Advanced quantum information: entanglement and nonlocality (part 2)

Alexander Streltsov, Jędrzej Kaniewski

June 15, 2022

Contents

1 Bell nonlocality

1.1 Preliminaries and notation

Throughout these notes all the Hilbert spaces, which we denote by $\mathcal{H}$, are assumed to be finite-dimensional unless specified otherwise. A pure quantum state is a vector $|\psi\rangle \in \mathcal{H}$ satisfying $\langle\psi|\psi\rangle = 1$. A mixed quantum state ρ is a linear, Hermitian and positive semidefinite operator acting on $\mathcal{H}$ which satisfies $\operatorname{Tr}\rho = 1$. A measurement with n outcomes is described by a set of n linear operators $\{F_j\}_{j=1}^n$ acting on $\mathcal{H}$, which are Hermitian, positive semidefinite and satisfy the normalisation condition

$$\sum_{j=1}^{n} F_j = \mathbb{1}. \tag{1.1}$$

A bipartite pure state on two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ is a vector $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$. A bipartite mixed state ρ_{AB} is a linear operator acting on $\mathcal{H}_A \otimes \mathcal{H}_B$ satisfying the conditions listed above.

1.2 Historical introduction

In the first part of the class we have focused on entanglement, which is an inherently quantum property. Therefore, one cannot use it to compare quantum mechanics against other, alternative physical theories. In this part we will talk about correlations between space-like separated devices. This field is usually referred to as Bell nonlocality after John S. Bell who was the first to give a formal description of this setup. However, these ideas can be traced back to the Einstein–Podolsky–Rosen (EPR) paradox first discussed in their famous 1935 paper.

The original EPR paradox considers the position and momentum of a quantum particle, but for our purposes we follow the variant proposed by Bohm which uses a two-level system, e.g. a spin-1/2 particle. Consider two spins in a maximally entangled state:

$$|\Phi_+\rangle_{AB} := \frac{1}{\sqrt{2}}\Big(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\Big). \tag{1.2}$$

The A spin is controlled by Alice, while the B spin is controlled by Bob and suppose that Alice and Bob are really far away. Now suppose that Alice performs a measurement in the standard basis $\{|0\rangle, |1\rangle\}$. If she obtains outcome "0", the state of Bob is given by $|0\rangle$, while if she obtains "1" his state is given by $|1\rangle$. Therefore, if Bob performs a measurement in the standard basis Alice can perfectly predict his outcomes. However, if he performs a measurement in the Hadamard basis $\{|+\rangle, |-\rangle\}$, Alice cannot predict his outcome. Clearly, the situation is reversed if Alice performs a measurement in the Hadamard basis instead.

The authors postulate that if a certain quantity can be predicted with certainty, there should exist an **element of reality** associated with it. Moreover, these elements of reality should be **local**, i.e. they should not be influenced by actions performed far away. While it is not exactly clear what these elements of reality should be, they are intended to capture some notion of objectivity, something that is independent of the observer or the state of anyone's knowledge. This postulate applied to the observations above implies that the spin of Bob should contain a separate element of reality for each possible measurement. However, this is not possible in the quantum formalism, which leads the authors to conclude that the quantum description should not be considered complete.

The postulate stated above, which according to Einstein, Podolsky and Rosen every "decent" physical theory should satisfy, is now known as the assumption of **local realism**. The reality part requires that objects have properties which can be assigned objective values regardless of whether a measurement is performed or not. In other words, performing a measurement simply reveals a pre-existing value. The locality part requires that these properties should be localised and should not be instantaneously affected by events happening somewhere else.

The final conclusion of the EPR paper is that the quantum-mechanical description should not be considered complete. Nowadays we interpret it differently: we simply admit that some of the predictions of quantum mechanics do not agree with our everyday intuition, i.e. that it is qualitatively different than all the pre-quantum physics.

This statement was formalised and proved in the seminal paper of John S. Bell published in 1964. He considered the scenario of two isolated parties, he formalised the notion of local realism and he showed that quantum mechanics indeed does not admit a local-realistic description. The study of such scenarios is now referred to as **Bell nonlocality** and constitutes the main topic of this course.

1.3 Local, quantum and no-signalling sets of correlations

In the simplest Bell scenario we consider a pair of devices controlled by two parties, which we will refer to as Alice and Bob. These devices could have interacted in the past, so they can be correlated, but they are not allowed to communicate during the Bell experiment (we will later formalise what we mean by this). Each device has a number of buttons which correspond to different measurements it can perform. Once a button is pressed the device produces an outcome. Note that the interaction with the devices is purely classical: we press a classical button and receive a classical outcome. What might not be classical is what happens inside the device, but we can only probe it through classical interaction.

The breakthrough discovery of Bell is that in such a simple setup we can conclusively distinguish between classical and quantum devices. More generally, we can think that the Bell setup allows us to make a fair comparison between different physical theories.

Suppose that Alice and Bob can choose one out of k measurements and let us denote the **measurement settings** of Alice and Bob by x and y, respectively. Each measurement produces **outcomes** from the set $[n] := \{1, 2, \ldots, n\}$ and let us denote the outcomes of Alice and Bob by a and b, respectively. Suppose, moreover, that they can repeat the experiment multiple times and that they are guaranteed that the devices will always behave in the same manner. This means that for every pair of measurement settings (x, y) we have a well-defined probability distribution over pairs of outcomes (a, b). Moreover, given enough statistics Alice and Bob can estimate it to arbitrary precision. We will denote this probability distribution by $P(ab|xy)$.[1] For a fixed pair of settings (x, y) we have a probability over n^2 outcomes which can be interpreted as a real vector with n^2 components. Since there are k^2 pairs of settings we can think of the entire statistics as a real vector of dimension n^2k^2. For convenience we will sometimes write $P \equiv \{P(ab|xy)\}_{abxy} \in \mathbb{R}^{n^2k^2}$ and call it a **probability point**.

Let us start with the case of classical devices. Consider a class of strategies which consists of a probability distribution $q(\lambda)$ over some finite[2] set L and two response functions:

$$\begin{aligned} r_A(a|x,\lambda) &: [n] \times [k] \times L \to \mathbb{R}_+, \\ r_B(b|y,\lambda) &: [n] \times [k] \times L \to \mathbb{R}_+ \end{aligned}$$

satisfying

$$\sum_{a=1}^{n} r_A(a|x,\lambda) = \sum_{b=1}^{n} r_B(b|y,\lambda) = 1$$

for all x, y, λ. Note that for fixed x and λ the response function $r_A(a|x,\lambda)$ is simply a probability distribution over $[n]$ and so is $r_B(b|y,\lambda)$ for fixed y and λ. Suppose that the classical devices function in the following manner: (a) before the Bell experiment the devices draw from the probability distribution λ and store the value, (b) during the experiment the outcomes are generated locally by the response functions based on the

[1] While this object is sometimes referred to as the "conditional probability distribution" one should not think of it as a conditional probability in the sense of probability theory since a priori there is no need to specify a probability distribution over inputs (x, y).

[2] Later we will see that allowing L to be infinite results in the same correlation set.

random variable λ and the local measurement setting only. The resulting statistics are given by:

$$P(ab|xy) = \sum_{\lambda \in L} q(\lambda)\, r_A(a|x,\lambda)\, r_B(b|y,\lambda). \tag{1.3}$$

If a probability point admits a description of the form given in Eq. (1.3) we say that it belongs to the **set of local-realistic correlations**, or the **local set** for short, denoted by $\mathcal{L}$. The variable λ is sometimes referred to as a **local hidden variable (LHV)** and the resulting decomposition as an **LHV model**.

Note that if we skipped the sum over λ in Eq. (1.3) we would obtain probability distributions that factorise between Alice and Bob:

$$P(ab|xy) = r_A(a|x)\, r_B(b|y). \tag{1.4}$$

In such distributions there are no correlations between Alice and Bob. Allowing for a sum over λ corresponds to taking convex combinations of such product distributions. Note that this is in exact analogy to the separable states which are defined as convex combinations of product states.

It should be clear that the definition above encompasses everything that classical devices are capable of. The notion of local realism discussed before, however, is phrased in a slightly different manner: it requires that all properties should simultaneously have well-defined values. In other words, we should be able to write down a joint probability distribution that contains the statistics of all possible measurements. Fortunately, it is not hard to prove that the two statements are equivalent, which is sometimes referred to as **Fine's theorem**. If our statistics can be written in the form given in Eq. (1.3), then a joint probability distribution is given by

$$P(a_1 a_2 \dots a_k b_1 b_2 \dots b_k) = \sum_{\lambda \in L} q(\lambda) \prod_{j=1}^{k} \Big[r_A(a_j|j,\lambda)\, r_B(b_j|j,\lambda) \Big]. \tag{1.5}$$

To see that given a joint probability distribution one can construct an LHV model note that we can simply take the hidden variable λ to contain all the variables, i.e. $\lambda = (a_1 a_2 \dots a_k b_1 b_2 \dots b_k)$. Then, the response functions simply pick out the right component of λ.

Having discussed the classical case let us move on to quantum devices. In the most general case these two devices will share a quantum state which we denote by ρ_{AB}. The measurement setting x on Alice's side corresponds to measurement operators $\{P_a^x\}_{a=1}^n$. The measurement setting y on Bob's side corresponds to measurement operators $\{Q_b^y\}_{b=1}^n$. Then, the Born rule tells us that

$$P(ab|xy) = \mathrm{Tr}(P_a^x \otimes Q_b^y \rho_{AB}). \tag{1.6}$$

The triple $\left\{\rho_{AB}, \{P_a^x\}, \{Q_b^y\}\right\}$ is often referred to as the **quantum realisation**. Let $\mathcal{Q}_{\text{fin}}$ be the set of correlations attainable by finite-dimensional realisations, i.e. a probability point P belongs to $\mathcal{Q}_{\text{fin}}$ if there exists a finite-dimensional quantum state and local measurements such that Eq. (1.6) holds. We then define the quantum set $\mathcal{Q}$ as the closure of $\mathcal{Q}_{\text{fin}}$, i.e. $\mathcal{Q}$ contains all the probability points which can be approximated arbitrarily well by finite-dimensional quantum realisations. Note that the quantum set is defined for a particular Bell scenario identified by the number of settings and outcomes but this is completely independent of the dimension of the quantum realisation.

It should not come as a surprise that quantum devices are at least as powerful as classical devices. To show this let us give an explicit construction that turns an LHV description of the form given in Eq. (1.3) into a particular quantum realisation. Let $d := |L|$ and let $\{|e_\lambda\rangle\}_{\lambda \in L}$ be an orthonormal basis on $\mathbb{C}^d$. Consider a quantum realisation acting $\mathbb{C}^d \otimes \mathbb{C}^d$ specified by:

$$\rho_{AB} := \sum_{\lambda \in L} q(\lambda)\, |e_\lambda\rangle\langle e_\lambda| \otimes |e_\lambda\rangle\langle e_\lambda|, \tag{1.7}$$

$$P_a^x := \sum_{\lambda \in L} r_A(a|x,\lambda)\, |e_\lambda\rangle\langle e_\lambda|, \tag{1.8}$$

$$Q_b^y := \sum_{\lambda \in L} r_B(b|y,\lambda)\, |e_\lambda\rangle\langle e_\lambda|. \tag{1.9}$$

It is easy to check that this indeed a valid quantum realisation and that it reproduces the statistics of the original LHV model. Hence, it immediately implies that $\mathcal{L} \subseteq \mathcal{Q}$.

In the description above we have allowed Alice and Bob to share a mixed quantum state and perform arbitrary measurements. However, it is clear that every mixed state can be purified, the purifying system can be given to one of the parties who can then ignore it in the measurement process. Therefore, the same statistics can be observed by measuring a pure state (although of a potentially larger dimension). More specifically, if $|\Psi\rangle_{ABB'}$ is a purification of ρ_{AB}, then we would say that both subsystems B and B' are in Bob's possession and that his new measurements are given by $Q_b^y \otimes \mathbb{1}$. Similarly, one can use Naimark's dilation to argue that we can without loss of generality assume that the measurements of Alice and Bob are projective. These two simplifications are often useful as they reduce the set of quantum realisations that we must consider.

An important feature of both the local and the quantum set is that they obey the **no-signalling conditions**:

$$\begin{aligned} \sum_b P(ab|xy) &= \sum_b P(ab|xy') \quad \text{for all} \quad a \in [n] \quad \text{and} \quad y, y' \in [k], \\ \sum_a P(ab|xy) &= \sum_a P(ab|x'y) \quad \text{for all} \quad b \in [n] \quad \text{and} \quad x, x' \in [k]. \end{aligned} \tag{1.10}$$

These conditions imply that the local distribution of outcomes on Alice's side does not depend on the setting chosen by Bob and vice versa. Clearly, this is necessary if we want to claim that Alice and Bob cannot communicate. It also implies that it is meaningful to talk about the local distribution of outcomes defined as

$$P(a|x) := \sum_{b} P(ab|xy), \tag{1.11}$$

$$P(b|y) := \sum_{a} P(ab|xy). \tag{1.12}$$

The fact that the two devices cannot signal to each other is one of the main assumptions of the Bell scenario. Hence, one could argue that any theory that we want to analyse in this framework must satisfy no-signalling. A logical next step is to ask: what about a theory in which no-signalling is the only restriction we impose on the probabilities? This surprisingly simple idea leads to the **no-signalling set of correlations**, which we denote by $\mathcal{NS}$. A probability point belongs to the no-signalling set if $P(ab|xy)$ corresponds to valid probability distributions, i.e.

$$P(ab|xy) \geq 0, \tag{1.13}$$

$$\sum_{ab} P(ab|xy) = 1 \quad \text{for all} \quad x, y, \tag{1.14}$$

and moreover it satisfies the no-signalling conditions given in Eq. (1.10).

So far we have defined three correlation sets: the local set, the quantum set and the no-signalling set. We have shown that the following inclusions hold:

$$\mathcal{L} \subseteq \mathcal{Q} \subseteq \mathcal{NS}, \tag{1.15}$$

and we will see that both of them are strict. The local and quantum sets capture what can be achieved when we restrict ourselves to classical and quantum systems, respectively. The no-signalling set can be seen as the largest set which is still consistent with the spirit of a Bell experiment. Alternatively, it can be seen as an outer approximation of the local or quantum set which admits a simple form.

