Lecture 8

The hierarchy of Navascués, Pironio, and Acín

In this lecture, we will define and study a class of strategies for nonlocal games known as *commuting measurement strategies*, or alternatively as *commuting operator strategies*. These strategies include all entangled strategies, in a sense that will be made more precise momentarily—and it was not long ago that this inclusion was proved to be proper by Slofstra [arXiv:1606.03140]. A proof that the *values* defined by the classes of entangled and commuting measurement strategies are different, where we take the supremum winning probability over the two classes of strategies, has only recently been announced by Ji, Natarajan, Vidick, Wright, and Yuen [arXiv:2001.04383]. But be warned—the paper is over 200 pages long. This refutes the famous Connes' embedding conjecture from the subject of von Neumann algebras, so it is worth every page it needs.

We will then analyze the semidefinite programming hierarchy of Navascués, Pironio, and Acín, better known as the *NPA hierarchy*, which provides us with a uniform family of semidefinite programs that converges to the *commuting measurement value* of any nonlocal game. This result is, in fact, a necessary ingredient in Ji, Natarajan, Vidick, Wright, and Yuen's proof.

8.1 Representing and comparing strategies

Let us fix question sets X and Y, and answer sets A and B, for a nonlocal game.

It is natural to think about *strategies* for nonlocal games having these question and answer sets as being represented by operators of the form

$$M \in \mathrm{L}(\mathbb{R}^X \otimes \mathbb{R}^Y, \mathbb{R}^A \otimes \mathbb{R}^B), \tag{8.1}$$

or equivalently as matrices whose columns are indexed by question pairs and whose rows are indexed by answer pairs. To be precise, the value $M(a, b|x, y)$ represents the probability that Alice and Bob answer the questions (x, y) with the answer (a, b).[1]

This representation is nice because not only does M store the probabilities with which Alice and Bob respond to each question pair (x, y) with answers (a, b), but also the action of M as a linear operator has meaning. For example, assuming the referee selects question pairs according to a probability vector π, we have that Alice and Bob's answers are distributed according to the vector $M\pi$. Perhaps more useful is to consider the vector

$$v = \sum_{(x,y)\in X\times Y} \pi(x, y) \, |x\rangle \otimes |y\rangle \otimes |x\rangle \otimes |y\rangle \tag{8.2}$$

and to observe that

$$\big(M \otimes \mathbb{1}_{X\times Y}\big)v \tag{8.3}$$

represents the joint probability distribution of quadruples (a, b, x, y) that arise from the selection of the questions (x, y) according to π together with Alice and Bob's answers to those questions.

Notice also that the probability for a strategy $M \in \mathrm{L}(\mathbb{R}^X \otimes \mathbb{R}^Y, \mathbb{R}^A \otimes \mathbb{R}^B)$ to win a particular nonlocal game $G = (X, Y, A, B, \pi, V)$ is equal to

$$\sum_{(x,y)\in X\times Y} \pi(x, y) \sum_{(a,b)\in A\times B} V(a, b|x, y)M(a, b|x, y), \tag{8.4}$$

which one may alternatively express as the inner product $\langle K, M\rangle$, where the operator $K \in \mathrm{L}(\mathbb{R}^X \otimes \mathbb{R}^Y, \mathbb{R}^A \otimes \mathbb{R}^B)$ is defined as

$$K(a, b|x, y) = \pi(x, y)V(a, b|x, y) \tag{8.5}$$

for all $x \in X$, $y \in Y$, $a \in A$, and $b \in B$.

Now, when we consider a particular *class* of strategies, such as the classical strategies or the entangled strategies, we are effectively defining a subset

$$\mathcal{S} \subset \mathrm{L}(\mathbb{R}^X \otimes \mathbb{R}^Y, \mathbb{R}^A \otimes \mathbb{R}^B) \tag{8.6}$$

of operators that represent strategies in the class under consideration. We then have the associated *value*

$$\omega_{\mathcal{S}}(G) = \sup_{M\in\mathcal{S}} \langle K, M\rangle \tag{8.7}$$

of a game G for the class $\mathcal{S}$.

