ICFO – The Institute of Photonic Sciences
Castelldefels (Barcelona), Spain

PhD thesis

Characterizing and witnessing multipartite correlations: from nonlocality to contextuality

Ana Belén Sainz

December 5, 2013

Thesis supervisor Dr. Antonio Acín

Acknowledgements

This PhD thesis is the result of more than four years of work, which I wouldn't have been able to do without the support of any of you. First, I would like to thank Toni, for giving me the great opportunity of doing my PhD with him in his group. Muchas gracias Toni, por todo el tiempo que me has dedicado y por todo lo que me has enseñado, tanto de física como de la vida. Gracias por las oportunidades que me has dado para crecer, y por la libertad y confianza que me has brindado.

Tobias, the great mind. Thanks for the patience, for the shared moments, for all the work we've done together. Thanks for being the first person who ever called me "colleague".

Anthony, the first postdoc I worked with. Even though the topic of our first project was a "road to nowhere", we eventually found a path to walk together, and with very nice results! Thanks for all the things you've taught me, and always answer my questions with a smile.

Jordi, Remik, Jonatan, Planeta: Thanks for being such great teams I got to be a part of during these four years.

Chirag, Ariel, Gonzalo, Matty, Iñaki: my dear friends, thanks for the talks, the hugs, the "mates" and the beers.

Lars and Elsa, with whom I shared every single hour of my time at the best office in ICFO. Larcito, thanks for the patience, the LaTeX, Matlab and Mathematica tutorials, and your particular sense of humour. Elsa, thanks for the talks, the laughters and the teas.

Thanks to all of you who are or have been at Toni's group these years, for the nice lunch breaks, groups meetings and groups dinners.

Vivi y Carli, gracias por haberme recibido con los brazos abiertos allá por julio del 2009. Por haberme dado un hogar en Barcelona y así no extrañar tanto mi casa en Argentina. Gracias por los mates, el fernet, los asados, y la amistad.

Edu, mi primer compañero de piso y desde entonces gran amigo. Gracias por ser vos.

A mi mamá y mi papá. Gracias por haberme dado la libertad de elegir mi camino, que me trajo hasta acá, y el apoyo constante para lograrlo.

A mis amigos en Argentina, con quienes pude ver que a pesar de la distancia

los sentimientos no cambian. Anita, Besma, Bere, Dioni, gracias por estar siempre.

A Marta y Juancho, Neus, Osamu, y todos con quienes pude compartir de mi vida en Barcelona estos años.

These four years and this PhD thesis have been a long road, and I couldn't have made it to the end without any of you. Thank you very much for always being there for me !!!

Abstract

In the past century, with the advances of technology, experimental discoveries have witnessed phenomena in Nature which challenge our everyday classical intuition. In order to explain these facts, quantum theory was developed, which so far has been able to reproduce the observed results. However, I believe that our understanding of quantum mechanics can be significantly improved by the search for an operational meaning behind its mathematical formulation, and that a better understanding of quantum physics is essential for identifying the limitations and possibilities of the theory for information processing.

An intriguing property of quantum theory is its intrinsic randomness. Indeed, Einstein, Podolsky and Rosen in their seminal paper of 1935 questioned the completeness of quantum theory. They argued the possibility of the existence of a broader complete theory where variables to which we have not access determine the behaviour of physical systems, and the randomness we observe in quantum mechanics is then due to our ignorance of these variables. These *hidden variable* theories, however, were proved not to be enough for explaining the predictions of quantum theory, as shown in the no-go theorems by Bell on quantum-nonlocality and by Kochen and Specker on quantum-contextuality.

In the past decades, many experiments have corroborated the nonlocal and contextual character of Nature. However, no intuition behind these phenomena has been found, in particular about what limits their strength. In fact, special relativity alone would allow for phenomena which are more nonlocal than what quantum theory allows. Hence, much effort has been devoted to finding physical properties that restrict these phenomena as predicted by quantum theory.

In this thesis, we study the constraints that arise on nonlocal and contextual phenomena when a certain exclusiveness structure compatible with quantum theory is imposed in the space of events. Here, an event denotes the situation where an outcome is obtained given that a specific measurement is performed on the physical system. Regarding nonlocality, we introduce a notion of orthogonality that states that events involving different outcomes of the same local measurement must be exclusive, and further construct constraints that the correlations among observers should satisfy. We denote this by *Local Orthogonality principle*, which is the first intrinsically multipartite principle for bounding quantum correlations. We prove that our principle helps identifying

the supra-quantum character of some bipartite and multipartite correlations, and gets close to the quantum boundary. Regarding contextuality, the events that correspond to different outcomes of the same measurement are considered (naturally) exclusive. However, the same abstract event may correspond to outcomes of *different* measurements, which introduces a non-trivial structure in the space of events. In order to study this, we develop a general formalism for contextuality scenarios in the spirit of the recent approach by Cabello, Severini and Winter. In our framework, quantum nonlocality arises as a particular case of quantum contextuality, which allows us to study the Consistent Exclusivity principle as a generalization of Local Orthogonality. Both in nonlocality and contextuality, we find close connections to problems in combinatorics, which allow us to use graph-theoretical tools for studying correlations.

The last part of this thesis is concerned with the problem of nonlocality detection in quantum systems. Most results on quantum nonlocality focus on few particles' experiments, while much less is known about the detection of quantum nonlocality in many-body systems. Standard many-body observables involve correlations among few particles, while there is still no multipartite Bell inequality to test nonlocality merely from these data. In this thesis, we provide the first proposal for nonlocality detection in many-body systems using two-body correlations. We construct families of Bell inequalities only from one and two-body correlators, which can detect nonlocality for systems with large number of constituents. In addition, we prove violations by systems which are relevant in nuclear and atomic physics, and show how some of these inequalities can be tested by measuring global spin components, hence opening the problem to experimental proposals and realizations.

List of Publications

- J. Tura, R. Augusiak, A. B. Sainz, T. Vértesi, M. Lewenstein and A. Acín, "Translationally invariant Bell inequalities with two-body correlators", Preprint at arXiv:1312.0265, December 2013. Submitted to the special issue "50 years of Bell's theorem" of *Journal of Physics A*.

- A. B. Sainz, T. Fritz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier and A. Acín, "Exploring the Local Orthogonality Principle", Preprint at arXiv:1311.6699, November 2013. Submitted to *Physical Review A*.

- Tobias Fritz, Anthony Leverrier and Ana Belén Sainz, "Probabilistic models on contextuality scenarios", Preprint at arXiv:1307.0145, July 2013. To appear in the proceedings of "Quantum Physics and Logic 2013", published at *Electronic Proceedings in Theoretical Computer Science*.

- J. Tura, R. Augusiak, A. B. Sainz, T. Vértesi, M. Lewenstein and A. Acín, "Detecting non-locality in multipartite quantum systems with two-body correlation functions", arXiv:1306.6860, June 2013. Submitted to *Science*.

- Tobias Fritz, Anthony Leverrier and Ana Belén Sainz, "A Combinatorial Approach to Nonlocality and Contextuality", Preprint at arXiv:1212.4084, December 2012. Submitted to *Communications in Mathematical Physics*.

- T. Fritz, A. B. Sainz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier and A. Acín, "Local Orthogonality as a multipartite principle for quantum correlations", *Nature Communications* **4**, 2263 (2013).

Contents

1. Introduction

It is a well accepted fact that the predictions of quantum theory are incompatible with those of classical physics. Nonlocality and Contextuality are phenomena which cannot be explained in the classical world, although they arise naturally in quantum mechanics. In the past decades, nonlocality and contextuality have become a fruitful topic of research, where new discoveries have revealed their intrinsic interdisciplinarity and striking applications in a wide range of topics far from the setting in which they were originally discovered.

In spite of the new insights on these phenomena, little is known about the structure of quantum nonlocality and contextuality, which is still a fundamental open problem. Indeed, no concise operational characterization of these quantum predictions has been found so far, and it is unclear whether such a characterization even exists. From the physical perspective, such a question has its root in Bell's theorem (Bel64) and the Kochen-Specker theorem (KS67), which are arguably some of the most fundamental lessons we have learned about Nature in the past decades.

Another interesting question focuses on the detection of these nonlocal properties. Even though the nonlocal nature of physical systems has already been observed in the lab, these experiments so far consist only of few-particles, whether a nonlocal test for quantum many-body systems is still missing. Such a test would provide insight on the role of nonlocality in systems with a large number of constituents, something that has hardly been studied.

Quantum Information Theory studies how to combine information theoretic concepts with quantum theory, in order to exploit these non-intuitive quantum phenomena to perform information tasks that are classically impossible. In this thesis, I will further combine notions of information theory to tackle the above mentioned fundamental problems, namely the characterization and detection of quantum nonlocality and contextuality.

In what follows, I present the different questions this thesis focuses on, and summarize the achieved contributions.

1.1. Motivation

Characterization of the Quantum Boundary for correlations

In the seminal paper of 1935, Einstein, Podolsky and Rosen (EPR35) noticed for the first time the phenomenon of quantum entanglement, which ultimately lead to the formal definition of Nonlocality by J. Bell in 1964 (Bel64). This phenomenon of nonlocality tells that correlations observed among the outcomes of spacelike separated measurements on a quantum system may be stronger than those predicted by classical mechanics. Since then, and up to loopholes, nonlocality has been observed many times in Nature (FC72; AGR82; RKM$^+$01; MMM$^+$08).

Bell's theorem relies on natural assumptions on the causal structure of experiments. These assumptions are that: far-apart observers cannot influence each other instantaneously (locality, also known as no-signaling), they can choose their respective measurements independently (free-will), and physical quantities have well-established values previous to any measurement (reality, also known as determinism). Based on these natural assumptions, Bell imposes restrictions on the correlations that the distant parties may obtain, and finds quantum states and measurements that do not satisfy them. Hence, quantum mechanics violates at least some of the assumptions behind Bell's construction. Surprisingly, quantum mechanics is not the only theory that exhibits these nonlocal features. Indeed, if the only assumption over the correlations is the well-founded principle of No Signaling (that it, that the parties cannot communicate instantaneously), these no-signaling correlations (NS) prove to be not only nonlocal, but also more nonlocal than quantum theory allows (PR94; Tsi80).

Pioneering work by Popescu and Rohrlich showed that the no-signaling principle alone does not suffice to recover the set of quantum correlations (PR94), as commented before. Indeed, they provided paradigmatic examples of correlations between two parties compatible with the no-signaling principle but without any quantum realization. Since then, several principles based on information theory have been proposed with the hope of characterizing the set of quantum correlations, e.g. non-trivial communication complexity (vD00; BBL$^+$06), Information Causality (PPK$^+$09), and Macroscopic Locality (NW10). Unfortunately, while being more restrictive than no-signaling, none of these principles seems to be sufficient to recover the quantum set exactly. One of the reasons behind it lies in the fact that intrinsically multipartite principles are essential to characterize the set of quantum correlations (GWAN11). It was indeed proven that there exist supra-quantum correlations for three parties that cannot be detected by any bipartite principle through the following protocol: the parties are split into two groups and the principle is applied to each such bipartition.

Unfortunately, most of the existing principles for quantum correlations are formulated in a bipartite setting and it is unclear whether they have more powerful multipartite generalizations.

Characterization of quantum contextuality

Similar to nonlocality, contextuality is another phenomenon inconsistent with the predictions of classical physics. The original ideas in this topic focused on the property of classical systems that when a set of observables is measured over them, then there is always possible to associate deterministic outcomes to each of them. In 1967, Kochen and Specker (KS67) proved that there exist quantum systems and measurement settings where the previous *noncontextual* property is no longer satisfied, that is, every possible assignment of deterministic outcomes to the observables is not consistent with the functional relations between these observables implied by quantum mechanics. Moreover, even if the assumption of determinism is relaxed and convex combinations of deterministic assignments are considered, the theoretical predictions cannot be explained. Hence, when performing measurements over a quantum system, the allowed conditional probability distributions over the outcomes form a set of probabilistic models which is larger than that allowed by classical mechanics.

The contextual character of quantum mechanics, however, is not as strong as general probabilistic theories allow. Indeed, if the only assumption over the probabilistic models is that they satisfy the No-Disturbance principle (generalization of No Signaling to contextuality scenarios, sometimes referred to as sheaf-condition (AB11)), the allowed models may be more contextual than quantum theory allows. Hence, similar to the case of quantum correlations, the search for principles that bound the set of quantum probabilistic models is a current interesting problem.

Until now, the study of contextuality seems to have been concerned with "small" proofs of the Kochen-Specker theorem (LBPC13) and particular examples of contextuality. However, a general theory has hardly been developed, apart from the study of test spaces in quantum logic (CMW00; Wil09), Spekkens' work on measurement and preparation contextuality (Spe05; LSW11), the graph-theoretic approach of Cabello, Severini and Winter (CSW10), and the sheaf-theoretic approach pioneered by Abramsky and Brandenburger (AB11).

Detection of nonlocality in many-body systems.

The nonlocal character of Nature has already been observed (up to loopholes) in different experiments. These usually consist of few particles (parties), whereas many-body systems have hardly been studied. The reason for this are the technical difficulties that such a problem encompasses, both theoretically and experimentally. On the one hand, finding all the Bell inequalities that characterize classical correlations for an arbitrary number of parties is a difficult task, since the complexity of the problem increases exponentially with the number of parties. On the other hand, in experiments involving many-body systems one has access only to few-body correlations, often two-body, and sometimes individual particles may not be addressed. This complicates the use of the known multipartite Bell inequalities (Ś03; BGP10; WW01; ZB02; LPZB04; AGCA12), since they usually involve products of observables of all parties.

Recently, Bell inequalities involving all-but-one parties have proven useful for detecting nonlocality in quantum systems (BSV12; WNZ12; WBA$^+$12), thus showing that all-partite correlations are sometimes not necessary. However, this improvement is still not enough for tackling many-body systems. In order to study these systems with current technology, one needs to ask a more demanding question: whether nonlocality detection is possible for systems of an arbitrary number of parties from the minimal information achievable in a Bell test, i.e. two–body correlations.

1.2. Main results and contributions

Local Orthogonality: a multipartite principle for quantum correlations.

We propose the first intrinsically multipartite principle to characterize quantum correlations. This is called *Local Orthogonality* (LO), and is presented in chapter 3. This principle is based on a definition of orthogonality (or exclusiveness) between events involving measurement choices and results by distant parties: we define some pairs of events to be orthogonal, or exclusive, whenever they involve different outcomes of the same local measurement by at least one of the parties. Imposing that the sum of the probabilities of mutually exclusive events is less than or equal to one implies a restriction on the possible correlations. These are Bell inequalities which we call LO inequalities, and one of our basic observations is that they are satisfied by quantum correlations. Violations of LO inequalities hence witness supra-quantum correlations.

The LO principle has a nice information-theoretical interpretation in terms of Distributed Guessing Problems and implies a highly non-trivial structure in the space of correlations. We show how to use the multipartite constraints to detect the nonquantumness of some supra-quantum bipartite NS correlations, and prove that the intrinsically multipartite formulation of the principle allows one to detect supra-quantum correlations for which any bipartite principle fails.

An important property of LO is its connection with Graph Theory, which proves very useful when computing the constraints that LO imposes on the space of correlations. Indeed, this problem is equivalent to computing some graph-theoretical invariants of what we call the *orthogonality graph* of the scenario.

A new framework for the study of contextuality.

We develop a graph-theoretical framework for Contextuality, similar to that of Cabello, Severini and Winter (CSW10), but which allows the study of both nonlocality and contextuality in a unified manner. Our approach, presented in chapter 4, demands the probabilistic models to be normalized and defines the notion of "product scenario", which allows Bell scenarios to arise as a particular case of general Contextuality scenarios.

Our framework is well suited to study probabilistic models under the Consistent Exclusivity principle (CE) (Hen12; FLS12; Cab13). In particular, we prove that within our definition of Contextuality scenarios, the Local Orthogonality principle and CE are equivalent. Similarly to the LO case, the CE principle imposes a highly non-trivial structure on the probability space, which until now had been unnoticed. Here we stress this structure by defining a hierarchy of sets of probabilistic models, each level satisfying stronger constraints formulated from CE.

In addition, our framework is also well suited to define a hierarchy of semidefinite programs (SDP) for probabilistic models on contextuality scenarios, similar to that of Navascués, Pironio and Acín (NPA07; NPA08; PNA10). This SDP hierarchy converges into the quantum set, and each level satisfies the CE principle.

Finally, being a graph-theoretically based framework, our approach profits from graph theory. Indeed, we are able to characterize in terms of graph theoretical invariants some sets of probabilistic models, such as the no-signaling set, all the sets in the CE hierarchy, the first level in the SPD hierarchy, and the classical set.

Bell inequalities from two-body correlations.

We propose in chapter 5 a solution for the problem of detecting nonlocality in many-body systems. First, we focus the study on Bell inequalities that contain only one and two-body correlators, simplifying the complexity of both the theoretical and the experimental problems. In principle one could argue the relevance of such inequalities, since in general the correlators that involve a large number of parties are those which carry detailed information about the correlations. Nevertheless, we find that one and two-body correlators are already useful for detecting nonlocality in physically relevant systems. These systems are the Dicke states, which arise as the ground state of the Lipkin-Meshkov-Glick Hamiltonian (LMG65). Moreover, these inequalities are strong enough to detect nonlocality in many-body systems. Finally, we show that in some cases the derived inequalities are experimentally friendly, as they can be tested through measurements of global observables such as the components of the total spin, which are routinely measured in atomic physics with great precision (HSP10; ERIR^{+}08). In particular, this makes our nonlocality criteria applicable in systems where individual particles cannot be addressed.

2. Preliminaries

In this chapter I present the key concepts needed to understand the topic of this thesis and its scope. I start with the notion of Nonlocality, including the definition of a Bell experiment and the relevant sets of correlations. Then, I move on to Contextuality, where I introduce the phenomenon and review relevant frameworks for its study. Finally, I briefly present the concepts from Graph Theory that will be of use throughout the thesis.

2.1. Nonlocality

Nonlocality is one feature of quantum mechanics that is not present in the classical world. It tells that the correlations observed among the outcomes of spacelike separated measurements on a quantum system may be stronger than those predicted by classical mechanics. This fact was first noticed by Einstein, Podolsky and Rosen in their seminal paper in 1935 (EPR35), but it was not until 1964 that J. Bell proved (Bel64) that the predictions of quantum theory are incompatible with those of classical physics. The idea behind Bell's theorem goes as follows. Consider, for simplicity, two distant parties, Alice and Bob, each of them having access to a physical system. Alice (Bob) is allowed to freely choose local measurements to perform on her (his) system. Let x and y denote the measurement choices of Alice and Bob, and a and b the corresponding outcomes. The actions one party performs are assumed to be space-like separated from those of the other party. We are interested in the joint conditional probability distribution $P(ab|xy)$, i.e. the probability that Alice and Bob obtain the outcomes a and b *given* they have measured x and y (in this thesis, I will also use the word *correlations* to refer to $P(ab|xy)$). The main point addressed by Bell is that there exist correlations which arise from measurements on quantum systems and that can not be explained by classical mechanics. In fact, imagine the case where the state that the parties share is entangled. Then, the correlations they achieve may be stronger than those attainable by any possible local strategy the parties could perform on classical systems, even when deciding in advance which strategy to use and sharing a source of randomness. These correlations incompatible with classical theory are called *nonlocal*. In this section, I present the notion of Bell-type experiment (or Bell scenario), and the relevant sets of conditional probability distributions that arise from it.

A general Bell Scenario involves n distant parties, each of them having access to a physical system (see Fig. 2.1). In principle, the number of measurements the parties have access to may differ from party to party, and so the number of outcomes that each party observes. We focus however on the case where the number of measurement choices is the same for all the parties, and moreover, that all the measurements have the same number of outcomes. In the scope of this thesis then, each party can perform m different measurements on his system, getting one out of d possible outcomes. This scenario is denoted by (n, m, d). The measurement applied by party i is denoted by x_i, and the corresponding outcome by a_i, with $i \in \{1, \ldots, n\}$, $x_i \in \{0, \ldots, m-1\}$, and $a_i \in \{0, \ldots, d-1\}$. The correlations among the parties are described by the joint conditional probability distribution $P(a_1 \ldots a_n | x_1 \ldots x_n)$, representing the probability for the parties to get outcomes $a_1, \ldots, a_n$ when making measurements $x_1, \ldots, x_n$. In order for P to be well defined, it should satisfy two basic constraints:

$$P(a_1 \ldots a_n | x_1 \ldots x_n) \geq 0, \forall \{a_1 \ldots a_n | x_1 \ldots x_n\} \quad (positivity) \tag{2.1}$$

and

$$\sum_{a_1 \ldots a_n} P(a_1 \ldots a_n | x_1 \ldots x_n) = 1, \forall \{x_1 \ldots x_n\} \quad (normalization) \tag{2.2}$$

In the particular case of two parties, i.e. $(2, m, d)$, we denote for simplicity the correlations by $P(ab|xy)$ instead of $P(a_1 a_2 | x_1 x_2)$.

2.1.1. Box World Scenarios

In this section I present the case where no assumptions are made on the systems the parties have or the specific way the measurements are implemented. This framework is usually called *Box World*, and used for the study of *device independent* tasks, such as device-independent quantum key distribution (ABG+07; VV12b) and device-independent random number generators (PAM+10; VV12a). Here, each party is thought of as having a device (box) in which by pressing a button produces an outcome. In order for these devices to be well defined objects, there should exist a conditional probability distribution $P(a_i|x_i)$ for each box, that represents its observed behavior. This constrains the correlations $P(a_1 \ldots a_n | x_1 \ldots x_n)$ allowed in the framework, like follows:

$$\sum_{a_{i+1} \ldots a_n} P(a_1 \ldots a_n | x_1 \ldots x_n) = P(a_1 \ldots a_i | x_1 \ldots x_i) \tag{2.3}$$

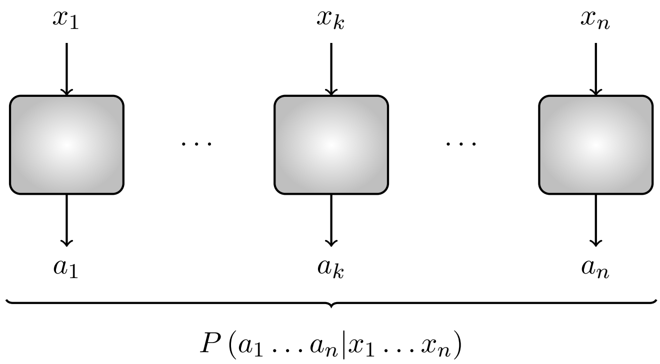


Figure 2.1.: Bell scenario: n distant parties each having access to a device (box). Each party performs a measurement x_k on his device, obtaining an outcome a_k. The correlations among the devices are studied through the conditional probability distribution $P(a_1 \dots a_n | x_1 \dots x_n)$.

for any splitting of the n parties into two groups. This assures that all the marginals of P are well defined, i.e. do not depend on the inputs of the parties that are traced out. Condition (2.3) is the so-called **No Signaling principle** (NS), and defines the set $\mathcal{NS}$ of no-signaling correlations. Besides this operational formulation, NS is often stated as the "impossibility of instantaneous communication among the parties", which originally relates it to special relativity or any other theory where the speed of communication is bounded.

In what follows, different sets of correlations that arise in the box world scenario are presented.

2.1.2. Classical Correlations

Classical correlations are those that arise when the parties have access each to a classical system, on which to perform measurements. In the spirit of (EPR35), these correlations can be explained in terms of local hidden variables as follows. In a local classical theory, the correlations are fully attributed to a common cause λ in the past of the measurements instances (BCP$^+$13). Hence, the statistics of each device reads $P_j(a_j|x_j, \lambda)$, and the conditional probability distribution factorizes as $P(a_1 \dots a_n | x_1 \dots x_n) = \Pi_{j=1}^n P_j(a_j|x_j, \lambda)$. However, when actually performing a Bell test, the parties go through many runs of performing measurement and obtaining outcomes, in order to produce the data from which to compute the statistics. Hence, in principle the common cause λ

may change between runs. The correlations may then be expressed as

$$P\left(a_1 \ldots a_n | x_1 \ldots x_n\right) = \int q(\lambda)\, \Pi_{j=1}^{n} P_j(a_j|x_j, \lambda) \mathrm{d}\lambda, \tag{2.4}$$

where $q(\lambda)$ denotes the probability distribution for the hidden variable λ.

Local correlations are equivalently defined in terms of *deterministic* local hidden variable models. First, consider a *deterministic strategy* for one party, i.e. a conditional probability distribution that assigns a deterministic output to each measurement. For this case, the behaviour of the devices is completely determined, and the observed statistics satisfy $D_j(a|x) := P_j(a|x) = \delta_{a,a_x}$, where a_x is the fixed output of measurement x for the device j. By $D_j(a|x)$ we denote such a deterministic conditional probability distribution for one party. A joint deterministic strategy is then given by a deterministic assignment for each party, i.e. $D\left(a_1 \ldots a_n | x_1 \ldots x_n\right) = \Pi_{j=1}^{n} D_j(a_j|x_j)$. If we consider now local hidden variables, the deterministic conditional probability distributions read $D(a|x) = D(a|x, \lambda)$, and hence the outcome of measurement x is completely determined by the information carried within the hidden variable λ and the measurement choice itself. In this sense, the correlations are expressed by means of the deterministic strategies as follows:

$$P\left(a_1 \ldots a_n | x_1 \ldots x_n\right) = \sum_{\lambda} q_\lambda\, \Pi_{j=1}^{n} D_j(a_j|x_j, \lambda), \tag{2.5}$$

where now λ labels the possible joint deterministic strategies, and $q_\lambda \geq 0$ their corresponding weights in the decomposition. It is easy to see that correlations in eq. (2.4) can always be written as in eq. (2.5), without loss of generality.

Given the product structure of $D\left(a_1 \ldots a_n | x_1 \ldots x_n\right)$, it is easy to see that local correlations satisfy the No Signaling principle. In fact, consider a local correlation defined by eq. (2.5). By tracing out one party, say n, normalization of D_j implies

$$\begin{aligned} P\left(a_1 \ldots a_{n-1} | x_1 \ldots x_{n-1}\right) &= \sum_{a_n} P\left(a_1 \ldots a_n | x_1 \ldots x_n\right) \\ &= \sum_{\lambda} q_\lambda\, \Pi_{j=1}^{n-1} D_j(a_j|x_j, \lambda) \sum_{a_n} D_n(a_n|x_n, \lambda) \\ &= \sum_{\lambda} q_\lambda\, \Pi_{j=1}^{n-1} D_j(a_j|x_j, \lambda). \end{aligned}$$

The same reasoning may be applied when tracing out more than one party.

2.1.3. Quantum Correlations

In the previous section, the devices that the parties have access to in a Bell-type experiment were assumed to be classical, which gave rise to the set of classical correlations. However, that assumption can be relaxed. This section presents the correlations that arise when considering devices of quantum nature.

In quantum theory, the state of a system is an element of a Hilbert space $\mathcal{H}$, represented by a positive semidefinite matrix ρ, usually called *density matrix*. A special class of states is that of *pure* quantum states, which correspond to vectors $|\Psi\rangle$ over the Hilbert space. In this case, the density matrix is given by $\rho = |\Psi\rangle\langle\Psi|$. The observables, moreover, are self-adjoint operators $\mathcal{A}$ on $\mathcal{H}$, whose expectation values are given by the Born's rule $\langle\mathcal{A}\rangle = \mathrm{tr}(\mathcal{A}\,\rho)$. The most general class of measurements over quantum systems is called *positive operator-valued measure* (POVM) (NC03). There, a measurement x is described by a set of nonnegative operators $\{M_a^x\}$ with the following properties: $\sum_a M_a^x = \mathbb{1}_{\mathcal{H}}$, and each operator M_a^x is associated to a possible outcome of the measurement, so that the probability of obtaining a when measuring x is given by $P(a|x) = \mathrm{tr}(\rho\, M_a^x)$. The nonnegativity of $\{M_a^x\}$ assures that the $P(a|x)$ are positive numbers, and the condition that $\{M_a^x\}$ sum up to the identity guarantees the probabilities $P(a|x)$ to be normalized. Note that these operators M_a^x need not be projectors over $\mathcal{H}$. When they are, the measurement belongs to a smaller family called *projective* or *von Neumann measurements*.

An interesting property is that, given a general state ρ and POVM $\{M_a^x\}$ in a Hilbert space $\mathcal{H}$, it is always possible to find a Hilbert space $\mathcal{H}'$ of larger dimension, a state ρ' and a projective measurement $\{\Pi_a^x\}$, with identical statistics for the measurement outputs, i.e. $P(a|x) = \mathrm{tr}(\rho\, M_a^x) = \mathrm{tr}(\rho'\,\Pi_a^x)$ (NC03). Indeed, suppose that we want to perform a measurement $\{M_a^x\}$ on the system ρ. Consider an ancillary system belonging to a Hilbert space $\mathcal{H}_b$, such that there exists a basis of orthonormal states $\{|a\rangle\}$ in $\mathcal{H}_b$ in one-to-one correspondence with the measurement outcomes of $\{M_a^x\}$ over ρ. This ancillary system can be thought of as a purely mathematical device appearing in the construction, or as an actual quantum physical system that helps in the measurement process. Operationally, the main idea then is to perform an entangling operation between the system ρ and the ancilla, which contains the information about the original POVM, and then perform a projective measurement $\{\mathbb{1}_{\mathcal{H}} \otimes |a\rangle\langle a|\}$ over the state of the ancilla. Formally, the construction of $\{\Pi_a^x\}$ from $\{M_a^x\}$ goes as follows. Since $\{M_a^x\}$ are positive semidefinite operators, they may be expressed as[1] $M_a^x = K_a^{x\dagger} K_a^x$. Consider now the initial joint state $\rho' = \rho \otimes |0\rangle\langle 0|$,

[1] The operators K_a^x are usually called *Kraus operators*. The decomposition of the elements of a POVM into its Kraus operators is not unique, since any unitaries acting on $\{K_a^x\}$

where $|0\rangle\langle 0|$ is the state of the ancilla, and define the unitary U which performs the entangling operation $U\rho' U^\dagger = \sum_{a,a'} K_a^x \rho K_{a'}^{x\dagger} |a\rangle\langle a'|$. Finally, the operators $\Pi_a^x = U^\dagger(\mathbb{1}_\mathcal{H} \otimes |a\rangle\langle a|)U$ indeed form a projective measurement over $\mathcal{H}' = \mathcal{H} \otimes \mathcal{H}_b$, with $\mathrm{tr}(\Pi_a^x \rho') = P(a|x)$.

Quantum correlations then arise via Born's rule $P(a|x) = \mathrm{tr}(M_a^x \rho)$ when performing measurements on a quantum device. In a Bell scenario, the situation of n parties measuring in their respective devices corresponds to performing each a POVM $\{M_a^{x_j}\}$. Hence, denoting by ρ the *joint* state of the n systems, the statistics read:

$$P(a_1 \dots a_n | x_1 \dots x_n) = \mathrm{tr}(M_{a_1}^{x_1} \otimes \dots \otimes M_{a_n}^{x_n} \rho). \tag{2.6}$$

We say that a conditional probability distribution $P(a_1 \dots a_n | x_1 \dots x_n)$ is *quantum*[2] whenever there exists a Hilbert space $\mathcal{H}$, a state ρ in $\mathcal{H}$, and a measurement $\{M_a^{x_j}\}$ for each party, such that correlations are recovered by eq. (2.6). Note that by the previous comments, if the dimension of the Hilbert space is not bounded, we can assume the $\{M_a^{x_j}\}$ to be projective measurements.

Quantum correlations do satisfy the No Signaling principle. Indeed, if we sum over the outcomes of one party, say n, the normalization of the POVM's imply

$$\begin{aligned}\sum_{a_n} P(a_1 \dots a_n | x_1 \dots x_n) &= \mathrm{tr}(M_{a_1}^{x_1} \otimes \dots \otimes M_{a_{n-1}}^{x_{n-1}} (\sum_{a_n} M_{a_n}^{x_n})\, \rho) \\ &= \mathrm{tr}(M_{a_1}^{x_1} \otimes \dots \otimes M_{a_{n-1}}^{x_{n-1}} \otimes \mathbb{1}_{\mathcal{H}_n}\, \rho),\end{aligned}$$

i.e. the result is independent of the measurement choice of party n. The same reasoning may be applied when tracing out more than one party.

2.1.4. Geometry of Correlations

In this section, we will introduce some basic concepts regarding the geometry of the space of probability distributions. First, note that a conditional probability distribution P arising from a Bell scenario (n, m, d) may be viewed as a vector $\mathbf{P} \in \mathbb{R}^{(md)^n}$. Indeed, every component of the vector corresponds to a choice of measurements and outcomes, and viceversa.

The set of No Signaling correlations then is the subset $\mathcal{NS} \subset \mathbb{R}^{(md)^n}$ of vectors $\mathbf{P}$ which satisfy the positivity and normalization conditions as well as the No Signaling principle (2.3). Since these are a finite number of linear

preserves the form of $\{M_a^x\}$.

[2]Equivalently, that it has a quantum realization.

constraints, the set $\mathcal{NS}$ is a convex set with a finite number of extreme points, i.e. a *polytope*.

The set of quantum correlations is the subset $\mathcal{Q} \subset \mathbb{R}^{(md)^n}$ of vectors $\mathbf{P}$ that have a quantum realization. Since quantum correlations satisfy No Signaling, $\mathcal{Q} \subset \mathcal{NS}$, and moreover $\mathcal{Q}$ is a convex set. Indeed, consider two quantum conditional probability distributions P_1 and P_2, and denote by $\mathcal{H}_k$, ρ_k and $\{\Pi_{a_j}^{x_j}\}_k$ the corresponding Hilbert spaces, states and measurements for party $j = 1 \dots n$. The conditional probability distribution given by a convex combination of P_1 and P_2, say $P = qP_1 + (1-q)P_2$ with $q > 0$, also has a quantum realization, given by the Hilbert space $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$, the state $\rho = q\rho_1 \oplus (1-q)\rho_2$, and measurements $\{\Pi_{a_j}^{x_j}\}$ of the form $\Pi_{a_j}^{x_j} = (\Pi_{a_j}^{x_j})_1 \oplus (\Pi_{a_j}^{x_j})_2$. However, the set $\mathcal{Q}$ is not itself a polytope and the inclusion $\mathcal{Q} \subset \mathcal{NS}$ is strict. Already for the simplest scenario $(2,2,2)$, there exist NS correlations that do not belong to the quantum set. One such example are PR correlations (PR94), that read $PR(ab|xy) = \frac{1}{2}\delta_{a\oplus b = xy}$, where the sum is taken mod 2. Popescu and Rohrlich noticed, using the result by Tsirelson (Tsi80) (see also (Tsi93)), that a hypothetical device behaving with such correlations may not have a quantum realization, although it would satisfy the NS princple. These PR correlations, also refered to as PR-box, are the unique[3] nonquantum extreme point of the $(2,2,2)$ no-signaling polytope.

Finally, the set of classical correlations, also refered to as *local correlations*, is the subset $\mathcal{C} \subset \mathbb{R}^{(md)^n}$ of vectors $\mathbf{P}$ which are obtained as a convex combination of deterministic points $\mathbf{P}_\mathrm{D}$. These deterministic points correspond to the deterministic strategies mentioned in section (2.1.2). The set $\mathcal{C}$ is convex by definition, with a finite number of extreme points $\mathbf{P}_\mathrm{D}$, and hence is a polytope. Moreover, $\mathcal{C} \subset \mathcal{Q}$. To see this, it suffices to prove that every deterministic point has a quantum realization. Consider then a vector $\mathbf{P}_\mathrm{D}$, and as Hilbert space $\mathcal{H} = \mathbb{C}$. For every choice of measurement settings $(x_1 \dots x_n)$, consider the projective measurement given by $\Pi_{a_1 \dots a_n}^{x_1 \dots x_n} = \mathbb{1}_\mathcal{H}$ if $(a_1 \dots a_n)$ is the deterministic output of $\mathbf{P}_\mathrm{D}$ (hence 0 otherwise). By setting the state $\rho = \mathbb{1}_\mathcal{H}$, the desired deterministic conditional probability distribution arises via Born's rule.

Figure 2.2 shows schematically the chain of inclusions $\mathcal{C} \subset \mathcal{Q} \subset \mathcal{NS}$. The set of correlations that belong to $\mathcal{NS}\backslash\mathcal{C}$, i.e. to $\mathcal{NS}$ but not to $\mathcal{C}$, are called *nonlocal*, while the ones belonging to $\mathcal{C}$ are also called local. Since $\mathcal{C}$ is a polytope, there are a finite number of inequalities which define the set. Indeed, each inequality divides the probability space into two semispaces, and the intersection of all the semispaces allowed by the inequalities defines the classical set.

[3]By unique, I mean that every other nonquantum extreme point may be obtained by a PR box by relabeling the inputs and/or outputs of the device.

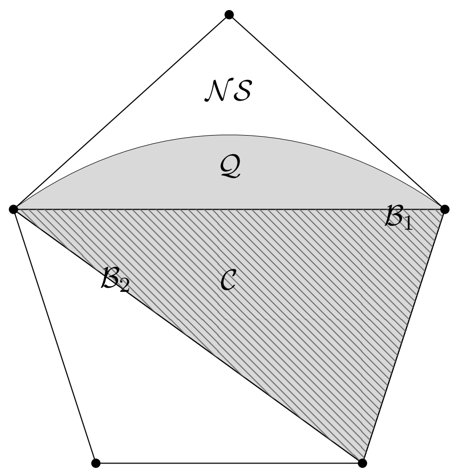


Figure 2.2.: Schematic representation of the sets of no-signaling ($\mathcal{NS}$ – pentagon), quantum ($\mathcal{Q}$ – gray area) and classical correlations ($\mathcal{C}$ – striped area). The lines $\mathcal{B}_1$ and $\mathcal{B}_2$ separating the set of classical correlations from the nonlocal ones are examples of tight Bell inequalities. While $\mathcal{B}_1$ is violated by some quantum correlations, $\mathcal{B}_2$ is only violated by supra-quantum nonlocal conditional probability distributions.

2.1.5. Bell Inequalities

The existence of quantum correlations violating Bell inequalities (BI) has been known since Bell's paper in 1964. Since then, much effort has been devoted to deriving various BI for different scenarios. Examples of such are the CHSH inequality for $(2,2,2)$ (CHSH69), the I_{3322} inequality for $(2,3,2)$ (CG04; Ś03), the chained inequality for $(2,m,d)$ (BKP06) and the CGLMP inequality for $(2,2,d)$ (CGL+02). The most famous BI for the $(2,2,2)$ scenario is the CHSH inequality

$$\begin{aligned} &P(00|00) - P(01|00) - P(10|00) + P(11|00) \\ &+P(00|01) - P(01|01) - P(10|01) + P(11|01) \\ &+P(00|10) - P(01|10) - P(10|10) + P(11|10) \\ &-P(00|11) + P(01|11) + P(10|11) - P(11|11) \leq 2, \end{aligned} \tag{2.7}$$

which is usually presented in *correlator* form

$$E_{00} + E_{01} + E_{10} - E_{11} \leq 2, \tag{2.8}$$

where the correlators are defined as $E_{ij} = P(a = b|ij) - P(a \neq b|ij)$. While classical correlations achieve a maximum value for CHSH of 2, quantum correlations achieve a value up to $2\sqrt{2}$, usually known as Tsirelson's bound (Tsi80). For NS correlations, the maximum value for CHSH is 4, which is the algebraic maximum of the inequality, and is achieved by PR boxes (PR94). It is worth noticing that the CHSH inequality (CHSH69) actually bounds the absolute value of the sums in (2.7) and (2.8), hence imposing two different constraints

$$-2 \leq E_{00} + E_{01} + E_{10} - E_{11} \leq 2.$$

However, the inequalities "$-2 \leq$" and "≤ 2" are equivalent under the relabelling of the measurements' outcomes of a single party.

In general, any inequality of the form $\sum c_j \mathbf{P}_j \geq -\beta_C$ (with real numbers $c_1 \ldots c_{(md)^n}$ and β_C) which is satisfied by classical correlations is called "Bell inequality". However, only some of them are interesting, in the sense that are useful for detecting nonlocality in quantum correlations. β_C is a constant usually called "classical bound", and sets the limit that quantum correlations sometimes violate. Moreover, when the Bell inequality coincides with a facet of the local polytope, it is called *tight* Bell inequality, and the local bound β_C is achieved by a certain number of deterministic points. Some examples of tight Bell inequalities are the CHSH inequality, I_{3322} and the CGLMP inequality.

There are many equivalent ways of writing down a Bell inequality. In fact, I previously presented CHSH in two different formulations. In what follows, I will comment on the no signaling form and the correlators form of a Bell inequality, and introduce the notion of Bell operator.

No signaling form: As mentioned in the previous section, from the geometrical picture of the probability space, finding all the BI inequalities for a given scenario may be seen as a convex hull problem. Even though this problem has been solved for simple scenarios, its complexity increases exponentially with the number of parties, becoming a computationally hard task. One way to simplify the problem is to reduce the dimension of the probability space under study. As shown in (CG04), the NS principle implies that a conditional probability distribution $\mathbf{P}$ of scenario (n, m, d) is completely specified by the values of $P\,(a_1 \ldots a_n | x_1 \ldots x_n)$, for $a_j = 0 \ldots (d-2)$, together with the values of all the k-partite marginals $(k = 1 \ldots n-1)$ where the outcome $d-1$ is not involved. For example, consider the $(2, 2, 2)$ case. In principle, the probability space has dimension $(md)^n = 16$. However, the following elements fully specify the complete conditional probability distribution $\mathbf{P} \in \mathbb{R}^{16}$:

$$\begin{aligned}\mathbf{P}' = [&P_1(0|0), P_1(0|1), P_2(0|0), P_2(0|1),\\ &P(00|00), P(00|01), P(00|10), P(00|11)].\end{aligned} \tag{2.9}$$

where P_j denotes the marginal of P for party j. Hence, studying no-signaling conditional probabilities distributions in the space $\mathbb{R}^{16}$, where positivity, normalization and no-signaling constraints are imposed, is equivalent to studying the objects $\mathbf{P}' \in \mathbb{R}^8$, with constraints being $\mathbf{P}'(j) \in [0, 1]\ \forall\, j = 1 \ldots 8$ and the positivity conditions $P_1(0|x) \geq P(00|xy)$ and $P_2(0|y) \geq P(00|xy)\ \forall x, y$.

Hence, a Bell inequality can also be written as a function of the components of $\mathbf{P}'$ rather that those of $\mathbf{P}$. For instance, the CHSH inequality (2.7) has an equivalent form given by

$$-1 \leq P(00|00) + P(00|01) + P(00|10) - P(00|11) - P_1(0|0) - P_2(0|0) \leq 0,$$

which is called CH inequality (CH74).

Correlator form: In scenarios with binary measurements, Bell inequalities may also appear in correlator form: as commented before, the CHSH inequality (2.7) is usually written as (2.8). For binary measurements, correlators are generally defined as

$$E_{x_1 \ldots x_k} = \sum_{a_1 \ldots a_k} (-1)^{a_1 + \ldots + a_k} P\,(a_1 \ldots a_k | x_1 \ldots x_k), \tag{2.10}$$

where the sum $a_1 + \ldots + a_k$ is taken mod 2. These correlators tell the probability of obtaining an even number of outcomes 1 minus the probability of obtaining an odd number of outcomes 1, given that the k parties have measured $x_1 \dots x_k$. Then, given a conditional probability distribution $P(a_1 \dots a_n | x_1 \dots x_n)$ in a scenario $(n, m, 2)$, one can define a set of correlators, whose elements are listed below:

- $E_{x_1 \dots x_n}$: m^n elements, corresponding to all the possible measurement choices.
- $E_{x_{i_1}} \dots x_{i_{n-1}}$: $n * m^{(n-1)}$ elements. They correspond to the $m^{(n-1)}$ measurement options of the n possible choices of $n-1$ parties $i_1 \dots i_{n-1}$. In order to compute these correlators, eq. (2.10) is used, where the conditional probability distribution considered in the sum is actually the marginal of the n-partite P over the one party which is not in the list $i_1 \dots i_{n-1}$.
- $E_{x_{i_1}} \dots x_{i_{n-2}}$: $\binom{n}{n-2} * m^{(n-2)}$ elements. They correspond to the $m^{(n-2)}$ measurement options of the $\binom{n}{n-2}$ possible choices of $n-2$ parties $i_1 \dots i_{n-2}$. In order to compute them, eq. (2.10) is used, where the conditional probability distribution considered in the sum is actually the marginal of the n-partite P over the two parties which are not in the list $i_1 \dots i_{n-2}$.
- Similarly, for all other $k \in 3 \dots n-1$, the $\binom{n}{n-k} * m^{(n-k)}$ elements of the form $E_{x_{i_1}} \dots x_{i_{n-k}}$, computed for the $n-k$ parties $i_1 \dots i_{n-k}$.

Finally, these correlators are arranged in a vector $\mathbf{E}$ of $(m+1)^n - 1$ components, starting from E_{i_1} to $E_{x_1 \dots x_n}$.

It is worth noticing that $\mathbf{P}$ can be also reconstructed from the correlators $\mathbf{E}$. For instance, in the scenario $(3, 2, 2)$ such an expression reads:

$$\begin{aligned} P(abc|xyz) = & \frac{1}{2^3}\Big\{1 + (-1)^a E_x + (-1)^b E_y + (-1)^c E_z + \\ & (-1)^{a+b} E_{xy} + (-1)^{a+c} E_{xz} + (-1)^{b+c} E_{yz} + (-1)^{a+b+c} E_{xyz}\Big\} \end{aligned}$$

Hence, for binary measurements, studying the objects $\mathbf{P}'$ in probability space is equivalent to studying the vector of correlators $\mathbf{E}$ in *correlator space.* Note however that the dimension of the problem remains the same.

Bell operator: this operator is useful when searching for (theoretical) quantum violations of a Bell inequality. As I mentioned in section 2.1.3, quantum correlations are of the form

$$P(a_1 \dots a_n | x_1 \dots x_n) = \mathrm{tr}(M_{a_1}^{x_1} \otimes \dots \otimes M_{a_n}^{x_n} \, \rho),$$

where $\{M_{a_j}^{x_j}\}_{j=1\ldots n}$ are the measurement operators of the n parties. Hence, a Bell inequality of the form

$$\sum c_{(a_1\ldots a_n, x_1\ldots x_n)} P\left(a_1\ldots a_n|x_1\ldots x_n\right) + \beta_C \geq 0$$

can be written as

$$\mathrm{tr}(\mathcal{B}\,\rho) \geq 0,$$

where $\mathcal{B}$ is the so called *Bell operator*:

$$\mathcal{B} = \sum c_{(a_1\ldots a_n, x_1\ldots x_n)} M_{a_1}^{x_1} \otimes \ldots \otimes M_{a_n}^{x_n} + \beta_C\, \mathbb{1}_{\mathcal{H}}.$$

This form is useful when we search for the existence of a quantum violation. Indeed, if the operator $\mathcal{B}$ has negative eigenvalues, the corresponding eigenvectors $|\lambda\rangle$ provide violations of $\mathrm{tr}(\mathcal{B}\,\rho) \geq 0$ when $\rho = |\lambda\rangle\langle\lambda|$.

Bell operators can be constructed for Bell inequalities also in no signaling or correlators form. In what follows I will comment on the later, since it will be useful in chapter 5.

In quantum mechanics, correlators are related to the expectation value of certain operators called *observables*. Indeed, $E_{x_1\ldots x_k} = \langle\mathcal{M}_{x_1\ldots x_k}\rangle = \mathrm{tr}(\mathcal{M}_{x_1\ldots x_k}\,\rho)$, where the operator $\mathcal{M}_{x_1\ldots x_k}$ is defined as

$$\mathcal{M}_{x_1\ldots x_k} = \sum_{a_1\ldots a_k} (-1)^{a_1+\ldots+a_k} M_{a_1}^{x_1} \otimes \ldots \otimes M_{a_k}^{x_k}.$$

Hence, any Bell inequality in correlators form $\sum_{\mathbf{j}} c_{\mathbf{j}} E_{\mathbf{j}} + \beta_C \geq 0$ can be written as $\mathrm{tr}(\mathcal{B}\,\rho) \geq 0$, where the Bell operator reads $\mathcal{B} = \sum_{\mathbf{j}} c_{\mathbf{j}}\, \mathcal{M}_{\mathbf{j}} + \beta_C\, \mathbb{1}_{\mathcal{H}}$.

2.1.6. Quantum boundary

Bell inequalities define the boundary between classical and nonlocal correlations. On the other hand, no-signaling correlations with no quantum realization are known to exist. A natural question then arises: is there a nice characterization of the boundary between quantum and general no-signaling supra-quantum correlations? Even though the answer is still currently unknown, much effort has been devoted to try to characterize the quantum set. In this subsection, I briefly comment on the state of the art relevant for this thesis.

An important contribution to the search of principles that bound the set of quantum correlations is due to van Dam (vD00). He introduced the idea that the existence of supra-quantum correlations, while not violating the no-signaling principle, could have implausible consequences from an information processing point of view. Particularly, he showed that distant parties having access to PR-boxes can render communication complexity trivial and argued that this could be a reason for the non-existence of these correlations in Nature. This principle is known as "non-trivial Communication Complexity" (CC) (see (vD00; BBL$^+$06)).

Another proposed principle is "Information Causality" (IC) (PPK$^+$09), which is formulated in terms of the following protocol. Alice is given a bit string $\mathbf{x}$ of length n, chosen uniformly at random, and Bob receives a random (integer) number $k \in [1, n]$. Alice then may send a bit string $\mathbf{a}$, of length m, which Bob may use to guess the k'th bit of Alice's original bit string, i.e. x_k. The parties may also share some no-signaling correlations, which can assist Bob's guessing. IC imposes that $\sum_{k=1}^{n} I(x_k : \beta_k) \leq m$, where β_k is Bob's guess of x_k, and $I(x : y)$ is the classical mutual information[4]. The intuition behind this bound is that the total information that Bob can access about Alice's bits cannot exceed the size of the message she sends. The main result of IC is that, if the no-signaling correlations the parties share violate CHSH by a value larger than Tsirelson's bound, then the IC condition is not satisfied. In particular, if the parties share a PR box, then the IC condition is maximally violated, and Bob can always guess with certainty Alice's bit. However, quantum correlations do satisfy IC. Although the initial formulation of IC is indeed based on a particular information-theoretic protocol between Alice and Bob, there have been reformulations of the principle where the figure of merit is expressed in terms of entropies (BBOC$^+$10; ASS11), and other types of games are considered.

Information Causality is an information principle which has a bipartite formulation. Indeed, when applying the principle, a protocol involving two parties should be performed. Hence, if we consider a general scenario (n, m, d), a natural way to extend IC is to divide the n parties into two groups, and perform the protocol considering each group of parties as one new party. The IC principle will then be satisfied for the n parties if it is satisfied for every partition of the parties into two groups. This way of generalizing IC to a multipartite scenario applies for any principle with a bipartite formulation. However, this generalization of bipartite principles for multipartite scenarios proves not to be enough for characterizing multipartite quantum correlations. Indeed, in (GWAN11) it was proven that there exist supra-quantum correlations for three parties which behave classically under any bipartition. Hence, intrinsically multipartite principles are essential to characterize the set of quantum correlations.

A different approach to the problem of characterizing quantum correlations is given by Navascués, Pironio and Acín (NPA) (see (NPA07; NPA08)). They in-

[4] $I(x : y) = H(x) + H(y) - H(x, y)$, where $H(x)$ is the classical Shannon entropy $H(\{q_i\}) = \sum_i q_i \log_2(q_i)$

troduced a hierarchy of semidefinite programming (SDP) tests to check whether a given **P** has a quantum representation. Their method applies to any Bell scenario and is independent of the dimension of the quantum systems. Instead of directly searching for a quantum state and measurements that reproduce the statistics of **P**, they consider a family of weaker conditions. Each of these conditions amounts to verify the existence of a positive semidefinite matrix partially constructed from the elements of **P**, whose structure depends on the algebraic properties satisfied by quantum states and measurements. The family of conditions forms in fact a hierarchy of increasingly stronger constraints, in the sense that if a conditional probability distribution **P** fails one of its levels, the immediate conclusion is that **P** does not have a quantum realization, and there is no need to test the following levels. If a conditional probability distribution **P** satisfies all the levels of the hierarchy, then it belongs to the quantum set. Even though the NPA hierarchy recovers the quantum boundary in the asymptotic limit, it may not seem operationally useful, since there is evidence that the hierarchy may not always converge in a finite number of steps. Indeed, the fact that a conditional probability distribution **P** satisfies one level does not imply that it is quantum. Hence, one would expect that the hierarchy is most useful when trying to detect correlations of supra-quantum nature, since one would just look for a failed test. However, this hierarchy proves also to be handy when computing the quantum bounds of Bell inequalities (PV09). Indeed, given an inequality, each level of the hierarchy can be used to estimate an upper bound on the value that quantum correlations can achieve. For a significantly large number of inequalities, this approach turns out to be very useful since the corresponding upper bounds converge to the real quantum bound in a finite number of steps.

It is worth mentioning that, even though tight Bell inequalities define the boundary between the classical and nonlocal set, sometimes they also define part of the quantum boundary. This unknown feature for bipartite scenarios was first noticed in (ABB$^+$10) for the multipartite case. There, the authors define a game called *guess your neighbour's input* (GYNI), in which each player receives a bit x_j, and assisted by some shared correlations he should guess the input bit of its neighbour. One further rule of the game is the promise that the input bits satisfy the condition $\sum_i x_i = 0$, where the sum is taken mod 2. This game has the following interesting property: players sharing classical correlations perform equally well than when sharing quantum correlations, that is, the overall probability of guessing correctly does not increase when upgrading the shared correlated devices from classical to quantum. However, when the players share supra-quantum no-signaling correlations, their guessing probability increases. In the $(3, 2, 2)$ scenario, for instance, the figure of merit is

$P_{\text{guess}} = \frac{1}{4}\left(P\left(000|000\right) + P\left(110|011\right) + P\left(011|101\right) + P\left(101|110\right)\right)$. For quantum or classical resources $P_{\text{guess}} \leq \frac{1}{4}$, while general no-signaling correlations can achieve $P_{\text{guess}} = \frac{1}{3}$. Condition $P_{\text{guess}} \leq \frac{1}{4}$ is usually referred to as GYNI's inequality for the tripartite scenario, and may be rewritten as

$$P\left(000|000\right) + P\left(110|011\right) + P\left(011|101\right) + P\left(101|110\right) \leq 1. \tag{2.11}$$

Since GYNI inequalities bound the set of classical correlations, they are also Bell inequalities, and since they are not violated by quantum correlations they also bound the quantum boundary. Moreover, for the case of odd number of parties GYNI inequalities prove to be tight Bell inequalities, i.e. they define facets of the local polytope.

2.2. Contextuality

Much effort has been devoted to understanding the peculiarities of quantum theory. In particular, this applies to the phenomenon known as *contextuality*. Formulated as the Kochen-Specker theorem (KS67), but also previously studied by Von Neumann (vN55), it states that quantum theory is at variance with any attempt to assign deterministic values to all observables in a way which would be consistent with the functional relationships between these observables.

Similarly to nonlocality, contextuality tests usually deal with the statistics of measurement outcomes and expectations values of observables. However, in this case there are no space-like separated parties but just a single system under study. The idea then is, given a set of measurements/observables, check the compatibility of the observed outcome's statistics with classical, quantum and no-signaling models. The definitions of section 2.1 naturally apply here, by considering the case of just a single party.

The original idea of Kochen and Specker (KS67) can be rephrased as follows. Consider, without loss of generality, a set of projective measurements and the following promises: (a) when measuring over a system the outcome associated to each projector can have either probability 0 or 1, independently of the measurement the projector belongs to, and hence every projector can be labelled by 0 or 1, (b) for each set of orthogonal projectors that sum up to the identity on the corresponding Hilbert space (i.e. for each complete measurement) only one vector can have the label 1 whereas the others should have the label 0. The question is, are these promises satisfied by any set of projective measurements? The seminal paper by Kochen and Specker (KS67) answered this as negative, and since then many works have been done to find the smallest set of projectors that violate the promises (CEGA96). The key point in this argument is to consider a set of projective measurements such that every measurement shares one projector with another one. Then, a logical contradiction is achieved when trying to assign the values 0 or 1 to these common projectors.

The study of contextuality so far often seems to have been concerned with particular examples of contextuality and "small" proofs of the Kochen-Specker theorem, while a general theory has hardly been developed. Some notable exceptions are the study of *test spaces* in quantum logic (CMW00; Wil09), Spekkens' work on *measurement and preparation contextuality* (Spe05; LSW11), the *sheaf-theoretic* approach pioneered by Abramsky and Brandenburger (AB11), and the *graph-theoretic* approach of Cabello, Severini and Winter (CSW10). In this section, I will comment on the last two.