Before moving on to a more detailed analysis let us make a brief comment on the foundational inconsistency between the concept of local realism and quantum mechanics. Recall that in a local-realistic theory we can interpret the measurement as simply revealing some pre-existing value. This should be contrasted with quantum mechanics in which the measurement outcome only comes into existence as a consequence of the measurement. Since in a local-realistic theory the measurement is a passive process, we can in principle perform an arbitrary number of measurements one after another (note that the order does not influence the observed statistics), which allows us to define a joint probability distribution as required by Fine's theorem. In quantum mechanics performing a measurement affects the state of the system, so afterwards we no longer have the original state. While we might be able to perform some measurement on the resulting state, it is not the same as performing it on the original state. This leads to the concept of incompatible measurements, i.e. measurements which cannot be performed simultaneously (on a single copy of the system) with the typical example being position and momentum of a quantum particle. It should not come as a surprise that one must use incompatible measurements to generate nonlocal correlations.

1.4 Basic properties of the three correlation sets

To continue our discussion of the correlation sets we need to introduce some basic concepts from convex geometry. Let $\mathcal{S}$ be a subset of $\mathbb{R}^n$. We say that $\mathcal{S}$ is **convex** if

$$x, y \in \mathcal{S} \implies px + (1 - p)y \in \mathcal{S} \tag{1.16}$$

for any $p \in [0, 1]$. In other words, we require the set to be closed under convex combinations.

Given an arbitrary set $\mathcal{S}$ we can make it convex by explicitly adding all possible convex combinations of points in $\mathcal{S}$. Such a procedure is known as taking the **convex hull** (or **convex envelope**) of $\mathcal{S}$ and we can think of it as finding the smallest convex set that contains $\mathcal{S}$.

We say that $z \in \mathcal{S}$ is an **extremal point** of $\mathcal{S}$ if the existence of $x, y \in \mathcal{S}$ such that $px + (1 - p)y = z$ for some $p \in (0, 1)$ implies that $x = y = z$. In other words, extremal points are those that do not admit a non-trivial convex decomposition.

Example. Consider the following convex subsets of $\mathbb{R}^2$:

$$\mathcal{S} := \{(x, y) \in \mathbb{R}^2 : |x| + |y| \leq 1\}, \tag{1.17}$$

$$\mathcal{T} := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}. \tag{1.18}$$

List the extremal points of $\mathcal{S}$ and $\mathcal{T}$. Note that these sets are simply unit balls in $\mathbb{R}^2$ according to the vector p-norm for $p = 1$ and $p = 2$, respectively.

What is important is that for convex sets which are compact[3] the knowledge of extremal points uniquely determines the set.

Krein–Millman's theorem. Every compact convex subset of a finite-dimensional vector space is equal to the convex hull of its extremal points.

An alternative statement of this theorem reads: every point of a compact convex set can be written as a convex combination of its extremal points.

Krein–Millman's theorem essentially tells us that knowing the extremal points allows us to reconstruct the entire set. In other words, if our goal is to understand some compact

[3] Recall that a subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

convex set, we may restrict our attention to its extremal points. To see that compactness is crucial here note that convex sets which are not compact are not even guaranteed to have any extremal points, e.g. consider $\mathcal{S} = \mathbb{R}$ or $\mathcal{S} = (0, 1) \subseteq \mathbb{R}$.

For our discussion it is convenient to interpret the correlation sets as subsets of $\mathbb{R}^{n^2k^2}$ since this enables us to use a number of standard tools. In the rest of this section we show that all three correlation sets are convex and compact and let us start with the former.

Consider two local-realistic points P_0 and P_1 which by definition can be written as

$$P_0(ab|xy) = \sum_{\lambda \in L_0} q_0(\lambda)\, r_{A,0}(a|x, \lambda)\, r_{B,0}(b|y, \lambda), \tag{1.19}$$

$$P_1(ab|xy) = \sum_{\lambda \in L_1} q_1(\lambda)\, r_{A,1}(a|x, \lambda)\, r_{B,1}(b|y, \lambda). \tag{1.20}$$

Since the actual values of the hidden variable λ do not matter (they merely serve as labels), we can without loss of generality assume that L_0 and L_1 are disjoint, i.e. $L_0 \cap L_1 = \emptyset$. It is easy to see that a convex combination $P = pP_0 + (1 - p)P_1$ can be written as

$$P(ab|xy) = \sum_{\lambda \in L} q(\lambda)\, r_A(a|x, \lambda)\, r_B(b|y, \lambda) \tag{1.21}$$

for

$$L := L_0 \cup L_1, \tag{1.22}$$

$$q(\lambda) := \begin{cases} pq_0(\lambda) & \text{if} \quad \lambda \in L_0, \\ (1-p)q_1(\lambda) & \text{if} \quad \lambda \in L_1, \end{cases} \tag{1.23}$$

$$r_A(a|x, \lambda) := \begin{cases} r_{A,0}(a|x, \lambda) & \text{if} \quad \lambda \in L_0, \\ r_{A,1}(a|x, \lambda) & \text{if} \quad \lambda \in L_1, \end{cases} \tag{1.24}$$

$$r_B(b|y, \lambda) := \begin{cases} r_{B,0}(b|y, \lambda) & \text{if} \quad \lambda \in L_0, \\ r_{B,1}(b|y, \lambda) & \text{if} \quad \lambda \in L_1. \end{cases} \tag{1.25}$$

To show that the quantum set $\mathcal{Q}$ is convex let us first prove the $\mathcal{Q}_{\text{fin}}$ is convex. Suppose we are given two quantum points from $\mathcal{Q}_{\text{fin}}$ along with their realisations

$$P_0 : [P_a^x]_0,\ [Q_b^y]_0,\ \rho_0, \tag{1.26}$$

$$P_1 : [P_a^x]_1,\ [Q_b^y]_1,\ \rho_1. \tag{1.27}$$

Suppose that the first realisation acts on $\mathcal{H}_{A_0} \otimes \mathcal{H}_{B_0}$, while the second realisation acts on $\mathcal{H}_{A_1} \otimes \mathcal{H}_{B_1}$. Since the two realisations can have distinct dimension and for us it is convenient to have realisations of the same dimension, let us first embed them in $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$, where

$$d_A := \max\{\dim(\mathcal{H}_{A_0}), \dim(\mathcal{H}_{A_1})\}, \tag{1.28}$$

$$d_B := \max\{\dim(\mathcal{H}_{B_0}), \dim(\mathcal{H}_{B_1})\}. \tag{1.29}$$

Since all the Hilbert spaces are finite-dimensional we are guaranteed that $d_A, d_B < \infty$. To simplify the notation we will not distinguish between the original realisations and the realisations embedded in $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$.

Consider a realisation that acts on $\mathcal{H}_{A_C} \otimes \mathcal{H}_{B_C} \otimes \mathcal{H}_{A_Q} \otimes \mathcal{H}_{B_Q}$, where $\dim(\mathcal{H}_{A_C}) = \dim(\mathcal{H}_{B_C}) = 2$, $\dim(\mathcal{H}_{A_Q}) = d_A$ and $\dim(\mathcal{H}_{B_Q}) = d_B$. The A registers belong to Alice, while the B registers belong to Bob. Consider a bipartite state of the form:

$$\rho_{AB} := p|0\rangle\langle 0| \otimes |0\rangle\langle 0| \otimes \rho_0 + (1-p)|1\rangle\langle 1| \otimes |1\rangle\langle 1| \otimes \rho_1, \tag{1.30}$$

while the measurements of Alice and Bob are given by[4]

$$P_a^x := |0\rangle\langle 0| \otimes [P_a^x]_0 + |1\rangle\langle 1| \otimes [P_a^x]_1, \tag{1.31}$$

$$Q_b^y := |0\rangle\langle 0| \otimes [Q_b^y]_0 + |1\rangle\langle 1| \otimes [Q_b^y]_1. \tag{1.32}$$

It is easy to verify that this leads to a convex combination of the original points given by $pP_0 + (1-p)P_1$. This construction necessarily increases the local dimensions of Alice and Bob, but this is not a problem since the definition of the quantum set does not put any restrictions on dimensionality. In fact, this feature turns out to be necessary: it is known that imposing a dimension bound in the definition of the quantum set might lead to non-convex sets.

This construction shows that $\mathcal{Q}_{\text{fin}}$ is a convex set and now it suffices to combine it with the standard result that the closure of a convex set is a convex set.

Convexity of $\mathcal{NS}$ is an immediate consequence of the linearity of all the conditions.

To show compactness recall that a subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded. To see that all three correlation sets are bounded note that they are contained in the unit ball according to the vector ∞-norm: $|P(ab|xy)| \leq 1$ for all a, b, x, y.

Since we defined $\mathcal{Q}$ to be the closure of $\mathcal{Q}_{\text{fin}}$, it is closed by definition. The no-signalling set is closed because it is an intersection of finitely many closed sets (every equality constraint can be replaced by two non-strict inequalities, $a = b \iff a \geq b \wedge a \leq b$, and every non-strict inequality defines a closed set). The local set is also closed, but to see it we need some extra understanding, which we develop in the next section.

[4] Since the operators of Alice and Bob act on $\mathcal{H}_{A_C} \otimes \mathcal{H}_{A_Q}$ and $\mathcal{H}_{B_C} \otimes \mathcal{H}_{B_Q}$, respectively, one has to be careful about the order of the tensor factors when computing the product of $P_a^x \otimes Q_b^y$ and ρ_{AB}.

1.5 The local and no-signalling sets as polytopes

A particularly simple class of compact convex sets consists of those that have only a finite number of extremal points. Such sets are known as **polytopes** and their extremal points are often referred to as **vertices**. Two out of three correlations sets defined above turn out to be polytopes.

To see that the local set is a polytope we have to realise that if a probability point can be written in the local form given in Eq. (1.3), it can also be written in the same form using only **deterministic** response functions, i.e. functions which only take values 0 and 1. This comes from the fact that we can "push back" all the randomness into the shared variable λ. In the worst case scenario we have to include an additional probability distribution over n outcomes for every measurement setting of Alice and Bob and for every value of λ. This leads to a larger but still finite hidden variable. The ability to restrict our attention to deterministic response functions is extremely convenient because there is only a finite number of them. More specifically, there are precisely n^k (distinct) functions from a set of size k to a set of size n. Therefore, we have n^k deterministic response functions for Alice and the same number for Bob, which gives rise to n^{2k} **deterministic points** in total. Each deterministic point is labelled by a pair of functions $f_A, f_B : [k] \to [n]$ and can be compactly written down as

$$P(ab|xy) = \begin{cases} 1 & \text{if} \quad a = f_A(x) \quad \text{and} \quad b = f_B(y), \\ 0 & \text{otherwise.} \end{cases} \tag{1.33}$$

The fact that every point in $\mathcal{L}$ can be written as a convex combination of points from a fixed finite set implies that $\mathcal{L}$ has a finite number of extremal points, i.e. it is a polytope. It is not hard to see that every deterministic point is extremal, i.e. the local polytope has precisely n^{2k} vertices.

The no-signalling set is defined as an intersection of a finite number of closed half-spaces. It is well-known in polytope theory that if a finite intersection of closed half-spaces yields a bounded set, then this set must be a polytope. As a consequence every polytope can be described in two equivalent but complementary ways: either by specifying its vertices or by specifying the half-spaces. It is worth pointing out that in both cases the minimal (shortest) descriptions turn out to be unique.

In our case the local set is naturally described by its extremal points (i.e. the deterministic points), while the no-signalling set is described by the half-spaces. It is therefore natural to try to compute the other description for each set. While there exist algorithms that allow us to go from one description to the other, they are not efficient which becomes a problem when the size of the problem grows. In the most general setting this problem has only been solved for the scenarios corresponding to $(n, k) = (2, 2), (2, 3)$ and $(3, 2)$.

If we compute the half-space description of the local set we will obtain three types of conditions. The first simply requires all the probabilities to be non-negative:

$$P(ab|xy) \geq 0. \tag{1.34}$$

Such conditions are known as **positivity facets** and they are not particularly interesting: clearly, they must be satisfied by any reasonable theory. The second type are the no-signalling conditions mentioned above (recall that equality $a = b$ is equivalent to the conjuction of $a \geq b$ and $b \geq a$). The remaining conditions are known as **facet Bell inequalities**. Since these constraints can (at least in principle) be violated by more general theories, we interpret them as limitations on the strength of correlations in local-realistic theories. Since any (no-signalling) probability point that does not belong to $\mathcal{L}$ must violate some facet Bell inequality, we often use the terms "generate nonlocal correlations" and "violate a Bell inequality" interchangeably.

Alternatively, if we compute a vertex description of the no-signalling set, we find that there are two types of vertices. First of all, we find all the vertices of the local set, i.e. the deterministic points. The remaining vertices are known as the **extremal no-signalling points**.

A brief summary of these observations can be found in the table below. What is important to remember is that the local set is smaller, so it has fewer extremal points (in the vertex description) and more linear constraints (in the half-space description). The no-signalling set is larger, so it has more extremal points and fewer constraints.

	local set	no-signalling set
vertex description	deterministic points	deterministic points extremal no-signalling points
half-space description	positivity facets no-signalling conditions facet Bell inequalities	positivity facets no-signalling conditions

1.6 The membership problem

Having defined the three sets of correlations it is natural to ask whether these definitions are convenient to work with. In other words, given a probability point P can we efficiently check whether it belongs to $\mathcal{L}$, $\mathcal{Q}$ or $\mathcal{NS}$?

Suppose we are given a probability point, which we already assume to be composed of valid probability distributions, i.e. non-negativity and normalisation conditions are satisfied. To check whether $P \in \mathcal{NS}$ it suffices to check the no-signalling conditions given in Eq. (1.10) and this can be done by evaluating $2nk^2$ linear equalities. This is an "easy" task because its complexity[5] scales polynomially with the input parameters n and k. Clearly, the half-space description is convenient for checking whether a point lies inside a set.