[1] Throughout the lecture, entries of operators having the form $M \in \mathrm{L}(\mathbb{R}^X \otimes \mathbb{R}^Y, \mathbb{R}^A \otimes \mathbb{R}^B)$ are expressed as $M(a, b|x, y)$ rather than $M((a, b),(x, y))$, to mirror the notation we have adopted for the referee's predicate $V(a, b|x, y)$.

$$\begin{pmatrix}1&1&1&1\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}\quad\begin{pmatrix}1&0&1&0\\0&1&0&1\\0&0&0&0\\0&0&0&0\end{pmatrix}\quad\begin{pmatrix}0&1&0&1\\1&0&1&0\\0&0&0&0\\0&0&0&0\end{pmatrix}\quad\begin{pmatrix}0&0&0&0\\1&1&1&1\\0&0&0&0\\0&0&0&0\end{pmatrix}$$

$$\begin{pmatrix}1&1&0&0\\0&0&0&0\\0&0&1&1\\0&0&0&0\end{pmatrix}\quad\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}\quad\begin{pmatrix}0&1&0&0\\1&0&0&0\\0&0&0&1\\0&0&1&0\end{pmatrix}\quad\begin{pmatrix}0&0&0&0\\1&1&0&0\\0&0&0&0\\0&0&1&1\end{pmatrix}$$

$$\begin{pmatrix}0&0&1&1\\0&0&0&0\\1&1&0&0\\0&0&0&0\end{pmatrix}\quad\begin{pmatrix}0&0&1&0\\0&0&0&1\\1&0&0&0\\0&1&0&0\end{pmatrix}\quad\begin{pmatrix}0&0&0&1\\0&0&1&0\\0&1&0&0\\1&0&0&0\end{pmatrix}\quad\begin{pmatrix}0&0&0&0\\0&0&1&1\\0&0&0&0\\1&1&0&0\end{pmatrix}$$

$$\begin{pmatrix}0&0&0&0\\0&0&0&0\\1&1&1&1\\0&0&0&0\end{pmatrix}\quad\begin{pmatrix}0&0&0&0\\0&0&0&0\\1&0&1&0\\0&1&0&1\end{pmatrix}\quad\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&1&0&1\\1&0&1&0\end{pmatrix}\quad\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&0\\1&1&1&1\end{pmatrix}$$

Figure 8.1: The 16 deterministic strategies for binary question and answer pairs.

Example 8.1. Suppose $X = Y = A = B = \{0, 1\}$. There are 16 deterministic strategies for games having these question and answer sets, and they are represented by the matrices shown in Figure 8.1.

It is natural to associate the classical strategies for these question and answer sets with the *convex hull* of these 16 deterministic strategies to account for the possibility that Alice and Bob make use of randomness.

Representing the CHSH game as an operator K, as described above, yields

$$K = \frac{1}{4}\begin{pmatrix}1&1&1&0\\0&0&0&1\\0&0&0&1\\1&1&1&0\end{pmatrix}. \tag{8.8}$$

It is now perhaps more clear by an inspection of this matrix together with those appearing in Figure 8.1 that the classical value of the CHSH game is $3/4$; if M ranges over the matrices representing deterministic strategies, it is possible to make three of the 1s in any of these matrices, but not all four, overlap the nonzero entries of K.

Now, if we were to represent the set of all (probabilistic) classical strategies by

$$\mathcal{P} \subset \mathrm{L}(\mathbb{R}^X \otimes \mathbb{R}^Y, \mathbb{R}^A \otimes \mathbb{R}^B) \tag{8.9}$$

and the set of all entangled strategies as

$$\mathcal{E} \subset \mathrm{L}(\mathbb{R}^X \otimes \mathbb{R}^Y, \mathbb{R}^A \otimes \mathbb{R}^B), \tag{8.10}$$

then there are various observations we could make. For example:

1. $\mathcal{P}$ and $\mathcal{E}$ are both convex sets.
2. $\mathcal{P}$ is a polytope.
3. $\mathcal{P} \subseteq \mathcal{E}$, and the inclusion is proper so long as X, Y, A, and B all have at least two elements.

As special cases of (8.7) we have

$$\omega(G) = \sup_{M \in \mathcal{P}} \langle K, M \rangle \quad \text{and} \quad \omega^*(G) = \sup_{M \in \mathcal{E}} \langle K, M \rangle. \tag{8.11}$$

8.2 Commuting measurement strategies

Now we will introduce a new set of strategies, called *commuting measurement strategies*. The basic idea is to drop the tensor product structure that is present in an entangled strategy, replacing it with the assumption that Alice and Bob's measurement operators *commute*. We also drop the requirement that the space $\mathcal{H}$ is finite dimensional.[2]