2.2.1. Observable-based approach

The observable-based approach to quantum contextuality and nonlocality was first explicitly studied by Abramsky and Brandenburger (AB11). They use the mathematical language of sheaf-theory to provide a unified framework where nonlocality arises as a particular form of contextuality.

In this framework, a contextuality scenario is defined by a set of observables X together with a finite set O of outcomes and a *measurement cover* $\mathcal{M}$. A measurement cover $\mathcal{M}$ is defined as a family of subsets of X: $\mathcal{M} \subseteq 2^X$, such that every observable occurs in some subset $C \in \mathcal{M}$ and the following property is satisfied: for any choice of two elements $C, C' \in \mathcal{M}$ (recall that both C and C' are sets of observables), if one is contained in the other, say $C \subseteq C'$, then the two elements are the same $C = C'$. The latter property guarantees a notion of "maximality" for the elements of $\mathcal{M}$. The $C \in \mathcal{M}$ are called **measurement contexts**, and specify the sets of jointly measurable observables.

Abramsky and Brandenburger further define the concepts of no-signaling, quantum and non-contextual (or local) empirical models, in the sense of section 2.1, but whose formulation I will skip since the sheaf-theoretic formalism is beyond the scope of this thesis.

This new insight into the problem of contextuality lead into novel results, such as the distinction among different degrees of strength of contextuality, and provided a wide variety of tools to tackle the problems of nonlocality and contextuality in a unified manner.

2.2.2. Graph-theoretical approach: CSW

Cabello, Severini and Winter (CSW10) studied contextuality scenarios from a graph-theoretical scope. They consider a contextuality scenario to be described as a collection of measurements which states how many outcomes each measurement has, and which measurements have which outcomes in common: here, different measurements may share outcomes, in the sense that outcomes of different measurements may define the same event. For instance, consider the case of projective measurements in quantum mechanics. It may happen that two measurements $\{\Pi^x_a\}_a$ and $\{\Pi^y_b\}_b$ have (at least) one common projector between their sets, say $\{\Pi'^x_a\} = \{\Pi'^y_b\}$. The outcomes associated to these two projectors then corresponds to the same physical property of the system, and hence are represented by the same vertex in the characterization of the contextuality scenario. This vertex will then appear in both the hyperedge corresponding to $\{\Pi^x_a\}$ and the one of $\{\Pi^y_a\}$. An important note on the CSW approach, is that they do not mind whether the measurements that describe the scenario are complete, i.e. $\sum_a p(a|x) \leq 1$. Hence, there might be measurement outcomes that are not considered when defining the scenario.

Formally, CSW defines a contextuality scenario as a hypergraph $H = (V, E)$. The set of vertices V encompasses all the considered outcomes of all the considered measurements, in the sense that every vertex $v \in V$ maps to a measurement outcome and vice-versa. I will usually refer to a vertex as *event*. The hyperedges $e = \{v_1, \cdots, v_k\} \in E$ are subsets of V, and each of them should be thought of as a measurement, where its vertices are the allowed measurement outcomes. Here, a natural notion of orthogonality (or *exclusiveness*) among events comes with the very definition of the scenario. Two events which are outcomes of the same measurement are naturally exclusive. Then, in the language of graph theory, two vertices u and v are orthogonal (denoted by $u \perp v$) if there exists a hyperedge $e \in E$ such that $u \in e$ and $v \in e$. By means of this notion of exclusiveness, CSW further defines the *orthogonality graph*[5] G of contextuality scenario $H = (V, E)$ as follows: the vertex set of G is that of H, and two vertices share an edge in G if there exists a hyperedge in H that contains both of them.

Their aim is to analyze the maximum value of the expression $\beta = \sum_{v \in V} p(v)$, for different families $\mathcal{E}$ of assignments $p : V \to [0, 1]$. The value $p(v)$ is thought of as the probability that the outcome v is obtained when a measurement it belongs to is performed. The results they find are expressed in terms of graph theoretical invariants presented in next section.

The first case of study is *classical assignments*. These are the ones obtained as convex combinations of the deterministic assignments p defined as follows:

[5] In the original CSW paper, it is called *adjacency graph*. Since it coincides with our definition of orthogonality graph of section 3.4, we will keep our notation, not to introduce redundant information.

$p(v) = \{0, 1\}$ and $p(u)\, p(v) = 0$ for all $u \perp v$. The set of classical assignments is denoted by $\mathcal{E}_{\mathrm{C}}$. The maximum of β over the elements of $\mathcal{E}_{\mathrm{C}}$ is $\beta_{\mathrm{C}} = \alpha(G)$ – the independence number of the orthogonality graph (see section 2.3 for definitions).

The second case of study is that of *quantum assignments*. In this formalism, an assignment p is quantum if there exists a Hilbert space $\mathcal{H}$, a state $\rho \in \mathcal{H}$, and projectors $P_v \in \mathcal{H}$ for every vertex $v \in V$, such that (a) if $u \perp v$ then $P_u \perp P_v$, (b) for every set C of mutually orthogonal events $\sum_{v \in C} P_v \leq \mathbb{1}_{\mathcal{H}}$, (c) and $p(v) = \mathrm{tr}(\rho\, P_v)$. The set of quantum assignments is denoted by $\mathcal{E}_{\mathrm{QM}}$. The maximum value of β for such a class of assignments is given by $\beta_{\mathrm{QM}} = \vartheta(G)$ – the Lovász function of G (see section 2.3 for definitions). The intuition behind it comes from the very definition of ϑ. There, a unit vector $|\Psi\rangle$ for graph G together with one unit vector $|\phi_v\rangle$ for each vertex $v \in V$ are found such that orthogonal vectors are assigned to exclusive events, i.e. adjacent vertices, and $\vartheta(G) = max_{|\Psi\rangle, \{|\phi_v\rangle\}} \sum_{v \in V} |\langle\Psi|\phi_v\rangle|^2$.

The third case of study is a *general probabilistic theory* class (GPT), defined as the assignments p satisfying $\sum_{v \in C} p(v) \leq 1$, for every set C of mutually orthogonal events. This requirement was later taken as a principle to bound the set of quantum models for contextuality scenarios, usually referred to as *Consistent Exclusivity* (Hen12; FLS12) or *Global Exclusivity* (Cab13). Hence, the use of GPT to denote this family of models does not relate to general No Signaling models. In any case, for this set, CSW shows $\beta_{\mathrm{GPT}} = \alpha^*(G)$ – the fractional packing number of the orthogonality graph (see section 2.3 for definitions).

CSW also extends the problem to arbitrary linear functions of p, namely $\vec{\lambda}\, \vec{p} = \sum_{v \in V} \lambda_v\, p(v)$, with $\lambda_v \geq 0$ for all $v \in V$. They show that finding the maximum of $\vec{\lambda}\, \vec{p}$ for quantum assignments $p \in \mathcal{E}_{\mathrm{QM}}$ is equivalent to solving a semidefinite program.

The question now is how do Bell Scenarios fit in this formalism. In this case, there are some extra constraints on the assignments p given by the normalization of the conditional probability distributions, which are not taken into account in this subnormalized approach. Hence, when studying Bell Scenarios, CSW includes normalization as an extra constraint in the previously mentioned convex optimization of $\vec{\lambda}\, \vec{p}$. When doing so, they define for every class of assignments $\mathrm{X} = \mathrm{C}, \mathrm{QM}, \mathrm{GPT}$ the new sets $\mathcal{E}^1_{\mathrm{X}}$, as the restriction of $\mathcal{E}_{\mathrm{X}}$ to those assignments that satisfy normalization. First, they prove that in bipartite scenarios this new set $\mathcal{E}^1_{\mathrm{GPT}}$ is the set of no-signaling correlations defined in section 2.1.1. Then, they analyze the maximal violations obtained for the Bell inequalities CHSH and I_{3322}. For CHSH they find $\beta_{\mathrm{QM}} = 2 = \sqrt{2}$, i.e. they recover Tsirelson's bound. However, for I_{3322} they obtain $\beta_{\mathrm{QM}} \sim 0.25147$ while the best know upper bound to I_{3322} is smaller: 0.25087 (see (CSW10) for details). The reason for this discrepancy is attributed to the bipartite structure of the Bell Scenario. When optimizing β_{QM}, one finds a state ρ and projectors Π_v such that $p(v) = tr(\rho\Pi_v)$. However, for a Bell Scenario, as the vertices $v = (v_A, v_B)$ refer to one local event for Alice and one for Bob, the projectors should also satisfy that locality condition $\Pi_v = \Pi_{v_A}\Pi_{v_B}$, with $[\Pi_{v_A}, \Pi_{v_B}] = 0 \quad \forall\, v_A, v_B$, as well as the no-signaling principle. These extra constraints are not taken into account in the SDP, which may lead to a larger value for β_{QM} than the one attainable for Bell Scenarios. The following paragraph may also be considered as a possible cause of this discrepancy.

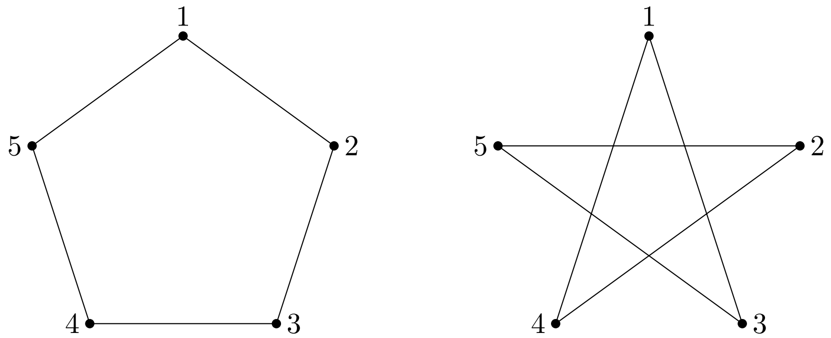
It is worth mentioning that the set of normalized quantum assignments $\mathcal{E}^1_{\text{QM}}$ is not the set of quantum correlations $\mathcal{Q}$ defined in section 2.1.3. Even though both cases are normalized at the level of p, they require different constraints at the level of the projectors P_v. On the one hand, CSW imposes $\sum_{v\in C} P_v \leq \mathbb{1}_{\mathcal{H}}$ for every set C of mutually orthogonal events, even for those C which correspond to the normalized measurements. Note that this is still consistent with the normalization of p, since the projectors may sum up to the identity in the subspace of $\mathcal{H}$ where ρ has support. On the other hand, quantum correlations on a Bell scenario require $\sum_{v\in C} P_v = \mathbb{1}_{\mathcal{H}}$ for those C that correspond to complete measurements. Hence, the orthogonality relations among the projectors must be satisfied in the the subspace of $\mathcal{H}$ where ρ has support. It naturally follows that $\mathcal{Q} \subseteq \mathcal{E}^1_{\text{QM}}$.

All in all, the CSW formalism is a powerful tool when studying general contextuality scenarios, and its connection to graph theory has proven very useful for testing the contextual character of quantum mechanics, both theoretically and experimentally.

2.3. Introduction to Graph Theory

In this section, I will introduce some concepts of graph theory that will be of use throughout the thesis. Formal mathematical definitions are presented in appendix A.

In this thesis, we consider a **graph** to be a collection of points, called *vertices*, some of them joined by edges. A graph is usually denoted by $G = (V, E)$, where V and E are the sets of vertices and edges, respectively. In addition, two vertices are called **adjacent** if they share an edge in E. For example, the *pentagon graph* of Fig. 2.3(a) is defined as $V = \{1, 2, 3, 4, 5\}$ and $E = \big\{\{1, 2\}, \{(2, 3\}, \{3, 4\}, \{4, 5\}, \{1, 5\}\big\}$. In what follows, I will present some useful concepts of graph theory, and use as example the pentagon graph of Fig. 2.3(a).



(a) Pentagon Graph (b) Complement

Figure 2.3.: Example: (a) pentagon graph and (b) its complement. Since (a) and (b) are isomorphic, the graph-theoretic invariants have the same values for both of them. These values are the following: independence number $\alpha = 2$, clique number $w = 2$, Shannon capacity $\Theta = \sqrt{5}$, Lovász number $\vartheta = \sqrt{5}$, and fractional packing number $\alpha^* = \frac{5}{2}$.

Empty graph: A graph $G = (V, E)$ is called *empty* if the edge set is empty, i.e. all the vertices are disconnected.

Subgraph: A subgraph of a graph $G = (V, E)$ is a graph $G' = (V', E')$ whose vertex set V' is a subset of V, and whose adjacency relations E' are those of E restricted to the subset V'. Example: the graph G' constructed from the vertex set $V' = \{1, 2, 3\}$ with edges $\{1, 2\}$ and $\{2, 3\}$ is a subgraph of the pentagon.

Complement graph: The complement of a graph G, denoted by $\overline{G}$, has the same vertex set as G, but complementary adjacency relations: two vertices in $\overline{G}$ are connected if and only if they do not share an edge in G. Example: the graph of Fig. 2.3(b) is the complement of the pentagon.

Independent set: is a set I of vertices of G such that no pair of elements of I share an edge on G. An independent set is **maximal** if it cannot be extended to a larger independent set.

Clique: is a set C of vertices of G such that every pair of elements of I share an edge on G. Note that a clique on G is an independent set on $\overline{G}$, and vice-versa. Similarly, a clique is **maximal** if it cannot be extended to a larger clique.

Cycle: is a sequence of vertices starting and ending at the same vertex, where two consecutive vertices in the sequence are adjacent to each other in the graph. The *length* of a cycle tells the number of vertices the sequence has, where the initial/final vertex is counted only once. Example: in the pentagon, the sequence "$1, 2, 3, 4, 5, 1$" is a cycle of length 5.

Weighted graphs: A weighted graph is a graph where each vertex is equipped with a "weight", i.e. a real number.

Isomorphic graphs: Two graphs G and G' are called isomorphic, if there exists a one-to-one correspondence between the vertex sets V and V' such that: two vertices in V are adjacent in G if and only if the corresponding vertices in V' are adjacent in G'. Example: the complement graph of the pentagon (Fig. 2.3(b)) is isomorphic to the pentagon (Fig. 2.3(a)) via the assignment: $(1, 2, 3, 4, 5) \leftarrow (1, 3, 5, 2, 4)$.

Orthonormal labelling: A set of unit vectors $\{\psi_v\}$, indexed by the vertices $v \in V$ of a graph G, is an *orthonormal labelling* of G if they follow the adjacency relations of G in this sense: whenever two vertices u and v do not share an edge in G, the corresponding vectors ψ_u and ψ_v are orthogonal.

Among the operations defined over graphs, there are many related to the notion of *product*. I will present the two which are used in this thesis.

Conormal product: Given two graphs G_1 and G_2, their *conormal product* is a graph $G = G_1 \cdot G_2$ with the following properties. The vertex set V is the cartesian product between the two sets V_1 and V_2, i.e. every vertex in V is indexed by a vertex in V_1 and one in $V_2 - v = (v_1, v_2)$. Then, two vertices (v_1, v_2) and (u_1, u_2) share an edge in G whenever v_1 and u_1 are connected in G_1 or v_2 and u_2 are connected in G_2.

Strong product: Given two graphs G_1 and G_2, their *strong product* is a graph $G = G_1 \boxtimes G_2$ where the vertex set V is the cartesian product between the two sets V_1 and V_2. The adjacency relations go as follows: two vertices (v_1, v_2) and (u_1, u_2) share an edge in G basically if both v_1 and u_1 are connected in G_1 and v_2 and u_2 are connected in G_2. In the case where $v_1 = u_1$ or $v_2 = u_2$, we only demand that the other two vertices be adjacent in the corresponding graph. Note thus that the edge set $E(G_1 \boxtimes G_2)$ is a subset of $E(G_1 \cdot G_2)$, since the constraints for adjacency are *stronger*.

There are other properties of graphs, sometimes referred to as *graph-theoretical invariants*, which play an important role in my thesis. These are explained in the following.

Independence Number: usually denoted by α, the *independence number* of a graph G is the size of the largest independent set on G. In the case of weighted graphs, it corresponds to the largest weight of an independent set on G, where the weight of I is the sum of the individual weights of the vertices that belong to I. Example: for the pentagon, $\alpha = 2$.

Clique Number: it denotes the size of the largest clique on G. In the case of weighted graphs, it corresponds to the largest weight of a clique on G. Example: for the pentagon, the clique number is $w = 2$. This is not surprising, since the clique number of a graph equals the independence number of its complement, and in the case of the pentagon its complement is isomorphic to itself.

Shannon Capacity: intuitively, the Shannon capacity can be thought of as a limit case of the independence number. The idea is to take the strong product of G with itself k times, compute the independence number α_k of the resulting graph, and by this construct the series $\sqrt[k]{\alpha_k}$. The limit of this series when the number of copies goes to infinite is the *Shannon Capacity* of G, and is denoted by Θ. Example: for the pentagon, Lovász proved in (Lov79) that its Shannon capacity is $\Theta = \sqrt{5}$.

Lovász Number: this quantity, first introduced by Lovász (Lov79), is related to the orthonormal labellings of the graph under consideration. Originally formulated as a bound for the Shannon capacity, it has nowadays many equivalent formulations (see (Knu94)), each of them with its own motivation. The one I find more intuitive to mention here is formalized in eq. (A.16) of appendix A. Consider an orthonormal representation $\{\psi_v\}$ of the *complement* of G, a unit vector Ψ, and compute the sum $S = \sum_{v \in V} (\Psi \cdot \psi_v)^2$. Then, the Lovász number ϑ is the maximum of S when optimizing over the choices of unit vectors Ψ and orthonormal representations $\{\psi_v\}$ of $\overline{G}$. Example: in the case of the pentagon, the Lovász number is $\vartheta = \sqrt{5}$ (Lov79). An orthonormal representation $\{\psi_v\}$ and unit vector Ψ which achieve the value can be constructed as follows: consider an "umbrella" with five "ribs", and associate its handle to the vector Ψ. Then, open the umbrella until the angle between the ribs is $\frac{\pi}{2}$: the ribs can now be associated with the five vectors $\{\psi_v\}$, since they satisfy the orthogonality relations. Note that, since the pentagon and its complement ase isomorphic, any orthonormal representation of the former is also an orthonormal representation of the latter after relabelling the vertices accordingly.

Fractional packing number: usually denoted by α^*, it tells which is the maximum possible total weight of the graph G, when the weights of the individual vertices are asked to satisfy some constraints, namely that no clique on G has a weight larger than 1. In the case when G is already a weighted graph equipped with vertex weights p_v, the fractional packing number maximizes the quantity $\sum_{v \in V} p_v q_v$ over all possible assignments q_v which satisfy that the sum of the values of q_v for any clique is upper bounded by one. Note that the unweighted version of α^* can be computed by its weighted counterpart by setting all the original weights p_v equal to one. Example: for the pentagon, the fractional packing number is $\alpha^* = 5/2$.

In this section I presented a summary of the main concepts I use throughout this thesis. Appendix A provides formal mathematical definitions and properties.

3. Local Orthogonality

How to characterize the boundary of the set of quantum correlations is an interesting and still open problem. Even though there are some proposals for information-theoretic based principles to bound the quantum set (see chapter 2), all of them have a bipartite formulation, while truly multipartite principles, yet to be discovered, are necessary for characterizing quantum correlations in a multipartite Bell scenario (GWAN11).

In this chapter I present the first proposal of an intrinsic multipartite principle to bound the set of quantum correlations: Local Orthogonality (LO). The principle is based on a definition of orthogonality (or exclusiveness) between events involving measurement choices and results by distant parties, and has a natural interpretation in terms of Distributed Guessing Problems. LO also implies a highly non-trivial structure in the space of correlations, specially for multipartite settings. In addition, LO can be used to detect the non-quantumn nature of some bipartite correlations, such as PR-boxes, and to get very close to the boundary of quantum correlations. Finally, I prove that the intrinsically multipartite formulation of the principle allows one to detect supra-quantum correlations for which any bipartite principle fails.

It is worth mentioning that although LO is useful for many tasks, Miguel Navascués proved that it does not recover the set of quantum correlations (NGHA13). In the next chapter I will provide a proof of this statement based on his original one, but also considering contextuality scenarios.

3.1. The Local Orthogonality principle

In this section, I present the formal definition of the LO principle.

A general Bell Scenario, in the device independent formalism, is defined in terms of few parameters: the number of parties, of measurements they have access to, and of outputs those may produce. Hence, little is assumed on the collective events $(a_1 \dots a_n|x_1 \dots x_n)$, which refer to the situation where the parties measure $(x_1 \dots x_n)$ and obtain $(a_1 \dots a_n)$. Then, is there any relation we can tell among the events of a Bell scenario (n, m, d) ? In this thesis, a notion of exclusiveness, or orthogonality, in introduced in this space of events. The idea is the following. Consider two different events e and e' given by

$e = (a_1 \dots a_n | x_1 \dots x_n)$ and $e' = (a'_1 \dots a'_n | x'_1 \dots x'_n)$. We call these two events *locally orthogonal* or simply *orthogonal*, if they involve different outputs of the same measurement by (at least) one party. That is, if for some i we have $a_i \neq a'_i$ while $x_i = x'_i$. We then call a collection of events $\{e_i\}$ orthogonal, or exclusive, if the events are pairwise orthogonal.

The motivation behind these definitions is twofold. First, the inputs and outputs in a device independent framework are thought of classical commands with classical results. Hence, it is natural to think that two situations where the same command is given by at least one party but different responses are obtained, are exclusive. Second, the notion of orthogonality is rather natural in the case of two events. In fact, consider two locally orthogonal events e, e' with $a_i \neq a'_i$ and $x_i = x'_i$ as before. These two events can be seen as different outcomes of a correlated measurement in which: (i) party i first measures x_i and announces the outcome to the other parties and (ii) the other parties apply measurements depending on this outcome, in particular they measure $x_1 \dots x_n$, if the outcome is a_i, and $x'_1 \dots x'_n$ otherwise. Note that, as e and e' are outcomes of the same (correlated) measurement, normalization of the conditional probability distribution implies that $P(e) + P(e') \leq 1$. The No Signaling principle is essential for this correlated measurement to be meaningful, as it is is possible to define the marginal probability for party i in the first step of the correlated measurement independently of the successive actions by the other parties.

The final ingredient of LO is to impose that for any set of orthogonal events, the sum of their probabilities must not be larger than one,

$$\sum_i P(e_i) \leq 1. \tag{3.1}$$

To summarize: the LO principle (i) introduces a notion of orthogonality between two events, (ii) imposes that any number of events are jointly exclusive whenever they are pairwise orthogonal, and (iii) requires that the inequality (3.1) is satisfied for any set of orthogonal events:

Definition 3.1 (Local Orthogonality). *Consider a Bell scenario defined by* (n, m, d). *The* **Local Orthogonality principle** *states that:*

1. *Given two events* $e = (a_1 \dots a_n | x_1 \dots x_n)$ *and* $e' = (a'_1 \dots a'_n | x'_1 \dots x'_n)$, e *is locally orthogonal to* e' *is there exists a party* i *such that* $x_i = x'_i$ *but* $a_i \neq a'_i$.

2. *Given a set of events, these are jointly exclusive whenever they are pairwise orthogonal.*

3. For every set S of exclusive events, the following holds: $\sum_{e \in S} P(e) \leq 1$.

In the previous paragraphs we showed that any two orthogonal events satisfy the inequality (3.1). The principle however becomes more restrictive when considering more events, where the previous reasoning does not apply any longer. As mentioned above, we extend the initial definition of orthogonality for two events to more events by demanding pairwise orthogonality. In addition, the formulation of the principle is independent of the number of parties, hence it is not a bipartite principle generalized to multipartite scenarios by testing every bipartition (see section 2.1.6). It is precisely this extended and intrinsically multipartite formulation that makes the principle non-trivial, because it involves summing probabilities conditioned on different measurements.

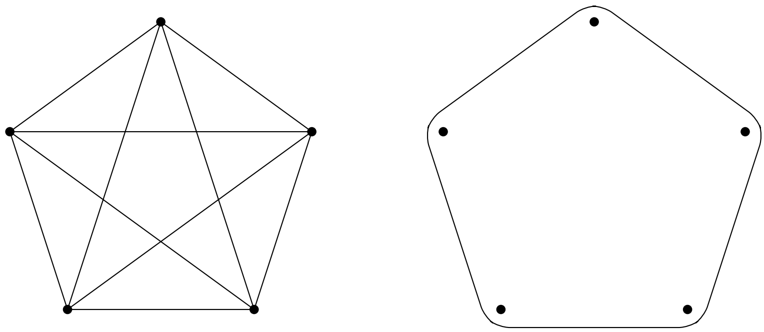
It is worth mentioning that in general probabilistic theories, Boole's axiom only demands that the sum of the probabilities of jointly exclusive events does not exceed 1 (Boo62), while pairwise exclusive events are not necessary jointly exclusive. A nice example of this is given by Henson (Hen12), and consists of the complete graph of 5 vertices, where all the events are orthogonal among each other (see Fig. (3.1)). This orthogonality, by Boole's axiom, implies that the sum of the probabilities of any two events is upper bounded by 1. Thus, a model that assigns a probability $\frac{1}{2}$ to each event satisfies the condition. However, if we now demand pairwise exclusive events to be jointly exclusive, the previous model is not allowed any more, since the sum over all vertices gives a value of $\frac{5}{2}$ in contradiction to Boole's axiom. We see then how the definition of joint orthogonality stated by LO imposes extra non-trivial constraints.

Moreover, it should be stressed that Boole's axiom applies only to probability distributions, whereas the LO principle bounds *conditional* probability distributions. The latter are probabilities *conditioned* on the given measurements, and this condition is usually different for each term in the inequality (3.1). Hence, in order to correctly apply Boole's axiom, one should make the probabilities $P(a_1 \dots a_n | x_1 \dots x_n)$ *unconditional* by considering the probabilities for the measurements $P(x_1 \dots x_n)$ and constructing

$$P(a_1 \dots a_n, x_1 \dots x_n) = P(a_1 \dots a_n | x_1 \dots x_n)\, P(x_1 \dots x_n).$$

Now, even if the original box $P(a_1 \dots a_n | x_1 \dots x_n)$ violates the inequality (3.1), the joint distribution $P(a_1 \dots a_n, x_1 \dots x_n)$ is a well-defined probability distribution and hence satisfies Boole's axiom.

Every (maximal) set of pairwise orthogonal events then gives rise to an inequality via eq. (3.1), which we denote by "LO inequality". All the restrictions on the conditional probability distributions implied by LO come from all the LO inequalities for the given (n, m, d) scenario. The set of LO correlations is then



(a) Pairwise exclusive (b) Jointly exclusive

Figure 3.1.: Example of exclusive events when demanding (a) pairwise exclusivity or (b) general joint exclusivity. In the first case, an assignment of a probability $\frac{1}{2}$ to each event is allowed, since for every pair of exclusive events, the sum of their probabilities is 1. For the second case, this assignment is no longer allowed, since the total sum over all events equals $\frac{5}{2}$, which contradicts Boole's axiom. Hence, the LO principle imposes that a set like (a) behaves like (b) regarding exclusivity properties, which as commented translates into non-trivial constraints.

the set of conditional probability distributions $P(a_1 \dots a_n | x_1 \dots x_n)$ satisfying all the LO inequalities, and is denoted by $\mathcal{LO}^1$.

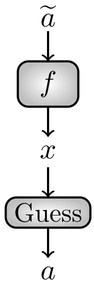
3.2. LO as an information task

Before moving on to the characterization of LO correlations, I provide an interpretation of the principle from an information processing viewpoint. To this end, the notion of a Distributed Guessing Problem (DGP) is introduced.

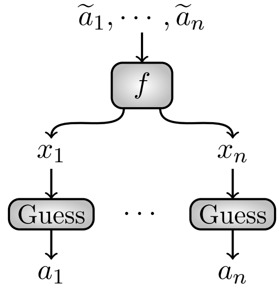
Guessing problems are ubiquitous in science. In the standard formulation (see Fig. 3.2(a)) an observer has access to some data x which depends on some initial parameter $\widetilde{a}$, that is $x = f(\widetilde{a})$. From the observed data, the observer should make a guess, a, of the initial parameter. His goal is to maximize the probability of guessing correctly.

Guessing problems can be easily adapted to distributed scenarios. It is convenient to present the distributed guessing problem as a game and to phrase it in terms of vectors of symbols. Consider then a non-local game in which a referee has access to a set of vectors of n symbols with values in $\{0, \dots, d-1\}$. Denote this set by S and by $|S|$ its size, which can be less than d^n in general. Now, the referee chooses a vector $(\widetilde{a}_1, \dots, \widetilde{a}_n)$ uniformly at random from S, and encodes it into a new vector of, again, n symbols using a function f. However, the new symbols can now take m values and, thus, $f : S \longrightarrow \{0, \dots, m-1\}^n$. The resulting vector is $(x_1, \dots, x_n) = f(\widetilde{a}_1, \dots, \widetilde{a}_n)$. These n symbols are distributed among n distant players who cannot communicate and must produce individual guesses $a_1, \dots, a_n$. Their goal is to guess the initial input to the function, that is, they win whenever $a_j = \widetilde{a}_j$ for all j. Note that the encoding function f and the set S are known in advance to all the players, who in addition may share some device that correlates them.

For some f, e.g. for $x_j = \widetilde{a}_j$, this task is very simple. However, there exist functions for which the task becomes extremely difficult. For a fixed size $|S|$, the most difficult functions are those for which the maximum guessing probability is equal to $1/|S|$. The players can always achieve this guessing probability by agreeing beforehand on one of the $|S|$ possible outputs, which they output regardless of the x_j. Since the input is uniform on S, their guess is correct with probability $1/|S|$. A DGP is thus *maximally difficult* whenever this strategy is optimal, that is, whenever it is impossible to provide a better estimate of the input than random guessing. For such an f, having access to the symbols x_j does not provide any useful information to the parties. Note that non-trivial maximally difficult functions are possible only in distributed scenarios. In standard single-observer guessing problems, the only maximally difficult function is



(a) Standard Guessing Problem



(b) Distributed Guessing Problem

Figure 3.2.: Guessing problems in the (a) standard and (b) distributed scenarios. In the standard scenario, an observer has to guess the value of a parameter $\widetilde{a}$ given only some function of it, $x = f(\widetilde{a})$. In the distributed scenario, the input parameter is a vector of n symbols, $(\widetilde{a}_1, \ldots, \widetilde{a}_n)$ and so is the data given to the players $(x_1, \ldots, x_n)$. Each of player has access to just one of the symbols x_j and has to guess the corresponding initial parameter $\widetilde{a}_j$. The game is won when all players guess correctly.

the one defined by a constant function f, which trivially erases any information about the input. An example of a difficult function in the $(n, 2, 2)$ scenario for odd n and classically correlated players is $f(a_1, \ldots, a_n) = (a_n, a_1, \ldots, a_{n-1})$ defined on the set S of inputs satisfying $a_1 \oplus \ldots \oplus a_n = 0$. This is the Guess-Your-Neighbour's-Input task considered in (ABB$^+$10).

Next, we prove that a DGP is maximally difficult for players sharing classical correlations (classical players) if, and only if, it corresponds to an LO inequality. Hence, in order to win the game with a probability larger than $1/|S|$, they need to share correlations violating LO. In particular, quantum correlations provide no advantage over the trivial strategy of randomly guessing the solution since they satisfy LO (see section 3.5.2).

Our goal then is to prove that imposing that correlations do not provide any advantage for DGP involving maximally difficult functions f is equivalent to LO. For such an f and any correlations $P(a_1 \ldots a_n | x_1 \ldots x_n)$, providing no advantage for the DGP defined by f means that

$$\frac{1}{|S|} \sum_{(a_1, \ldots, a_n) \in S} P\left(a_1 \ldots a_n | f_1(a_1, \ldots, a_n) \ldots f_n(a_1, \ldots, a_n)\right) \leq \frac{1}{|S|}, \tag{3.2}$$

where $f_1, \ldots, f_n$ refer to the components of the vector f, and $x_j = f_j(a_1, \ldots, a_n)$

is the input that party j receives. Note that, for simplicity, and since the goal of the parties is to provide a correct guess of the initial parameters, we slightly abuse notation and replace all $\widetilde{a}_j$ by a_j. In order to prove the correspondence, we now show that functions f that are maximally difficult for classical players are precisely those which have the property that if $f(a_1, \ldots, a_n)$ and $f(a'_1, \ldots, a'_n)$ are both defined (that is, both $(a_1, \ldots, a_n)$ and $(a'_1, \ldots, a'_n)$ belong to S), then there exists some j for which $a_j \neq a'_j$, but $f_j(a_1, \ldots, a_n) = f_j(a'_1, \ldots, a'_n)$. Given that f varies over all those partial functions having this property, the DGP inequalities (3.2) define all LO inequalities in the (n, m, d) scenario.

We first prove the 'only if' direction by contradiction. Assume there exist $a_1, \ldots, a_n$ and $a'_1, \ldots, a'_n$, both on which f is defined, such that for every party j, either $a_j = a'_j$ or $f_j(a_1, \ldots, a_n) \neq f_j(a'_1, \ldots, a'_n)$ holds true. Then the following classical strategy performs better than random guessing: for those j with $a_j = a'_j$, let them output this particular value independently of their input; for those with $a_j \neq a'_j$ and $f_j(a_1, \ldots, a_n) \neq f_j(a'_1, \ldots, a'_n)$, choose some function g_j such that $a_j = g_j(f_j(a_1, \ldots, a_n))$ and $a'_j = g_j(f_j(a'_1, \ldots, a'_n))$, and let them output $g_j(x_j)$. This strategy recovers all correct values both for the $a_1, \ldots, a_n$ as well as for the $a'_1, \ldots, a'_n$ and therefore performs better than random guessing.

Conversely, we need to show that if f has this property, then using local operations only cannot be more successful than random guessing. Thanks to convexity, it is enough to consider deterministic local strategies. If a deterministic strategy is better than random guessing, there needs to exist at least $a_1, \ldots, a_n$ and $a'_1, \ldots, a'_n$ such that the strategy works on both of these. In particular, this means that, for each party j, either $a_j = a'_j$, or party j needs to be able to tell the two cases apart via x_j, so that $f_j(a_1, \ldots, a_n) \neq f_j(a'_1, \ldots, a_n)$. This implies that f cannot have the property described above.

3.3. LO in networks: a hierarchy of sets

In a device independent formalism, the parties operate over devices (boxes) by pressing buttons and obtaining outcomes. These devices can moreover be distributed among a larger number of parties, forming a network. As an example, consider a "bipartite" box, i.e. a device that is accessed by two parties (namely one that accepts two inputs) and produces two outputs, one for each party. The statistics of such a device is given by the conditional probability distribution $P(ab|xy)$. In addition, two copies of this box could be distributed among four parties as shown in Fig. 3.3. These two bipartite devices may now be used by the four parties to produce four outputs when provided by four inputs, and hence the conditional probability distribu-

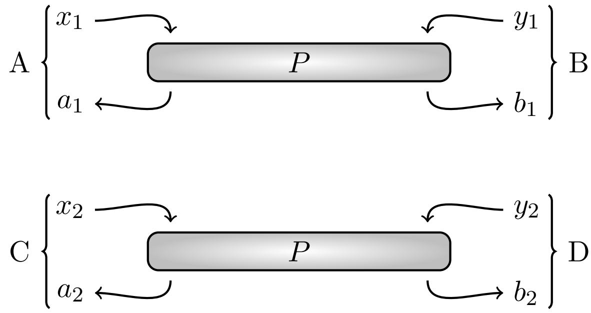


Figure 3.3.: Two copies of a bipartite device characterized by the conditional probability distribution P distributed among four parties: Alice, Bob, Charlie and Dani.

tion $P'(a_1b_1a_2b_2|x_1y_1x_2y_2) = P(a_1b_1|x_1y_1)\,P(a_2b_2|x_2y_2)$ on scenario $(4,2,2)$ is studied. It is natural to assume that if a bipartite device characterized by P exists in nature, so should a fourpartite device constructed in this way characterized by P'. In this section I will formalize the problem of how to restrict the possible conditional probability distributions P by imposing LO constraints on P' in the larger scenario.

First, we define the sets of correlations that obey LO for a certain number of copies in the following sense. A given conditional probability distribution for the (n,m,d) scenario can be thought of as provided by some device shared between the n parties each having access to one input and output of the device. If the correlations provided by the device are compatible with LO, a natural question is whether a larger conditional probability distribution coming from several copies of such a device distributed among more parties necessarily satisfies LO. As we will show in section 3.5.3, the answer to this question is negative. That is, LO displays activation effects and, hence, we have a hierarchy of sets.

The largest set in this hierarchy, denoted $\mathcal{LO}^1$, is the set of all correlations in the (n,m,d) scenario which obey the LO inequalities for this scenario. We have referred to this as "LO set" in the previous section, but when emphasizing its place in the hierarchy we will refer to it as LO^1.

Now consider k copies of a device characterized by a conditional probability distribution P, distributed among kn parties, each of which has access to one input of only one device. See for instance Fig. (3.5), where $2k$ parties share k copies of the bipartite box PR. If the conditional probability distribution P^k of the kn-partite global device obeys all the LO inequalities for the scenario

(kn, m, d), we say that P satisfies LO^k, and belongs to $\mathcal{LO}^k$. These sets satisfy a chain of inclusions:

$$\mathcal{LO}^\infty \subset \ldots \subset \mathcal{LO}^k \subset \mathcal{LO}^{k-1} \subset \ldots \subset \mathcal{LO}^1$$

We denote by $\mathcal{LO}^\infty$ the set of correlations for (n, m, d) which obey the LO inequalities for any number of copies.

It is worth mentioning that when defining this hierarchy of LO sets, independent copies of the same device are considered, where no subset of boxes are combined to produce outputs. This latter case will be partially addressed at the end of the chapter.

3.4. LO and Graph Theory

Having stated the LO principle, our main goal is the study of the sets $\mathcal{LO}^k$ of LO correlations. As we shall see, the LO principle turns out to be very powerful for ruling out non-quantum correlations. As in the case of contextuality (CSW10), graph theory is perfectly suited for our purposes. We consider the $(md)^n$ possible events in the (n, m, d) scenario and map them onto a graph with $(md)^n$ vertices. The edges of the graph arise from the orthogonality relations among the corresponding events, in the following way.

First, we define the *orthogonality graph* of scenario (n, m, d), which we denote by G. It consists of $(md)^n$ vertices, where two vertices are connected by an edge if and only if the corresponding events are locally orthogonal. For instance, Fig. 3.4 shows the orthogonality graph of the $(2, 2, 2)$ scenario. In graph theory, a *clique* in a graph $G = (V, E)$ is a subset of vertices $C \subseteq V(G)$ such that the subgraph induced by C is complete, i.e. such that all pairs of vertices in C are connected by an edge in G. A clique is *maximal* if it cannot be extended to another clique by including a new vertex. Clearly, any clique in the orthogonality graph of events gives rise to an LO inequality (and vice versa), as all events in the clique are connected and, thus, are pairwise orthogonal. Therefore, the problem "find all the optimal LO inequalities" is equivalent to "find all maximal cliques of the associated orthogonality graph". While the problem of finding all maximal cliques of a graph is known to be NP-hard (Kar72), there exist software packages (Uno05; NO10) that provide the solution for small graphs. We have used these packages to derive and partly classify LO inequalities for various Bell scenarios (see section 3.5.1 and appendix C). Note however that in principle, while the problem of finding the maximal cliques is NP-hard for general graphs, this may no longer be the case for graphs associated to correlations among distant parties. Indeed, these graphs may represent a subset of all possible graphs that does not include the hard instances of the problem.

Consider now a Bell scenario (kn, m, d), motivated by the hierarchy of LO sets defined in section 3.3. A natural question is how to relate the orthogonality graph of (kn, m, d) (G_k) with that of (n, m, d) (G_1). The answer is simple, and is given by the conormal product of graphs, formally defined in section 2.3, denoted by " $\cdot$ ": given $G_1 = (V_1, E_1)$, the orthogonality graph G_k is the conormal product of G_1 with itself k times, $G_k = G_1^{\cdot k}$. The intuition behind the conormal product is the following: given two vertices (u_1, u_2) and (v_1, v_2) of the product graph, they share and edge if either u_1 and v_1 are connected in G_1 or u_2 and v_2 are connected in G_2. This is consistent with our notion of orthogonality, where two events are orthogonal if at least for one local part of the event the corresponding situations are orthogonal.

Another graph of interest is the complement of the orthogonality graph, which we call *non-orthogonality graph* of Bell Scenario (n, m, d) and denote by $\mathrm{NO}(B_{n,m,d})$[1]. The preferred use of $\mathrm{NO}(B_{n,m,d})$ over G began with the study of LO correlations in contextuality scenarios (see chapter 4). The non-orthogonality graph is thus defined as the graph $\mathrm{NO}(B_{n,m,d}) = (V, E)$, where the vertices V correspond to the $(md)^n$ events of the Bell Scenario, and two vertices share and edge if the corresponding events are not orthogonal. In graph theory, an *independent set* of a graph is a subset I of vertices such that none of them are connected by and edge, i.e. the sugbraph induced by I is empty. An independent set is *maximal* if it cannot be extended to another independent set by including a new vertex. Similar to the case of the orthogonality graph, any independent set in the non-orthogonality graph defines an LO inequality, and vice-versa. Hence, the problem "find all the optimal LO inequalities" is equivalent to "find all maximal independent sets of the associated non-orthogonality graph". While the problem of finding maximal independent sets in $\mathrm{NO}(B_{n,m,d})$ is as difficult as finding maximal cliques in G, the advantage in the use of $\mathrm{NO}(B_{n,m,d})$ appears in the relation of its graph invariants with the hierarchy of LO sets for contextuality scenarios, as explained in sections 4.6 4.8.

We now ask how to relate the non-orthogonality graph of (kn, m, d) with that of (n, m, d). The answer is given by the strong product of graphs defined in section 2.3, denoted by " $\boxtimes$ ": given $\mathrm{NO}(B_{n,m,d})$, the non-orthogonality graph $\mathrm{NO}(B_{kn,m,d})$ is the strong product of $\mathrm{NO}(B_{n,m,d})$ with itself k times, $\mathrm{NO}(B_{kn,m,d}) = \mathrm{NO}(B_{n,m,d})^{\boxtimes k}$. The intuition behind the strong product is the following: two vertices (u_1, u_2) and (v_1, v_2) of the product graph share and edge if both u_1 and v_1 are connected in G_1 and u_2 and v_2 are connected in G_2, which

[1]This in principle complicated notation proves useful in chapter 4

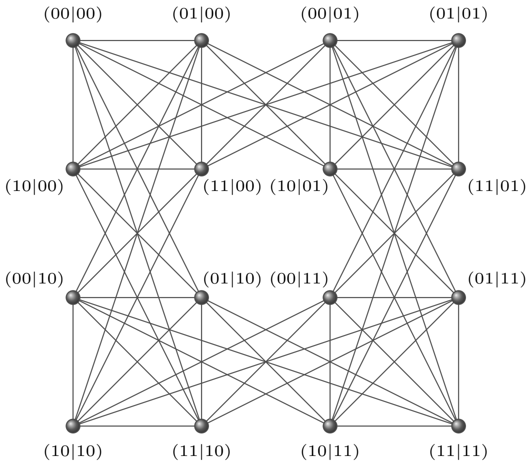


Figure 3.4.: Orthogonality graph of the $(2, 2, 2)$ scenario. As mentioned in the text, each possible event corresponds to a node, while the edges connect locally orthogonal events.

is also similar to our notion of orthogonality.

In this work, we will relate graph invariants of the orthogonality and non-orthogonality graphs to some problems regarding correlations, and we refer the reader to section 2.3 and appendix A for their formal definitions.

3.5. LO and Correlations

In this section, I study how the LO principle imposes constraints over correlations. I first investigate the connection between LO and the NS principle. Then, I show that the LO principle is naturally satisfied by quantum correlations, and hence classical correlations as well. I then investigate the use of LO as a tool to detect post-quantum no-signaling correlations. Clearly, those correlations violating GYNI are in contradiction with LO as well. However, the situation turns out to be much richer already for two parties. In principle, one might think that LO would be useless for the detection of supra-quantum bipartite correlations because of the equivalence of LO^1 with NS. However, this intuition is not correct. I show that postulating LO also on the many-party level leads to detection of non-quantumness of bipartite correlations. The idea is similar in spirit to the network approach to non-locality presented in Ref. (CASA11).

3.5.1. No-signaling correlations

The first question we ask when characterizing LO correlations is how they relate to the set of no-signaling correlations.

For the case of bipartite scenarios, the answer is surprising since the No Signaling principle defines the same set of correlations as LO. This was first noticed by Cabello, Severini and Winter (CSW10), and an alternative proof in given in appendix B.

However, this equivalence between LO^1 and NS breaks down when moving on to the multipartite scenario. In fact, consider the scenario $(3, 2, 2)$. Exploiting the graph-theoretical approach presented in section 3.4, we derive all the LO inequalities for this scenario. These inequalities are then organized into equivalence classes under the symmetries of relabelling of inputs/outputs, permuting the parties and imposing no-signaling constrains (see appendix C). We find one (and only one) class of non-trivial LO inequalities, where non-trivial means that the inequalities are violated by some NS correlations. This inequality turns out to be the GYNI inequality (ABB^+10), which in the tripartite case reads $P(000|000) + P(110|011) + P(011|101) + P(101|110) \leq 1$. It is easy to see by simple inspection that GYNI is an LO inequality. As shown in (ABB^+10), the maximum of the GYNI inequality over $\mathcal{NS}$ is equal to 4/3, which proves the existence of NS correlations violating LO^1.

Our numerical data suggest that the gap between $\mathcal{LO}^1$ and $\mathcal{NS}$ increases with the number of parties: in the $(4, 2, 2)$ scenario, we find already 35 equivalence classes, which are presented in appendix C. Unfortunately, for more parties ($n > 4$), even the simplest scenario $(n, 2, 2)$ becomes computationally intractable due to the large size of the orthogonality graph. Nevertheless, examples of such inequalities for larger n as well as m and d are known and can be constructed from unextendible product bases (BDM^+99) by using the method discussed in (SFA^+13).

3.5.2. Quantum correlations

Quantum correlations, presented in section 2.1.3, arise via Born's rule

$$P(a_1 \dots a_n | x_1 \dots x_n) = tr\left(\rho\, \Pi_{a_1}^{x_1} \otimes \cdots \otimes \Pi_{a_n}^{x_n}\right),$$

where $\Pi_{a_j}^{x_j}$ are the projectors associated to the measurements outputs. In this section we move on to prove that LO inequalities are satisfied by the set of quantum correlations, that is, $\mathcal{Q} \subseteq \mathcal{LO}^\infty$. First, we analyze an inequality involving two exclusive events, where the statement is straightforward, and then generalize the proof to an arbitrary LO inequality, which also means any number of copies.

Consider two locally orthogonal events $e_1 = (a_1^1 \dots a_n^1 | x_1^1 \dots x_n^1)$ and $e_2 = (a_1^2 \dots a_n^2 | x_1^2 \dots x_n^2)$ with $a_i^1 \neq a_i^2$ while $x_i^1 = x_i^2$ for some party i, and the corresponding inequality $p\,(e_1) + p\,(e_2) \leq 1$. The maximization of the sum of these two probabilities over quantum correlations reads

$$\max_{|\Psi\rangle, \{\Pi^{x_j^1}_{a_j^1}\}} \langle\Psi| \, (\Pi^{x_1^1}_{a_1^1} \otimes \dots \otimes \Pi^{x_i^1}_{a_i^1} \otimes \dots \otimes \Pi^{x_n^1}_{a_n^1} + \Pi^{x_1^2}_{a_1^2} \otimes \dots \otimes \Pi^{x_i}_{a_i^2} \otimes \dots \otimes \Pi^{x_n^2}_{a_n^2}) \, |\Psi\rangle \, , \quad (3.3)$$

where the maximization runs over all possible states $|\Psi\rangle$ and projectors $\{\Pi^{x_j^i}_{a_j^i}\}$ acting on an arbitrary Hilbert space. Indeed, the maximization should act on

$$\mathrm{tr}\left(\rho(\Pi^{x_1^1}_{a_1^1} \otimes \dots \otimes \Pi^{x_i^1}_{a_i^1} \otimes \dots \otimes \Pi^{x_n^1}_{a_n^1} + \Pi^{x_1^2}_{a_1^2} \otimes \dots \otimes \Pi^{x_i}_{a_i^2} \otimes \dots \otimes \Pi^{x_n^2}_{a_n^2})\right),$$

but since the Hilbert space is of arbitrary dimension, we may consider ρ to be a pure state $|\Psi\rangle\langle\Psi|$. The term in the parenthesis of eq. (3.3) is equal to the sum of two orthogonal projectors, since $\Pi^{x_i}_{a_i^1}\Pi^{x_i}_{a_i^2} = 0$. Thus, this sum is upper bounded by the identity operator, and the LO inequality follows.

Now we move on to $k > 2$ exclusive events. By a similar argument, we want to maximize the quantity:

$$\max_{|\Psi\rangle, \{\Pi^{x_l}_{a_l}\}} \langle\Psi| \left(\sum_{j=1}^{k} \Pi^{x_1^j}_{a_1^j} \otimes \dots \otimes \Pi^{x_i^j}_{a_i^j} \otimes \dots \otimes \Pi^{x_n^j}_{a_n^j} \right) |\Psi\rangle$$
$$= \max_{|\Psi\rangle, \{\Pi^{x_l}_{a_l}\}} \langle\Psi| \left(\sum_{j=1}^{n} \Pi_j \right) |\Psi\rangle \, , \quad (3.4)$$

where $\Pi_j = \Pi^{x_1^j}_{a_1^j} \otimes \dots \otimes \Pi^{x_i^j}_{a_i^j} \otimes \dots \otimes \Pi^{x_n^j}_{a_n^j}$ is the projector associated to event e_j. Since the events in the LO inequality are pairwise orthogonal, the projectors Π_j that appear in the sum are pairwise orthogonal as well. Hence $\sum_{j=1}^{k} \Pi_j \leq \mathbb{1}$, and the LO inequality follows.

Note that the simplicity of the above argument may mislead the strength of the constraints imposed by LO, since it is a non-trivial property of quantum mechanics that pairwise orthogonality (of projectors) implies orthogonality of all the projectors, i.e. that pairwise orthogonal events are naturally jointly exclusive.

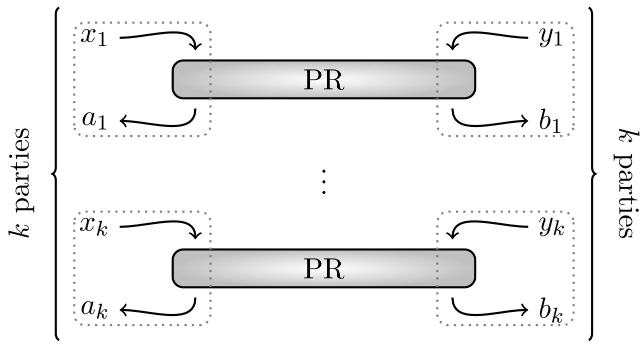


Figure 3.5.: k copies of a PR-box shared among $2k$ parties. Each party has access to one part of a box.

3.5.3. Supra-quantum bipartite correlations

We now move on to show that postulating LO also on the many-party level leads to detection of non-quantumness of bipartite correlations. Given some bipartite correlations, as explained in section 3.3, the main idea consists in distributing k copies of these among $2k$ parties, such that one party has access to one part (input and output) of only one bipartite box. In the resulting $2k$-partite scenario, the LO principle is stronger than the NS principle. Thus, it may happen that the initial bipartite correlations violate LO when distributed among different parties in a network. In what follows, we will show that the supra quantum devices called PR boxes violate LO^2, and hence are ruled out by the Local Orthogonality principle.

A PR-box is a hypothetical device taking binary inputs and giving binary outputs which obey $PR(ab|xy) = 1/2$ if $a \oplus b = xy$ and $PR(ab|xy) = 0$ otherwise (PR94). These boxes are known to be more non-local than what quantum theory allows. For instance, they provide a violation of the Clauser-Horne-Shimony-Holt Bell inequality (CHSH69) larger than Tsirelson's bound for quantum correlations (Tsi80). PR-boxes are bipartite no-signaling devices, and therefore might naïvely be expected to satisfy LO due to the equivalence between LO^1 and NS. However, we now prove that when distributed in networks they violate LO. Consider k copies of a PR-box, distributed among $2k$ parties as shown in Fig. 3.5. The conditional probability distribution is:

$$P(a_1 b_1 \cdots a_k b_k | x_1 y_1 \cdots x_k y_k) = \prod_{j=1}^{k} PR(a_j b_j | x_j y_j)\,, \tag{3.5}$$

where j labels the k boxes. Already for $k = 2$, we find LO inequalities violated by these two copies of the PR-box, and hence PR boxes do not satisfy LO[2]. One example of such an inequality is $P(0000|0000) + P(1110|0011) + P(0011|0110) + P(1101|1011) + P(0111|1101) \leq 1$. For a PR-box, the left-hand side is equal to 5/4. One way to search for such a violation (or LO inequality) is the following. We first classify events as either "possible" or "not possible" for this many-copy box: an event $(ab|xy)$ for one PR-box is possible if $a \oplus b = xy$. Hence, PR-box correlations may be written as

$$PR(ab|xy) = \begin{cases} \frac{1}{2} & \text{if the event is possible,} \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

In the case of k boxes, an event $(a_1 b_1 \cdots a_k b_k | x_1 y_1 \cdots x_k y_k)$ is possible iff $a_j \oplus b_j = x_j y_j$ for all $j \in \{1, \ldots, k\}$. Then, the general form for the k-box probability (3.5) is

$$P(a_1 b_1 \cdots a_k b_k | x_1 y_1 \cdots x_k y_k) = \begin{cases} 2^{-k} & \text{if the event is possible,} \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

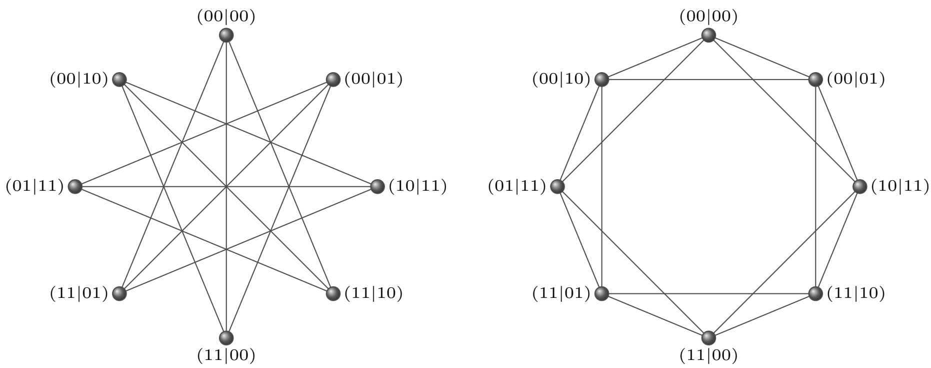
Consider a clique $C \subseteq V$ in the orthogonality graph $G = (V, E)$ of scenario $(2k, 2, 2)$ and the corresponding LO inequality $LO(C)$[2]. Define the set $C_p \subseteq C$ to be the subset of possible events in C. Then, the multipartite box (3.7) violates $LO(C)$ if, and only if, it violates $LO(C_p)$. In particular, in order to exclude the PR-box, it is sufficient to find a clique of size larger than 2^k in the orthogonality graph $G_{\text{poss}} = (V_p, E_p)$ of possible events for box (3.7). This problem becomes significantly easier, since $|V_p| = 8^k$, compared to $|V| = 16^k$ for the initial graph (compare figs. 3.4 and 3.6(a)). Already for $k = 2$, there exist cliques of size larger that 4. We found that all of them have size 5, and one example is given by:

$$\{(0000|0000),\ (1110|0011),\ (0011|0110),\ (1101|1011),\ (0111|1101)\},$$

which corresponds to the LO inequality previously mentioned.

We can also analyze noisy versions of the PR-box given by $P_q = q\,PR + (1-q)P_{\mathbb{I}}$, where $P_{\mathbb{I}}(ab|xy) = 1/4$ for all a, b, x, y. Since we search for violations of LO, we focus first on the two copy case. Now all events become possible and one should consider the full list of LO inequalities for $(4, 2, 2)$. However, the five-term inequalities found for the noiseless case are non-maximal cliques

[2]We denote by $LO(C)$ the LO inequality that arises from the LO principle by the set of exclusive events C imposing condition (3.1).



