description of ρ. Equation (1.11) is known as the Von-Neumann equation and reduces to Schrödinger equation in the pure ρ case.

As in (1.10), the Born's rule is now trace based. If a measurement is projective, then the probabilities are obtained by

$$p_i = \text{tr}\ (\rho P_i)$$

and the post-measurement state is given by

$$\frac{P_i\ \rho\ P_i}{\text{tr}\ (\rho P_i)}.$$

Similarly, if a measurement is performed by a POVM, then

$$p_i = \text{tr}\ (\rho E_i)$$

and the post-measurement state is given by

$$\frac{E_i^{1/2} \rho E_i^{1/2}}{\text{tr}\ (\rho E_i)}.$$

The composition rule is trivially extended by the use of the tensor product between the density matrices representing their respective subsystems.

The density operator has some useful properties compared to the kets. You can easily verify that ρ is pure by taking its square, then if $\rho^2 = \rho$, ρ is pure. Also, all density operators are Hermitian, which suffices to say that their eigenvalues are real.

It is also convenient to define the partial trace operation. Let us say that $\rho \in \mathcal{H}_A \otimes \mathcal{H}_B$ and define $\{|a_i\rangle\}$ and $\{|b_i\rangle\}$ orthonormal basis of $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. The partial trace over the B system is defined as the mapping $\mathrm{tr}_B : A \otimes B \mapsto A \,\mathrm{tr}\, B$, for any $A \in \mathcal{H}_A$ and $B \in \mathcal{H}_B$. The partial trace over the A system is defined similarly. If we decompose $\rho = \sum_{ijkl} p_{ijkl} \, |a_i\rangle\langle a_j| \otimes |b_k\rangle\langle b_l|$, then

$$\begin{aligned} \rho^A &= \mathrm{tr}_B \, \rho = \sum_{ijkl} p_{ijkl} \, \langle b_k|b_l\rangle \, |a_i\rangle\langle a_j| \quad \text{and} \\ \rho^B &= \mathrm{tr}_A \, \rho = \sum_{ijkl} p_{ijkl} \, \langle a_i|a_j\rangle \, |b_k\rangle\langle b_l| \,, \end{aligned}$$

where ρ^A and ρ^B are also called of **reduced** states.

1.4 Entanglement

Differently from the former sections, this one does not talk about textbook quantum mechanics. Entanglement is a subject which is not studied in many undergraduate or even graduate courses of quantum mechanics. The separation between this final section and the others highlights this distinction: while the role of the former sections is to introduce the fundamental ingredients needed to the development of this thesis, here we also introduces one of these fundamental ingredients, but entanglement, despite being a fundamental resource for the development of current physics [9], is regarded as a corollary of the other fundamental principles, even though it is a present resource, among all physical theories, only on quantum mechanics.

Historically, the study of entanglement starts in the seminal paper by Einstein, Podolsky and Rosen [10], which later came to be known as the enunciator of the EPR paradox. In the language of EPR,

> "the quantum-mechanical description of physical reality given by wave functions is not complete".

Though the content of the EPR article is much more intricate than the merely exposition of its conclusion, it was precisely this contradictory statement that motivated the

investigation of the entanglement phenomenom.

As on EPR description, let us start with a two-part composite system. For these cases, the definition of entanglement is as follows: a joint physical system is entangled if the joint state $\rho^{\mathrm{AB}} \in \mathcal{H}_A \otimes \mathcal{H}_B$ cannot be written as a convex combination of product states,

$$\rho^{AB} \neq \sum_i p_i \, \rho_i^A \otimes \rho_i^B, \tag{1.12}$$

where $\rho_i^A \in \mathcal{H}_A$, $\rho_i^B \in \mathcal{H}_B$ and $p_i \geq 0$ for all i. These two parts, A and B, are conventionally called in the literature by Alice and Bob, respectively. If a third part, denominated Charlie, is considered, definition (1.12) can be trivially extended by adding a third reduced state ρ_i^C on the summation. We can also specify (1.12) in order to make it immediately recognizable for pure states. Again, consider $\{|a_i\rangle\}$ a basis of $\mathcal{H}_A$ and $\{|b_i\rangle\}$ a basis of $\mathcal{H}_B$. An entangled pure state $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ is such that

$$|\psi\rangle_{AB} \neq \sum_{i,j} c_{ij} \, |a_i\rangle \otimes |b_j\rangle \,, \tag{1.13}$$

where there is a set of coefficients $a_i,\ b_j \in \mathbb{C}$ providing $c_{ij} = a_i b_j,\ \forall\, i, j$. Equation (1.13) summarizes the case of all of the elements of the Bell basis, as (1.7). For this, (1.7) is also denominated an entangled basis.

The magic of using entanglement as a resource for quantum information protocols can be easily seen with an ordinary example. Consider that Alice and Bob share a two qubit system known in the literature as **singlet**. It has exactly the same form as $|\Psi^-\rangle$ state:

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}} \left(|01\rangle - |10\rangle\right). \tag{1.14}$$

The convention is to consider that the first qubit will be possessed by Alice, the second one by Bob and so on. Then, if Alice measures her qubit on σ_z basis, she can get +1 or -1 with 50% of probability each. Suppose the result +1 is achieved; then Bob's reduced state will be $|1\rangle$. If, instead, the result -1 is obtained, then Bob's reduced state will be $|0\rangle$.

At first sight, nothing too surprising can be deduced from this phenomenon, but there are two points worth to be noted. The first one is that if the state is a product state, then Alice cannot steer Bob's state. From (1.13),

$$|\psi\rangle_{AB} = \sum_{i,j=0}^{1} a_i\, b_j\, |i\rangle \otimes |j\rangle = \sum_{i=0}^{1} a_i\, |i\rangle \sum_{j=0}^{1} b_j\, |j\rangle \equiv |\psi\rangle_A\, |\psi\rangle_B\,,$$

where $|\psi\rangle_A$ and $|\psi\rangle_B$ are the reduced states of Alice and Bob, respectively. Then, no measurement done on Alice's laboratory will influence Bob's reduced state.

The second point is that Alice can exert influence on Bob's reduced state just by performing a measurement on her laboratory. This fact is quite obvious since the singlet is produced anti correlated. What is not explicit is that this description of the experiment does not take into account any dynamical description, that is, the laboratories of Alice and Bob can be arbitrarily apart from each other and the description would still be the same as presented above.

Although it smells like communication at a distance, note that neither Alice can choose the result of her measurement nor force Bob to use a specific measurement direction, i.e., if Bob chooses to perform his measurement on σ_x basis, the best description that Alice could obtain would be $|0\rangle_x$ with 50% of probability or $|1\rangle_x$ with another 50%.

This kind of protocol is currently known as EPR steering as it was first suggested in the EPR paper, which used, instead of qubits, the position and momentum basis of two entangled particles. Despite being used as the motivating problem of this thesis, the steering protocol is not our focus and the interested reader can consult [11]. There are also many other protocols in which entangled pairs can exhibit non-classical features such as this non-local steering example presented above. Again, for more detailed reading, see this characterization of arbitrary protocols on reference [12].

A satisfactory answer to the EPR argument about the completeness of quantum mechanics was only possible on the 1964 paper by John Bell [13]. Bell cleverly reformulated the local realism hypothesis originally coined by EPR which enabled an experimental test based on a local hidden variables (LHV) model. A hidden variable model was a completeness proposal suggested by David Bohm [14, 15], an Einstein's pupil which continued EPR's result. On Bell's argument, any completion LHV based is wrong, that is, either the LHV models are impossible or either they evolve non-locally. Our next immediate step is to explore the modern view of Bell's formulation, later known as nonlocality. In the next chapter our approach, instead of focusing on a specific state and measurement operators, will answer the question: for an arbitrary given state and POVM, can this set exhibit nonlocality?

2 Bell nonlocality

Picture a shoe factory with a weird behavior. Instead of just producing shoe pairs and delivering them to the stores, this particular factory produces the pair and packages each shoe in a box. The boxes are then delivered to two different stores: Alice's store and Bob's store. The experiment finishes when either Alice and Bob open their boxes. Without surprise, if Alice opens her box and finds a left shoe, then Bob will find a right shoe.

This silly experiment can be seen in optics of a probabilistic description. Let us say that the event where Alice finds a left shoe in her box is A_l. The same way A_r is the event when Alice finds a right shoe. This notation will also be extended for Bob's findings. It is easy to check that A_l and B_r are not independent. In fact,

$$p(A_l,\, B_r) > p(A_l)p(B_r). \tag{2.1}$$

If A_l and A_r are equiprobable - same for B_l and B_r - then, $p(A_l,\, B_r) + p(A_r,\, B_l) = 1$ and $p(A_r,\, B_r) = p(A_l,\, B_l) = 0$, and you can easily check that $p(A_l,\, B_r) = 1/2$ and $p(A_l)\, p(B_r) = 1/4$.

Instead of considering the description of this experiment in terms of absolute probabilities, we can express equation (2.1) by using conditionals. Once the producer of the shoes finishes his job, he needs to make a choice and send for Alice and Bob one of the boxes he knows that contains the right or left shoe. Suppose he chooses to send the left shoe to Alice. This way, he has no alternative than sending Bob the right shoe, and the equality on (2.1) is achieved, if conditioned to the choice of the producer.

This ordinary example can be summarized as follows [16]:

Reichenbach's common cause principle

Consider events A and B with causal dependency such as

$$p(A, B) > p(A)\,p(B).$$

Then, one of the following possibilities is true: either A is cause of B, or B is cause of A, or there is a common cause to A and B, let us say C, such that

$$p(A, B\,|\,C) = p(A\,|\,C)\,p(B\,|\,C).$$

The above result is attributed to Hans Reichenbach on his posthumous work "The direction of time" published on 1956 [17].

A conclusion to this example cannot be other than infer that there is indeed a common cause to the dependency between A_l and B_r on (2.1). Although this conclusion is supported by Reichenbach's principle, it is due to the hypothesis of the example.

2.1 Bell scenario

Let us diverge a little the focus and start with another *gedankenexperiment*. This will be baptized of Bell scenario. On a typical Bell scenario experiment we must have, like the former case, more than one box, which are apart from each other in such a manner that a box cannot send a signal to the others, i.e., the boxes are space-like separated. In this case the boxes do not contain a shoe, but an unknown experiment. On each box we can find some buttons and a display. Once you press one of the available buttons, the experiment inside the box is performed yielding a result, which is displayed on the screen, as in figure 2.1b.

Every Bell scenario can be easily referred as the triple of numbers (N, m, r), where N indicates the number of boxes involved in the experiment. If the experiment is said uniform, then the number of measurements m per part is the same for all boxes and the number of results r per measurement, which is assumed to be discrete and finite, is the same for all measurements as well.

As in the shoes box experiment, we will treat a Bell scenario with probabilities. Let us say that Alice chooses a measurement x in the set of m measurements which provides

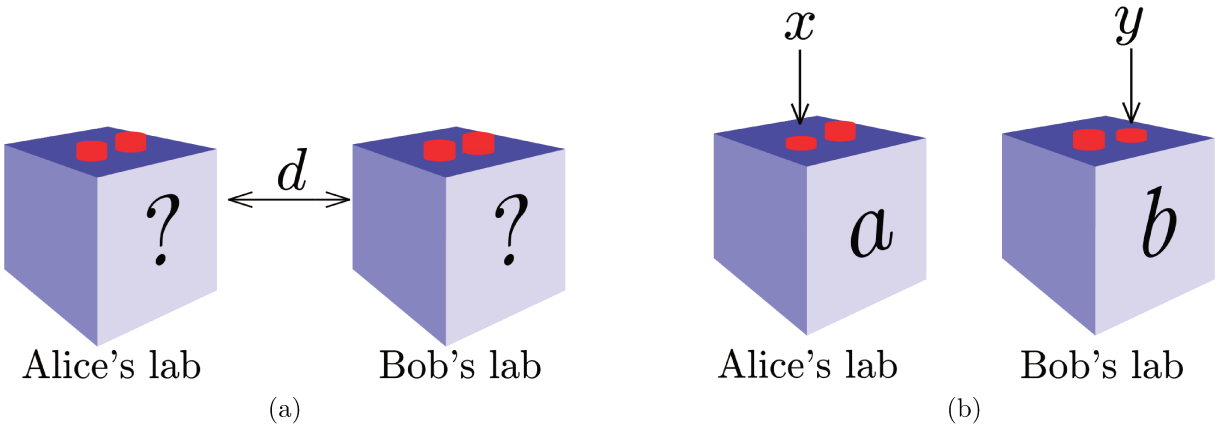


Figure 2.1: (a) A typical Bell scenario with two parts - Alice and Bob - and two measurements per part. A condition needed for every Bell scenario is that the boxes need to be kept in a distance d which is enough to secure that no signal can be sent from one of the parts to another. (b) Encoding the choice of Alice, we attribute x variable to the measurement. The results of Alice's box are encoded by a. Bob's measurement is represented by y variable while his result is represented by b variable.

result a in a set of r possible results. For Bob, we will consider the same notation as illustrated in figure 2.1b. The objective of the experiment conducted by Alice and Bob is to check if their results are dependent from each other and, more specifically, if any causal influence can be found on Bob's results once Alice has already performed her experiment or vice versa. Then, in the general case, for each round of the experiment

$$p(a, b|x, y) \geq p(a|x)p(b|y). \tag{2.2}$$

If (2.2) is strictly positive, than we say that Alice's and Bob's experiment are positively correlated. Otherwise, if the equality is achieved, they are uncorrelated.

We also assumed that the boxes are far apart from each other and a restriction can be obtained over the probabilities from this. If the marginal probabilities, $p(a|x)$ and $p(b|y)$, can only be locally calculated then Alice's free choice of measurement cannot influence Bob's marginal probability and vice versa. That is,

$$\begin{aligned} p(a|x) &= \sum_{b\in r_y} p(a, b|x, y) = \sum_{b\in r_{y'}} p(a, b|x, y') \\ p(b|y) &= \sum_{a\in r_x} p(a, b|x, y) = \sum_{a\in r_{x'}} p(a, b|x', y) \qquad \forall\, a, b, x, x', y, y', \end{aligned} \tag{2.3}$$

where r_i is the set of results for the i-th measurement, for $i = x, x', y, y'$. This condition is named as **nonsignalling** and is assumed for all Bell scenarios.

2.2 Locality condition

Let us now make the assumption that there is a causal influence between Alice's and Bob's experiment. Unlike the previous experiment where we supposed that there was a producer of the shoes who is the responsible for sending the packages for Alice's and Bob's store, here we have not even assumed what is the nature of the experiment conducted on Alice's and Bob's box.

According to the Reichenbach's principle, if the events are causally dependent, one of the statements is true: either Alice's experiment causally influences Bob's experiment, or Bob's experiment causally influences Alice's experiment, or there is a common cause that influences both Alice's and Bob's experiment. By hypothesis, the first and second sentences are not true. Still, we cannot establish, by elimination of alternatives, that the last sentence states the correct causality among the events. The aim of this section is no more than creating a model to verify plausibility of the existence of a common cause to Alice's and Bob's experiment. We will name it as a **local hidden variable** (or, shortly, LHV) model.