Checking whether $P \in \mathcal{L}$ is slightly harder. Recall that $\mathcal{L}$ can be defined through its extremal points (the deterministic vertices) so essentially we are asking whether there exists a convex combination of the deterministic vertices which yields precisely the target point P. If we denote the deterministic vertices by $\{D_j\}_{j=1}^N$, we are simply looking for a probability distribution $\{q_j\}_{j=1}^N$ such that

$$P = \sum_{j=1}^{N} q_j D_j. \tag{1.35}$$

This is an instance of an optimisation problem known as a **linear program (LP)** and such problems are considered "easy" from the complexity-theoretic point of view because they can be solved using resources polynomial in the size of the program. However, we observed earlier that $N = n^{2k}$, i.e. the size of our problem is already exponential in the number of settings. Therefore, solving such problems quickly becomes infeasible as k increases. An alternative approach would be to first find the half-space description of $\mathcal{L}$ and then check whether P satisfies all the linear inequalities. However, as mentioned before finding the half-space description of $\mathcal{L}$ is hard and even if we manage to do so, it might happen that the number of inequalities to check will be exponential.

Checking whether $P \in \mathcal{Q}$ turns out to be even harder and this is largely due to the fact that we know nothing about the dimension of the quantum realisations to consider. If we wanted to find out whether P has a quantum realisation where the systems of Alice and Bob have dimension d, we could in principle enumerate/parameterise all possible quantum realisations acting on $\mathbb{C}^d \otimes \mathbb{C}^d$. This is exponentially hard but in principle could be done (given access to unlimited computational power). However, the fact that we have to do this for all $d \geq 2$ renders the task completely infeasible. Since there exist probability points P which can only be achieved in the limit of $d \to \infty$, searching for finite-dimensional realisations does not necessarily help us in answering the original question. The situation is, however, not entirely hopeless and later we will discuss some tools which allow us to study the quantum set.

1.7 Basic notions of convex geometry

The local and no-signalling sets are polytopes which in particular implies that their geometry is rather simple. However, the quantum set is not a polytope and to describe its geometry we need some additional notions from convex geometry.

[5] In this context complexity can be thought of as the number of operations one must perform to accomplish a computation.

The **dimension of a convex set** is defined as the dimension of the smallest affine subspace which contains the entire set.[6] For instance a polytope in $\mathbb{R}^3$ whose 4 vertices are given by $(x, y, z) = (\pm 1, \pm 1, 1)$ is two-dimensional because it is contained in the two-dimensional affine subspace given by $z = 1$. Intuitively, the dimension of a convex set is the minimal number of real parameters required to uniquely identify every point in that set.

The concept of extremal points introduced in Section 1.4 admits an elegant generalisation. Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a convex set and $\mathcal{F}$ be a non-empty convex subset of $\mathcal{S}$. We say that $\mathcal{F}$ is a **face** of $\mathcal{S}$ if for every point $z \in \mathcal{F}$ the fact that $x, y \in \mathcal{S}$ satisfy $z = px + (1-p)y$ for some $p \in (0, 1)$ implies that $x, y \in \mathcal{F}$. In other words, points in $\mathcal{F}$ can only be convexly-decomposed into other points in $\mathcal{F}$. This captures the notion that the whole set $\mathcal{F}$ lies at the boundary of $\mathcal{S}$. Since every face of $\mathcal{S}$ is a convex set, its dimension is well-defined and we can group faces according to their dimension. Zero-dimensional faces, i.e. those composed of a single point, are precisely the extremal points of $\mathcal{S}$. The largest dimension of a face equals the dimension of $\mathcal{S}$, which as shown above can be strictly smaller than the dimension of the space in which $\mathcal{S}$ is embedded. It is not hard to see that the unique face of the maximal dimension is the set $\mathcal{S}$ itself. To exclude this possibility we often talk about **proper faces**, i.e. faces satisfying $\mathcal{F} \neq \mathcal{S}$.

Example. List all faces of a unit cube and a unit ball in $\mathbb{R}^3$ and order them by dimension.

Polytopes exhibit a particularly simple facial structure. For a polytope of dimension d we are guaranteed to find faces of every dimension from 0 to d. Vertices correspond to zero-dimensional faces, while the largest proper faces, i.e. faces of dimension $d - 1$, are known as **facets**. Facets are important because they appear in the **minimal half-space description** of the polytope, i.e. the description in terms of the smallest number of linear inequalities. Another concept that turns out to be useful in the study of convex sets is that of a (real) linear functional, i.e. a function $f : V \to \mathbb{R}$ which satisfies linearity. For our purposes it suffices to consider the case of $V = \mathbb{R}^n$ and then a linear functional is represented by a real vector $f = \{f_j\}_{j=1}^n \in \mathbb{R}^n$. Evaluating the functional on a particular point $x = \{x_j\}_{j=1}^n \in \mathbb{R}^n$ corresponds to simply taking the inner product: $\langle f, x \rangle = \sum_{j=1}^n f_j x_j$.

One reason why linear functionals are useful is the fact that they lead to a more detailed classification of points of a convex set. We have previously introduced the notion of an extremal point and let us now define two additional notions. Let $\mathcal{S} \in \mathbb{R}^n$ be a convex set. We say that $P \in \mathcal{S}$ is a **boundary point** if there exists a linear functional which takes distinct values on the points of $\mathcal{S}$ and which is maximised at P (the first condition is necessary to exclude trivial functionals which are simultaneously maximised by all points in $\mathcal{S}$). We say that P is an **exposed point** of $\mathcal{S}$ if there exists a linear functional which is maximised uniquely by P. While every exposed point is extremal, not every extremal point is exposed. Fig. 1.1 shows different types of points of a convex set.

[6] It is appropriate to consider affine rather than linear subspaces because the dimension should be invariant under shifts. Recall that the dimension of the affine space generated by vectors $\{x_1, x_2, \dots, x_n\}$ is precisely the dimension of the linear space generated by vectors $\{x_2 - x_1, x_3 - x_1, \dots, x_n - x_1\}$.

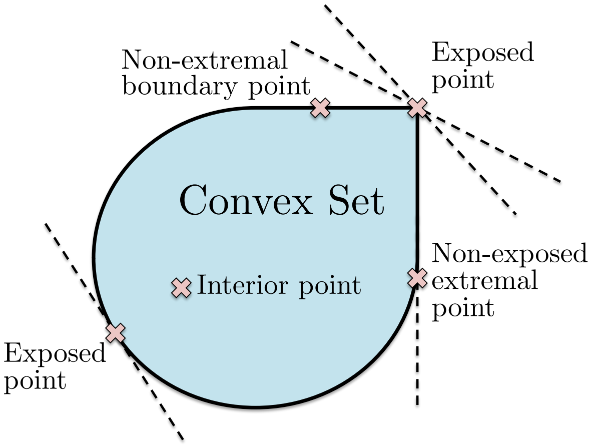

Figure 1.1: Different types of points of a convex set. (Figure taken from `arXiv:1710.05892` reproduced with the authors' permission.)

Another important application of linear functionals is the hyperplane separation theorem (in its more general variants also known as the Hahn–Banach theorem).

Hyperplane separation theorem. Let $\mathcal{S}$ be a closed convex set in $\mathbb{R}^n$ and let $x \in \mathbb{R}^n$ be a point outside of $\mathcal{S}$. Then, there exists a linear functional that separates x from $\mathcal{S}$, i.e. there exists $f \in \mathbb{R}^n$ and $c \in \mathbb{R}$ such that $\langle f, y\rangle \leq c$ for all $y \in \mathcal{S}$ but $\langle f, x\rangle > c$.

What this says is that whenever a point does not belong to a closed convex set this can be compactly demonstrated by finding a suitable linear functional. That is why we are often interested in computing the maximal value of a fixed functional over a set:

$$\beta := \sup_{y \in \mathcal{S}} \langle f, y\rangle. \tag{1.36}$$

If we then encounter a point x such that $\langle f, x\rangle > \beta$, we immediately deduce that $x \notin \mathcal{S}$.

A simple corollary of the hyperplane separation theorem is that every closed convex set can be described as an intersection (possibly infinite) of closed half-spaces.

Example. Give a description of the unit disc defined in Eq. (1.18) in terms of closed half-spaces.

We have earlier said that every polytope can be described in two complementary ways: either by extremal points or by half-spaces. We now see that this statement holds for every compact convex set. If at least one of the descriptions is finite (i.e. a finite number of extremal points or a finite intersection of half-spaces), then so is the other description and we are dealing with a polytope. On the other hand, if our convex set is not a polytope, no finite description (of either type) can be found. This is precisely why studying and understanding the quantum set, which is not a polytope, poses such a challenge.

Finally, let us briefly explain how for a polytope one can derive the halfspace description from the vertex description and vice versa.

Suppose we are given a polytope $\mathcal{P} \subseteq \mathbb{R}^n$ and for simplicity let us assume that its dimension equals n, i.e. it is full-dimensional. Given its vertices $\{V_j\}_{j=1}^m$ our goal is to find all its facets, i.e. hyperplanes of dimension $n-1$ which delimit the polytope. Every facet contains a certain number of vertices so to find all facets we can simply try every subset of the vertices. For every subset we have to see whether it defines a unique hyperplane and whether the entire set lies on one side of the hyperplane. While not every subset of vertices leads to a facet and the same facet can arise from distinct subsets, this already shows that there is only a finite number of facets. Similarly, if we are given a halfspace description of a polytope and we want to find the vertices we have to realise that vertices are points which saturate the maximal number of halfspace inequalities. Here again we can try all subsets of halfspace conditions and check which of them are saturated only by a single point. Then, we would check whether this point actually belongs to $\mathcal{P}$. In this way we are guaranteed to find all the vertices of $\mathcal{P}$.

The procedures explained above are clearly not optimal, but they show that these problems can be solved in a finite number of steps and lead to a finite solution.

1.8 Dimension of the correlation sets

In the previous section we have defined the dimension of a convex set. Let us now show that all three correlation sets have the same dimension given by:

$$D := 2(n-1)k + (n-1)^2k^2. \tag{1.37}$$

The argument consists of two parts. First, we show that any no-signalling point can be parametrised by D real numbers, which implies that $\dim(\mathcal{NS}) \leq D$. Then, we give an explicit choice of $D+1$ points from the local set and show that they are affinely independent, which allows us to conclude that $\dim(\mathcal{L}) \geq D$. Since $\mathcal{L} \subseteq \mathcal{Q} \subseteq \mathcal{NS}$, we deduce that $\dim(\mathcal{L}) = \dim(\mathcal{Q}) = \dim(\mathcal{NS}) = D$.

Let us first argue that given the following quantities:

$$P(a|x) \quad \text{for} \quad a \in [n-1], x \in [k], \tag{1.38}$$

$$P(b|y) \quad \text{for} \quad b \in [n-1], y \in [k], \tag{1.39}$$

$$P(ab|xy) \quad \text{for} \quad a, b \in [n-1], x, y \in [k] \tag{1.40}$$

we are able to reconstruct all the probabilities of a no-signalling point. The missing marginal terms can be computed from normalisation, e.g.:

$$\sum_{a=1}^{n} P(a|x) = 1 \implies P(n|x) = 1 - \sum_{a=1}^{n-1} P(a|x). \tag{1.41}$$

To compute $P(nb|xy)$ for $b \leq n-1$ or $P(an|xy)$ for $a \leq n-1$ we use the no-signalling condition, e.g.:

$$\sum_{a=1}^{n} P(ab|xy) = P(b|y) \implies P(nb|xy) = P(b|y) - \sum_{a=1}^{n-1} P(ab|xy). \tag{1.42}$$

Finally, to compute $P(nn|xy)$ we use the normalisation condition $\sum_{ab} P(ab|xy) = 1$. Note that Eqs. (1.38)–(1.40) specify precisely D real parameters, which implies that $\dim(\mathcal{NS}) \leq D$.

Let us now specify $D+1$ points from $\mathcal{L}$ which are affinely independent. It will be convenient to use the parametrisation introduced in Eqs. (1.38)–(1.40), i.e. we represent these local points as vectors in $\mathbb{R}^D$. The first point corresponds to the vector of all zeroes. Then, we have $2(n-1)k$ points which have a single 1 in a coordinate corresponding to one of the marginals (of either Alice or Bob). Finally, we have $(n-1)^2k^2$ points which have a 1 in exactly one coordinate corresponding to a correlator term and another two 1s in the matching marginal terms (this is required since $P(ab|xy) = 1 \implies P(a|x) = P(b|y) = 1$). Let us denote these points by x_j for $j = \{0, 1, \ldots, D\}$ and recall that these points are affinely independent if and only if the points $\{x_1 - x_0, x_2 - x_0, \ldots, x_D - x_0\}$ are linearly independent. Since $x_0 = 0$, this reduces to showing that vectors $\{x_1, x_2, \ldots, x_D\}$ are linearly independent and this is easily proved by inspection by first looking at the correlator coordinates and then the marginal coordinates.

1.9 Bell functionals and Bell inequalities

A **Bell functional** is a real linear functional acting on the space of probability points. We will represent it by a real vector $F = \{f_{abxy}\}_{abxy} \in \mathbb{R}^{n^2k^2}$ and the action of the functional on a probability point is given by

$$\langle F, P \rangle := \sum_{abxy} f_{abxy} P(ab|xy). \tag{1.43}$$

For every Bell functional F we define the **local**, **quantum** and **no-signalling value** as the largest value achieved by probability points in that set. More specifically, we have

$$\beta_L := \max_{P \in \mathcal{L}} \langle F, P \rangle, \tag{1.44}$$

$$\beta_Q := \max_{P \in \mathcal{Q}} \langle F, P \rangle, \tag{1.45}$$

$$\beta_{NS} := \max_{P \in \mathcal{NS}} \langle F, P \rangle, \tag{1.46}$$

By the trivial inclusions $\mathcal{L} \subseteq \mathcal{Q} \subseteq \mathcal{NS}$ we immediately deduce that $\beta_L \leq \beta_Q \leq \beta_{NS}$.