(a) Orthogonality Graph (b) Non-Orthogonality Graph

Figure 3.6.: Possible events for the PR box. (a) 4-antiprism graph corresponding to the orthogonality graph of possible events for a single PR-box. It coincides with Fig. 2 in (SBBC11), where the authors study the CHSH inequality. (b) Circulant graph $Ci_4(1,2)$ corresponding to the non-orthogonality graph of possible events for a PR-box. It is the complement of the graph in (a).

in the orthogonality graph of the $(4,2,2)$ scenario, which can be completed to maximal cliques. By doing so, we obtain inequalities with additional terms corresponding to events that are impossible for the PR-box, but which are possible for a noisy PR-box. We have found that the conditional probability distribution

$$P(a_1b_1a_2b_2|x_1y_1x_2y_2) = \left(q\cdot \mathrm{PR}(a_1b_1|x_1y_1) + \tfrac{(1-q)}{4}\right)\cdot\left(q\cdot \mathrm{PR}(a_2b_2|x_2y_2) + \tfrac{(1-q)}{4}\right) \tag{3.8}$$

violates LO down to $q \approx 0.72$, which is close to Tsirelson's bound $q = 1/\sqrt{2} \approx 0.707$ (meaning that noisy boxes with $q \leq 1/\sqrt{2}$ can be simulated with quantum states and measurements). An example of such an LO inequality is given by the following set of ten LO events:

$$\{(1111|0000),(1100|1010),(0100|1100),(0011|0001),(0010|0111),(1011|0000),$$
$$(0101|1100),(1101|1100),(1010|0110),(1001|0100)\}.$$

It appears plausible to conjecture that the generalization of the previous approach to an arbitrary number of parties converges to Tsirelson's bound $q =$

$1/\sqrt{2} \approx 0.707$ in the limit of an infinite number of parties, although we did not yet find any proof. A study of noisy PR-boxes for more than two copies is presented in appendix D, together with a discussion for more general boxes.

An immediate consequence of these results is that LO indeed rules out all extremal boxes in the $(2,2,d)$ and the $(2,m,2)$ scenarios. For the first case, it follows from the fact that using such d-outcome extremal boxes, one can always simulate correlations of the PR-box arbitrarily well (BLM[+]05). More precisely, in the case of even d, a single copy of such a box is enough (one only needs to relabel the outcomes of the box), and in the case of odd d, by using enough copies of a d-outcome extremal box it is possible to simulate the PR-box with arbitrarily small error. For the second case, we use the characterization given in (JM05) (see also (BP05)). There, any extremal bipartite binary box is a PR box when restricting the measurement settings for each party only to the first two out of the m possible choices. Hence, any LO inequality violated by a PR box, is also violated by these extremal bipartite binary boxes.

As a final remark, it is also interesting to compare LO with Information Causality (IC) (PPK[+]09), another proposal for a physical principle to single out quantum correlations. A natural question then is if and when LO can do better than IC in ruling out supra-quantum correlations. Following Allcock *et al.* (ABPS09), we study LO predictions for different families of no-signaling correlations in the $(2,2,2)$ scenario, by mixing PR-box correlations with different types of noise. We search for violations of the 10-term LO inequalities for scenario $(4,2,2)$ by two copies of the bipartite box defined by $P(ab|xy) = \xi PR(ab|xy) + \gamma P_L(ab|xy) + (1-\xi-\gamma)P_{\mathbb{I}}$, where $P_L(ab|xy) = \delta_{a,0}\delta_{b,0}$. In some situations LO provides a bound to the set of quantum correlations which is tighter than the known bounds obtained from applications of IC, as can be seen in Fig. 3.7. Hence, LO rules out correlations that were not excluded before by IC. However, along the isotropic line ($\xi = 1$) IC recovers Tsirelson's bound, a feat which LO cannot achieve at the two-copy level.

3.5.4. Supra-quantum tripartite correlations

In the previous sections, we prove that not only all extremal bipartite correlations in the $(2,2,d)$ and $(2,m,2)$ scenarios but also other bipartite supra-quantum correlations are ruled out by LO. Nevertheless, due to its intrinsically multipartite formulation, we expect the LO principle to be of particular relevance for the study of genuine multipartite correlations among more than two parties. In what follows, we study the case of three parties having access to two binary measurements.

We focus on the question of how well LO performs for extremal no-signaling

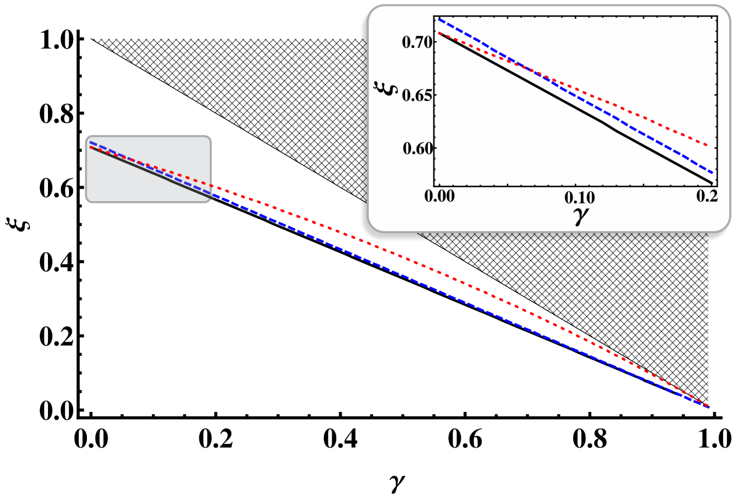


Figure 3.7.: Comparison of IC and LO^2 for detecting supra-quantum correlations in the (2,2,2) scenario. We consider the family of correlations parametrized as $P(ab|xy) = \xi PR(ab|xy) + \gamma P_L(ab|xy) + (1 - \xi - \gamma)P_{\mathbb{I}}$, where $P_L(ab|xy) = \delta_{a,0}\delta_{b,0}$. The curves show the bounds provided by the 1+AB level of the NPA hierarchy (NPA07; NPA08) (black), LO^2 (blue dashed), IC (red dotted), and the edge of the crossed-out region corresponds to NS correlations and bounds the allowed parameter space. Note that when $\gamma \to 0$ (see inset), IC approximates the quantum set better than LO^2, which is consistent with the fact that IC recovers Tsirelson's bound for $\gamma = 0$, while LO reaches ≈ 0.72 for two copies of the device. However, LO beats the known IC bound for other parameter values, ruling out correlations that were not excluded before by applications of IC.

nonlocal correlations. All these extremal NS boxes for the $(3,2,2)$ scenario were computed in Ref. (PBS11) and any NS correlations in this scenario can be obtained by mixing them. These extremal boxes can be grouped into 46 equivalence classes under symmetries, the first class corresponding to deterministic local points, while the other 45 are non-local. The latter can be interpreted as the maximally non-local correlations in the $(3,2,2)$ scenario compatible with the NS principle. We will prove that LO rules out **all** these nonlocal extremal boxes.

In the spirit of the previous section, consider the orthogonality graph of possible events G_{poss} for each of the 45 boxes. The conditional probability distribution of some of them satisfy $P_j(e) = c_j$ if e is a possible event for box j (and 0 otherwise), where the constant c_j depends only on the box and not on the event. Hence, we search for cliques of size larger than c_j^{-1} on $G_{\text{poss},j}$. In the cases where no such clique is found, we move on to two copies of the corresponding boxes in the $(6,2,2)$ scenario. There, we search for cliques of size larger than c_j^{-2} on $G_{\text{poss},j}^{\cdot 2}$.

However, some boxes within the 45 require a slightly different approach. These are the ones whose conditional probability distribution does not assign the same value to all the vertices of the orthogonality graph of possible events. In this case, we construct the weighted orthogonality graph of possible events $G_{\text{poss},j}(w)$, where each vertex v has a weight given by the probability of the corresponding event $w(v) = P_j(e_v)$. The weight of a clique is given by the sum of the individual weights over all the vertices of the clique. Hence, we search for cliques of weight larger than one, which guarantees the violation of LO by the given box j. In the cases where no such clique is found, we move on to two copies of the corresponding boxes in the $(6,2,2)$ scenario. Now, the weight of vertex $(v_1, v_2) \in V(G_{\text{poss},j}^{\cdot 2})$ in the two-copy orthogonality graph is given by $w(v_1, v_2) = P_j(e_{v_1})P_j(e_{v_2})$. Similarly, we search for cliques of weight larger than one.

We find that every maximally nonlocal box in $(3,2,2)$ violates either LO^1 or LO^2, and hence cannot have a quantum realization. Table 3.1 displays the results for the 45 nonlocal boxes in the numbering scheme of (PBS11). The columns show the number of copies needed for finding the violation, the number of terms in the violated inequality and the value given by the box for that inequality (which is always larger than 1, thereby proving violation).

We remark that this is an example of how the intrinsically multipartite formulation of LO allows detecting correlations for which any bipartite principle fails. Box number 4 in (PBS11) is a tripartite no-signaling box which cannot be ruled out by any bipartite principle (YCA[+]12) but which violates LO.

box	copies	terms	value	box	copies	terms	value
2	2	5	5/4	25	1	4	4/3
3	2	17	17/16	26	2	22	37/36
4	2	17	17/16	27	2	17	17/16
5	1	4	5/4	28	1	4	7/6
6	2	17	17/16	29	1	4	4/3
7	2	17	17/16	30	2	14	26/25
8	2	9	9/8	31	2	13	26/25
9	1	4	9/8	32	1	4	6/5
10	1	4	9/8	33	1	4	6/5
11	1	4	5/4	34	2	13	37/36
12	2	17	9/8	35	1	4	7/6
13	1	4	7/6	36	2	13	66/64
14	1	4	7/6	37	1	4	9/8
15	1	4	9/8	38	1	4	7/6
16	1	4	7/6	39	2	22	37/36
17	1	4	7/6	40	2	17	17/16
18	1	4	9/8	41	2	17	17/16
19	2	17	17/16	42	2	17	17/16
20	1	4	6/5	43	2	14	50/49
21	2	17	17/16	44	2	17	17/16
22	2	19	37/36	45	2	17	17/16
23	1	4	7/6	46	2	17	17/16
24	1	4	7/6				

Table 3.1.: Table of LO violations for the extremal tripartite NS boxes in the numbering scheme of (PBS11). The columns show the number of copies needed for finding the violation, the number of terms in the violated inequality and the value given by the box for that inequality. Note that since it is always larger than 1, they violate either LO^1 or LO^2, and hence LO.

3.6. LO and wirings

Wirings (ABL+09) are operations that can be applied to one or more no-signaling boxes in order to produce a new no-signaling box. For example, a wiring may consist of party 1 communicating his/her outcome a_1 to party 2, who uses this outcome as his/her measurement setting and obtains an outcome a_2. This defines a new no-signaling box upon identifying parties 1 and 2 with a new joint party with measurement setting x_1 and joint outcome $g(a_1, a_2)$, where g can be any function of a_1 and a_2.

In this section, we first define wirings. Then we show that if a no-signaling box P satisfies LO^∞, then so does any other box which can be obtained from P by applying wirings to copies of P distributed among many parties. In fact, we show that if M copies of a box violate LO^1 when wired together, then the same M copies violate LO^1 as independent copies without any wiring, meaning that the original single-copy box violates LO^M. Hence, in order to find violations of LO^∞, it is enough to consider only distributed copies of P. In this sense, wirings are useless for detecting violations of LO^∞. Since our wirings are more general than those of (ABL+09), this implies also that $\mathcal{LO}^\infty$ is a set of NS-boxes which is closed in the sense of (ABL+09).

We start by considering a restricted, but especially instructive, class of wirings and will generalize later. We call these wirings *static* since the "wires" are fixed throughout the protocol rather than dynamically determined. A static wiring protocol for an n-partite box P is obtained as follows: first, we distribute M copies of P among Mn parties, where M is arbitrary. This results in the box $P^{\otimes M}$. These Mn parties may now assemble into s groups; these groups will become the parties of the "wired" box

$$P_{\mathrm{wired}}(b_1 \dots b_s | y_1 \dots y_s),$$

which they are about to construct. Here, y_i denotes the input which all parties in group i receive *jointly*, while b_i stands for the joint output that they are going to obtain/produce. We may intuitively think of the parties composing a group as physically meeting up at the same location, where each party brings with them his/her part of the box to which they have access; see Figure 3.8 for an example situation. In this way, a particular group of l parties has access to l input-output devices.

Second, there is a subprotocol for each group which specifies how the parties within that group communicate and coordinate the use of their boxes. For the sake of concreteness, let us assume that the first group is formed by parties $1, \dots, l$. For these l parties, we now need to specify an ordering among them, corresponding to the temporal order in which the parties use their devices. For

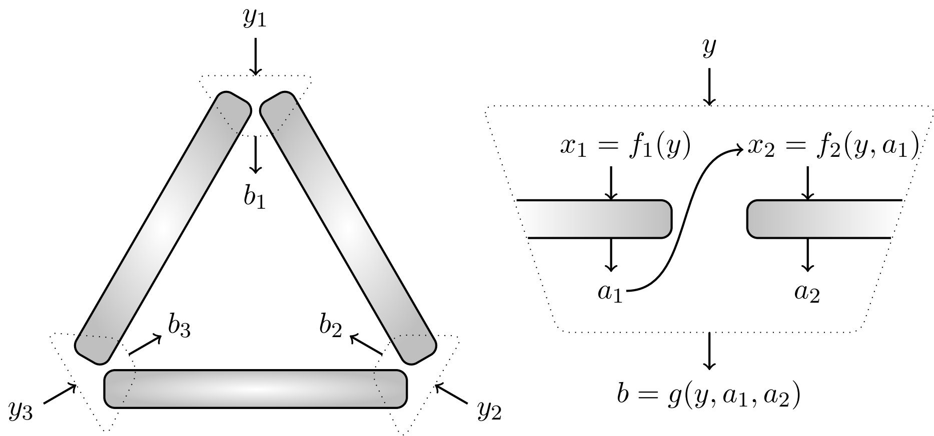


Figure 3.8.: Three copies of a bipartite box wired between three groups of two parties each. The resulting wired box is 3-partite, where each new party has access to one end of two boxes. To provide an example, consider a party receives input y, and chooses to use first the box on his right hand side. He inputs $f_1(y)$ to this box and obtains outcome a_1. Then, he uses the box in the left hand side. He inputs $f_2(y, a_1)$ and obtains an outcome a_2. Finally, he outputs $b = g(y, a_1, a_2)$ as a result of the protocol.

notational convenience, we relabel the parties in the group such that this total ordering is precisely given by the enumeration $1, \ldots, l$. Party 1 starts by choosing a measurement setting $f_1(y_1)$ on his/her device, where f_1 is any function which is part of the specification of the protocol, and obtains an outcome a_1. Afterwards, party 2 continues by inputting $f_2(y_1, a_1)$ into his/her device where f_2 is likewise some fixed function, and gets an outcome a_2. In general, party $j+1$ operates after party j and uses the setting $f_{j+1}(y_1, a_1, \ldots, a_j)$ on his/her device getting an outcome a_{j+1}. After all parties have used their devices in this way, the group announces the total outcome $g(y_1, a_1, \ldots, a_l)$, where g is again a fixed function. In general, such a subprotocol exists for the parties within each group separately.

A trivial example of such a static wiring protocol consists in taking $M = 1$, having each party form their own group, and letting each such party apply certain functions to their inputs and outputs. This type of wiring corresponds to the special case in which P_{wired} is obtained from P via deterministic *local operations*.

We now proceed to show that if P satisfies LO^∞, then so does any P_{wired} obtained from such a P via static wirings. In other words, the set $\mathcal{LO}^\infty$ is *closed under wirings*.

First of all, it is sufficient to show that the wired box satisfies LO^1, for the following reason: if a box P_{wired} can be constructed from P via wirings, then so can any of its powers $P_{\text{wired}}^{\otimes k}$. Indeed, if P_{wired} can be constructed from M copies of P via wirings, then so can any $P_{\text{wired}}^{\otimes k}$, even if the number of copies of P increases k-fold. We will show that all boxes constructed from P via wirings satisfy LO^1, which in particular applies to all the $P_{\text{wired}}^{\otimes k}$. Therefore, $P_{\text{wired}}^{\otimes k} \in \mathcal{LO}^1$ for all k, which means by definition that $P_{\text{wired}} \in \mathcal{LO}^\infty$.

In order to show that P_{wired} satisfies LO^1, it is enough to consider the case where only the first l parties form a non-trivial group applying a non-trivial wiring. The reason is that the same argument can be applied to the resulting box, and another non-trivial group can be formed, and this argument can be repeated until all the desired groups have formed. Assuming this and using the notation from above, the resulting wired box is r-partite with $r = Mn - l + 1$. We enumerate the parties in such a way that the non-trivial group contains the parties $1, \ldots, l$. We make this whole assumption to keep things conceptually simple and not to clutter our notation.

As explained above, the conditional probability distribution P_{wired} of such a wiring of M boxes has the form

$$P_{\text{wired}}(b_1 \dots b_r | y_1 \dots y_r) \\ = \sum_{\substack{a_1,\dots a_l \\ \text{s.t. } g(a_1,\dots,a_l)=b_1}} P^{\otimes M}(a_1 \dots a_l \, b_2 \dots b_r | f_1(y_1) \dots f_l(y_1, a_1, \dots, a_l) \, y_2 \dots y_r). \tag{3.9}$$

For fixed b_1 and y_1, all events occurring in this sum are orthogonal: for any two different terms in the sum represented by indices $a_1, \dots, a_l$ and $a'_1, \dots, a'_l$, respectively, let i be the smallest party index for which $a_i \neq a'_i$. Then the two settings of party i are both equal to $f(y_1, a_1, \dots, a_{i-1})$, while the outcomes are different. This witnesses orthogonality.

Now consider a given set of mutually orthogonal events $(b_1 \dots b_r | y_1 \dots y_r)$ which represents an LO inequality for P_{wired}. We claim that upon substituting eq. (3.9) into this inequality, we obtain an LO inequality for $P^{\otimes M}$. Checking this means that we need to consider a pair of events that may occur in such an inequality, say

$$\begin{aligned} &(a_1 \dots a_l b_2 \dots b_r | f_1(y_1) f_2(y_1, a_1) \dots f_l(y_1, a_1, \dots, a_l) \, y_2 \dots y_r), \\ &(a'_1 \dots a'_l b'_2 \dots b'_r | f_1(y'_1) f_2(y'_1, a'_1) \dots f_l(y'_1, a'_1, \dots, a'_l) \, y'_2 \dots y'_r). \end{aligned} \tag{3.10}$$

The case where $b_j = b'_j$ and $y_j = y'_j$ for all j, i.e. where the two events occur in the same sum (3.9), was already considered above, where we found them to be orthogonal. Otherwise, there exists some j for which $y_j = y'_j$ and $b_j \neq b'_j$, due to the assumption that the original inequality for P_{wired} is a LO inequality. If $j \in \{2, \dots, r\}$, then the two events (3.10) are clearly orthogonal as well. If $j = 1$, then there has to exist some index i for which $a_i \neq a'_i$; upon considering the smallest i with this property, we again find the events (3.10) to be orthogonal in the same manner as above. Now due to the assumption that $P^{\otimes M}$ satisfies all LO inequalities, we conclude that also P_{wired} satisfies the given LO inequality.

This shows that the set $\mathcal{LO}^{\infty}$ is closed under static wirings. In other words, if a wired box violates LO^{∞}, then so does the original box from which it was constructed.

We now generalize to "dynamic" wirings in which the temporal ordering of the parties within a group is itself determined during the execution of the protocol. Again we take parties $1, \dots, l$ to form the only non-trivial group. After receiving their input y_1, the party which measures his/her box first is given by a function $i_1(y_1)$. This party $i_1(y_1)$ performs the measurement $x_{i_1} = f_1(y_1)$ and obtains an outcome a_{i_1}. This outcome, together with the initial input, determines the second party in the protocol to be the party $i_2(y_1, a_{i_1})$. Similarly, this party then chooses the setting $x_{i_2} = f_2(y_1, a_{i_1})$ and obtains an outcome a_{i_2}. The third party in the protocol then is $i_3(y_1, a_{i_1}, a_{i_2})$, and so on. When all parties have finished, the group announces their joint outcome $g(y_1, a_{i_1}, \ldots, a_{i_l})$.

In the case of such a dynamic wiring, the explicit form of the sum in eq. (3.9) is considerably messier to write down explicitly and we refrain from doing so. Nevertheless, all events occurring in the corresponding sum for fixed b_j and y_j also satisfy the property of being orthogonal. Indeed, consider two events e and e' in this sum. Both events originated by party $i_1(y_1)$ applying a measurement. Now consider the temporally first step t of the wiring protocol at which the protocol realizations of e and e' differ. Since the protocols are deterministic except for the randomness in the boxes, this difference of the realizations must originate from the previous step $t-1$ by one box having produced different outcomes, $a_{i_{t-1}} \neq a'_{i'_{t-1}}$, although the parties were the same, $i'_{t-1} = i_{t-1}$, and the settings were the same, $x_{i_{t-1}} = x'_{i'_{t-1}}$. Hence, the two events e and e' are orthogonal. All other statements which we made for static wirings apply directly to dynamic wirings as well, and $\mathcal{LO}^\infty$ is in particular also closed under dynamic wirings.

In this section, all the wirings that we have considered have been *deterministic*: no randomness is allowed in the protocols in the sense that the functions f_j, i_j and g are required to be deterministic. The case were shared and/or local randomness is provided for the protocol is discussed in (SFA$^+$13), where we conclude that if P satisfies LO$^\infty$ then the same applies to any box constructed from P via stochastic wirings.

4. Contextuality: a new framework

Contextuality is another feature of quantum mechanics that is not observed in the classical world. So far, the study of this phenomenon has mainly focused on particular examples of contextuality or "small" proofs of the Kochen-Specker theorem, although a general theory has not frequently been studied nor developed (see chapter 2).

In this chapter, I present a framework that allows the study of both nonlocality and contextuality in a unified manner. Our approach, similar in spirit to that of Cabello, Severini and Winter (CSW10), studies measurements and events from a graph-theoretic angle. However, our framework somehow refines that of (CSW10), since from its very definition it focuses on complete measurements, i.e. normalized probabilistic models.

The relation between our framework and the observable-based approach of Abramsky and Brandenburger (AB11) (see section 2.2.1) will be discussed in appendix E.

4.1. Contextuality scenarios and Probabilistic models

In our formalism, a contextuality scenario is completely described by a hypergraph like follows.

Definition 4.1. *A* ***contextuality scenario*** *is a hypergraph H such that no hyperedge contains another one:*

$$e_1, e_2 \in E(H),\ e_1 \subseteq e_2\ \Rightarrow e_1 = e_2, \tag{4.1}$$

and $\bigcup_{e\in E(H)} e = V(H)$.

The reason for postulating (4.1) is related to the normalization of probability: if all outcomes of a measurement e_1 are also outcomes of a measurement e_2, then the additional outcomes of e_2 necessarily have probability 0 and can therefore be disregarded. On the other hand, condition $\bigcup_{e\in E(H)} e = V(H)$ simply states that each outcome should occur in at least one measurement, i.e. the hypergraph does not have isolated vertices.

Throughout the thesis, when speaking of vertices and hyperedges I will sometimes be referring to their physical interpretation, meaning outcomes and measurements, hence both notations will be commonly used. Moreover, for simplicity I will sometimes refer to the hyperedges by "edges".

It is worth mentioning that definition 4.1 differs from the formalisms proposed in (AB11) and (CF12; FC13). These works also provide a formalization of contextuality phenomena in terms of hypergraphs, but the vertices of the hypergraph represent observables rather than outcomes, while the hyperedges stand for (maximal) jointly measurable sets of observables. In appendix E I present a more detailed discussion on the connection between both approaches.

In a concrete physical situation, every outcome happens with some probability. Moreover, our formalism deals with complete measurements, hence the sum of these outcome probabilities over all outcomes in a measurement is 1. This motivates the following definition:

Definition 4.2. *Let H be a contextuality scenario. A* ***probabilistic model*** *on H is an assignment $p : V(H) \to [0,1]$ of a probability $p(v)$ to each vertex $v \in V(H)$ such that*

$$\sum_{v \in e} p(v) = 1 \quad \forall e \in E(H). \tag{4.2}$$

It is important to keep in mind that each $p(v)$ is actually a *conditional* probability: it stands for the probability of getting the outcome v *given that* a measurement $e \ni v$ is conducted. We denote by $\mathcal{G}(H) \subseteq [0,1]^{|V|}$ the set of probabilistic models for the scenario H, which by construction is a polytope (and may be empty). This notation is supposed to suggest the reading "general probabilistic" in the sense of **general probabilistic theories** (Mac63; Lud85; Bar07). Note that in section 2 we do not denote p by "probabilistic model" but rather by "assignment", since we want to stress that a probabilistic model is properly normalized.

Some basic examples of contextuality scenarios and their probabilistic models are the following.

Example 4.3. Figure 4.1 displays the triangle scenario Δ. Its only probabilistic model is $p(v_1) = p(v_2) = p(v_3) = \frac{1}{2}$, since this is the only solution to the system of normalization equations

$$p(v_1) + p(v_2) = 1, \qquad p(v_2) + p(v_3) = 1, \qquad p(v_1) + p(v_3) = 1.$$

Contextuality scenarios having a unique probabilistic model, like Δ does, will be of particular importance in Theorem 4.8.

From the Δ scenario, another one can be constructed, such that it admits no probabilistic model:

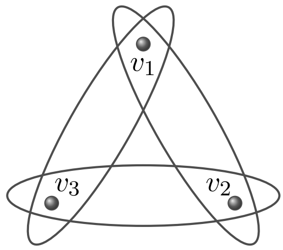


Figure 4.1.: The triangle scenario Δ.

Example 4.4. Figure 4.2 displays a contextuality scenario H_0 with $\mathcal{G}(H_0) = \emptyset$. Indeed, each of the outer triangles corresponds to a copy of the scenario Δ of Figure 4.1 and admits a unique probabilistic model where each vertex is assigned a probability 1/2. This is incompatible with the three-outcome measurement depicted in a dashed line which imposes that the probabilities associated with the three corresponding vertices should sum to 1.

In section 4.2, we will introduce the concept of product scenario and relate it to Bell scenarios. There, the notion of contextuality scenario for one single party in a Bell-type experiment plays an important role. Such a scenario is described as follows.

Example 4.5. Figure 4.3 displays the contextuality scenario defined by m measurements with d outcomes each, such that no two measurements share any outcome. Note that no assumption is made on the type or nature of the measurements, only that no outcome of any measurement defines the same event as any outcome of the others. Hence, such scenarios are particularly relevant for describing "box" experiments where an observer can press one of m buttons and record the corresponding measurement outcome.

Finally, as I mentioned in section 2.2, the study of contextuality begun with the Kochen-Specker theorem, whose "smallest" proof is given by 18 *projectors*. This scenario is depicted in our formalism like shown in Fig. 4.4 , where the 18 vertices map to the 18 projectors of the proof.

We now move on to the characterization of $\mathcal{G}(H)$. For fixed H, the set $\mathcal{G}(H) \subseteq \mathbb{R}^{V(H)}$ is defined in terms of finitely many linear inequalities with rational coefficients. Therefore, it is a convex polytope. A natural question then is: what are its extreme points? For example, for the CHSH scenario $B_{2,2,2}$ that we will discuss in Section 4.2, $\mathcal{G}(B_{2,2,2})$ is the no-signaling polytope, and hence its extreme points are the 16 deterministic boxes together with the 8

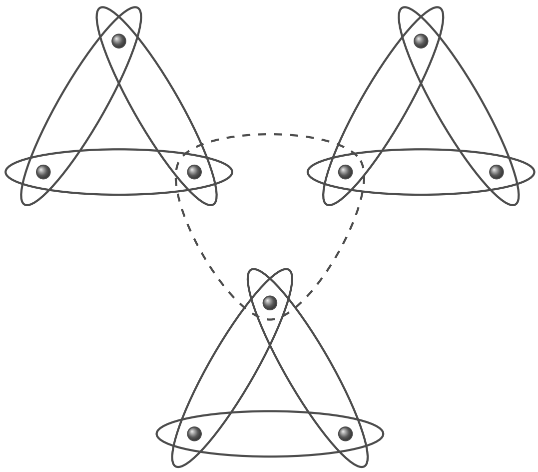


Figure 4.2.: Example of a scenario H_0 without any probabilistic model: $\mathcal{G}(H_0) = \emptyset$.

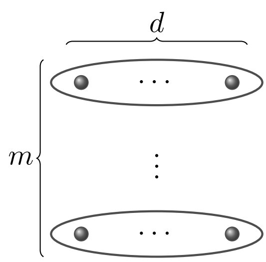


Figure 4.3.: The contextuality scenario $B_{1,m,d}$: a "Bell scenario" with only one party. It describes m measurements with d outcomes each, where no assumption is made on the type or nature of the measurements.

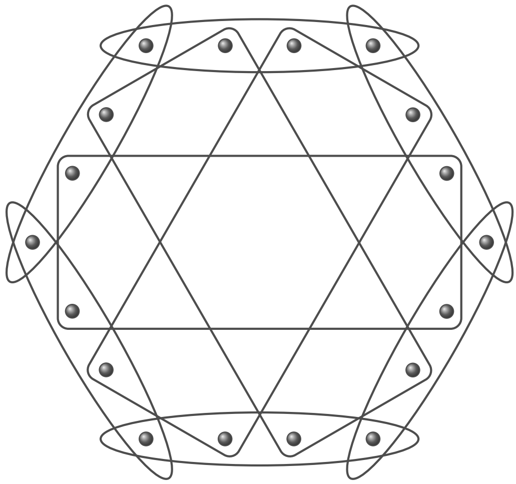


Figure 4.4.: The contextuality scenario H_{KS} proving the Kochen-Specker theorem (CEGA96; Cab08).

variants of the PR-box. In what follows, I present an abstract characterization of these extremal models which applies to every contextuality scenario. In order to do so, I will first introduce the notion of *induced subscenario*, which basically describes a family of subhypergraphs of the original scenario.

Definition 4.6. *Let H be a contextuality scenario. We say that a non-empty set $W \subseteq V(H)$* ***induces a subscenario*** *if $e_1 \cap W \subseteq e_2 \cap W$ implies that $e_1 = e_2$ for all $e_1, e_2 \in E(H)$. In this case, H_W with*

$$V(H_W) = W; \qquad E(H_W) = \{e \cap W \ : \ e \in E(H)\}$$

is the ***subscenario induced by*** *W.*

The assumption on W guarantees that H_W is also a contextuality scenario. In particular, it implies that $e \cap W \neq \emptyset$ for all $e \in E(H)$, meaning that W is a **transversal** of the hypergraph H. The intuition behind the subscenario H_W is the following: it is constructed by dropping all vertices which do not belong to W and restricting all edges accordingly. In doing this, the subset $W \subseteq V(H)$ is assumed to guarantee that no two different edges have equal restrictions or one restriction containing the other.

Example 4.7. The Triangle scenario 4.1 is as induced subscenario of the scenario $H = (V, E)$ depicted in Fig. 4.5. Indeed, if we take $W = \{v_1, v_2, v_3\}$,

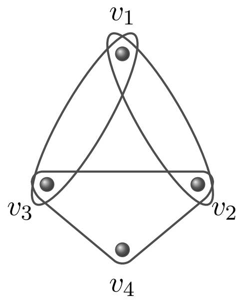


Figure 4.5.: Contextuality scenario of example 4.7. It has as an induced subscenario H_W the Triangle, depicted in Fig. 4.1, defined by $V_W = \{v_1, v_2, v_3\}$ and $E_W = \Big\{\{v_1, v_2\}, \{v_2, v_3\}, \{v_3, v_1\}\Big\}$.

the restrictions of the hyperedge set $E = \Big\{\{v_1, v_2\}, \{v_2, v_3\}, \{v_3, v_4, v_1\}\Big\}$ to the vertices of W gives $E_W = \Big\{\{v_1, v_2\}, \{v_2, v_3\}, \{v_3, v_1\}\Big\}$, which satisfies the properties of def. 4.6.

H_W may be interpreted as the same scenario H when all outcomes not in W have been forbidden. Indeed, we have already made implicit use of these concepts in section 3.5.3, where we restricted the study of the complete Bell scenario $(2, 2, 2)$ to only the "possible events" for the PR-box. The set of these events corresponds to W, and the orthogonality graph of possible events relates to the hypergraph H_W.

In addition, every probabilistic model p_W on H_W extends to H by setting

$$p(v) := \begin{cases} p_W(v) & \text{if } v \in W \\ 0 & \text{if } v \notin W \end{cases}.$$

When this happens, we say that p is the **extension** of p_W to H.

The extreme points of the polytope of probabilistic models on a contextuality scenario H can then be characterized as follows.

Theorem 4.8. *$p \in \mathcal{G}(H)$ is extremal if and only if it is the extension of $p_W \in \mathcal{G}(H_W)$ from some induced subscenario H_W which has p_W as its unique probabilistic model.*

Proof. If H has a unique probabilistic model, i.e. if $\mathcal{G}(H) = \{p\}$, then there is nothing to prove.

Otherwise, the extreme points of $\mathcal{G}(H)$ are precisely the extreme points of the facets of $\mathcal{G}(H)$. Since $\mathcal{G}(H)$ is defined by

$$p(v) \geq 0 \;\; \forall v \in V(H), \qquad \sum_{v \in e} p(v) = 1 \;\; \forall e \in E(H),$$

for every facet of $\mathcal{G}(H)$ there exists some $v \in V$ such that the facet contains exactly those $p \in \mathcal{G}(H)$ with $p(v) = 0$. We fix such a v and set

$$W = \{w \in V(H) \mid \exists\, p \in \mathcal{G}(H) \text{ s.t. } p(v) = 0 \;\wedge\; p(w) \neq 0\}.$$

In particular, $v \in W$, and W induces a subscenario H_W. By construction, $\mathcal{G}(H_W)$ is the facet of $\mathcal{G}(H)$ defined by $p(v) = 0$.

The assertion then follows by repeatedly applying this process to the induced subscenarios constructed in this way. At each step one obtains an induced subscenario of the original H, since if $H_{W,W'}$ is an induced subscenario of H_W, and H_W one of H, then $H_{W,W'}$ is an induced subscenario of H. This recursion necessarily ends with a scenario which admits a unique probabilistic model, since the dimension of $\mathcal{G}(H_W)$ decreases by 1 in each step. □

As the proof shows, a similar statement also holds for all faces of $\mathcal{G}(H)$: they all are of the form $\mathcal{G}(H_W)$ for some induced subscenario H_W.

We conclude that an extreme point $p \in \mathcal{G}(H)$ is uniquely determined by the set of vertices $W = \{v \in V(H) \mid p(v) \neq 0\}$, which induces a subscenario H_W with a unique probabilistic model corresponding to forgetting the zeros of p.

The deterministic models of Definition 4.19 are a special case of this. Clearly, every deterministic model is an extreme point of $\mathcal{G}(H)$. In terms of Theorem 4.8, p is deterministic if and only if each measurement in the associated H_W has only one outcome. Those extreme points which are not deterministic are the **maximally contextual** models in the scenario H.

4.2. Products of contextuality scenarios

As explained in section 2.1.1, a general scenario for non-locality (Bell Scenario) is defined as a set of distant parties, each of which has access to a device, and by pressing a button obtains an outcome. However, one may think that each party in fact operates on a contextuality scenario. Then, the natural question is: can the global (multi-party) situation be described as a contextuality scenario, and how? The affirmative answer to this question is based on the *product* of contextuality scenarios. The appropriate way to describe the joint scenario for two parties is given by the product of the two individual scenarios as follows:

Definition 4.9 ((FR81))**.** *Consider two hypergraphs $H_A = (V_A, E_A)$ and $H_B = (V_B, E_B)$. The **Foulis-Randall product (FR-product)** is the scenario $H_A \otimes H_B$ with*

$$V(H_A \otimes H_B) = V_A \times V_B, \qquad E(H_A \otimes H_B) = E_{A\to B} \cup E_{A\leftarrow B}$$

where

$$\begin{aligned} E_{A\to B} &:= \left\{ \bigcup_{v_A \in e_A} \{v_A\} \times f(v_A) \;:\; e_A \in E_A,\; f : e_A \to E_B \right\}, \\ E_{A\leftarrow B} &:= \left\{ \bigcup_{v_B \in e_B} f(v_B) \times \{v_B\} \;:\; e_B \in E_B,\; f : e_B \to E_A \right\}. \end{aligned} \tag{4.3}$$

Intuitively, an element of $E_{A\to B}$ is the following: first, an edge $e_A \in E(H_A)$ representing a measurement conducted by Alice; second, a function $f : e_A \to E(H_B)$ which determines the subsequent measurement of Bob as a function of Alice's outcome. This function f maps each vertex $a \in e_A$ to an edge $f(a) \in E_B$. Similarly for $E_{A\leftarrow B}$, where we think of Bob measuring first and communicating his outcome to Alice, who then chooses her measurement as a function of Bob's outcome. Both possibilities are feasible ways to operate on the joint system and therefore should be considered as measurements conductible on the joint system. Indeed, this specific form of the hyperedges encompasses the notion of correlated measurements defined in section 3.1. In this way, an edge in $H_A \otimes H_B$ is an element of $E_{A\to B}$, $E_{A\leftarrow B}$, or of both sets. For example, Figure 4.6(f) displays the FR-product of 4.6(a) with 4.6(b), which is another copy of 4.6(a). $E_{A\to B}$ contains the edges of Figure 4.6(c) and 4.6(d), while $E_{B\to A}$ consists of 4.6(c) and 4.6(e) (see example 4.14 below).

In the terminology of this chapter, the No Signaling principle is equivalently stated as follows.

Definition 4.10. *A probabilistic model $p \in \mathcal{G}(H_A \times H_B)$ is **no-signaling** if*

1. *For every $w \in V(H_B)$,*

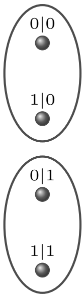
$$\sum_{v\in e} p(v,w) = \sum_{v\in e'} p(v,w)$$

for all $e, e' \in E(H_A)$;

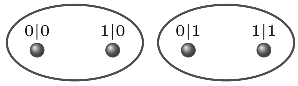
2. *For every $v \in V(H_A)$,*

$$\sum_{w\in e} p(v,w) = \sum_{w\in e'} p(v,w)$$

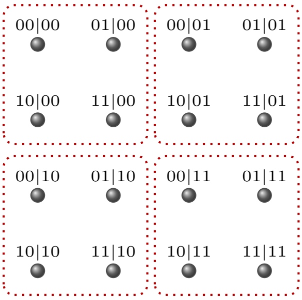
for all $e, e' \in E(H_B)$.



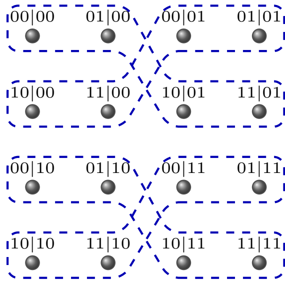
(a) Alice's two binary measurements $B_{1,2,2}$.



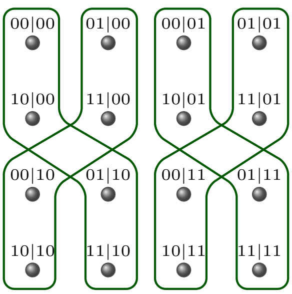
(b) Bob's two binary measurements $B_{1,2,2}$.



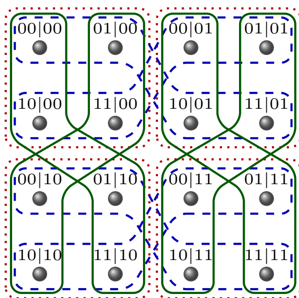
(c) Simultaneous measurements.



(d) Bob's measurement choice depends on Alice's outcome.



(e) Alice's measurement choice depends on Bob's outcome.



(f) Foulis-Randall product: the CHSH scenario $B_{2,2,2} = B_{1,2,2} \otimes B_{1,2,2}$.

Figure 4.6.: Construction of the CHSH scenario $B_{2,2,2}$ as a Foulis-Randall product $B_{2,2,2} = B_{1,2,2} \otimes B_{1,2,2}$.

Hence, a natural question is, do the allowed probabilistic models on $\mathcal{G}(H_A \times H_B)$ satisfy the No Signaling principle? A stronger answer to this question is due to Barnum, Fuchs, Renes and Wilce:

Proposition 4.11 ((BFRW05, Cor. 3.5)). *$\mathcal{G}(H_A \otimes H_B)$ is exactly the set of no-signaling models.*

It is in this sense that $H_A \otimes H_B$ automatically incorporates the no-signaling requirement of special relativity, and hence we regard it as the "right" product of contextuality scenarios.

One can also do all this for the case of *unidirectional* no-signaling: defining a product of H_A and H_B by only using the $E_{A\to B}$ of (4.3) gives probabilistic models which are no-signaling from Bob to Alice (see (BFRW05) for details). The resulting product contextuality scenario may be interpreted as describing a temporal succession of operating on H_B after having operated on H_A.

Given two contextuality scenarios H_A and H_B together with probabilistic models $p_A \in \mathcal{G}(H_A)$ and $p_B \in \mathcal{G}(H_B)$, there should exist a probabilistic model $p_A \otimes p_B$ on $H_A \otimes H_B$ having the interpretation of placing physical systems behaving as p_A and p_B "side by side" so that measurements can be conducted on both in parallel, revealing no correlations between the two systems, but independent statistics. To this end, we define

$$p_A \otimes p_B \; : \; V(H_A) \times V(H_B) \longrightarrow [0,1]$$

as a function which assigns to vertex $(v_A, v_B) \in V(H_A \otimes H_B)$ a probability given by $p(v_A, v_B) = p_A(v_A)p_B(v_B)$.

Proposition 4.12. *This $p_A \otimes p_B$ is a probabilistic model on $H_A \otimes H_B$.*

Proof. We need to prove that $\sum_{v\in e} p_A \otimes p_B(v) = 1$ for each edge $e \in E(H_A \otimes H_B)$. Without loss of generality, we can assume $e \in E_{A\to B}$, i.e. $e = \bigcup_{a\in e_A}\{a\} \times f(a)$ for some $e_A \in E_A$ and some $f : e_A \mapsto E_B$, which maps each vertex in e_A to an edge in H_B. Therefore,

$$\begin{aligned}\sum_{v\in e} p_A \otimes p_B(v) &= \sum_{a\in e_A} \sum_{b\in f(a)} p_A(a)p_B(b) \\ &= \sum_{a\in e_A} p_A(a) \sum_{b\in f(a)} p_B(b) = \sum_{a\in e_A} p_A(a) \cdot 1 = 1,\end{aligned}$$

since p_B and p_A are probabilistic models on H_B and H_A, respectively. □

We write $\mathcal{G}(H_A) \otimes \mathcal{G}(H_B)$ for the set of all probabilistic models of the form $p_A \otimes p_B$. Note however that for arbitrary contextuality scenarios H_A and H_B not every element of $\mathcal{G}(H_A \otimes H_B)$ is necessarily in the product form $p_A \otimes p_B$, i.e. $\mathcal{G}(H_A) \otimes \mathcal{G}(H_B) \subseteq \mathcal{G}(H_A \otimes H_B)$. Indeed, for the Bell scenario $B_{2,2,2} = B_{1,2,2} \otimes B_{1,2,2}$ discussed below, the PR-box is an element of $\mathcal{G}(B_{1,2,2} \otimes B_{1,2,2})$, but does not lie in the convex hull of $\mathcal{G}(B_{1,2,2}) \otimes \mathcal{G}(B_{1,2,2})$.

It is easy to check that the Foulis-Randall product "⊗" is a commutative binary operation on contextuality scenarios. However, when we move on to products of more than two scenarios, the Foulis-Randall product is not associative. Indeed, given three scenarios H_A, H_B, H_C, one possibility is to form the product $H_A \otimes H_B$ and then the product of this with H_C, which gives $(H_A \otimes H_B) \otimes H_C$. In this case, a hyperedge belongs to one of the four sets

$$E_{(A \to B) \to C}, \qquad E_{(A \leftarrow B) \to C}, \qquad E_{(A \to B) \leftarrow C}, \qquad E_{(A \leftarrow B) \leftarrow C} \tag{4.4}$$

where $E_{(A \to B) \to C}$ is defined to be the collection of all sets of the form

$$\left\{ (a,b,c) \in V(H_A) \times V(H_B) \times V(H_C) \mid a \in e_A,\ b \in f(a),\ c \in g(a,b) \right\} \tag{4.5}$$

where $e_A \in E(H_A)$ is fixed and $f : V(H_A) \to E(H_B)$ and $g : V(H_A) \times V(H_B) \to E(H_C)$ are any functions (and similarly for the other three sets). In this way, every one of the four sets (4.4) contains all those measurements associated to a certain ordering of the three parties; these four orderings are

$$A \overset{⏱}{\rightsquigarrow} B \overset{⏱}{\rightsquigarrow} C, \qquad B \overset{⏱}{\rightsquigarrow} A \overset{⏱}{\rightsquigarrow} C, \qquad C \overset{⏱}{\rightsquigarrow} A \overset{⏱}{\rightsquigarrow} B, \qquad C \overset{⏱}{\rightsquigarrow} B \overset{⏱}{\rightsquigarrow} A. \tag{4.6}$$

On the other hand, the bracketing $H_A \otimes (H_B \otimes H_C)$ is based in a similar way on four sets of edges

$$E_{A \to (B \to C)}, \qquad E_{A \leftarrow (B \to C)}, \qquad E_{A \to (B \leftarrow C)}, \qquad E_{A \leftarrow (B \leftarrow C)} \tag{4.7}$$

which represent the time orderings

$$A \overset{⏱}{\rightsquigarrow} B \overset{⏱}{\rightsquigarrow} C, \qquad B \overset{⏱}{\rightsquigarrow} C \overset{⏱}{\rightsquigarrow} A, \qquad A \overset{⏱}{\rightsquigarrow} C \overset{⏱}{\rightsquigarrow} B, \qquad C \overset{⏱}{\rightsquigarrow} B \overset{⏱}{\rightsquigarrow} A. \tag{4.8}$$

These four time orderings are different from (4.6); therefore, in general, $H_A \otimes (H_B \otimes H_C)$ contains different edges than $(H_A \otimes H_B) \otimes H_C$. This proves that the Foulis-Randall product is not associative.

Since we are interested on working with products of more than two contextuality scenarios, one way around this non-associativity issue is to define the $n-$fold product of hypergraphs as follows:

Definition 4.13. *Consider n hypergraphs $H_j = (V_j, E_j)$, $j \in \{1, n\}$. The n-fold product is the hypergraph $H_1 \otimes \cdots \otimes H_n$ with vertex set*

$$V(H_1 \otimes \cdots \otimes H_n) = V_1 \times \cdots \times V_n,$$

and edge set

$$E(H_1 \otimes \cdots \otimes H_n) = \bigcup_{\sigma, \boldsymbol{f}} \Big\{ (a_1, \ldots, a_n) \mid a_{\sigma^{-1}(i)} \in f_{\sigma^{-1}(i)}(a_{\sigma(1)}, \ldots, a_{\sigma(\sigma^{-1}(i)-1)}), \quad \text{for} \quad 1 \leq i \leq n \Big\}. \tag{4.9}$$

where σ runs over all possible permutations of the parties. The i-th outcome (a_i) belongs to a measurement defined by the $\sigma(i) - 1$ outcomes

$$(a_{\sigma^{-1}(1)}, \cdots, a_{\sigma^{-1}(\sigma(i)-1)})$$

via the function $f_{\sigma^{-1}(i)}$, and $\boldsymbol{f}$ varies over all possible choices of the functions $\{f_{\sigma^{-1}(i)}\}_i$.

The intuition behind this definition comes from constructing each edge in $H_1 \otimes \cdots \otimes H_n$ from an $n-$partite correlated measurement. Given a certain order of the parties, we think of the first one performing a measurement e_1 and sending the outcome a_1 to the the other parties. Then, party 2 performs a measurement e_2 that depends on outcome a_1, and obtains an output a_2. Party 3 then receives the outputs a_1 and a_2, and decides a measurement e_3. He sends his outcome to all the remaining parties. Then, the total output of the protocol is given by $(a_1, \cdots, a_n)$, where the measurement that party j performs may depend on the outputs he receives from the $j-1$ previous parties. In the bipartite case, we decomposed the vertex set as $E_{A\to B} \cup E_{A\leftarrow B}$, since we had only two sequential orders for the correlated measurements, namely A, B and B, A. For n parties, we need to consider all their possible orderings in order to encompass all possible correlated measurements. Hence, the need of the permutations σ in the definition of $n-$fold product.

We now explain how Bell scenarios are examples of contextuality scenarios. The Bell scenario $B_{n,m,d}$ consists of n parties having access to m measurements each, each of which has d possible outcomes. At the single-party level, the outcomes form a contextuality scenario $B_{1,m,d}$ as depicted in Figure 4.3. As contextuality scenarios, we define

$$B_{n,m,d} := \underbrace{B_{1,m,d} \otimes \cdots \otimes B_{1,m,d}}_{n}, \tag{4.10}$$

and we will see in the following how this leads to the usual concepts studied as "nonlocality". It is straightforward to generalize this definition and all our upcoming results to scenarios where the parties have access to different numbers of measurements and outcomes per measurement, but we will not consider this explicitly.

Example 4.14 (The CHSH scenario)**.** Figure 4.6 illustrate how $B_{2,2,2}$ arises as $B_{1,2,2} \otimes B_{1,2,2}$. A vertex $ab|xy$ represents the event where Alice (resp. Bob) chooses measurement x (resp. y) and obtains output a (resp. b). In this scenario, the edges are as follows:

- For simultaneous measurements, the f of (4.3) are constant, and the measurements are as in Figure 4.6(c):

$$\{00|00,\ 01|00,\ 10|00,\ 11|00\},$$
$$\{00|01,\ 01|01,\ 10|01,\ 11|01\},$$
$$\{00|10,\ 01|10,\ 10|10,\ 11|10\},$$
$$\{00|11,\ 01|11,\ 10|11,\ 11|11\}.$$

- If Alice measures first and Bob's choice of setting depends on her outcome, then the events are of the form $ab|xf(a)$, where f is not a constant. Thus we have two possibilities: $f(a) = a$ or $f(a) = 1 - a$. In the first case we obtain the edges

$$\{00|00,\ 01|00,\ 10|01,\ 11|01\},$$
$$\{00|10,\ 01|10,\ 10|11,\ 11|11\},$$

  and in the second case,

$$\{00|01,\ 01|01,\ 10|00,\ 11|00\},$$
$$\{00|11,\ 01|11,\ 10|10,\ 11|10\}.$$

  These are the dashed blue edges in figures 4.6(d) and 4.6(f).

- Similarly, Bob measuring first with Alice's subsequent choice of setting depending on his outcome gives rise to the edges

$$\{00|00,\ 01|10,\ 10|00,\ 11|10\},$$
$$\{00|01,\ 01|11,\ 10|01,\ 11|11\},$$
$$\{00|10,\ 01|00,\ 10|10,\ 11|00\},$$
$$\{00|11,\ 01|01,\ 10|11,\ 11|01\}.$$

  These are the solid green edges in Figures 4.6(e) and 4.6(f).

In what follows, we will see that the allowed probabilistic models on $B_{n,m,d}$ are exactly those no-signaling correlations on (n, m, d) in the framework of Non Locality.

Proposition 4.15. *Let $B_{n,k,m}$ be a Bell scenario. Then $\mathcal{G}(B_{n,k,m})$ is the standard no-signaling polytope containing all no-signaling boxes of type (n, m, d).*

Proof. While this follows from an application of the multipartite version of Proposition 4.11, we believe that an independent proof is more instructive.

We identify the vertices of $B_{n,k,m}$ with the events

$$a_1 \dots a_n | x_1 \dots x_n, \quad a_i \in \{1, \dots, m\},\ x_i \in \{1, \dots, k\}$$

in the usual Bell scenario notation.

We show first that a non-signaling box of type (n, k, m) is indeed a probabilistic model on $B_{n,k,m}$. Such a box is an assignment of a probability $p(\vec{a}|\vec{x})$ to each event $\vec{a}|\vec{x}$ such that the no-signaling equations

$$\sum_{a_i} p(a_1 \dots a_n | x_1 \dots x_n) = \sum_{a_i} p(a_1 \dots a_n | x_1 \dots x_i' \dots x_n) \tag{4.11}$$

hold (where the right-hand side is the same except that the setting x_i has been replaced by some other setting x_i'), as well as the normalization condition

$$\sum_{a_1, \dots, a_n} p(a_1 \dots a_n | x_1 \dots x_n) = 1. \tag{4.12}$$

Now we consider any edge in the scenario $B_{n,k,m}$. Without loss of generality, we take the underlying total order of the parties to be the numerical one, so that the temporal order of the parties' measurements is simply $1, \dots, n$. The settings used by the parties are then determined by functions $x_i = f_i(a_1, \dots, a_{i-1})$, and we need to consider

$$\sum_{a_1, \dots, a_n} p(a_1 \dots a_n | f_1() \dots f_n(a_1, \dots a_{n-1})),$$

where $x_1 = f_1()$ is a function without arguments, i.e. a constant. Since the vector of settings does not depend on a_n, the no-signaling equations imply that the last function $f_n(a_1, \dots, a_{n-1})$ can be replaced by an arbitrary constant setting x_n without changing the value of the sum. After applying this modification, the vector of settings does not depend on a_{n-1}, and then the setting of party $n-1$ can be taken to be some fixed x_{n-1}. Applying this procedure repeatedly eventually replaces all functions $f_i(a_1, \dots, a_{i-1})$ by constant settings x_i. Then the normalization equation implies that the sum has the value 1, as has been claimed.

Conversely, suppose that p is a probabilistic model on $B_{n,k,m}$. Then p satisfies the normalization equation since taking all functions f_i to be constants x_i gives precisely (4.12). In order to prove the no-signaling equation, we fix arbitrary outputs b_j and choose all functions to be constants $f_j = x_j$, except for

$$f_n(a_1, \ldots, a_{n-1}) = \begin{cases} x_n & \text{if } a_j = b_j \text{ for all } j < n, \\ x'_n & \text{otherwise,} \end{cases}$$

which gives the equation

$$\sum_{a_n} p(b_1 \ldots b_{n-1} a_n | x_1 \ldots x_n) + \sum_{a_n} \sum_{(a_1,\ldots,a_{n-1}) \neq (b_1,\ldots,b_{n-1})} p(a_1 \ldots a_n | x_1 \ldots x'_n) = 1.$$

Upon combining this with the already proven normalization equation

$$\sum_{a_n} p(b_1 \ldots b_{n-1} a_n | x_1 \ldots x'_n) + \sum_{a_n} \sum_{(a_1,\ldots,a_{n-1}) \neq (b_1,\ldots,b_{n-1})} p(a_1 \ldots a_n | x_1 \ldots x'_n) = 1.$$

we obtain (4.11) with $i = n$ and $b_1 \ldots b_{n-1}$ in place of $a_1 \ldots a_{n-1}$. The other no-signaling equations can be obtained in the same way, choosing different orders of the parties. □

In particular, this proof shows explicitly how the non-trivial edges occurring in the definition of "$\otimes$" give rise to the no-signaling property.

4.3. Non-orthogonality graphs

An important feature of our formalism is that it allows to relate different classes of probabilistic models with concepts of graph theory. In chapter 3 we studied correlations using graph invariants of both the orthogonality and the non-orthogonality graph. In this section, we will extend the definition of the latter to general contextuality scenarios.

Definition 4.16 (Non-orthogonality graph)**.** *Let H be a contextuality scenario. The* ***non-orthogonality graph*** $\mathrm{NO}(H)$ *is the undirected graph with the same vertices as H and adjacency relation*

$$u \sim v \iff \nexists e \in E(H) \text{ with } \{u, v\} \subseteq e.$$

We say that two different vertices u and v of H are orthogonal, which we denote by $u \perp v$, if they are not adjacent in $\mathrm{NO}(H)$, i.e. if they do belong to a common edge in H.

In the case of Bell scenarios, this definition of non-orthogonality graph coincides with that of chapter 3. Indeed, as explained in section 4.2 every hyperedge in $B_{n,m,d}$ is associated to a correlated measurement among the n parties, and vice-versa. This feature of the hyperedge set $E(B_{n,m,d})$ guarantees that the orthogonality relations which arise in the definition of $B_{n,m,d}$ are exactly those determined by the LO principle. The formal proof is presented below.

Lemma 4.17. *The events $u, v \in V(B_{n,k,m})$ are locally orthogonal if and only if $u \perp v$. In words, the orthogonality between two events in the sense of setion 3.1 naturally arises from the FR product.*

Proof. Suppose that $u = a_1 \dots a_n | x_1 \dots x_n$ and $v = a'_1 \dots a'_n | x'_1 \dots x'_n$ are locally orthogonal. By relabelling the parties, we can arrange for $a_1 \neq a'_1$ and $x_1 = x'_1$. Now choose any functions $f_2, \dots, f_n$ with $f_i(a_1) = x_i$ and $f_i(a'_1) = x'_i$. Then the set of events of the form

$$b_1 \dots b_n | x_1 f_2(b_1) \dots f_n(b_1)$$

defines an edge in $B_{n,k,m}$ containing both u and v. Intuitively, Alice communicates her outcome to the other parties who then choose their measurement settings as a function of that outcome.