As we assumed nonsignalling between Alice and Bob, suppose there is an event Λ, encoded by variable λ, in the common past of Alice's and Bob's measurement events such that

$$p(a, b|x, y, \lambda) = p(a|x, \lambda)p(b|y, \lambda). \tag{2.4}$$

At first, there is no need to know λ, as the description of the experiment is made in terms of $p(a, b|x, y)$. Thus let us marginalized (2.2) w.r.t λ. With the aid of the Bayes' rule,

$$\begin{aligned} p(a, b|x, y) &= \sum_{\lambda} p(a, b, \lambda|x, y) \\ &= \sum_{\lambda} p(a, b|x, y, \lambda)p(\lambda|x, y). \end{aligned}$$

As Λ precedes the other events, $p(\lambda|x, y) = p(\lambda)$. Using (2.4),

$$p(a, b|x, y) = \sum_{\lambda} p(a|x, \lambda)p(b|y, \lambda)p(\lambda). \tag{2.5}$$

The above equation is known, as suggested by the title of this section, as **locality con-**

dition[1]. Every LHV model can be summarized in a probabilistic description satisfying equations (2.3) and (2.5). In addition, in a LHV model,

$$\sum_{a,b} p(a, b|x, y) = 1 \quad \text{and} \tag{2.6}$$

$$p(a, b|x, y) \geq 0, \quad \forall\, a, b, x, y, \tag{2.7}$$

which are known as normalization and non-negativity conditions, respectively. While (2.6) and (2.7) are trivial probability conditions, (2.3) represents the spatial separation between the boxes. However, equation (2.5) is special. If any correlation between Alice's and Bob's experiment can be localy explained, then its marginals, conditioned to an unknown variable λ, can be localy calculated as well. As the marginals are still λ-dependent we cannot say much about their general shape. There is a little trick which is attributed to Arthur Fine [18], where he proves that it is possible to obtain a deterministic LHV model if $p(a, b|x, y, \lambda)$ is factorizable as in (2.4). Let Λ_A and Λ_B be random variables such that for each combination of measurement, x and y, there is a distribution which is completely determined. Then,

$$d_A(a|x, \lambda_a) = \delta_{a,f_A(x,\lambda_a)} \quad \text{and} \quad d_B(b|y, \lambda_b) = \delta_{b,f_B(y,\lambda_b)},$$

where f_a and f_b denotes the λ_a-th and λ_b-th deterministic assignment of x to a and y to b, respectively. If so, there are r^m different assignments for both f_a and f_b, in such a way we are able to write the marginals $p(a|x, \lambda)$ and $p(b|y, \lambda)$ as

$$p(a|x, \lambda) = \sum_{\lambda_a}^{r^m} d_A(a|x, \lambda_a)p(\lambda_a|\lambda) \quad \text{and} \quad p(b|y, \lambda) = \sum_{\lambda_b}^{r^m} d_B(b|y, \lambda_b)p(\lambda_b|\lambda). \tag{2.8}$$

Substituting (2.8) in (2.5), then

$$\begin{aligned} p(a, b|x, y) &= \sum_{\lambda,\lambda_a,\lambda_b} d_A(a|x, \lambda_a)d_B(b|y, \lambda_b)p(\lambda_a|\lambda)p(\lambda_b|\lambda)p(\lambda) \\ &= \sum_{\lambda_a,\lambda_b} d_A(a|x, \lambda_a)d_B(b|y, \lambda_b)\left(\sum_{\lambda} p(\lambda_a|\lambda)p(\lambda_b|\lambda)p(\lambda)\right) \end{aligned}$$

[1]The original locality condition, the one formulated by John Bell in 1964, consider the distribution over the λ's as continuous. Actually there is no problem on considering λ as a discrete variable and, although the locality condition is written commonly with $p(\lambda)$ continuous, here we will adopt the discrete notation for convenience, as you will see soon.

$$:= \sum_{\lambda_a,\lambda_b} d_A(a\,|\,x,\,\lambda_a)d_B(b\,|\,y,\,\lambda_b)p(\lambda_a,\,\lambda_b), \tag{2.9}$$

where we defined the term in parenthesis as a new probability distribution which is λ_a, λ_b-dependent. Reshaping the pair $(\lambda_a,\,\lambda_b)$ into λ, we got

$$p(a,\,b\,|\,x,\,y) = \sum_{\lambda}^{r^{2m}} d_A(a\,|\,x,\,\lambda)d_B(b\,|\,y,\,\lambda)p(\lambda).$$

In other words, by considering the deterministic assigments d_A and d_B, we can insert the randomness of variables Λ_A and Λ_B into the hidden variable λ and write locality condition in terms of a countable set, d_A and d_B.

2.3 The local polytope

A very common way of understanding the probabilistic description of a Bell scenario consists in a beautiful geometrical interpretation of its probabilities called local polytope, attributed to Itamar Pitowsky [19]. To address this formalism, we will make a pause in the description of a generic Bell scenario and make a simple specification to continue.

Consider a (2, 2, 2) Bell scenario. On this scenario, it is possible to perform rounds of experiments and obtain a description in terms of 16 probabilities. We will group all of these in a vector:

$$\mathbf{p} = [\,p(0,0|0,0) \quad p(0,1|0,0) \quad ... \quad p(1,1|1,1)\,]^{\mathsf{T}}, \tag{2.10}$$

where $a,\,b,\,x,\,y \in \{0,1\}$, such that $p(0,0|0,0)$ denotes the probability of obtaining results $a, b = 0$ if the zero-th measurement is performed both by Alice and Bob. It is worth noting that this notation is no more than a label, which can be changed if convenient. The vector built on (2.10) is known as **correlation** or **behavior** of the (2, 2, 2) Bell scenario and describes the realization of the entire experiment after many rounds are performed in such a way the statistics is big enough to establish the full description of the experiment in function of its probabilities.

If settled as in (2.10), locality condition can be written as

$$\mathbf{p} = \sum_{\lambda} \mathbf{d}_A(\lambda) \otimes \mathbf{d}_B(\lambda)p(\lambda), \tag{2.11}$$

where $\mathbf{d}_A$ and $\mathbf{d}_B$ represent the local deterministic behaviors of Alice and Bob, respectively. The notation displayed in (2.11) induces us to see behaviors of a Bell scenario that satisfy the locality condition as points $\mathbf{p} \in \mathbb{R}^{16}$ which can be written as a convex combination weighted by the distribution of the hidden variables. This definition resembles the one of polytopes, the multidimensional analogue of polygons and polyhedra. A polytope can be thought as the geometrical site defined by the convex hull of a finite set of points. This way, as the combination on (2.11) is convex, this inspires us to define the set of behaviors which satisfy the locality condition as the **local polytope**.

Now, let $\mathbf{p}$ be a generic multidimensional point. According to Minkowsky-Weyl theorem [20, 21] a polytope can be equivalently represented as

1. the convex hull of a finite number V of points $\mathbf{v}_i$, these named as extremal points or vertices. Formally, if D is the dimension of the space in which the polytope $\mathcal{P}$ belongs then $\mathbf{p} \in \mathcal{P}$ if, and only if

$$\mathbf{p} = \sum_i^V c_i \mathbf{v}_i, \quad \text{with} \quad \sum_i^V c_i = 1 \quad \text{and} \quad c_i \geq 0 \quad \forall i. \tag{2.12}$$

   This is defined as V-representation.

2. the intersection of a finite number H of hyperplanes or half-spaces - a space with dimension $D-1$ which divides the space in two halves. Thus, if $\mathbf{p} \in \mathcal{P}$, then $\mathbf{p}$ satisfies all H inequalities such as

$$\mathbf{h}_j \cdot \mathbf{p} \leq h_j, \tag{2.13}$$

   where the pair $(\mathbf{h}_j, h_j)$ characterizes the j-th hyperplane. These hyperplanes are named as the facets of the polytope and equation (2.13) is defined as H-representation.

At this point of the chapter, we can finally establish a clear way to decide whether a specific behavior is or not explained by a LHV model. Once we picked a Bell scenario, it is possible to obtain the deterministic assignments by finding all of the r^{2m} combinations of extremal points in this scenario. Then, using the behavior of the experiment, we can check the factibility of a convex combination of their extremal points. If there is such a combination, then this behavior admits a LHV model, otherwise, LHV model will be disproved.

Despite the efficiency of the above method, there is a much more formal way which we can decide analytically the plausability of a LHV model. Using the set of extremal points we can convert the V-representation of the polytope to a H-representation by a process named facet enumeration, which will be addressed in more detail on chapter 4. If all of the facets of the local polytope provided by facet enumeration are satisfied by the behavior of the experiment, then, again, it admits a LHV model.

Facet enumeration process gives us two distinct kinds of facets. The first ones are the trivial facets. These have already been shown in equation (2.7), so they are redundant. In fact, when performing facet enumeration we must consider all constraints involved in a Bell scenario: non-negativity, normalization, nonsignalling and, of course, locality. So, it is expected the process returns trivial facets. The non-trivial facets receive a special name, the celebrated **Bell inequalities**, our main interest of investigation.

2.3.1 A remarkable case: the CHSH scenario

The simplest Bell scenario we can approach was already mentioned in the beginning of this section. The (2, 2, 2) Bell scenario is also given a special name in honor to Clauser, Horne, Shimony and Holt (CHSH), the first who dedicated themselves to study it in detail [22]. In simple terms, the CHSH scenario possesses a single inequality, known by the homonym name, which represents the non-trivial facet of the local polytope of the (2, 2, 2) Bell scenario. It is given by the expression

$$\begin{aligned} |\, p(a=b|0,0) - p(a\neq b|0,0) + p(a=b|0,1) - p(a\neq b|0,1) \\ + p(a=b|1,0) - p(a\neq b|1,0) - p(a=b|1,1) + p(a\neq b|1,1)\,| \leq 2, \end{aligned} \tag{2.14}$$

where $a = b$ denotes the sum of terms $a, b = \{(0,0),(1,1)\}$ and $a \neq b$ denotes the sum of $a, b = \{(0,1),(1,0)\}$. In fact, all Bell inequalities - let it be represented by I_{Bell} - can be written as

$$I_{\text{Bell}} = \sum_{a,b,x,y} c_{a,b,x,y}\, p(a,\, b\,|\, x,\, y) = \mathbf{c}\cdot\mathbf{p} \leq b_l, \tag{2.15}$$

where the sum runs over all combinations of $\{a,\, b\,|\, x,\, y\}$. Once they are the representation of the local polytope facets, then all I_{Bell} are linear on $\mathbf{p}$ and a single inequality can be given by a vector of coefficients $\mathbf{c}$ and a local bound b_l.

Inequality (2.14), however, is not the most known form of CHSH inequality. Defining

$\langle A_x B_y \rangle := p(a = b|\, x,\, y) - p(a \neq b|\, x,\, y)$ we rewrite

$$|\, \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \,| \leq 2. \tag{2.16}$$

The interest feature on equation (2.16) is that, if you trust on textbook quantum mechanics, we can easily disprove the LHV model for the CHSH scenario.

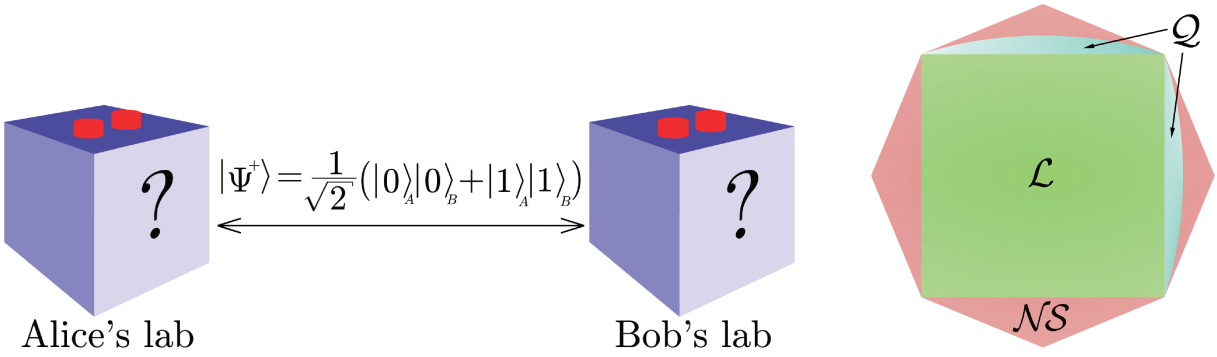
2.4 Bell's theorem

The aim of this entire section is to add one more assumption in a Bell scenario: the experiment contained in the boxes possesses quantum nature. For this reason, let us assume there is a two-part system described by ρ which is shared among Alice and Bob, as in figure 2.2a. The most general case is to consider that the measurement performed in both boxes is a POVM, then the probabilities obtained are given by

$$p(a,\, b|\, x,\, y) = \mathrm{tr}\left(\rho \mathbf{A}_{a|x} \otimes \mathbf{B}_{b|y}\right), \tag{2.17}$$

where $\mathbf{A}_{a|x}$ is the effect associated to the POVM $\{\mathbf{A}_{a|x}\}$ given by the choice x of Alice. The same notation holds for $\mathbf{B}_{b|y}$.

If written in form (2.17), the correlations obtained by quantum measurements repro-



(a) (b)

Figure 2.2: (a) Alice and Bob share the $|\Psi^+\rangle$ state. The subindices A and B refer to the qubits possessed by Alice and Bob, respectively. As $|\Psi^+\rangle$ is pure, then a density matrix can be represented by $\rho = |\Psi^+\rangle\langle\Psi^+|$ (b) A bidimensional representation of local, quantum and nonsignalling sets ($\mathcal{L}$, $\mathcal{Q}$ and $\mathcal{NS}$, respectively). The following inclusion relation is implicit: $\mathcal{L} \subset \mathcal{Q} \subset \mathcal{NS}$. For the sake of simplicity, we will represent $\mathcal{L}$ as a square, although an actual local polytope has tens of extremal points. As represented in figure, $\mathcal{L}$ and $\mathcal{NS}$ are polytopes. Although $\mathcal{Q}$ may have several faces, its number of extremal points is not finite, then not being a polytope.

duce all correlations which satisfies equation (2.5). Let us suppose a separable state ρ^{AB} is shared among Alice and Bob, as in (1.12). Then,

$$\begin{aligned} p(a,b\,|\,x,y) &= \mathrm{tr}\left(\rho\,\mathbf{A}_{a|x}\otimes\mathbf{B}_{b|y}\right) \\ &= \mathrm{tr}\left[\left(\sum_i p_i\,\rho_i^A\otimes\rho_i^B\right)\mathbf{A}_{a|x}\otimes\mathbf{B}_{b|y}\right] \\ &= \sum_i p_i\,\mathrm{tr}\left(\rho_i^A\mathbf{A}_{a|x}\right)\mathrm{tr}\left(\rho_i^B\mathbf{B}_{b|y}\right) \\ &= \sum_i p_i\,p(a\,|\,x,i)p(b\,|\,y,i). \end{aligned} \tag{2.18}$$

The wise choice of ρ_i^A, ρ_i^B and measurements $\{\mathbf{A}_{a|x}\}$ and $\{\mathbf{B}_{b|y}\}$, for all i, x and y, enable us to reproduce the local correlations. Quantum correlations also satisfy nonsignalling conditions. We can always marginalize (2.17) by taking the sum over the other part's result and performing identity measurement, whatever is this result.

This way, let $\mathcal{L}$ be the local polytope of an arbitrary scenario. $\mathcal{L}$ is such that it contains correlations satisfying equations (2.3), (2.5), (2.6) and (2.7), as already mentioned. Similarly, we will define $\mathcal{NS}$ as the nonsignalling set. $\mathcal{NS}$ is less restrictive than $\mathcal{L}$, once condition (2.5) is not required for correlations belonging to $\mathcal{NS}$. It is also known that the $\mathcal{NS}$ also forms a polytope and some of its extremal points are coincident with $\mathcal{L}$ [23]. The other extremal points of the $\mathcal{NS}$ polytope are known as PR boxes [24], a non-local correlation which still satisfies nonsignalling condition. Finally, let $\mathcal{Q}$ be the set of quantum correlations. By the previous considerations, then is possible to infer that $\mathcal{L}\subset\mathcal{Q}\subset\mathcal{NS}$ as suggested in figure 2.2b. The intriguing question is that if $\mathcal{L}\supset\mathcal{Q}$, or equivalently $\mathcal{L}=\mathcal{Q}$.