We are particularly interested in functionals for which $\beta_Q > \beta_L$ as they can be used as certificates of non-classicality: if we are given a probability point which satisfies $\langle F, P \rangle > \beta_L$ we immediately know it must be outside of the local set.[7] It is worth pointing out that in some texts any condition of the form $\langle F, P \rangle \leq \beta_L$ is referred to as a **Bell inequality**, but one should remember that this is different from the facet Bell inequalities which arise when analysing the local polytope.

Computing the no-signalling value of a Bell functional is easy as it can be cast as a linear program. To compute the local value note that the maximum of a linear function over a compact convex set is always achieved at some extremal point, i.e. it suffices to maximise over the deterministic vertices:

$$\beta_L = \max_{j \in [N]} \langle F, D_j \rangle. \tag{1.47}$$

However, as $N = n^{2k}$ this quickly becomes infeasible.

Not surprisingly computing the quantum value is even harder. If we restrict ourselves to a fixed dimension we could in principle enumerate all the quantum realisations as explained in Section 1.6. However, since now we actually have a linear functional to optimise, we can be slightly smarter and use the so-called **see-saw algorithm**. Recall that a quantum realisation consists of three components: the quantum state $|\psi\rangle$, the measurements of Alice $\{P_a^x\}$ and the measurements of Bob $\{Q_b^y\}$. The see-saw method is based on the observation that optimising one component can be done efficiently if the other two components are kept unchanged. For instance to find the optimal state for fixed measurements we first construct the **Bell operator**:

$$W = \sum_{abxy} f_{abxy} P_a^x \otimes Q_b^y. \tag{1.48}$$

It should now be clear that determining the optimal state is equivalent to computing the largest eigenvalue of the Bell operator and determining the corresponding eigenspace. Optimising over the measurements of one party is slightly harder, but turns out to be an instance of a **semidefinite program (SDP)**, a generalisation of linear programming

[7] This is analogous to the concept of entanglement witnesses discussed in the first part of the course.

which can still be solved efficiently. We can now alternate over optimising the three components until we reach a local maximum. This might not be the global maximum so to make the final result slightly more trustworthy we should repeat the entire procedure multiple times with randomly-generated starting points. However, even in simple scenarios (small number of settings, small number of outcomes and quantum realisations of low dimension) performing a truly exhaustive search is not possible. Hence, this method should only be used as a way of obtaining lower bounds on the quantum value or obtaining candidates for optimal realisations, while upper bounds must be obtained through a different approach.

Let us now briefly discuss one of the simplest methods to derive an upper bound on the quantum value a Bell functional. Our goal is to give an upper bound on $\langle W, \rho_{AB} \rangle$ which holds for all possible quantum realisations. It is clear that

$$\langle W, \rho_{AB} \rangle \leq \lambda_{\max}(W), \tag{1.49}$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a Hermitian operator. Therefore, our task reduces to bounding the spectrum of W from above, which can be conveniently written as an operator inequality. Given two Hermitian operators X, Y acting on the same Hilbert space we write $X \geq Y$ to mean $X - Y \geq 0$. It is clear that if one of the operators is proportional to the identity, this inequality is equivalent to a bound on the spectrum of the other operator. In our case the inequality $\lambda_{\max}(W) \leq \lambda$ for some $\lambda \in \mathbb{R}$ is equivalent to

$$W \leq \lambda \, \mathbb{1}. \tag{1.50}$$

Hence, our goal is to show that for all choices of measurements operators we have

$$\lambda \, \mathbb{1} - W \geq 0. \tag{1.51}$$

One way of proving that a real-valued function $f(x)$ is non-negative is to find a **sum-of-squares (SOS) decomposition**, i.e. a family of real-valued functions $\{p_j(x)\}_j$ such that

$$f(x) = \sum_j [p_j(x)]^2. \tag{1.52}$$

Analogously, to show that $\lambda \, \mathbb{1} - W \geq 0$ we will look for Hermitian operators $\{L_j\}_j$, which now have to depend on the measurement operators, such that

$$\lambda \, \mathbb{1} - W = \sum_j L_j^2. \tag{1.53}$$

Finding such a decomposition which is valid for all measurements operators of Alice and Bob implies that $\beta_Q \leq \lambda$. If in addition we find a quantum realisation which achieves this upper bound, we have proven that $\beta_Q = \lambda$.

So far we have dedicated our time to formalising the scenario and familiarising ourselves with some basic tools to tackle it. We should now be able to appreciate the complexity of the problem. Indeed, despite a large body of works dedicated to Bell nonlocality several important problems remain open. In the next section we will see that already the simplest non-trivial Bell scenario turns out to be quite complicated.

2 The CHSH scenario

2.1 Preliminaries

The simplest scenario in which the three correlation sets differ corresponds to two measurement setting and two outputs and let us refer to it as the **CHSH scenario**. Recall that the observed statistics are fully described by $P = \{P(ab|xy)\} \in \mathbb{R}^{16}$ and in the CHSH scenario it is customary to think of the settings and outcomes as bits, so let us for this section assume that $a, b, x, y \in \{0, 1\}$. The formula given in Section 1.8 implies that in this scenario all three correlation sets are 8-dimensional. While we could use the representation given in Section 1.8, there exists a more convenient parametrisation in terms of 8 real parameters. For $x, y \in \{0, 1\}$ define

$$\begin{aligned} \langle A_x \rangle &= P(a = 0|x) - P(a = 1|x), \\ \langle B_y \rangle &= P(b = 0|y) - P(b = 1|y), \\ \langle A_x B_y \rangle &= P(a = b|xy) - P(a \neq b|xy). \end{aligned} \tag{2.1}$$

The first two terms depend only on the marginal distributions (these are well-defined thanks to the no-signalling condition), while the last term captures the correlations between Alice and Bob. We will refer to the terms $\langle A_x \rangle$ and $\langle B_y \rangle$ as **marginals** and to the terms $\langle A_x B_y \rangle$ as **correlators**. Collectively we will refer to these variables as the **reduced coordinates**.

It is clear that all these numbers range from -1 to 1. To see that knowing these 8 real parameters allows us to reconstruct the entire distribution note that

$$P(ab|xy) = \frac{1}{4}\Big(1 + (-1)^a \langle A_x \rangle + (-1)^b \langle B_y \rangle + (-1)^{a+b} \langle A_x B_y \rangle\Big). \tag{2.2}$$

The transformation which takes us from probabilities to marginals and correlators is a linear transformation, which implies that geometric properties like extremality or exposedness remain unchanged. Note, however, that it is not **isometric**, which means that lengths and angles between vectors are not necessarily preserved.

In our case this transformation maps an 8-dimensional subspace of $\mathbb{R}^{16}$ onto $\mathbb{R}^8$. This is convenient because for two reasons: (a) in the smaller space the correlation sets are full-dimensional and (b) the origin plays the special role of a point of no correlations.

Note that the transformation above relies only on the fact that there are two possible outcomes and, therefore, it works for any number of settings. There exist generalisations to a higher number of outcomes but then some convenient and elegant features are lost.

From now we will think of a probability point P as a vector in $\mathbb{R}^8$ defined as

$$P := \big(\langle A_0\rangle, \langle A_1\rangle, \langle B_0\rangle, \langle B_1\rangle, \langle A_0B_0\rangle, \langle A_0B_1\rangle, \langle A_1B_0\rangle, \langle A_1B_1\rangle\big). \tag{2.3}$$

Similarly, a Bell functional on the reduced coordinates corresponds to 8 real numbers $\{a_x, b_y, c_{xy}\}_{x,y=0,1}$ and its value on the probability point specified above equals

$$\begin{aligned}\beta &= a_0\langle A_0\rangle + a_1\langle A_1\rangle + b_0\langle B_0\rangle + b_1\langle B_1\rangle \\ &\quad + c_{00}\langle A_0B_0\rangle + c_{01}\langle A_0B_1\rangle + c_{10}\langle A_1B_0\rangle + c_{11}\langle A_1B_1\rangle.\end{aligned} \tag{2.4}$$

As explained before a deterministic point can be fully specified by listing the outcomes of Alice and Bob for every measurement setting. Let use denote the outcome of Alice for measurement setting x by $a_x \in \{0,1\}$ and the outcome of Bob for measurement setting y by b_y. It is a simple exercise to see that the corresponding probability point is given by

$$P = \big((-1)^{a_0}, (-1)^{a_1}, (-1)^{b_0}, (-1)^{b_1}, (-1)^{a_0+b_0}, (-1)^{a_0+b_1}, (-1)^{a_1+b_0}, (-1)^{a_1+b_1}\big). \tag{2.5}$$

There are 16 deterministic points and let us denote them $D_1, \dots, D_{16}$. It is easy to check that

$$\sum_{j=1}^{16} D_j = (0,0,0,0,0,0,0,0) =: P_0. \tag{2.6}$$

In this sense the origin lies at the centre of the local set. If we compute the corresponding probability distribution, we obtain $P(ab|xy) = \frac{1}{4}$ for all a, b, x, y, i.e. the outcomes of Alice and Bob are maximally random and uncorrelated.

Example. If the local set were to have a centre of symmetry, P_0 would be a natural candidate. Does the reflection of $P_1 = (1,1,1,1,1,1,1,1)$ about P_0 belong to $\mathcal{L}$?

Example. Given a particular deterministic strategy which achieves P investigate how the following transformations of the strategy affect the resulting coordinates: (a) $a'_x := 1 - a_x$ and $b'_y = b_y$ and (b) $a'_x := 1 - a_x$ and $b'_y = 1 - b_y$. Do they imply any symmetries of the local polytope?

The local polytope in this scenario is defined by the 16 deterministic vertices of the form specified in Eq. (2.5) and turns out to have 24 facets. The first 16 of them correspond to positivity constraints $P(ab|xy) \geq 0$, which in the new coordinate system read

$$1 + (-1)^a\langle A_x\rangle + (-1)^b\langle B_y\rangle + (-1)^{a+b}\langle A_xB_y\rangle \geq 0 \tag{2.7}$$

for all a, b, x, y. Since the new coordinates are no-signalling by definition, we will not find any no-signalling conditions among the facets. All the remaining facets are known as the Clauser–Horne–Shimony-Holt (CHSH) facets and the first representative reads

$$\langle A_0B_0\rangle + \langle A_0B_1\rangle + \langle A_1B_0\rangle - \langle A_1B_1\rangle \leq 2. \tag{2.8}$$

Another facet inequality reads

$$\langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \geq -2. \tag{2.9}$$

Note that these two inequalities in some sense correspond to the "opposite" facets of the local set. Realising that the single minus sign can go with any of the four correlators explains why there are 8 CHSH facets in total. The CHSH facets turn out to be important, so it is convenient to define the corresponding functional. The **CHSH functional**, denoted by F_{CHSH}, reads

$$\langle F_{\mathrm{CHSH}}, P \rangle := \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle. \tag{2.10}$$

Equivalently in terms of the coefficients specified in Eq. (2.4) it is given by

$$a_x = b_y = 0 \quad \text{and} \quad c_{xy} = (-1)^{xy}. \tag{2.11}$$

Then, the inequality given in Eq. (2.8) can be simply written as $\langle F_{\mathrm{CHSH}}, P \rangle \leq 2$ and we say that the local value of the CHSH functional equals 2. The no-signalling set in this scenario is defined by positivity and no-signalling constraints. However, the coordinates we have chosen automatically take care of the no-signalling constraints, which can be seen directly from Eq. (2.2). In other words, in the new coordinate system the only constraints we need to impose to recover the no-signalling set are the 16 positivity constraints. The resulting polytope turns out to have 24 vertices: 16 of them are the deterministic points mentioned before while the remaining 8 are known as **Popescu–Rohrlich (PR) boxes**. The standard PR box reads

$$P_{\mathrm{PR}} := (0, 0, 0, 0, 1, 1, 1, -1) \tag{2.12}$$

and it is easy to check $\langle F_{\mathrm{CHSH}}, P_{\mathrm{PR}} \rangle = 4$. It is clear that this is the largest value of the CHSH functional possible simply because every correlator in the definition of the functional has modulus of at most 1. Therefore, the no-signalling value of the CHSH functional equals 4. It is easy to check that the corresponding probability distribution admits a particularly elegant description:

$$P(ab|xy) = \begin{cases} \frac{1}{2} & \text{if} \quad a \oplus b = xy, \\ 0 & \text{otherwise.} \end{cases} \tag{2.13}$$

There is also a PR box "on the other side" of the no-signalling set given by

$$(0, 0, 0, 0, -1, -1, -1, 1). \tag{2.14}$$

As before the minus sign can go with any of the four correlators which gives rise to the total of 8 PR boxes.

In the CHSH scenario there is an elegant duality between the facet Bell inequalities and the extremal no-signalling boxes: every facet Bell inequality is violated by exactly one

PR box so we may think that these objects come in pairs. Unfortunately, this property does not hold in larger scenarios and in general there is no one-to-one mapping between facet Bell inequalities and no-signalling extremal boxes.

The beauty of polytopes lies in the fact that they are simple. Once we have both the vertex and the half-space description of a polytope, nothing else can be added to enhance our understanding. Therefore, we have concluded our investigation of the local and no-signalling sets in the CHSH scenario.

2.2 The quantum set

The quantum set, on the other hand, turns out to be much more complicated. So far we only know it is a closed convex set which sits in between the two polytopes. Before proceeding any further let us reveal a convenient link between the reduced coordinates introduced above and the quantum realisation.

A measurement with two outcomes is given by two positive semidefinite operators $\{P_0, P_1\}$, but these are not independent: they are constrained by the normalisation condition $P_0 + P_1 = \mathbb{1}$. Therefore, such a measurement can be fully described by a single operator and for our purposes it is convenient to choose

$$A := P_0 - P_1. \tag{2.15}$$

We will call A a **binary observable** or simply **observable** representing the two-outcome measurement $\{P_0, P_1\}$. To see that the observable allows us to reconstruct the original measurement note that

$$P_0 = \frac{\mathbb{1} + A}{2} \quad \text{and} \quad P_1 = \frac{\mathbb{1} - A}{2}. \tag{2.16}$$

Observables are Hermitian operators and since

$$A \leq P_0 \leq \mathbb{1} \quad \text{and} \quad A \geq -P_1 \geq -\mathbb{1}, \tag{2.17}$$

we immediately see that their eigenvalues are contained in the interval $[-1, 1]$. This can be compactly written as $\|A\| \leq 1$, where $\|\cdot\|$ denotes the Schatten ∞-norm. If the original measurement is projective, the only allowed eigenvalues of A are ± 1, which implies that $A^2 = \mathbb{1}$. For non-projective measurements we can only deduce that $A^2 \leq \mathbb{1}$.