Conversely, $u \perp v$ means that there is an edge $e \in E(B_{n,k,m})$ with $u, v \in e$. More concretely, this states that there is an ordering of the parties $\sigma(1), \dots, \sigma(n)$ and functions $f_{\sigma(i)}(b_{\sigma(1)}, \dots, b_{\sigma(i-1)})$ such that e contains exactly those events which have the form

$$b_{\sigma(1)} \dots b_{\sigma(n)} | f_{\sigma(1)}() \dots f_{\sigma(n)}(b_{\sigma(1)}, \dots, b_{\sigma(n-1)})$$

where we have now written the parties in the order given by the permutation σ. Since both given events $u = a_1 \dots a_n | x_1 \dots x_n$ and $v = a'_1 \dots a'_n | x'_1 \dots x'_n$ are assumed to be of this form, we know that $x_{\sigma(i)} = f_{\sigma(i)}(a_{\sigma(1)}, \dots, a_{\sigma(i-1)})$ and $x'_{\sigma(i)} = f_{\sigma(i)}(a'_{\sigma(1)}, \dots, a'_{\sigma(i-1)})$. Now let $\sigma(j)$ be the smallest index with $a_{\sigma(j)} \neq a'_{\sigma(j)}$. Then, since $x_{\sigma(j)}$ and $x'_{\sigma(j)}$ only depend on $a_{\sigma(i)}$ and $a'_{\sigma(i)}$ with $i < j$, we conclude that $x_{\sigma(j)} = x'_{\sigma(j)}$, which proves the claim. $\square$

For general product scenarios, a relation between the non-orthogonality graphs of the involved scenarios holds similarly to the one presented in section 3.4 for Bell scenarios.

Lemma 4.18. *Let H_A and H_B be contextuality scenarios. Then,*

$$\mathrm{NO}(H_A \otimes H_B) = \mathrm{NO}(H_A) \boxtimes \mathrm{NO}(H_B).$$

Proof. Clearly both sides are graphs having $V(H_A) \times V(H_B)$ as their set of vertices, so what needs to be shown is that the adjacency relations coincide.

We first prove that if $(u_A, u_B) \perp (v_A, v_B)$ in $\mathrm{NO}(H_A \otimes H_B)$, then these two vertices are also not adjacent in $\mathrm{NO}(H_A) \boxtimes \mathrm{NO}(H_B)$. The assumption means that there is an edge $e \in E(H_A \otimes H_B)$ which contains both (u_A, u_B) and (v_A, v_B); this edge has one of the two forms of (4.3). If it is in $E_{A \to B}$, then $u_A, v_A \in e_A$, meaning that $u_A \perp v_A$. Similarly, if the edge is in $E_{A \leftarrow B}$, then $u_B \perp v_B$. The conclusion follows from either case.

For proving the opposite implication, we show that $(u_A, u_B) \perp (v_A, v_B)$ in $\mathrm{NO}(H_A) \boxtimes \mathrm{NO}(H_B)$ implies the same in $\mathrm{NO}(H_A \otimes H_B)$. The assumption means that $u_A \perp v_A$ or $u_B \perp v_B$; by symmetry, it is enough to consider the case $u_A \perp v_A$. Then, there exists some $e_A \in E(H_A)$ with $u_A, v_A \in e_A$. Now choose $e_B, e'_B \in E_B$ such that $u_B \in e_B$ and $v_B \in e'_B$, and some function $f : e_A \to E_B$ with $f(u_A) = e_B$ and $f(v_A) = e'_B$. Then

$$\bigcup_{a \in e_A} \{a\} \times f(a)$$

is an edge in $H_A \otimes H_B$ containing (u_A, u_B) and (v_A, v_B), which proves the claim. □

In next sections, I will present relevant sets of probabilistic models, and relate most of them with graph-theoretical invariants of the non-orthogonality graphs.

4.4. Classical models

For each scenario H, one can define several relevant subsets of the set of $\mathcal{G}(H)$. In the following, we will define these and study some of their properties in some detail, starting with set of classical models $\mathcal{C}(H)$.

Definition 4.19. *Let H be a contextuality scenario.*

1. *A probabilistic model $p \in \mathcal{G}(H)$ is **deterministic** if $p(v) \in \{0, 1\}$ for all $v \in V(H)$.*

2. *A probabilistic model $p \in \mathcal{G}(H)$ is **classical** if it is a convex combination of deterministic ones.*

This definition of "classical" encompasses the idea of hidden variables like in the works by Bell (Bel64), Fine (Fin82) and Kochen-Specker (KS67).

For finite H there are only finitely many deterministic models, hence the set of classical models is a polytope. We denote this polytope by $\mathcal{C}(H)$.

Example 4.20 ((CEGA96))**.** Let H be the contextuality scenario of Figure 4.4. We will show that $\mathcal{C}(H) = \emptyset$. To see this, let V_1 be the set of vertices to which a given deterministic model assigns a 1. The set V_1 is required to intersect every edge in one and only one vertex. Since every vertex appears in precisely two edges, $2|V_1|$ should be equal to the number of edges. Since the latter is odd, it implies that no deterministic model exists, which means that $\mathcal{C}(H) = \emptyset$.

A concept that plays an important role for classical models, as we will see below, is that of *exact transversal*:

Definition 4.21 ((Eit94))**.** *An* **exact transversal** *of a hypergraph* $H = (V, E)$ *is a subset of vertices* $V' \subseteq V$ *such that each hyperedge* $e \in E$ *intersects* V' *at exactly one element.*

As we just exemplified, a deterministic model p is determined by the set of vertices $V_1 = \{v \in V \mid p(v) = 1\}$. By definition of deterministic model, V_1 has the property that it intersects every edge in exactly one vertex, hence V_1 is an exact transversal. Conversely, every exact transversal V_1 defines a deterministic model in this way. Therefore, $\mathcal{C}(H) \neq \emptyset$ if and only if H has an exact transversal.

In the case of Bell scenarios, the set of classical models $\mathcal{C}(B_{n,m,d})$ coincides with the standard Bell polytope. Indeed, one way to define the Bell polytope is as the convex hull of deterministic models, and a deterministic model in the contextuality scenario $B_{n,m,d}$ is the same as a local deterministic model in the Bell sense.

More generally, given two arbitrary contextuality scenarios H_A and H_B, the classical models on their product are characterized as follows.

Proposition 4.22.

$$\mathcal{C}(H_A \otimes H_B) = \mathrm{conv}\left(\mathcal{C}(H_A) \otimes \mathcal{C}(H_B)\right),$$

where conv(S) *denotes the convex hull of the elements in* S*.*

This is supposed to be seen in contrast to the case of general no-signaling models, where we saw that $\mathcal{G}(H_A) \otimes \mathcal{G}(H_B) \subseteq \mathcal{G}(H_A \otimes H_B)$.

Proof. Let $p_A \in \mathcal{C}(H_A)$ and $p_B \in \mathcal{C}(H_B)$ be deterministic models. Then also $p_A \otimes p_B$ is a deterministic model on $H_A \otimes H_B$, which proves $\mathcal{C}(H_A \otimes H_B) \supseteq \mathrm{conv}\left(\mathcal{C}(H_A) \otimes \mathcal{C}(H_B)\right)$ by convexity of $\mathcal{C}(H_A \otimes H_B)$.

Conversely, consider a deterministic model p_{AB} on $H_A \otimes H_B$. Let V_1 be the set of vertices in $H_A \otimes H_B$ for which $p_{AB}(v) = 1$, and define $p_A \in \mathcal{C}(H_A)$ and $p_B \in \mathcal{C}(H_B)$ as follows: for each $v_A \in V_A$, set $p_A(v_A) = 1$ if and only if there exists $v_B \in V_B$ such that $(v_A, v_B) \in V_1$, and $p_A(v_A) = 0$ otherwise. Similarly, define p_B. We want to check that these are indeed probabilistic models, i.e. show that $\sum_{v_A \in e_A} p_A(v_A) = 1$ and $\sum_{v_B \in e_B} p_B(v_B) = 1$ for every edge e_A of H_A and e_B of H_B. As V_1 is an exact transversal of $H_A \otimes H_B$, no two elements of V_1 belong to the same edge. This implies that if both $(u_A, u_B), (u'_A, u'_B) \in V_1$, then there is no $e_A \in E(H_A)$ with $\{u_A, u'_A\} \subseteq e_A$: for if there where, then we could construct an edge as in the proof of Lemma 4.18 which contains both (u_A, u_B) and (u'_A, u'_B). It follows for each edge $e_A \in E_A$, there is at most one vertex $v_A \in e_A$ with $p_A(v_A) = 1$. In fact, there is exactly one such vertex, since $e_A \times e_B$ is an edge on $H_A \otimes H_B$ for any $e_B \in E(H_B)$, and this edge intersects with V_1. Hence, p_A is a probabilistic model on H_A. Similarly, p_B is a probabilistic model. Since $p_{AB} = p_A \otimes p_B$ by construction, the claim follows by convexity. □

Finally, I will show how to detect classicality using a graph-theoretic invariant from section 2.3. Indeed, since the classical deterministic models are characterized by exact transversals on H, which in turn form cliques in $\mathrm{NO}(H)$, we expect these graph-theoretical objects to be relevant in the characterization of $\mathcal{C}(H)$. In what follows we will see that the graph invariant defined in terms of cliques that plays the role is the fractional packing number.

Proposition 4.23. *A probabilistic model $p \in \mathcal{G}(H)$ is in $\mathcal{C}(H)$ if and only if $\alpha^*(\mathrm{NO}(H), p) \leq 1$.*

Note that the normalization $\sum_{v \in e} p(v) = 1$ for every $e \in E(H)$ implies that $\alpha^*(\mathrm{NO}(H), p) \geq 1$, so that the condition $\alpha^*(\mathrm{NO}(H), p) \leq 1$ is actually equivalent to $\alpha^*(\mathrm{NO}(H), p) = 1$.

Proof. By definition, $\alpha^*(\mathrm{NO}(H), p) \leq 1$ means that if $q : V(H) \to [0, 1]$ are vertex weights satisfying $\sum_{v \in C} q_v \leq 1$ for all cliques $C \subseteq \mathrm{NO}(H)$, then also

$$\sum_{v \in V(H)} q_v p(v) \leq 1. \tag{4.13}$$

In order to prove the claim for all classical p, it is sufficient to consider deterministic p. In this case, the associated set $V_1 = \{v \in V(H) \mid p(v) = 1\}$ is itself a clique in $\mathrm{NO}(H)$, while all other $p(v)$ vanish, and hence (4.13) follows from the assumption on q.

For the other direction, we use the dual formulation (A.8) of the weighted fractional packing number: there exists a number $x_C \geq 0$ associated to every clique $C \subseteq \mathrm{NO}(H)$ such that $p(v) \leq \sum_{C \ni v} x_C$ and $\sum_C x_C = 1$. We claim that every C for which $x_C \neq 0$ corresponds to a deterministic model via its characterization as the set of vertices $V_1 = \{v \in V \mid p(v) = 1\}$. In other words, we will see that if $x_C \neq 0$, then $|e \cap C| = 1$ for every $e \in E(H)$. First, $|e \cap C| \leq 1$, since e is an independent set in $\mathrm{NO}(H)$ while C is a clique. Second, the chain of inequalities

$$1 = \sum_{v \in e} p(v) \leq \sum_{v \in e} \sum_{C \ni v} x_C = \sum_{C \text{ with } C \cap e \neq \emptyset} x_C \leq \sum_C x_C = 1$$

actually needs to be a chain of equalities. $\sum_{C \text{ with } C \cap e \neq \emptyset} x_C = \sum_C x_C$ implies that if $x_C \neq 0$, then $|e \cap C| = 1$ for every $e \in E(H)$, i.e. every clique C with $x_C \neq 0$ is an exact transversal (hence deterministic model) on H. In addition, we also conclude that $p(v) = \sum_{C \ni v} x_C$, i.e. $p = \sum_C x_C \mathbb{1}_C$, which is an explicit decomposition of p as a convex combination of deterministic models. □

4.5. Quantum models

In this section I will present the probabilistic models that arise when the contextuality scenario represents measurements on a quantum system.

Let $\mathcal{H}$ be the Hilbert space under consideration. We denote by $\mathcal{B}(\mathcal{H})$ the set of all bounded operators on $\mathcal{H}$, and $\mathcal{B}_+(\mathcal{H})$ the subset of positive semi-definite operators. A quantum state ρ is given by a normalized density operator, i.e. by some $\rho \in \mathcal{B}_{+,1}(\mathcal{H})$, where $\mathcal{B}_{+,1}(\mathcal{H}) := \{\rho \in \mathcal{B}_+(\mathcal{H}) \mid \operatorname{tr} \rho = 1\}$. We define a quantum probabilistic model as follows.

Definition 4.24. *Let H be a contextuality scenario. A probabilistic model $p \in \mathcal{G}(H)$ is a **quantum model** if there exists a Hilbert space $\mathcal{H}$, a quantum state $\rho \in \mathcal{B}_{+,1}(\mathcal{H})$ and a projection operator $P_v \in \mathcal{B}(\mathcal{H})$ associated to every $v \in V$ which constitute projective measurements in the sense that*

$$\sum_{v \in e} P_v = \mathbb{1}_{\mathcal{H}} \quad \forall e \in E(H), \tag{4.14}$$

and reproduce the given probabilities,

$$p(v) = \operatorname{tr}(\rho P_v) \quad \forall v \in V(H). \tag{4.15}$$

We denote by $\mathcal{Q}(H)$ the set of quantum models, which is convex. Indeed, let $p_1, p_2 \in \mathcal{Q}(H)$ be quantum models implemented by Hilbert spaces $\mathcal{H}_1$, $\mathcal{H}_2$, projection operators $P_{1,v}$, $P_{2,v}$ and states ρ_1, ρ_2 on the respective Hilbert space. Then, for any coefficient $\lambda \in [0,1]$ we can construct a quantum representation of $\lambda p_1 + (1-\lambda)p_2$ by setting $\mathcal{H} := \mathcal{H}_1 \oplus \mathcal{H}_2$, $\quad P_v := P_{1,v} \oplus P_{2,v}$, and $\quad \rho := \lambda\rho_1 \oplus (1-\lambda)\rho_2$. It is immediate to verify that this is indeed a quantum representation of $\lambda p_1 + (1-\lambda)p_2$. Moreover, every classical model is a quantum model: $\mathcal{C}(H) \subseteq \mathcal{Q}(H)$. This follows from the convexity of $\mathcal{Q}(H)$, upon showing that every deterministic model is quantum: a deterministic model p can be seen to be quantum by setting $\mathcal{H} = \mathbb{C}$, $P_v = p(v) \cdot \mathbb{1}$ and $\rho = \mathbb{1}$.

In the case of product scenarios, the set of quantum models $\mathcal{Q}(H_A \otimes H_B)$ satisfies the following property.

Proposition 4.25. *Let H_A and H_B be two contextuality scenarios. Then $p \in \mathcal{Q}(H_A \otimes H_B)$ if and only if there is a Hilbert space $\mathcal{H}$, a quantum state $\rho \in \mathcal{B}_{+,1}(\mathcal{H})$ and projection operators $P_{A,u} \in \mathcal{B}(\mathcal{H})$, $P_{B,v} \in \mathcal{B}(\mathcal{H})$ assigned to every $u \in V(H_A)$, $v \in V(H_B)$ such that*

$$\sum_{u \in e_A} P_{A,u} = \mathbb{1}_{\mathcal{H}} = \sum_{v \in e_B} P_{B,v} \quad \forall e_A \in E(H_A),\ e_B \in E(H_B),$$

$$[P_{A,u}, P_{B,v}] = 0 \qquad\qquad \forall u \in V(H_A),\ v \in V(H_B),$$

and the given probabilistic model is reproduced,

$$p(u,v) = \operatorname{tr}(\rho\, P_{A,u} P_{B,v}) \qquad \forall u \in V(H_A),\ v \in V(H_B). \tag{4.16}$$

Proof. We start from (4.16) and assign to every vertex $(u,v) \in V(H_A \otimes H_B)$ the projection

$$P_{(u,v)} := P_{A,u} P_{B,v},$$

so that (4.15) holds by (4.16). By symmetry, it is sufficient to show (4.14) for an edge $e \in E_{A \to B}$ given by

$$e := \bigcup_{a \in e_A} \{a\} \times f(a) \ \text{ with } \ e_A \in E_A,\ f : e_A \to E_B.$$

In this case,

$$\sum_{w \in e} P_w = \sum_{u \in e_A} P_{A,u} \sum_{v \in f(u)} P_{B,v} = \sum_{u \in e_A} P_{A,u} \cdot \mathbb{1}_{\mathcal{H}} = \mathbb{1}_{\mathcal{H}},$$

which is analogous to the computation in the proof of Proposition 4.12.

Conversely, one can construct the "local" observables $P_{A,u}$ and $P_{B,v}$ from a quantum model on $\mathcal{Q}(H_A \otimes H_B)$ by noting that the operators

$$P_{A,u} := \sum_{v \in e_B} P_{(u,v)}, \qquad P_v := \sum_{u \in e_A} P_{(u,v)} \tag{4.17}$$

do not depend on the choice of $e_B \in E(H_B)$ or $e_A \in E(H_A)$, respectively. To see this, it is enough to prove that

$$\sum_{v\in e_B} P_{(u,v)} = \sum_{v\in e'_B} P_{(u,v)} \tag{4.18}$$

for any $u \in V(H_A)$ and $e_B, e'_B \in E(H_B)$, which is analogous to the proof of Proposition 4.15. Choose some $e_A \ni u$ and consider the function $f : e_A \to E(H_B)$ with

$$f(u') = \begin{cases} e_B & \text{if } u' = u, \\ e'_B & \text{otherwise.} \end{cases}$$

An application of (4.14) to the edge defined by f as well as the edge $e_A \times e'_B$ gives

$$\sum_{v\in e_B} P_{(u,v)} + \sum_{u'\in e_A\setminus\{u\}} \sum_{v\in e'_B} P_{(u',v)} = \mathbb{1}_{\mathcal{H}} = \sum_{u'\in e_A} \sum_{v\in e'_B} P_{(u',v)},$$

which reduces to (4.18) after cancelling terms. This shows that the "local" operators (4.17) are well-defined.

The normalization condition $\sum_u P_{A,u} = \mathbb{1}_{\mathcal{H}} = \sum_v P_{B,v}$ now is an immediate consequence of (4.14). Finally, the commutativity $[P_{A,u}, P_{B,v}] = 0$ follows again from the normalization

$$\sum_{u'\in e_A} \sum_{v'\in e_B} P_{(u',v')} = \mathbb{1}_{\mathcal{H}},$$

taken for some $e_A \ni u$ and $e_B \ni v$: the terms in this sum are necessarily mutually orthogonal, and hence commute pairwise; but now both $P_{A,u}$ and $P_{B,v}$ are partial sums of this big sum, and therefore these commute as well. Also, mutual orthogonality implies $P_{(u,v)} = P_{A,u}P_{B,v}$, which yields the desired probabilities (4.16). □

This characterization of quantum models on product scenarios somehow generalizes the **commutativity paradigm** of quantum correlations in Bell scenarios (JNP+11; Fri12). A straightforward generalization of this result for quantum models on n-fold products proves that $\mathcal{Q}(B_{n,m,d})$ is the set of quantum correlations in the Bell sense in the commutativity paradigm.

Another immediate consequence of proposition 4.25 is that $\mathcal{Q}(H_A)\otimes\mathcal{Q}(H_B) \subseteq \mathcal{Q}(H_A \otimes H_B)$. Again, the CHSH scenario $B_{2,2,2} = B_{1,2,2} \otimes B_{1,2,2}$ exemplifies that this is not an equality in general.

4.6. A hierarchy of relaxations

For Bell scenarios, quantum correlations with the commutativity paradigm are characterized by a sequence ("hierarchy") of semidefinite programs, due to Navascués, Pironio and Acín (NPA07; NPA08). In this section, I will extend this hierarchy to one that suits general contextuality scenarios in our framework. Our hierarchy may be considered as a special case of the general hierarchy for noncommutative polynomial optimization (PNA10).

A moment matrix of order n associated with a contextuality scenario $H = (V, E)$ is a symmetric matrix M_n whose rows and columns are indexed by **words** of size at most n written in the alphabet formed by V. More explicitly, the indices of M_n belong to the set $V^{*n} = \bigcup_{k \leq n} V^k$. $\emptyset \in V^{*n}$ is the empty string of length 0, and by $V^* = \bigcup_{k \in \mathbb{N}} V^k$ we denote the set of all (arbitrary large) strings over V. Moreover, we choose the normalization $M_k(\emptyset, \emptyset) = 1$. In what follows, I denote the strings by vectors, i.e. $\mathbf{v} = v_1 \dots v_n$ and its reverse by $\mathbf{v}^\dagger = v_n \dots v_1$. Given two strings $\mathbf{v} \in V^*$ and $\mathbf{w} \in V^*$, I write their concatenation simply as $\mathbf{vw} \in V^*$, and I also use $v_1 \dots \hat{v}_i \dots v_n$ as a shorthand for $v_1 \dots v_{i-1} v_{i+1} \dots v_n$.

Definition 4.26. *A matrix M_n is a **certificate of order** n for the probabilistic model $p \in \mathcal{G}(H)$ if*

1. *it is positive semidefinite: $M_n \succeq 0$,*
2. *for every $v \in V$,*
$$M_n(v, \emptyset) = p(v) \tag{4.19}$$

In addition, we can define some other conditions for the matrices M_n to satisfy.

Definition 4.27. *Let M_n be a moment matrix.*

1. *M_n is **normalized** with respect to the contextuality scenario $H = (V, E)$ if for every two strings $\mathbf{v} \in V^{*(n-1)}$ and $\mathbf{w} \in V^{*n}$, and every hyperedge $e \in E$, the following condition holds:*
$$\sum_{u \in e} M(\mathbf{v}u, \mathbf{w}) = M(\mathbf{v}, \mathbf{w}). \tag{4.20}$$
2. *M_n is **orthogonal** with respect to $H = (V, E)$ if for every $e \in E$, and every $\mathbf{v}, \mathbf{w} \in V^{*(n-1)}$, if $v, w \in e$ then*
$$M(\mathbf{v}v, \mathbf{w}w) = 0. \tag{4.21}$$

Our hierarchy of relaxation is then defined as follows.

Definition 4.28. *Let H be a contextuality scenario. We say that $p \in \mathcal{G}(H)$ is a $\mathcal{Q}_n$-model if there exists a certificate of order n for p which satisfies Normalization (4.20) and Orthogonality (4.21).*

By definition, testing whether a given probabilistic model lies in $\mathcal{Q}_n$ is a semidefinite programming problem of size $\frac{|V|^{n+1}-1}{|V|-1} \times \frac{|V|^{n+1}-1}{|V|-1}$, that is, of order $|V(H)|^n \times |V(H)|^n$. By making judicious use of the equations (4.20) and the upcoming (4.25), this size can be significantly reduced if H has many edges; any practical computation should take this into account. Furthermore, it can be assumed that all matrix entries are actually in $\mathbb{R}$, i.e. no imaginary components are needed.

It is straightforward to see that the sets $\mathcal{Q}_n$ form a sequence or hierarchy. Indeed, every matrix M showing that p is a $\mathcal{Q}_{n+1}$-model can be restricted to a matrix showing that p is a $\mathcal{Q}_n$-model, hence $\mathcal{Q}_{n+1}(H) \subseteq \mathcal{Q}_n(H)$. Since each $\mathcal{Q}_n(H)$ is defined in terms of a semidefinite program, we say that this represents a **hierarchy of semidefinite programs**.

Moment matrices which are positive semidefinite and satisfy Normalization (4.20) and Orthogonality (4.21), also satisfy other useful properties, which I list below (for simplicity, I will omit the subscript n when writing M_n).

Remark 4.29. Let $\mathbf{v}, \mathbf{w}, \mathbf{v}', \mathbf{w}'$ be strings, and further denote the elements of $\mathbf{v}$ by $v_1 \ldots v_m$. Then,

1. If $\mathbf{v}\mathbf{w}^\dagger = \mathbf{v}'\mathbf{w}'^\dagger$, then
$$M(\mathbf{v}, \mathbf{w}) = M(\mathbf{v}', \mathbf{w}'). \tag{4.22}$$
This follows by induction from $M(v_1 \ldots v_m, \mathbf{w}) = M(v_1 \ldots v_{m-1}, \mathbf{w}v_m)$, which in turn is a consequence of (4.20) and (4.21) upon choosing some $e \ni v_m$,
$$\begin{aligned} M(v_1 \ldots v_m, \mathbf{w}) &\overset{(4.20)}{=} \sum_{x \in e} M(v_1 \ldots v_m, \mathbf{w}x) \overset{(4.21)}{=} M(v_1 \ldots v_m, \mathbf{w}v_m) \\ &\overset{(4.21)}{=} \sum_{x \in e} M(v_1 \ldots v_{m-1}x, \mathbf{w}v_m) \\ &\overset{(4.20)}{=} M(v_1 \ldots v_{m-1}, \mathbf{w}v_m). \end{aligned}$$
Equation (4.22) implies in particular that all matrix entries $M(\mathbf{v}, \mathbf{w})$ are determined by those of the "first row", i.e. those of the form $M(\emptyset, \mathbf{v})$, although this requires $\mathbf{v} \in V(H)^{*2n}$.

2. Repeating one letter in the index string gives the same matrix entry,

$$M(v_1 \dots v_i \dots v_m, \mathbf{w}) = M(v_1 \dots v_i v_i \dots v_m, \mathbf{w}). \tag{4.23}$$

Upon using (4.22), this follows from a very similar argument.

3. For every $e \in E(H)$,

$$\sum_{v_i \in e} M(\mathbf{v}, \mathbf{w}) = M(v_1 \dots \hat{v}_i \dots v_m, \mathbf{w}). \tag{4.24}$$

This is a consequence of (4.20) and (4.22).

4. Having subsequent orthogonal indices makes the matrix entry vanish,

$$v_i \perp v_{i+1} \quad \Longrightarrow \quad M(v_1 \dots v_i v_{i+1} \dots v_m, \mathbf{w}) = 0. \tag{4.25}$$

This follows from (4.24) together with (4.23).

5. Choosing some $e \ni v$ and applying (4.20) and (4.21) also shows that

$$M(v, \emptyset) = M(v, v). \tag{4.26}$$

In particular, if M is a certificate for a probabilistic model p, $p(v) = M(v, v)$ follows from (4.19).

It worth mentioning that in the case of infinite matrices M with entries $M(\mathbf{v}, \mathbf{w})$ indexed by strings of arbitrary length $\mathbf{v}, \mathbf{w} \in V(H)^*$, the definition of a moment matrix still holds if we take positive semidefiniteness to mean that

$$\sum_{\mathbf{v}, \mathbf{w} \in V(H)^*} x_{\mathbf{v}}^* M(\mathbf{v}, \mathbf{w}) x_{\mathbf{w}} \geq 0$$

for all finitely supported $(x_{\mathbf{v}})_{\mathbf{v} \in V(H)^*}$.

Proposition 4.30. *Let H be a contextuality scenario, and consider $p \in \mathcal{G}(H)$. If there exists such an infinite matrix as a certificate for p which satisfies Normalization (4.20) and Orthogonality (4.21), then $p \in \mathcal{Q}$.*

Proof. Such an infinite matrix M can be understood to be a ($*$-algebraic) state ϕ on the $*$-algebra with generators $\{P_v,\ v \in V(H)\}$ and relations

$$P_v = P_v^2 = P_v^*, \qquad \sum_{v \in e} P_v = \mathbb{1} \quad \forall e \in E(H) \tag{4.27}$$

via the assignment

$$\phi\left(P_{v_1}\dots P_{v_n}\right) := M(v_1\dots v_n,\emptyset).$$

and extending by linearity. Then, the GNS construction (see e.g. (KR83)) turns this into a quantum representation satisfying (4.19). For this reason, a probabilistic model is quantum if and only if there exists such an infinite matrix M which is a certificate for p having the properties of def. 4.27.

More concretely, this works as follows. First, we claim that

$$\sum_{\mathbf{v},\mathbf{w}\in V(H)^*} x_{\mathbf{v}}^* M(\mathbf{v}u,\mathbf{w}u) x_{\mathbf{w}} \leq \sum_{\mathbf{v},\mathbf{w}\in V(H)^*} x_{\mathbf{v}}^* M(\mathbf{v},\mathbf{w}) x_{\mathbf{w}}. \tag{4.28}$$

To see this, choose any $e \ni u$ and write

$$\begin{aligned}\sum_{\mathbf{v},\mathbf{w}\in V(H)^*} & x_{\mathbf{v}}^*\big(M(\mathbf{v},\mathbf{w}) - M(\mathbf{v}u,\mathbf{w}u)\big)x_{\mathbf{w}} \\ &\overset{4.29}{=} \sum_{\mathbf{v},\mathbf{w}\in V(H)^*} x_{\mathbf{v}}^* \left(\sum_{u'\in e,\, u'\neq u} M(\mathbf{v}u',\mathbf{w}u')\right) x_{\mathbf{w}} \\ &= \sum_{u'\in e,\, u'\neq u}\ \sum_{\mathbf{v},\mathbf{w}\in V(H)^*} x_{\mathbf{v}}^* M(\mathbf{v}u',\mathbf{w}u') x_{\mathbf{w}} \geq 0,\end{aligned}$$

where the last inequality is due to positive semidefiniteness of M. This proves (4.28).

Now we start the construction with the infinite-dimensional vector space spanned by all strings, $\mathcal{H}_0 := \mathrm{lin}_{\mathbb{C}}\left(V(H)^*\right)$. The formula

$$\left\langle \sum_{\mathbf{v}\in V(H)^*} x_{\mathbf{v}}\mathbf{v},\ \sum_{\mathbf{w}\in V(H)^*} y_{\mathbf{w}}\mathbf{w} \right\rangle := \sum_{\mathbf{v},\mathbf{w}\in V(H)^*} x_{\mathbf{v}}^* M(\mathbf{v},\mathbf{w}) y_{\mathbf{w}}.$$

defines a positive semidefinite inner product on $\mathcal{H}_0$ in terms of the matrix M. The Cauchy-Schwarz inequality shows that

$$\mathcal{N} := \left\{ \sum_{\mathbf{v}\in V(H)^*} x_{\mathbf{v}}\mathbf{v} \in \mathcal{H}_0 \;\middle|\; \left\langle \sum_{\mathbf{v}} x_{\mathbf{v}}\mathbf{v}, \sum_{\mathbf{v}} x_{\mathbf{v}}\mathbf{v} \right\rangle = 0 \right\}$$

is a linear subspace of $\mathcal{H}_0$. The inner product on the quotient $\mathcal{H}_0/\mathcal{N}$ then is positive definite by definition. We take $\mathcal{H}$ to be the completion of $\mathcal{H}_0/\mathcal{N}$ with respect to the norm coming from this inner product.

Now for $u \in V(H)$, the operator P_u is defined to act on $\mathcal{H}_0$ as

$$P_u \left(\sum_{\mathbf{v} \in V(H)^*} x_{\mathbf{v}} \mathbf{v} \right) := \sum_{\mathbf{v} \in V(H)^*} x_{\mathbf{v}} \mathbf{v} u.$$

Thanks to (4.28), this descends to a well-defined operator in $\mathcal{B}(\mathcal{H})$, which we also denote by P_u. The equation $M(\mathbf{v}u, \mathbf{w}) = M(\mathbf{v}, \mathbf{w}u)$ guarantees that P_u is self-adjoint, while $M(\mathbf{v}uu, \mathbf{w}) = M(\mathbf{v}u, \mathbf{w})$ shows that $P_u^2 = P_u$ since

$$\sum_{\mathbf{v} \in V(H)^*} x_{\mathbf{v}} \left(\mathbf{v}uu - \mathbf{v}u \right) \in \mathcal{N}.$$

The equation $\sum_{u \in e} P_u = \mathbb{1}_{\mathcal{H}}$ holds since

$$\sum_{\mathbf{v} \in V(H)^*} x_{\mathbf{v}} \left(\mathbf{v} - \sum_{u \in e} \mathbf{v}u \right) \in \mathcal{N}$$

thanks to (4.20). Finally, the rank-one density operator associated to the empty string $\emptyset \in \mathcal{H}$ is the desired quantum state, since

$$\langle \emptyset, P_u \emptyset \rangle = M(\emptyset, u) = p(u).$$

This ends the GNS construction. □

From this reasoning, we find that the sequence of sets $(\mathcal{Q}_n)_{n \in \mathbb{N}}$ converges in the following sense:

Theorem 4.31.

$$\mathcal{Q} = \bigcap_{n \in \mathbb{N}} \mathcal{Q}_n.$$

Proof ((NPA08)). It needs to be shown that if $p \in \mathcal{Q}_n$ for all $n \in \mathbb{N}$, then $p \in \mathcal{Q}$. To this end, we show that if a matrix $(M^n_{\mathbf{v},\mathbf{w}})_{\mathbf{v},\mathbf{w} \in V(H)^{*n}}$ exists with the required properties for every n, then there also exists a corresponding infinite matrix $(M^\infty_{\mathbf{v},\mathbf{w}})_{\mathbf{v},\mathbf{w} \in V(H)^*}$.

For $\mathbf{v} \in V(H)^{*n}$, positive semidefiniteness gives the estimate

$$\left(M^{2n}_{\mathbf{v},\mathbf{v}} \right)^2 \overset{(4.22)}{=} \left(M^{2n}_{\mathbf{v}\mathbf{v}^\dagger,\emptyset} \right)^2 \leq M^{2n}_{\mathbf{v}\mathbf{v}^\dagger,\emptyset} \cdot M^{2n}_{\emptyset,\emptyset} = M^{2n}_{\mathbf{v},\mathbf{v}},$$

which implies $M^{2n}_{\mathbf{v},\mathbf{v}} \leq 1$, and hence

$$|M^{2n}_{\mathbf{v},\mathbf{w}}|^2 \leq M^{2n}_{\mathbf{v},\mathbf{v}} M^{2n}_{\mathbf{w},\mathbf{w}} \leq 1$$

again thanks to positive semidefiniteness. We obtain $M^k_{\mathbf{v},\mathbf{w}} \in [-1,+1]$ for all $\mathbf{v},\mathbf{w} \in V(H)^{*n}$ with $n \leq 2k$.

Now consider the truncation of any M^{2n} to a matrix indexed by $\mathbf{v},\mathbf{w} \in V(H)^{*n}$. Upon filling this truncation up with 0's, we obtain an infinite matrix indexed by $\mathbf{v},\mathbf{w} \in V(H)^*$ with all elements in $[-1,+1]$. In this way, every matrix M^n becomes an element of the space $[-1,+1]^{V(H)^* \times V(H)^*}$. This space, equipped with the product topology, is second countable, and also compact thanks to Tychonoff's theorem. Hence, the sequence $(M^n)_{n\in\mathbb{N}}$ has a convergent subsequence, and we write M^∞ for its limit. By construction, this M^∞ is an infinite matrix indexed by $\mathbf{v},\mathbf{w} \in V(H)^*$ having all the desired properties. The claim now follows from Proposition 4.30. □

By means of the previous convergence theorem, we see that the hierarchy $\mathcal{Q}_n(H)$ of semidefinite programs characterizes $\mathcal{Q}(H)$.

One relevant set in this hierarchy is actually the first level $\mathcal{Q}_1$. In particular, in appendix F we prove that our set $\mathcal{Q}_1(B_{2,m,d})$ coincides with the set Q^{1+AB} of (NPA08). Moreover, as I will prove in section 4.7, $\mathcal{Q}_1$ also coincides with the set $\mathcal{E}^1_{\mathrm{QM}}$ defined in the CSW approach (CSW10). In the following proposition, I present other equivalent characterizations of $\mathcal{Q}_1$ which provide a better intuition on the properties of the set.

Proposition 4.32. *Let H be a contextuality scenario. For $p \in \mathcal{G}(H)$, the following are equivalent:*

1. $p \in \mathcal{Q}_1(H)$;
2. *There exists a Hilbert space $\mathcal{H}$, a unit vector $|\Psi\rangle \in \mathcal{H}$ and a vector $|\phi_v\rangle$ for every $v \in V(H)$ such that*
   a) $u \perp v \implies \langle\phi_u|\phi_v\rangle = 0$,
   b) $\sum_{v\in e} |\phi_v\rangle = |\Psi\rangle \quad \forall e \in E(H)$,
   c) $p(v) = \langle\phi_v|\phi_v\rangle$;
3. *There exists a Hilbert space $\mathcal{H}$, a unit vector $|\Psi\rangle \in \mathcal{H}$ and a unit vector $|\psi_v\rangle$ for every $v \in V(H)$ such that*
   a) $u \perp v \implies \langle\psi_u|\psi_v\rangle = 0$,
   b) $p(v) = |\langle\psi_v|\Psi\rangle|^2$;
4. *There exists a Hilbert space $\mathcal{H}$, a unit vector $|\Psi\rangle \in \mathcal{H}$ and a projection P_v for every $v \in V(H)$ such that*
   a) $u \perp v \implies P_u \perp P_v$,

b) $p(v) = \langle\Psi|P_v|\Psi\rangle \quad \forall v \in V(H)$*;*

5. There exists a Hilbert space $\mathcal{H}$*, a unit vector* $|\Psi\rangle \in \mathcal{H}$ *and a projection* P_v *for every* $v \in V(H)$ *such that*

a) $\sum_{v\in e} P_v \leq \mathbb{1}_{\mathcal{H}} \quad \forall e \in E(H)$*,*

b) $p(v) = \langle\Psi|P_v|\Psi\rangle \quad \forall v \in V(H)$*;*

In all cases, $\mathcal{H}$ *can also be taken to be the real Hilbert space* $\mathbb{R}^{|V(H)|}$*.*

Proof. 1⇒2: By positive semidefiniteness, we can write M as a Gram matrix, so that there exist vectors $|\Psi\rangle, |\phi_v\rangle$ in $\mathcal{H} = \mathbb{R}^{|V(H)|}$ such that

$$M(\emptyset,\emptyset) = \langle\Psi|\Psi\rangle, \qquad M(\emptyset, v) = \langle\Psi|\phi_v\rangle, \qquad M(u,v) = \langle\phi_u|\phi_v\rangle,$$

from which 2a and 2c follow.

Now we fix $e \in E(H)$ and show 2b. We decompose $|\Psi\rangle$ into orthogonal components $|\Psi\rangle = |\Psi^{\|}\rangle + |\Psi^{\perp}\rangle$, where $|\Psi^{\|}\rangle \in \mathrm{lin}_{\mathbb{C}}\{|\phi_v\rangle \,:\, v \in e\}$. Due to (4.20) and (4.21), the vectors satisfy:

$$\langle\phi_v|\Psi\rangle = M(v,\emptyset) = \sum_{u\in e} M(v,u) = M(v,v) = \langle\phi_v|\phi_v\rangle.$$

Then the equations

$$\langle\phi_v|\Psi\rangle = \langle\phi_v|\phi_v\rangle$$

imply that $|\Psi^{\|}\rangle = \sum_{v\in e}|\phi_v\rangle$. On the other hand,

$$\langle\Psi^{\|}|\Psi^{\|}\rangle + \langle\Psi^{\perp}|\Psi^{\perp}\rangle = M(\emptyset,\emptyset) = \sum_{v\in e} M(\emptyset, v)$$

$$= \sum_{v,u\in e} M(u,v) = \sum_{v,u\in e}\langle\phi_u|\phi_v\rangle = \langle\Psi^{\|}|\Psi^{\|}\rangle$$

shows that $|\Psi^{\perp}\rangle = 0$, so that $\sum_{v\in e}|\phi_v\rangle = |\Psi\rangle$, as desired.

2⇒3: Normalizing the $|\phi_v\rangle$ to $|\psi_v\rangle := \frac{1}{\sqrt{\langle\phi_v|\phi_v\rangle}}|\phi_v\rangle$ guarantees the orthogonality relations, and choosing some edge $e \in E(H)$ with $v \in e$ gives

$$|\langle\psi_v|\Psi\rangle|^2 = \frac{1}{\langle\phi_v|\phi_v\rangle}\left|\left\langle \phi_v \middle| \sum_{u\in e}\phi_u \right\rangle\right|^2 = \frac{1}{\langle\phi_v|\phi_v\rangle}\langle\phi_v|\phi_v\rangle^2 = \langle\phi_v|\phi_v\rangle,$$

due to the orthogonality relations.

3⇒4: Define $P_v = |\psi_v\rangle\langle\psi_v|$.

4⇒5: This is clear since for fixed $e \in E(H)$, all projections P_v for $v \in e$ are mutually orthogonal, which implies $\sum_{v\in e} P_v \leq \mathbb{1}_{\mathcal{H}}$.

5⇒1: Define $M(v, w) = \langle\Psi|P_v P_w^\dagger|\Psi\rangle$. We check that M satisfies conditions (4.20) and (4.21), $M(\emptyset,\emptyset) = 1$, and is positive semidefinite:

$M(\emptyset,\emptyset) = 1$: $M(\emptyset,\emptyset) = \langle\Psi|\Psi\rangle = 1$, since $|\Psi\rangle$ is a unit vector.

(4.20) Consider an edge $e \in E$. Since $p(v)$ is a probabilistic model,

$$\langle\Psi|\Psi\rangle = 1 = \sum_{v\in e} p(v) = \langle\Psi|\sum_{v\in e} P_v|\Psi\rangle,$$

which implies $\sum_{v\in e} P_v|\Psi\rangle = |\Psi\rangle$. Then,

$$\sum_{v\in e} M(v, w) = \langle\Psi|\sum_{v\in e} P_v P_w|\Psi\rangle = \langle\Psi|P_w|\Psi\rangle = M(\emptyset, w).$$

(4.21) If $v \perp w$, then there is an edge $e \in E(H)$ with $v, w \in e$. Hence, $P_v \perp P_w$, so that $M(v, w) = \langle\Psi|P_v P_w|\Psi\rangle = 0$.

$\succeq 0$: It needs to be shown that for any vector $x \in \mathbb{C}^{V(H)^{*n}}$ with components $x_{\mathbf{v}} \in \mathbb{C}$, $\mathbf{v} \in V(H)^{*n}$, the expression

$$\sum_{\mathbf{v},\mathbf{w}} x_{\mathbf{v}}^* M(\mathbf{v}, \mathbf{w}) x_{\mathbf{w}}.$$

is nonnegative. By the definition, this is equal to

$$\sum_{\mathbf{v},\mathbf{w}} \langle\Psi|\, x_{\mathbf{v}}^* P_{\mathbf{v}} P_{\mathbf{w}}^\dagger x_{\mathbf{w}}\, |\Psi\rangle.$$

With $Q = \sum_{\mathbf{v}} x_{\mathbf{v}} P_{\mathbf{v}}^\dagger$, this is of the form $\langle\Psi|\, Q^\dagger Q\, |\Psi\rangle$, and therefore indeed nonnegative.

□

Finally, I present a necessary and sufficient condition for a model to belong to $\mathcal{Q}_1$, in terms of a graph-theoretic invariant from section 2.3. Prop. 4.32 proposes a characterization (2) of $\mathcal{Q}_1$ in terms of unit vectors $\{|\phi_v\rangle\}_{v\in V}$, which satisfy orthogonality relations similar to those of an orthonormal representation of $\mathrm{NO}(H)$. In what follows, we will see that the weighted Lovász number ϑ succeeds in picking up the characteristic properties of the probabilistic models in $\mathcal{Q}_1$.

Proposition 4.33. *A probabilistic model $p \in \mathcal{G}(H)$ is in $\mathcal{Q}_1$ if and only if $\vartheta(\mathrm{NO}(H), p) \leq 1$.*

Proof. We use the characterization of $\mathcal{Q}_1(H)$ given in Proposition 4.32-3. Assuming $p \in \mathcal{Q}_1(H)$, we choose corresponding vectors $|\psi_v\rangle, |\Psi\rangle \in \mathbb{R}^{|V(H)|}$; then, by Definition A.15,

$$\vartheta(\mathrm{NO}(H), p) \leq \max_{v \in V} \frac{p(v)}{|\langle\Psi|\psi_v\rangle|^2} = \frac{p(v)}{p(v)} = 1.$$

Conversely, if $\vartheta(\mathrm{NO}(H), p) \leq 1$, then there is an orthonormal labelling $(|\psi_v\rangle)_{v \in V}$ and a vector $|\Psi\rangle \in \mathbb{R}^{|V|}$ such that $|\langle\Psi|\psi_v\rangle|^2 \geq p(v)\ \forall v$. By choosing $\mathcal{H} = \mathbb{R}^{|V(H)|} \oplus \mathbb{R}^{|V(H)|}$ and setting

$$|\psi'_v\rangle := \frac{\sqrt{p(v)}}{|\langle\Psi|\psi_v\rangle|}|\psi_v\rangle \oplus \sqrt{1 - \frac{p(v)}{|\langle\Psi|\psi_v\rangle|^2}}|e_v\rangle \ \in \mathcal{H}$$

where the $|e_v\rangle$ form the standard basis of $\mathbb{R}^{|V(H)|}$, one obtains $|\langle\Psi|\psi'_v\rangle|^2 = p(v)$ with unit vectors $|\psi'_v\rangle$, as desired. □

This relation to graph theory has a simple first application:

Proposition 4.34.

$$\mathcal{Q}_1(H_A) \otimes \mathcal{Q}_1(H_B) \subseteq \mathcal{Q}_1(H_A \otimes H_B) \tag{4.29}$$

Proof. Combine Proposition 4.33 with

If $p_A \in \mathcal{Q}_1(H_A)$ and $p_B \in \mathcal{Q}_1(H_B)$ then both $\vartheta(\mathrm{NO}(H_A), p_A) \leq 1$ and $\vartheta(\mathrm{NO}(H_B), p_B) \leq 1$. Hence, multiplicativity of ϑ (corollary A.25) implies that $\vartheta(G_1 \boxtimes G_2, p_1 \otimes p_2) = \vartheta(G_1, p_1)\vartheta(G_2, p_2) \leq 1$. It follows that $p_1 \otimes p_2 \in \mathcal{Q}_1(H_A \otimes H_B)$. □

4.7. Relation to the CSW approach

The Cabello, Severini and Winter approach to contextuality (CSW10) also has it basis in a graph-theoretic formulation, and has in part inspired the formalism presented in this chapter. In this section, I will elaborate on the connection between the two approaches.

The main difference between the approaches lies in the normalization of the probabilistic models: while we demand that $\sum_{v \in e} p(v) = 1$ for every measurement $e \in E$, CSW rather asks $\sum_{v \in e} p(v) \leq 1$, i.e. their measurements may not be complete. In their approach this is not an issue, due the kind of problems they focus on. The question then is how to compare these two approaches which differ in such a basic property. In what follows, I present two connections.

First, consider the set $\mathcal{E}^1_{\rm QM}$ defined in (CSW10). This set contains all the quantum assignments (in the CSW sense) that further satisfy normalization. Rephrasing (CSW10) and section 2.2,

Definition 4.35. *Let $H = (V, E)$ be a contextuality scenario. An assignment $p : V \to [0, 1]$ belongs to the set $\mathcal{E}^1_{\rm QM}$ if there exists a Hilbert space $\mathcal{H}$, a state $\rho \in \mathcal{B}_{+,1}(\mathcal{H})$, and projectors $P_v \in \mathcal{H}$ for every vertex $v \in V$, such that*

1. *if $u, v \in e$ for some hyperedge $e \in E$, then $P_u \perp P_v$.*
2. *$\sum_{v\in e} P_v \leq \mathbb{1}_{\mathcal{H}}$ for every hyperedge $e \in E$.*
3. *$p(v) = \mathrm{tr}(\rho\, P_v)$ for every vertex $v \in V$.*
4. *$\sum_{v\in e} p(v) = 1$ for every hyperedge $e \in E$.*

Note that the requirement that the assignment is normalized is not in contradiction with the subnormalized character of the projectors, since they may still be normalized in the subspace where the state ρ has support on.

By comparing the previous definition with proposition 4.32, it follows that def. 4.35 is another characterization of the set $\mathcal{Q}_1$. Hence, we see that the set of normalized quantum assignments $\mathcal{E}^1_{\rm QM}(H)$ in the CSW formalism is equivalent to the first level in our hierarchy $\mathcal{Q}_1(H)$. In particular, this applies to Bell scenarios, which were introduced in CSW by defining the sets $\mathcal{E}^1_{\rm X}$. Hence quantum models on a Bell scenario in CSW satisfy $\mathcal{E}^1_{\rm QM}((B_{n,m,d})) \equiv \mathcal{Q}_1(B_{n,m,d})$.

The CSW formalism however focuses on contextuality scenarios, where there is no need a priori to demand normalization of the assignments p. The set of not normalized quantum models $\mathcal{E}_{\rm QM}$ is then defined as in def. 4.35 without the last requirement, and hence cannot be directly studied in our formalism. In what follows we show, given a CSW scenario H together with a CSW-quantum assignment p, how to construct a scenario H' where p extends to H' as a quantum model. The sketch of the construction is the following. The main idea is to add to each hyperedge a new vertex, referred to as "no-detection event". Given an assignment $p \in \mathcal{E}_{\rm QM}$, there are many ways to consistently assign projectors to these no-detection events, such that the extended model p' is normalized, and all of them correspond to models in $\mathcal{Q}_1$. However, there is one that corresponds to a model in $\mathcal{Q}$. Note that this construction does not apply to Bell scenarios, since in that case the possible events are uniquely determined by the number of parties, measurements and outcomes, and no no-detection events can be added to the hyperedges. This idea is formalized in the following proposition.

Proposition 4.36. *Let H be a contextuality scenario and $p \in \mathcal{E}_{\text{QM}}(H)$. Then, there exists a contextuality scenario H' such that the extension p' of p to H' belongs to $\mathcal{Q}(H')$.*

Proof. Construct a contextuality scenario H' from H by adding for each $e \in E(H)$ one **no-detection event** w_e,

$$V(H') := V(H) \cup \{w_e \,:\, e \in E(H)\}, \qquad E(H') := \{e \cup \{w_e\} \,:\, e \in E(H)\}\,.$$

The assignment p is extended to p' like follows,

$$p'(v) := p(v) \quad \text{if } v \in V, p'(w_e) := 1 - \sum_{v \in e} p(v).$$

In the particular case where p is normalized, i.e. when $p \in \mathcal{E}^1_{\text{QM}}(H) \subset \mathcal{E}_{\text{QM}}(H)$, the no-detection events have probability 0, which justifies the name. Moreover, in this case the notion of "extension" of p to H' coincides with the definition of extension of section 4.1.

Now we will see that p' is a quantum model on H'. Consider the Hilbert space $\mathcal{H}$, the state ρ and projectors P_v, $v \in V$, in the definition of p. Define:

$$P'_v := P_v \quad \text{if } v \in V, P'_{w_e} := 1 - \sum_{v \in e} P_v.$$

First, since $\sum_{v \in e} P_v \leq \mathbb{1}_{\mathcal{H}}$, the projectors P'_{w_e} are well defined and satisfy $P'_{w_e} \perp P'_v$ for all $v \in V$. Second, the completeness relation for the hyperedges in $E(H')$ holds for definition. Third,

$$\operatorname{tr}(\rho\, P'_{w_e}) = \operatorname{tr}(\rho) - \sum_{v \in e} \operatorname{tr}(\rho\, P_v) = 1 - \sum_{v \in e} p(v) = p'(w_e).$$

Hence, $p' \in \mathcal{Q}(H')$ □

To summarize, given a CSW-quantum assignment $p \in \mathcal{E}_{\text{QM}}$,

1. if it is normalized, i.e. $p \in \mathcal{E}^1_{\text{QM}}$, it is a $\mathcal{Q}_1$ model. This is the case for CSW-Bell scenarios. Note that these p are thus not necessary quantum.

2. if it does not correspond to a model on a Bell scenario, it can be extended to a model p' on a larger hypergraph such that p' is quantum.

This construction of a quantum model from a CSW-quantum model will play a key role in the next section. Further properties of hypergraphs equipped with no-detection events are presented in section 4.11.

4.8. Consistent Exclusivity and Local Orthogonality

The search for principles to bound the set of quantum models $\mathcal{Q}$ also extends to contextuality scenarios. One such a candidate is the "Consistent Exclusivity" principle (CE), defined as follows:

Definition 4.37 ((Hen12)). *A probabilistic model $p \in \mathcal{G}(H)$ satisfies **Consistent Exclusivity** if*

$$\sum_{v \in I} p(v) \leq 1 \tag{4.30}$$

holds for any independent set $I \subseteq V(\mathrm{NO}(H))$. We write $\mathscr{CE}^1(H) \subseteq \mathcal{G}(H)$ for the set of probabilistic models satisfying CE.

We also write CE^1 for this version of CE in order to distinguish it from the upcoming refinement termed CE^n. Intuitively, CE is saying that the total probability of any collection of pairwise exclusive outcomes is upper-bounded by 1. In this formulation, CE may almost sound like a trivial consequence of the laws of probability. However, this is not the case, since the probabilities $p(v)$ of a probabilistic model are *conditional* probabilities representing the probability that outcome v occurs *given that* a measurement e with $v \in e$ has been performed: in general such a collection of pairwise orthogonal events is not necessarily jointly exclusive (see section 3.1 for a similar argument).

This principle considers general contextuality scenarios, hence in our formalism it also applies to Bell scenarios. In chapter 3 I presented the "Local Orthogonality" principle for quantum correlations, which after introducing a notion of orthogonality in Bell scenarios, imposes the same constraint of eq. 4.30 as in CE. The natural question then is how do this two principles relate. As proved in Lemma 4.17, the orthogonality relations that arise in our definition of a Bell scenario are exactly those imposed by the LO principle, hence in our framework the Consistent Exclusivity principle and the Local Orthogonality principle are equivalent.

Note that the set of quantum models satisfy CE^1. Indeed, if $p \in \mathcal{Q}(H)$, the projectors in $\{P_v\}_{v \in I}$ are pairwise orthogonal for every independent set $I \subseteq V(\mathrm{NO}(H))$. This implies that $\sum_{v \in I} P_v \leq \mathbb{1}_{\mathcal{H}}$, hence condition (4.30) is automatically satisfied.

The triangle scenario Δ of Fig. 4.1 is an example of how non trivial the CE principle is. Indeed, $V(\Delta)$ is itself an independent set in $\mathrm{NO}(\Delta)$, hence its unique probabilistic model $p = \frac{1}{2}$ violates CE: $\sum_{v \in V(\Delta)} p(v) = \frac{3}{2}$. We see that $\mathscr{CE}^1(\Delta) = \emptyset$, although $\mathcal{G}(\Delta) = \{p\}$.

Similar to the case of Local Orthogonality (see section 3.3), we define a hierarchy of CE sets as follows.

Definition 4.38 (CE hierarchy of sets). *Let H be a contextuality scenario and $p \in \mathcal{G}(H)$. We write $p \in \mathscr{CE}^n(H)$ if and only if $p^{\otimes n} \in \mathscr{CE}^1(H^{\otimes n})$. Furthermore,*

$$\mathscr{CE}^\infty(H) := \bigcap_{n \in \mathbb{N}} \mathscr{CE}^n(H).$$

If $p \in \mathscr{CE}^k(H)$, then we also say that p satisfies CE^k. In particular, $p \in \mathscr{CE}^\infty(H)$ if and only if $p \in \mathscr{CE}^n(H)$ for all $n \in \mathbb{N}$, in which case we say that p satisfies CE^∞. From the observation above, it also follows that $\mathscr{CE}^k(B_{n,m,d}) = \mathcal{LO}^k(n, m, d)$.

We now relate the $\mathscr{CE}^*$ family of sets to the graph-theoretical invariants of section 2.3. First, note that $\mathscr{CE}^1$ imposes constraints on the total weight of any independent set in $\text{NO}(H)$, hence it is natural to relate it to the weighted independence number α. The other sets $\mathscr{CE}^k$ then impose constraints on the total weight of the independent sets of $\text{NO}(H^{\otimes k})$, thus we will relate them to the independence number of the orthogonality graph for the corresponding product scenarios. Finally, since the Shannon capacity is a limit instance of the independence number, it is natural to relate it to the limit set $\mathscr{CE}^\infty$.

Lemma 4.39.

1. *$p \in \mathscr{CE}^n(H)$ if and only if $\alpha(\text{NO}(H)^{\boxtimes n}, p^{\otimes n}) \leq 1$.*
2. *$p \in \mathscr{CE}^\infty(H)$ if and only if $\Theta(\text{NO}(H), p) \leq 1$, or, equivalently,*

   if $\alpha(\text{NO}(H), p) = \Theta(\text{NO}(H), p) = 1$.

Proof.

1. By definition, $p \in \mathscr{CE}^n(H)$ if and only if $p^{\otimes n} \in \mathscr{CE}^1(H^{\otimes n})$, i.e. $\sum_{v \in I} p^{\otimes n}(v) \leq 1$, where I is an independent set on $H^{\otimes n}$. This implies that $\alpha(\text{NO}(H^{\otimes n}), p^{\otimes n}) \leq 1$. The claim now follows from Lemma 4.18, which states that $\text{NO}(H^{\otimes n}) = \text{NO}(H)^{\boxtimes n}$.