For this purpose, let us substitute (2.17) on (2.15). Thus, all Bell inequalities can be written in the quantum regime as

$$I_{\mathrm{Bell}} = \sum_{a,b,x,y} c_{a,b,x,y}\,\mathrm{tr}\left(\rho\,\mathbf{A}_{a|x}\otimes\mathbf{B}_{b|y}\right) \equiv \mathrm{tr}\left(\rho\,G_{\mathrm{Bell}}\right) \leq b_l, \tag{2.19}$$

where we defined $G_{\mathrm{Bell}} \coloneqq \sum_{a,b,x,y} c_{a,b,x,y}\,\mathbf{A}_{a|x}\otimes\mathbf{B}_{b|y}$ as the Bell observable. On notation of equation (2.16), the Bell observable for the CHSH scenario can be written as

$$G_{\mathrm{CHSH}} = A_0\otimes B_0 + A_0\otimes B_1 + A_1\otimes B_0 - A_1\otimes B_1.$$

If, for the CHSH scenario, we associate the following observables for Alice and Bob

$$A_0 = \sigma_x, \quad A_1 = \sigma_z, \quad B_0 = \frac{1}{\sqrt{2}}(\sigma_x + \sigma_z) \quad \text{and} \quad B_1 = \frac{1}{\sqrt{2}}(\sigma_x - \sigma_z),$$

and take the shared state among then as $\rho = |\Psi^+\rangle\langle\Psi^+|$, then, following the notation on (2.19), we obtain

$$2\sqrt{2} > b_l = 2,$$

which violates the CHSH local bound. This counterexample for the LHV model on the CHSH scenario is conclusion of the **Bell's theorem**. In other words, Bell's theorem states that a LHV model is not plausible for the CHSH scenario, but much more than a mere conclusion for a single scenario: it breaks the logic behind the Reichenbach's common cause principle, that is, we established a pair of events which are correlated between and, as one is spatially apart from the other, then the unique probable alternative, which would be to explain the correlation by a common cause, is discarded by Bell's theorem. This phenomenon is called **Bell**[2] **nonlocality** and responds our question: $\mathcal{L} \not\supset \mathcal{Q}$.

This result is so strong, that it changes the status of quantum mechanics from local to nonlocal theory. That is, it still needed an experimental verification, which was also performed many times since Bell's theorem was enunciated in 1964. A satisfactory end to the experimental verification to a violation of the CHSH inequality was only provided in 2015 with three independent significant[3]-loophole-free experiments [25, 26, 27]. Previously, some honorable mentions are worth to be said.

The first experimental attempt to prove any violation of Bell inequalities was conducted by Stuart Freedman and John Clauser (the 'C' of CHSH), who tried to find violations of Clauser-Horne inequality - a variant by labels changing of CHSH inequality [28]. They were followed by Alain Aspect and colleagues who found a violation of CHSH inequality [29]. In Aspect's experiment, two main loopholes were identified and pointed out, these being the detection loophole and the locality loophole. The first one is related to the fact that the detection is not fully efficient, and there may be undetected samples that would compromise the value of the inequality beyond the local bound. Detection

[2]The use of the word 'Bell' preceding 'nonlocality' is made to differentiate an arbitrary non-local action to the nonlocality allowed by quantum mechanics in the framework we have just presented.

[3]The suffix significant from significant-loophole-free, refers to detection and locality loopholes. You might include as many loopholes as your paranoia allows you and, in some cases, they would not be possible to be tested, therefore being not significant.

loophole was completely closed on a 2001 experiment [30] which used detectors with efficiency above 90%. The locality loophole remains in situations in which the parts of the experiment could be close enough to the point that there could be some kind of information exchange. It was closed in a 1998 experiment led by Anton Zeinlinger [31], where they used a clever mechanism to ensure the choice of measurement was completely random, then eliminating any correlation which could remain on the measurement apparatus.

2.4.1 How to achieve nonlocality?

We have just provided an example of a set of state and measurement operators which are able to violate (2.16). Of course it was a not a shot or a set of operators selected randomly, but a didactic model with all the ingredients needed to exemplify nonlocality, leaving unanswered, therefore, the question that entitles this subsection.

These ingredients are two. The first of them, which must have already become obvious at this point, is entanglement. This is clearly shown on equation (2.18), as a separable state always provides a factorisable joint probability. Although entangled states are necessary to obtain nonlocality, they are not sufficient as shown by Reinhard Werner [32].

The second one is also a necessary condition and was deducted by the aforementioned A. Fine on [18]. The celebrated work of Fine demonstrates that, if a Bell inequality is satisfied, then it must have a joint probability distribution for all the observables involved in the experiment. In the CHSH scenario, for instance, if there exists a distribution $p(A_0, A_1, B_0, B_1)$, the CHSH inequality is satisfied. Such a distribution is possible if, and only if, all the observables are simultaneously measurable. In fact, the observables possessed by Alice are always simultaneously measurable with the Bob's observables. Thus, the existence of such a distribution arises on the fact that Alice's observables must be mutually compatible among themselves. The same holds for Bob's observables. Fine's condition can be summarized requiring that the observables of each of the parts are necessarily incompatible with each other.

2.4.2 A limit for quantum violations

Let us proceed with another intriguing question. So far, we have shown that we indeed can obtain non-local behaviors for a Bell scenario. However, a regular characterization of this description must also include the question: how non-local can a behavior be? A gross

maximization of (2.14) suggest that if all of the positive terms are equal to one and the negative ones are zero, then we can achieve four on inequality. This is in fact achievable by some of the extremal points of the $\mathcal{NS}$ set; for instance, the behavior

$$p(a, b\,|\,x, y) = \begin{cases} 1/2, & \text{if } a \oplus b = xy \\ 0, & \text{otherwise} \end{cases}, \tag{2.20}$$

where the $\oplus$ symbol denotes sum modulo two. Equation (2.20) represents one of the aforementioned PR boxes, that is, one of the extremal points of the $\mathcal{NS}$ polytope which achieves the **geometrical bound** of the CHSH inequality.

Until now, we have shown that the inclusion relations between $\mathcal{L}$, $\mathcal{Q}$ and $\mathcal{NS}$ are such that $\mathcal{L} \subset \mathcal{Q} \subset \mathcal{NS}$ and $\mathcal{L} \not\supset \mathcal{Q}$. The result that will be presented in this subsection resolves the inclusion relation $\mathcal{Q} \not\supset \mathcal{NS}$ and it is known as Tsirelson's bound [33]. Again, let us start with the CHSH scenario and consider, as before, the Bell observable G_{CHSH} where the spectrum of observables A_0, A_1, B_0 and B_1 corresponds to ± 1. The square of G_{CHSH} can be written as

$$G^2_{\text{CHSH}} = 4 \times \mathbb{1}_{4\times 4} + [A_0,\, A_1] \otimes [B_0,\, B_1].$$

Now, consider that $\|A_0\|$ denotes the greatest eigenvalue of A_0. Then, the Cauchy-Scharwz inequality applied to the commutator of A_0 and A_1 produces $\|[A_0,\, A_1]\| \leq 2\|A_0\|\,\|A_1\|$. If considered the similar inequality for B_0 and B_1, then

$$\|G^2_{\text{CHSH}}\| \leq 4 + 4\|A_0\|\,\|A_1\|\,\|B_0\|\,\|B_1\| \leq 8, \quad \text{and}$$
$$\|G_{\text{CHSH}}\| \leq 2\sqrt{2}. \tag{2.21}$$

Despite this simple verification was not the original method used by Boris Csirel'son on [33], it efficiently find out a quantum bound for quantum violations of CHSH inequality, i.e.,

$$|\,\langle A_0B_0\rangle + \langle A_0B_1\rangle + \langle A_1B_0\rangle - \langle A_1B_1\rangle\,| \overset{b_l}{\leq} 2 \overset{b_q}{\leq} 2\sqrt{2} \overset{b_g}{\leq} 4,$$

where b_l, b_q and b_g denotes the local, quantum and geometrical bounds, respectively.

2.5 What about Charlie?

All what we have been showing so far considered a Bell scenario composed by only two parts. As the title of this thesis clearly refers to nonlocality on multipartite systems it is

already time to present scenarios with more than two parts. By the way, it is suitable to denote the common language for multipartite scenario. From now on, we will refer to the two parts scenarios as **bipartite**. For a three parts scenario we will use the terminology **tripartite** and so on.

The simplest tripartite Bell scenario is the (3, 2, 2). Like the CHSH scenario it also receives a special name: the **Śliwa** scenario, due to the study conducted on [34] by Cezary Śliwa. As it is composed of one more part, the degree of complexity of the Śliwa scenario is much larger than its bipartite analogous. Indeed, the resolution of this scenario is far more recent, it comes from 2003, while CHSH inequality dates from 1969. Unlike the former case, which has a single non-trivial inequality, Śliwa scenario has 46 classes[4] which will not be shown for obvious reasons. The interested reader could see, e.g., reference [34, e-print].

There are some specificities which turn multipartite scenarios more interesting than bipartite ones. Some of the constraints presented before are trivial and can be straightly extended, such as locality condition for three parts,

$$p(a,\, b,\, c\,|\,x,\, y,\, z\,) = \sum_{\lambda} p(a\,|\,x,\, \lambda)p(b\,|\,y,\, \lambda)p(c\,|\,z,\, \lambda)p(\lambda), \tag{2.22}$$

where the simple addition of a marginal for Charlie is needed. But some conditions change a lot with the addition of the third part, so the extension is no longer trivial. For instance, w.r.t. the nonsignalling condition,

$$\begin{aligned}
p(a\,|\,x) &= \sum_{\substack{b\in r_y \\ c\in r_z}} p(a,\, b,\, c\,|\,x,\, y,\, z\,) = \sum_{\substack{b\in r_{y'} \\ c\in r_{z'}}} p(a,\, b,\, c\,|\,x,\, y',\, z') \\
p(b\,|\,y) &= \sum_{\substack{a\in r_x \\ c\in r_z}} p(a,\, b,\, c\,|\,x,\, y,\, z\,) = \sum_{\substack{a\in r_{x'} \\ c\in r_{z'}}} p(a,\, b,\, c\,|\,x',\, y,\, z') \\
p(c\,|\,z) &= \sum_{\substack{a\in r_x \\ b\in r_y}} p(a,\, b,\, c\,|\,x,\, y,\, z\,) = \sum_{\substack{a\in r_{x'} \\ b\in r_{y'}}} p(a,\, b,\, c\,|\,x',\, y',\, z\,) \qquad \forall\, a,b,c,x,x',y,y',z,z',
\end{aligned} \tag{2.23}$$

where these are the nonsignalling for the single marginals. We can still consider a

[4] Check on appendix A for a clarifying discussion about **classes** of inequalities.

nonsignalling condition for the marginals two by two, as it follows:

$$\begin{aligned} p(a,\,b|\,x,\,y) &= \sum_{c\in r_z} p(a,\,b,\,c|\,x,\,y,\,z) = \sum_{c\in r_{z'}} p(a,\,b,\,c|\,x,\,y,\,z') \\ p(a,\,c|\,x,\,z) &= \sum_{b\in r_y} p(a,\,b,\,c|\,x,\,y,\,z) = \sum_{b\in r_{y'}} p(a,\,b,\,c|\,x,\,y',\,z) \\ p(b,\,c|\,y,\,z) &= \sum_{a\in r_x} p(a,\,b,\,c|\,x,\,y,\,z) = \sum_{a\in r_{x'}} p(a,\,b,\,c|\,x',\,y,\,z) \qquad \forall\, a,b,c,x,x',y,y',z,z'. \end{aligned} \tag{2.24}$$

Then, as you can see, with the addition of one more part, we obtain several more nonsignalling equations than in the bipartite case.

One interesting phenomenon which is not implicit in bipartite cases is that entanglement and nonlocality are not limited to the parts two by two. The idea of entanglement for any tripartite case can be obtained extending equation (1.12) for three parts,

$$\rho^{ABC} \neq \sum_i p_i\, \rho_i^A \otimes \rho_i^B \otimes \rho_i^C. \tag{2.25}$$

Despite of being the correct extended notion of tripartite entanglement, a stronger form of equation (2.25) can be also obtained.

$$\rho^{ABC} \neq \sum_i \left(p_{i1}\, \rho_i^{AB} \otimes \rho_i^C + p_{i2}\, \rho_i^{AC} \otimes \rho_i^B + p_{i3}\, \rho_i^A \otimes \rho_i^{BC} \right), \tag{2.26}$$

where $\sum_{i,j} p_{ij} = 1$, $j = 1,\, 2,\, 3$, and $p_{ij} \geq 0$. This is the definition of **genuinely tripartite entanglement** and attached to it is the concept of **genuinely tripartite nonlocality**, that is, a behavior which does not admit a decomposition such

$$p(a,\,b,\,c|\,x,\,y,\,z) = p_1\, p(a,\,b|\,x,\,y)p(a|\,x) + p_2\, p(a,\,c|\,x,\,z)p(b|\,y) + p_3\, p(b,\,c|\,y,\,z)p(c|\,z), \tag{2.27}$$

where $p_i \geq 0$ and $\sum_i p_i = 1$, for i =1, 2, 3. This definition is attributed to George Svetlichny in his 1987 work [35].

Of course there is a huge path between establishing Bell inequalities - as the 46 classes of Śliwa inequalities - and testifying the plausibility of nonlocality for each of them. Differently of the CHSH case, where we could easily propose a set of state and measurement operators which violates CHSH inequality, for a tripartite scenario this task is reasonably hindered. There are, however, some numerical solutions you can check on appendix B.

In 1987, Svetlichny proposed a pair of inequalities which could only be violated by a genuine tripartite entangled state:

$$\begin{aligned} |\langle A_0B_0C_0\rangle + \langle A_0B_0C_1\rangle + \langle A_1B_0C_0\rangle - \langle A_1B_0C_1\rangle + \langle A_0B_1C_0\rangle \\ - \langle A_0B_1C_1\rangle - \langle A_1B_1C_0\rangle - \langle A_1B_1C_1\rangle| \leq 4 \quad \text{and} \\ |\langle A_0B_0C_0\rangle - \langle A_0B_0C_1\rangle - \langle A_1B_0C_0\rangle - \langle A_1B_0C_1\rangle - \langle A_0B_1C_0\rangle \\ - \langle A_0B_1C_1\rangle - \langle A_1B_1C_0\rangle + \langle A_1B_1C_1\rangle| \leq 4. \end{aligned} \tag{2.28}$$

Inequalities (2.28) also belong to the (3, 2, 2) Śliwa scenario, but differently from the remaining inequalities, any non-local behavior which violates the above expresssions must be more strongly correlated than in (2.25).

This short description provided in the end of this chapter finishes our considerations about multipartite nonlocality. Our next step will be the fully characterization of scenarios with causal relaxations, which is the main subject of this thesis. On this new framework, all the interesting scenarios are multipartite and the immediate goal is to find their inequalities.

3 Super nonlocality

In the ending of last chapter we briefly presented a notion of nonlocality which is stronger than the usual nonlocality idea in a tripartite scenario - the genuine multipartite nonlocality. On this chapter, another concept of nonlocality, also stronger than the common idea, will be presented with the - not that original - name of **super nonlocality**.

When studying tripartite scenarios an evident feature is its grown complexity. Whereas in the bipartite CHSH scenario we have a single representative inequality, for the Śliwa scenario this number grows to 46. This means not only an increase in the computational complexity of a multipartite problem, but also in the characterization of a scenario, according as more structure is added. For this reason, this chapter is intended to present a framework known as causal relaxations which can be understood as a second layer of causal structure in the standard Bell scenario.

To introduce this new and rich concept, chapter three will be divided in two parts: the first one is dedicated to explore some preliminary concepts necessary to the introduction of causal relaxations. The second half will explain the concept of causal relaxations and the notion of nonlocality that emerges from it and entitles this chapter.

3.1 A sensible definition of causal structure

The concepts which will be addressed in this section are based in the seminal work by Judea Pearl [36] and some other interesting related works, which will be opportunely referred in the body of the text.

Up to this point, the reader must have already noted that this work is about inferring causal correlations among a set of events. Given a pair of events, we may be interested in

determining if they are directly correlated, i.e., one event causes the second, if they are indirectly correlated, as in Reinchenbach's principle, or even if they are anyway uncorrelated. We have already explored this question in section 2.4 and now we will provide it a slightly different formalism.