The observables relevant in the CHSH scenario are

$$A_x := P_0^x - P_1^x \quad \text{and} \quad B_y := Q_0^y - Q_1^y \tag{2.18}$$

and we can now think that the quantum realisation is fully described by the state ρ_{AB}, two observables of Alice A_0, A_1 and two observables of Bob B_0, B_1. One should now be able to see that the reduced coordinates introduced in Eq. (2.1) are in fact quantum-inspired. It is easy to check that

$$\begin{aligned} \langle A_x \rangle &= \mathrm{Tr}\,[(A_x \otimes \mathbb{1})\rho_{AB}], && (2.19)\\ \langle B_y \rangle &= \mathrm{Tr}\,[(\mathbb{1} \otimes B_y)\rho_{AB}], && (2.20)\\ \langle A_x B_y \rangle &= \mathrm{Tr}\,[(A_x \otimes B_y)\rho_{AB}]. && (2.21) \end{aligned}$$

Moreover, the Bell operator corresponding to a Bell functional specified in Eq. (2.4) reads

$$\begin{aligned} W = {} & a_0 A_0 \otimes \mathbb{1} + a_1 A_1 \otimes \mathbb{1} + b_0 \mathbb{1} \otimes B_0 + b_1 \mathbb{1} \otimes B_1 && (2.22)\\ & + c_{00} A_0 \otimes B_0 + c_{01} A_0 \otimes B_1 + c_{10} A_1 \otimes B_0 + c_{11} A_1 \otimes B_1. && (2.23) \end{aligned}$$

Since the only difference between the local and no-signalling set are the CHSH facets, it is natural to start the investigation of the quantum set by computing the quantum value of the CHSH functional. The Bell operator corresponding to the CHSH functional reads

$$W = A_0 \otimes (B_0 + B_1) + A_1 \otimes (B_0 - B_1). \tag{2.24}$$

To derive an upper bound on the quantum value we will provide a SOS decomposition as explained in Section 1.9. If

$$\begin{aligned} L_0 &:= A_0 \otimes \mathbb{1} - \mathbb{1} \otimes \frac{B_0 + B_1}{\sqrt{2}}, && (2.25)\\ L_1 &:= A_1 \otimes \mathbb{1} - \mathbb{1} \otimes \frac{B_0 - B_1}{\sqrt{2}}, && (2.26) \end{aligned}$$

then a simple calculation shows that

$$L_0^2 + L_1^2 = 4\,\mathbb{1} \otimes \mathbb{1} - \sqrt{2}\,W, \tag{2.27}$$

where we have assumed that the measurements are projective, i.e. $A_x^2 = \mathbb{1}$ and $B_y^2 = \mathbb{1}$. This implies that $W \leq 2\sqrt{2}\,\mathbb{1}$, which is equivalent to $\beta_Q \leq 2\sqrt{2}$. This upper bound can be saturated by performing the following measurements:

$$\begin{aligned} A_0 &:= \mathsf{X}, \quad B_0 := \frac{\mathsf{X} + \mathsf{Z}}{\sqrt{2}}, && (2.28)\\ A_1 &:= \mathsf{Z}, \quad B_1 := \frac{\mathsf{X} - \mathsf{Z}}{\sqrt{2}} && (2.29) \end{aligned}$$

on the maximally entangled state of two-qubits:

$$|\Phi^+\rangle := \frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big). \tag{2.30}$$

This implies that the quantum value of the CHSH functional equals $2\sqrt{2}$, which should be compared with the local and no-signalling values equal to 2 and 4, respectively. This implies that already in the CHSH scenario we have strict inclusions $\mathcal{L} \subsetneq \mathcal{Q} \subsetneq \mathcal{NS}$.

Computing the quantum value turns out to be easy for the CHSH functional but this is not always the case. Even in the CHSH scenario we do not have an analytic description of the quantum set or a simple analytic expression for the quantum value of an arbitrary Bell functional. Fortunately, there exist some numerical tools to tackle the problem and let us discuss one of them which relies on an important result from linear algebra known as **Jordan's lemma**.

Jordan's lemma. Let P and Q be two projectors acting on a separable Hilbert space $\mathcal{H}$. Then, there exists a basis on $\mathcal{H}$ such that P and Q are block-diagonal with blocks of size either 2×2 or 1×1.

Intuitively Jordan's lemma tells us that all the interesting features of how two projectors interact can already be found in the qubit case. A reformulation of Jordan's lemma states that the identity on $\mathcal{H}$ can be decomposed into projectors $\{\Pi_j\}_j$ such that

$$\operatorname{Tr}\Pi_j = 1 \quad \text{or} \quad \operatorname{Tr}\Pi_j = 2 \tag{2.31}$$

and

$$[P, \Pi_j] = [Q, \Pi_j] = 0. \tag{2.32}$$

We will now see that Jordan's lemma has deep implications on the structure of the quantum set in the CHSH scenario.

Consider a quantum realisation given by $\{P_a^x\}, \{Q_b^y\}$ and ρ_{AB}. As mentioned before we can without loss of generality assume that the measurements are projective. Let us now take two projectors of Alice which correspond to distinct measurement settings, e.g. P_0^0 and P_0^1, and apply Jordan's lemma to them. This yields a family of projectors which we denote by $\Pi_{A,j}$. Since $P_1^x = \mathbb{1} - P_0^x$, these projectors commute with all the measurement operators of Alice, not just the two we started with. This allows us to write

$$\begin{aligned}
P(ab|xy) = \operatorname{Tr}(P_a^x \otimes Q_b^y\, \rho_{AB}) &= \operatorname{Tr}\Big(\sum_j \Pi_{A,j} P_a^x \otimes \sum_k \Pi_{B,k} Q_b^y\, \rho_{AB}\Big) &(2.33)\\
&= \sum_{jk} \operatorname{Tr}\Big(\Pi_{A,j} P_a^x \Pi_{A,j} \otimes \Pi_{B,k} Q_b^y \Pi_{B,k}\, \rho_{AB}\Big) &(2.34)\\
&= \sum_{jk} \operatorname{Tr}\Big[P_a^x \otimes Q_b^y\, (\Pi_{A,j} \otimes \Pi_{B,k}\, \rho_{AB}\, \Pi_{A,j} \otimes \Pi_{B,k})\Big], &(2.35)
\end{aligned}$$

where we have used the completeness relation, then projectivity and commutativity and finally the cyclic property of the trace.

Let

$$q_{jk} := \operatorname{Tr}(\Pi_{A,j} \otimes \Pi_{B,k}\, \rho_{AB}\, \Pi_{A,j} \otimes \Pi_{B,k}). \tag{2.36}$$

Since the trace of a positive semidefinite operator vanishes if and only if the actual operator vanishes, we note that if $q_{jk} = 0$, then we may ignore the corresponding term in the sum given in Eq. (2.35). In the remaining cases, i.e. whenever $q_{jk} > 0$, let us define

$$\tau_{AB,jk} := \frac{1}{q_{jk}} (\Pi_{A,j} \otimes \Pi_{B,k}\, \rho_{AB}\, \Pi_{A,j} \otimes \Pi_{B,k}) \tag{2.37}$$

and note that

$$P(ab|xy) = \sum_{jk} q_{jk} \operatorname{Tr}(P_a^x \otimes Q_b^y\, \tau_{AB,jk}), \tag{2.38}$$

where the sum is taken only over pairs for which $q_{jk} > 0$. Clearly, $\tau_{AB,jk}$ is a normalised quantum state. If either of the projectors is rank-1 it necessarily has the product form, i.e. $\tau_{AB,jk} = \tau_A \otimes \tau_B$, and hence the outcomes of Alice and Bob will be uncorrelated. If both projectors are rank 2, the resulting state is a two-qubit state (although embedded in a higher-dimensional Hilbert space). This implies that any probability point in the CHSH scenario is a convex combination of points which can be achieved using two-qubit states. One one hand this implies that all extremal points of the quantum set can be achieved by two-qubit states. On the other hand, combining this with Carathéodory's theorem[1] implies that every point of $\mathcal{Q}$ can be achieved using a finite-dimensional realisation, i.e. that in the CHSH scenario $\mathcal{Q}_{\text{fin}} = \mathcal{Q}$.

Since all the extremal points can be achieved by two-qubit realisations, the quantum value of any Bell functional can be found by optimising over such realisations. As mentioned before finding the optimal state for fixed observables (measurements) correspond to finding the largest eigenvalue of the Bell operator. Therefore, our task reduces to parametrising the local observables of Alice and Bob. First of all, we only need to parameterise observables corresponding to projective measurements, i.e. satisfying $A^2 = \mathbb{1}$. If any observable equals $\pm\mathbb{1}$, then the statistics will be local (because commuting measurements give rise to local statistics). Therefore, the only non-trivial case happens when all the observables have exactly 1 eigenvalue of each sign. However, it is then easy to see that any pair of such observables is unitarily equivalent to

$$A_0 = \mathsf{X}, \tag{2.39}$$

$$A_1 = \cos a\, \mathsf{X} + \sin a\, \mathsf{Z} \tag{2.40}$$

for some $a \in [0, \pi]$.[2] Since applying local unitaries does not change the spectrum, we can assume that the observables of Alice are of this form. Similarly, we can assume that the observables of Bob are given by

$$B_0 = \mathsf{X}, \tag{2.41}$$

$$B_1 = \cos b\, \mathsf{X} + \sin b\, \mathsf{Z} \tag{2.42}$$

[1] Carathéodory's theorem states that if a point $x \in \mathbb{R}^n$ lies in the convex hull of some set $\mathcal{S}$, then it can be written as a convex combination of at most $n + 1$ points from $\mathcal{S}$.

[2] Observe that such observables can be interpreted as unit vectors in the Bloch sphere and the only property of a pair of vectors on a Bloch sphere which is invariant under rotations is the angle between them.

for some $b \in [0, \pi]$. Using this parametrisation we can construct the Bell operator which we denote by $W(a, b)$. It should now be clear that

$$\beta_Q = \max_{a,b \in [0,\pi]} \lambda_{\max}(W(a, b)). \tag{2.43}$$

This method allows us to numerically estimate the quantum value of any Bell functional in the CHSH scenario to arbitrary precision.

We have seen before that for a closed convex set the ability to compute the value of every functional constitutes a complete description of the set. It might not, however, be the most convenient description. In fact, it does not seem particularly helpful in deciding whether a probability point P belongs to $\mathcal{Q}$ or not (note that since P might not be extremal, it does not suffice to look at two-qubit realisations). While researchers are still looking for an analytic closed-form description of the quantum set, this is unlikely to exist. There is, however, a certain subproblem that admits a closed-form solution.

Suppose that we only care about the correlators and not the marginals, i.e. we want to decide whether for a specified combination of $\langle A_0 B_0 \rangle, \langle A_0 B_1 \rangle, \langle A_1 B_0 \rangle, \langle A_1 B_1 \rangle$ we can find marginals $\langle A_0 \rangle, \langle A_1 \rangle, \langle B_0 \rangle, \langle B_1 \rangle$ such that the point

$$P = \{\langle A_0 \rangle, \langle A_1 \rangle, \langle B_0 \rangle, \langle B_1 \rangle, \langle A_0 B_0 \rangle, \langle A_0 B_1 \rangle, \langle A_1 B_0 \rangle, \langle A_1 B_1 \rangle\} \tag{2.44}$$

belongs to the quantum set. Mathematically speaking we are asking about the **projection** of $\mathcal{Q}$ onto its last 4 coordinates. Recall that a projection of a convex set is a convex set and in this case let us denote the resulting set by $\mathcal{Q}_{\mathrm{cor}} \subseteq \mathbb{R}^4$. It was shown by Tsirelson in 1980 that if the answer is positive, we can without loss of generality choose the marginals to be unbiased, i.e. $\langle A_0 \rangle = \langle A_1 \rangle = \langle B_0 \rangle = \langle B_1 \rangle = 0$, which implies that the projection question is in fact equivalent to describing a particular **slice** of the quantum set.[3] Moreover, Tsirelson proved that for the unbiased marginals this probability point can be obtained from a quantum realisation based on the maximally entangled state of two qubits. It turns out this question has a closed-form characterisation. According to the celebrated **Tsirelson–Landau–Masanes** criterion the 4 correlators belong to $\mathcal{Q}_{\mathrm{cor}}$ if and only if

$$1 + \prod_{xy} \langle A_x B_y \rangle + \prod_{xy} \sqrt{1 - \langle A_x B_y \rangle^2} - \frac{1}{2} \sum_{xy} \langle A_x B_y \rangle^2 \geq 0, \tag{2.45}$$

where the sums and products go over $x, y \in \{0, 1\}$. If the left-hand side is strictly positive, the point in question is an interior point of $\mathcal{Q}_{\mathrm{cor}}$. Otherwise it lies on the boundary. Boundary points such that at most 1 correlator is of unit modulus are known to be extremal.

[3] It is easy to see that the same is true for the local set.

2.3 Hardy paradox

In the section above we have seen that comparing the local and quantum values of specific Bell functionals is an elegant approach to demonstrate that the local set is a strict subset of the quantum set. Let us conclude with an alternative approach, known as the Hardy paradox, which is based on looking at a particular slice of these high-dimensional sets.