2. If $p \in \mathscr{CE}^\infty(H)$, then $p \in \mathscr{CE}^n(H)$ for all n. Hence, $\alpha(\text{NO}(H)^{\boxtimes n}, p^{\otimes n}) \leq 1$ for all n, which by the definition of Θ (A.6) implies that $\Theta(\text{NO}(H), p) \leq 1$.

   For the converse, start from $\Theta(\text{NO}(H), p) \leq 1$ and assume there exists a k such that $\alpha(\text{NO}(H)^{\boxtimes k}, p^{\otimes k}) > 1$. From corollary A.26 follows that $\alpha(\text{NO}(H)^{\boxtimes k*m}, p^{\otimes k*m}) \geq \alpha(\text{NO}(H)^{\boxtimes k}, p^{\otimes k})^m$, hence

   $$\sqrt[km]{\alpha(\text{NO}(H)^{\boxtimes k*m}, p^{\otimes k*m})} \geq \sqrt[k]{\alpha(\text{NO}(H)^{\boxtimes k}, p^{\otimes k})} > 1$$

   for all m. This implies that the limit of the sequence is larger than one, i.e. $\Theta(\text{NO}(H), p) > 1$, which contradicts the original assumption.

   From corollary A.24 we know that $\alpha(\text{NO}(H), p) \leq \Theta(\text{NO}(H), p)$. Since p is normalized, $\alpha(\text{NO}(H), p) \geq 1$, and since p satisfies CE^∞, $\Theta(\text{NO}(H), p) \leq 1$. This implies $1 = \alpha(\text{NO}(H), p) = \Theta(\text{NO}(H), p) = 1$.

□

These sets of CE models and the quantum quantum models are related as follows.

Lemma 4.40. *For every $k, n \in \mathbb{N}$, the following inclusions hold:*

$$\mathscr{CE}^\infty(H) \subseteq \ldots \subseteq \ldots \mathscr{CE}^n(H) \subseteq \ldots \subseteq \mathscr{CE}^1(H).$$

Moreover, $\mathcal{Q}(H) \subseteq \mathscr{CE}^\infty(H)$.

Proof. We choose any $p \in \mathscr{CE}^1(H)$. Thanks to Corollary A.26, we know

$$\alpha(\mathrm{NO}(H)^{\boxtimes n}, p^{\otimes n}) \geq \alpha(\mathrm{NO}(H)^{\boxtimes(n-1)}, p^{\otimes(n-1)}) \cdot \alpha(\mathrm{NO}(H), p).$$

In addition, $\alpha(\mathrm{NO}(H), p) = 1$ since p is normalized, which implies that the sequence $\left(\alpha(\mathrm{NO}(H)^{\boxtimes n}, p^{\otimes n})\right)_{n\in\mathbb{N}}$ is monotonically nondecreasing. The first claim now follows from Lemma 4.39.

Consider now $p \in \mathcal{Q}(H)$. The last remark in section 4.5 implies that $p^{\otimes n} \in \mathcal{Q}(H)^{\otimes n} \subset \mathcal{Q}(H^{\otimes n}) \subset \mathscr{CE}^1(H^{\otimes n})$ for all n. Hence $p \in \mathscr{CE}^\infty(H)$. □

In chapter 3 I mentioned that Local Orthogonality does not recover the set of quantum correlations, a fact that was first noticed by Miguel Navascués before this formalism had been set up. In what follows I show that his statement still holds for general contextuality scenarios.

Proposition 4.41 (Navascués)**.** *For every H,*

$$\mathcal{Q}_1(H) \subseteq \mathscr{CE}^\infty(H). \tag{4.31}$$

Proof. Proposition 4.33 states that $p \in \mathcal{Q}_1$ if and only if $\vartheta(\mathrm{NO}(H), p) \leq 1$, while by Lemma 4.39 $p \in \mathscr{CE}^\infty(H)$ if and only if $\Theta(\mathrm{NO}(H), p) \leq 1$. Corollary A.24 states that $\Theta(\mathrm{NO}(H), p) \leq \vartheta(\mathrm{NO}(H), p)$, hence every $\mathcal{Q}_1$ model satisfies CE^∞. □

In particular, together with $\mathcal{Q}(H) \subseteq \mathcal{Q}_1(H)$, this proves that $\mathcal{Q}(H) \subseteq \mathscr{CE}^\infty(H)$. All together, the sets of probabilistic models defined in this chapter satisfy this chain of inclusions depicted in Fig. 4.7.

In (NPA08), the authors prove that $\mathcal{Q}(B_{2,2,2}) \subsetneq \mathcal{Q}_1(B_{2,2,2})$, which implies $\mathcal{Q}(B_{2,2,2}) \subsetneq \mathscr{CE}^\infty(B_{2,2,2})$. Hence, already in the CHSH scenario, the LO principle does not characterize quantum models. A natural question is whether this happens only for Bell scenarios, i.e. if other types of contextuality scenarios may be characterized by CE. The following theorem proves this is not the case.

Theorem 4.42. *There are contextuality scenarios H for which*

$$\mathcal{Q}_1(H) \subsetneq \mathscr{CE}^\infty(H).$$

$$
\begin{array}{ccccccccccccccc}
\mathcal{C} & \subseteq & \mathcal{Q} & \overset{4.31}{=} & \mathcal{Q}_\infty & \subseteq & \cdots & \subseteq & \mathcal{Q}_1 & \overset{4.41}{\subseteq} & \mathscr{CE}^\infty & \overset{4.40}{\subseteq} & \cdots & \overset{4.40}{\subseteq} & \mathscr{CE}^1 & \subseteq & \mathcal{G} \\
\updownarrow{\scriptstyle 4.23} & & & & & & & & \updownarrow{\scriptstyle 4.33} & & \updownarrow{\scriptstyle 4.39} & & & & \updownarrow{\scriptstyle 4.39} & & \\
\alpha^* & & & & \geq & & & & \vartheta & \geq & \Theta & & \geq & & \alpha & &
\end{array}
$$

Figure 4.7.: Chain of inclusions between sets of probabilistic models and corresponding inequalities between graph invariants.

Proof. Our Proposition 4.33 and Lemma 4.39 suggests that this is related to the existence of graphs G for which $\alpha(G) = \Theta(G) < \vartheta(G)$. Indeed, we will turn Haemers' example (Hae81) of this phenomenon into an example of a contextuality scenario J_n with a probabilistic model $p_J \in \mathscr{CE}^\infty(H)$ with $p_J \notin \mathcal{Q}_1(H)$.

Let $n \geq 12$ be an integer divisible by 4. Let J_n have vertices $V(J_n)$ being all 3-element subsets of $\{1, \ldots, n\}$. An edge of J_n is given in terms of a partition of $\{1, \ldots, n\}$ into 4-element subsets; a vertex (3-element subset) belongs to the edge if and only if it is contained in one of the subsets of the partition.

By construction, all $e \in E(J_n)$ have cardinality $|e| = n$, since every partition consists of $n/4$ subsets and each subset hosts 4 vertices. Therefore, assigning a weight of $\frac{1}{n}$ to each vertex defines a probabilistic model p_J. Now the non-orthogonality graph $\mathrm{NO}(H_J)$ consists of the 3-element subsets of $\{1, \ldots, n\}$ two of which are adjacent if and only if they have exactly one element in common. This is the graph that was considered by Haemers (Hae81), who showed that

$$\alpha(\mathrm{NO}(H_J)) = \Theta(\mathrm{NO}(H_J)) = n < \vartheta(\mathrm{NO}(H_J)).$$

Since the probabilistic model p_J has constant weights $\frac{1}{n}$, this means that

$$\alpha(\mathrm{NO}(H_J), p_J) = \Theta(\mathrm{NO}(H_J), p_J) = 1 < \vartheta(\mathrm{NO}(H_J), p_J),$$

and hence $p_J \in \mathscr{CE}^\infty(H_J)$, but $p_J \notin \mathcal{Q}_1(H_J)$. □

We now move on to study under which conditions $\mathcal{C}(H)$ coincides with $\mathscr{CE}^1(H)$. This is an interesting case, since it implies that the whole hierarchy of sets – but the general probabilistic models $\mathcal{G}$ – collapse into the classical one. This means not only that Consistent Exclusivity recovers the quantum set for these scenarios, but also that these quantum models do not exhibit contextual features. In order to study this phenomenon, I will first present the concept of *perfect graphs*, and then relate them to it.

Proposition 4.43. *A graph G is called* **perfect** *if the chromatic number*[1] *of any induced subgraph is equal to the clique number of this subgraph (Ber61).*
If $\mathrm{NO}(H)$ *is perfect, then* $\mathcal{C}(H) = \mathscr{CE}^1(H)$.

Proof. By the weak perfect graph theorem of Lovász (Lov72), we can as well assume the complement $\overline{\mathrm{NO}(H)}$ to be perfect. A probabilistic model $p \in \mathscr{CE}^1(H)$ can be interpreted as vertex weights $p(v)$ for $v \in V(H)$ with $\sum_{v \in C} p(v) \leq 1$ for every clique C in $\overline{\mathrm{NO}(H)}$. Then, perfection guarantees (Knu94, Thm. 31) that p is a convex combination of indicator functions of independent sets in $\overline{\mathrm{NO}(H)}$, i.e. there are cliques $U_1, \ldots, U_k$ in $\mathrm{NO}(H)$ and coefficients $\lambda_i \in [0,1]$ with $\sum_i \lambda_i = 1$ such that

$$p = \sum_{i=1}^{k} \lambda_i \mathbb{1}_{U_i}. \tag{4.32}$$

We now claim that every $\mathbb{1}_{U_i}$ is a deterministic model. Since its weights clearly take values in $\{0,1\}$, it is enough to verify the normalization condition $\sum_{v \in e} \mathbb{1}_{U_i}(v) = 1$ for all $e \in E(H)$. But this follows from (4.32) together with $\sum_{v \in e} p(v) = 1$. □

In (CRST06) the authors prove that a graph G is perfect if and only if neither G nor $\overline{G}$ contains an induced subgraph which is a cycle of odd length ≥ 5. Hence, if neither $\mathrm{NO}(H)$ nor $\overline{\mathrm{NO}(H)}$ contains an odd cycle of length ≥ 5 as an induced subgraph, then $\mathcal{C}(H) = \mathscr{CE}^1(H)$.

On the other hand, the converse to proposition 4.43 is not true:

Proposition 4.44. *For the scenario depicted in Figure 4.8,* $\mathcal{G}(H_0) = \mathcal{C}(H_0)$*, although* $\mathrm{NO}(H_0)$ *is not perfect.*

Proof. $\mathrm{NO}(H_0)$ is not perfect since its complement $\overline{\mathrm{NO}(H_0)}$ contains the pentagon ⬠ as an induced subgraph in the left part.

On the other hand, every probabilistic model p on H_0 is guaranteed to satisfy $p(v) = 1$ due to the structure on the right. Hence, $p(u) = 0$ for all u in the pentagon. Therefore, both $\mathcal{G}(H_0)$ and $\mathcal{C}(H_0)$ can be identified with their counterparts for the right part H_R of Figure 4.8. Since every maximal independent set in $\mathrm{NO}(H_R)$ is itself an edge, we get $\mathscr{CE}^1(H_R) = \mathcal{G}(H_R)$, and since $\mathrm{NO}(H_R)$ is perfect, we have $\mathcal{C}(H_R) = \mathscr{CE}^1(H_R)$. □

[1] The chromatic number of a graph, usually denoted by χ, is the smallest number of colors required to color the vertices of a graph such that adjacent vertices do not have the same color.

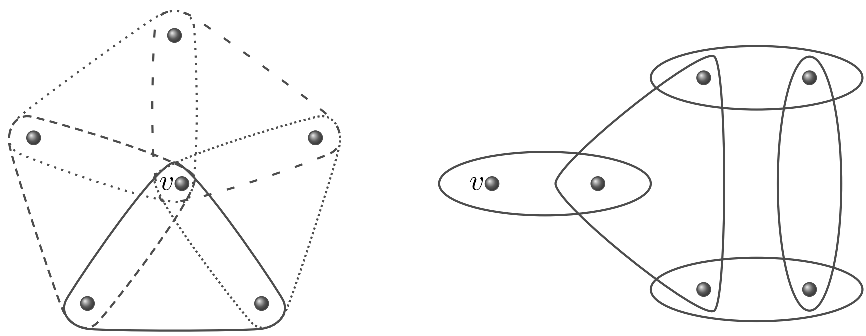


Figure 4.8.: A scenario H_0 with $\mathcal{G}(H_0) = \mathcal{C}(H_0)$, although $\mathrm{NO}(H_0)$ is not perfect. The two nodes labelled v represent the same vertex.

4.9. Non-convexity of $\mathscr{CE}^\infty$

Since $\mathscr{CE}^1$ is defined in terms of linear inequalities, it is naturally convex. However, the situation changes for $\mathscr{CE}^n$ for $n \geq 2$ and $\mathscr{CE}^\infty$. In this section, I will study some properties regarding the convexity of the CE sets. In particular, I will prove, by constructing a counter-example, that the set $\mathscr{CE}^\infty$ is not convex in general.

Theorem 4.45. *For all contextuality scenarios H, H_A, H_B, the following statements are equivalent:*

1. $\mathscr{CE}^\infty(H_A) \otimes \mathscr{CE}^\infty(H_B) \subseteq \mathscr{CE}^1(H_A \otimes H_B)$*;*

2. $\mathscr{CE}^\infty(H_A) \otimes \mathscr{CE}^\infty(H_B) \subseteq \mathscr{CE}^\infty(H_A \otimes H_B)$*;*

In addition, they both imply:

3. $\mathscr{CE}^\infty(H)$ *is convex.*

Proof. Property 2 clearly implies 1. For the converse, suppose that we have $p_A \in \mathscr{CE}^\infty(H_A)$ and $p_B \in \mathscr{CE}^\infty(H_B)$ with $p_A \otimes p_B \notin \mathscr{CE}^\infty(H_A \otimes H_B)$. Then there exists some $n \in \mathbb{N}$ with $(p_A \otimes p_B)^{\otimes n} \notin \mathscr{CE}^1(H_A^{\otimes n} \otimes H_B^{\otimes n})$. This would mean that $p_A^{\otimes n} \in \mathscr{CE}^\infty(H_A^{\otimes n})$ and $p_B^{\otimes n} \in \mathscr{CE}^\infty(H_B^{\otimes n})$ was a counterexample to 1.

Concerning the implication from 1 to3, we consider $p_1, p_2 \in \mathscr{CE}^\infty(H)$ and deduce $p_1^{\otimes k} \otimes p_2^{\otimes(n-k)} \in \mathscr{CE}^1(H^{\boxtimes n})$ from assumption 1. Due to convexity of

$\mathscr{CE}^1(H^{\otimes n})$, this shows that for any $\lambda \in (0,1)$,

$$(\lambda p_1 + (1-\lambda)p_2)^{\otimes n} = \sum_{k=0}^{n} \binom{n}{k} \lambda^k (1-\lambda)^{n-k} p_1^{\otimes k} \otimes p_2^{\otimes (n-k)} \in \mathscr{CE}^1(H^{\otimes n}),$$

so that $(\lambda p_1 + (1-\lambda)p_2) \in \mathscr{CE}^n(H)$. Since n was arbitrary, this means that $(\lambda p_1 + (1-\lambda)p_2) \in \mathscr{CE}^\infty(H)$, as was to be shown. □

In what follows, we prove that $\mathscr{CE}^\infty$ is not convex in general, which implies that the statements in Theorem 4.45 are not always satisfied. Briefly, the sketch for the construction of an explicit counter-example goes as follows:

1. Consider a contextuality scenario H_A together with a probabilistic model p_A such that $p_A \in \mathscr{CE}^\infty(H_A) \setminus \mathcal{Q}_1(H_A)$. An explicit construction is given in Theorem 4.42.

2. Construct a second scenario H_B together with a probabilistic model $p_B \in \mathcal{Q}_1(H_B)$. Proposition 4.41 implies that $p_B \in \mathscr{CE}^\infty(H_B)$.

3. Then, show that the probabilistic model $p_A \otimes p_B$ does not satisfy the Consistent Exclusivity principle by exhibiting a set of events $C := \{v_1, \cdots, v_k\}$ in $H_A \otimes H_B$ such that $v_i \perp v_j$ for all $i \neq j$ and $\sum_{v_i \in C} p_A \otimes p_B(v_i) > 1$. The latter implies that $p_A \otimes p_B \notin \mathscr{CE}^1(H_A \otimes H_B)$.

4. Finally, by means of Theorem 4.45, the existence of $p_A \in \mathscr{CE}^\infty(H_A)$ and $p_B \in \mathscr{CE}^\infty(H_B)$ with $p_A \otimes p_B \notin \mathscr{CE}^1(H_A \otimes H_B)$ imply that $\mathscr{CE}^\infty(H_A \otimes H_B)$ is not convex.

The formal proof is presented below.

Theorem 4.46. *There exists a contextuality scenario H for which $\mathscr{CE}^\infty(H)$ is not convex.*

Proof. Consider first a contextuality scenario $H_A = (V_A, E_A)$ which satisfies $\mathcal{Q}_1(H_A) \subsetneq \mathscr{CE}^\infty(H_A)$. The existence of such a scenario is guaranteed by Theorem 4.42. For this scenario, we choose a probabilistic model p_A contained in $\mathscr{CE}^\infty(H_A)$ but not in $\mathcal{Q}_1(H_A)$. Lemmas (4.39) and (4.33) state respectively that

$$\Theta(\mathrm{NO}(H_A), p_A) = 1 \quad \text{and} \quad \vartheta(\mathrm{NO}(H_A), p_A) > 1.$$

The characterization of the Lovász number ϑ of a graph given in def. A.16 states that there exist an orthonormal representation $|\phi_v\rangle$ of $\overline{\mathrm{NO}}(H_A)$ and a normalized state $|\Psi\rangle$ such that

$$\vartheta(\mathrm{NO}(H_A), p_A) = \sum_{v \in V_A} p_A(v) |\langle \Psi | \phi_v \rangle|^2.$$

Hence, for this choice of orthonormal representation and state $|\Psi\rangle$, the following holds:

$$\sum_{v \in V_A} p_A(v) |\langle \Psi | \phi_v \rangle|^2 > 1. \tag{4.33}$$

We now wish to interpret the numbers $|\langle \Psi | \phi_v \rangle|^2$ as the probabilities of events in a contextuality scenario H_B. Consider the graph $G_B := \overline{\mathrm{NO}}(H_A)$. We now follow the construction in Prop. 4.36, and define a contextuality scenario $H_B = (V_B, E_B)$ from the graph G_B i the following way:

- $V(H_B) := V_A \cup \{ w_e \, : \, e \in E(\overline{G_B}) \}$,
- $E(H_B) := \{ e \cup \{w_e\} \, : \, e \in E(\overline{G_B}) \}$.

Here, the event w_e can be interpreted as the no-detection event associated to measurement e.

Let us finally introduce the following probabilistic model p_B on H_B:

$$p_B(v) := \begin{cases} |\langle \Psi | \phi_v \rangle|^2 & \text{if} \quad v \in V_A, \\ 1 - \sum_{v \in e} |\langle \Psi | \phi_v \rangle|^2 & \text{if} \quad v = w_e. \end{cases} \tag{4.34}$$

As shown in Prop. 4.36, $p_B \in \mathcal{Q}(H_B)$.

Consider now the probability model $p_A \otimes p_B$ on $H_A \otimes H_B$ defined as:

$$p_A \otimes p_B(v_A, v_B) := p_A(v_A) p_B(v_B). \tag{4.35}$$

By construction, it holds that $p_A \otimes p_B \in \mathscr{CE}^{\infty}(H_A) \otimes \mathscr{CE}^{\infty}(H_B)$. We will show, following Yan's argument (Yan13), that $p_A \otimes p_B \notin \mathscr{CE}^{1}(H_A \otimes H_B)$. Consider any couple of vertices $u \neq v$ in V_A, and the associated events $(u,u), (v,v) \in V(H_A \otimes H_B)$. These events are necessary orthogonal, meaning that there exists a measurement $e \in E(H_A \otimes H_B)$ containing both (u,u) and (v,v). In particular, the set $C := \{(v,v) \, : \, v \in V_A\}$ forms an independent set in $\mathrm{NO}(H_A \otimes H_B)$ [2]. By Lemma 4.39, a necessary condition for $p_A \otimes p_B$ to belong to $\mathscr{CE}^{1}(H_A \otimes H_B)$ is

$$\sum_{v \in C} p_A \otimes p_B(v) \leq 1.$$

However, it is clear that this sum can be rewritten

$$\sum_{v \in V_A} p_A(v) p_B(v) = \sum_{v \in V_A} p_A(v) |\langle \Psi | \phi_v \rangle|^2 > 1,$$

[2] We note that Yan's idea of looking at the *diagonal* set of vertices is not new in the context of the study of the Lovász number. Indeed, it was already present in Lovász's original paper on the subject (Lov79).

thereby proving that $p_A \otimes p_B \notin \mathscr{CE}^1(H_A \otimes H_B)$. From Theorem 4.45 follows that $\mathscr{CE}^\infty(H_A \otimes H_B)$ is not convex.

□

Failure of convexity leads to a natural way to strengthen the CE principle: the collection of physically realizable probabilistic models should be both convex and closed under $\otimes$. Therefore, if some physically realistic $q \in \mathscr{CE}^\infty(H)$ can be combined with some $p \in \mathscr{CE}^\infty(H)$ by using convex combinations and $\otimes$-products such that the combination is not in $\mathscr{CE}^\infty$, then p itself should be considered to violate the CE principle in a certain extended form. I elaborate on this ideas in the next section.

4.10. Extended Consistent Exclusivity Principle

In the previous section I showed that the set of probabilistic models $\mathscr{CE}^\infty$ is neither convex nor closed under $\otimes$. However, a natural assumption is that the collection of physically realizable probabilistic models should satisfy both properties, since the set $\mathcal{Q}$ of quantum models indeed does. This motivates the following definition:

Definition 4.47. *A probabilistic model p on a contextuality scenario H satisfies the Extended Consistent Exclusivity principle if for all contextuality scenarios H' and $q \in \mathcal{Q}(H')$,*

$$p \otimes q \in \mathscr{CE}^\infty(H \otimes H').$$

We write $\widetilde{CE}^\infty(H)$ for the set of probabilistic models satisfying the extended consistent exclusivity principle.

One way to interpret this definition is the following: given a set that is convex and closed under $\otimes$ and which satisfies CE, we want to see which are the probabilistic models that do not belong to it, but when combined to any element of the set they jointly satisfy CE. However, $\mathcal{Q}$ is not the only set that is convex and closed under $\otimes$ and satisfies CE. For instance, $\mathcal{Q}_1$ also satisfies those properties. An interesting problem is to see how the set $\widetilde{CE}^\infty(H)$ changes when considering a set other than $\mathcal{Q}$ for its definition. The proposed definition has the following nice implication:

Corollary 4.48. $\widetilde{CE}^\infty(H) = Q_1(H)$.

Proof. The construction presented in the proof of non-convexity of $\mathscr{CE}^\infty$ in section 4.9 shows that $\widetilde{CE}^\infty(H) \subseteq Q_1(H)$. The other inclusion is a consequence of proposition 4.34: $\mathcal{Q}_1(H_A) \otimes \mathcal{Q}_1(H_B) \subseteq \mathcal{Q}_1(H_A \otimes H_B)$. □

This means that there are no probabilistic models outside $\mathcal{Q}_1$ which, after combined with one in $\mathcal{Q}$, jointly satisfy CE. Moreover, if we had used the set $\mathcal{Q}_1$ for the definition of ECE, we would have arrived to the same conclusion, i.e. there are no probabilistic models outside $\mathcal{Q}_1$ which, after combined with one in $\mathcal{Q}_1$, jointly satisfy CE. This result was somehow implicit in the one by Yan (Yan13). There, Yan shows that the maximum violation of a noncontextuality inequality given by models that satisfy ECE is the same as the maximum "quantum" violation in the CSW formalism. Here we see that a model satisfies ECE if and only if it is a normalized CSW-quantum assignment (see section 4.7).

4.11. Examples

In the previous sections, I developed the abstract theory of contextuality scenarios in some detail, and have also exemplified some of the concepts and results for the case of Bell scenarios. In particular, this illustrates how our formalism makes precise the intuition that nonlocality is a special case of contextuality. Also, I related our approach to the graph-theoretical one of Cabello, Severini and Winter (CSW10) and the observable-based one of Abramsky and Brandenburger (AB11). Now I move on to considering other more concrete cases. The examples that have already been considered in the quantum foundations literature are too numerous to list, so in this section I focus on a few particularly appealing classes. First, I will discuss *Hypergraphs with no-detection events*: here, the whole hierarchy of $\mathcal{Q}_n$ sets collapses into the first level, hence the quantum set is characterized by the Lovász number. Then, I'll move on to *n-circular hypergraphs* and *antiprism scenarios*, where we find both an infinite family of hypergraphs for which all the sets of probabilistic models defined in this chapter collapse into the classical set, and an infinite family of scenarios for which there exist contextual models.

Hypergraphs with no-detection events

Hypergraphs equipped with no-detection events have been useful throughout this chapter, when relating our approach to CSW's or when proving nonconvexity of $\mathscr{CE}^{\infty}$. In what follows, I will present another property of this family of hypergraphs, namely that $\mathcal{Q}_1(H_{nd}) = \mathcal{Q}(H_{nd})$.

Let us begin by recalling the definition of a Hypergraph with no-detection events.

Definition 4.49. *Let H be a contextuality scenario. We say that H is equipped with no-detection events, and denote it by H_{nd}, when for every hyperedge $e \in$*

$E(H_{nd})$ there exists a vertex $w_e \in V(H_{nd})$ which is not contained in any other hyperedge, i.e. $w_e \in e'$ if and only if $e = e'$. These vertices w_e are called no-detection events.

Note that a hyperedge in H_{nd} may contain more than one no-detection event. However, without loss of generality, we will consider in the following that each hyperedge on H_{nd} contains exactly one no-detection event.

Proposition 4.50. $\mathcal{Q}(H_{nd}) = \mathcal{Q}_1(H_{nd})$.

Proof. The inclusion $\mathcal{Q}(H_{nd}) \subseteq \mathcal{Q}_1(H_{nd})$ follows by definition, so we need to prove that $\mathcal{Q}(H_{nd}) \supseteq \mathcal{Q}_1(H_{nd})$.

Consider a probabilistic model $p \in \mathcal{Q}_1(H_{nd})$. The characterization 4.32-4 of $\mathcal{Q}_1(H_{nd})$ implies that there exists a state $|\Psi\rangle$ and projections P_v for all $v \in V(H_{nd})$ such that $u \perp v \implies P_u \perp P_v$ and $p(v) = \langle\Psi|P_v|\Psi\rangle$.

Define the sets $V_{nd} = \{w_e \mid e \in E(H_{nd})\}$, i.e. the set of no-detection events, and $V' = V \setminus V_{nd}$. We now define

$$P'_{w_e} := \mathbb{1}_{\mathcal{H}} - \sum_{v\in e,\, v\neq w_e} P_v$$

and claim that these, together with $\{P_v \mid v \in V'\}$ and the state $|\Psi\rangle$, form a quantum model for p. First, due to $\sum_{v\in e} P_v \leq \mathbb{1}_{\mathcal{H}}$, the operator P'_{w_e} is also a projection. Second, the completeness relation for hyperedges in $E(H_{nd})$ then holds by definition. Third,

$$\langle\Psi|P'_{w_e}|\Psi\rangle = \langle\Psi|\Psi\rangle - \sum_{v\in e}\langle\Psi|P_v|\Psi\rangle = 1 - \sum_{v\in e,\, v\neq w_e} p(v) = p(w_e).$$

as claimed. Hence, $p \in \mathcal{Q}(H_{nd})$. □

In words, the fact that every hyperedge has a no-detection event, allows us to define new projectors for these events such that the associated probabilities remain unchanged but the model becomes normalized at the level of projectors as well.

The description of the sets $\mathcal{Q}_n$ for this family of hypergraphs is then particularly simple: the whole semidefinite hierarchy collapses to the first level. The advantage of this is that proposition 4.33 on the relation between $\mathcal{Q}_1$ and the Lovász number now also applies to quantum models. Hence, this family of hypergraphs forms a very special and well-behaved subclass of all contextuality scenarios. The n-circular hypergraphs that we consider next arise in this way. However, many of the more interesting contextuality scenarios –like Bell scenarios– are not of this form.

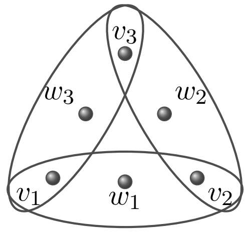


Figure 4.9.: The 3-circular hypergraph Δ_3

.

n-circular hypergraphs

The n-circular hypergraphs generalize the "pentagon" idea of Klyachko-Can-Binicioğlu-Shumovsky (KCB08), and are defined as follows.

Definition 4.51. *For $n \geq 3$, the n-**circular hypergraph** Δ_n is given by*

$$V(\Delta_n) := \{v_1, \ldots, v_n, w_1, \ldots, w_n\},$$
$$E(\Delta_n) := \{\{v_1, w_1, v_2\}, \ldots, \{v_n, w_n, v_1\}\}.$$

In words, Δ_n has $2n$ vertices and n edges such that, if all vertices are evenly distributed on a circle in the order $v_1, w_1, \ldots, v_n, w_n, v_1$, then every second triple of adjacent vertices, namely those of the form $\{v_j, w_j, v_{j+1}\}$, is an edge. We write $v_{n+1} = v_1$. Figure 4.9 displays Δ_3, and Δ_5 is the "pentagon" scenario on which the KCBS inequality (KCB08) is defined.

We now extend some of these results to arbitrary n.

Proposition 4.52. *Let $n \geq 3$.*

1. $\dim(\mathcal{C}(\Delta_n)) = \dim(\mathcal{G}(\Delta_n)) = n$.
2. *If n is even, then $\mathcal{C}(\Delta_n) = \mathcal{G}(\Delta_n)$.*
3. *If n is odd, then $\mathcal{C}(\Delta_n) \subsetneq \mathcal{G}(\Delta_n)$ is determined by the inequality*

$$\sum_i p(v_i) \leq \frac{n-1}{2}. \tag{4.36}$$

There is one extreme point of $\mathcal{G}(\Delta_n)$ which violates this inequality. It is the probabilistic model $p_x \in \mathcal{G}(\Delta_n)$ with

$$p_x(v_i) = \tfrac{1}{2}\,\forall i, \qquad p_x(w_i) = 0\,\forall i. \tag{4.37}$$

In particular, $\mathcal{G}(\Delta_n)$ has one vertex more than $\mathcal{C}(\Delta_n)$.

Proof. We consider all vertex indices modulo n, so that $v_{n+1} = v_1$, etc.

1 The equations imposed on the probabilities $p(v_i)$ and $p(w_i)$ by the normalization constraints are just

$$p(w_i) = 1 - p(v_i) - p(v_{i+1}), \tag{4.38}$$

which implies $\dim(\mathcal{G}(\Delta_n)) \leq n$. The conclusion follows if we can produce $n+1$ linearly independent deterministic models. This is simple: the set of models

$$p_j(v_i) := \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases}$$

where the $p_j(w_i)$ are uniquely determined thanks to (4.38) and with $j \in \{1, \ldots, n\}$, is linearly independent. Furthermore, adding to this set the model p_0 with $p_0(v_i) = 0$ for all i preserves linear independence. This is the desired collection of $n+1$ linearly independent deterministic models.

2 $\mathcal{C}(\Delta_n) = \mathscr{CE}^1(\Delta_n)$ follows from Prop. 4.43, and $\mathscr{CE}^1(\Delta_n) = \mathcal{G}(\Delta_n)$ because the maximal independent sets of $\mathrm{NO}(\Delta_n)$ are precisely the edges on Δ_n. In particular, while (4.37) is also a probabilistic model for even n, in this case it has to be a convex combination of deterministic models.

Note that this argument has not used 1.

3 We apply Theorem 4.8 in combination with Prop. 4.43. Any induced subscenario H_W with $\mathcal{C}(H_W) \neq \mathcal{G}(H_W)$ needs to contain an induced (anti-)cycle in $\mathrm{NO}(H_W)$. This is possible only if W contains all v_i. If W also contains one or more of the w_i's, then H_W does not have a unique probabilistic model. Therefore, there can be at most one nonclassical extreme point of $\mathcal{G}(H)$, namely the one associated to the induced subscenario on $W = \{v_1, \ldots, v_n\}$. Now this H_W does indeed have a unique probabilistic model given by $p_x(v_i) = \frac{1}{2}$, which yields (4.37) upon extension to Δ_n. This proves that $\mathcal{G}(\Delta_n)$ has p_x as its sole nonclassical extreme point without ever using any inequalities.

We now give an independent proof showing that (4.36) defines $\mathcal{C}(\Delta_n)$. Thanks to (4.38), it is enough to consider the values $p(v_i)$ only. Now the deterministic models correspond to the independent sets in the cycle graph C_n; upon identifying each vertex with the edge adjacent on its left, an independent set in C_n gets identified with a set of edges in C_n no two of which are adjacent at the same vertex, i.e. with a **matching** on C_n. Now it is known (Sch03) that the polytope of all matchings on C_n corresponds to

$$p(v_i) \geq 0, \qquad p(v_i) + p(v_{i+1}) \leq 1, \qquad \sum_{i=1}^{n} p(v_i) \leq \frac{n-1}{2}.$$

This is precisely the description of $\mathcal{C}(\Delta_n)$ that was to be proven.

□

For $n = 5$, the set of classical models is bounded by the inequality $\sum_{i=1}^{5} p(v_i) \leq 2$, which is precisely the inequality which has been studied in (KCB08).

Proposition 4.53. *$\mathcal{C}(\Delta_3) = \mathscr{CE}^1(\Delta_3) \subsetneq \mathcal{G}(\Delta_3)$. For all other n, $\mathscr{CE}^1(\Delta_n) = \mathcal{G}(\Delta_n)$.*

Proof. Since $\{v_1, v_2, v_3\}$ is the only independent set in $\mathrm{NO}(\Delta_3)$ which is not an edge of Δ_3, we find that $\mathscr{CE}^1(\Delta_3)$ as a subset of $\mathcal{G}(\Delta_3)$ is given by imposing the inequality $p(v_1) + p(v_2) + p(v_3) \leq 1$. This is precisely the inequality that determines $\mathcal{C}(\Delta_3)$ in 4.52-3. For $n \geq 4$, however, every independent set in $\mathrm{NO}(\Delta_n)$ is of the form $\{v_i, w_i, v_{i+1}\}$, i.e. is itself an edge. □

Antiprism scenarios

The antiprism scenarios are a variant of the circular hypergraph scenarios with some additional edges thrown in such that there is a symmetry exchanging the v_i with the w_i. Again, we consider all vertex indices modulo n. The antiprism scenarios are supposed to illustrate that an interesting looking hypergraph is not necessarily an interesting contextuality scenario.

Definition 4.54. *Let $n \geq 3$. The n-**antiprism scenario** AP_n is*

$$\begin{aligned} V(AP_n) &:= \{v_1, \ldots, v_n, w_1, \ldots, w_n\}, \\ E(AP_n) &:= \{\{v_1, w_1, v_2\}, \ldots, \{v_n, w_n, v_1\}\} \\ &\quad \cup \{\{w_1, v_2, w_2\}, \ldots, \{w_n, v_1, w_1\}\}. \end{aligned}$$

Proposition 4.55. *If n is divisible by 3, then $\mathcal{C}(AP_n) = \mathcal{G}(AP_n)$ is a 2-dimensional triangle. Otherwise, AP_n has a unique probabilistic model which is not classical.*

Proof. We show that $p(v_1)$ and $p(v_2)$ determine all other probabilities $p(v_i)$ and $p(w_i)$ by induction on i:

$$p(v_{i+1}) = 1 - p(v_i) - p(w_i), \qquad p(w_{i+1}) = 1 - p(w_i) - p(v_{i+1}).$$

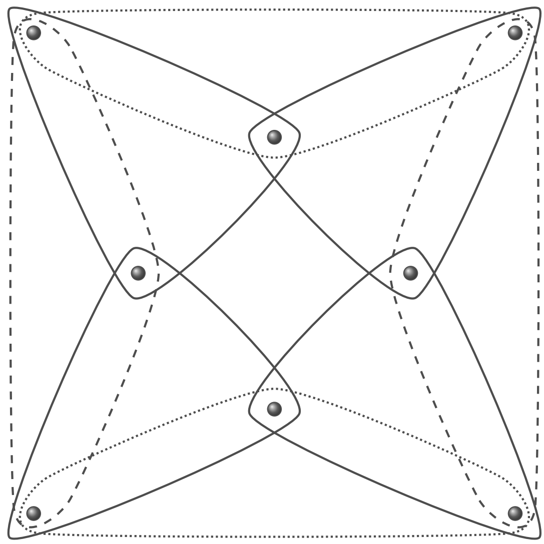


Figure 4.10.: The contextuality scenario AP_4.

In fact, this shows that

$$p(v_{3j+1}) = p(w_{3j+2}) = p(v_1), \quad p(v_{3j+2}) = p(w_{3j}) = p(v_2)$$

$$p(v_{3j}) = p(w_{3j+1}) = 1 - p(v_1) - p(v_2).$$

Now if n is divisible by 3, then this is consistent upon "going around the cycle", so that $\mathcal{G}(AP_n)$ can be identified with the triangle

$$p(v_1) \geq 0, \qquad p(v_2) \geq 0, \qquad p(v_1) + p(v_2) \leq 1.$$

Clearly, the extreme points of this triangle are deterministic.

If n is not divisible by 3, then the above recurrence relations imply that $p(v_1) = p(v_2) = \frac{1}{3}$, so that $\mathcal{G}(AP_n)$ degenerates to a single point. $\mathcal{C}(AP_n) = \emptyset$ since there is no deterministic model. □

We now give another example application of our methods, for the hypergraph AP_4. Note that for this scenario, its non-orthogonality graph $\mathrm{NO}(AP_4)$ coincides with the orthogonality graph of possible events for a PR box (Fig. 3.6(a)) studied in section 3.5.3, which is itself the complement of the 4-antiprism graph $\bowtie_4$ (Fig. 3.6(b)).

Proposition 4.56. *$\mathcal{Q}_1(AP_4) = \emptyset$, although $\mathscr{CE}^1(AP_4) = \mathcal{G}(AP_4)$.*

Proof. Direct inspection shows that every maximal independent set in $\mathrm{NO}(AP_4)$ is an edge, so that the unique probabilistic model given by $p(v_i) = p(w_i) = \frac{1}{3}$ is in $\mathscr{C}\mathscr{E}^1(AP_4)$.

It remains to show that the unique probabilistic model is not in $\mathcal{Q}_1(AP_4)$. By Proposition 4.33, this boils down to showing that $\frac{1}{3}\vartheta(\mathrm{NO}(AP_4)) > 1$. Now $\mathrm{NO}(AP_n)$ is the complement of the 4-antiprism graph $\bowtie_4$. Since $\bowtie_4$ is vertex-symmetric, we deduce (Knu94, Thm. 25) that $\vartheta(\bowtie_4)\vartheta(\mathrm{NO}(AP_4)) = 8$. Now $\vartheta(\bowtie_4)$ is known (BPT11) to equal $8 - 4\sqrt{2}$, so that

$$\vartheta(\mathrm{NO}(AP_4)) = \frac{8}{8 - 4\sqrt{2}} = \frac{2}{2 - \sqrt{2}} = 2 + \sqrt{2} > 3,$$

as was to be shown. □

5. Bell inequalities from two-body correlation functions

Detecting the nonlocal character of correlations observed in an experiment is an interesting problem. In principle, one needs to consider the local polytope of the corresponding Bell scenario and check whether the conditional probability distribution lies inside or outside of it. However, from a practical point of view this approach is inconvenient for large scenarios, since the dimensionality of the polytope increases exponentially with the number of parties, which makes the problem computationally intractable.

In this chapter, I aim at simplifying the problem by focusing the study on Bell inequalities that contain only one and two-body correlators. In principle one could argue the relevance of such inequalities, since in general the correlators that involve a large number of parties are those which carry detailed information about the correlations. Contrary to this intuition, we found that one and two-body correlators are already useful for detecting nonlocality in physically relevant systems. For this, we have further restricted the two-body correlators Bell inequalities to those that satisfy certain symmetries regarding the labelling of the parties: on the one hand permutational invariance, and on the other translational invariance.

5.1. Bell inequalities from two-body correlators

In the previous chapters, the notion of classical and quantum models was studied in terms of (conditional) probability distributions. However, as mentioned in section 2.1.5, Bell inequalities can also be written in terms of correlators, an equivalent representation in the case of dichotomic measurements that proves useful in this chapter. Within this representation, the objects that correspond to the probabilistic models on scenario $(n, m, 2)$ are now vectors $\mathbf{E}$, whose components are presented in section 2.1.5. In this chapter, we further interpret each of the components of these correlators $\mathbf{E}$ as *expectation values of physical observables*, like mentioned in section 2.1.5. Indeed, given a set of dichotomic observables $\{\mathcal{M}_k^{(i)}\}_k$ for each party i, each component of the vector $\mathbf{E}$ can be written as $E_{x_1 \dots x_k} = \langle \mathcal{M}_{x_1}^{(1)} \cdots \mathcal{M}_{x_k}^{(k)} \rangle$. In this chapter, we will use the notation $\mathbf{M}$ over $\mathbf{E}$, since it stresses the actual observables under consideration.

The components of the vectors $\mathbf{M} \in \mathbb{R}^{(m+1)^n - 1}$ are expressed (see section 2.1.5) as follows:

- First, $n * m$ components which correspond to the $n * m$ single-party correlators $\langle \mathcal{M}_k^{(i)} \rangle = P_i(0|k) - P_i(1|k)$. Here, $P_i(a|k)$ denotes the probability that party i obtains outcome a when measuring k. $\mathcal{M}_k^{(i)}$ is usually referred to as the "observable" k measured by party i.
- Second, $\binom{n}{2} m^2$ components which correspond to the two-party correlators $\langle \mathcal{M}_{k_1}^{(i_1)} \mathcal{M}_{k_2}^{(i_2)} \rangle = \sum_{a_1, a_2} (-1)^{a_1 \oplus a_2} P_{i_1 i_2}(a_1 a_2 | k_1 k_2)$.
- Continue, for each $j = 3 \dots n$, with $\binom{n}{j} m^j$ components which correspond to the j-party correlators

$$\langle \mathcal{M}_{k_1}^{(i_1)} \dots \mathcal{M}_{k_j}^{(i_j)} \rangle = \sum_{a_1 \dots a_j} (-1)^{a_1 \oplus \dots \oplus a_j} P_{i_1 \dots i_j}(a_1 \dots a_j | k_1 \dots k_j). \quad (5.1)$$

It is straightforward to check that $\sum_{j=1}^{n} \binom{n}{j} m^j = (m+1)^n - 1$, as the dimension of the vector $\mathbf{M}$. Similar to the case of the $\mathbf{P}$ representation of correlations (see section 2.1.5), for classical models the correlators $\langle \mathcal{M}_{k_1}^{(i_1)} \dots \mathcal{M}_{k_j}^{(i_j)} \rangle$ take a product form $\langle \mathcal{M}_{k_1}^{(i_1)} \rangle \cdot \ldots \cdot \langle \mathcal{M}_{k_j}^{(i_j)} \rangle$. Hence, the set of classical correlations is characterized by the convex hull of the deterministic correlators $\mathbf{M}_D$, defined as those $\langle \mathcal{M}_{k_1}^{(i_1)} \rangle \cdot \ldots \cdot \langle \mathcal{M}_{k_j}^{(i_j)} \rangle$ with local mean values being $\langle \mathcal{M}_{k_l}^{(i_l)} \rangle = \pm 1$. The set of classical correlations is again a polytope that we denote by $\mathbb{P}$. As mentioned in section 2.1.5, the facets of this polytope correspond to the *tight* Bell inequalities of the scenario $(n, m, 2)$.

Most of the known constructions of multipartite Bell inequalities contain highest-order correlators, i.e., those with $j = n$ in eq. (5.1). However, throughout this chapter we will see how to design Bell inequalities that witness nonlocality only from one and two body[1] expectation values. In addition, we will focus on the case of two measurements per party. The general form of such a Bell inequality is

$$\sum_{i=1}^{n} (\alpha_i \langle \mathcal{M}_0^{(i)} \rangle + \beta_i \langle \mathcal{M}_1^{(i)} \rangle) + \sum_{i<j}^{n} \gamma_{ij} \langle \mathcal{M}_0^{(i)} \mathcal{M}_0^{(j)} \rangle +$$
$$+ \sum_{i \neq j}^{n} \delta_{ij} \langle \mathcal{M}_0^{(i)} \mathcal{M}_1^{(j)} \rangle + \sum_{i<j}^{n} \varepsilon_{ij} \langle \mathcal{M}_1^{(i)} \mathcal{M}_1^{(j)} \rangle + \beta_C \geq 0, \quad (5.2)$$

[1]In this chapter, I will use the words n-party and n-body interchangeably.

where $\alpha_i, \beta_j, \gamma_{ij}, \delta_{ij}$, and ε_{ij} are some real parameters, while β_C is the so-called classical bound. The corresponding polytope $\mathbb{P}_2$ of classical correlations is then constructed from the elements of $\mathbb{P}$ by neglecting correlators of order higher than two. Indeed, we take all elements $\mathbf{M}$ of $\mathbb{P}$ and simply remove the components which correspond to j-party correlators with $j \geq 3$. Similarly, the vertices of $\mathbb{P}_2$ are those collections of correlators $\mathbf{M}_2$ for which $\langle \mathcal{M}_k^{(i)} \mathcal{M}_l^{(j)} \rangle = \langle \mathcal{M}_k^{(i)} \rangle \cdot \langle \mathcal{M}_l^{(j)} \rangle$, where local mean values are ± 1.

Although $\dim \mathbb{P}_2 = 2n^2$ is much smaller than $3^n - 1$ (the dimension of $\mathbb{P}$), it still grows with the number of parties, which difficults the task of determining the facets of $\mathbb{P}_2$. One way to overcome this problem is to restrict the study to Bell inequalities that obey some symmetries. For instance, one could consider translationally invariant Bell inequalities or, in the spirit of Ref. (BGP10), those that are invariant under any permutation of the parties. In section 5.4 I will comment on the former, while the latter is presented in sections 5.2 and 5.3.

5.2. Fully-symmetric two-body Bell Inequalities

In this section I study a specific type of Bell inequalities, namely those which are symmetric under permutation of the parties and further contain only one and two-body correlators.

Given a Bell inequality, imposing permutational symmetry means that when we exchange the label (order) of any party the equation remains the same. Mathematically, for a two-body correlators Bell inequality in the form (5.2) this implies that the expectation values $\langle \mathcal{M}_k^{(i)} \rangle$ and $\langle \mathcal{M}_k^{(i)} \mathcal{M}_l^{(j)} \rangle$, with fixed k, l and different i, j, appear in the Bell inequality (5.2) with the same "weights", i.e. $\alpha_i = \alpha$, $\beta_i = \beta$, and so on. Hence, the general form of a symmetric Bell inequality with one- and two-body correlators is

$$I := \alpha \mathcal{S}_0 + \beta \mathcal{S}_1 + \frac{\gamma}{2} \mathcal{S}_{00} + \delta \mathcal{S}_{01} + \frac{\varepsilon}{2} \mathcal{S}_{11} \geq -\beta_C, \tag{5.3}$$

where $\alpha, \beta, \gamma, \delta, \varepsilon$ are real parameters, and $\mathcal{S}_k$ and $\mathcal{S}_{kl}$ (with $k, l = 0, 1$) denote the one- and two-body correlators symmetrized over all observers, i.e.,

$$\mathcal{S}_k = \sum_{i=1}^{n} \langle \mathcal{M}_k^{(i)} \rangle, \qquad \mathcal{S}_{kl} = \sum_{i \neq j = 1}^{n} \langle \mathcal{M}_k^{(i)} \mathcal{M}_l^{(j)} \rangle. \tag{5.4}$$

Geometrically, the polytope $\mathbb{P}_2$ is mapped under permutational symmetry onto a simpler one $\mathbb{P}_2^S$, which independently of the number of parties, is always five-dimensional and it elements are the vectors $(\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_{00}, \mathcal{S}_{01}, \mathcal{S}_{11})$. Note that the number of vertices is significantly reduced, from 2^{2n} for $\mathbb{P}_2$ to $2(n^2+1)$ for $\mathbb{P}_2^S$.

The idea now is to find all the tight Bell inequalities of the form (5.3), i.e. the facets of $\mathbb{P}_2^S$. In order to do so, we begin by finding its extremal points. First, we present a parametrization of the vertices of $\mathbb{P}_2^S$ in terms of three natural numbers, and then find a necessary and sufficient condition for those parameters to define such an extremal point.

Characterization of the extreme points of $\mathbb{P}_2^S$.

Denote by V the set of vertices of $\mathbb{P}_2$, and by V_S that of $\mathbb{P}_2^S$. For every element of V we denote by

$$x_i = \langle \mathcal{M}_0^{(i)} \rangle, \qquad y_i = \langle \mathcal{M}_1^{(i)} \rangle, \tag{5.5}$$

the pair of local deterministic expectation values for party i (which by definition have value ± 1) and by $\{x_i, y_i\}$ the corresponding local strategy. Note that permutational symmetry guarantees that the values of $\mathcal{S}_0$ and $\mathcal{S}_1$ do not depend on the particular local strategies applied by each party, but rather on the number of parties which have applied each strategy. Hence, we introduce the following parametrization:

$$\begin{aligned} a &= \#\{i \in \{1,\ldots,n\} \,|\, x_i = 1, y_i = 1\} \\ b &= \#\{i \in \{1,\ldots,n\} \,|\, x_i = 1, y_i = -1\}, \\ c &= \#\{i \in \{1,\ldots,n\} \,|\, x_i = -1, y_i = 1\}, \\ d &= \#\{i \in \{1,\ldots,n\} \,|\, x_i = -1, y_i = -1\}. \end{aligned} \tag{5.6}$$

In words, given a vertex in $\mathbb{P}_2$, the parameters a, b, c, and d represent the number of parties who apply the local strategy $\{1,1\}$, $\{1,-1\}$, $\{-1,1\}$, and $\{-1,-1\}$, respectively. By definition, $a+b+c+d=n$.

Following these parameters, the symmetrized local expectation values $\mathcal{S}_k$ ($k = 0,1$) may be expressed as

$$\mathcal{S}_0 = a+b-c-d, \qquad \mathcal{S}_1 = a-b+c-d. \tag{5.7}$$

Moreover, since for every element of V

$$\mathcal{S}_{xy} = \mathcal{S}_x\mathcal{S}_y - \sum_{i=1}^{n} \langle \mathcal{M}_x^{(i)} \rangle \langle \mathcal{M}_y^{(i)} \rangle \qquad (x,y=0,1), \tag{5.8}$$

the two-body symmetrized expectation values may be expressed as

$$\begin{aligned}
\mathcal{S}_{00} &= \mathcal{S}_0^2 - n = (a+b-c-d)^2 - n, \\
\mathcal{S}_{11} &= \mathcal{S}_1^2 - n = (a-b+c-d)^2 - n, \\
\mathcal{S}_{01} &= \mathcal{S}_0\mathcal{S}_1 - \sum_{i=1}^{n} \langle \mathcal{M}_x^{(i)} \rangle \langle \mathcal{M}_y^{(i)} \rangle \\
&= (a+b-c-d)(a-b+c-d) - (a-b-c+d).
\end{aligned} \tag{5.9}$$

Hence, all vertices of $\mathbb{P}_2$ are mapped under symmetrization onto elements of $\mathbb{P}_2^S$ parametrized by the previously defined $\{a,b,c,d\}$. Geometrically, these parameters belong to the set $\mathbb{T}_n = \{(a,b,c,d) \in \mathbb{N}^4 \,|\, a+b+c+d = n\}$, which is isomorphic to a tetrahedron in $\mathbb{N}^3$

$$\mathbb{T}_n = \{(a,b,c) \in \mathbb{N}^3 \,|\, a+b+c \leq n\}. \tag{5.10}$$

Even though every element in $\mathbb{T}_n$ is an extreme point in $\mathbb{P}_2$, not every vertex of $\mathbb{P}_2$ is an extreme point of $\mathbb{P}_2^S$. In what follows, we show that vertices of $\mathbb{P}_2^S$ are uniquely represented by all those 4-tuples from $\mathbb{T}_n$ that belong to its boundary $\partial\mathbb{T}_n$, i.e. those for which the condition $abcd = 0$ is satisfied.

Theorem 5.1. *Let $\varphi : \mathbb{T}_n \mapsto \mathbb{P}_2^S$ be the previously defined parametrization*

$$\varphi((a,b,c,d)) = (\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_{00}, \mathcal{S}_{01}, \mathcal{S}_{11}). \tag{5.11}$$

Then $\varphi(p)$ is a vertex of $\mathbb{P}_2^S$ iff $p \in \partial\mathbb{T}_n$.

Proof. We start from the "only if" part. Assume on the contrary that $p = (a,b,c,d)$ belongs inside $\mathbb{T}_n$, i.e. that all its components are larger than zero $(a,b,c,d \geq 1)$. Consider a vector $v = (1,-1,-1,1) \notin \mathbb{T}_n$ and notice that the values of $\mathcal{S}_k$ and $\mathcal{S}_{kk}$ with $k = 0,1$ are constant along the line $p + \lambda v$ for any $\lambda \in \mathbb{R}$, while $\mathcal{S}_{01}(p+\lambda v) = \mathcal{S}_{01}(p) - 4\lambda$. Hence, for any $\alpha, \beta > 0$,

$$\begin{aligned}
&\alpha\varphi(p+\beta v) + \beta\varphi(p-\alpha v) \\
&\quad = \alpha(\mathcal{S}_0(p), \mathcal{S}_1(p), \mathcal{S}_{00}(p), \mathcal{S}_{01}(p) - 4\beta, \mathcal{S}_{11}(p)) \\
&\quad\quad + \beta(\mathcal{S}_0(p), \mathcal{S}_1(p), \mathcal{S}_{00}(p), \mathcal{S}_{01}(p) + 4\alpha, \mathcal{S}_{11}(p)) \\
&\quad = (\alpha+\beta)(\mathcal{S}_0(p), \mathcal{S}_1(p), \mathcal{S}_{00}(p), \mathcal{S}_{01}(p), \mathcal{S}_{11}(p)) \\
&\quad = (\alpha+\beta)\varphi(p),
\end{aligned} \tag{5.12}$$

which allows us to express $\varphi(p)$ as

$$\varphi(p) = \frac{\alpha}{\alpha+\beta}\varphi(p+\beta v) + \frac{\beta}{\alpha+\beta}\varphi(p-\alpha v). \tag{5.13}$$

Now choose $\alpha = \min\{a, d\}$ and $\beta = \min\{b, c\}$. Hence, both $p + \beta v$ and $p - \alpha v$ belong to the boundary of $\mathbb{T}_n$, which implies that $\varphi(p) \in \mathbb{P}_2^S$ may be written as a convex combination of two other elements $\varphi(p + \min\{a, d\}v)$ and $\varphi(p - \min\{b, c\}v)$ of $\mathbb{P}_2^S$. It follows that $\varphi(p)$ is not extremal.