Now we will encode the random variables by nodes in a graph. Whenever an event v_1 direct influence an event v_2, we will note it as $v_1 \to v_2$, the graphical representation of a directed graph. A directed graph $G = (V, E)$ is a set of vertices V (or nodes) combined with a set of relating edges E. In a directed graph, such as $v_1 \xrightarrow{e} v_2$, v_1 is denoted as source of e. Similarly, v_2 is denoted as target of e. The relation between v_1 and v_2 is also named as follows: the set of parents of v_2 is represented as $\mathrm{pa}(v_2) = \{v_1\}$ and, similarly, $\mathrm{ch}(v_1) = \{v_2\}$ denotes the set of children of v_1.

In the language of [37, def. 3.1], a directed path between v_1 and v_2 is given by a finite number N of nodes $n_1, ..., n_N$ such that $v_1 \to n_1 \to ... \to n_N \to v_2$, where $v_1, v_2, n_1, ..., n_N \in V$. We can also use the shorthand $v_1 \rightsquigarrow v_2$ to indicate the existence of a directed path between v_1 and v_2. It also worth to include that if $v_1 \rightsquigarrow v_2$ then they can be said transitive. In the light of these concepts one can state that

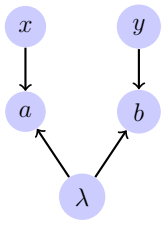
Definition 3.1 (causal structure)

Any causal structure can be represented by a graph $G = (V, E)$ such that $v \not\rightsquigarrow v$, for all $v \in V$. Such a graph is denoted as a directed acyclic graph (DAG). For the non-transitivity of the vertices, zero length paths must not be implied.

In summary, definition 3.1 claims that any causal structure cannot contain a cycle, in such a way an event may be associated to its own cause. Henceforth, whenever a Bell scenario is presented, there will be an associated DAG representing its causal structure. But before we go into this, let us extract a second formalism from the causal structures.

3.1.1 Classical bayesian networks

In the Bell scenarios we have explored so far, the causal structure inherent to them is already implicit. For instance, consider a generic bipartite Bell scenario, where the locality condition is given by equation (2.5). Alice's and Bob's local distribution indicate that a and b variables are directly influenced by x and y, respectively, and jointly by λ, as we can see in the associated DAG:



(3.1)

In this way, the locality condition is no more than a consequence of the LHV model causal structure, reason why we can associate, for every locality condition, in any Bell scenario, a respective DAG. We can attach these two concepts using a very convenient definition, as in [38, 39]:

Definition 3.2 (bayesian networks)

1. (Markov condition) A DAG $G(V, E)$ is Markov relative to a distribution p if

$$p(v_1, ..., v_n) = \prod_i^n p(v_i \,|\, \mathrm{pa}(v_i)), \tag{3.2}$$

   where $v_1, ..., v_n \in V$.
2. A classical Bayesian network is defined by a DAG G and its Markov related distribution, p.

The application of definition 3.2 is straightforward. Again, for a bipartite Bell scenario, we can use equations (3.2) and (3.1) to write

$$p(a, b, x, y, \lambda) = p(a|x, \lambda)\, p(b|y, \lambda)\, p(x)\, p(y)\, p(\lambda).$$

Marginalizing over λ and using Bayes' rule gives us

$$p(a, b|x, y) = \frac{p(a, b, x, y)}{p(x)\, p(y)} = \sum_\lambda p(a|x, \lambda)\, p(b|y, \lambda)\, p(\lambda), \tag{3.3}$$

which equals equation (2.5).

3.2 Causal relaxations

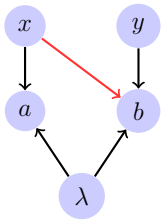
Let us now put aside the causal structures for a moment and go back a few steps in the formulation of a Bell *gedankenexperiment*. In a typical Bell scenario there are some causal assumptions we make which are natural for the development of the concept, such as the free will of Alice and Bob to choose their own measurements - an assumption we denote

as measurement independence - and the locality assumption[1], which is implicit in the nonsignalling conditions.

Several works aimed to investigate Bell scenarios when one of these causal assumptions are given up totally or partially. A Bell scenario in which one of the causal assumptions is omitted is said relaxed, thus giving rise to two new frameworks: locality causal relaxations and measurement independence causal relaxations. While the latter is based on how much independent Alice's and Bob's measurements must be in order to still achieve nonlocality [40, 41], the former concerns on how communicating parts affect the result of a nonlocality experiment, either simulating [42, 43] or quantifying it [44].

We can define a locality causal relaxation as any direct influence among the parts, like the communication of one bit or any analogous strategy, which must be local to both of the involved parts, and thus can be clearly represented by DAG's. With respect to measurement independence causal relaxation, its name is self-explanatory; any nonrandom strategy of choice in the measurement set of observables or a hidden variables correlated strategy may be considered measurement dependent [40].

The formal description of a relaxed Bell scenario is provided by a Bayesian network. Let us propose a scenario in which Alice classically communicates with Bob - but not the opposite - which configures a locality causal relaxation. It can be represented by the DAG



$$\Rightarrow \quad p(a,\, b|x,\, y) = \sum_{\lambda} p(a|x,\, \lambda)\, p(b|x,\, y,\, \lambda)\, p(\lambda) \tag{3.4}$$

where a red arrow was used to indicate the classical communication. Also, we must exclude one of the nonsignalling equations, since Bob's marginal is now x-dependent, $p(b|x,\, y)$, and the nonsignalling condition implies only

$$p(a\,|\,x) = \sum_{b \in r_y} p(a,\, b\,|\,x,\, y) = \sum_{b \in r_{y'}} p(a,\, b\,|\,x,\, y') \qquad \forall\, a, b, x, y, y'. \tag{3.5}$$

[1]There is also a third assumption, realism, which implies in the previous existence of a value for any physical measurable quantities. This assumption is usually encompassed in the locality assumption under the local-realism hypothesis. The overall conclusion in a Bell experiment is that you must either abdicate of locality or realism. As we are not interested in this specificity, we shall consider the locality assumption as one.

In a scenario like the one depicted in the above relaxed DAG, equation (3.4) works as a new locality-like condition which must be compatible with condition (3.5): we must admit some of the nonsignalling correlations that are simulated by the communication link from Alice to Bob. In this relaxed scenario, any non-local like behavior is considered as **super non-local** and the set of correlations satisfied by this model will be referred as the relaxed polytope. The relaxed polytope is no more than an inflated local polytope; while condition (3.4) also admits the usual local behaviors, condition (3.5) abdicates of some of the nonsignalling extremal points.

Also, the relaxed scenario on (3.4) is known in the literature as causal parameter-dependent (CPD) model [44]. For the purpose of this thesis it also is appropriate to characterize the CPD models as **boring**. For instance, in the (2, 2, 2) Bell scenario, all of the nonsignalling correlations - including its extremal points, the PR boxes - can be reproduced with the communication of a single classical bit between the parts [43]. Indeed, for several bipartite cases similar to the CPD model, either the relaxed scenario can be considered boring, for a practical number of inputs and outputs, or their characterization based on relaxed Bell inequalities has already been explored [44].

3.2.1 Hierarchy for tripartite relaxed scenarios

However, in the tripartite case, the number of combinations of causal relaxations grows severally as well as the number of possible resultant scenarios, thus there are many scenarios which remain unexplored. For this reason, it is necessary to establish some criteria that reduce the total number of cases in a few representative ones. This is, briefly, the result of [45] which creates a hierarchy of Bell relaxed scenarios in terms of the amount of causal relaxations established among the variables.

Although the result of [45] is valid for the multipartite case, we will address specifically the tripartite scenarios. The hierarchy is obtained by a couple of lemmas which are based on the concept of **nonsignalling equivalence**. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be the DAG's of two distinct tripartite relaxed scenarios. If every nonsignalling correlation produced by the Bayesian network of $\mathcal{G}_1$ is also produced by the Bayesian network of $\mathcal{G}_2$, then $\mathcal{G}_1$ nonsignalling implies $\mathcal{G}_2$ and we denote as $\mathcal{G}_1 \rightarrow \mathcal{G}_2$. Moreover, if $\mathcal{G}_2 \rightarrow \mathcal{G}_1$, then $\mathcal{G}_1$ and $\mathcal{G}_2$ are nonsignalling equivalent.

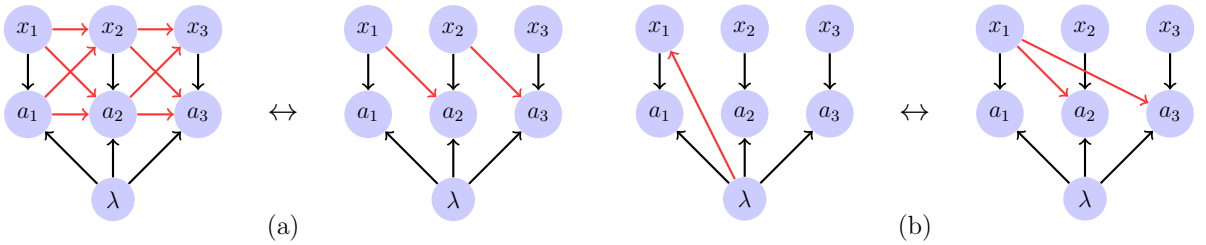


Figure 3.1: (a) Representation of lemma 3.3. Again, the red arrows represent the locality causal relaxations among the parts. We can observe generic locality causal relaxations between parts 1 and 2 and 2 and 3, but not 3 and 1. Its nonsignalling equivalent graph is shown at its right, with locality causal relaxations between x_1 and a_2 and x_2 and a_3 (b) Representation of lemma 3.4 with a single equivalence (also, we can consider the inverse of the left hand graph as equivalent, $\text{pa}(\lambda) = \{x_1\}$). Now, the red arrows are representing a measurement independence causal relaxation between the local variable λ and the measurement free choice. Any measurement independence causal relaxation is nonsignalling equivalent to the broadcasting of inputs among the parts.

Consider a generic locality causal relaxation in a tripartite scenario[2]. Worth to define what do we mean by generic, i.e., a locality causal relaxation between parts i and j such that $\text{pa}(x_i) \supseteq \{a_j, x_j\}$ or $\text{pa}(a_i) \supseteq \{a_j, x_j\}$ for any $i > j$, with $i, j = 1, 2, 3$. (see figure 3.1a). This way we are able to enunciate the below lemma.

Lemma 3.3

Any generic tripartite relaxed scenario between parts i and j posses a nonsignalling equivalent relaxed scenario which is obtained by performing a locality causal relaxation from the input measurement x_j to its output response a_i such that $\text{pa}(a_i) \supseteq \{x_j\}$, as depicted in figure 3.1a.

Proof. As it involves parts i and j we can consider the proof over the subgraph involving only parts i and j. A marginalization over λ provides

$$\begin{aligned} p(a_j,\, a_i \,|\, x_j,\, x_i) &= \sum_\lambda p(a_j,\, a_i,\, \lambda \,|\, x_j,\, x_i) \\ &= \sum_\lambda p(a_j,\, a_i \,|\, x_j,\, x_i,\, \lambda)\, p(\lambda \,|\, x_j,\, x_i), \end{aligned}$$

where Bayes' rule was used over λ. As x_j is neither parent nor children of λ, its influence is trivial and

$$p(a_j,\, a_i \,|\, x_j,\, x_i) = \sum_\lambda p(a_j,\, a_i \,|\, x_j,\, x_i,\, \lambda)\, p(\lambda \,|x_i)$$

[2]For the enunciation of lemmas 3.3 and 3.4, we will use a slightly different notation for the parts. Let x_1 be the measurement performed by the 1st part, Alice, and a_1 be the result of its measurement. A similar notation holds for Bob and Charlie.

$$
\begin{aligned}
&= \sum_{\lambda} p(a_j,\, a_i \,|\, x_j,\, x_i,\, \lambda)\, \frac{p(x_i\,|\lambda)}{p(x_i)}\, p(\lambda) \\
&= \sum_{\lambda} p(a_i \,|\, a_j,\, x_j,\, x_i,\, \lambda)\, p(a_j \,|\, x_j,\, x_i,\, \lambda)\, \frac{p(x_i\,|\lambda)}{p(x_i)}\, p(\lambda),
\end{aligned}
$$

where Bayes' rule was used twice in $p(\lambda\,|x_i)$ and $p(a_j,\, a_i \,|\, x_j,\, x_i,\, \lambda)$. We can associate the randomness of a_j to λ and make a_j deterministic to the choice of its measurement, x_j, then

$$
p(a_j,\, a_i \,|\, x_j,\, x_i) = \sum_{\lambda} p(a_i \,|\, x_j,\, x_i,\, \lambda)\, p(a_j \,|\, x_j,\, \lambda)\, \frac{p(x_i\,|\lambda)}{p(x_i)}\, p(\lambda).
$$

Note that the above equation represents a Bayesian network associated to a generic relaxed scenario as defined earlier. As λ is parent of a_j and $\mathrm{pa}(x_i) \supseteq \{a_j, x_j\}$, then we cannot omit its influence over $p(x_i\,|\lambda)$. Instead, consider a marginalization of $p(x_i\,|\lambda)$ in terms of a_j; therefore, using Bayes' rule one more time,

$$
\begin{aligned}
p(x_i\,|\lambda) &= \sum_{a_j} p(x_i,\, a_j \,|\, \lambda) \\
&= \sum_{a_j} p(x_i \,|\, a_j,\, \lambda) p(a_j \,|\, \lambda) \\
&= \sum_{a_j} p(x_i \,|\, a_j) p(a_j \,|\, \lambda),
\end{aligned}
$$

since x_i is neither parent nor children of λ.

Returning to, $p(a_j,\, a_i \,|\, x_j,\, x_i)$,

$$
p(a_j,\, a_i \,|\, x_j,\, x_i) = \sum_{\lambda, a'_j} p(a_i \,|\, x_j,\, x_i,\, \lambda)\, p(a_j \,|\, x_j,\, \lambda)\, \frac{p(x_i \,|\, a'_j)}{p(x_i)}\, p(a'_j \,|\, \lambda)\, p(\lambda),
$$

note that, from Bayes' rule, the following identity

$$
p(x_i \,|\, a'_j) = \frac{p(a'_j \,|\, x_i)\, p(x_i)}{p(a'_j)} \quad \Rightarrow \quad \frac{p(x_i \,|\, a'_j)}{p(x_i)} = \frac{p(a'_j \,|\, x_i)}{p(a'_j)},
$$

equals one; once $\mathrm{pa}(a_j) \not\supseteq x_i$, $p(a'_j \,|\, x_i) \equiv p(a'_j)$ and

$$
\begin{aligned}
p(a_j,\, a_i \,|\, x_j,\, x_i) &= \sum_{\lambda, a'_j} p(a_i \,|\, x_j,\, x_i,\, \lambda)\, p(a_j \,|\, x_j,\, \lambda)\, p(a'_j \,|\, \lambda)\, p(\lambda) \\
&= \sum_{\lambda} p(a_i \,|\, x_j,\, x_i,\, \lambda)\, p(a_j \,|\, x_j,\, \lambda)\, p(\lambda),
\end{aligned}
$$

where we can infer that $\mathrm{pa}(a_i) \supseteq x_j$. ■

Lemma 3.3 reduces all of the generic locality relaxed scenarios to a single relaxation among the involved parts which are sourced from an input measurement and targeted to another part's result. It simplifies a lot the total number of combinations of locality causal relaxations, but it does not address other important case when we have a causal relaxation between the hidden variable λ and the choice of a measurement for the j-th part, that is, it does not consider any measurement independence causal relaxation, which is the subject of the next lemma.

Lemma 3.4

Consider three different tripartite relaxed Bell scenarios. The first and second scenarios are such that posses a single relaxation, $\mathrm{pa}(x_i) = \{\lambda\}$ and $\mathrm{pa}(\lambda) = \{x_i\}$, for $i = 1, 2, 3$. The third one is such that $x_i \in \mathrm{pa}(a_j)$ for $j = 1, 2, 3$. These three scenarios are nonsignalling equivalent, as indicated on figure 3.1b.