Let us look at the slice given by the following three conditions:

$$P(0,0|0,0) = 0, \tag{2.46}$$

$$P(0,1|1,0) = 0, \tag{2.47}$$

$$P(1,0|0,1) = 0. \tag{2.48}$$

Suppose now that $P \in \mathcal{L}$ and hence can be written as a convex combination of the deterministic points. Since the conditions above force some probabilities to vanish, they must hold simultaneously for every term present in the convex combination. A straightforward analysis of the local vertices shows that there are only 5 vertices satisfying the constraints above and each of them satisfies $P(0,0|1,1) = 0$. Therefore, for a local point the three conditions above imply that $P(0,0|1,1) = 0$.

This turns out not be true in quantum mechanics, which is precisely the paradox. Consider the following realisation:

$$\begin{aligned} |\psi\rangle &= \sqrt{\frac{1-a^2}{2}}\big(|01\rangle + |10\rangle\big) + a|11\rangle, \\ A_0 &= B_0 = 2a\,\mathsf{X} + \sqrt{1-4a^2}\,\mathsf{Z}, \\ A_1 &= B_1 = -\mathsf{Z}, \end{aligned}$$

where $a := \sqrt{\sqrt{5}-2}$. It is easy to verify that the three constraints are satisfied, while $P(0,0|1,1) = (5\sqrt{5}-11)/2 \approx 0.09$. This is the largest violation of the Hardy paradox and, perhaps surprisingly, it is achieved by a non-maximally entangled state. The unique point that allows for this maximal violation is an interesting example from the geometric point of view: it is an extremal point of the quantum set but it is not exposed.

2.4 Visualising the three correlation sets

To conclude this section let us try to visualise the three correlation sets. Since the correlation sets are 8-dimensional we cannot simply plot them. The best we can do is to visualise specific 2-dimensional (or maybe 3-dimensional) slices of the full object, but it should be clear that one cannot hope to see all the relevant features in a single figure. Fig. 2.1 shows the most commonly used slice of the correlation sets in which the quantum set is simply a disc sandwiched in between two squares. Indeed, in this highly-symmetric slice the quantum set admits a closed-form characterisation. However, looking at other slices reveals the true complexity of the problem, see Figs. 2.2, 2.3 and 2.4. Indeed, the quantum set of correlations turns out to be as complex as a convex set can be.

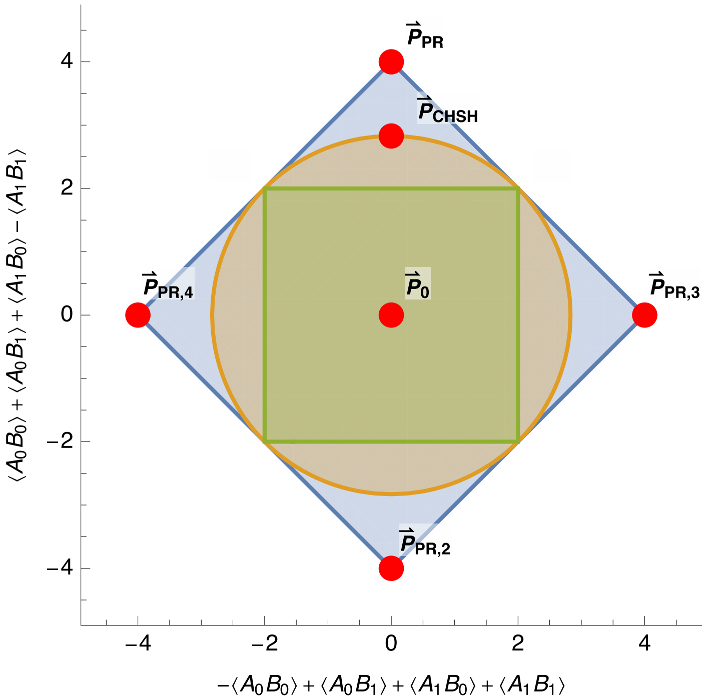

Figure 2.1: The most common visual representation of the three correlation sets. The local, quantum and no-signalling sets are shown in green, orange and blue, respectively. (Figure taken from `arXiv:1710.05892` reproduced with the authors' permission.)

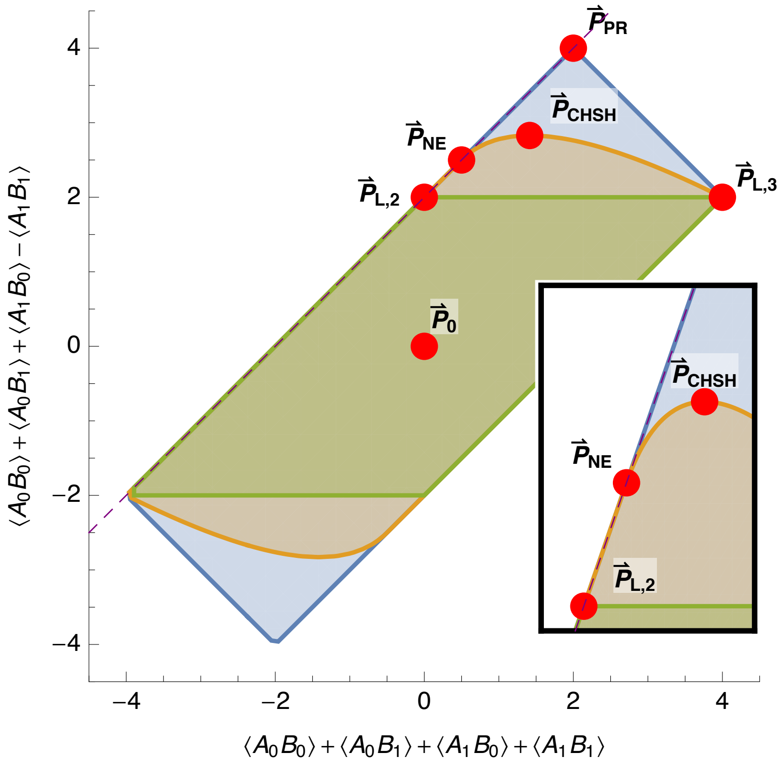

Figure 2.2: A slice of unbiased marginals which demonstrates that the quantum set contains points which are extremal but not exposed. (Figure taken from arXiv:1710.05892 reproduced with the authors' permission.)

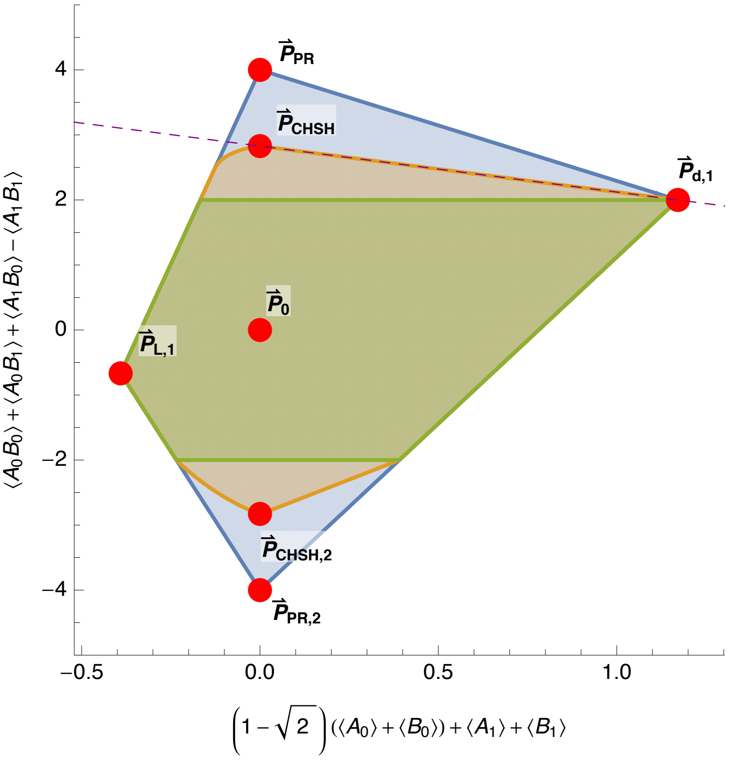

Figure 2.3: A slice showing a non-trivial exposed face of the quantum set which contains both local and nonlocal points. (Figure taken from `arXiv:1710.05892` reproduced with the authors' permission.)

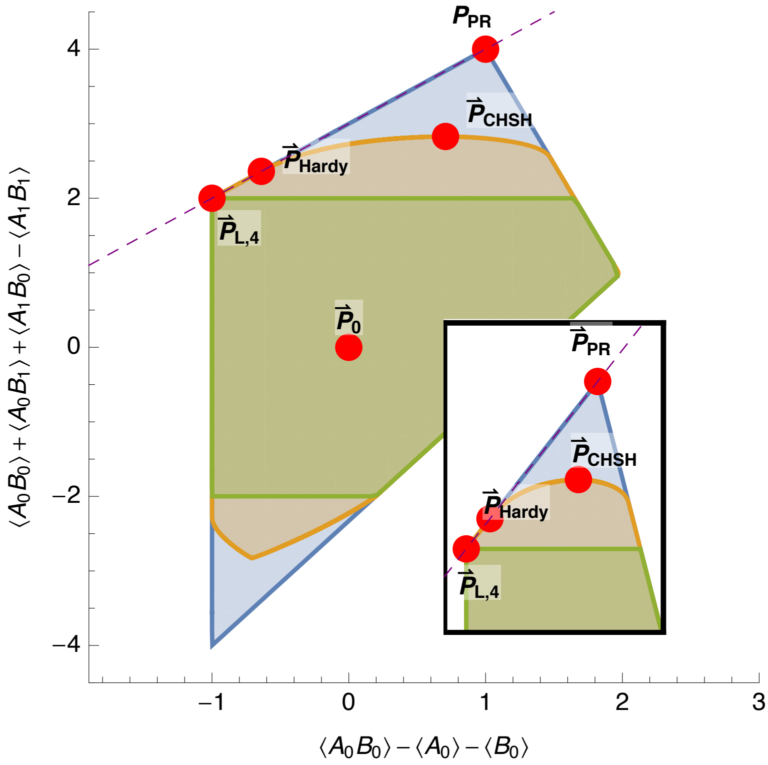

Figure 2.4: A slice showing the geometry of the quantum set around the Hardy point, which turns out to be extremal but not exposed. (Figure taken from arXiv:1710.05892 reproduced with the authors' permission.)

3 Beyond the CHSH scenario

While one could ask for some additional results in the CHSH scenario (e.g. an analytic description of the quantum set or a parametrisation of the extremal points), it is fair to say that our understanding there is almost complete. As one might expect things become significantly more complicated when we move on to Bell scenarios with more settings and more outcomes.

Let us first point out that the correlation sets corresponding to different number of settings and outcomes exhibit a nested structure. For instance, given the correlation sets of k settings and n outcomes we can obtain the correlation sets for $k-1$ setting by disregarding one of the settings. We can also reduce the number of outcomes by restricting our attention to distributions where certain outcomes do not appear. This implies that all the features we have observed in the CHSH case must necessarily be present in all larger Bell scenarios. Similarly, we can take a Bell functional from the CHSH scenario and interpret it as a functional in some larger Bell scenario and it is not hard to see that the local, quantum and no-signalling values are preserved.[1] Such procedures are sometimes referred to as **liftings**. Therefore, whenever analysing larger Bell scenarios we will necessarily see features already present in smaller scenarios, but of course we are mainly interested in new findings.

Moreover, when listing all facet Bell inequalities or extremal no-signalling boxes it is convenient to introduce some equivalence relations. We say that two probability distributions or Bell functionals are equivalent if they are related by some combination of the following operations: (a) swapping the roles of Alice and Bob, (b) relabelling the settings and (c) relabelling the outcomes. It is easy to see that all 8 facet Bell inequalities in the CHSH scenario belong to the same equivalence class.

So what is known about the correlation sets in larger Bell scenarios? Unfortunately, not that much. Finding the facet Bell inequalities of $\mathcal{L}$ and the extremal no-signalling boxes of $\mathcal{NS}$ is relatively easy in the CHSH scenario, but the difficulty of performing this task grows unexpectedly fast. In fact, the only other scenario where all the facets of $\mathcal{L}$ have been found corresponds to 3 settings and 2 outcomes and it turns out that in this scenario there are only two equivalence classes: liftings of the CHSH inequality and the so-called I_{3322} functional. However, already in the scenario with 4 settings and 2

[1] We implicitly assume that all three values are non-negative. Every Bell functional can be shifted to satisfy this condition.

outcomes the number of distinct classes is not known (it is known that there are at least 26 classes). Finding all the extremal boxes of the no-signalling set is slightly easier: we have a compact characterisation of extremal no-signalling boxes for scenarios whenever either $k = 2$ or $n = 2$ (in both cases they turn out to be simple generalisations of the PR box). Not much is known beyond that.

Given that in larger scenarios we cannot even completely describe the two polytopes, it might seem that there is no hope to describe the quantum set. Moreover, the only reliable tool for studying the quantum set presented so far crucially depended on Jordan's lemma which is only relevant for the scenario with two settings and two outcomes. It is clear that in order to make progress we must introduce a completely new approach.

3.1 Describing the quantum set through a hierarchy of optimisation problems

In Section 1.9 we described a heuristic method which allows us to find a lower bound on the quantum value of a Bell functional. However, since achieving the quantum value might require quantum realisations of an arbitrarily large dimension, that method by itself cannot be considered conclusive (even given access to unlimited computational power). What we are missing is a way of obtaining upper bounds and the procedure described below does precisely that.