In order to prove the "if" part, assume that $p \in \partial\mathbb{T}_n$ and that the vector $\varphi(p) = (\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_{00}, \mathcal{S}_{01}, \mathcal{S}_{11})$ is not a vertex of $\mathbb{P}_2^S$. Then, $\varphi(p)$ can be decomposed into a convex combination of vertices of $\mathbb{P}_2^S$ represented by $p_i = (a_i, b_i, c_i, d_i) \in \mathbb{T}_n$, i.e.

$$\varphi(p) = \sum_{i=0}^{k} \lambda_i \varphi(p_i) \tag{5.14}$$

with $0 < \lambda_i < 1$ summing up to unity, and

$$\varphi(p_i) = (\mathcal{S}_0^{(i)}, \mathcal{S}_1^{(i)}, \mathcal{S}_{00}^{(i)}, \mathcal{S}_{01}^{(i)}, \mathcal{S}_{11}^{(i)}). \tag{5.15}$$

By combining eqs. (5.11) and (5.15), eq. (5.14) is equivalent to the following five equations:

$$\mathcal{S}_l = \sum_{i=0}^{k} \lambda_i \mathcal{S}_l^{(i)}, \qquad \mathcal{S}_{ll} = \sum_{i=0}^{k} \lambda_i \mathcal{S}_{ll}^{(i)} \tag{5.16}$$

for $l = 0, 1$, and

$$\mathcal{S}_{01} = \sum_{i=0}^{k} \lambda_i \mathcal{S}_{01}^{(i)}. \tag{5.17}$$

Since for all vertices of $\mathbb{P}_2$ holds that $\mathcal{S}_{ll}^{(i)} = [\mathcal{S}_l^{(i)}]^2 - n$ (see eqs. (5.9)), eqs. (5.16) imply that $\mathcal{S}_l^{(i)}$ must satisfy

$$\sum_i \lambda_i \left(\mathcal{S}_l^{(i)}\right)^2 = \left(\sum_i \lambda_i \mathcal{S}_l^{(i)}\right)^2 \qquad (l = 0, 1). \tag{5.18}$$

On the one hand, eq. (5.18) can be thought of as being a quadratic equation for a particular $\mathcal{S}_l^{(m)}$, i.e.

$$\begin{aligned} &\lambda_m(\lambda_m - 1)\left(\mathcal{S}_l^{(m)}\right)^2 + 2\lambda_m \mathcal{S}_l^{(m)} \sum_{i \neq m} \lambda_i \mathcal{S}_l^{(i)} \\ &\quad + \left(\sum_{i \neq m} \lambda_i \mathcal{S}_l^{(i)}\right)^2 - \sum_{i \neq m} \lambda_i \left(\mathcal{S}_l^{(i)}\right)^2 = 0 \end{aligned} \tag{5.19}$$

This equation has real solutions if and only if its discriminant is nonnegative, which holds iff

$$-4\lambda_0 \sum_{\substack{i<j \\ i,j\neq m}} \lambda_i\lambda_j \left(\mathcal{S}_l^{(i)} - \mathcal{S}_l^{(j)}\right)^2 \geq 0. \tag{5.20}$$

Since all λ's are positive, the above condition is fulfilled iff $\mathcal{S}_l^{(i)} = \mathcal{S}_l^{(j)}$ for all $i, j \neq m$ and $l = 0, 1$. Moreover, eq. (5.20) should be obeyed for any m, hence

$$\mathcal{S}_l^{(i)} = \mathcal{S}_l^{(j)} = \mathcal{S}_l \tag{5.21}$$

for any $i, j = 1, \ldots, k$ and $l = 0, 1$.

On the other hand, the assumption that $\varphi(p)$ is not a vertex of $\mathbb{P}_2^S$, i.e. that it can be decomposed as in (5.14), means that $\mathcal{S}_{01}^{(i)}$ cannot be all equal, since otherwise $\varphi(p_i)$ are all the same. Hence, from the expressions $\mathcal{S}_{01} = \mathcal{S}_0\mathcal{S}_1 - (a - b - c + d)$ and $\mathcal{S}_{01}^{(i)} = \mathcal{S}_0^{(i)}\mathcal{S}_1^{(i)} - (a_i - b_i - c_i + d_i)$, combined with eq. (5.21), follows

$$a - b - c + d = \sum_i \lambda_i(a_i - b_i - c_i + d_i). \tag{5.22}$$

If we further note that

$$a_i + b_i + c_i + d_i = n = a + b + c + d \tag{5.23}$$

should hold for any i, it follows that each a, b, c, and d is a convex combination of a_i, b_i, c_i, and d_i, respectively, hence

$$p = \sum_{i=1}^{k} \lambda_i p_i. \tag{5.24}$$

In order to reach the contradiction with the original assumption, it is enough to notice that $p \in \mathrm{int}\mathbb{T}_n$, i.e. not all p_i can belong to the same facet of $\mathbb{T}_n$. Indeed, if all p_i belong to the same facet of the tetrahedron, one of their coordinates (the same one for all i) should be zero (for instance, $a_i = 0$). Then, from eqs. (5.21) and (5.23) it follows that all p_i's are equal, contradicting the assumption that (5.14) is a proper convex combination. Hence, p belongs to the interior of the tetrahedron, which contradicts the assumption that $p \in \partial\mathbb{T}_n$ and completes the proof. □

One immediate consequence of this theorem relates to the computation of the classical bound β_C. Indeed, since all the symmetric one and two body correlators $\mathcal{S}_k$ and $\mathcal{S}_{kl}$ are parametrized by (a, b, c, d), finding the classical bound of the Bell inequality (5.3) is equivalent to minimizing I being a function of a, b, c, and d over the boundary of $\mathbb{T}_n$, i.e. $\beta_C = -\min_{\partial \mathbb{T}_n} I$.

A class of Bell inequalities.

Using the previous characterization of the symmetric polytope of two-body local models, we can now search for particular Bell inequalities violated by multipartite quantum states. For sufficiently low number of parties, all Bell inequalities corresponding to the facets of $\mathbb{P}_2^S$ can be listed (see appendix G). Indeed, the number of parties for which the problem is still computationally tractable is much larger than for the complete polytope $\mathbb{P}$. In what follows, we provide a general class of few-parameter symmetric Bell inequalities and show that they reveal nonlocality in quantum states for any n.

We start from the general form of a symmetric two-body Bell inequality (5.3), and consider a particular parametrization of the coefficients:

$$\begin{aligned} \gamma &= x^2 \\ \varepsilon &= y^2 \\ \delta &= \sigma x y \\ \alpha_\pm &= x[\sigma\mu \pm (x+y)] \end{aligned} \tag{5.25}$$

where x, y are positive natural numbers, and $\sigma = \pm 1$ defines the sign of δ. We assume that $\mu \equiv \beta/y$ is an integer with opposite parity to ε for odd n and to γ for even n. In what follows, we will find an expression for the classical bound β_C in terms of these parameters x, y, σ, μ and the number of parties n.

First, note that for all local deterministic models, the left-hand side of (5.3) equals to

$$I = \alpha \mathcal{S}_0 + \beta \mathcal{S}_1 + \frac{\gamma}{2}(\mathcal{S}_0^2 - n) + \delta(\mathcal{S}_0 \mathcal{S}_1 - z) + \frac{\varepsilon}{2}(\mathcal{S}_1^2 - n), \tag{5.26}$$

where $z = a - b - c + d$. In terms of the parameters (5.25), I can be rewritten as

$$\begin{aligned} I \;\; = \;\; & \frac{x^2}{2}\mathcal{S}_0^2 + \sigma x y \mathcal{S}_0 \mathcal{S}_1 + \frac{y^2}{2}\mathcal{S}_1^2 - \frac{n}{2}(x+y) - \sigma x y z \\ & + x[\sigma\mu \pm (x+y)]\mathcal{S}_0 + \beta \mathcal{S}_1 \,, \end{aligned} \tag{5.27}$$

which in turn is equivalent to

$$\begin{aligned} I &= \frac{1}{2}\left(x\mathcal{S}_0+\sigma y\mathcal{S}_1+\sigma\mu\pm x\right)^2-\frac{1}{2}(\sigma\mu\pm x)^2 \\ &\quad +xy(\pm\mathcal{S}_0\mp\sigma\mathcal{S}_1-\sigma z)-\frac{1}{2}n(x+y). \end{aligned} \tag{5.28}$$

Moreover, the single body mean values $\mathcal{S}_j$ may also be parametrized in terms of a, b, c and d, as we did when studying the vertices of $\mathbb{P}_2^S$. Hence, combining parametrizations (5.7) and (5.25) we find that $\pm\mathcal{S}_0\mp\sigma\mathcal{S}_1-\sigma z=4r-n$, where r depends on α and the sign of δ (i.e., σ) as follows:

$$r=\begin{cases} b, & \text{for} \quad \alpha_+,\sigma=1 \\ a, & \text{for} \quad \alpha_+,\sigma=-1 \\ c, & \text{for} \quad \alpha_-,\sigma=1 \\ d, & \text{for} \quad \alpha_-,\sigma=-1 \end{cases}. \tag{5.29}$$

All together, for deterministic models the left-hand side of (5.3) equals

$$\begin{aligned} I &= \frac{1}{2}\left(x\mathcal{S}_0+\sigma y\mathcal{S}_1+\sigma\mu\pm x\right)^2+4xyr \\ &\quad -\frac{1}{2}\left[(\sigma\mu\pm x)^2+n(x+y)^2\right]. \end{aligned} \tag{5.30}$$

The goal is to prove that the classical bound in (5.3) is

$$\beta_C=\frac{1}{2}\left[n(x+y)^2+(\sigma\mu\pm x)^2\right]-\frac{1}{2}. \tag{5.31}$$

Hence, it is enough to show that

$$\left(x\mathcal{S}_0+\sigma y\mathcal{S}_1+\sigma\mu\pm x\right)^2+8xyr\geq 1. \tag{5.32}$$

Since both x and y are positive integers, the above inequality is trivially satisfied if $r\neq 0$. For $r=0$ (i.e., when optimizing over the facets of $\mathbb{T}_n$), the expression inside the parentheses is integer, and in what follows we will prove that the above assumptions guarantee that its parity is always odd. Let us consider the cases of odd and even n separately. For odd n, both $\mathcal{S}_0$ and $\mathcal{S}_1$ are odd. Hence, $x\mathcal{S}_0+\sigma y\mathcal{S}_1+\sigma\mu\pm x$ has the same parity as $y+\mu$. By assumption, μ has opposite parity to ε. Since $\varepsilon=y^2$, and both y^2 and y have the same parity, it follows that $y+\mu$ is odd. Therefore, the expression inside the parentheses is odd. For even n, $\mathcal{S}_0$ as well as $\mathcal{S}_1$ are even, implying that $x\mathcal{S}_0+y\sigma\mathcal{S}_1$ is even. As before, the assumptions imply that $\sigma\mu\pm x$ is odd, hence the expression inside the parentheses is odd. In particular, this means that for any n, it can never

n	# Bell inequalities in the class	Total # of tight Bell inequalities
5	16	152
10	272	2018
15	1208	7744
20	3592	21274

Table 5.1.: The number of facets (second column) of $\mathbb{P}_2^S$ that are grasped by our class of Bell inequalities for various numbers of parties n. The third column contains the total number of facets of $\mathbb{P}_2^S$.

take the value zero, which implies that eq. (5.32) is satisfied. It follows that all classical correlations satisfy

$$\begin{aligned} I := & x[\sigma\mu \pm (x+y)]\mathcal{S}_0 + \mu y \mathcal{S}_1 + \frac{x^2}{2}\mathcal{S}_{00} + \sigma x y \mathcal{S}_{01} + \frac{y^2}{2}\mathcal{S}_{11} \\ \geq & \frac{1}{2} - \frac{1}{2}\left[n(x+y)^2 + (\sigma\mu \pm x)^2\right]. \end{aligned} \tag{5.33}$$

where x and y are positive natural number, $\sigma = \pm 1$, and μ is an integer with opposite parity to y^2 for odd n and to x^2 for even n.

In order to see that (5.31) is a classical bound, in the sense of $\beta_C = -\min_{\partial\mathbb{T}_n} I$, we still need to prove that it is saturated by a particular number of deterministic correlations. However, this is not the case in general, hence the inequality (5.33) is usually not a tight Bell inequality. Table (5.1) summarizes, for few cases, the number of Bell inequalities in this family that define facets of $\mathbb{P}_2^S$. Even though in general the inequalities in the form of (5.33) may not be optimal, they are still useful for detecting quantum nonlocality, as we will see in the following example.

Quantum violations.

A particular case of a Bell inequality of this class arises from the choice of parameters $x = y = -\sigma = 1$, and $\alpha_- = -2$. From eq. (5.31) the classical bound is $\beta_C = 2n$, and the resulting Bell inequality reads

$$-2\mathcal{S}_0 + \frac{1}{2}\mathcal{S}_{00} - \mathcal{S}_{01} + \frac{1}{2}\mathcal{S}_{11} + 2n \geq 0. \tag{5.34}$$

In order to search for quantum violations of ineq. (5.34), we assume that all parties measure the same pairs of observables, i.e., $\mathcal{M}_j^{(i)} = \mathcal{M}_j$ for every $i = 1, \ldots, n$. Without loss of generality, we take them as $\mathcal{M}_0 = \sigma_z$ and

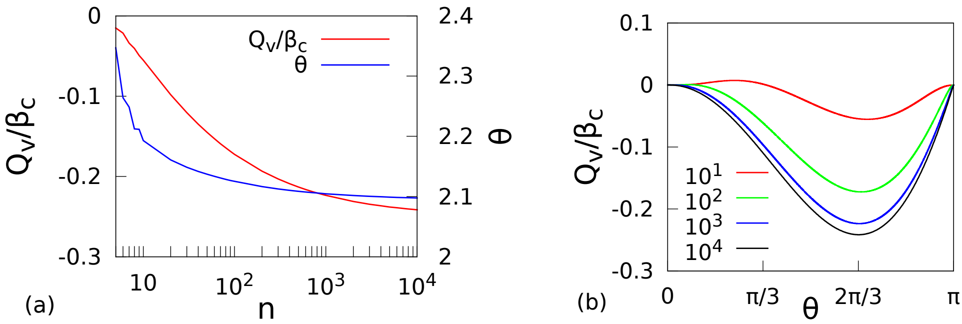


Figure 5.1.: (a) The effective (divided by the classical bound) maximal violation of Ineq. (5.34) (red line) and the corresponding angle θ in $\mathcal{M}_1$ (blue line) as functions of n. (b) Effective violation of Ineq. (5.34) as a function of θ for $n = 10^k$ with $k = 1, 2, 3, 4$. For large n the violation is robust against misalignments of the second observable.

$\mathcal{M}_1(\theta) = \cos\theta\sigma_z + \sin\theta\sigma_x$ for $\theta \in [0, \pi]$ (Mas05), and denote by $\mathcal{B}_n(\theta)$ the Bell operator constructed from these observables (see section 2.1.5). The inequality (5.34) is then violated if there exists θ such that $\mathcal{B}_n(\theta) \not\succeq 0$. Fig. 5.1 presents the lowest negative eigenvalue of $\mathcal{B}_n(\theta)$ for various values of n. Numerically, we see that the effective violation (i.e. the violation divided by the classical bound) grows with n, and becomes more robust against misalignments of θ for large n. In addition, the corresponding eigenstate of $\mathcal{B}_n(\theta)$, i.e. the pure state maximally violating (5.34) is always symmetric[2], that is, it belongs to the symmetric subspace of an n-qubit Hilbert space. Since any (also mixed) n-qubit symmetric state is entangled if and only if it is genuinely multipartite entangled (ESBL02; ATSL12), our Bell inequalities detect genuinely multipartite entangled states. Another interesting feature of these quantum violations is that they are attained by states which are symmetric under any permutation of the parties. Since for these states all the two-body reduced states are local in the considered scenario, the corresponding marginals of the probability distributions are local as well. Hence, our Bell inequalities are able to detect nonlocal correlations only from their local two-body marginals. However, these nonlocal correlations need not be truly multipartite.

[2]Note however that there also exist antisymmetric states violating this inequality.

5.3. Nonlocality of physically relevant systems

In the previous section we showed how to construct Bell inequalities from one and two-body correlators, which are symmetric under permutation of the parties and detect genuinely multipartite entanglement. In this section we will see that these inequalities are powerful enough to reveal nonlocality in "physically relevant" states, such as ground states of spin models that naturally appear in many-body physics.

In what follows, we focus on inequalities violated by a family of states called Dicke states (Dic54), hence I first briefly comment on the latter. Dicke states are n-qubit states which remain invariant under any permutation of the qubits. They span the $(n+1)$-dimensional symmetric subspace of $(\mathbb{C}^2)^{\otimes n}$ and read

$$\left|D_n^k\right\rangle = \mathcal{S}(|\{0, n-k\}, \{1, k\}\rangle) \quad (k = 0, \ldots, n), \tag{5.35}$$

where $|\{0, n-k\}, \{1, k\}\rangle$ is any pure product vector with $n-k$ qubits in the state $|0\rangle$ and k in the state $|1\rangle$, while $\mathcal{S}$ denotes symmetrization over all parties. It is worth mentioning that $\left|D_n^k\right\rangle$ are genuinely multipartite entangled for any $k \neq 0, n$. Moreover, their entanglement properties have been extensively studied in the literature (see (LORV05; BG13) and references therein), and the state $\left|D_6^3\right\rangle$ was recently generated experimentally (HHR$^+$05; KST$^+$07).

In many-body physics, the Dicke states arise naturally as the lowest-energy eigenstates of the isotropic Lipkin-Meshkov-Glick Hamiltonian (LMG65):

$$H = -\frac{\lambda}{n} \sum_{\substack{i,j=1 \\ i<j}}^{n} \left(\sigma_x^{(i)} \sigma_x^{(j)} + \sigma_y^{(i)} \sigma_y^{(j)} \right) - h \sum_{i=1}^{n} \sigma_z^{(i)}, \tag{5.36}$$

which describes n spins interacting through the two-body ferromagnetic coupling ($\lambda > 0$), embedded into the magnetic field acting along the z direction of strength $h \geq 0$. Again, $\sigma_a^{(i)}$ ($a = x, y, z$) are the Pauli matrices acting at site i. In what follows we consider the case of weak magnetic field applied to the system, precisely $h \leq \lambda/n$. Hence, the ground state of H is $\left|D^{n/2}\right\rangle$ for even n and $\left|D^{\lceil n/2 \rceil}\right\rangle$ for odd n, except for the case of $h = 0$ and odd n, for which the lowest energy is two-fold degenerate and the corresponding subspace is spanned by $\left|D_n^k\right\rangle$, with $k = \lfloor n/2 \rfloor$ and $k = \lceil n/2 \rceil$.

In what follows, we present a new class of tight two-body symmetric Bell inequalities that we will use to detect nonlocality of the above mentioned Dicke states. This family is obtained by setting in eq. (5.3), for each n, the following coefficients:

$$\begin{aligned} \alpha_n &= n(n-1)(\lceil n/2 \rceil - n/2) \\ \beta_n &= \alpha_n / n \\ \gamma_n &= n(n-1)/2 \\ \delta_n &= n/2 \\ \varepsilon_n &= -1 \end{aligned} \quad . \tag{5.37}$$

We will see that, for any number of parties, this choice of parameters is consistent with the classical bound

$$\beta_C^n = \frac{1}{2} n(n-1) \left\lceil \frac{n+2}{2} \right\rceil . \tag{5.38}$$

These Bell inequalities are independent of the class presented in section 5.2, and for $n = 2$ they reproduce the CHSH Bell inequality (CHSH69).

Proposition 5.2. *The classical bound of a two-body symmetric Bell inequality*

$$I := \alpha \mathcal{S}_0 + \beta \mathcal{S}_1 + \frac{\gamma}{2} \mathcal{S}_{00} + \delta \mathcal{S}_{01} + \frac{\varepsilon}{2} \mathcal{S}_{11} \geq -\beta_C,$$

with coefficients given by

$$\begin{aligned} \alpha_n &= n(n-1)(\lceil n/2 \rceil - n/2) \\ \beta_n &= \alpha_n / n \\ \gamma_n &= n(n-1)/2 \\ \delta_n &= n/2 \\ \varepsilon_n &= -1 \end{aligned}$$

is $\beta_C^n = \frac{1}{2} n(n-1) \left\lceil \frac{n+2}{2} \right\rceil$. This family of Bell inequalities defines facets of $\mathbb{P}_2^S$, i.e. the inequalities are tight for every n.

Proof. The explicit form of the inequality with coefficients given by (5.37) is, for even and odd n,

$$I_n^e = \frac{n(n-1)}{4} \mathcal{S}_{00} + \frac{n}{2} \mathcal{S}_{01} - \frac{1}{2} \mathcal{S}_{11} \geq -\beta_C, \tag{5.39}$$

$$\begin{aligned} I_n^o \;\; &= \;\; \frac{1}{2} \binom{n}{2} \mathcal{S}_{00} + \frac{n}{2} \mathcal{S}_{01} - \frac{1}{2} \mathcal{S}_{11} \\ &\quad + \frac{n(n-1)}{2} \mathcal{S}_0 + \frac{n-1}{2} \mathcal{S}_1 \geq -\beta_C . \end{aligned} \tag{5.40}$$

We aim to minimize the left hand size of these inequalities, over all possible deterministic strategies. First, note that for classical correlations

$$\begin{aligned} & -n \leq \mathcal{S}_{00}, \mathcal{S}_{11} \leq n(n-1), \\ & |\mathcal{S}_{01}| \leq n(n-1), \\ & |\mathcal{S}_k| \leq n, \quad \text{for} k = 0, 1. \end{aligned} \tag{5.41}$$

Hence, the first term in eqs. (5.39) and (5.40) is the dominant one: it is of order four in n, while the remaining ones are of second or third order in n. This suggests to make the term containing $\mathcal{S}_{00}$ small in order to minimize $I_n^{e/o}$. Since $\mathcal{S}_{00} = \mathcal{S}_0^2 - n$, we will consider $\mathcal{S}_0$ as a parameter, and among all the solutions parametrized by $\mathcal{S}_0$ we will choose the smallest one.

For reasons that will soon become clear, recall the parametrization of the vertices of $\mathbb{P}_2^S$ in terms of a, b, c and d defined in (5.6). Since $\mathcal{S}_0 = a + b - c - d$ and $a + b + c + d = n$, it follows

$$a = \frac{1}{2}(n + \mathcal{S}_0) - b, \qquad c = \frac{1}{2}(n - \mathcal{S}_0) - d, \tag{5.42}$$

Hence, the two-body expectation values are rewritten as

$$\mathcal{S}_{00} = \mathcal{S}_0^2 - n, \qquad \mathcal{S}_{11} = [n - 2(b + d)]^2 - n, \tag{5.43}$$

$$\mathcal{S}_{01} = \mathcal{S}_0[n - 1 - 2(b + d)] + 2(b - d), \tag{5.44}$$

where we now consider b and d as free variables that are nonnegative integers constrained as

$$0 \leq b \leq \frac{1}{2}(n + \mathcal{S}_0), \qquad 0 \leq d \leq \frac{1}{2}(n - \mathcal{S}_0). \tag{5.45}$$

From theorem 5.1 it follows that, in order to find β_C, it suffices to minimize $I_n^{e/o}$ over the 4-tuples (a, b, c, d) belonging to the boundary of the tetrahedron, i.e., those for which $abcd = 0$. Eqs. (5.42) imply that the cases of $a = 0$ and $d = 0$ are now equivalent to $b = (1/2)(n + \mathcal{S}_0)$ and $d = (1/2)(n - \mathcal{S}_0)$, respectively. Within this framework, treating $\mathcal{S}_0$ as a parameter means that we intersect the three-dimensional tetrahedron with hyperplanes of constant $\mathcal{S}_0$ and look for the minimal value of $I_n^{e/o}$ for points lying on the boundary of the resulting two-dimensional object. Then, we choose the optimal solution among those parametrized by $\mathcal{S}_0$.

In what follows we will compute $\min_{\partial \mathbb{T}_n} I_n^{e/o}$ separately for the cases of even and odd n, and in each one we consider all the facets of the tetrahedron separately. We will prove that

$$\min_{\partial \mathbb{T}_n} I_n^{e/o} = -\frac{1}{4} \begin{cases} n(n-1)(n+2), & n \,\text{even} \\ n(n-1)(n+3), & n \,\text{odd}, \end{cases} \tag{5.46}$$

Even n.– First, we express I_n^e as a function of b, d and $\mathcal{S}_0$, from equations (5.43), (5.44) and (5.39):

$$\begin{aligned} I_n^e(b,d;\mathcal{S}_0) = \frac{1}{2}\Big\{ & \frac{n(n-1)}{2}(\mathcal{S}_0^2-n) - [n-2(b+d)]^2 \\ & + n\,[\mathcal{S}_0(n-2(b+d)-1)+2(b-d)] + n\Big\}. \end{aligned} \tag{5.47}$$

Case a=0. This case is equivalent to $b=(n+\mathcal{S}_0)/2$. Direct evaluation of eq. (5.47) gives

$$\begin{aligned} I_n^e(\tfrac{n+\mathcal{S}_0}{2},d;\mathcal{S}_0) &= \frac{1}{4}(n^2-3n-2)(\mathcal{S}_0^2-n) \\ &\quad -d[2d+\mathcal{S}_0(n+2)+n]. \end{aligned}$$

This is a quadratic function in d which, since its second derivative with respect to d is negative, has a local maximum. Hence, it attains its minimal value either at $d=0$ or $d=(n-\mathcal{S}_0)/2$.

In the first case, $I_n^e(\frac{n+\mathcal{S}_0}{2},0;\mathcal{S}_0)=(1/4)(\mathcal{S}_0^2-n)(n^2-3n-2)$, which is minimal for $\mathcal{S}_0=0$. The value for this minimum is $I_n^e(\frac{n+\mathcal{S}_0}{2},0;0)=-(1/4)n(n^2-3n-2)$, which is larger that that presented in eq. (5.46).

In the second case $d=(n-\mathcal{S}_0)/2$, i.e.

$$I_n^e(\tfrac{n+\mathcal{S}_0}{2},\tfrac{n-\mathcal{S}_0}{2};\mathcal{S}_0) = -\frac{n(n-1)}{4}[n+2-\mathcal{S}_0(\mathcal{S}_0-2)]. \tag{5.48}$$

This expression attains its minimum value at either $\mathcal{S}_0=0$ or $\mathcal{S}_0=2$, which corresponds to $I_n^e(\frac{n+\mathcal{S}_0}{2},\frac{n-\mathcal{S}_0}{2};\mathcal{S}_0)=-\frac{1}{4}n(n-1)(n+2)\quad(\mathcal{S}_0=0,2)$, i.e. the value in eq. (5.46).

Therefore, two different 4-tuples

$$\left(0,\frac{n}{2},0,\frac{n}{2}\right), \qquad \left(0,\frac{n}{2}+1,0,\frac{n}{2}-1\right), \tag{5.49}$$

attain the value of I_n^e given in eq. (5.46).

Case b=0. Direct evaluation of eq. (5.47) gives

$$\begin{aligned} I_n^e(0,d;\mathcal{S}_0) = \frac{1}{2}\Big\{ & \frac{n(n-1)}{2}(\mathcal{S}_0^2-n) - (n-2d)^2 + n \\ & + n[\mathcal{S}_0(n-2d-1)-2d]\Big\}. \end{aligned} \tag{5.50}$$

The second derivative of $I_n^e(0,d;\mathcal{S}_0)$ with respect to d is negative, and therefore we look for its minimal value at the boundary of the range of d. For $d=0$

the expression reduces to the right-hand side of eq. (5.48), which as previously mentioned, has minima for $\mathcal{S}_0 = 0$ and $\mathcal{S}_0 = 2$. Hence, there are two additional elements of $\mathbb{T}_4$ for which I_n^e attains the value in eq. (5.46):

$$\left(\frac{n}{2} - 1, 0, \frac{n}{2} + 1, 0\right), \qquad \left(\frac{n}{2}, 0, \frac{n}{2}, 0\right) \tag{5.51}$$

For $d = (n - \mathcal{S}_0)/2$, direct evaluation of eq. (5.50) gives $I_n^e(0, \frac{n-\mathcal{S}_0}{2}; \mathcal{S}_0) = \frac{1}{4}(\mathcal{S}_0^2 - n)(n^2 + n - 2)$, which attains a minimum at $\mathcal{S}_0 = 0$. The value of this minimum is that of eq. (5.46), hence we found the fifth point saturating the Bell inequality (5.47):

$$\left(\frac{n}{2}, 0, 0, \frac{n}{2}\right). \tag{5.52}$$

Cases c=0 or d=0. The case $c = 0$ is equivalent to $d = (n - \mathcal{S}_0)/2$. Hence, a similar argument as before implies that, for both $d = (n - \mathcal{S}_0)/2$ or $d = 0$, the minimum value of I_n^e is $-(1/4)n(n-1)(n+2)$, i.e that of eq. (5.46). This value is attained at the five vectors (5.49), (5.51), and (5.52).

Odd n.– First, we express I_n^o as a function of b, d and $\mathcal{S}_0$, from equations (5.43), (5.44) and (5.40):

$$\begin{aligned} I_n^o(b, d; \mathcal{S}_0) = &\frac{1}{4}n(n-1)[\mathcal{S}_0(\mathcal{S}_0 + 4) - n] - 2(b^2 + d^2) \\ &- b[4d - 1 + n(\mathcal{S}_0 - 2)] - d(n\mathcal{S}_0 - 1) \end{aligned} \tag{5.53}$$

Notice that since n is odd, $\mathcal{S}_0$ is also odd.

Case a=0. This case is equivalent to $b = (n + \mathcal{S}_0)/2$. Direct evaluation of eq. (5.53) gives

$$\begin{aligned} I_n^o(\tfrac{n+\mathcal{S}_0}{2}, d; \mathcal{S}_0) &= -d[2d + \mathcal{S}_0(n+2) + 2n - 1] \\ &\quad + \frac{1}{4}(n^2 - 3n - 2)(\mathcal{S}_0^2 - n) \\ &\quad + \frac{1}{2}(n-1)^2\mathcal{S}_0 \\ &= I_n^e(\tfrac{n+\mathcal{S}_0}{2}, d; \mathcal{S}_0) + d + \frac{1}{2}(n-1)^2\mathcal{S}_0. \end{aligned}$$

Since the second derivative of $I_N^o(\frac{n+\mathcal{S}_0}{2}, d; \mathcal{S}_0)$ with respect to d is negative for any $\mathcal{S}_0$, it attains its minimum value either at $d = 0$ or $d = (n - \mathcal{S}_0)/2$.

For the first case $I_n^o(\frac{n+\mathcal{S}_0}{2}, 0; \mathcal{S}_0) = \frac{1}{4}(\mathcal{S}_0^2 - n)(n^2 - 3n - 2) + \frac{1}{2}(n-1)^2\mathcal{S}_0$, which is minimal at $\mathcal{S}_0 = 0$. The value for this minimum is $I_n^o(\frac{n+\mathcal{S}_0}{2}, 0; 0) = -(1/4)n(n^2 - 3n - 2)$, which is larger that that of eq. (5.46).

For $d = (n - \mathcal{S}_0)/2$,

$$I_n^o(\tfrac{n+\mathcal{S}_0}{2}, \tfrac{n-\mathcal{S}_0}{2}; \mathcal{S}_0) = -\frac{1}{4}n(n-1)(n+4-\mathcal{S}_0^2). \tag{5.54}$$

This expression is minimal for $\mathcal{S}_0 = 0$, but since $\mathcal{S}_0$ must be odd, we obtain the lowest value for $\mathcal{S}_0 = \pm 1$. As a result, I_n^o attains as minimum value that of eq.(5.46), at the following two elements of $\mathbb{T}_n$

$$\left(0, \frac{n \pm 1}{2}, 0, \frac{n \mp 1}{2}\right). \tag{5.55}$$

Case b=0. Direct evaluation of eq. (5.53) gives

$$\begin{aligned} I_n^o(0, d; \mathcal{S}_0) &= \frac{1}{4}n(n-1)[\mathcal{S}_0(\mathcal{S}_0+4) - n] - 2d^2 \\ &\quad -d(n\mathcal{S}_0 - 1). \end{aligned} \tag{5.56}$$

This is a quadratic function in d that has a local maximum. Hence, $I_n^o(0, d; \mathcal{S}_0)$ attains its minimal value at the boundary of the range of d, i.e. either at $d = 0$ or $d = (n - \mathcal{S}_0)/2$.

For $d = 0$, $I_n^o(0, 0; \mathcal{S}_0) = \frac{1}{4}n(n-1)[\mathcal{S}_0(\mathcal{S}_0+4) - n]$, which, since $\mathcal{S}_0$ must be odd, is minimal at $\mathcal{S}_0 = -3$ or $\mathcal{S}_0 = -1$. The corresponding minimum value is $-(1/4)n(n-1)(n+3)$, which coincides with eq.(5.46). This value is attained at the following two vertices:

$$\left(\frac{n-1}{2}, 0, \frac{n+1}{2}, 0\right), \quad \left(\frac{n-3}{2}, 0, \frac{n+3}{2}, 0\right) \tag{5.57}$$

For $d = (n - \mathcal{S}_0)/2$, direct evaluation of eq. (5.56) gives $I_n^o(0, \frac{n-\mathcal{S}_0}{2}; \mathcal{S}_0) = -\frac{n-1}{4}[(n+2)(n-\mathcal{S}_0^2) - 2(n+1)\mathcal{S}_0]$. Since $\mathcal{S}_0$ is odd, I_n^o is minimal at $\mathcal{S}_0 = -1$. The corresponding minimum value is $-(1/4)n(n-1)(n+3)$, i.e. that of eq.(5.46), and is attained by

$$\left(\frac{n-1}{2}, 0, 0, \frac{n+1}{2}\right). \tag{5.58}$$

Cases c=0 or d=0. Similarly, for either $c = 0$ or $d = 0$, the minimal value of I_n^o is $-(1/4)n(n-1)(n+3)$, which coincides with eq.(5.46). This minimum is attained by the five vertices (5.55), (5.57), and (5.58).

We have proved that the lowest value of $I_n^{e/o}$ for both even and odd n is the one given in eq. (5.46) and is realized by five elements of $\mathbb{T}_n$: (5.49), (5.51), and (5.52) for even n, and (5.55), (5.57), and (5.58) for odd n. Hence, for any n the Bell inequality $I_n^{e/o}$ is tangent to $\mathbb{P}_2^S$ on five vertices. Since these are linearly independent, $I_n^{e/o}$ indeed represents a facet of $\mathbb{P}_2^S$, which completes the proof. □

We now move on to proving that this class of tight two-body symmetric Bell inequalities is indeed violated by Dicke states. Similar to section 5.2, we assume that each observer has the same pair of qubit observables, that is, $\mathcal{M}_0^{(i)} = \mathcal{M}_0$ and $\mathcal{M}_1^{(i)} = \mathcal{M}_1$ for every $i = 1, \ldots, n$. Hence, the resulting Bell operator $\mathcal{B}_n$ (see section 2.1.5) is permutationally invariant, which together with the fact that the Dicke states are also fully symmetric, significantly simplifies the problem. Indeed, an immediate consequence is

$$\langle D_n^k|\mathcal{B}_n|D_n^k\rangle = \mathrm{tr}(\rho_n^k \widetilde{\mathcal{B}}_n), \tag{5.59}$$

where $\widetilde{\mathcal{B}}_n$ stands for the two-qubit "reduced Bell operator"

$$\begin{aligned} \widetilde{\mathcal{B}}_n \;=\; & \beta_C^n \mathbb{1}_4 + \tfrac{n}{2}\alpha_n(\mathcal{M}_0 \otimes \mathbb{1}_2 + \mathbb{1}_2 \otimes \mathcal{M}_0) \\ & + \tfrac{n(n-1)}{2}\left[\gamma_n \mathcal{M}_0 \otimes \mathcal{M}_0 + \varepsilon_n \mathcal{M}_1 \otimes \mathcal{M}_1 \right. \\ & \qquad \left. + \delta_n \left(\mathcal{M}_0 \otimes \mathcal{M}_1 + \mathcal{M}_1 \otimes \mathcal{M}_0\right)\right] \\ & + \tfrac{n}{2}\beta_n(\mathcal{M}_1 \otimes \mathbb{1}_2 + \mathbb{1}_2 \otimes \mathcal{M}_1), \end{aligned} \tag{5.60}$$

with $\mathbb{1}_d$ being a $d \times d$ identity matrix, and ρ_n^k denotes the reduced state of any two–qubit subsystem of $|D_n^k\rangle$. The state ρ_n^k can be computed analytically (see (WM02)), and for $k = \lceil n/2 \rceil$ has the following form

$$\rho_n^{\lceil n/2 \rceil} = \frac{1}{n(n-1)} \begin{pmatrix} p_n & 0 & 0 & 0 \\ 0 & q_n & q_n & 0 \\ 0 & q_n & q_n & 0 \\ 0 & 0 & 0 & r_n \end{pmatrix}, \tag{5.61}$$

where $p_n = (\lfloor n/2 \rfloor - 1)\lfloor n/2 \rfloor$, $q_n = \lfloor n/2 \rfloor \lceil n/2 \rceil$, and $r_n = (\lceil n/2 \rceil - 1)\lceil n/2 \rceil$.

Without loss of generality, we set the observables as $\mathcal{M}_0 = \sigma_z$ and $\mathcal{M}_1 = \cos\theta\sigma_z + \sin\theta\sigma_x$ with $\theta \in [0, \pi]$. Hence,

$$\langle D_n^{\lceil n/2 \rceil}|\mathcal{B}_n|D_n^{\lceil n/2 \rceil}\rangle = 4\lfloor n/2 \rfloor \sin^2(\theta/2)[(\lceil n/2 \rceil + 1)\sin^2(\theta/2) - 1], \tag{5.62}$$

which attains its minimum at

$$\theta_{\min}^n = \pm \arccos\left(\frac{\lceil n/2 \rceil}{\lceil n/2 \rceil + 1}\right). \tag{5.63}$$

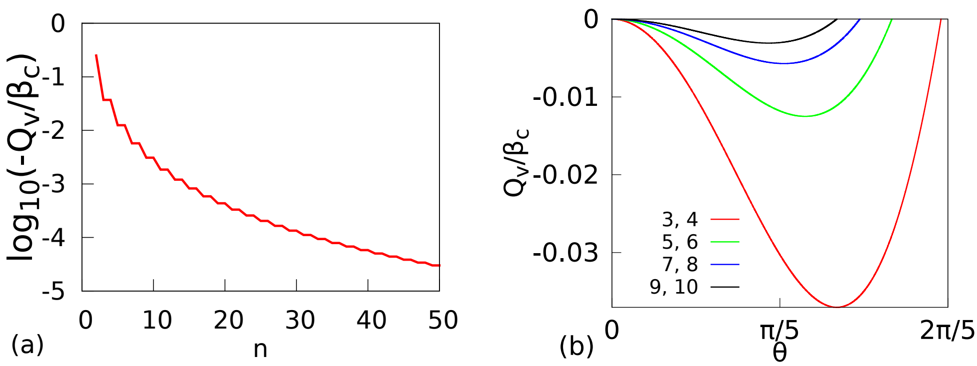


Figure 5.2.: (a) Effective violation of Ineq. (5.34) by the Dicke states $\left|D_n^{\lceil n/2\rceil}\right\rangle$ as a function of the number of parties n. The violation decays with n as $1/n^3$. (b) Effective violation as a function of θ for various values of n.

The value at this minimum is

$$\langle D_N^{\lceil n/2\rceil}|\mathcal{B}_N(\theta)|D_N^{\lceil n/2\rceil}\rangle = -\frac{\lfloor n/2\rfloor}{\lceil n/2\rceil + 1}. \tag{5.64}$$

which proves quantum violation.

The same quantum violation can be achieved for $k = \lfloor n/2\rfloor$. Indeed, $\left|D_N^{\lfloor n/2\rfloor}\right\rangle$ is obtained from $\left|D_n^{\lceil n/2\rceil}\right\rangle$ by swapping the elements of the computational basis $\{|0\rangle, |1\rangle\}$. Hence, it suffices to modify the observables $\mathcal{M}_0$ and $\mathcal{M}_1$ accordingly.

We see then that this class of Bell inequalities is violated by Dicke states for any n, although the effective violation decays with n as $1/n^3$ (see Fig. 5.2). It should be stressed that, even though the previous analysis may suggest that this violation is purely bipartite, this is certainly not the case. Dicke states are symmetric, and therefore any marginal bipartite correlations obtained from them in a Bell experiment with the same two dichotomic observables per site are local; otherwise all bipartite marginal correlations would be nonlocal, contradicting the fact that in this case quantum correlations are monogamous (TV06). Hence, our results provide further examples of local marginal bipartite correlations that are only compatible with global nonlocal correlations, like those obtained in the previous section.

One final comment before moving on to translational invariance. Both in this and the previous section we have focused on Bell inequalities that are invariant under any permutation of the parties, in particular on violations attained when all the parties measure the same pair of observables. In this case, the one and two-body correlators which appear in the Bell inequalities can be expressed in terms of total spin operators $S_\alpha = (1/2)\sum_{i=1}^{n} \sigma_\alpha^{(i)}$ with $\alpha = x, y, z$ and their combinations $\boldsymbol{m} \cdot \boldsymbol{S}$ in any direction $\boldsymbol{m}$, where $\boldsymbol{S} = [S_x, S_y, S_z]$. To be precise, in the previous examples we studied the observables $\mathcal{M}_0 = \sigma_z$ and $\mathcal{M}_1 = \boldsymbol{m}\cdot\boldsymbol{\sigma}$, with $\boldsymbol{m} = [\sin\theta, 0, \cos\theta]$ and $\boldsymbol{\sigma} = [\sigma_x, \sigma_y, \sigma_z]$. Direct evaluation leads to:

$$
\begin{aligned}
\mathcal{S}_0 &= 2\langle S_z\rangle, \\
\mathcal{S}_1 &= 2\langle \boldsymbol{m}\cdot\boldsymbol{S}\rangle, \\
\mathcal{S}_{00} &= 4\langle S_z^2\rangle - n, \\
\mathcal{S}_{11} &= 4\langle(\boldsymbol{m}\cdot\boldsymbol{S})^2\rangle - n, \\
\mathcal{S}_{01} &= (1/4)[\langle(S_z + \boldsymbol{m}\cdot\boldsymbol{S})^2\rangle - \langle(S_z - \boldsymbol{m}\cdot\boldsymbol{S})^2\rangle.
\end{aligned}
$$

Hence, the violations of our Bell inequalities can be computed from quantities that can be measured experimentally with current technologies (see the case of spin polarization spectroscopy (HSP10; ERIR$^+$08)). This reveals another advantage of our approach, that is of great experimental convenience.

5.4. Translationally invariant two-body Bell inequalities

In sections 5.2 and 5.3 I studied two-body correlators Bell inequalities which satisfy the additional assumption of permutational symmetry. In this section, I move on to study another type of two-body correlators Bell inequalities, namely those that are translationally invariant.

Translational invariance imposes that the Bell inequalities remain the same if the following transformations are simultaneously applied:

$$\mathcal{M}_j^{(i)} \to \mathcal{M}_j^{(i+1)} \qquad (j = 0, 1), \tag{5.65}$$

where the convention that $\mathcal{M}_j^{(n+k)} = \mathcal{M}_j^{(k)}$ for any $k = 1, \ldots, n$ and $j = 0, 1$ is assumed. This translates into certain conditions on the parameters appearing in ineq. (5.2):

$$\alpha_i = \alpha_{i+1}, \quad \beta_i = \beta_{i+1} \qquad (i = 1, \ldots, n-1). \tag{5.66}$$

These imply that all α_i and β_i are equal. Hence, γ_{ij}, δ_{ij}, and ϵ_{ij} satisfy the following cycles of equalities:

$$
\begin{aligned}
\gamma_{1,1+k} &= \gamma_{2,2+k} = \ldots = \gamma_{n-k,n} \\
&= \gamma_{1,n-k+1} = \gamma_{2,n-k+2} = \ldots = \gamma_{k,n}
\end{aligned} \tag{5.67}
$$

and

$$\begin{aligned} \epsilon_{1,1+k} &= \epsilon_{2,2+k} = \ldots = \epsilon_{n-k,n} \\ &= \epsilon_{1,n-k+1} = \epsilon_{2,n-k+2} = \ldots = \epsilon_{k,n} \end{aligned} \tag{5.68}$$

with $k = 1, \ldots, \lfloor n/2 \rfloor$, and also

$$\begin{aligned} \delta_{1,1+k} &= \delta_{2,2+k} = \ldots = \delta_{n-k,n} \\ &= \delta_{n-k+1,1} = \delta_{n-k+2,2} = \ldots = \delta_{n,k} \end{aligned} \tag{5.69}$$

with $k = 1, \ldots, n-1$.

In order to find the general form of a translationally invariant Bell inequality, denote $\alpha := \alpha_i$ and $\beta := \beta_i$ for any $i = 1, \ldots, n$, and by γ_k, ϵ_k ($k = 1, \ldots, \lfloor n/2 \rfloor$), and δ_k ($k = 1, \ldots, n-1$) those parameters that form cycles in Eqs. (5.67), (5.68), and (5.69) enumerated by k. Then, any translationally invariant Bell inequality can be written as

$$\alpha \mathcal{S}_0 + \beta \mathcal{S}_1 + \sum_{k=1}^{\lfloor n/2 \rfloor} \left(\gamma_k \mathcal{T}_{00}^{(k)} + \epsilon_k \mathcal{T}_{11}^{(k)} \right) + \sum_{k=1}^{n-1} \delta_k \mathcal{T}_{01}^{(k)} \geq -\beta_C, \tag{5.70}$$

with $\mathcal{S}_j$ being symmetrized local expectation values:

$$\mathcal{S}_j = \sum_{m=1}^{n} \langle \mathcal{M}_j^{(m)} \rangle \qquad (j = 0, 1), \tag{5.71}$$

and $\mathcal{T}_{ij}^{(k)}$ all translationally invariant two-body correlators:

$$\mathcal{T}_{ij}^{(k)} = \sum_{m=1}^{n} \langle M_i^{(m)} M_j^{(m+k)} \rangle \qquad (i \leq j = 0, 1), \tag{5.72}$$

with $k = 1, \ldots, \lfloor n/2 \rfloor$ for $i = j$ and $k = 1, \ldots, n-1$ for $i < j$. We have used the convention that for any $i, j = 0, 1$, $\langle M_i^{(m)} M_j^{(k)} \rangle \equiv \langle M_j^{(k)} M_i^{(m)} \rangle$ if $m > k$.

The polytope $\mathbb{P}_2$ is mapped under the translational symmetry onto the "translationally invariant" polytope $\mathbb{P}_{2,n}^T$ whose elements are vectors

$$(\mathcal{S}_0, \mathcal{S}_1, \mathcal{T}_{00}^{(1)}, \ldots, \mathcal{T}_{00}^{(\lfloor \frac{n}{2} \rfloor)}, \mathcal{T}_{01}^{(1)}, \ldots, \mathcal{T}_{01}^{(n-1)}, \mathcal{T}_{11}^{(1)}, \ldots, \mathcal{T}_{11}^{(\lfloor \frac{n}{2} \rfloor)}) \tag{5.73}$$

computed for all elements of $\mathbb{P}_2$. Note that, contrary to the fully symmetric case $\mathbb{P}_2^S$, here we make explicit use of the number of parties n when denoting the polytope $\mathbb{P}_{2,n}^T$, since its dimension is $n + 1 + 2\lfloor n/2 \rfloor$, i.e. it grows linearly with the number of parties n. In what follows, I present the complete description of $\mathbb{P}^T_{2,n}$ for $n = 3, 4$, and one class of translationally invariant two-body Bell inequalities for all n.

For the case $n = 3$, the general formula (5.70) reduces to

$$\alpha\mathcal{S}_0 + \beta\mathcal{S}_1 + \gamma\mathcal{T}_{00} + \epsilon\mathcal{T}_{11} + \delta_1\mathcal{T}_{01}^{(1)} + \delta_2\mathcal{T}_{01}^{(2)} + \beta_C \geq 0, \tag{5.74}$$

where we denote γ_1 and ϵ_1 by γ and ϵ, respectively, and also skip the superscripts in $\mathcal{T}_{00}^{(1)}$ and $\mathcal{T}_{11}^{(1)}$. The latter, as well as $\mathcal{T}_{01}^{(1)} + \mathcal{T}_{01}^{(2)}$, is permutationally invariant, hence for $\delta_1 = \delta_2$, one obtains a symmetric Bell inequality (see section 5.2). The dimension of $\mathbb{P}^T_{2,3}$ is 6 and there are $2^{2\cdot 3} = 64$ local deterministic strategies. However, *modulo* translational invariance, this number is reduced to $\frac{1}{3}(1\cdot 4^3 + 2\cdot 4) = 24$ different strategies. By using `cdd` (Fuk97), we found the 38 facets of $\mathbb{P}^T_{2,3}$ and also its 24 vertices, which implies that every local strategy corresponds to a vertex in $\mathbb{P}^T_{2,3}$. These facets are grouped into 6 equivalence classes (see appendix G). Interestingly, only class #6 is violated by quantum states, and its no-signaling bound is $\beta_{NS} = 13$. The other classes satisfy $\beta_C = \beta_{NS}$.

For the case $n = 4$, the general form of a translationally invariant two-body Bell inequality (5.70) is

$$\begin{aligned} &\alpha\mathcal{S}_0 + \beta\mathcal{S}_1 + \gamma_1\mathcal{T}_{00}^{(1)} + \gamma_2\mathcal{T}_{00}^{(2)} + \epsilon_1\mathcal{T}_{11}^{(1)} + \epsilon_2\mathcal{T}_{11}^{(2)} \\ &\delta_1\mathcal{T}_{01}^{(1)} + \delta_2 + \mathcal{T}_{01}^{(2)} + \delta_3\mathcal{T}_{01}^{(3)} + \beta_C \geq 0. \end{aligned} \tag{5.75}$$

Here, $\dim \mathbb{P}^T_{2,4} = 9$, while the number of different local deterministic strategies is $2^{2\cdot 4} = 256$. *Modulo* translational invariance, this number is reduced to $(1\cdot 4^4 + 1\cdot 4^2 + 2\cdot 4)/4 = 70$. By using `cdd` (Fuk97), we find that there are 1038 tight Bell inequalities, which we group in 103 classes (see appendix G). Similarly, we find that 68 out of the 70 local strategies correspond to extremal vertices of $\mathbb{P}^T_{2,4}$. The reason for this is that the deterministic local strategies $\{A = (+,+), B = (-,+), C = (+,-), D = (-,-)\}$ and $\{A = (+,+), B = (-,-), C = (+,-), D = (-,+)\}$ give exactly the same 1 and 2-body translationally invariant correlators, as well as $\{A = (+,+), B = (+,-), C = (-,+), D = (-,-)\}$ and $\{A = (+,+), B = (-,-), C = (-,+), D = (+,-)\}$ do. Hence, they correspond to the same vertex of $\mathbb{P}^T_{2,4}$. We computed the maximal quantum bound β_Q by optimizing measurements over qubits, since this proves sufficient in a scenario with 2 dichotomic observables (TV06). It is worth noticing the existence of 4 classes of inequalities which have no-signaling violation but for which quantum physics does not provide an advantage: These are classes $\#28, 63, 74, 76$. Hence, we seem to have found an example of an information task with no quantum advantage that is not in LO form (see section

3.5.1). However, these inequalities are not tight Bell inequalities when translated into the larger space of $\mathbb{P}_4$, hence they do not disprove our conjecture of section 3.5.1.

For the case of 5 parties, $\dim \mathbb{P}^T_{2,5} = 10$, with 1024 different local strategies, which *modulo* translational invariance become $(1 \cdot 4^5 + 4 \cdot 4)/5 = 208$. This corresponds to a polytope with 34484 facets, which can be grouped into 4198 different classes.

Since the complexity of the problem increases with n, a fact that we already notice in this case of 5 parties, we further simplify the problem by considering Bell inequalities that contain only correlators between nearest neighbours. This assumption is justified from an experimental point of view, since in many-body experiments one generally has access to statistic between nearest neighbours.

In this class of inequalities $\gamma_2 = \epsilon_2 = \delta_2 = \delta_3 = 0$, and the dimension of $\mathbb{P}^{T,NN}_{2,5}$ is constraint to be 6. In such scenario, the Bell inequality

$$35 - 2\mathcal{S}_0 - 6\mathcal{S}_1 - 2\mathcal{T}^{(1)}_{00} + 2\mathcal{T}^{(1)}_{01} + 4\mathcal{T}^{(4)}_{01} + 5\mathcal{T}^{(1)}_{11} \geq 0 \tag{5.76}$$

is violated by a translationally invariant qubit state, setting all pairs of observables to be the same for all parties. However, maximal quantum violation is achieved by breaking this symmetry.

Equation (5.76) can be generalized for any odd $n = 2k+1$ to:

$$8k^2 + 2k - 1 - k\mathcal{S}_0 - 3k\mathcal{S}_1 - k\mathcal{T}^{(1)}_{00} + k\mathcal{T}^{(1)}_{01} + 2k\mathcal{T}^{(2k)}_{01} + (2k+1)\mathcal{T}^{(1)}_{11} \geq 0. \tag{5.77}$$

In what follows, we will see that 5.77 is indeed a Bell inequality and search for quantum violations.

Classical bound.

We start by proving that the classical bound of the inequality

$$I_k := -k\mathcal{S}_0 - 3k\mathcal{S}_1 - k\mathcal{T}^{(1)}_{00} + k\mathcal{T}^{(1)}_{01} + 2k\mathcal{T}^{(2k)}_{01} + (2k+1)\mathcal{T}^{(1)}_{11} \geq -\beta_C, \tag{5.78}$$

is indeed given by

$$\beta_C = 8k^2 + 2k - 1. \tag{5.79}$$

Since the minimization of I_k is performed over deterministic classical correlators, we will introduce a new notation and rewrite I_k for deterministic points in a more convenient way. Similar to equation (5.5) in the fully symmetric case, here we denote by

$$x_i = \langle \mathcal{M}^{(i)}_0 \rangle, \qquad y_i = \langle \mathcal{M}^{(i)}_1 \rangle, \tag{5.80}$$

the pair of local deterministic expectation values for party i, which by definition take values ± 1. Hence, I_k can be written as

$$I_k = \sum_{i=1}^{n} f(\{x_i, y_i\}, \{x_{i+1}, y_{i+1}\}),$$

where the function f reads:

$$\begin{aligned} f(\{x_i, y_i\}, \{x_{i+1}, y_{i+1}\}) := & -\frac{k}{2}(x_i + x_{i+1}) - \frac{3k}{2}(y_i + y_{i+1}) \\ & - k\, x_i\, x_{i+1} + k\, x_i\, y_{i+1} \\ & + 2k\, y_i\, x_{i+1} + (2k+1)\, y_i\, y_{i+1}. \end{aligned}$$

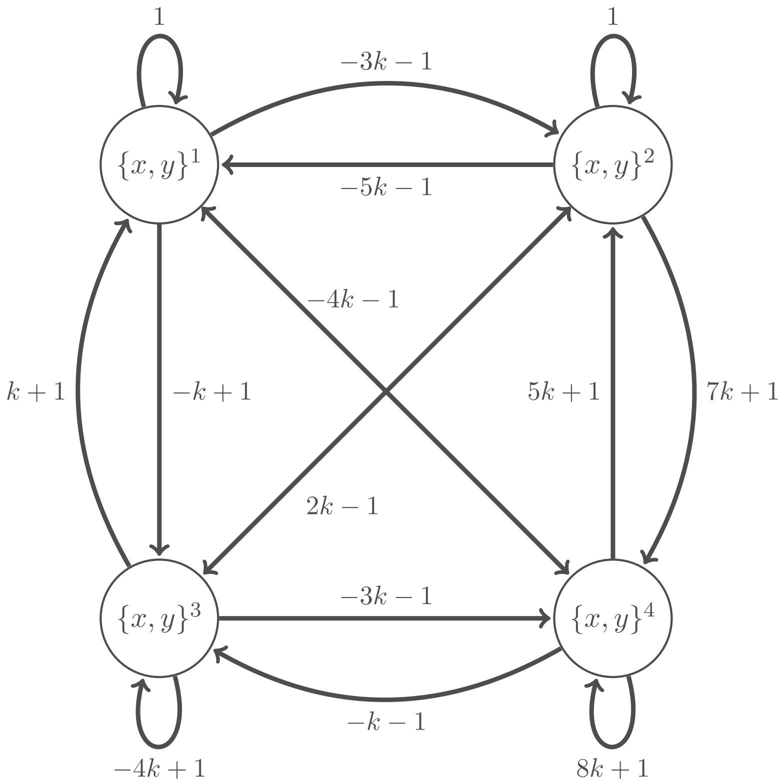
One may naïvely think that, in order to minimize I_k it suffices to take the minimum of f and multiply it by n. However, this is not the case, since the solution should satisfy the boundary conditions $x_{n+1} = x_1$ and $y_{n+1} = y_1$. In what follows, we present a method to compute such a constrained optimization.

First, note that each party i has four possible deterministic strategies $\{x_i, y_i\}$, given by $\{+, +\}$, $\{+, -\}$, $\{-, +\}$ and $\{-, -\}$. We will denote them by $\{x_i, y_i\}^j$ for $j = 1, 2, 3, 4$ respectively. Then, a global deterministic strategy for the n parties is given by a sequence $\{x_1, y_1\}^{j_1} \ldots \{x_n, y_n\}^{j_n}$. The idea now is to construct such a sequence that minimizes I_k.

The possible values of the function f in terms of the strategies $\{x_i, y_i\}$ and $\{x_{i+1}, y_{i+1}\}$ are

$\{x_i, y_i\} \backslash \{x_{i+1}, y_{i+1}\}$	$++$	$+-$	$-+$	$--$
$++$	1	$-3k-1$	$-k+1$	$-4k-1$
$+-$	$-5k-1$	1	$2k-1$	$7k+1$
$-+$	$k+1$	$2k-1$	$-4k+1$	$-3k-1$
$--$	$-4k-1$	$5k+1$	$-k-1$	$8k+1$

The important step now is to notice that every global deterministic strategy corresponds to a cycle of length n (see section 2.3) in the following graph:



Indeed, the first vertex in the cycle corresponds to the strategy chosen by party 1, the second vertex to the one of party 2, and so on. Since the cycle has length n, the boundary condition $\{x_{n+1}, y_{n+1}\} = \{x_1, y_1\}$ is satisfied. Note that every edge in the graph has a "weight" given by the value of the function f. In the cases where f is not symmetric, the corresponding edge was divided into two "directed edges", one weighted by $f(\{x,y\}^i, \{x,y\}^j)$ and the other by $f(\{x,y\}^j, \{x,y\}^i)$.