Proof. Consider the condition produced by the Bayesian network associated to the right hand graph on figure 3.1b:

$$p(a_1,\, a_2,\, a_3|x_1,\, x_2,\, x_3) = \sum_{\lambda} p(a_1|x_1,\, \lambda)\, p(a_2|x_1,\, x_2,\, \lambda)\, p(a_3|x_1,\, x_3,\, \lambda)\, p(\lambda).$$

As we did on (2.9), consider the deterministic assignments to each of the marginals involved,

$$\begin{aligned} d_1(a_1\,|\,x_1,\, \lambda_1) &= \delta_{a_1, f_1(x, \lambda_1)}, \\ d_2(a_2\,|\,x_1,\, x_2,\, \lambda_2) &= \delta_{a_2, f_2(x_1, x_2, \lambda_2)} \quad \text{and} \\ d_3(a_3\,|\,x_1,\, x_3,\, \lambda_3) &= \delta_{a_3, f_3(x_1, x_3, \lambda_3)}, \end{aligned}$$

where f_1 denotes the λ_1-th deterministic assignment of x_1 to a_1 and f_i denotes the λ_i-th deterministic assignment of x_i and x_1 to a_i, for $i = 2, 3$. The above deterministic assignments provide

$$p(a_1\,|\,x_1,\, \lambda) = \sum_{\lambda_1}^{r_1^{m_1}} d_1(a_1\,|\,x_1,\, \lambda_1) p(\lambda_1\,|\,\lambda),$$

$$
\begin{aligned}
p(a_2\,|\,x_1,\,x_2,\,\lambda) &= \sum_{\lambda_2}^{r_2^{m_1\times m_2}} d_2(a_2\,|\,x_1,\,x_2,\,\lambda_2)p(\lambda_2\,|\,\lambda)\\
p(a_3\,|\,x_1,\,x_3,\,\lambda) &= \sum_{\lambda_3}^{r_3^{m_1\times m_3}} d_3(a_3\,|\,x_1,\,x_3,\,\lambda_3)p(\lambda_3\,|\,\lambda),
\end{aligned}
$$

where m_i is the measurement set of the i-th part and r_i is the associated result set to the m_i-th measurement of the i-th part, for $i = 1, 2, 3$.

Similarly to (2.9), the former definitions yield,

$$
\begin{aligned}
p(a_1,\,a_2,\,a_3|x_1,\,x_2,\,x_3) &= \sum_{\lambda} p(a_1|x_1,\,\lambda)\,p(a_2|x_1,\,x_2,\,\lambda)\,p(a_3|x_1,\,x_3,\,\lambda)\,p(\lambda)\\
&= \sum_{\lambda_1,\lambda_2,\lambda_3} d_1(a_1\,|\,x_1,\,\lambda_1)\,d_2(a_2\,|\,x_1,\,x_2,\,\lambda_2)\,d_3(a_3\,|\,x_1,\,x_3,\,\lambda_3)\,p(\lambda_1,\,\lambda_2,\,\lambda_3).
\end{aligned}
$$

Now, note that fixing λ_i and x_1, f_i becomes a single argument assignment function of x_i into a_i, for $i = 2, 3$. Once λ_i and x_1 posses $r_i^{m_1\times m_i}$ and m_1 different values respectively, then fixing λ_i and x_1 produces $r_i^{m_1\times m_i}\times m_1/r_i^{m_i}$ same single argument assignment function of x_i into a_i. Thus, let us relabel

$$
f_i^*(x_i,\,\lambda_i^*) \coloneqq f_i(x_1,\,x_i,\,\lambda_i) \quad \text{with} \quad \lambda_i^* \coloneqq \lambda_i^*(x_1,\,\lambda_i),
$$

for $i = 2, 3$, where λ_i^* maps all of the different pairs of λ_i and x_1, $r_i^{m_1\times m_i}\times m_1$, into different $r_i^{m_i}$ single argument assignment functions, f_i^*. Then,

$$
\begin{aligned}
&p(a_1,\,a_2,\,a_3\;|\;x_1,\,x_2,\,x_3)\\
&\qquad= \sum_{\lambda_1,\lambda_2,\lambda_3} d_1(a_1\,|\,x_1,\,\lambda_1)\,d_2(a_2\,|\,x_1,\,x_2,\,\lambda_2)\,d_3(a_3\,|\,x_1,\,x_3,\,\lambda_3)\,p(\lambda_1,\,\lambda_2,\,\lambda_3)\\
&\qquad= \sum_{\substack{\lambda_1,\lambda_2^*,\lambda_3^*\\ \lambda_i:\lambda_i^*,\,i=2,3}} d_1(a_1\,|\,x_1,\,\lambda_1)\,d_2(a_2\,|\,x_1,\,x_2,\,\lambda_2^*)\,d_3(a_3\,|\,x_1,\,x_3,\,\lambda_3^*)\,p(\lambda_1,\,\lambda_2,\,\lambda_3)\\
&\qquad= \sum_{\lambda_1,\lambda_2^*,\lambda_3^*} d_1(a_1\,|\,x_1,\,\lambda_1)\,d_2(a_2\,|\,x_1,\,x_2,\,\lambda_2^*)\,d_3(a_3\,|\,x_1,\,x_3,\,\lambda_3^*) \sum_{\substack{\lambda_i:\lambda_i^*\\ i=2,3}} p(\lambda_1,\,\lambda_2,\,\lambda_3)\\
&\qquad\coloneqq \sum_{\lambda_1,\lambda_2^*,\lambda_3^*} d_1(a_1\,|\,x_1,\,\lambda_1)\,d_2(a_2\,|\,x_1,\,x_2,\,\lambda_2^*)\,d_3(a_3\,|\,x_1,\,x_3,\,\lambda_3^*)\,p^*(\lambda_1,\,\lambda_2^*,\,\lambda_3^*|\,x_1),
\end{aligned}
$$

where p^* denotes the normalized conditional probability w.r.t. $p(\lambda_1,\,\lambda_2,\,\lambda_3)$. The above condition represents the decomposition in deterministic assignments of a tripartite relaxed scenario in which $\mathrm{pa}(\lambda) = \{x_i\}$. The scenario corresponding to the inverse measurement

independence causal relaxation, $\mathrm{pa}(x_i) = \{\lambda\}$, can be obtained by using Bayes' rule,

$$
\begin{aligned}
& p\left(a_1, a_2, a_3 \,|\, x_1, x_2, x_3\right) \\
&= \sum_{\lambda_1, \lambda_2^*, \lambda_3^*} d_1(a_1 \,|\, x_1, \lambda_1)\, d_2(a_2 \,|\, x_1, x_2, \lambda_2^*)\, d_3(a_3 \,|\, x_1, x_3, \lambda_3^*)\, p^*(\lambda_1, \lambda_2^*, \lambda_3^* \,|\, x_1) \\
&= \sum_{\lambda_1, \lambda_2^*, \lambda_3^*} d_1(a_1 \,|\, x_1, \lambda_1)\, d_2(a_2 \,|\, x_1, x_2, \lambda_2^*)\, d_3(a_3 \,|\, x_1, x_3, \lambda_3^*)\, \frac{p^*(x_1 \,|\, \lambda_1, \lambda_2^*, \lambda_3^*)}{p(x_1)}\, p^*(\lambda_1, \lambda_2^*, \lambda_3^*). \blacksquare
\end{aligned}
$$

If used together, lemmas 3.3 and 3.4 shows an equivalence of any kind of tripartite relaxed scenario with another tripartite relaxed scenario with the latter possessing a causal relaxation of locality sourced in an input measurement and targeted in other part's output result. This conclusion motivates definition 3.5, as follows:

Defintion 3.5

Each tripartite relaxed scenario possessing only locality causal relaxation sourced in an input measurement and targeted in other part's output result defines a causal class of tripartite relaxed scenarios. It is represented by its associated Bayesian network and an input-to-output Bell DAG (IO-BDAG), as in figure 3.2. A causal class contains all of the nonsignalling equivalents scenarios as well as those obtained by relabeling parts, measurements and results. All of the IO-BDAG's and its associated Bayesian networks and scenarios are ranked by the number of locality causal relaxations, forming an hierarchy.

Definition 3.5 enable us to reduce the amount of tripartite relaxed scenarios in a total of 16, divided into six levels of a hierarchy. Furthermore, we can distinct each class of the hierarchy as being boring - in the case that the class reproduce all of the nonsignalling correlations for its scenario, independent of the number of inputs per part and outputs per measurement, as defined earlier - and **interesting**. An interesting class may behave as boring for certain relaxed scenarios, but this is not true when considered for an arbitrary number of inputs and outputs. More importantly, no boring class may present super nonlocality, neither one can assure an interesting class possesses super nonlocality.

The main result of [45], which is summarized in its theorem 4, is to categorize the 16 causal classes among boring or interesting. Among the 16 classes, 6 are definitely boring and 6 others are definitely interesting, that is, there were already known examples of super nonlocality (see figure 3.2). With respect to the four remaining classes, three of

them are nonsignalling equivalent and referred as to **star** class, as the output result a_3 is causally connected to all of the others possible variables - the rays of the "star". The last remaining causal class will be referred as to **circle**. Both star and circle classes were not

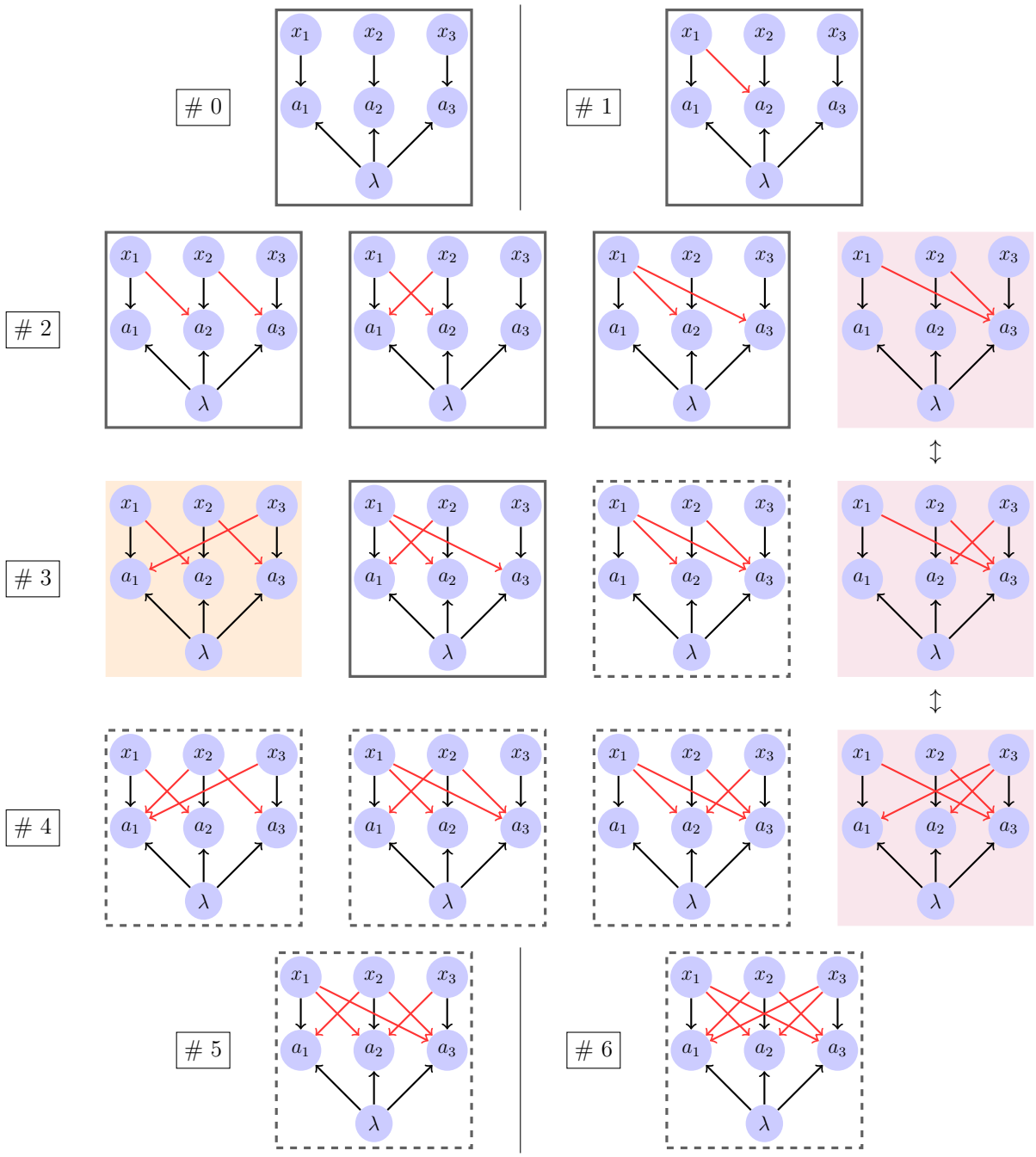


Figure 3.2: The complete causal hierarchy of relaxed tripartite scenarios containing all of the 16 classes of IO-BDAG's. The boxed IO-BDAG's into a continuum square are known to satisfy Svetilchny's inequality [46], which, on the other hand, is known to be violated by quantum correlations [35]; therefore, these classes are indeed interesting. All of the dashed boxes indicates that the IO-BDAG's contained into are boring, as shown in [45]. The box filled in light orange contains the class that will be referred to as the circle. The three remaining boxes, filled in light purple were shown to be nonsignalling equivalent [45] and will be referred to as the star class.

known to be interesting or boring so far.

3.2.2 In the search of super nonlocality in star and circle classes

Before finishing this chapter, let us clarify about the situation of star and circle classes. In [45], it was discovered the star class is boring in a tripartite relaxed (3, 2, 2) scenario. However, it was also found they satisfy a family of inequalities expressed by

$$I_3 = I_2(A, B) + I_2(A', C) + I_2(B', C') \leq b_l + 2b_g, \tag{3.6}$$

where b_l and b_g denote the local and geometrical bounds of I_2, an arbitrary bipartite Bell expression such that $I_2 \leq b_l$ constitutes a Bell inequality. The above inequality express a scenario in which each part's output may be indexed in two different labels, for instance, A and A' are two different labels of Alice's results. Remarkably, not only the star class has been found to be interesting, but examples of quantum violations of (3.6) were found, but with a high number of inputs per part and outputs per measurement.

With respect to circle class, a Bell inequality was found in [45], but no quantum violations of this inequality were verified.

Finally, we are in a condition to expose the main goal of this thesis: firstly, verify if the circle class is indeed interesting or not and, in the affirmative case, find an example of super nonlocality in it. Secondly, we want to find other examples of super nonlocality in the star class, but these in simpler relaxed scenarios, i.e., with a smaller number of inputs and outputs. Although this is the subject of next chapter, we were able to testify that the circle class is interesting and find examples of its super nonlocality.

4 Methods and results

In this last chapter, we will expose our methods and the obtained results. We have made two main attempts in order to find examples of super nonlocality in the circle class.

The first obvious method we have used was the **facet enumeration**. Although it is capable of providing some ordinary results, such as the full characterization of the CHSH and Śliwa scenarios, it has proved to be impracticable for the purpose of this thesis. Secondly, we have started to develop a method based on reference [47] and a result of convex optimization known as **separating hyperplane theorem** [48, sec. 2.5], method which culminated in our main results.

This chapter is organized as follows: firstly, we present both methods as well as a short theoretical introduction and then finalize with the results.

4.1 Facet enumeration

As already mentioned in section 2.3, the Minkowsky-Weyl theorem states that every polytope is finitely generated and can be equivalently represented either by its extremal points (or vertices) or by its facets. Now, if we stick to the definition of what is facet enumeration, any method which is able to obtain the whole set of facets of a specific polytope from more fundamental information about it can be understood as facet enumeration.

However, in the study on Bell nonlocality, this *more fundamental information* of the polytope use to be its vertices, which are, in general, easily obtainable. This way, the method must be understood here as the usage of Minkowsky-Weyl: the conversion of a V-representation to a H-representation of the same polytope. Worth to mention that the inverse problem is also relevant for some cases and, although we will not address it here,

it is known as vertex enumeration.