The approach we will focus on now employs a hierarchy of optimisation problems to approximate the quantum set of correlations. The variant presented below is usually referred to as the **Navascués–Pironio–Acín (NPA) hierarchy** and is based on a simple observation. Suppose we are given a particular quantum realisation $\{|\psi\rangle, P_a^x, Q_b^y\}$ in the scenario with k measurement settings and n measurement outcomes. Since we only care about whether some probability point is achievable by quantum systems or not, we may without loss of generality assume that the state is pure and that the measurements are projective. We can now generate a set of $2nk$ vectors by considering

$$P_a^x \otimes \mathbb{1}\,|\psi\rangle \quad \text{for all} \quad x \in [k], a \in [n], \tag{3.1}$$

$$\mathbb{1} \otimes Q_b^y\,|\psi\rangle \quad \text{for all} \quad y \in [k], b \in [n]. \tag{3.2}$$

Let Γ be the **Gram matrix** of this set, i.e. a square matrix of size $2nk$ whose entries are given by the inner product of vectors, which by definition is positive semidefinite (every Gram matrix is positive semidefinite). This matrix contains terms of the form $\langle\psi|P_a^x \otimes Q_b^y|\psi\rangle$, which correspond to probabilities, but also other terms, which cannot be given physical meaning, e.g. $\langle\psi|P_{a'}^{x'} P_a^x \otimes \mathbb{1}|\psi\rangle$. Nevertheless, the normalisation and orthogonality of measurement operators impose certain linear constraints even on these "unphysical" entries, for instance:

$$\sum_{a'} \langle\psi|P_{a'}^{x'} P_a^x \otimes \mathbb{1}|\psi\rangle = \langle\psi|P_a^x \otimes \mathbb{1}|\psi\rangle \tag{3.3}$$

and

$$\langle\psi|P_{a'}^x P_a^x \otimes \mathbb{1}|\psi\rangle = \delta_{aa'} \langle\psi|P_a^x \otimes \mathbb{1}|\psi\rangle. \tag{3.4}$$

The bottom line here is that if a quantum realisation exists, we can write down a positive semidefinite matrix Γ satisfying all these constraints. This is convenient because searching for a positive semidefinite matrix satisfying a set of linear constraints is an instance of an optimisation problem which can be solved efficiently. Such problems constitute a generalisation of linear programs discussed in Section 1.6 and are called **semidefinite programs (SDPs)**.

Now given a probability point P we can ask whether there exists a positive semidefinite matrix Γ whose physical entries coincide with P and whose unphysical entries satisfy the required linear constraints. Most importantly, this can be checked efficiently using a numerical algorithm. If the answer turns out to be negative, we are guaranteed that P lies outside of the quantum set. If the answer is positive, we still cannot be sure whether $P \in \mathcal{Q}$ or not. Let us denote the set of probability points for which a valid Γ matrix can be found by $\mathcal{Q}_1$ and we will say that $\mathcal{Q}_1$ represents the first level of the NPA hierarchy. The observation above implies that $\mathcal{Q} \subseteq \mathcal{Q}_1$ and it should be easy to see that the set $\mathcal{Q}_1$ is compact and convex. Moreover, since no-signalling conditions are included in the linear constraints we immediately deduce that $\mathcal{Q}_1 \subseteq \mathcal{NS}$.

In the first level the rows and columns of the Gram matrix were labelled by first-degree mononomials in $\{P_a^x \otimes \mathbb{1}, \mathbb{1} \otimes Q_b^y\}$ and let us denote such a Γ matrix by Γ_1. To construct the second level of the hierarchy let us add rows and columns labelled by second-degree monomials and let us denote the resulting matrix by Γ_2. It should now be clear how to extend this to an arbitrary level and let us denote the corresponding Gram matrix by Γ_n. Moreover, let $\mathcal{Q}_n$ be the set of probability points for which a valid Γ_n matrix can be found. Since a valid Γ_n matrix contains a valid Γ_{n-1} matrix as a submatrix, we immediately deduce that for any $n \in \mathbb{N}$ we have

$$\mathcal{Q} \subseteq \mathcal{Q}_n \subseteq \mathcal{Q}_{n-1} \subseteq \ldots \subseteq \mathcal{Q}_1. \tag{3.5}$$

In other words we are dealing with a non-increasing sequence of compact convex sets which provide a better and better approximation of the quantum set.

So far we have argued that semidefinite programming allows us to efficiently check whether a given probability point belongs to a certain level of the NPA hierarchy, i.e. check the membership in $\mathcal{Q}_n$. Using similar methods we can also maximise any Bell functional over $\mathcal{Q}_n$ for any n. The result will only be an upper bound on the actual quantum value, but if a matching lower bound is found (e.g. by looking for explicit quantum realisations), we have identified the quantum value of this particular Bell functional.

Having understood the idea behind the NPA hierarchy it is natural to ask whether the sequence of sets $\{\mathcal{Q}_n\}_{n\in\mathbb{N}}$ converges to $\mathcal{Q}$. Note that this would imply that the values of a fixed Bell functional computed at increasing levels of the hierarchy would necessarily converge to the quantum value.

The answer turns out to be subtle: the NPA hierarchy indeed converges but to a quantum set defined in a slightly different manner. So far we have always worked in the **tensor-product paradigm** in which we start with a Hilbert space for Alice and a Hilbert space for Bob, denoted by $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, and the combined Hilbert space is given by $\mathcal{H}_A \otimes \mathcal{H}_B$. The measurements of Alice are given by operators acting on $\mathcal{H}_A$, while the measurements of Bob are given by operators acting on $\mathcal{H}_B$. The action of Alice on the combined Hilbert space is given by $P_a^x \otimes \mathbb{1}$, while the action of Bob is given by $\mathbb{1} \otimes Q_b^y$. Such operators necessarily commute:

$$[P_a^x \otimes \mathbb{1}, \mathbb{1} \otimes Q_b^y] = P_a^x \otimes Q_b^y - P_a^x \otimes Q_b^y = 0. \tag{3.6}$$

This paradigm makes sense if we can think of the systems of Alice and Bob as two discrete systems, e.g. two photons or atoms, which can be assigned a Hilbert space of their own. On the other hand, there are situations where the entire system is described by a single Hilbert space. There, the restriction that Alice and Bob should act on their own system can be replaced by the requirement that their actions commute. More specifically, all the measurement operators act on the same Hilbert space, but we require that every measurement operator of Alice commutes with every measurement operator of Bob, i.e. $[P_a^x, Q_b^y] = 0$ for all $x, y \in [k]$ and $a, b \in [n]$. This formulation is known as the **commuting paradigm** and let us denote the resulting quantum set by $\mathcal{Q}_c$. One can show that the NPA hierarchy converges to $\mathcal{Q}_c$. While it is immediately clear that $\mathcal{Q} \subseteq \mathcal{Q}_c$ the question of whether the two sets are equal was an important open problem known as the **Tsirelson's problem**. In January 2020 it was proven that $\mathcal{Q} \neq \mathcal{Q}_c$ by showing that there exists a Bell functional whose quantum values in the two paradigms are different. While the proof proceeds through a construction, this construction seems to be hard to implement. In particular, the authors do not give any estimates on how many measurement settings and outcomes one needs to observe the difference between the two sets.

Since for practical purposes the distinction between $\mathcal{Q}$ and $\mathcal{Q}_c$ does not seem to play any role, the NPA hierarchy is the main tool used to study the quantum set in larger Bell scenarios. One aspect of the NPA hierarchy that is poorly understood is its convergence: at the moment we have no way of quantifying how close $\mathcal{Q}_n$ is to $\mathcal{Q}$ or $\mathcal{Q}_c$.

To gain more practical understanding of the NPA hierarchy, let us work through an example. More specifically, we will use the first level of the NPA hierarchy to compute the quantum value of the CHSH functional.

Recall that in the CHSH scenario there are two measurement settings and two measurement outcomes. Hence, the rows and columns of the Gram matrix are labelled by vectors

$$P_a^x \otimes \mathbb{1} \, |\psi\rangle \quad \text{and} \quad \mathbb{1} \otimes Q_b^y \, |\psi\rangle \tag{3.7}$$

for $a, b, x, y \in \{0, 1\}$, which gives rise to an 8×8 matrix. However, we have seen before that when dealing with two-outcome measurements it is convenient to use observables instead of measurement operators. Therefore, we will instead consider the Gram matrix of the following vectors:

$$|\psi\rangle, A_x \otimes \mathbb{1}\,|\psi\rangle, \mathbb{1} \otimes B_y\,|\psi\rangle, \tag{3.8}$$

which gives rise to a 5×5 matrix. One can show that the two optimisation problems are equivalent because the 8 vectors given in Eq. (3.7) and the 5 vectors given in Eq. (3.8) span the same linear subspace.

An additional advantage of working in the observable-based picture is that the linear constraints become simpler. In fact, the linear constraints present at the first level of the NPA hierarchy simply reduce to setting all the diagonal entries to 1. Writing down the matrix explicitly yields:

$$\Gamma = \begin{pmatrix} 1 & \langle A_0 \rangle & \langle A_1 \rangle & \langle B_0 \rangle & \langle B_1 \rangle \\ \langle A_0 \rangle & 1 & \langle A_0 A_1 \rangle & \langle A_0 B_0 \rangle & \langle A_0 B_1 \rangle \\ \langle A_1 \rangle & \langle A_1 A_0 \rangle & 1 & \langle A_1 B_0 \rangle & \langle A_1 B_1 \rangle \\ \langle B_0 \rangle & \langle A_0 B_0 \rangle & \langle A_1 B_0 \rangle & 1 & \langle B_0 B_1 \rangle \\ \langle B_1 \rangle & \langle A_0 B_1 \rangle & \langle A_1 B_1 \rangle & \langle B_1 B_0 \rangle & 1 \end{pmatrix}.$$

The green entries are precisely the reduced coordinates introduced before, i.e. these are the "physical" entries. The red entries, for which we have used the shorthand notation $\langle A_x A_{x'} \rangle := \langle \psi | A_x A_{x'} \otimes \mathbb{1} | \psi \rangle$ and $\langle B_y B_{y'} \rangle := \langle \psi | \mathbb{1} \otimes B_y B_{y'} | \psi \rangle$, are the "unphysical" entries.

Having written down the first level of the hierarchy we can use it to compute an upper bound on the quantum value of an arbitrary Bell functional. If we choose the CHSH functional we reach the following optimisation problem:

$$\begin{aligned} &\text{maximise} \quad \Gamma_{24} + \Gamma_{25} + \Gamma_{34} - \Gamma_{35} && (3.9)\\ &\text{over} \quad \Gamma \geq 0 && (3.10)\\ &\text{satisfying} \quad \Gamma_{jj} = 1 \quad \text{for} \quad j \in \{1, 2, \ldots, 5\}, && (3.11) \end{aligned}$$

where Γ_{jk} is the relevant entry of Γ. This is a semidefinite program which can be solved using freely available numerical packages. These packages output the value of the problem and also a particular Γ matrix achieving it. In this case the value equals $2\sqrt{2}$ and one choice of a Γ matrix achieving it is given by:[2]

$$\Gamma = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 & 1 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 1 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 1 \end{pmatrix}. \tag{3.12}$$

[2] Note that most packages will only return numerical values, hence, recognising their analytical form might require some effort.

It is easy to check that this is a valid solution to the optimisation problem. In fact, the only condition that is not obvious from inspection is the positivity of Γ. To show that $\Gamma \geq 0$ we write it as a sum of positive semidefinite rank-1 operators. Indeed, it is easy to check that

$$\Gamma = \sum_{j=1}^{3} |e_j\rangle\langle e_j|$$

for

$$|e_1\rangle := \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad |e_2\rangle := \begin{pmatrix} 0 \\ 1 \\ 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad |e_3\rangle := \begin{pmatrix} 0 \\ 0 \\ 1 \\ \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{pmatrix}. \tag{3.13}$$

Note that vectors $|e_2\rangle$ and $|e_3\rangle$ correspond precisely to the terms appearing in the SOS decomposition given in Eqs. (2.25) and (2.26). Indeed, one can show that solving the semidefinite program is equivalent to searching for a SOS decomposition for the corresponding Bell operator.

The solution given in Eq. (3.12) is precisely the Gram matrix arising from the quantum realisation presented in Eqs. (2.28), (2.29) and (2.30). In particular, the resulting probability point (in the reduced coordinates) is given by:

$$P = \left(0, 0, 0, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right). \tag{3.14}$$

One can show that this is indeed the unique probability point which achieves the quantum value of $2\sqrt{2}$. However, this cannot be deduced from the first level of the NPA hierarchy. One can find other valid Γ matrices which correspond to probability points which do not belong to $\mathcal{Q}$. This shows that while the first level of the NPA hierarchy is sufficient to determine exactly the quantum value of the CHSH functional, it does not provide a tight characterisation of the quantum set in the CHSH scenario, i.e. $\mathcal{Q} \neq \mathcal{Q}_1$.

3.2 What is needed to generate nonlocal statistics?

Having realised that quantum states can be used to generate nonlocal statistics, it is natural to ask whether all quantum states are capable of doing so. This question was posed and answered in a seminal paper of Reinhard Werner in 1991. Werner defined the notion of separable states and it is clear that his motivation was to write down the largest class of states which must necessarily generate local statistics. In that sense one can think of separable states as classical (or non-useful for nonlocality purposes). In the same paper he defined a family of highly-symmetric states, now known as Werner states, and he showed that some of them cannot generate nonlocal correlations despite being entangled. On the other hand, it is easy to show that all pure entangled states can generate nonlocal correlations (in fact, they can violate the CHSH inequality).

A particularly curious is the case of PPT entangled states. Consider a state ρ_{AB} which is PPT and suppose measurements $\{P_a^x\}$ and $\{Q_b^y\}$ are performed on it. It is easy to see that the same statistics can be obtained by performing measurements $\{P_a^x\}$ and $\{[Q_b^y]^{\mathrm{T}}\}$ on the state $\rho_{AB}^{\mathrm{T}_B}$. In other words, the states ρ_{AB} and $\rho_{AB}^{\mathrm{T}_B}$ are indistinguishable when it comes to generating nonlocal correlations. This might come as a surprise as there are known examples where these two states have strikingly different entanglement properties.

Perhaps this is what led Asher Peres to conjecture that PPT states cannot generate nonlocal correlations. This statement, known as Peres conjecture, attracted significant attention, but progress was limited. Until recently the only known rigorous result in this direction was due to Werner and Wolf, who showed that PPT states do not violate a particular class of Bell inequalities (to which CHSH happens to belong). The problem was solved in 2014, when Vértesi and Brunner gave an example of a two-qutrit PPT state which violates some Bell inequality (but only by a little bit). The fact that finding such an example took the community so long and that the violation is tiny seems to suggest that this really is an unusual situation, but we do not have any analytic understanding of this phenomenon. Perhaps the difficulty arose from the fact that we had to go beyond the simplest, most-studied scenarios, i.e.: two qubits (there are no entangled PPT states of two qubits) and the CHSH Bell scenario (PPT states do not violate CHSH). One could define $\mathcal{Q}_{\mathrm{PPT}}$ as the set of correlations achievable using PPT states and any rigorous result allowing us to compare $\mathcal{Q}_{\mathrm{PPT}}$ to either $\mathcal{L}$ or $\mathcal{Q}$ would be extremely interesting.