The next step is to notice that, for every deterministic global strategy, the value of I_k is given by the "weight" of the corresponding cycle. Indeed, for a global strategy $\{x_1, y_1\}^{j_1} \ldots \{x_n, y_n\}^{j_n}$ the cycle's weight is the sum of the weights of the edges of consecutive vertices, i.e. $\sum_{i=1}^n f(\{x_i, y_i\}^{j_i}, \{x_{i+1}, y_{i+1}\}^{j_{i+1}})$, which coincides with I_k. Hence, constructing a global deterministic strategy that minimizes I_k is equivalent to finding a cycle of length n with minimum weight in the previous graph. Such a cycle has weight $-8k^2 - 2k + 1$, which proves the classical bound of the Bell inequality (5.78) is $\beta_C = 8k^2 + 2k - 1$.

Quantum violations.
Now that we have a family of translationally invariant two-body correlators Bell inequality, with nearest-neighbours interactions, we move on to finding quantum violations. Similar to the fully symmetric case, we consider the one-qubit dichotomic observables at site i given by $\mathcal{M}_0^{(i)} = \sigma_z^{(i)}$ and $\mathcal{M}_1^{(i)} = cos(\theta_i)\sigma_x^{(i)} + sin(\theta_i)\sigma_z^{(i)}$. Numerical optimization over the angles θ_i suggest that for $n \geq 5$ the maximal quantum violation is achieved by $n-2$ consecutive parties applying $\theta_i = 0$. Hence, consider the state

$$\rho = |01010\cdots 10\rangle \langle 01010\cdots 10| \otimes \rho_{n-1,n},$$

i.e. a fixed pure state for the first $n-2$ parties, and an arbitrary mixed state for the last two parties. Assume further that the first $n-2$ parties have $\theta_i = 0$. Let $\mathcal{B}$ be the Bell operator associated to the Bell inequality (5.78) (see section 2.1.5). Finding a negative eigenvalue of $\mathcal{B}$ (i.e. a quantum violation) is equivalent to finding a negative eigenvalue of the reduced Bell operator $\mathcal{B}_{n-1,n} = \langle 01010\cdots 10| \mathcal{B} |01010\cdots 10\rangle$. Its Characteristic Polynomial is

$$Q_M(\lambda) = \lambda^4 - 4(2n-1)\lambda^3 - 2(n-1)(n(\cos\theta_n - 9) + \cos\theta_{n-1}(1-\cos\theta_n))\lambda^2 + \\ -8(n-1)^2(1+2n-\cos\theta_{n-1})\sin^2\frac{\theta_n}{2}\lambda - 4(n-1)^3 n \sin^2\theta_{n-1}\sin^2\theta_n,$$

and has the property that for $N > 1$ and $\forall\theta_{N-1} \neq n\pi$, $\forall\theta_N \neq n\pi$, the signs of its coefficients are $+\lambda^4, -\lambda^3, +\lambda^2, -\lambda^1, -\lambda^0$. Hence, by Descartes' rule of signs, the polynomial has a negative root, assuring quantum violation. In the limit of large number of parties, the minimum eigenvalue of $\mathcal{B}$ is $\lambda_{\min} \approx -0.309343(n-1)$, hence the quantum violation grows linearly with the number of parties.

6. Conclusions and Future Work

In this thesis we have addressed different problems on foundations of quantum mechanics. First, we have studied correlations in Bell-type experiments, focusing on the reasons for nature to forbid them to be as nonlocal as the No Signaling principle allows. We have proposed an intrinsically multipartite principle, called Local Orthogonality, to characterize the set of quantum correlations, and profit from its multipartite formulation to study both bipartite and multipartite nonlocality. Second, we have developed a framework for the study of contextuality, where Bell scenarios emerge as a special case of general scenarios. This approach allowed us to define a hierarchy of semidefinite programs (SDP) for probabilistic models on contextuality scenarios and study the constraints imposed by the Consistent Exclusivity principle (CE), as well as to take advantage of graph-theoretical tools to characterize sets of probabilistic models. Finally, we moved on to the more practical problem of how to detect and study nonlocality in many-body systems. We have proposed families of Bell inequalities that detect nonlocality for an arbitrary number of particles, with the advantageous property that they are formulated in terms of one and two-body correlators only. These Bell inequalities proved relevant for the study of nonlocality in interesting physical systems while also being experimentally friendly, in the sense that they could be evaluated in experiments where individual particles cannot be addressed.

The results achieved during the development of this thesis not only provided new insight into the topics of Nonlocality and Contextuality, but also arose several new questions. In what follows, I comment on some directions for future work.

Characterization of the Quantum Boundary.–

Throughout the study of the constraints that Local Orthogonality imposes on the space of correlations, we have worked not only with multipartite probability distributions, but also with independent copies of them. This approach, from which we defined our hierarchy of LO sets, is based on the existence of a large family of wirings for which any possible operational combination of a set of devices is equivalent to a tensor product model over networks when studying LO properties. However, it is not entirely clear whether any operation over a set of devices (rather than wirings) satisfies this property. Whether it is possible to define other hierarchies of LO sets, in terms of more general operations, that recover the set of quantum correlations, is still an open question.

Even though LO proves not to single out the set of quantum correlations in its current formulation, it captures some of its special properties, hence other natural ways to strengthen it could be explored. From the results on contextuality, one could argue a generalization similar to the Extended Consistent Exclusivity principle. However, the construction presented in this thesis does not directly apply to Nonlocality, since the auxiliary contextuality scenario may fail to be a Bell scenario. Hence, it remains as an open question how to find the corresponding extension to Nonlocality.

It is worth mentioning that, although information-based principles are useful for studying quantum correlations, it is not known whether they are sufficient to capture all the properties of quantum mechanics that make their set so particular. Hence, an open problem is whether a characterization merely from this approach is even possible. One recent approach that could shed light on the problem is that of characterizing quantum mechanics from principles that are influenced by quantum information theory (Har01; CDP11; DB11; MM11; Har11). However, this task studies the most-demanding question of deriving the whole structure of quantum theory, and not just its possible correlations.

Finally, interesting questions also arise from possible operational interpretations of LO. Indeed, all the known examples of non-trivial (i.e. defining a Bell inequality which is violated by some no-signaling correlations), tight (in the sense of defining a tight Bell inequality) information tasks with no quantum advantage, given in Refs. (ABB$^+$10; ASH$^+$11; AFK$^+$12; AAAB12), are examples of LO inequalities. In fact, these are the only known examples of non-trivial tasks with that property which also define tight Bell inequalities. It is an interesting working conjecture to prove that any non-trivial and tight information task with no quantum advantage defines an LO inequality. In particular, this would imply that any non-trivial tight Bell inequality in a bipartite scenario has quantum violations.

A framework for Contextuality.–

There are many directions for future work, since this framework is still under development. First, other possible definitions of contextuality scenarios could also be explored. Indeed, we have considered one way of extending the bipartite Foulis-Randall product into more factors, which preserves the No Signaling property of probabilistic models and recovers the traditional sets of correlations when restricted to Bell scenarios. However, it is not clear whether other generalizations with these properties also exist. This could give new insights into the characterization of the sets of probabilistic models.

Regarding the connection between Nonlocality and Contextuality in this framework, by further studying the CE principle we could revisit the previous question of whether information-based principles can capture all the characteristic properties of quantum correlations. The first level in our SDP hierarchy has a very similar formulation to that of quantum models, hence whether information-based principles are able to tell this subtle difference is still a non-trivial open problem.

Finally, during this thesis we have shown how to use graph theoretical tools to solve problems in quantum physics. Hence, another line of future work consists in doing the opposite: explore problems in combinatorics that would profit from known results on contextuality, in particular, how to relate some graph-theoretical invariants with examples of probabilistic models on already studied contextuality scenarios.

Detection of Nonlocality in many-body systems.–

As commented before, our Bell inequalities have the advantage of being formulated in terms of one and two-body correlators, which are easily accessible in a many-body setup. In addition, these are violated by the ground states of experimentally realizable models, such as Lipkin-Meshkov-Glick–like models with long-range interactions (for ionic spin 1/2 and spin 1 realizations see (PC04; GJDKL13) and for cold atoms in nanophotonic waveguides see (CCK13)), or degenerated ground states of the ferromagnetic Heisenberg model (Sac11). In addition, when all observers measure the same pair of observables, the Bell violation can be estimated via collective measurements of total spin operators and their projection into some direction, which enables the study on setups where individual particles cannot be addressed. Nowadays, it is possible to measure these quantities in atomic systems with current experimental technologies, such as spin polarization spectroscopy (HSP10; ERIR$^+$08). Hence, a future line of work would be to study particular experimental setups and the optimal Bell inequalities for testing the systems. This could provide a broad set of possibilities to study both theoretically and experimentally the nonlocal nature of many-body entangled states.

In some experimental setups, however, the exact number of particles in the system is not known with certainty. Our Bell inequalities, in turn, are well defined for a fixed n, since the classical bound (and in some cases also the coefficients) is always a function of the number of particles. An interesting line of work hence is to study whether there exist Bell inequalities formulated from two-body correlators which are valid for any n, i.e. where both the coefficients and the classical bound do not scale with n. This would allow the study of the nonlocal properties of systems where the number of constituents is subjected to experimental errors.

Finally, our fully-symmetric two-body Bell inequalities prove to detect nonlocality from correlations with local bipartite marginals. Even though in some cases the corresponding quantum states are genuinely mulipartite entangled, it is still an open question whether the correlations are truly multipartite nonlocal in the sense of (WBA+12). In order to study if this is the case, one would need to include time-ordered-bilocal (TOBL) models in the study of these inequalities and their corresponding TOBL bounds. More generally, it would be also interesting to study the power of two-body correlators for witnessing multipartite nonlocality and to what extent such a task is possible with the minimal resources available. In addition, one could argue the inclusion of higher-order correlators into new classes of Bell inequalities, and study the their capability for detecting nonlocal and truly multipartite nonlocal correlations, as well as the trade-off between theoretical improvements and experimental disadvantages.

A. Background on Graph Theory

This section reviews standard material on the invariants of graphs which are of relevance to the main text, first for unweighted and then for weighted graphs. In this thesis, a **graph** is an undirected simple graph without isolated vertices. When G is a graph, we denote its set of vertices by $V(G)$. For $u, v \in V(G)$, we write $u \sim v$ whenever u and v share an edge (are **adjacent**) in G.

There are many ways to take products of graphs (IK00). For us, the relevant ones are the following:

Definition A.1. *Let G_1 and G_2 be graphs. Their* ***strong product*** *is the graph $G_1 \boxtimes G_2$ with*

$$V(G_1 \boxtimes G_2) := V(G_1) \times V(G_2)$$

and $(u_1, u_2) \sim (v_1, v_2)$ whenever

$$(u_1 \sim v_1 \wedge u_2 \sim v_2) \vee (u_1 \sim v_1 \wedge u_2 = v_2) \vee (u_1 = v_1 \wedge u_2 \sim v_2) .$$

For $n \in \mathbb{N}$, we write $G^{\boxtimes n}$ for the n-fold strong product of G with itself.

Definition A.2. *Let G_1 and G_2 be graphs. Their* ***conormal product*** *is the graph $G_1 \cdot G_2$ with*

$$V(G_1 \cdot G_2) := V(G_1) \times V(G_2)$$

and $(u_1, u_2) \sim (v_1, v_2)$ whenever

$$(u_1 \sim v_1) \vee (u_2 \sim v_2) .$$

For $n \in \mathbb{N}$, we write $G^{\cdot n}$ for the n-fold conormal product of G with itself.

A.1. Relevant invariants of unweighted graphs

Since later in this chapter I consider graphs equipped with vertex weights, I also use the term "unweighted graph" when working with plain graphs in order to emphasize the distinction.

Recall that an **independent set** in a graph G is a subset $I \subseteq V(G)$ such that no two vertices in I share an edge. I is an independent set in G if and only if it is a **clique** in the complement graph $\overline{G}$. An independent set I is **maximal** if there is no other independent set $I' \subseteq V(G)$ with $I \subsetneq I'$. The **independence number** $\alpha(G)$ is the size of the largest any independent set in G.

Lemma A.3. *Let $I_1 \subseteq G_1$ and $I_2 \subseteq G_2$ be maximal independent sets. Then $I_1 \times I_2 \subseteq G_1 \boxtimes G_2$ is also a maximal independent set.*

Proof. The definition of adjacency in $G_1 \boxtimes G_2$ implies immediately that $I_1 \times I_2$ is also an independent set in $G_1 \boxtimes G_2$.

We now show maximality of $I = I_1 \boxtimes I_2$. For any $v = (v_1, v_2) \in V(G_1 \boxtimes G_2) \setminus I$, the following cases are possible:

1. Case $v_1 \notin I_1$ and $v_2 \notin I_2$: by maximality of I_1 and I_2, there are $u_1 \in I_1$ with $u_1 \sim v_1$ and $u_2 \in I_2$ with $u_2 \sim v_2$. Hence $(u_1, u_2) \sim (v_1, v_2)$.
2. Case $v_1 \notin I_1$ and $v_2 \in I_2$: by maximality of I_1, there is $u_1 \in I_1$ with $u_1 \sim v_1$. Hence $(u_1, v_2) \in I$ and $(u_1, v_2) \sim (v_1, v_2)$.
3. Case $v_1 \in I_1$ and $v_2 \notin I_2$: Similar to the previous case.

In either case, the conclusion is that v is adjacent to some vertex in I, and hence I is a maximal independent set. □

Corollary A.4.

$$\alpha(G_1 \boxtimes G_2) \geq \alpha(G_1)\alpha(G_2)$$

In particular, this implies

$$\alpha(G^{\boxtimes(n+m)}) \geq \alpha(G^{\boxtimes n})\alpha(G^{\boxtimes m}) \qquad \forall m, n \in \mathbb{N}. \tag{A.1}$$

Remark A.5. Despite this inequality, the sequence $\left(\sqrt[n]{\alpha(G^{\boxtimes n})}\right)_{n\in\mathbb{N}}$ is not monotonically increasing in general; this happens, for example, for the pentagon graph (or 5-cycle) ⬠, for which

$$\alpha(⬠) = 2, \qquad \alpha(⬠^{\boxtimes 2}) = 5, \qquad \alpha(⬠^{\boxtimes 3}) = 10.$$

See (AL06) for more results on the behaviour of $\left(\sqrt[n]{\alpha(G^{\boxtimes n})}\right)_{n\in\mathbb{N}}$.

In combination with Fekete's Lemma (Fek23), (A.1) guarantees the existence of the following limit:

Definition A.6 (Shannon capacity)**.** *The **(unweighted) Shannon capacity** $\Theta(G)$ is*

$$\Theta(G) := \lim_{n\to\infty} \sqrt[n]{\alpha(G^{\boxtimes n})}. \tag{A.2}$$

Intuitively, $\Theta(G)$ is an asymptotic version of the independence number $\alpha(G)$. This number can be interpreted in terms of information theory as follows. Consider the problem of two parties, Alice and Bob, where Alice wants to send a message to Bob through a noisy channel. Here, some symbols in the alphabet that Alice uses may be confused when reaching Bob's side. The problem is which is the size of the largest message she can successfully send to Bob with no error. To answer this question, the normal approach is to define the **confusability graph** G of the channel, where the vertices $V(G)$ are given by the letters of Alice's alphabet, and $u \sim v$ if and only if u and v have non-trivial probability to produce the same channel output. Then this channel can asymptotically transfer $\log_2 \Theta(G)$ bits of perfect information per channel use. This is the context in which Θ was originally introduced by Shannon (Sha56). The use of the logarithm here differs from the standard information-theoretic definitions of capacities, which usually already include it in their definition.

Not much is known about the values of Θ for particular graphs, not even $\Theta(C_7)$, where C_7 is the 7-cycle (CGR03).

Definition A.7. *Let G_1 and G_2 be graphs. Their disjoint union is the graph $G_1 + G_2$ with*

$$V(G_1 + G_2) = V(G_1) \bigcup V(G_2)$$

and

$$E(G_1 + G_2) = E(G_1) \bigcup E(G_2)$$

The Shannon capacity of the product and disjoint union of graphs has the following properties:

Lemma A.8 ((Sha56))**.** *1.*

$$\Theta(G_1 + G_2) \geq \Theta(G_1) + \Theta(G_2). \tag{A.3}$$

2.

$$\Theta(G_1 \boxtimes G_2) \geq \Theta(G_1)\Theta(G_2). \tag{A.4}$$

Finding examples in which these inequalities are not tight is surprisingly difficult. The following results are due to Haemers and Alon.

Theorem A.9 ((Hae79; Alo98))**.** *There exist graphs G_1 and G_2 such that*

1. $\Theta(G_1 \boxtimes G_2) > \Theta(G_1)\Theta(G_2)$.
2. $\Theta(G_1 + G_2) > \Theta(G_1) + \Theta(G_2)$,

Of particular relevance for our considerations in the main text are graphs whose independence number coincides with their Shannon capacity:

Definition A.10. *A graph G is **single-shot** if $\alpha(G) = \Theta(G)$.*

G is single-shot precisely when the sequence $\left(\sqrt[n]{\alpha(G^{\boxtimes n})}\right)_{n\in\mathbb{N}}$ is constant. Our terminology is motivated by the information-theoretic interpretation alluded to above: if a communication channel has a confusability graph which is single-shot, then there exists a zero-error code for this channel which operates on the single-shot level.

Due to standard results (Knu94), every perfect graph is single-shot. The Petersen graph is not perfect, but nevertheless single-shot since its Lovász number (see below) coincides with its independence number (Knu94, p. 31).

Another relevant graph invariant is the so called **Lovász number** (Lov79). It has many equivalent definitions (Knu94), and I use the following:

Definition A.11 (Lovász number (Lov79))**.**
1. *An **orthonormal labelling** of G is an assignment $v \mapsto |\psi_v\rangle$ of a unit vector $|\psi_v\rangle \in \mathbb{R}^{|V(G)|}$ to every $v \in V(G)$ such that $u \not\sim v$ and $u \neq v$ implies $|\psi_u\rangle \perp |\psi_v\rangle$.*

2. *The **Lovász number** $\vartheta(G)$ is*

$$\vartheta(G) := \min_{|\Psi\rangle,\,|\psi_v\rangle} \max_{v\in V} \frac{1}{|\langle\Psi|\psi_v\rangle|^2}$$

where $|\Psi\rangle \in \mathbb{R}^{|V(G)|}$ ranges over all unit vectors and $(|\psi_v\rangle)_{v\in V(G)}$ over all orthonormal labellings.

Multiplicativity of ϑ is one of its many useful properties:

Proposition A.12 ((Lov79))**.**

$$\vartheta(G_1 \boxtimes G_2) = \vartheta(G_1)\vartheta(G_2).$$

Finally, another graph invariant is the **fractional packing number**, defined as follows:

Definition A.13. *The **fractional packing number** $\alpha^*(G)$ is*

$$\alpha^*(G) := \max_q \sum_v q_v$$

where $q : V(G) \to [0,1]$ ranges over all vertex weightings satisfying $\sum_{v\in C} q_v \leq 1$ for all cliques $C \subseteq V(G)$.

The fractional packing number can be regarded as the linear relaxation of the independence number. For this reason, it is sometimes also called **fractional independence number**.

The previously mentioned graph invariants relate to each other in the following way.

Proposition A.14 ((Lov79))**.**

$$\alpha(G) \le \Theta(G) \le \vartheta(G) \le \alpha^*(G).$$

In general, none of these inequalities is an equality. This is most difficult to see for $\Theta(G) \le \vartheta(G)$, for which it was shown by Haemers (Hae79) after having been posed as an open problem by Lovász (Lov79).

A.2. Relevant invariants of weighted graphs

This section generalizes the definitions to graphs equipped with vertex weights, i.e. to graphs G equipped with a **weight function** $p : V(G) \to \mathbb{R}_+$.

Given two graphs G_1 and G_2, together with their weight functions $p_1 : V(G_1) \to \mathbb{R}_+$ and $p_2 : V(G_2) \to \mathbb{R}_+$, the strong product of the graphs has an associated weight function in product form:

$$p_1 \otimes p_2 : V(G_1 \boxtimes G_2) \to \mathbb{R}_+, \quad (v_1, v_2) \mapsto p_1(v_1)p_2(v_2).$$

In this way, $p^{\otimes n}$ is a weight function on $G^{\boxtimes n}$. The same holds for conormal product of graphs. Similarly, there is an obvious weight function $p_1 + p_2$ defined on the disjoint union $G_1 + G_2$. In the particular case when p_1 and p_2 are defined on the same graph, we use the same notation $p_1 + p_2$ for the pointwise sum; despite this ambiguous notation, the meaning will always be clear from the context.

The relevant graph invariants may then be generalized to the weighted case:

Definition A.15. *Let G be a graph equipped with vertex weights p.*

1. *The **weighted independence number** $\alpha(G, p)$ is the largest total weight of an independent set in G.*

2. *The **weighted Lovász number** $\vartheta(G, p)$ is*

$$\vartheta(G, p) := \min_{|\Psi\rangle,\, |\psi_v\rangle} \max_{v \in V} \frac{p(v)}{|\langle \Psi | \psi_v \rangle|^2} \tag{A.5}$$

where $|\Psi\rangle \in \mathbb{R}^{|V(G)|}$ ranges over all unit vectors and $(|\psi_v\rangle)_{v \in V(G)}$ over all orthonormal labellings.

3. *The **weighted Shannon capacity** $\Theta(G,p)$ is*

$$\Theta(G,p) = \lim_{n\to\infty} \sqrt[n]{\alpha(G^{\boxtimes n}, p^{\otimes n})}. \tag{A.6}$$

4. *The **weighted fractional packing number** $\alpha^*(G,p)$ is*

$$\alpha^*(G,p) := \max_q \sum_{v\in V} p(v)q(v).$$

where $q : V(G) \to \mathbb{R}_+$ ranges over all vertex weights satisfying

$$\sum_{v\in C} q(v) \le 1$$

for all cliques $C \subseteq V(G)$.

The fraction in (A.5) uses the convention $\frac{0}{0} = 0$. Note that of these quantities specialize to their unweighted counterparts by choosing unit weights $p = \mathbb{1}$.

As mentioned in the previous section, (Knu94) presents several equivalent definitions of $\vartheta(G,p)$. One of relevant importance in this thesis is the forth characterization of ϑ given in Section 10 of (Knu94):

Definition A.16. *The **weighted Lovász number** $\vartheta(G,p)$ is*

$$\vartheta(G,p) := \max_{|\Psi\rangle, |\phi_v\rangle} \sum_{v\in V} p(v) |\langle\Psi|\phi_v\rangle|^2 \tag{A.7}$$

where $|\Psi\rangle \in \mathbb{R}^{|V(G)|}$ ranges over all unit vectors and $(|\phi_v\rangle)_{v\in V(G)}$ over all orthonormal labellings of $\overline{G}$, i.e. the complement of G.

The weighted fractional packing number can also be characterized by the dual of the linear program that appears on its definition, which leads to the equivalent formulation:

Proposition A.17. *Let $\mathrm{Cl}(G)$ denote the set of all cliques on G.*

$$\alpha^*(G,p) = \min_x \sum_{C\in \mathrm{Cl}(G)} x(C) \tag{A.8}$$

where x ranges over all functions $x : \mathrm{Cl}(G) \to \mathbb{R}_+$ with $p(v) \le \sum_{C\ni v} x(C)\, \forall v$.

Similar to the unweighted case, the Shannon capacity has the following properties:

Lemma A.18. *1.*

$$\Theta(G_1 + G_2, p_1 + p_2) \geq \Theta(G_1, p_1) + \Theta(G_2, p_2). \tag{A.9}$$

2.

$$\Theta(G_1 \boxtimes G_2, p_1 \otimes p_2) \geq \Theta(G_1, p_1)\Theta(G_2, p_2). \tag{A.10}$$

Proof. As in the unweighted case (Sha56). □

Since these inequalities are not tight in general in the unweighted case (Hae79; Alo98), neither can they be tight in the weighted case. One might expect simpler counterexamples to exist in the weighted case, but we have still not been successful in finding any.

When p_1, p_2 are weight functions on the same graph G, superadditivity no longer holds for trivial reasons: e.g. for $G = K_2$, the graph on two adjacent vertices $\{u, v\}$ with $p_1 = \mathbb{1}_u$ and $p_2 = \mathbb{1}_v$, we have

$$1 = \Theta(G, p_1 + p_2) < \Theta(G, p_1) + \Theta(G, p_2) = 2.$$

A.3. Relation between invariants of unweighted and weighted graphs

Many statements about the invariants of weighted graphs can be reduced to statements about their unweighted counterparts using a technique we call **blow-up**. Applying this technique requires the vertex weights to be rational. Therefore, we begin by proving a continuity result which allows us to reduce many problems to the case of rational weights.

Lemma A.19. *Let (G, p) be a weighted graph and $\overline{K}_m$ the empty graph on m vertices with weights q. Then,*

$$X(G + \overline{K}_m, p + q) = X(G, p) + \sum_{v \in V(\overline{K}_m)} q(v) \tag{A.11}$$

for all four invariants $X \in \{\alpha, \Theta, \vartheta, \alpha^*\}$.

Proof. This is trivial for $X = \alpha$. For $X = \vartheta$, it is a special case of (Knu94, eq. (18.2)). For $X = \alpha^*$, it follows from an application of Proposition A.17. It remains to treat the case $X = \Theta$.

Since $\Theta(\overline{K}_m, q) = \sum_v q_v$, the inequality "$\geq$" is an instance of superadditivity (A.9) of Θ. To also show "$\leq$", we choose any independent set I in $(G + \overline{K}_m)^{\boxtimes n}$ and partition it into a disjoint union

$$I = \bigcup_{\vec{s} \in \{0,1\}^n} I_{\vec{s}}$$

where each $I_{\vec{s}}$ contains only vertices $(v_1, \ldots, v_n)$ with $v_i \in V(G)$ if $s_i = 0$ and $v_i \in V(\overline{K}_m)$ if $s_i = 1$. Then upon dropping all components i with $s_i = 1$, such an $I_{\vec{s}}$ becomes an independent set in some $G^{\boxtimes k}$. In this way, we get the estimate

$$\begin{aligned}
\alpha\left((G + \overline{K}_m)^{\boxtimes n}, (p+q)^{\otimes n}\right) &\leq \sum_{k=0}^{n} \binom{n}{k} \alpha(G^{\boxtimes k}, p^{\otimes k}) \left(\sum_i q_i\right)^{n-k} \\
&\leq \sum_{k=0}^{n} \binom{n}{k} \Theta(G,p)^k \left(\sum_i q_i\right)^{n-k} \\
&= \left(\Theta(G,p) + \sum_i q_i\right)^n,
\end{aligned}$$

which implies the desired inequality upon taking the n-th root and then $n \to \infty$. □

Another interesting question is what happens to the graph invariants when the weight of a single vertex is increased. In this case, we see that the following inequalities hold.

Lemma A.20. *Let (G, p) be a weighted graph, $v \in G$ a vertex, $q \in \mathbb{R}_+$ and $X \in \{\alpha, \Theta, \vartheta, \alpha^*\}$. Then*

$$X(G,p) \leq X(G, p + q\mathbb{1}_v) \leq X(G,p) + q. \tag{A.12}$$

Proof. The first inequality is clear since $X(G,p)$ is a non-decreasing function of p.

Since adding additional edges cannot increase the value of X and two vertices with exactly the same neighbours can be identified to one vertex by adding the weights (for $X = \vartheta$, see (Knu94, Lemma 16)), we have $X(G, p + q\mathbb{1}_v) \leq X(G + \overline{K}_1, p + q)$. Now the second inequality follows from the previous Lemma with $m = 1$. □

This Lemma directly gives the desired continuity result:

Corollary A.21. *For any graph G and any $X \in \{\alpha, \Theta, \vartheta, \alpha^*\}$, the function $p \mapsto X(G,p)$ is continuous.*

We can now introduce the blow-up technique which can be used to translate problems from the weighted case to the unweighted setting.

Definition A.22. *Let (G,p) be a weighted graph with $p(v) \in \mathbb{N}\,\forall v$. Then the **blow-up** $\mathrm{Blup}(G,p)$ is the unweighted graph with vertex set*

$$\big\{(v,k)\ :\ v \in G, k \in \{1,\dots,p(v)\}\big\},$$

where we take (v,k) and (v',k') to be adjacent if and only if $v \sim v'$ in G.

Intuitively speaking, $\mathrm{Blup}(G,p)$ is constructed by replacing every vertex v in G by $p(v)$ many non-adjacent vertices. In particular, if $p(v) = 0$, the vertex v simply gets removed from the graph. Blow-ups have also been considered in (Knu94, Sec. 16), although not under that name.

The Blup relates to the product, union and invariants like follows.

Lemma A.23. *For vertex weights in $\mathbb{N}$,*

1. $\mathrm{Blup}(G_1 + G_2, p_1 + p_2) = \mathrm{Blup}(G_1, p_1) + \mathrm{Blup}(G_2, p_2)$.
2. $\mathrm{Blup}(G_1 \boxtimes G_2, p_1 \otimes p_2) = \mathrm{Blup}(G_1, p_1) \boxtimes \mathrm{Blup}(G_2, p_2)$*;*
3. $X(\mathrm{Blup}(G,p)) = X(G,p)$ *for every* $X \in \{\alpha, \Theta, \vartheta, \alpha^*\}$*.*

Hence, by using the properties of the blow-up technique, the following statements have clear proofs:

Corollary A.24.

$$\alpha(G,p) \le \Theta(G,p) \le \vartheta(G,p) \le \alpha^*(G,p).$$

Proof. Combine Lemma A.23 with Proposition A.14. □

Corollary A.25 ((Knu94, (20.5)))**.**

$$\vartheta(G_1 \boxtimes G_2, p_1 \otimes p_2) = \vartheta(G_1, p_1)\vartheta(G_2, p_2)$$

Proof. Combine Lemma A.23 with Proposition A.12. □

Corollary A.26.

$$\alpha(G_1 \boxtimes G_2, p_1 \otimes p_2) \ge \alpha(G_1, p_1)\alpha(G_2, p_2)$$

Proof. Combine Lemma A.23 with Lemma A.4. □

B. Bipartite scenarios: $\mathcal{LO}^1 \equiv \mathcal{NS}$

In this appendix, I will show that $\mathcal{LO}^1 \equiv \mathcal{NS}$ for bipartite scenarios. Although this was already noticed in (CSW10), here I give a slightly different proof which emphasizes the connection to LO.

In what follows, measurements and results by the two parties are labelled by x, y and a, b, so that correlations read $P(ab|xy)$. The no signaling conditions are:

$$\sum_{b=0}^{d-1} P(ab|xy) = \sum_{b=0}^{d-1} P(ab|xy')\,, \quad \sum_{a=0}^{d-1} P(ab|xy) = \sum_{a=0}^{d-1} P(ab|x'y), \tag{B.1}$$

Let us start by characterizing the possible sets of locally orthogonal events. Recall that two events are locally orthogonal if for at least one party the settings are identical but the outcomes are different. Consider a set of locally orthogonal events which contains $(ab|xy)$ and $(a'b'|x'y)$ with $x' \neq x$. Then, this set cannot contain any event of the form $(a''b''|x''y')$ with $y' \neq y$, because it could not be locally orthogonal to both other events. From this intuition, we find that the sets of pairwise orthogonal events are either

$$\{(ab|x\omega_A(a)) : a, b = 0, \ldots, d-1\}$$

for fixed x, or

$$\{(ab|\omega_B(b)y) : a, b = 0, \ldots, d-1\}$$

for fixed y, with $\omega_W : \{0, \ldots, d-1\} \longrightarrow \{0, \ldots, m-1\}$ $(W = A, B)$ being some map.

We start by showing that sets of the first kind have the desired property; the proof for sets of the second kind is analogous. Take two such events $(ab|x\omega_A(a)) \neq (a'b'|x\omega_A(a'))$. Then either $a \neq a'$ and orthogonality holds on Alice's side, or $b \neq b'$ and orthogonality follows from Bob. This proves that we have a set of exclusive events. To see that the set is maximal, consider an arbitrary event $(\widetilde{ab}|\widetilde{xy})$. If $\widetilde{x} = x$ and $\widetilde{y} = \omega_A(\widetilde{a})$, then this event is already in the set. Otherwise, LO fails between $(\widetilde{ab}|\widetilde{xy})$ and $(\widetilde{ab}|x\omega_A(\widetilde{a}))$. Hence it is impossible to add any event to the set, i.e. the set is maximal.

Now we prove that every maximal LO^1 set is of one of these two forms. It is enough to show that every LO^1 set is contained in a set of this form. As noted above, for events in an LO set, one of the parties is restricted to using a single input. Hence, without loss of generality, we can take x to be fixed. Since every two orthogonal events differ on at least one outcome, there exists a function $\omega(a,b)$ such that every element in the set is of the form $(ab|x\omega(a,b))$. We complete the proof by showing that $\omega(a,b)$ does not depend on b. The existence of a and b, b' with $\omega(a,b) \neq \omega(a,b')$ would imply that $(ab|x\omega(a,b))$ and $(ab'|x\omega(a,b'))$ are not orthogonal, contradicting the assumption.

We now prove that $\mathcal{LO}^1 = \mathcal{NS}$, in the present bipartite setting.

$\mathcal{LO}^1 \subseteq \mathcal{NS}$: all optimal LO^1 inequalities are of the form

$$\sum_{a,b=0}^{d-1} P(ab|x\omega_A(a)) \leq 1,$$

modulo exchanging the parties. We fix any a_0, y and y' and consider the function $\omega_A(a) = y$ if $a = a_0$ and $\omega_A(a) = y'$ if $a \neq a_0$. The LO inequality yields

$$\sum_{a\neq a_0,b=0}^{d-1} P(ab|xy') + \sum_{b=0}^{d-1} P(a_0b|xy) \leq 1.$$

Together with the normalization equation $\sum_{a,b=0}^{d-1} P(ab|xy') = 1$, this implies

$$\sum_{b=0}^{d-1} P(a_0b|xy) \leq \sum_{b=0}^{d-1} P(a_0b|xy').$$

Since the same inequality can be derived with y and y' interchanged, we find that it actually needs to be an equality, which is (B.1).

$\mathcal{NS} \subseteq \mathcal{LO}^1$: start from the normalization condition $\sum_{a,b=0}^{d-1} P(ab|xy) = 1$. Using the no-signaling equations (B.1), we can transform it into an equality of the form $\sum_{a,b=0}^{d-1} P(ab|x\omega_A(y)) = 1$ for any given x and ω_A. It follows that $\mathcal{NS} \subseteq \mathcal{LO}^1$, and thus $\mathcal{LO}^1 = \mathcal{NS}$ in the bipartite scenario.

C. Families of LO inequalities

In this appendix, I explain how to sort LO inequalities into classes and present the list all LO inequalities for the $(3, 2, 2)$, $(3, 2, 3)$ and $(4, 2, 2)$ scenarios.

C.1. Defining and computing equivalence classes

Starting from the orthogonality graph for a given scenario (n, m, d), as explained in section 3.4, we can generate a list of all the corresponding LO inequalities by employing standard methods from graph theory. For sufficiently small scenarios, this computation is feasible with the existing software packages for clique enumeration (Uno05; NO10), and here we describe the results of our computations along these lines. We write $\mathbf{a} := a_1 \dots a_n$ and $\mathbf{x} := x_1 \dots x_n$ as shorthands.

Each LO inequality is of the form

$$\sum_{\mathbf{a},\mathbf{x}} c_{\mathbf{a},\mathbf{x}} P(\mathbf{a}|\mathbf{x}) \leq 1, \tag{C.1}$$

with $c_{\mathbf{a},\mathbf{x}} \in \{0, 1\}$, that is, each LO inequality corresponds simply to a list of the terms which are present, i.e. the terms for which $c_{\mathbf{a},\mathbf{x}} = 1$. However, for the purpose of understanding the structural aspects of the LO principle, many of these inequalities can be considered equivalent. More concretely, if two inequalities with respective coefficients $c_{\mathbf{a},\mathbf{x}}$ and $c'_{\mathbf{a},\mathbf{x}}$ can be transformed into each other by relabelling the parties or the measurement choices and outcomes, or by making use of the normalization and no-signaling equations, or by combining such transformations, then they really represent different instances of the same basic inequality. So we consider two LO inequalities to be equivalent if one can be transformed into the other under a combination of the following transformations:

(i) *Permutation of parties.* For some permutation σ of n objects, the $c'_{\mathbf{a},\mathbf{x}}$ of the second inequality can be obtained from the $c_{\mathbf{a},\mathbf{x}}$ of the first inequality as $c'_{\mathbf{a},\mathbf{x}} = c_{\mathbf{a}',\mathbf{x}'}$, where $a'_i = a_{\sigma(i)}$ and $x'_i = x_{\sigma(i)}$.

(ii) *Relabelling of measurement choices.* For some set of permutations $\sigma_1, \dots, \sigma_n$ of m objects, the coefficients of the second inequality can be obtained as $c'_{\mathbf{a},\mathbf{x}} = c_{\mathbf{a},\mathbf{x}'}$, where $x'_i = \sigma_i(x_i)$.

(iii) *Relabelling of outcomes.* For some set of permutations $\sigma_{1,1}, \ldots, \sigma_{n,m}$ of d objects indexed by parties and measurement choices, the coefficients of the second inequality can be obtained as $c'_{\mathbf{a},\mathbf{x}} = c_{\mathbf{a}',\mathbf{x}}$, where $a'_i = \sigma_{i,x_i}(a_i)$.

(iv) *No-signaling and normalization.* The second inequality can be obtained from the first one by adding a linear combination of the no-signaling equations (2.3) and the normalization condition $\sum_{\mathbf{a}} P(\mathbf{a}|\mathbf{x}) = 1$.

Since the clique enumeration software enumerated *all* cliques in the respective orthogonality graphs, the corresponding sets of LO inequalities had large redundancy in the sense that many inequalities were equivalent to each other under these symmetry transformations. In the following, we describe how we eliminated this redundancy by computing one unique representative of each symmetry class.

First, we considered each inequality for n parties with r terms as an $(r \times 2n)$-matrix. Each row in the matrix corresponds to a term of the inequality by listing the corresponding outcomes and measurement choices $a_1 \ldots a_n x_1 \ldots x_n$. Two such matrices which differ only by the order of their rows trivially represent the same inequality, and hence we will choose the lexicographically smallest ordering as a *normal form* with respect to this equivalence: two matrices represent the same inequality if and only if they have the same normal form. In all subsequent steps, an inequality was always represented as a matrix whose rows are lexicographically ordered. More generally, we always reduced the elimination of equivalences to the computation of normal forms.

Next we eliminated the equivalences under transformations of types (i)–(iii), again by computing a normal form with respect to these transformations for each inequality. For each party, the measurement choices were ordered according to their multiplicity, i.e. according to the number of terms in which they appear. They were then relabelled such that the measurement choice which occurred most often was assigned the lowest label, and so on for the following measurement choices. Similarly, for each party and each measurement choice, the outputs were relabelled according to their multiplicity. Whenever multiple measurement choices or outcomes occurred with the same multiplicity, all possible relabellings were applied, resulting in a list of equivalent inequalities. Next, all possible permutations of the parties were applied, resulting in an even longer list of inequalities. Then again, for each matrix representing an inequality in the list, the rows were ordered lexicographically—corresponding to a permutation of the terms in the inequality—and then the matrices themselves were ordered lexicographically. The first matrix in this reordered list was taken to be the normal form representating the whole equivalence class. The relabellings of measurement choices and outcomes defined in this way are invariant under permutation of parties and terms, since such permutations cannot change the multiplicity of a given measurement choice or outcome. This ensures that the representative is unique. This defines a normal form with respect to the equivalences (i)–(iii), as well as an algorithm to compute it. In this way, we eliminated these equivalences using a piece of MATHEMATICA code. This produced a smaller list of inequalities given in the form (C.1).

Finally, we had to eliminate equivalences under transformations which also include those of type (iv). To this end, a normal form for general Bell inequalities and a method for computing this normal form had previously been described in (BGP10). This normal form expresses the inequalities in terms of generalized correlators (see Appendix A of (BGP10) and also (PBS11)). A MATLAB package for computing this normal form has been developed by Bancal and was kindly provided to us. Although this software is capable of eliminating equivalences of all types (i)–(iv), our strategy of first eliminating (i)–(iii) has turned out to be advantageous: in contrast to general Bell inequalities, our LO inequalities are very sparse and all of their coefficients are 0 or 1. This is a feature that we have exploited in our MATHEMATICA code, which does not store the inequalities as large arrays of coefficients, but as $(r \times 2n)$-matrices as explained above, which led to a significant speed-up. We applied Bancal's MATLAB software to the list of inequalities obtained in the previous step, which resulted in further elimination of equivalences, this time finally under all of (i)–(iv). In the end, the representative of each equivalence class in its matrix representation was taken to be the first inequality of the class in the sorted output from MATHEMATICA.

C.2. All LO inequalities for the $(3,2,2)$, $(3,2,3)$ and $(4,2,2)$ scenarios

Using the method of the previous section, we were able to completely classify all LO inequalities for the scenarios (3,2,3) and (4,2,2). In the tables below we list the normal form representative of each of the non-trivial equivalence classes. Here, an inequality is non-trivial if it can be violated by some no-signaling box. All the other inequalities are trivial, i.e. represent the normalization of probabilities or the no-signaling condition, and thereby are equivalent under (iv) to the tautological inequality $0 \leq 0$.

The (3,2,2) scenario. The GYNI inequality (2.11) represents the only class in this scenario.

The (3,2,3) scenario. In what follows, I list the 4 equivalence classes found for the scenario $(3,2,3)$. As explained in the main text these four inequalities correspond, respectively, to maximal cliques of 12, 13, 14, and 15 vertices.

$$\begin{aligned}&\big\{(000|000),(001|000),(002|110),(010|000),\\&(011|000),(012|110),(102|110),(112|110),\\&(120|011),(220|011),(221|101),(222|101)\big\}\end{aligned}\tag{C.2}$$

$$\begin{aligned}&\big\{(000|001),(001|001),(002|111),(010|001),\\&(011|001),(110|010),(120|010),(121|100),\\&(122|100),(210|010),(220|010),(221|100),(222|100)\big\},\end{aligned}\tag{C.3}$$

$$\begin{aligned}&\big\{(000|000),(001|000),(002|110),(010|000),\\&(011|000),(012|110),(100|000),(101|000),(110|000),\\&(111|000),(120|101),(220|101),(221|011),(222|011)\big\},\end{aligned}\tag{C.4}$$

$$\begin{aligned}&\big\{(000|000),(001|000),(002|110),(010|000),\\&(011|000),(012|110),(100|000),(101|000),(102|110),\\&(110|000),(111|000),(112|110),(220|011),(221|011),(222|101)\big\}.\end{aligned}\tag{C.5}$$

The (4,2,2) scenario. In what follows, I list all 35 equivalence classes found for the scenario $(4,2,2)$. First, 30 inequivalent inequalities with 8 terms each, second two with 9 terms each, then two with 10 terms each, and finally one inequality with 12 terms.

$$\begin{aligned}&\big\{(0000|0000),(0001|0000),(0010|1100),(0101|1010),\\&(1010|1101),(1100|0110),(1110|0111),(1111|1011)\big\}\end{aligned}\tag{C.6}$$

$$\big\{(0000|0000),(0001|0010),(0010|1100),(0011|1110),\\(1100|0101),(1101|0111),(1110|1001),(1111|1011)\big\} \tag{C.7}$$

$$\big\{(0000|0000),(0001|0010),(0010|1100),(0011|1110),\\(1100|0101),(1101|1011),(1110|1001),(1111|0111)\big\} \tag{C.8}$$

$$\big\{(0000|0000),(0001|0010),(0010|1100),(0100|1001),\\(1001|0101),(1100|1011),(1101|0111),(1111|1110)\big\} \tag{C.9}$$

$$\big\{(0000|0000),(0001|0010),(0010|1100),(0100|1001),\\(1011|0100),(1101|0111),(1110|1010),(1111|1110)\big\} \tag{C.10}$$

$$\big\{(0000|0000),(0001|0010),(0010|1100),(0100|1001),\\(1011|1110),(1101|1011),(1110|0101),(1111|0111)\big\} \tag{C.11}$$

$$\big\{(0000|0000),(0001|0010),(0010|1100),(0101|1000),\\(1011|0100),(1100|0110),(1110|1010),(1111|1110)\big\} \tag{C.12}$$

$$\big\{(0000|0000),(0001|0010),(0010|1100),(0101|1000),\\(1011|0100),(1100|1010),(1110|0110),(1111|1110)\big\} \tag{C.13}$$

$$\big\{(0000|0000),(0001|0010),(0010|1100),(0101|1000),\\(1011|1110),(1100|1010),(1110|0110),(1111|0100)\big\} \tag{C.14}$$

$$\big\{(0000|0000),(0001|0010),(0100|1010),(0101|1000),\\(1010|0110),(1011|0100),(1110|1100),(1111|1110)\big\} \tag{C.15}$$

$$\big\{(0000|0000),(0001|0010),(0100|1010),(0101|1000),\\(1010|1100),(1011|0100),(1110|0110),(1111|1110)\big\} \tag{C.16}$$

$$\big\{(0000|0000),(0001|0010),(0100|1010),(0111|1000),\\(1001|0100),(1010|0110),(1110|1100),(1111|1110)\big\} \tag{C.17}$$

$$\big\{(0000|0000),(0001|0110),(0010|1100),(0011|1010),\\(1100|0001),(1101|0111),(1110|1101),(1111|1011)\big\} \tag{C.18}$$

$$\big\{(0000|0000), (0001|0110), (0010|1100), (0011|1010), (1100|0110), (1101|0000), (1110|1010), (1111|1100)\big\} \tag{C.19}$$

$$\big\{(0000|0000), (0001|0110), (0010|1100), (0011|1010), (1100|1010), (1101|0000), (1110|0110), (1111|1100)\big\} \tag{C.20}$$

$$\big\{(0000|0000), (0001|0110), (0010|1100), (0011|1010), (1100|1010), (1101|1100), (1110|0110), (1111|0000)\big\} \tag{C.21}$$

$$\big\{(0000|0000), (0001|0110), (0010|1100), (0110|0101), (1011|0010), (1100|0001), (1101|0111), (1111|1011)\big\} \tag{C.22}$$

$$\big\{(0000|0000), (0001|0110), (0010|1100), (0111|1010), (1001|0000), (1100|0110), (1110|1010), (1111|1100)\big\} \tag{C.23}$$

$$\big\{(0000|0000), (0001|0110), (0010|1100), (0111|1010), (1001|0000), (1100|1010), (1110|0110), (1111|1100)\big\} \tag{C.24}$$

$$\big\{(0000|0000), (0001|0110), (0010|1100), (0111|1100), (1001|0000), (1100|1010), (1110|0110), (1111|1010)\big\} \tag{C.25}$$

$$\big\{(0000|0000), (0001|0110), (0110|0011), (0111|0101), (1000|0110), (1001|0000), (1110|0101), (1111|0011)\big\} \tag{C.26}$$

$$\big\{(0000|0000), (0001|0110), (0110|0011), (0111|1011), (1000|0110), (1001|0000), (1110|0101), (1111|1101)\big\} \tag{C.27}$$

$$\big\{(0000|0000), (0001|0110), (0110|1011), (0111|1101), (1000|0110), (1001|0000), (1110|1101), (1111|1011)\big\} \tag{C.28}$$

$$\big\{(0000|0000), (0011|0100), (0101|1000), (0110|1100), (1001|0010), (1010|0110), (1100|1010), (1111|1110)\big\} \tag{C.29}$$

$$\big\{(0000|0000), (0001|0000), (0010|0000), (0100|1011), (0111|1011), (1001|0111), (1010|0111), (1111|1100)\big\} \tag{C.30}$$

$$\big\{(0000|0000), (0001|0000), (0010|1100), (0011|1100), (1100|0110), (1101|1010), (1110|0111), (1111|1011)\big\} \tag{C.31}$$

$$\big\{(0000|0000), (0001|0000), (0010|1100), (0101|1010), (1011|0110), (1100|0110), (1110|1010), (1111|1100)\big\} \tag{C.32}$$

$$\big\{(0000|0000), (0001|0000), (0010|1100), (0101|1010), (1011|0110), (1100|1010), (1110|0110), (1111|1100)\big\} \tag{C.33}$$

$$\big\{(0000|0000), (0001|0000), (0010|1100), (0011|1100), (1100|0110), (1101|0110), (1110|1010), (1111|1010)\big\} \tag{C.34}$$

$$\big\{(0000|0000), (0001|0000), (0010|1100), (0011|1100), (1100|0110), (1101|1010), (1110|1010), (1111|0110)\big\} \tag{C.35}$$

$$\big\{(0000|0001), (0010|0100), (0011|1000), (0100|1000), (0101|0010), (1000|0010), (1001|0100), (1110|0001), (1111|1111)\big\} \tag{C.36}$$

$$\big\{(0000|0000), (0001|0000), (0010|0000), (0011|1100), (0100|0001), (1000|0110), (1001|0000), (1110|0101), (1111|1011)\big\} \tag{C.37}$$

$$\big\{(0000|0000), (0001|0000), (0010|0000), (0011|1100), (0100|0001), (0110|1000), (1000|0101), (1010|0000), (1101|0110), (1111|1011)\big\} \tag{C.38}$$

$$\big\{(0000|0000), (0001|0000), (0010|0000), (0011|0000), (0100|0000), (0101|1010), (1000|0100), (1001|0010), (1110|1001), (1111|0111)\big\} \tag{C.39}$$

$$\big\{(0000|0000), (0001|0000), (0010|0000), (0011|0000), (0100|0000), (0101|0000), (0110|0000), (0111|0000), (1000|0000), (1001|0110), (1110|0011), (1111|0101)\big\} \tag{C.40}$$

D. LO and Noisy boxes

In this section, I study further the connection between noisy multipartite boxes and graph-theoretical invariants. First, I elaborate on the example of noisy PR-boxes presented in section 3.5.3, and then briefly discuss more general cases.

In the spirit of section 3.5.3, consider k copies of a noisy PR-box distributed among $2k$ parties as in Fig. D.1. The joint conditional probability distribution is:

$$\begin{aligned} P_q(a_1 b_1 \cdots a_k b_k | x_1 y_1 \cdots x_k y_k) &= \prod_{j=1}^{k} P_q(a_j b_j | x_j y_j) \\ &= \prod_{j=1}^{k} \left[q \cdot PR(a_j b_j | x_j y_j) + (1-q)\frac{\mathbb{I}}{4} \right] . \end{aligned} \tag{D.1}$$

Denote by q_k the maximum value of q for which P_q satisfies LO^k. In what follows, we focus on bounding the values of q_k by means of graph invariants (namely, the Shannon Capacity), and work with the non-orthogonality graph defined in section 3.4. To simplify the notation, here we denote only by NO the non-orthogonal graph, and leave implicit the scenario which it refers to.

Similar to section 3.5.3, we focus on the possible events of the (noiseless) box $P_1^k(a_1 b_1 \cdots a_k b_k | x_1 y_1 \cdots x_k y_k)$, and study then the *non-orthogonality graph of possible events*, $\mathrm{NO}_{\mathrm{poss}}$. In the one-copy case ($k = 1$), $\mathrm{NO}^1_{\mathrm{poss}}$ corresponds to the graph of Fig. 3.6(b) (which is the complement of the graph of Fig. 3.6(a)), and in the many-copy case ($k > 1$), $\mathrm{NO}^k_{\mathrm{poss}} = (\mathrm{NO}^1_{\mathrm{poss}})^{\boxtimes k}$ (see section 3.4 and appendix A). A maximal independent set in $\mathrm{NO}^k_{\mathrm{poss}}$ corresponds to an LO inequality in the complete scenario (kn, m, d), though it may not be maximal. Therefore, the maximum value of q for which P_q^k satisfies all LO inequalities generated from $\mathrm{NO}^k_{\mathrm{poss}}$, denoted by q_k^*, only gives an upper bound for q_k.

The conditional probability distribution defined in eq. (D.1) assigns to each vertex of the graph $\mathrm{NO}^k_{\mathrm{poss}}$ a weight $w(q) = [(1+q)/4]^k$. If we denote by α_k the unweighted independence number (see section 2.3 and appendix A) of $\mathrm{NO}^k_{\mathrm{poss}}$, direct evaluation of the maximal LO inequality by P_q^k gives $\alpha_k \, w(q) \leq 1$, which implies that $q_k^* = (4/\sqrt[k]{\alpha_k}) - 1$.

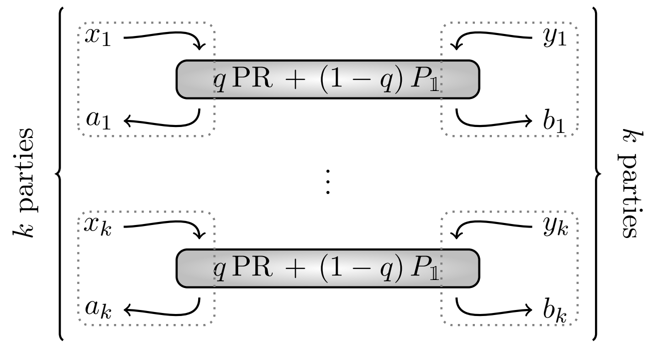


Figure D.1.: k copies of a noisy PR-box shared among $2k$ parties. Each party has access to one part of a box.

E.g. for $k = 2$, we have $\alpha_2 = 5$, so $q_2^* \approx 0.79$. In section 3.5.3, we focused on the 2 copy case, and analyzed the non-orthogonality graph of all events, not only the possible. There we find the value $q_2 \approx 0.72$, which is consistent with $q_2 \leq q_2^*$, and is significantly closer to Tsirelson's bound $q = 1/\sqrt{2} \approx 0.707$.

In the limit of infinite number of copies of the noisy PR-box, the critical noise is related to the unweighted Shannon capacity of $\mathrm{NO}^1_{\mathrm{poss}}$:

$$q^*_\infty := \frac{4}{\Theta(\mathrm{NO}^1_{\mathrm{poss}})} - 1\,. \tag{D.2}$$

and this upper bounds all q_k^*.

The graph of interest $\mathrm{NO}^1_{\mathrm{poss}}$ (depicted in Fig. 3.6(b)) is the Cayley graph of the cyclic group $\mathbb{Z}_8$ with respect to the generating set $\{1, 2, 6, 7\}$. Alternatively, it can be regarded as the circulant graph $Ci_4(1, 2)$, or more specifically, as the 4-antiprism graph (GL04). Unfortunately, to the best of our knowledge, its Shannon capacity is not known.

Consider now the case of a multipartite box in a general scenario (n, m, d) having the property that each probability $P(e)$ has a constant value c if the event e is possible, and 0 otherwise. Then, the method described above for the PR-box can be applied here as well, and a similar relation between the critical noise level q^*_∞ and a Shannon capacity $\Theta(\mathrm{NO}^1_{\mathrm{poss}})$ is found, where $\mathrm{NO}^1_{\mathrm{poss}}$ is the non-orthogonality graph associated to the possible events of the box. This result may be understood from two different perspectives. On the one hand, if the Shannon capacity $\Theta(\mathrm{NO}^1_{\mathrm{poss}})$ happens to be known, then an upper bound on the critical noise level is found. On the other hand, lower bounds on q^* follow from finding quantum representations of the box at a certain noise level, and these bounds translate into upper bounds on $\Theta(\mathrm{NO}^1_{\mathrm{poss}})$. These bounds, however, may be dominated by the Lovász number $\vartheta(\mathrm{NO}^1_{\mathrm{poss}})$ (see appendix A).

For instance, in the case of the noisy PR-box, the existence of a quantum representation at $q = 1/\sqrt{2}$ implies that $q^*_\infty \geq 1/\sqrt{2}$, which eq. (D.2) turns into $\Theta(\mathrm{NO}^1_{\mathrm{poss}}) \leq 4(2 - \sqrt{2})$, a bound which coincides with the Lovász number $\vartheta(\mathrm{NO}^1_{\mathrm{poss}})$. If one could prove this bound to be the exact value, then eq. (D.2) would recover Tsirelson's bound. Similar considerations can be found in (Cab13).

Computing the boundary of $\mathcal{LO}^\infty$ is difficult, and even approximating it is computationally costly. We have seen that it can be related to a purely combinatorial problem, namely the evaluation of the Shannon capacity of certain graphs. In fact, the Shannon capacity gives a bound on the amount of noise required to make a conditional probability distribution quantum, and in turn, this could lead to bounds on the Shannon capacity of graphs. This is an interesting connection to graph theory, similar to those already found in (CSW10) and chapter 4.