The first step in order to perform facet enumeration of a given polytope is to identify their vertices. We know, beforehand, that the vertices of the local polytope are expressed by their deterministic behaviors, as in equation (2.11), for a generic bipartite scenario. Thus, all of the work that must be done is to list all the r^m deterministic assignments for each of the parts and combining them as in (2.11), where m is the part's number of measurements and r is the number of output results for each possible measurement. Care must be taken when listing the deterministic assignments of a relaxed scenario such as the ones depicted in figure 3.2, where a part's output receives a signal from other part's input. In this case, there are $r_i^{m_i \times m_j}$ deterministic assignments for each part, where i and j indicate the involved parts. Also, attention must be paid when assembling the deterministic assignments; equation (2.11) is no longer valid, but the condition obtained by the associated Bayesian network.

Once this process is done, r^{2m} extremal points are obtained for a bipartite uniform scenario. The task is continued by inputing all of the extremal points in a polytope representation conversion software. There are several of these, for instance, the polymake [49], the lrs [50], the Parallel AdjaceNcy Decomposition Algorithm (PANDA) [51] and the POlyhedron Representation Transformation Algorithm (PORTA) [52], being the latter two the mainly explored in this thesis.

Both PORTA and PANDA are optimized implementations of the Fourier-Motzkin elimination method, a method for solving a system of linear inequalities which consists on eliminating a single variable for each iteration of the algorithm. A detailed explanation about the Fourier-Motzkin elimination may be found on [53]. Also, the PANDA allow the user to input information about the symmetries of the polytope, which enables the software to find not only the whole wanted facets but also a representative for each class of them (see appendix A).

4.1.1 Some preliminary results with facet enumeration

Using the facet enumeration method we were able to quickly reproduce some known results, such the entire set of 24 inequalities in the CHSH scenario (with trivial inequalities included), as well as the 53 856 inequalities of the Śliwa scenario belonging to 46 distinct classes [54].

Our next natural step was, of course, to address the circle class, represented by the orange filled DAG in figure 3.2. In this step of the thesis we considered, for simplicity, the circle class in the (3, 2, 2) scenario. Some glaring differences were already manifested at the beginning of our study such as the number of extremal points, for instance. While CHSH and Śliwa scenario posses 16 and 64 extremal points, respectively, the circle class has a total of 4096 extremal points. Its study quickly proved to be difficult when we tried to enumerate its facets. For most of our executions of PANDA and PORTA, a 16 cores processor equipped with 32 GB of RAM was used [55]. For CHSH and Śliwa scenarios, we were able to enumerate all facets in a single round of few seconds. The same execution for the circle class took 4 days and returned no practical result.

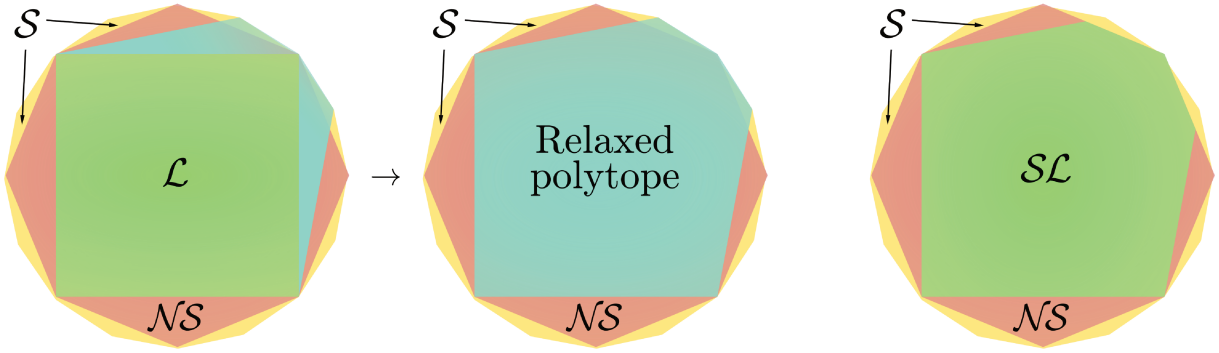
This impediment motivated us to look for alternatives for the circle class. Remember that in any relaxed scenario of the causal hierarchy, like the ones depicted in figure 3.2, we are considering that a part's choice of measurement is able to signalize to other part's result, which implies that some of the behaviors contained in the relaxed politope are actually outside of the nonsignalling polytope, as in figure 4.1a. Since $\mathcal{Q} \subseteq \mathcal{NS}$, these signalling behaviors are not interesting for our investigation. Indeed, all of the interesting information about the circle class is contained in an intersection of the relaxed and nonsignalling polytopes, which, in turn, was termed as the signalling-local polytope, or siglocal, for short, as suggested by figure 4.1b.

Using this insight, we performed some attempts to still enumerate the facets of the circle class. Firstly, consider this new representation (see appendix A) for the behaviors in a relaxed scenario:

$$\begin{gathered}
\langle A\rangle_{x,z} = \sum_a a\, p_{a|x,z}, \quad \langle B\rangle_{x,y} = \sum_b b\, p_{b|x,y}, \quad \langle C\rangle_{y,z} = \sum_c c\, p_{c|y,z}, \\
\langle AB\rangle_{x,y,z} = \sum_{a,b} ab\, p_{a,b|x,y,z}, \quad \langle AC\rangle_{x,y,z} = \sum_{a,c} ac\, p_{a,c|x,y,z}, \\
\langle BC\rangle_{x,y,z} = \sum_{b,c} bc\, p_{b,c|x,y,z}, \quad \text{and} \quad \langle ABC\rangle_{x,y,z} = \sum_{a,b,c} abc\, p_{a,b,c|x,y,z},
\end{gathered} \tag{4.1}$$

where a, b, $c \in \{-1, 1\}$, x, y, $z \in \{0, 1\}$. It can be inverted for the probabilities if

$$\begin{aligned}
p_{a,b,c|x,y,z} = \frac{1}{8}[1 &+ (-1)^a \langle A\rangle_{x,z} + (-1)^b \langle B\rangle_{x,y} + (-1)^c \langle C\rangle_{y,z} \\
&+ (-1)^{a+b} \langle AB\rangle_{x,y,z} + (-1)^{a+c} \langle BC\rangle_{x,y,z} + (-1)^{b+c} \langle BC\rangle_{x,y,z}
\end{aligned}$$



(a) (b)

Figure 4.1: (a) The construction of the relaxed polytope (a bidimensional representation). In a scenario whose nonsignalling conditions are relaxed, we must consider some of the signalling extremal points - the extremal distributions of the signalling polytope, represented by the $\mathcal{S}$ hexadecagon in orange - as the extremal of the relaxed polytope. This way the relaxed polytope can be seen as the convex hull of the local and the additional signalling extremal points. (b) The bidimensional representation of the signalling-local polytope, or siglocal, for short. This new polytope is the intersection of the relaxed and the nonsignalling polytopes. The advantage of using this polytope instead of the relaxed one is that its dimension is smaller than the former. However, its obtainment can be even more difficult.

$$+ (-1)^{a+b+c} \langle ABC \rangle_{x,y,z}]. \quad (4.2)$$

Note that, for equation (4.1) running upon all of the subindices, there are 44 correlators, while in the former representation of probabilities we had 64 possible variables. The correlation vector is now described by 44 grouped signalling correlators[1]. This reduction of 20 dimensions in the representation is a consequence that in the circle class some - but not all - nonsignalling conditions are relaxed. That is, other signalling are still possible, but not present in circle class.

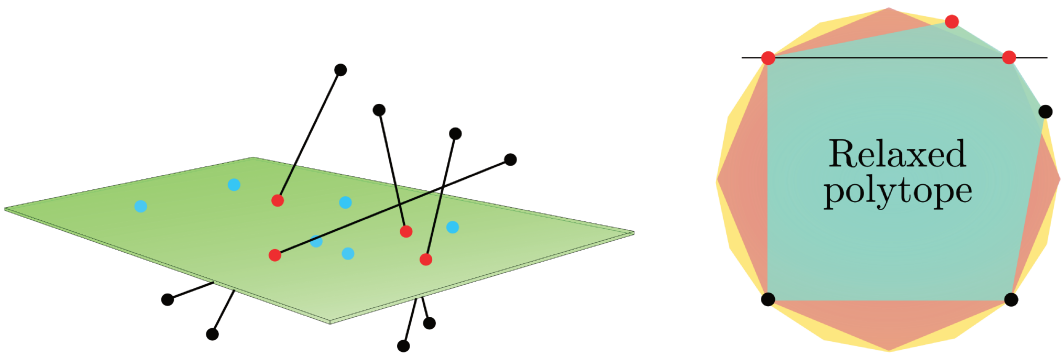
This way, the polytope with causal relaxations of circle class is 44-dimensional. Once the extremal points of the polytope with causal relaxations were obtained in this new representation, we continued with facet enumeration. However, for a second time, the execution of the facet enumeration software proved once again to be inconveniently long and for one more time we could not proceed with facet enumeration even with the reduction of 20 dimensions of the problem. It turns out that for the Śliwa scenario we were able to solve the enumeration of facets even using the 64-dimensional representation of probabilities. However, Śliwa's local polytope has fewer extremal points and has dimension 26. That is,

[1]Remember that the Śliwa polytope is a 26-dimensional object. This is easily checked if the dimension of the correlator vector is counted, which corresponds to 64. Then, subtracting the 8 normalizations and the 30 nonsignalling conditions (equations (2.23) and (2.24)) - considering that each constraint reduces a single dimension - we got 26. This same counting for the relaxed polytope reaches 44.

the main determinant for the execution of the enumeration of facets is the effective size of the polytope, not the representation used. Thus, the following attempts were made to reduce the practical number of points or the effective dimension of the polytope.

It is a fact that we are not looking for the facets of the relaxed polytope, but rather, its intersection with the nonsignalling polytope, the siglocal polytope - as both are virtually the same problem. Thus, we built an algorithm which aimed to perform the intersection given only the vertices of the relaxed one. The central idea of this algorithm is based on the fact that the relaxed polytope has dimension 44 and the nonsignalling polytope, 26. Therefore, the nonsignalling polytope is contained in 18 hyperplanes more than the relaxed polytope, if we take into account the representation of signalling correlators. We started by dividing the extremal points of the relaxed polytope in three groups. Picking one of the aforementioned 18 hyperplanes in which the nonsignalling polytope is contained, we grouped the extremal points of the relaxed polytope among those above this particular hyperplane, those below and those accidentally contained in the hyperplane.

From the groups of points above and below one of the hyperplanes, we take the convex combination among them, two by two. The point of intersection of each line segment with the hyperplane is maintained and the points above and below the hyperplane are discarded, as explained in figure 4.2a. The repetition of this procedure for each of the



(a) (b)

Figure 4.2: (a) Tridimensional illustration of the intersection algorithm. The groups of points above and below the plane are in black joined by line segments. The light blue points are those already contained in the plane - they do not participate in this iteration of the algorithm. The red points represent the intersection of the segment with the plane. While the black dots are discarded, the red ones take their place in the next iteration of the algorithm. (b) Bidimensional illustration of the second algorithm. The same color pattern as the figure 4.1a is used here. The line secant to all polytopes represents one of Śliwa's inequalities. The points in red are the extremal points of the smaller polytope separated by the inequality. The black points are the other extremal of the relaxed polytope. All points below the line are local w.r.t this specific inequality.

hyperplanes gives us the extremal points of the desired intersection polytope.

This approach, despite returning extremal points at the intersection of the two desired polytopes, also returns points on the facets of the intersecting polytope, which are redundant. Then, for each iteration, the number of points increased dramatically and, even taking the convex hull among all the resulting points in each iteration, the execution time also proved to be excessively long and this procedure was also abandoned.

The last addressed attempt was to section the relaxed polytope into pieces. The procedure consisted of dividing the polytope using the known Śliwa's inequalities. Since the interest is to find out only super non-local behaviors, then it is trivial that these behaviors are above the Bell's inequalities for the corresponding scenario. Thus, the "nucleus" of the polytope is uninteresting. This way, the problem was broken into pieces. Each Bell's inequality for the Śliwa scenario cuts the relaxed polytope in two haves. The extremal points above or contained in the hyperplane designated by a single Bell inequality forms a smaller polytope than the original, as in figure 4.2b. Nevertheless, even for this "small" polytope, the process of facet enumeration proved to be unproductive. The number of extremal points for this new polytope has not decreased considerably, which has not changed the conditions for the execution of the facet enumeration software.

4.2 Linear programming

After several failed trials to find out any super non-local behavior (as we had not even got close to the inequalities), we considered a more radical changing in our original plans. Instead of insisting in the facet enumeration process, we picked the separating hyperplane theorem as our current method.

Before going deeper into this theorem, let us, firstly, introduce the base needed to address the separating hyperplane method, which corresponds to linear programming. More generally, linear programming is a class of problems we denote as an optimization problem. Every optimization problem can be immediately recognized as

$$\begin{aligned} \text{minimize} \quad & o(\mathbf{x}) \\ \text{subject to} \quad & c_i(\mathbf{x}) \leq b_i, \quad i = 1, ..., N, \end{aligned} \tag{4.3}$$

where $o : \mathbb{R}^n \mapsto \mathbb{R}$ is denoted as the objective and the N inequalities $c_i : \mathbb{R}^n \mapsto \mathbb{R}$ bounded

respectively by $b_i \in \mathbb{R}$ are denoted as the constraints of the problem. For every $\mathbf{x}$ which satisfies all of the N constraints, we denote it as a feasible point and, if the problem is not infeasible, i.e., the solution $\mathbf{x}^*$ of (4.3) can be obtained, then $\mathbf{x}^*$ is the optimal point of $o(\mathbf{x})$. It is worth noting that the restrictions should not necessarily be an inequality. An equality constraint also constitutes a valid restriction.

In this way, for different forms of the functions $o(\mathbf{x})$ and $c_i(\mathbf{x})$, we establish a different class of problems. Naturally, if $o(\mathbf{x})$ and $c_i(\mathbf{x})$ are linear w.r.t. $\mathbf{x}$, then the task defined in (4.3) is regarded as linear program.

In a more pleasant manner, every linear program can be written as

$$\begin{aligned} \text{minimize} \quad & \mathbf{o} \cdot \mathbf{x} \\ \text{subject to} \quad & C\,\mathbf{x} \leq \mathbf{b}, \end{aligned} \tag{4.4}$$

where $\mathbf{b} \in \mathbb{R}^N$ is the vector containing the b_i bounds of (4.3) and $\mathbf{o} \in \mathbb{R}^n$ and $C \in \mathbb{R}^{N \times n}$ now represent the objective and the constraints, respectively. Also, in a linear program, if there is $\mathbf{x}^*$, then the set $\mathcal{F}$ of feasible points is represented by a convex polytope, being $\mathbf{x}^*$ an extremal point of $\mathcal{F}$.

Seeing this way, optimization problems and, specifically, linear programming seem to enjoy a common solving method, which is not true. Actually, for a specific optimization problem, there can be several applicable methods. For linear programming, the most used algorithm is known as Dantzig's simplex algorithm [56] whose operation is based on the fact that if an extremal point does not reach the maximum (or minimum) of the objective, then its value can always be increased (or decreased) by walking along an edge containing the initial extremal point and ending on a second one [57, sec. 3.8]. The execution is finished when this process reaches a maximum (or minimum) value of $\mathbf{o} \cdot \mathbf{x}$ or when the current edge is unbounded, in the case that the problem has no solution.

There are several applications of linear programs that can be used in the study of nonlocality. The following linear program, for instance, can be used to determine whether a given point $\mathbf{p}$ is or not within a polytope whose vertices are expressed by $\mathbf{v}_i$. Then,

$$\begin{aligned} \text{maximize} \quad & \sum_i a_i \\ \text{subject to} \quad & \sum_i a_i \leq 1, \;\; \sum_i a_i \mathbf{v}_i \leq \mathbf{p}, \;\; a_i \geq 0 \quad \forall i, \end{aligned} \tag{4.5}$$

If the answer of the objective function is equal to 1, then $\mathbf{p}$ is inside the given polytope.