Just like entanglement is what is required from the state, **incompatibility** is a necessary property of measurements. A pair of measurements acting on $\mathcal{H}$, denoted by $\{F_a\}_{a=1}^{n_A}$ and $\{G_b\}_{b=1}^{n_B}$ is called **compatible** or **jointly measurable** if there exists a measurement $\{H_{ab}\}_{ab}$ whose outcomes are labelled by pairs of (a, b), where $a \in [n_A]$ and $b \in [n_B]$, such that

$$\sum_b H_{ab} = F_a \quad \text{for all} \quad a \in [n_A], \tag{3.15}$$

$$\sum_a H_{ab} = G_b \quad \text{for all} \quad b \in [n_B]. \tag{3.16}$$

What this means is that the two original measurements can be performed simultaneously by performing the last measurement, often referred to as the **parent measurement**. This allows us to define a consistent joint probability distribution, which by Fine's theorem implies that the resulting statistics are local.

We have now seen what is obviously necessary to produce nonlocal correlations: the state must be entangled and both parties must perform incompatible measurements. Therefore, observing nonlocal correlations can be seen as certificate that the state is entangled and that the measurements are incompatible. The next step would be to make these conclusions stronger, e.g. by deducing how entangled the state is. This is precisely the idea behind **self-testing** or **device-independent certification of quantum devices**, which we will discuss in the last chapter.

4 Self-testing of quantum systems and device-independent cryptography

We have seen that in order to generate nonlocal correlations one must perform incompatible measurements on an entangled state. These conditions are necessary but they are not sufficient. As mentioned above there exist entangled states which are not capable of generating nonlocal statistics and even for highly entangled state one must carefully choose the measurements. For instance it is easy to check that performing $A_0 = B_0 = \mathsf{X}$ and $A_1 = B_1 = \mathsf{Z}$ on the maximally entangled state $|\Phi^+\rangle$ does not violate any Bell inequality.

Having realised that nonlocality is a rather special phenomenon which only occurs under particular circumstances, it is natural to ask whether it can be used for certification, we can ask: "given that I have observed certain statistics, what can I deduce about the underlying quantum system?". In most areas of quantum physics we start with a model and try to predict its properties, e.g. quantities that can be measured in an experiment. Here we are dealing with an inverse problem: we start with the observed data and we try to draw conclusions about the physical system.

The term certification is often used in the tomographic setting, e.g. we use trusted measurement devices to certify a source of quantum states or vice versa. We would call such an approach **device-dependent certification** because ultimately we rely on having at least some devices whose functioning we trust. The direction we are pursuing here is rather different.

Recall that the only assumption of the Bell scenario is that there are two devices which are not allowed to communicate. If we observe a Bell violation, we can immediately conclude that these devices do not admit a local-realistic description. Having ruled a local-realistic description it seems natural to assume that our devices are quantum. Since in this approach we only assume that we have two non-communicating quantum devices but we do not rely on a detailed characterisation, we will refer to this approach as **device-independent certification**.

4.1 How to maximally violate the CHSH inequality?

Let us start with the standard example of device-independent certification. Suppose we are given a pair of unknown quantum systems on which we can perform a Bell experiment. In the experiment we see that these two devices produce the maximal violation of the CHSH inequality. What can we say about these devices?

This question boils down to describing all the quantum realisations which achieve the maximal CHSH violation. Perhaps surprisingly, this question has a simple answer: all these realisations are "equivalent" to the quantum realisation given in Eqs. (2.28), (2.29) and (2.30). Let us now go through a proof of this statement which will give us the precise sense in which these realisations are "equivalent".

In Section 2.2 we have shown that a SOS decomposition for the CHSH operator W can be written using the following operators:

$$L_0 := A_0 \otimes \mathbb{1} - \mathbb{1} \otimes \frac{B_0 + B_1}{\sqrt{2}}, \tag{4.1}$$

$$L_1 := A_1 \otimes \mathbb{1} - \mathbb{1} \otimes \frac{B_0 - B_1}{\sqrt{2}}. \tag{4.2}$$

Previously we were only interested in deriving the quantum value of the CHSH functional and, hence, we could without loss of generality assume that the measurements are projective (recall that observables corresponding to projective measurements satisfy $A_x^2 = \mathbb{1}$ and $B_y^2 = \mathbb{1}$). Now our goal is to characterise all quantum realisations achieving the maximal violation, so we would like to drop this assumption. Without the projectivity assumption the SOS decomposition reads:

$$W = \frac{1}{\sqrt{2}}\Big((A_0^2 + A_1^2) \otimes \mathbb{1} + \mathbb{1} \otimes (B_0^2 + B_1^2)\Big) - \frac{1}{\sqrt{2}}(L_0^2 + L_1^2).$$

It is now easy to see that $\langle W, \rho_{AB} \rangle = 2\sqrt{2}$ implies that the following conditions hold:

$$\langle A_x^2, \rho_A \rangle = 1 \quad \text{for} \quad x \in \{0, 1\}, \tag{4.3}$$

$$\langle B_y^2, \rho_B \rangle = 1 \quad \text{for} \quad y \in \{0, 1\}, \tag{4.4}$$

$$\langle L_j^2, \rho_{AB} \rangle = 0 \quad \text{for} \quad j \in \{0, 1\}. \tag{4.5}$$

The first two conditions tell us that the measurements of Alice and Bob are projective on the support of the state. This is a inherent feature of device-independent certification: measurements can only be certified on the support of the state. To simplify the notation it is convenient to take the reduced states ρ_A and ρ_B to be full-rank. Note that this is not an assumption, we simply truncate the irrelevant (unoccupied) dimensions of the Hilbert space. Then, Eqs. (4.3) and (4.4) imply that

$$A_x^2 = \mathbb{1} \quad \text{and} \quad B_y^2 = \mathbb{1}. \tag{4.6}$$

In other words, we have deduced in a device-independent manner that the measurements are projective. Projectivity is a local property of the measurement, while the last condition allows us to deduce some relations between the operators of Alice and the operators of Bob. Since

$$\langle L_j^2, \rho_{AB} \rangle = \left\| L_j \rho_{AB}^{1/2} \right\|_F^2, \tag{4.7}$$

where $\|\cdot\|_F$ is the Frobenius norm, we see that Eq. (4.5) implies that $L_j \rho_{AB}^{1/2} = 0$ and finally $L_j \rho_{AB} = 0$. Writing this relation out for $j = 0$ gives:

$$(A_0 \otimes \mathbb{1})\rho_{AB} = \frac{1}{\sqrt{2}}\Big[\mathbb{1} \otimes (B_0 + B_1)\Big]\rho_{AB}. \tag{4.8}$$

Combining this with Eq. (4.6) implies that

$$\rho_{AB} = (A_0^2 \otimes \mathbb{1})\rho_{AB} = \frac{1}{\sqrt{2}}\Big[A_0 \otimes (B_0 + B_1)\Big]\rho_{AB} = \frac{1}{2}\Big[\mathbb{1} \otimes (B_0 + B_1)^2\Big]\rho_{AB} = \rho_{AB} + \frac{1}{2}[\mathbb{1} \otimes \{B_0, B_1\}]\rho_{AB} \tag{4.9}$$

and finally

$$[\mathbb{1} \otimes \{B_0, B_1\}]\rho_{AB} = 0. \tag{4.10}$$

Taking partial trace over Alice's system gives

$$\{B_0, B_1\}\rho_B = 0. \tag{4.11}$$

Finally, since ρ_B is full-rank, we can right-multiply this equation by the inverse ρ_B^{-1} to obtain

$$\{B_0, B_1\} = 0. \tag{4.12}$$

This turns out to be quite a powerful conclusion. One can show that if we are given Hermitian operators B_0 and B_1 acting on $\mathcal{H}_B$ satisfying $B_0^2 = B_1^2 = \mathbb{1}$ and $\{B_0, B_1\} = 0$, then the Hilbert space can be written as $\mathcal{H}_B \equiv \mathbb{C}^2 \otimes \mathcal{H}_{B'}$ and one can choose a basis such that

$$B_0 = \mathsf{X} \otimes \mathbb{1}, \tag{4.13}$$

$$B_1 = \mathsf{Z} \otimes \mathbb{1}. \tag{4.14}$$

This is a standard result in representation theory, but it can be proven using elementary methods, e.g. using Jordan's lemma.

Since the CHSH functional is symmetric with respect to swapping Alice and Bob, the same conclusions holds for the observables of Alice. More specifically, we deduce that the Hilbert space of Alice decomposes as $\mathcal{H}_A \equiv \mathbb{C}^2 \otimes \mathcal{H}_{A'}$, but here it is more convenient to choose a local basis such that

$$A_0 = \frac{\mathsf{X} + \mathsf{Z}}{\sqrt{2}} \otimes \mathbb{1}, \tag{4.15}$$

$$A_1 = \frac{\mathsf{X} - \mathsf{Z}}{\sqrt{2}} \otimes \mathbb{1}. \tag{4.16}$$

Having reconstructed the local observables we are ready to write down the Bell operator. A simple calculation shows that:

$$W = W_{\text{CHSH}} \otimes \mathbb{1}_{A'} \otimes \mathbb{1}_{B'}, \tag{4.17}$$

where

$$W_{\text{CHSH}} = \sqrt{2}(\mathsf{X} \otimes \mathsf{X} + \mathsf{Z} \otimes \mathsf{Z}) \tag{4.18}$$

is a two-qubit operator. Finally, we need to determine for which states ρ_{AB} the equality $\langle W, \rho_{AB} \rangle = 2\sqrt{2}$ holds. This reduces to identifying the eigenspace of W which corresponds to the largest eigenvalue of $2\sqrt{2}$. Since all the corresponding eigenstates are of the form

$$|\Phi^{+}\rangle \otimes |\psi_{\text{aux}}\rangle_{A'B'}, \tag{4.19}$$

we conclude that ρ_{AB} must be of the form:

$$\rho_{AB} = \Phi^{+} \otimes \sigma_{A'B'}. \tag{4.20}$$

This is quite remarkable as we have managed to give an almost-complete description of the unknown quantum system based solely on observing the maximal CHSH violation. In fact, there are only two aspects in which this description is different from the previously given ideal realisation: (a) we have to allow for auxiliary degrees on freedom on which the measurements act trivially and (b) we had to choose the local bases appropriately (which corresponds to applying local unitaries to the original realisation). It is easy to see that in the device-independent scenario these two equivalences are always present and we have to accept them.

The characterisation given in Eqs. (4.13), (4.14), (4.15), (4.16) and (4.20) is often referred to as a **self-testing statement** for the CHSH inequality. Intuitively, it means that the maximal violation of the CHSH inequality can be achieved in an essentially unique manner.

From the self-testing statement we immediately see that the CHSH inequality is maximally violated only by the probability point given in Eq. (3.14), which implies that it is an exposed point of the quantum set.

4.2 Device-independent cryptography

The fact that Alice and Bob can almost-completely characterise unknown quantum devices (under rather weak assumptions) turns out to have remarkable consequences for cryptographic applications.

Let us start with the task of **randomness generation**. Suppose that Alice wants to buy a random number generator from a potentially untrusted vendor. If she decides to buy a classical device, she can use it several times and check that the output passes various statistical tests. However, she can never rule out the scenario in which the device deterministically outputs a fixed string of bits generated in advance. In particular, this string could be completely known to the vendor.

This inherent limitation of classical devices can be overcome using quantum systems. If Alice buys a pair of quantum devices from an untrusted party, she simply needs to check that these two devices are capable of generating the maximal CHSH violation. To do so she simply needs to ensure that the devices cannot communicate during the Bell experiment, which can be done by placing them at distant locations. If this is the case, she can determine what is happening inside the device. In particular, since the optimal measurements correspond to Pauli measurements on a maximally entangled state, no external party can be correlated with the outcomes. This is a simple consequence of the fact that the only class of tripartite states compatible with Eq. (4.20) is given by:

$$\rho_{ABE} = \Phi^{+} \otimes \sigma_{A'B'E}. \tag{4.21}$$

In other words, the eavesdropper can only be correlated with auxiliary degrees of freedom which are ignored by the measurements of Alice and Bob. This implies that whenever Alice observes the maximal violation, she can be guaranteed that her devices are generating 1 perfect bit of fresh randomness unknown to anyone in the Universe.

The goal of randomness generation is to generate randomness unknown to anyone. In a slightly more complex task known as **quantum key distribution (QKD)** Alice and Bob try to generate a secret key unknown to the eavesdropper Eve. In the standard (entanglement-based) QKD Alice and Bob have access to trusted qubit measurements X and Z but the state, which we denote by ρ_{ABE}, is provided by Eve. The security argument relies on the fact that if Alice and Bob observe perfect correlations in the X and Z bases, i.e.

$$\langle \mathsf{X} \otimes \mathsf{X}, \rho_{AB} \rangle = \langle \mathsf{Z} \otimes \mathsf{Z}, \rho_{AB} \rangle = 1, \tag{4.22}$$

they can conclude that $\rho_{AB} = \Phi^{+}_{AB}$, which immediately implies that Eve must be uncorrelated, i.e. $\rho_{ABE} = \Phi^{+}_{AB} \otimes \rho_{E}$.

The standard quantum key distribution has a tomographic flavour: we use trusted measurement devices to characterise an unknown state. In the device-independent approach we can drop this assumption. As shown above Alice and Bob can take a pair of untrusted devices and use them to violate the CHSH inequality. If they observe the maximal violation, they know that Eve cannot be correlated with their measurement outcomes. Using standard classical postprocessing techniques these correlations can then be turned into a perfect key shared between Alice and Bob and completely unknown to Eve.