E. Relation to the observable-based approach

The observable-based approach to quantum contextuality and nonlocality has first been studied explicitly by Abramsky and Brandenburger (AB11). In this appendix, our goal is to show how the observable-based approach can be embedded into our formalism. A converse construction should be possible upon augmenting the observable-based approach by additional constraints as in (AB11, Sec. 7). In this sense, the two formalisms are essentially equivalent. We believe that both approaches have their merits; for example, in both cases the relation to sophisticated mathematical methods can be exploited. In the observable-based approach, this has been done in (AMSB11); for the hypergraph-based approach, this has been started in (CSW10) and further developed in this thesis.

The following definition blends the terminology of (AB11) with the one of (FC13).

Definition E.1. *A* ***marginal scenario*** $(X, O, \mathcal{M})$ *is defined by a finite set* $X = \{A_1, \dots, A_n\}$*, the elements of which we call* ***observables****, together with a finite set* O *of outcomes and a* ***measurement cover*** $\mathcal{M}$*, which is a family of subsets* $\mathcal{M} \subseteq 2^X$ *such that*

(i) every element of X *occurs in some* C*:* $\bigcup_{C\in\mathcal{M}} C = X$*.*

(ii) $\mathcal{M}$ *is an anti-chain:* $C, C' \in \mathcal{M}$*,* $C \subseteq C' \implies C = C'$*.*

The $C \in \mathcal{M}$ *are called* ***measurement contexts****.*

We denote a marginal scenario $(X, O, \mathcal{M})$ simply by X, at least when O and $\mathcal{M}$ are clear from the context.

From the mathematical point of view, the maximal sets of compatible observables form also a hypergraph, but the physical interpretation is quite different. The subsets in $\mathcal{M}$ represent the maximal sets of jointly measurable observables. Figure E.1 displays the hypergraph corresponding to the CHSH scenario (CHSH69), in which the four pairs $\{A_1, B_1\}$, $\{A_1, B_2\}$, $\{A_2, B_1\}$, $\{A_2, B_2\}$

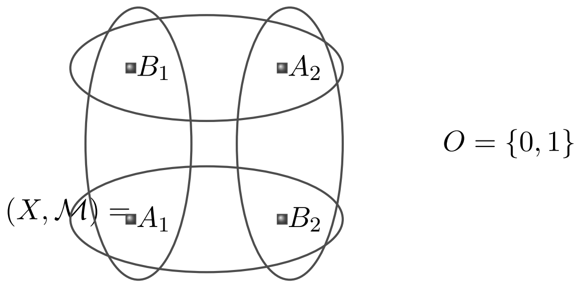


Figure E.1.: The CHSH scenario as a marginal scenario. We now draw the vertices as squares in order to indicate that the interpretation differs from the one of all other illustrations of hypergraphs in this paper.

are jointly measurable, but no other pairs or triples of observables are jointly measurable.

As noted in (AB11), it is not a substantial restriction to assume that all observables take values in the same set of outcomes O. We assume this mainly for convenience of notation and note that all of our considerations and results can easily be extended to the general case in which each measurement $A \in X$ takes values in an associated finite set of outcomes O_A depending on A.

In the following, we want to consider measurements of compatible observables which are conducted in a certain temporal order. The situation where we have already measured some observable $A \in X$, defines a marginal scenario which encodes all the possibilities for subsequent measurements in the following sense.

Definition E.2. *Given an observable $A \in X$, the* ***induced marginal scenario*** *$X\{A\}$ is the marginal scenario having observables*

$$X\{A\} = \left\{ A' \in X \;\middle|\; A' \neq A,\ \exists C \in \mathcal{M} \text{ s.t. } \{A, A'\} \subseteq C \right\}$$

and measurement contexts defined to be the restrictions of those $C \in \mathcal{M}$ with $A \in C$ down to $X\{A\}$.

By definition, any $X\{A\}$ has a smaller number of observables than the original X. In particular, iterating this construction by taking an induced marginal scenario of an induced marginal scenario, one eventually ends up with an empty scenario and the process terminates.

This idea motivates the following recursive definition of measurement protocol:

Definition E.3. *A* ***measurement protocol*** *T on a marginal scenario X is* *(i)* $T = \emptyset$ *if* $X = \emptyset$*;*

(ii) otherwise, $T = (A, f)$*, where* $A \in X$ *is an observable and* $f : O \to \mathrm{MP}(X\{A\})$ *is a function, where* $\mathrm{MP}(X\{A\})$ *is the set of all measurement protocols on the scenario* $X\{A\}$*.*

Intuitively, a measurement protocol consists of a choice of observable and an assignment of a new measurement protocol to each outcome of the observable, where the new measurement protocol lives on the induced marginal scenario.

Upon unravelling the recursive structure of this definition, one finds that a measurement protocol specifies sequences of measurements which can be applied to the system, where the choices of subsequent measurements f are allowed to depend on the outcomes of the earlier ones. These measurement sequences have the additional property that all measurements in a sequence are compatible and that no measurement can occur twice in the same sequence. We use the letter "T" to indicate the tree-like appearance of this structure. Note that every measurement sequence is automatically maximal in the sense that it contains all observables of a certain measurement context.

The set of outcomes $\mathrm{Out}(T)$ of a measurement protocol T is also defined recursively: if $T = \emptyset$, then there is only a single outcome which we denote by "$*$", so that $\mathrm{Out}(\emptyset) = \{*\}$. Otherwise, we have $T = (A, f)$ and put

$$\mathrm{Out}(T) := \{\, (a, \alpha) \;:\; a \in O,\ \alpha \in \mathrm{Out}(f(a)) \,\}\,.$$

In this way, an element of $\mathrm{Out}(T)$ corresponds to a measurement sequence in T together with an associated sequence of outcomes for these measurements such that applying the protocol to any outcome in the sequence results in the following measurement, except if the outcome is the last one in the sequence.

In terms of this concepts, we define a contextuality scenario H associated to a marginal scenario X as follows.

Definition E.4. *The contextuality scenario* $H[X]$ *associated to a marginal scenario* X *has vertices*

$$V(H[X]) := \left\{ s \in O^C \;:\; C \in \mathcal{M} \right\}$$

and edges

$$E(H[X]) := \{\mathrm{Out}(T) \;:\; T \in \mathrm{MP}(X)\}\,.$$

We write P for an **empirical model** on X (AB11). This means that for each $C \in \mathcal{M}$, P_C is a probability distribution over O^C, such that the **sheaf condition** holds:

$$P_{C|C\cap C'} = P_{C'|C\cap C'} \quad \forall C, C' \in \mathcal{M}, \tag{E.1}$$

where $P_{C|C\cap C'}$ stands for the marginal distribution of P_C associated to the observables in $C \cap C'$. For an assignment of outcomes $s \in O^C$, $P_C(s)$ is to be thought of as the probability of obtaining the joint outcome s when jointly measuring all observables in C. The sheaf condition is a generalization of the no-signaling condition in this observable-based approach.

To see the equivalence with our framework, we further need to prove that every empirical model P is associated to a well defined probabilistic model p, and vice-versa. The idea then is to associate to an empirical model P a probabilistic model on $H[X]$ by setting, for each $C \in \mathcal{M}$ and each $s \in O^C$,

$$p(s : C \to O) := P_C(s). \tag{E.2}$$

It will need to be verified that this actually is a probabilistic model, i.e. that these probabilities are suitably normalized for every edge in $E[X]$. Conversely, given a probabilistic model p on $H[X]$, we claim that (E.2) defines an empirical model P on X.

Theorem E.5. *This defines a linear bijection between empirical models on X and probabilistic models on $H[X]$.*

Proof. We first verify that (E.2) turns an empirical model P into a probabilistic model p. It needs to be shown that

$$\sum_{s\in\mathrm{Out}(T)} P_C(s) = 1 \tag{E.3}$$

for any measurement protocol T. In order to prove this, we introduce the notion of **post-measurement** empirical model. Suppose that a measurement has resulted in an outcome $a \in O$ for an observable $A \in X$. Then for the subsequent measurements in the scenario $X\{A\}$, we expect the posterior probabilities

$$P_C^{\mathrm{post}(a)}(s) = \frac{P_C(s)}{P_{\{A\}}(a)}.$$

We now use induction on the size of X in order to prove (E.3). The base case is $X = \emptyset$, in which there is nothing to prove. For the induction step, we decompose $T = (A, f)$ and use the induction assumption on each $P^{\mathrm{post}(a)}$ for those $a \in O$ with $P_{\{A\}}(a) \neq 0$. Then

$$\sum_{s\in\mathrm{Out}(T)} P_C(s) = \sum_a \sum_{\alpha\in\mathrm{Out}(f(a))} P_{\{A\}}(a) P_C^{\mathrm{post}(a)}(\alpha) = \sum_a P_{\{A\}}(a) = 1,$$

as was to be shown.

Conversely, we need to prove that if p is a probabilistic model on $H[X]$, then the associated P is an empirical model, i.e. that it satisfies (E.1). It is sufficient to consider the case $C \cap C' \neq \emptyset$, for otherwise (E.1) is vacuous. Let $s_0 \in O^{C \cap C'}$ be an arbitrary joint outcome of the observables $C \cap C'$. Then we consider a measurement protocol T given by conducting the measurements $C \cap C'$, and then conducting the measurements $C \setminus C'$ if the joint outcome was s_0, and conducting the measurements $C' \setminus C$ otherwise. Then the normalization equation associated to this measurement protocol reads

$$\sum_{t \in O^{C \setminus C'}} p(s_0 \cup t) + \sum_{s_0 \neq s \in O^{C \cap C'}} \sum_{t' \in O^{C' \setminus C}} p(s \cup t') = 1.$$

Comparing this with the normalization equation associated to the measurement protocol which simply measures all observables in C' and outputs their joint outcome,

$$\sum_{s \in O^{C \cap C'}} \sum_{t' \in O^{C' \setminus C}} p(s' \cup t') = 1,$$

gives, upon splitting the latter equation into the $s = s_0$ part and the $s \neq s_0$ part,

$$\sum_{t \in O^{C \setminus C'}} p(s_0 \cup t) = \sum_{t' \in O^{C' \setminus C}} p(s_0 \cup t'),$$

as was to be shown. □

There are analogous correspondence theorems for quantum models and classical models.

F. Hierarchies of SDPs: $\mathcal{Q}_1(B_{n,m,d}) \equiv \tilde{\mathcal{Q}}$

In this section, I present the connection between the hierarchy of probabilistic models introduced in section 4.6 and the NPA hierarchy (NPA08; NPA07). In particular, I will introduce another characterization of the $\mathcal{Q}_1$ set, and prove that it is equivalent to the extension of the original "1+AB" level of NPA to multipartite scenarios (NGHA13).

Definition F.1. *Let $H = (V, E)$ be a contextuality scenario. Consider $\mathcal{A}$ to be the *-algebra with generators $\{P_v, v \in V\}$ and relations*

$$P_v = P_v^2 = P_v^*; \qquad P_u\, P_v = P_v\, P_u = 0 \quad \text{if}\, u \sim v; \qquad \sum_{u \in e} P_u = 1 \quad \forall\, e \in E;$$

Let $\mathcal{A}^k \subset \mathcal{A}$ be the set of elements of $\mathcal{A}$ which are written as polynomials of order at most k over the generators of $\mathcal{A}$.

We define $\mathcal{L}^k(H)$ to be the subspace of linear functionals $L : \mathcal{A}^{2k} \to \mathbb{R}$ with the following properties:

(i) $L(1) = 1$.

(ii) $L(q^* q) \geq 0 \quad \forall\, q \in \mathcal{A}^k$.

It is immediate that a probabilistic model $p : V(H) \to [0, 1]$ is a $\mathcal{Q}_k$-model in the sense of def. 4.28 if and only if there exists an linear functional $L_p \in \mathcal{L}^k(H)$ such that $L_p(v) = p(v)$. Indeed, the positive semidefinite matrix M of def. 4.28 is linked to L_p through: $M_{q_1,q_2} = L(q_1 q_2^\dagger)$.

For Bell scenarios $B_{n,m,d}$, the set $\tilde{\mathcal{Q}}(B_{n,m,d})$ of "almost quantum" correlations (NGHA13), corresponding to the level "1 + AB" of the NPA hierarchy, can be alternatively defined as follows:

Definition F.2 ((Nav13))**.** *Consider the multipartite Bell scenario $B_{n,m,d}$. For each party j, let $\mathscr{A}_j$ be the abstract unital *-algebra with hermitian generators $\{E_j^{x,a}\}_{\substack{a=0\ldots d-1 \\ x=0\ldots m-1}}$ subject to the identities*

$$E_j^{x,a}\, E_j^{x,\tilde{a}} = \delta_{a,\tilde{a}} E_j^{x,a}, \qquad \sum_{a=0}^{d-1} E_j^{x,a} = \mathbb{1}. \tag{F.1}$$

Construct then the extended algebra $\mathscr{A} = \mathscr{A}_1 \otimes \ldots \otimes \mathscr{A}_n$*, with generators* $\{E_j^{x_j,a_j}, j = 1 \ldots n\}_{\substack{a_j=0\ldots d-1 \\ x_j=0\ldots m-1}}^{j=1\ldots n}$ *satisfying conditions (F.1) for every* j*, together with the commutation relations*

$$[E_i^{x,a}, E_j^{\tilde{x},\tilde{a}}] = 0 \quad \text{for } i \neq j. \tag{F.2}$$

Define $\mathscr{A}^k$*, with* $k \in \mathbb{N}$*, as the set of elements of* $\mathscr{A}$ *which are expressed as a polynomial* $\mathcal{P}$ *over the generators of* $\mathscr{A}$*, such that each term in* $\mathcal{P}$ *is of degree at most* k *in the generators of* $\mathscr{A}_j$ *for every* $j = 1 \ldots n$*.*

Define the subspace $\mathscr{L}^k$ *of linear functionals* $L : \mathscr{A}^{2k} \to \mathbb{R}$ *such that*

(i) $L(\mathbb{1}) = 1$.

(ii) $L(q^*q) \geq 0 \quad \forall\, q \in \mathscr{A}^k$.

Then, a probabilistic model $p : V(H) \to [0,1]$ *is a* $\tilde{\mathcal{Q}}$*-model in the sense of (NGHA13), i.e. it corresponds to the 1+AB level of the NPA hierarchy, if and only if there exists an linear functional* $L_p \in \mathscr{L}^1(H)$ *such that* $L_p(v) = p(v)$*.*

Theorem F.3 (Equivalence between the two hierarchies)**.** *For Bell scenarios,*

$$\mathscr{L}^k(B_{n,m,d}) = \mathcal{L}^k(B_{n,m,d}).$$

Proof. Consider a Bell scenario $B_{n,m,d}$, we will show that the two algebras $\mathcal{A}^k$ and $\mathscr{A}^k$ are isomorphic. Indeed, consider a vertex $v = (a_1 \ldots a_n | x_1 \ldots x_n)$ of the scenario. Starting with generators $E_j^{x_j,a_j}$ of $\mathscr{A}_j$, define:

$$P_v = \prod_{j=1}^{m} E_j^{x_j,a_j}.$$

One easily checks that the relations of the algebra $\mathcal{A}$ are satisfied, namely:

(i) $P_v = P_v^2 = P_v^*$ since the elements $E_j^{x_j,a_j}$ are Hermitian projectors;

(ii) if $v \sim v'$, it means that one party, say party 1, is such that $x_1 = x_1'$, but $a_1 \neq a_1'$. In particular,

$$P_v P_u = E_1^{x_1,a_1} E_1^{x_1,a_1'} \prod_{j=2}^{n} E_j^{x_j,a_j} E_1^{x_j',a_j'} = 0 = P_u P_v;$$

(iii) Consider an edge e and assume without loss of generality that party $(j+1)$ measures after party j and that their measurement choice is $x_{j+1} = f_{j+1}(a_1 \dots a_j, x_1 \dots x_j)$, then:

$$\sum_{v \in e} P_v = \sum_{a_1} E_1^{x_1,a_1} \sum_{a_2} E_2^{f_2(a_1,x_1),a_2} \sum_{a_2} \dots = 1,$$

by noticing that $\sum_{a_j} E_j^{f(a_1 \dots a_{j-1}, x_1 \dots x_{j-1})),a_j} = 1$.

Conversely, start with generators P_v of the algebra $\mathcal{A}$, then one can define generators of $\mathscr{A}_j$ as:

$$E_j^{x_j,a_j} = \sum_{i \neq j} \sum_{a_i=0}^{d} P_{(a_1 \cdots a_n | x_1 \cdots x_n)},$$

where the choice of x_i for $i \neq j$ is arbitrary, as can be shown thanks to the relation $\sum_{v \in e} P_v = 1$. The relations given by F.1, as well as the commutativity of the elements on different algebras $\mathscr{A}_i$ and $\mathscr{A}_j$ are easily seen to hold.

With these constructions, we also notice that the degrees of equivalent polynomials coincide in both hierarchies, showing that $\mathcal{A}^k$ is isomorphic to $\mathscr{A}^k$. We conclude that the sets of linear functionals $\mathcal{L}^k$ and $\mathscr{L}^k$ also coincide, which in particular proves that the first level $\mathcal{Q}_1$ in our hierarchy is equivalent to the "almost quantum" set $\tilde{\mathcal{Q}}$. □

G. Facets of $\mathbb{P}_2^S$ and $\mathbb{P}_{2,n}^T$

G.1. Fully symmetric two-body polytope

In the case of fully symmetric two-body Bell Inequalities, the polytope $\mathbb{P}_2^S$ is already by construction invariant under any permutation of the parties. Hence, when classifying its facets info equivalence classes, we focus on the following symmetries:

- Renaming of observables by applying the transformation $\mathcal{M}_0^{(i)} \leftrightarrow \mathcal{M}_1^{(i)}\ \forall i$. This operation interchanges both $\alpha \leftrightarrow \beta$ and $\gamma \leftrightarrow \varepsilon$ in (5.3).
- Renaming of outcomes by applying the transformation $\mathcal{M}_x^{(i)} \leftrightarrow -\mathcal{M}_x^{(i)}\ \forall i$ for a particular $x \in \{0,1\}$. For $x = 0$ it implies $\alpha \leftrightarrow -\alpha$ and $\delta \leftrightarrow -\delta$ in (5.3), while for $x = 1$, $\beta \leftrightarrow -\beta$ and $\delta \leftrightarrow -\delta$.

From Table G.1 to Table G.5, I present the equivalence classes for the facets of $\mathbb{P}_2^S$ for $n = 3$ to $n = 6$.

Table G.1.: Equivalence classes for the facets of $\mathbb{P}_2^S$ for $n = 3$.

#	β_C	α	β	γ	δ	ϵ
1	1	0	0	0	0	1
2	18	-2	6	-2	-3	6
3	3	0	0	1	1	0
4	6	2	2	0	1	0
5	3	2	0	1	0	0
6	3	0	0	0	0	-1

G.2. Translationally invariant two-body polytope

In the case of tranlationally invariant two-body Bell Inequalities, the facets of $\mathbb{P}_{2,n}^T$ obey the following symmetries:

Table G.2.: Equivalence classes for the facets of $\mathbb{P}_2^S$ for $n = 4$.

#	β_C	α	β	γ	δ	ϵ
1	2	0	-1	0	0	1
2	42	-9	12	1	-6	6
3	30	-3	-12	-1	2	6
4	54	-6	12	-1	-8	12
5	20	-3	5	0	-3	4
6	18	0	0	-1	-2	6
7	12	-3	-3	1	2	1
8	6	0	3	0	0	1
9	8	1	3	0	1	2
10	6	0	0	0	1	1
11	8	-2	0	1	1	1
12	12	3	3	0	1	0
13	6	0	0	0	0	-1

- Renaming of parties in a cyclical way $A_j^{(i)} \to A_j^{(i+1)}\ \forall i, j$. This symmetry leaves $\mathbb{P}_{2,n}^T-$invariant, by construction.
- Renaming of observables for all parties $A_0^{(i)} \leftrightarrow A_1^{(i)}\ \forall i$. This symmetry changes $\alpha \leftrightarrow \beta$, $\gamma_k \leftrightarrow \epsilon_k$ and $\delta_k \leftrightarrow \delta_{n-k}$ in (5.70).
- Renaming of j-th observable outcomes for all parties: $A_j^{(i)} \leftrightarrow -A_j^{(i)}\ \forall i$. This symmetry changes in (5.70) $\delta_k \leftrightarrow -\delta_k$ and $\alpha \leftrightarrow -\alpha$ if $j = 0$ or $\delta_k \leftrightarrow -\delta_k$ and $\beta \leftrightarrow -\beta$ if $j = 1$.
- Renaming of parties by applying the symmetry $A_j^{(i)} \leftrightarrow A_j^{(n-i+1)}\ \forall i, j$. This symmetry changes in (5.70) $\delta_k \leftrightarrow \delta_{n-k}$.

The above symmetries are valid in the most general case, when the coefficients of the Bell inequality are unconstrained. However, if some of them are 0, then translational invariance needs not be preserved for the corresponding correlator and further symmetries can be exploited. For example, if $\alpha = \beta = 0$ and n is even, then applying $A_j^{(i)} \leftrightarrow -A_j^{(i)}\ \forall i$ even , $\forall j$, leads to the symmetry changes in (5.70) $\gamma_k \leftrightarrow -\gamma_k, \delta_k \leftrightarrow -\delta_k, \epsilon_k \leftrightarrow -\epsilon_k$ for all odd k. This was taken into account when classifying the Bell Inequalities for $n = 3, 4$ into the families of Tables G.6 to G.9.

Table G.3.: Equivalence classes for the facets of $\mathbb{P}_2^S$ for $n = 5$.

#	β_C	α	β	γ	δ	ϵ
1	4	0	-2	0	0	1
2	8	0	0	0	-1	2
3	24	-6	6	1	-2	1
4	14	-4	2	1	-1	1
5	400	-36	60	2	-45	60
6	160	-12	-60	-2	5	20
7	90	-20	22	3	-8	5
8	80	4	-20	-2	-5	20
9	20	-2	8	0	-1	3
10	130	-28	-36	3	10	8
11	110	-24	-30	3	9	7
12	10	0	2	1	1	1
13	20	0	0	3	3	1
14	20	4	4	0	1	0
15	70	-20	-14	5	5	1
16	20	-4	0	3	2	1
17	20	4	4	1	2	1
18	200	-60	-24	30	15	-2
19	40	8	12	1	3	3
20	30	4	10	1	2	3
21	40	6	12	0	3	5
22	10	4	0	1	0	0
23	2	0	0	1	0	0
24	10	0	0	0	0	-1

Table G.4.: Equivalence classes for the facets of $\mathbb{P}_2^S$ for $n = 6$.

#	β_C	α	β	γ	δ	ϵ
1	7	0	-3	0	0	1
2	16	-4	2	1	-1	1
3	15	1	-4	0	-1	3
4	240	15	-45	-1	-18	45
5	132	-25	27	4	-9	6
6	180	-10	-60	-1	3	15
7	70	-18	12	3	-3	1
8	52	-15	7	3	-2	1
9	156	-35	31	5	-8	3
10	375	-30	45	8	-36	45
11	90	5	-30	-1	-3	15
12	300	0	0	1	-27	45
13	34	-7	7	1	-2	1
14	24	-3	1	1	-2	3
15	114	-15	19	4	-11	12
16	129	-20	24	5	-12	12
17	225	-54	41	8	-10	3
18	192	-46	36	7	-9	3
19	39	3	-8	0	-3	7
20	112	12	-26	1	-9	17
21	156	17	-37	1	-12	23
22	42	-9	-9	1	2	1
23	3	0	1	0	0	1
24	120	-20	30	1	-5	5
25	240	-45	-55	4	13	10
26	165	-20	-40	1	12	20
27	24	-6	-4	1	1	1
28	195	-20	60	-1	-8	20
29	24	-3	7	0	-1	2
30	60	0	0	-1	3	15
31	87	-15	-22	1	4	4
32	12	-2	0	1	1	1
33	30	-5	-5	2	3	2
34	15	0	5	0	0	1
35	42	3	15	0	1	4
36	24	1	5	1	2	3
37	45	5	10	1	4	6
38	51	4	15	1	3	6
39	99	11	20	2	9	13

Table G.5.: Equivalence classes for the facets of $\mathbb{P}_2^S$ for $n = 6$ – *cont.*

#	β_C	α	β	γ	δ	ϵ
40	273	32	60	5	24	36
41	105	-30	-15	6	4	1
42	87	-27	-10	6	3	1
43	51	-16	-5	4	2	1
44	48	-13	-5	4	3	2
45	30	0	0	1	3	3
46	39	12	-3	4	-2	1
47	30	5	5	0	1	0
48	15	0	0	0	0	-1

Table G.6.: Equivalence classes for the facets of $\mathbb{P}_{2,3}^T$.

#	β_C	α	β	γ_1	δ_1	δ_2	ϵ_1
1	1	0	0	0	0	0	1
2	3	0	0	0	1	-1	-1
3	3	0	0	1	1	1	0
4	3	1	1	0	1	0	0
5	3	2	0	1	0	0	0
6	9	-1	-3	-1	1	2	3

Table G.7.: Equivalence classes for the facets of $\mathbb{P}_{2,4}^T$.

#	β_{NS}	β_Q	β_C	α	β	γ_1	δ_1	δ_3	ϵ_1	γ_2	δ_2	ϵ_2
1	4	4.0000	4	0	-2	0	0	0	2	0	0	1
2	4	4.0000	4	0	0	0	0	0	-2	0	0	1
3	4	4.0000	4	0	0	0	0	1	1	0	1	0
4	4	4.0000	4	0	0	0	1	1	0	0	0	1
5	4	4.0000	4	0	2	0	0	0	0	0	0	1
6	4	4.0000	4	0	2	0	0	0	1	0	0	0
7	4	4.0000	4	1	1	0	0	0	0	0	1	0
8	4	4.0000	4	1	1	0	0	1	0	0	0	0
9	8	8.0000	8	-2	-2	0	2	2	0	1	0	1
10	48/5	8.4230	8	-2	-2	1	0	1	1	0	1	0
11	8	8.0000	8	-2	-2	1	1	1	1	0	2	0
12	10	8.8284	8	-2	0	1	1	-1	1	0	0	0
13	8	8.0000	8	-1	-1	-1	1	1	1	1	1	0
14	8	8.0000	8	-1	1	1	1	1	1	0	1	1
15	8	8.0000	8	-1	3	0	-1	-1	2	0	-1	1
16	12	9.2665	8	0	0	-2	-1	-1	2	1	0	1
17	16	11.3137	8	0	0	0	0	0	0	-1	2	1
18	8	8.0000	8	0	2	-1	-1	1	1	0	0	0
19	8	8.0000	8	1	1	0	1	1	-2	0	-1	1
20	16	13.6569	12	-2	-2	1	1	1	0	-1	2	1
21	44/3	12.5951	12	-1	-1	-3	1	2	1	2	0	0
22	44/3	12.5155	12	-1	3	-2	-1	-2	2	1	0	1
23	52/3	13.6021	12	0	0	-1	-1	-2	4	0	-1	2
24	20	14.7703	12	0	0	0	-1	-1	4	-1	-2	2
25	20	14.4234	12	0	0	1	1	2	2	-1	3	1
26	76/5	12.9645	12	0	2	-3	-1	-1	1	1	2	0
27	76/5	12.2591	12	0	2	-2	0	0	2	1	2	0
28	44/3	12.0000	12	0	2	-2	0	2	2	1	0	0
29	16	13.6569	12	0	2	-2	2	2	2	2	0	1
30	16	13.6569	12	0	2	-1	0	2	0	-1	-2	1
31	16	13.6569	12	0	2	1	0	2	2	-1	2	1
32	96/5	16.7214	16	-4	-4	1	2	2	0	0	2	1
33	96/5	16.5951	16	-4	-2	2	-1	3	1	0	2	-1
34	16	16.0000	16	-4	0	2	2	2	2	1	2	1
35	18	16.5549	16	-3	-1	1	-2	3	1	0	2	-1
36	96/5	16.4461	16	-2	-2	1	-2	4	1	-1	2	-1
37	96/5	16.5968	16	-2	-2	1	2	2	-4	0	0	3
38	24	18.3698	16	-2	-2	2	2	2	2	-1	4	1
39	20	17.6569	16	-1	-5	0	-1	2	3	-1	2	1

Table G.8.: Equivalence classes for the facets of $\mathbb{P}_{2,4}^{T}$ – *cont.*

#	β_{NS}	β_Q	β_C	α	β	γ_1	δ_1	δ_3	ϵ_1	γ_2	δ_2	ϵ_2
40	64/3	17.2516	16	-1	-1	-5	2	2	1	3	-1	0
41	24	18.0188	16	0	0	-2	-3	-3	4	1	0	3
42	20	17.6569	16	0	4	-2	2	2	2	1	-2	1
43	20	17.6569	16	0	4	0	-2	2	2	-1	2	1
44	16	16.0000	16	0	4	0	2	2	4	1	2	1
45	24	17.5024	16	1	3	1	2	2	3	-1	3	2
46	56/3	16.5849	16	2	2	-2	1	-1	-2	1	2	1
47	16	16.0000	16	2	2	-2	3	1	-2	1	-2	1
48	24	21.4272	20	-4	2	2	-4	2	2	0	-2	-1
49	20	20.0000	20	-3	5	0	-3	-3	4	0	-3	2
50	24	21.4272	20	-2	-8	-2	2	2	4	1	0	2
51	24	20.8420	20	-2	-8	-1	1	1	4	0	2	2
52	24	20.7714	20	-2	-8	-1	2	2	4	0	0	2
53	24	21.4272	20	-2	-8	0	0	2	4	-1	2	2
54	116/5	20.3609	20	-2	-4	0	2	2	4	1	4	0
55	24	21.2376	20	-2	4	-2	-2	-4	4	1	-2	2
56	68/3	20.4573	20	-2	4	-1	-2	-4	4	0	-2	2
57	28	21.1792	20	-2	4	0	-2	-2	4	-1	-4	2
58	32	23.3137	20	-2	4	1	-2	-2	4	-2	-4	2
59	24	21.3099	20	-2	6	1	-3	0	4	-1	-3	1
60	20	20.0000	20	-1	5	-1	-2	-2	5	1	-3	1
61	28	21.8851	20	0	2	-6	2	2	2	3	-2	0
62	28	21.9339	20	0	2	-4	-2	-2	0	2	4	-1
63	116/5	20.0000	20	0	4	-1	-3	3	3	-1	2	0
64	28	25.3099	24	-8	-4	3	3	0	1	1	3	-1
65	28	24.4670	24	-4	8	1	-2	-2	4	-1	-4	2
66	32	25.3003	24	-2	-2	-6	4	4	2	5	0	1
67	32	27.3137	24	-2	6	-4	-4	-4	4	3	0	3
68	32	25.0090	24	0	4	-2	2	4	0	-1	-4	3
69	32	24.9956	24	0	4	2	2	4	4	-1	4	3
70	156/5	28.4819	28	-6	8	0	-4	-4	4	1	-4	2
71	156/5	28.4107	28	-6	8	1	-4	-4	4	0	-4	2
72	36	29.2933	28	-4	-4	0	3	3	4	2	6	-1
73	32	28.4038	28	-2	-8	-4	0	0	4	1	4	2
74	32	28.0000	28	-2	-8	-4	0	4	4	1	0	2
75	36	31.3137	28	-2	-8	-4	4	4	4	2	-2	3
76	32	28.0000	28	-2	-8	-2	-2	4	4	-1	2	2
77	36	31.3137	28	-2	-8	-1	-2	2	4	-2	4	2
78	40	30.8543	28	-2	4	-8	-4	-4	4	5	0	2

Table G.9.: Equivalence classes for the facets of $\mathbb{P}_{2,4}^T$ – *cont.*

#	β_{NS}	β_Q	β_C	α	β	γ_1	δ_1	δ_3	ϵ_1	γ_2	δ_2	ϵ_2
79	40	30.6198	28	-2	4	-4	-6	-6	4	3	0	4
80	36	29.2538	28	-2	4	-2	-6	-4	6	1	-2	4
81	100/3	29.1453	28	-2	4	-2	-2	-6	6	0	-2	3
82	44	31.8400	28	-2	4	0	-2	-2	6	-2	-6	3
83	44	31.7274	28	-2	4	1	-3	-3	7	-2	-6	3
84	36	31.3137	28	-2	8	1	-4	0	6	-2	-4	2
85	36	29.2004	28	-1	-7	-2	4	4	2	0	-3	4
86	40	33.6435	32	-4	0	0	4	4	-4	-1	-6	3
87	192/5	32.6923	32	-1	-7	-1	-4	5	5	-2	4	1
88	44	39.3137	36	-6	-12	-4	4	4	4	3	0	2
89	44	39.3137	36	-6	-12	1	0	4	4	-2	4	2
90	124/3	36.5739	36	-4	8	-1	-5	-6	8	0	-5	4
91	44	36.8908	36	-4	8	0	-5	-5	8	-1	-6	4
92	44	38.6969	36	-2	-8	-8	4	4	4	3	-4	2
93	52	39.1060	36	-2	4	-4	-8	-8	6	3	0	6
94	60	42.8189	36	2	-4	2	-4	-4	8	-3	-8	4
95	48	40.9178	40	-4	8	-2	-8	-6	8	1	-4	5
96	56	42.3181	40	-4	8	1	-5	-5	9	-2	-8	4
97	268/5	46.3398	44	-10	12	3	-4	-4	4	-2	-8	2
98	64	50.4924	48	-4	8	-4	-10	-10	8	3	-2	7
99	500/7	62.8908	60	-14	16	-4	-8	-8	4	5	-4	2
100	500/7	61.8324	60	-10	12	-12	-8	-8	4	9	0	2
101	80	69.7517	64	-4	-16	-8	10	10	8	3	-6	7
102	88	68.9090	64	-4	8	-8	-14	-14	8	5	2	9
103	104	74.4957	64	-4	8	4	-8	-8	12	-5	-14	7

Bibliography

[AAAB12] A. Acín, M. L. Almeida, R. Augusiak, and N. Brunner. Guess your neighbour's input: no quantum advantage but an advantage for quantum theory. May 2012. Preprint at http://arxiv.org/abs/1205.3076.

[AB11] Samson Abramsky and Adam Brandenburger. The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, 13(11):113036, 2011.

[ABB+10] M. L. Almeida, J.-D. Bancal, N. Brunner, A. Acín, N. Gisin, and S. Pironio. Guess Your Neighbor's Input: A Multipartite Nonlocal Game with No Quantum Advantage. *Phys. Rev. Lett.*, 104:230404, Jun 2010.

[ABG+07] Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, Stefano Pironio, and Valerio Scarani. Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.*, 98:230501, Jun 2007.

[ABL+09] Jonathan Allcock, Nicolas Brunner, Noah Linden, Sandu Popescu, Paul Skrzypczyk, and Tamás Vértesi. Closed sets of nonlocal correlations. *Phys. Rev. A*, 80:062107, 2009.

[ABPS09] J. Allcock, N. Brunner, M. Pawłowski, and V. Scarani. Recovering part of the boundary between quantum and nonquantum correlations from information causality. *Phys. Rev. A*, 80:040103, Oct 2009.

[AFK+12] R. Augusiak, T. Fritz, Ma. Kotowski, Mi. Kotowski, M. Pawłowski, M. Lewenstein, and A. Acín. Tight Bell inequalities with no quantum violation from qubit unextendible product bases. *Phys. Rev. A*, 85:042113, Apr 2012.

[AGCA12] Leandro Aolita, Rodrigo Gallego, Adán Cabello, and Antonio Acín. Fully nonlocal, monogamous, and random genuinely multipartite quantum correlations. *Phys. Rev. Lett.*, 108:100401, Mar 2012.

[AGR82] Alain Aspect, Philippe Grangier, and Gérard Roger. Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities. *Phys. Rev. Lett.*, 49(2):91–94, Jul 1982.

[AL06] Noga Alon and Eyal Lubetzky. The shannon capacity of a graph and the independence numbers of its powers. *IEEE Trans. Inf. Theory*, 52(5), 2006.

[Alo98] Noga Alon. The Shannon capacity of a union. *Combinatorica*, 18(3):301–310, 1998.

[AMSB11] Samson Abramsky, Shane Mansfield, and Rui Soares Barbosa. The cohomology of non-locality and contextuality. In *Proceedings 8th International Workshop on Quantum Physics and Logic (Nijmegen, 2011)*. 2011.

[ASH+11] R. Augusiak, J. Stasińska, C. Hadley, J. K. Korbicz, M. Lewenstein, and A. Acín. Bell Inequalities with No Quantum Violation and Unextendable Product Bases. *Phys. Rev. Lett.*, 107:070401, Aug 2011.

[ASS11] Sabri W. Al-Safi and Anthony J. Short. Information causality from an entropic and a probabilistic perspective. *Phys. Rev. A*, 84:042323, Oct 2011.

[ATSL12] R. Augusiak, J. Tura, J. Samsonowicz, and M. Lewenstein. Entangled symmetric states of n qubits with all positive partial transpositions. *Phys. Rev. A*, 86:042316, 2012.

[Bar07] Jonathan Barrett. Information processing in generalized probabilistic theories. *Phys. Rev. A*, 75(3):032304, 2007.

[BBL+06] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger. Limit on Nonlocality in Any World in Which Communication Complexity Is Not Trivial. *Phys. Rev. Lett.*, 96:250401, Jun 2006.

[BBOC+10] Howard Barnum, Jonathan Barrett, Lisa Orloff Clark, Matthew Leifer, Robert Spekkens, Nicholas Stepanik, Alex Wilce, and

Robin Wilke. Entropy and information causality in general probabilistic theories. *New Journal of Physics*, 12(3):033024, 2010.

[BCP+13] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner. Bell nonlocality. 2013. Preprint at http://arxiv.org/abs/1303.2849.

[BDM+99] C. H. Bennett, D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin, and B. M. Terhal. Unextendible Product Bases and Bound Entanglement. *Phys. Rev. Lett.*, 82:5385–5388, 1999.

[Bel64] John S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.

[Ber61] C. Berge. Färbung von graphen, deren sämtliche bzw. deren ungerade kreise starr sind. *Wiss. Z. Martin-Luther-Univ. Halle-Wittenberg Math.-Natur. Reihe*, 10(114), 1961.

[BFRW05] Howard Barnum, Christopher A. Fuchs, Joseph M. Renes, and Alexander Wilce. Influence-free states on compound quantum systems. 2005. Preprint at http://arxiv.org/abs/quant-ph/0507108.

[BG13] Marcel Bergmann and Otfried Gühne. Entanglement criteria for Dicke states. 2013. Preprint at http://arxiv.org/abs/1305.2818.

[BGP10] J. D. Bancal, N. Gisin, and S. Pironio. Looking for symmetric Bell inequalities. *Journal of Physics A: Mathematical and Theoretical*, 43(38):385303, 2010. Data available at http://www.gapoptic.unige.ch/Publications/bellinequalities.

[BKP06] Jonathan Barrett, Adrian Kent, and Stefano Pironio. Maximally nonlocal and monogamous quantum correlations. *Phys. Rev. Lett.*, 97:170409, Oct 2006.

[BLM+05] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts. Nonlocal correlations as an information-theoretic resource. *Phys. Rev. A*, 71:022101, 2005.

[Boo62] G. Boole. On the theory of probabilities. *Phil. Trans. R. Soc. Lond.*, 152:225, 1862.

[BP05] Jonathan Barrett and Stefano Pironio. Popescu-Rohrlich correlations as a unit of nonlocality. *Phys. Rev. Lett*, 95:140401, 2005.

[BPT11] Christine Bachoc, Arnaud Pécher, and Alain Thiéry. On the theta number of powers of cycle graphs. 2011. Preprint at http://arxiv.org/abs/1103.0444.

[BSV12] Nicolas Brunner, James Sharam, and Tamás Vértesi. Testing the structure of multipartite entanglement with bell inequalities. *Phys. Rev. Lett.*, 108:110501, Mar 2012.

[Cab08] Adán Cabello. Experimentally testable state-independent quantum contextuality. *Phys. Rev. Lett.*, 101(21):210401, Nov 2008.

[Cab13] Adán Cabello. Simple explanation of the quantum violation of a fundamental inequality. *Phys. Rev. Lett.*, 110:060402, Jan 2013.

[CASA11] D. Cavalcanti, M. L. Almeida, V. Scarani, and A. Acín. Quantum networks reveal quantum nonlocality. *Nat. Commun.*, 2:184, February 2011.

[CCK13] D. E. Chang, J. I. Cirac, and H. J. Kimble. Self-organization of atoms along a nanophotonic waveguide. *Phys. Rev. Lett.*, 110:113606, Mar 2013.

[CDP11] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Informational derivation of quantum theory. *Phys. Rev. A*, 84:012311, 2011.

[CEGA96] Adán Cabello, José M. Estebaranz, and Guillermo García-Alcaine. Bell-Kochen-Specker theorem: a proof with 18 vectors. *Phys. Lett. A*, 212(4):183–187, 1996.

[CF12] Rafael Chaves and Tobias Fritz. Entropic approach to local realism and noncontextuality. *Phys. Rev. A*, 85(3):032113, Mar 2012.

[CG04] Daniel Collins and Nicolas Gisin. A relevant two qubit bell inequality inequivalent to the chsh inequality. *Journal of Physics A: Mathematical and General*, 37(5):1775, 2004.

[CGL$^+$02] Daniel Collins, Nicolas Gisin, Noah Linden, Serge Massar, and Sandu Popescu. Bell inequalities for arbitrarily high-dimensional systems. *Phys. Rev. Lett.*, 88:040404, Jan 2002.

[CGR03] B. Codenotti, I. Gerace, and G. Resta. Some remarks on the shannon capacity of odd cycles. *Ars Combinatoria*, 66:243–257, 2003.

[CH74] John F. Clauser and Michael A. Horne. Experimental consequences of objective local theories. *Phys. Rev. D*, 10:526–535, Jul 1974.

[CHSH69] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed Experiment to Test Local Hidden-Variable Theories. *Phys. Rev. Lett.*, 23(15):880–884, Oct 1969.

[CMW00] Bob Coecke, David Moore, and Alexander Wilce. *Operational quantum logic: an overview*, volume 111 of *Fund. Theories Phys.* Kluwer Acad. Publ., Dordrecht, 2000.

[CRST06] Maria Chudnovsky, Neil Robertson, Paul Seymour, and Robin Thomas. The strong perfect graph theorem. *Ann. of Math. (2)*, 164(1):51–229, 2006.

[CSW10] A. Cabello, S. Severini, and A. Winter. (Non-)Contextuality of Physical Theories as an Axiom. October 2010. Preprint at http://arXiv.org/abs/1010.2163.

[DB11] Borivoje Dakic and Caslav Brukner. Quantum Theory and Beyond: Is Entanglement Special? 2011. in *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, Ed. H. Halvorson (Cambridge University Press, Cambridge, UK), pp 365-392.

[Dic54] R. H. Dicke. Coherence in Spontaneous Radiation Processes. *Phys. Rev.*, 93:99 – 110, 1954.

[Eit94] Thomas Eiter. Exact transversal hypergraphs and application to Boolean μ-functions. *J. Symbolic Comput.*, 17(3):215–225, 1994.

[EPR35] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, May 1935.

[ERIR+08] Kai Eckert, Oriol Romero-Isart, Mirta Rodriguez, Maciej Lewenstein, Eugene S. Polzik, and Anna Sanpera. Quantum nondemolition detection of strongly correlated systems. *Nature Physics*, 4:50, 2008.

[ESBL02] K. Eckert, J. Schliemann, D. Bruss, and M. Lewenstein. Quantum Correlations in Systems of Indistinguishable Particles. *Annals of Physics*, 299:88 – 127, 2002.

[FC72] Stuart J. Freedman and John F. Clauser. Experimental test of local hidden-variable theories. *Phys. Rev. Lett.*, 28:938–941, Apr 1972.

[FC13] Tobias Fritz and Rafael Chaves. Entropic inequalities and the marginal problem. *IEEE Trans. on Information Theory*, pages 803 – 817, 2013.

[Fek23] Michael Fekete. Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. *Math. Z.*, 17(1):228–249, 1923.

[Fin82] Arthur Fine. Hidden variables, joint probability, and the Bell inequalities. *Phys. Rev. Lett.*, 48(5):291–295, 1982.

[FLS12] T. Fritz, A. Leverrier, and A.B. Sainz. A Combinatorial Approach to Nonlocality and Contextuality. Dic 2012. Preprint at http://arxiv.org/abs/1212.4084.

[FR81] D. J. Foulis and C. H. Randall. Empirical logic and tensor products. 5:9–20, 1981.

[Fri12] Tobias Fritz. Tsirelson's problem and kirchberg's conjecture. *Rev. Math. Phys.*, 24:1250012, 2012.

[Fuk97] Komei Fukuda. cdd/cdd+ reference manual. ETH Zurich, Switzerland, 1997.

[GJDKL13] Tobias Graß, Bruno Juliá-Díaz, Marek Kuś, and Maciej Lewenstein. Quantum Chaos in SU(3) Models with Trapped Ions. *Phys. Rev. Lett.*, 111:090404, Aug 2013.

[GL04] Mordecai J. Golin and Yiu Cho Leung. Unhooking circulant graphs: a combinatorial method for counting spanning trees and other parameters. In *Graph-theoretic concepts in computer science*, volume 3353 of *Lecture Notes in Comput. Sci.*, pages 296–307. Springer, Berlin, 2004.

[GWAN11] R. Gallego, L. E. Würflinger, A. Acín, and M. Navascués. Quantum Correlations Require Multipartite Information Principles. *Phys. Rev. Lett.*, 107:210403, Nov 2011.

[Hae79] Willem Haemers. On some problems of Lovász concerning the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25(2):231–232, 1979.

[Hae81] Willem Haemers. An upper bound for the Shannon capacity of a graph. In *Algebraic methods in graph theory, Vol. I, II (Szeged, 1978)*, volume 25 of *Colloq. Math. Soc. János Bolyai*, pages 267–272. North-Holland, Amsterdam, 1981.

[Har01] Lucien Hardy. Quantum theory from five reasonable axioms. 2001. Preprint at http://arxiv.org/abs/quant-ph/0101012.

[Har11] Lucien Hardy. Reformulating and reconstructing quantum theory. 2011. Preprint at http://arxiv.org/abs/1104.2066.

[Hen12] Joe Henson. Quantum contextuality from a simple principle? Oct 2012. Preprint at http://arXiv.org/abs/1210.5978.

[HHR+05] H. Häffner, W. Hänsel, C. F. Roos, J. Benhelm, D. Chek-al kar, M. Chwalla, T. Körber, U. D. Rapol, M. Riebe, P. O. Schmidt, C. Becher, O. Gühne, W. Dür, and R. Blatt. Scalable multiparticle entanglement of trapped ions. *Nature*, 438:643 – 646, 2005.

[HSP10] Klemens Hammerer, Anders S. Sørensen, and Eugene S. Polzik. Quantum interface between light and atomic ensembles. *Rev. Mod. Phys.*, 82:1041–1093, Apr 2010.

[IK00] Wilfriedand Imrich and Sandi Klavžar. *Product graphs: structure and recognition.* Wiley-Intersciene, New York, 2000.

[JM05] Nick S. Jones and Lluis Masanes. Interconversion of Nonlocal Correlations. *Phys. Rev. A*, 72:052312, 2005.

[JNP+11] Marius Junge, Miguel Navascués, Carlos Palazuelos, David Pérez-García, Volkher B. Scholz, and Reinhard F. Werner. Connes embedding problem and Tsirelson's problem. *J. Math. Phys.*, 52(1):012102, 12, 2011.

[Kar72] R. M. Karp. Reducibility Among Combinatorial Problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. New York: Plenum, 1972.

[KCB08] A. A. Klyachko, M. A. Can, and A. S. Binicioglu, S. andShumovsky. A simple test for hidden variables in spin-1 system. *Phys. Rev. Lett.*, 101:020403–020406, 2008.

[Knu94] Donald E Knuth. The sandwich theorem. *The Electronic Journal of Combinatorics*, 1(1):A1, 1994.

[KR83] Richard V. Kadison and John R. Ringrose. *Fundamentals of the theory of operator algebras. Vol. I*, volume 100 of *Pure and Applied Mathematics*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983. Elementary theory.

[KS67] Simon Kochen and E. P. Specker. The problem of hidden variables in quantum mechanics. *J. Math. Mech.*, 17:59–87, 1967.

[KST$^+$07] N. Kiesel, C. Schmid, G. Tóth, E. Solano, and H. Weinfurter. Experimental Observation of Four-Photon Entangled Dicke State with High Fidelity. *Phys. Rev. Lett.*, 98:063604, 2007.

[LBPC13] Petr Lisonek, Piotr Badziag, Jose R. Portillo, and Adan Cabello. The simplest Kochen-Specker set. 2013. Preprint at http://arxiv.org/abs/1308.6012.

[LMG65] H. J. Lipkin, N. Meshkov, and A. J. Glick. Validity of many-body approximation methods for a solvable model: (I). Exact solutions and perturbation theory. *Nuclear Physics*, 62:188 – 198, 1965.

[LORV05] José I. Latorre, Román Orús, Enrique Rico, and Julien Vidal. Entanglement entropy in the Lipkin-Meshkov-Glick model. *Phys. Rev. A*, 71:064101, 2005.

[Lov72] László Lovász. Normal hypergraphs and the perfect graph conjecture. *Discrete Math.*, 2(3):253–267, 1972.

[Lov79] László Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25(1):1–7, 1979.

[LPZB04] Wiesław Laskowski, Tomasz Paterek, Marek Żukowski, and Časlav Brukner. Tight multipartite Bell's inequalities involving many measurement settings. *Phys. Rev. Lett.*, 93:200401, Nov 2004.

[LSW11] Yeong-Cherng Liang, Robert W. Spekkens, and Howard M. Wiseman. Specker's parable of the overprotective seer: a road to contextuality, nonlocality and complementarity. *Phys. Rep.*, 506(1-2):1–39, 2011.

[Lud85] G. Ludwig. *Foundations of Quantum Mechanics*, volume I and II. Springer-Verlag, New York, 1983 and 1985.

[Mac63] George W. Mackey. *The mathematical foundations of quantum mechanics*. W. A. Benjamin Inc., New York, 1963.

[Mas05] Lluis Masanes. Extremal quantum correlations for N parties with two dichotomic observables per site. 2005. Preprint at http://arxiv.org/abs/quant-ph/0512100.

[MM11] Lluis Masanes and Markus P. Mueller. A derivation of quantum theory from physical requirements. *New J.Phys*, 13:063001, 2011.

[MMM+08] D. N. Matsukevich, P. Maunz, D. L. Moehring, S. Olmschenk, and C. Monroe. Bell inequality violation with two remote atomic qubits. *Phys. Rev. Lett.*, 100:150404, Apr 2008.

[Nav13] Miguel Navascués, 2013. Private communication.

[NC03] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, The Edinburgh Building, Cambridge CB2 8RU, UK, 2003.

[NGHA13] Miguel Navascués, Yelena Guryanova, Matty J. Hoban, and Antonio Acín. 2013. In preparation.

[NO10] S. Niskanen and P. Östergard. Cliquer, 2010. version 1.21.

[NPA07] Miguel Navascués, Stefano Pironio, and Antonio Acín. Bounding the set of quantum correlations. *Phys. Rev. Lett.*, 98(1):010401, 2007.

[NPA08] Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008.

[NW10] M. Navascués and H. Wunderlich. A glance beyond the quantum model. *Proc. R. Soc. A*, 466(2115):881–890, 2010.

[PAM+10] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe. Random numbers certified by bell's theorem. *Nature*, 464:1021, 2010.

[PBS11] S. Pironio, J.-D. Bancal, and V. Scarani. Extremal correlations of the tripartite no-signaling polytope. *J. Phys. A: Math. and Theor.*, 44(6):065303, 2011. Data available at homepages.ulb.ac.be/∼spironio/tripartite_boxes/.

[PC04] D. Porras and J. I. Cirac. Effective quantum spin systems with trapped ions. *Phys. Rev. Lett.*, 92:207901, May 2004.

[PNA10] Stefano Pironio, Miguel Navascués, and Antonio Acín. Convergent relaxations of polynomial optimization problems with noncommuting variables. *SIAM Journal on Optimization*, 20(5):2157–2180, 2010.

[PPK+09] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski. Information causality as a physical principle. *Nature*, 461(7267):1101–1104, October 2009.

[PR94] S. Popescu and D. Rohrlich. Quantum Nonlocality as an Axiom. *Found. Phys.*, 24:379, 1994.

[PV09] Károly F. Pál and Tamás Vértesi. Quantum bounds on bell inequalities. *Phys. Rev. A*, 79:022120, Feb 2009.

[RKM+01] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland. Experimental violation of a Bell's inequality with efficient detection. *Nature*, 409:791, 2001.

[Ś03] Cezary Śliwa. Symmetries of the bell correlation inequalities. *Physics Letters A*, 317(3–4):165 – 168, 2003.

[Sac11] S. Sachdev. *Quantum phase transitions.* Cambridge University Press, Cambridge, second ed. edition, 2011.

[SBBC11] M. Sadiq, P. Badziag, M. Bourennane, and A. Cabello. The most elementary Bell inequalities. Jun 2011. Preprint at http://arxiv.org/abs/1106.4754.

[Sch03] Alexander Schrijver. *Combinatorial optimization. Polyhedra and efficiency.*, volume 24 of *Algorithms and Combinatorics.* Springer-Verlag, Berlin, 2003.

[SFA+13] A. B. Sainz, T. Fritz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier, and A. Acín. Exploring the Local Orthogonality Principle. 2013. Preprint at http://arxiv.org/abs/1311.6699.

[Sha56] Claude E. Shannon. The zero error capacity of a noisy channel. *Institute of Radio Engineers, Transactions on Information Theory,*, IT-2(September):8–19, 1956.

[Spe05] Robert W. Spekkens. Contextuality for preparations, transformations, and unsharp measurements. *Phys. Rev. A*, 71(5):052108, 2005.

[Tsi80] Boris Tsirelson. Quantum generalizations of bell's inequality. *Lett. Math. Phys.*, 4:93–100, 1980.

[Tsi93] B.S. Tsirelson. Some results and problems on quantum Bell-type inequalities. *Hadronic Journal Supplement*, 8:329–345, 1993.

[TV06] Benjamin Toner and Frank Verstraete. Monogamy of Bell correlations and Tsirelson's bound. 2006. Preprint at http://arxiv.org/abs/quant-ph/0611001.

[Uno05] Takeaki Uno. Mace: Maximal clique enumerator, 2005. version 2.0.

[vD00] W. van Dam. *Nonlocality and Communication Complexity.* PhD thesis, University of Oxford, 2000. Preprint at http://arxiv.org/abs/quant-ph/0501159 (2005).

[vN55] John von Neumann. *Mathematical foundations of quantum mechanics.* Princeton University Press, Princeton, 1955.

[VV12a] Umesh Vazirani and Thomas Vidick. Certifiable quantum dice: or, true random number generation secure against quantum adversaries. In *Proceedings of the 44th symposium on Theory of Computing*, pages 61–76. ACM, 2012.

[VV12b] Umesh Vazirani and Thomas Vidick. Fully device independent quantum key distribution. 2012. Preprint at http://arxiv.org/abs/1210.1810.

[WBA+12] Lars Erik Würflinger, Jean-Daniel Bancal, Antonio Acín, Nicolas Gisin, and Tamás Vértesi. Nonlocal multipartite correlations from local marginal probabilities. *Phys. Rev. A*, 86:032117, Sep 2012.

[Wil09] Alexander Wilce. *Test spaces.* Elsevier/North-Holland, Amsterdam, 2009.

[WM02] X. Wang and K. Mølmer. Pairwise entanglement in symmetric multi-qubit systems. *Eur. Phys. J. D*, 18:385 – 391, 2002.

[WNZ12] Marcin Wieśniak, Mohamed Nawareg, and Marek Żukowski. n-particle nonclassicality without n-particle correlations. *Phys. Rev. A*, 86:042339, Oct 2012.

[WW01] R. F. Werner and M. M. Wolf. All-multipartite bell-correlation inequalities for two dichotomic observables per site. *Phys. Rev. A*, 64:032112, Aug 2001.

[Yan13] Bin Yan. Quantum correlations are tight bounded by exclusivity principle. *Phys. Rev. Lett.*, 110:260406, June 2013.

[YCA$^+$12] T. H. Yang, D. Cavalcanti, M. L. Almeida, C. Teo, and V. Scarani. Information-causality and extremal tripartite correlations. *New J. Phys.*, 14(1):013061, 2012.

[ZB02] Marek Żukowski and Časlav Brukner. Bell's theorem for general N -qubit states. *Phys. Rev. Lett.*, 88:210401, May 2002.