Every linear program such (4.4) may be also referred as the **primal** form of the problem, that is, we can also establish a second linear program, which is known as the **dual** and may be written as

$$\begin{aligned} &\text{maximize} && \mathbf{b} \cdot \boldsymbol{\nu} \\ &\text{subject to} && C^{\mathsf{T}} \boldsymbol{\nu} \geq \mathbf{o}. \end{aligned} \tag{4.6}$$

where $\boldsymbol{\nu} \in \mathbb{R}^N$. We say that a linear program is strongly dual if both of its solutions, $\mathbf{x}^*$ and $\boldsymbol{\nu}^*$ - the primal and the dual, respectively -, are such that both of the objective are equal, i.e., $\mathbf{o} \cdot \mathbf{x}^* = \mathbf{b} \cdot \boldsymbol{\nu}^*$. Additionally, if the program is strongly dual, then if the primal is infeasible, the dual is unbounded and vice versa.

Another good application of linear programs in nonlocality is that we can compute the distance of a point $\mathbf{p}$ to a given polytope with N vertices $\mathbf{v}_i$, or, more interestingly, obtain, by its dual form, the separating hyperplane between them by writing

$$\begin{aligned} &\underset{a}{\text{minimize}} && \mathbf{a}^{\mathsf{T}} \cdot \mathbf{p} \\ &\text{subject to} && \mathbf{a}^{\mathsf{T}} \cdot \mathbf{v}_i \geq 1, \;\; \mathbf{a} \geq 0, \;\; i = 1, ..., N, \end{aligned} \tag{4.7}$$

where $\mathbf{a} \geq 0$ must be understood as element wise and $\mathbf{a} \in \mathbb{R}^n$ represent the coefficients obtained by optimization which characterizes the hyperplane $\mathbf{a}^{\mathsf{T}} \cdot \mathbf{x} \geq 1$, for $\mathbf{x} \in \mathbb{R}^n$. Note that, here, we are interested on finding the hyperplane $\mathbf{a}^{*\mathsf{T}} \cdot \mathbf{x} \geq 1$ which maximally separates the point $\mathbf{p}$ from the vertices $\mathbf{v}_i$.

The above result is supported by a theorem known in the literature as the **separating hyperplane theorem**. It can be generically enunciated by the existence of $\mathbf{a}^*$ such that the hyperplane $\mathbf{a}^{*\mathsf{T}} \cdot \mathbf{x} = 1$ separates two given convex sets $\mathcal{A}$ and $\mathcal{B}$, with $\mathcal{A} \cap \mathcal{B} = \varnothing$. As this is a result of convex optimization, which, in its turn, is not contained in our main topic of investigation, it will suffice to enunciate it. The interested reader can consult a judicious proof of it on [48, ex. 2.22] as well as a richer introduction to optimization problems.

4.2.1 A super non-local behavior

In light of the concepts just shown, we can finally present our main results. The original inspiration was a work by Cope and Colbeck [47] which is, in its turn, based on the following insight: every local behavior is contained in a convex set (the local polytope) which is contained in another convex set, the nonsignalling polytope. The extremal behaviors of the $\mathcal{NS}$ set are either outside the local polytope or are also extremal of the local polytope [23].

Once we can check whether an extremal point is or not within the local polytope, by using (4.5), then one is able to filter only the disjoint extremal behaviors of the $\mathcal{NS}$ set and $\mathcal{L}$. Therefore, the idea is use (4.7) systematically to obtain the separating hyperplane between the extremal nonsignalling behaviors which are disjoint to the local set $\mathcal{L}$. The minimization over $\mathbf{a}^{\mathsf{T}} \cdot \mathbf{p}$ force this hyperplane to be a face of $\mathcal{L}$ and a later verification distinguishes the facets from the faces.

This systematical usage of (4.7) is performed in [47] considering that not only the extremal behaviors are able to be used. Note that, since you have got a single separating hyperplane by performing a single iteration of (4.7), then you can select extremal local behaviors which lie on the obtained hyperplane. Any convex combination of these still yield an outsider point, w.r.t. the local polytope. This specific method is able to find hundreds of thousands of Bell inequalities in an acceptable running time of the algorithm.

Despite of the potential to find out several new Bell inequalities, we actually solved our problem with a single iteration of the Cope and Colbeck method. The first step of implementation consists on obtaining the extremal behaviors of the nonsignalling polytope for the desired scenario: circle class in the (3, 2, 2) configuration. Alternatively, it is possible to generate some random nonsignalling points outside of $\mathcal{L}$, but as the extremal points of $\mathcal{NS}$ are an already known result [58], we could start straightly from (4.5).

The (3, 2, 2) $\mathcal{NS}$ polytope possesses 53 856 extremal points, belonging to 46 inequivalent classes[2]. Running (4.5) for a representative of each of the classes, we concluded that the 27th, 29th, 43rd and 44th of the classes defined in [58] are not within the (3, 2, 2) relaxed polytope for the circle configuration. If the Cope and Colbeck method is employed, then these special points are the starting iteration of the method. It happened that the

[2]Coincidence or not, this is also the number of Bell inequalities for this Bell scenario: 53 856 inequalities, belonging to 46 inequivalent classes

behavior associated to the 29th class provided the following separating hyperplane:

$$\begin{aligned}
&\langle A\rangle_{00} + \langle A\rangle_{11} + \langle B\rangle_{00} + \langle B\rangle_{11} + \langle C\rangle_{00} + \langle C\rangle_{11} + \langle AB\rangle_{000} + 2\langle AB\rangle_{001} \\
&-2\langle AB\rangle_{101} - \langle AB\rangle_{111} + \langle AC\rangle_{000} + 2\langle AC\rangle_{010} - 2\langle AC\rangle_{011} - \langle AC\rangle_{111} \\
&+\langle BC\rangle_{000} + 2\langle BC\rangle_{100} - 2\langle BC\rangle_{110} - \langle BC\rangle_{111} - \langle ABC\rangle_{000} - \langle ABC\rangle_{111} \leq 12, \qquad (4.8)
\end{aligned}$$

which is a facet of the relaxed polytope.

After finding (4.8), no further iterations were needed to show that it can be violated. We can do so considering the projection of (4.8) onto the nonsignalling polytope, then

$$\begin{aligned}
&\langle A_0\rangle + \langle A_1\rangle + \langle B_0\rangle + \langle B_1\rangle + \langle C_0\rangle + \langle C_1\rangle + 3\langle A_0B_0\rangle \\
&-2\langle A_1B_0\rangle - \langle A_1B_1\rangle + 3\langle A_0C_0\rangle - 2\langle A_0C_1\rangle - \langle A_1C_1\rangle \\
&+3\langle B_0C_0\rangle - 2\langle B_1C_0\rangle - \langle B_1C_1\rangle - \langle A_0B_0C_0\rangle - \langle A_1B_1C_1\rangle \leq 12 \qquad (4.9)
\end{aligned}$$

and equation (4.8) can be written in nonsignalling correlators representation, such as equation (2.16). Now, let

$$\begin{aligned}
A_{0|0} = B_{0|0} = C_{0|0} = |0\rangle\langle 0|\,, \\
A_{1|0} = B_{1|0} = C_{1|0} = |1\rangle\langle 1|
\end{aligned}$$

be the projective measurement operators related to measurement x, y, $z = 0$ and

$$\begin{aligned}
A_{0|1} = B_{0|1} = C_{0|1} = |\varphi\rangle\langle\varphi|\,, \\
A_{1|1} = B_{1|1} = C_{1|1} = \mathbb{1} - |\varphi\rangle\langle\varphi|
\end{aligned}$$

be the projective measurement operators related to measurement x, y, $z = 1$, where

$$|\varphi\rangle = \cos\left(\frac{\pi}{3}\right)|0\rangle + \sin\left(\frac{\pi}{3}\right)|1\rangle\,.$$

If $\rho = |\psi\rangle\langle\psi|$, with

$$|\psi\rangle = \cos\alpha\left[\cos\beta\,|000\rangle + \sin\beta\,|111\rangle\right] + \sin\alpha\left[\cos\gamma\,|W_0\rangle + \sin\gamma\,|W_1\rangle\right],$$

where $\alpha \approx {}^{1}\!/_{100}$, $\beta \approx {}^{5\pi}\!/_{6}$, $\gamma \approx {}^{5\pi}\!/_{8}$ and

$$|W_0\rangle = {}^{1}\!/_{\sqrt{3}}\left(|100\rangle + |010\rangle + |001\rangle\right),$$

$$|W_1\rangle = {}^1/\sqrt{3}\,(|011\rangle + |101\rangle + |110\rangle)\,,$$

then, using the NPA hierarchy (see appendix B), inequality (4.9) can be violated up to 13.897.

This result confirms that the circle class, for the (3, 2, 2) Bell scenario, is indeed interesting and finishes our presentation of results so far.

4.3 Final considerations

This thesis aimed to answer some questions opened in [45]. The main of them is if the circle class, as defined in figure 3.2, level #3, is interesting or boring. By using a method slightly based on the Cope and Colbeck method [47], we derived the following inequality

$$\begin{aligned}
&\langle A_0\rangle + \langle A_1\rangle + \langle B_0\rangle + \langle B_1\rangle + \langle C_0\rangle + \langle C_1\rangle + 3\,\langle A_0B_0\rangle \\
&-2\,\langle A_1B_0\rangle - \langle A_1B_1\rangle + 3\,\langle A_0C_0\rangle - 2\,\langle A_0C_1\rangle - \langle A_1C_1\rangle \\
&+3\,\langle B_0C_0\rangle - 2\,\langle B_1C_0\rangle - \langle B_1C_1\rangle - \langle A_0B_0C_0\rangle - \langle A_1B_1C_1\rangle \leq 12,
\end{aligned}$$

which can be violated up to 13.897 as verified by the NPA hierarchy.

The above inequality represents a facet of the relaxed polytope of the (3, 2, 2) Bell scenario in the just mentioned circle configuration. Its violation represents that there can be quantum correlations which cannot be simulated even when the parts signal their measurements input to one of its neighbor, in a cyclical manner. Because the set of correlations simulated in this classical configuration is not the local set anymore, but a larger one, we classified this strong quantum correlation as super non-local.

Furthermore, a result which is not explicit on the main text is that the facet enumeration is not anymore a recommended method for obtaining practical results on Bell scenarios, i.e., facet enumeration effectively can reproduce lots of known results, but when applied in new scenarios it hardly produce something new.

A

Transformations in a Bell inequality

Every reader slightly acquainted with the concept of Bell inequalities should be able to recognize the following inequality as the standard form of the CHSH inequality [22]:

$$|\langle A_0B_0\rangle + \langle A_0B_1\rangle + \langle A_1B_0\rangle - \langle A_1B_1\rangle| \leq 2, \tag{A.1}$$

where the mathematical entity defined by $\langle A_xB_y\rangle = p(a{=}b\,|\,x,\,y){-}p(a{\neq}b\,|\,x,\,y)$ is known as the **correlator** between Alice's and Bob's choice of measurement and $a,\,b,\,x,\,y \in \{0,\,1\}$. Hence, a trivial expansion of A.1 in terms its probabilities yields

$$\begin{aligned}
&|p(0,0|0,0) + p(1,1|0,0) - p(0,1|0,0) - p(1,0|0,0)\\
&+p(0,0|0,1) + p(1,1|0,1) - p(0,1|0,1) - p(1,0|0,1)\\
&+p(0,0|1,0) + p(1,1|1,0) - p(0,1|1,0) - p(1,0|1,0)\\
&-p(0,0|1,1) - p(1,1|1,1) + p(0,1|1,1) - p(1,0|1,1)| \leq 2.
\end{aligned} \tag{A.2}$$

which is no more than (A.1) but written in an extensive manner and representing the same Bell inequality as in [22].

The first appearance of (A.1) in the literature was in the aforementioned work by Clauser, Horne, Shimony and Holt in 1969 in the - not that familiar - form of

$$|\langle A_0B_0\rangle - \langle A_0B_1\rangle| \leq 2 - \langle A_1B_0\rangle - \langle A_1B_1\rangle\,, \tag{A.3}$$

where the correlator is now written as $\langle A_xB_y\rangle = \sum_{a,b}(-1)^{a+b}p(a,\,b\,|\,x,\,y)$, with $a,\,b \in \{1,\,2\}$. A few years later, in 1974 two of the same authors of the CHSH inequality stated

what would come to be known as CH inequality [59] for the (2, 2, 2) Bell scenario. It can be written as

$$p(1,1|0,0) - p(1,1|0,1) + p(1,1|1,0) + p(1,1|1,1) - p_A(1|1) - p_B(1|0) \leq 0, \qquad \text{(A.4)}$$

using the same labels as (A.3), where the subindex A and B denotes the marginal probability of Alice and Bob, respectively. Both (A.1), (A.2), (A.3) and (A.4) are different representations of the same Bell inequality of the (2, 2, 2) scenario, which was already mentioned as being unique.

It turns out that every single Bell inequality can be differently written as to consider the convenience of the problem being adressed. This way, in a scenario such as (2, 2, 2), where there is a single representative Bell inequality, one can write several different versions of this same expression.

Briefly, there are three different ways we can transform a given inequality. Firstly, we are able to use any of the normalization and nonsignalling constraints. For a bipartite Bell scenario, for instance, there are m^N (where N is the number of parts and m the number of measurements per part) normalization conditions of the form (2.6) that can be used for this purpose. In (A.4) we can retrieve (A.2) by using the fact that every marginal related to a measurement set defines a nonsignalling restriction over the probabilities and then p_A can be written as $p_A(a\,|\,x) = \sum_b p(a,\, b\,|\,x,\, y),\, \forall\, y$.

Secondly, we are also free to change the labels of the Bell scenario. There are also three kinds of relabelings that can be done, which consists on changing the names of the parties, the label of the measurements and the label of results. An example can be pictured by using, again, the CHSH inequality, as in (A.1). Actually this inequality possesses quite a lot of simmetry and some of the changings will not be effective. For instance, turning $\langle A_x B_y \rangle$ into $\langle B_x A_y \rangle$ has the same effect than changing Bob's and Alice's measurements and the Bell observable will remain unchanged

$$\mathcal{G}_{\text{Bell}} = A_0B_0 + A_0B_1 + A_1B_0 + A_1B_1 \quad \rightarrow B_0A_0 + B_0A_1 + B_1A_0 + B_1A_1 = \mathcal{G}_{\text{Bell}}, \quad \text{(A.5)}$$

since $\langle B_x A_y \rangle = \langle A_y B_x \rangle$. However, if only Alice's (or Bob's) measurement is relabeled,

then we achieve a new inequality:

$$\langle A_0B_0\rangle + \langle A_0B_1\rangle + \langle A_1B_0\rangle - \langle A_1B_1\rangle \leq 2 \;\rightarrow\; \langle A_1B_0\rangle + \langle A_1B_1\rangle + \langle A_0B_0\rangle - \langle A_0B_1\rangle \leq 2. \tag{A.6}$$

Indeed, the above inequality is not (physically) different of (A.1) as we expect that an arbitrary change in the labels of Alice's measurement will not generate a new scenario. Actually, there are others changings which can be done in (A.1) that provide new inequalities. Exclusively for (A.1), we have eight different inequalities that can be obtained by a similar changing of labels. As they are the same inequality up to this changing, we denote that these eight belong to the same **class**[1] of inequalities.

CHSH scenario possesses a single non-trivial class, which is represented by its homonym inequality, and a trivial class which can be written as $p(a,\, b\,|\,x,\, y) \geq 0$, $\forall\, a$, b, x and y. Similarly, Śliwa scenario possesses 46 classes, which one of them is trivial.

Finally, it is also possible to change the representation of a inequality. We have already done it by transforming (A.1) into (A.2). While equation (A.1) is said to be represented in the correlators representation, (A.2) representation is named as a probabilistic.

