Advanced quantum information: entanglement and nonlocality (part 2)

Alexander Streltsov, Jędrzej Kaniewski

April 28, 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:

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

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

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

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

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

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