The importance of changing the representation of a inequality does not lie only in the fact that we are able to write the inequality in a shorthand manner. Also, we achieve a reduction of the redundancies in the multidimensional space in which the polytope is contained. For instance, in the (2, 2, 2) Bell scenario the local polytope is contained in a 16-dimensional space in the probabilistic representation as each behavior is entirely described by a 16-component real vector $\mathbf{p}_{\text{prob}} = [\,p(0,0|0,0)\; p(0,0|0,1)\; ...\,]^{\mathsf{T}}$. When in the correlators representation we can say that the behavior is now represented by a 8-component real vector

$$\mathbf{p}_{\text{corr}} = [\,\langle A_0B_0\rangle\;\; ...\;\; \langle A_1B_1\rangle\;\; \langle A_0\rangle\;\; \langle A_1\rangle\;\; \langle B_0\rangle\;\; \langle B_1\rangle]^{\mathsf{T}}\,, \tag{A.7}$$

where $\langle A_xB_y\rangle = \sum_{a,b} ab\, p(a,\, b\,|\,x,\, y)$, $\langle A_x\rangle = \sum_a a\, p(a\,|\,x)$ and $\langle B_y\rangle = \sum_b b\, p(b\,|\,y)$ if we label $a,\, b \in \{-1,\, 1\}$.

The dimension of the subspace defined by the correlators representation is precisely the subspace spanned by the nonsignalling subspace. It considers the nonsignalling conditions in its composition by adopting the marginals of the probabilities in the marginal

[1] For those who are familiar with group theory the preferable name is orbit.

correlators $\langle A_x \rangle$ and $\langle B_y \rangle$. This is less redundant as for each measurement set of each marginal we have a corresponding nonsignalling condition.

This trick is also explored in a third representation known in the literature as the Collins-Gisin representation [60]. In this representation, the behavior is written by omitting the probabilities of the results $a = b = 1$ (could be $a = b = 0$, as well) and including the marginals related to the results $a = b = 0$ of Alice and Bob (again, could be $a = b = 1$):

$$\mathbf{p}_{\text{CG}} = [p_A(0|0)\ p_A(0|1)\ p_B(0|0)\ p_B(0|1); p(0,0|0,0)\ p(0,0|0,1); p(0,0|1,0)\ p(0,0|1,1)]^{\mathsf{T}}. \tag{A.8}$$

It has a clear advantage on the correlators representation as it does not require that the results of each measurements to be dichotomous.

The way we can see a representation changing is as a coordinate changing in the space of the probabilities. As we can transform the coordinates of the space of probabilities, we are able to do that in a form that some of the coordinates become redundant. This way we can focus in the subspace which excludes the redundancy.

B

Numerical ways to verify a quantum violation

Once you got a new Bell inequality is desirable to know if this inequality is or not violated by the quantum mechanics and quantify this violation. Moreover, we are interested to find out, if there is a quantum violation, then what is the maximum amount we can achive for it.

For an arbitrary scenario, given a Bell observable $\mathcal{G}_{\text{Bell}}$, is quite easy to determine its value for a given state ρ and POVMs $\{\mathbf{A}_{a|x}\}$, $\{\mathbf{B}_{b|y}\}$, $\{\mathbf{C}_{c|z}\}$, ... All we got to do is to substitute these quantities in the desired observable and calculate its value. However, it is not possible to garantee, obviously, that the random initial choice of state and POVMs will reach the local bound of the inequality. Fortunately, there are two numerical approaches which accomplish this task. Both of them, the see-saw algorithm and the NPA hierarchy are based in a optimization problem known as semi-definite programming.

Like linear programs, semi-definite programs (SDPs) are a class of optimization problems which share a common property between them: the optimization the objective function, which is also linear, is taken over the intersection of the positive semidefinite cone of matrices with the problem space. Typically, every SDP may be written in the following form:

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \text{tr}\,(CX) \\ \text{subject to} \quad & \text{tr}(A_i X) \leq b_i, \;\; i = 1, ..., N. \\ & X \succeq 0, \end{aligned} \tag{B.1}$$

where X, C anb B_i are square complex matrices with d lines and $b_i \in \mathbb{C}$, for $i = 1, ..., N$.

Here, such as in the linear programming case, we also establish a strong duality of the SDPs, i.e., the optimal solution of (B.1) applied in its objective function must be equal to the optimal value of the dual objective function.

See-saw algorithm works by using SDPs for several times [61]. Defined the inequality you aim to explore and the dimension of the state and POVMs, you randomly pick positive semi-definite matrices to represent the POVMs. The next step is to express your Bell observable $\mathcal{G}_{\text{Bell}}$ in terms of the matrices you have picked the step before. With $\mathcal{G}_{\text{Bell}}$ in hands, is possible to obtain its eigenvectors and define ρ as the state composed by the eigenvector $|\psi\rangle$ possessing the biggest eigenvalue, that is, $\rho = |\psi\rangle\langle\psi|$. Now, we maximize $\text{tr}(\rho G_{\text{Bell}})$ w.r.t the first POVM, let us say $\{\mathbf{A}_{a|x}\}$, as in equation (2.19). This step is iterated for all of the parts and repeated until it converges to a local maximum. Threfore, is recommended to run the algorithm more than once, considering different starting matrices in a way you will be achieving more probably a global maximum value for the desired inequality.

Similarly, to the see-saw algorithm the NPA hierarchy also provides a bound to the violation of a desired inequality, but instead of a maximization of the Bell observable, it provides a superior bound for the violation, being, thus, complementary to the see-saw optimization.

Let $\mathcal{M} = \{M_1, ..., M_n\}$ be a set of linear operators acting in the Hilbert space $\mathcal{H}$ of ρ. Then, let us define the matrix Γ whose components are written as

$$\Gamma_{ij} = \text{tr}\,(\rho M_i M_j)\,. \tag{B.2}$$

If written such (B.2), Γ matrix is a semi-definite matrix, as for all $|\psi\rangle \in \mathcal{H}$,

$$\begin{aligned}\langle\psi\,|\,\Gamma\,|\,\psi\rangle &= \sum_{i,j} \psi_i^* \text{tr}\,(\rho M_i M_j)\,\psi_j \\ &= \text{tr}\left(\rho \sum_i \psi_i^* M_i \sum_j \psi_j M_j\right).\end{aligned} \tag{B.3}$$

As $\left(\sum_j \psi_j M_j\right)^\dagger = \sum_i \psi_i^* M_i$, the whole argument of the above expression is positive semi-definite [7, ex. 2.25].

Now, suppose that the set o linear operators is given by $\mathcal{M}^{(1)} = \{\mathbb{1}, \{\mathbf{A}_{a|x}\}_{a,x}, \{\mathbf{B}_{b|y}\}_{b,y}\}$

for the first level of the hierarchy. If considered so, then some of the components of matrix $\Gamma^{(1)}$ are easily calculated as the components of the behavior:

$$\Gamma^{(1)}_{ij} = \text{tr}\left(\rho \mathbf{A}_{a|x} \otimes \mathbf{B}_{b|y}\right) = p(a,\, b \,|\, x,\, y), \tag{B.4}$$

while for some other components, let us say i' and j', $\Gamma^{(1)}_{i',j'}$ are undefined: $\Gamma^{(1)}_{i',j'} = \text{tr}\left(\rho \mathbf{A}_{a|x} \otimes \mathbf{A}_{a|x}\right)$. The method works by ensuring that if the behavior $\mathbf{p}$ is a quantum behavior, then $\Gamma^{(1)}$ is a positive semi-definite matrix [62, criterion 6]. Although this is a necessary condition for the existence of $\Gamma^{(1)} \succeq 0$, it is not sufficient, then some of the behaviors which turns $\Gamma^{(1)}$ positive semi-definite are not contained in the set of quantum correlations and we denote the set of all $\mathbf{p}$ such that $\Gamma^{(1)} \succeq 0$ as $\mathcal{Q}^{(1)}$.

The following level is repeated but considering $\mathcal{M}^{(2)} = \mathcal{M}^{(1)} \cup \{\{\mathbf{A}_{a|x}\mathbf{A}_{a'|x'}\}_{a,x,a',x'}, \{\mathbf{B}_{b|y}\mathbf{B}_{b'|y'}\}_{b,y,b',y'}, \{\mathbf{A}_{a|x}\mathbf{B}_{b|y}\}_{a,b,x,y}\}$ and a second set $\mathcal{Q}^{(2)}$ is obtained associated to the second level. The method provides that each set $\mathcal{Q}^{(n+1)}$ is contained on its precursor $\mathcal{Q}^{(n)}$ and in the limit that $n \to \infty$,

$$\lim_{n\to\infty} \mathcal{Q}^{(n)} = \mathcal{Q}, \tag{B.5}$$

the hierarchy yields the set $\mathcal{Q}$ of quantum behaviors.

Obviously, there is no need to evaluate several levels of this hierarchy and reach the quantum set. As it is an exterior approximation of $\mathcal{Q}$, then a few levels of the hierarchy must provide a good superior bound to the violation we aim to find out.

Bibliography

[1] V.J. Wright, S. Weigert, Found. Phys. **49**, 594–606 (2019).

[2] R.W. Spekkens, Found. Phys. **44**, 1125–1155 (2014).

[3] D.M. Lee, Rev. Mod. Phys. **69**, 645 (1997).

[4] Physics Today **54**, 12, 14 (2001).

[5] C. Cohen-Tannoudji, B. Diu, F. Laloë, *Quantum Mechanics, vol. 1*, Wiley-Interscience (2006).

[6] M. Born, Z. Physik **37**, 863–867 (1926). A translated version can be found here.

[7] M.A. Nielsen, I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, 2nd ed. (2010).

[8] J.B. Altepeter, D.F.V. James, P.G. Kwiat, *Quantum State Estimation*, chap. 4. Lect. Notes Phys. **649**, 113, Springer, Berlin Heidelberg (2004).

[9] R. Horodecki, P. Horodecki, M. Horodecki, K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).

[10] A. Einstein, B. Podolsky, N. Rosen, Phys. Rev. **47**, 777 (1935).

[11] D. Cavalcanti, P. Skrzypczyk, Rep. Prog. Phys. **80**, 024001 (2016).

[12] D. Rosset, D. Schmid, F. Buscemi, e-print: arXiv:1911.12462 [quant-ph].

[13] J.S. Bell, Physics Physique Fizika **1**, 195 (1964).

[14] D. Bohm, Phys. Rev. **85**, 166 (1952).

[15] D. Bohm, Phys. Rev. **85**, 180 (1952).

[16] M. Rédei, *Reichenbach's Common Cause Principle and Quantum Correlations*. In: T. Placek, J. Butterfield (eds), *Non-locality and Modality* **64**, Springer Netherlands, Dordrecht (2002).

[17] H. Reichenbach, *The Direction of Time*, University of California Press, Berkeley (1956).

[18] A. Fine, Phys. Rev. Lett. **48**, 291 (1982).

[19] I. Pitowski, Math. Prog. **50**, 395-414 (1991).

[20] K. Fukuda, C. Weibel, *Computing All Faces of the Minkowski Sum of $\mathcal{V}$-Polytopes*. In: Proceedings of the 17th Canadian Conference on Computational Geometry, 256-259 (2005).

[21] An excellent consulting material can be found on `www.cs.mcgill.ca/~fukuda/soft/polyfaq/`.

[22] J.F. Clauser, M.A. Horne, A. Shimony, R.A. Holt, Phys. Rev. Lett. **23**, 880 (1969).

[23] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, D. Roberts, Phys. Rev. A **71**, 022101 (2005).

[24] S. Popescu, D. Rohrlich, Found. Phys. **24**, 379-385 (1994).

[25] B. Hensen, H. Bernien, A. Dréau et al. Nature **526**, 682-686 (2015).

[26] M. Giustina et al. Phys. Rev. Lett. **115**, 250401 (2015).

[27] L.K. Shalm et al. Phys. Rev. Lett., **115** 250402 (2015).

[28] S.J. Freedman, J.F. Clauser, Phys. Rev. Lett. **28**, 938 (1972).

[29] A. Aspect, J. Dalibard, G. Roger, Phys. Rev. Lett. **49**, 1804 (1982).

[30] M.A. Rowe et al, Nature **409**, 791-794 (2001).

[31] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, A. Zeilinger, Phys. Rev. Lett. **81**, 5039 (1998).

[32] R.F. Werner, Phys. Rev. A **40**, 4277 (1989).

[33] B.S. Cirel'son, Lett. Math. Phys. **4**, 93-100 (1980).

[34] C. Śliwa, Physics Lett. A **317**, 3-4 (2003). E-print: arXiv:quant-ph/0305190.

[35] G. Svetlichny, Phys. Rev. D **35**, 3066 (1987).

[36] J. Pearl, *Causality: models, reasoning, and inference*, Cambridge University Press, Cambridge, 2nd ed. (2009).

[37] T. Fritz, Commun. Math. Phys. **341**, 391 (2016).

[38] J. Henson, R. Lal, M.F. Pusey, New J. Phys. **16**, 113043 (2014).

[39] C.J. Wood, R.W. Spekkens, New J. Phys. **17**, 033002 (2015).

[40] J. Barrett, N. Gisin, Phys. Rev. Lett. **106**, 100406 (2011).

[41] G. Pütz, D. Rosset, T.J. Barnea, Y.-C. Liang, N. Gisin, Phys. Rev. Lett. **113**, 190402 (2014).

[42] G. Brassard, R. Cleve, A. Tapp, Phys. Rev. Lett. **83**, 1874 (1999).

[43] B.F. Toner, D. Bacon, Phys. Rev. Lett. **91**, 187904 (2003).

[44] J.B. Brask, R. Chaves, J. Phys. A: Math. Theor. **50**, 094001 (2017).

[45] R. Chaves, D. Cavalcanti, L. Aolita, Quantum **1**, 23 (2017).

[46] N.S. Jones, N. Linden, S. Massar, Phys. Rev. A **71**, 042329 (2005).

[47] T. Cope, R. Colbeck, Phys. Rev. A **100**, 022114 (2019).

[48] S. Boyd, L. Vandenberghe, *Convex optimization*, Cambridge University Press, Cambridge (2004).

[49] E. Gawrilow, M. Joswig, *polymake: a Framework for Analyzing Convex Polytopes*. In: G. Kalai, G.M. Ziegler (eds), *Polytopes - Combinatorics and Computation* **29**, Birkhäuser, Basel (2000). Available on `www.polymake.org/doku.php`.

[50] D. Avis, *A Revised Implementation of the Reverse Search Vertex Enumeration Algorithm*. In: Kalai G., Ziegler G.M. (eds), *Polytopes — Combinatorics and Computation* **29**, Birkhäuser, Basel. Available on `http://cgm.cs.mcgill.ca/~avis/C/lrs.html`

[51] S. Lörwald, G. Reinelt, EURO J. Comput. Optim. **3**, 297-308 (2015). Available on `http://comopt.ifi.uni-heidelberg.de/software/PANDA/`

[52] T. Christof, A. Löbel, M. Stoer, *Porta - polyhedron representation transformation algorithm* (version 1.4.1) (1997).

[53] G.B. Dantzig, B.C. Eaves, *Fourier-Motzkin Elimination and Its Dual with Application to Integer Programming*. In: B. Roy (eds) *Combinatorial Programming: Methods and Applications* **19**, Springer, Dordrecht.

[54] These results can be checked on `https://github.com/gabirelpa/Cenario_CHSH` and `https://github.com/gabirelpa/Cenario_Sliwa`.

[55] The cluster webpage is available on `https://sites.ifi.unicamp.br/hpc/clusters/feynman/` in portuguese.

[56] G.B. Dantzig, *Linear Programming and Extensions*, Princeton University Press, New Jersey (1963).

[57] K.G. Murty, *Linear programming*, John Wiley & Sons, New York (1983).

[58] S. Pironio, J.-D. Bancal, V. Scarani, J. Phys. A: Math. Theor. **44**, 065303 (2011).

[59] J.F. Clauser, M.A. Horne, Phys. Rev. D **10**, 526 (1974)

[60] D. Collins, N. Gisin, J. Phys. A: Math. Gen. **37**, 1775 (2004).

[61] R.F. Werner, M.M. Wolf, e-print: arXiv: 0107093 [quant-ph].

[62] M. Navascués, S. Pironio, A. Acín, New J. Phys. **10**, 073013 (2008).