Definition 8.2. For a given choice of question sets X and Y and answer sets A and B, we say that an operator $M \in \mathrm{L}(\mathbb{R}^X \otimes \mathbb{R}^Y, \mathbb{R}^A \otimes \mathbb{R}^B)$ represents a *commuting measurement strategy* if there exists a complex Hilbert space $\mathcal{H}$, a unit vector $u \in \mathcal{H}$, and projection operators

$$\{P_a^x : x \in X,\, a \in A\} \quad \text{and} \quad \{Q_b^y : y \in Y,\, b \in B\} \tag{8.12}$$

acting on $\mathcal{H}$, such that the following properties are satisfied. The projections represent measurements, in the sense that

$$\sum_{a \in A} P_a^x = \mathbb{1}_{\mathcal{H}} \quad \text{and} \quad \sum_{b \in B} Q_b^y = \mathbb{1}_{\mathcal{H}} \tag{8.13}$$

[2] If we restrict the definition to finite-dimensional $\mathcal{H}$, then we obtain precisely the entangled strategies. It is not the case, in contrast, that allowing the spaces $\mathcal{A}$ and $\mathcal{B}$ appearing in the definition of entangled strategies to be infinite dimensional makes entangled strategies equivalent to commuting measurement strategies.

for all $x \in X$ and $y \in Y$, the two collections of projections commute in pairs, meaning that $\left[P_a^x, Q_b^y\right] = 0$ for all $x \in X$ and $y \in Y$, and we have

$$M(a, b|x, y) = \langle u, P_a^x Q_b^y u \rangle \tag{8.14}$$

for all $x \in X$, $y \in Y$, $a \in A$, and $b \in B$.

Hereafter we will write

$$\mathcal{C} \subset \mathrm{L}(\mathbb{R}^X \otimes \mathbb{R}^Y, \mathbb{R}^A \otimes \mathbb{R}^B) \tag{8.15}$$

to denote the set of all commuting measurement strategies for the question and answer sets X, Y, A, and B. The *commuting measurement value* of G, which is denoted $\omega^c(G)$, is the supremum value of the winning probability for G taken over all commuting measurement strategies for Alice and Bob:

$$\omega^c(G) = \sup_{M \in \mathcal{C}} \langle K, M \rangle, \tag{8.16}$$

assuming again that K is defined from G as in (8.5).

Here are a couple of known facts about the set $\mathcal{C}$ of commuting measurement strategies and its relationship to the entangled strategies $\mathcal{E}$, stated without proof:

1. $\mathcal{C}$ is compact and convex.
2. $\mathcal{E} \subseteq \mathcal{C}$.

Proving that $\mathcal{C}$ is convex and includes $\mathcal{E}$ is a good exercise that I will leave to you. The fact that $\mathcal{C}$ is compact happens to follow from what we will prove later in the lecture (as does convexity). The containment $\mathcal{E} \subseteq \mathcal{C}$ is proper, as was mentioned at the start of the lecture.

8.3 The NPA hierarchy

This section is devoted to a description of the semidefinite programming hierarchy of Navascués, Pironio, and Acín—the *NPA hierarchy*—which is principally concerned with the commuting measurement strategies. Its convergence, and what exactly that means, will be discussed in the section following this one.

Basics of strings

When defining the NPA hierarchy in precise terms, it is helpful to make use of some elementary concepts concerning *strings* of symbols. In particular, we will be discussing matrices and vectors whose entries are indexed by strings of varying lengths. These concepts are elementary and quite familiar within theoretical computer science, and it will take just a moment to become familiar with them in case you are not.

Suppose that an alphabet Σ, which is a finite and nonempty set whose elements are viewed as symbols, has been fixed. A *string* over Σ is any finite, ordered sequence of symbols from Σ. (Infinite sequences of symbols are *not* to be considered as strings.) The *length* of a string is the total number of symbols, counting all repetitions, appearing in that string.

For example, if $\Sigma = \{0, 1\}$, then 0, 0110, and 1000100100011 are examples of strings over Σ; the string 0 has length 1, the string 0110 has length 4, and the string 1000100100011 has length 13. There is a special string that has length 0, and this string is called the *empty string*. We use the Greek letter ε to denote this string.

For every nonnegative integer n, we write $\Sigma^{\leq n}$ to denote the set of all strings having length at most n and we write Σ^* to denote the set of all strings, of any (finite) length, over Σ. For every alphabet Σ, the set Σ^* is countably infinite.

Lastly, given any string $s \in \Sigma^*$, we let s^{R} denote the string obtained by *reversing* the ordering of the symbols in s. For example, if $s = 00010$, then $s^{\mathsf{R}} = 01000$.

Intuition behind the NPA hierarchy

Next, let us introduce the basic idea behind the NPA hierarchy.

When we think about a particular commuting measurement strategy, represented by a complex Hilbert space $\mathcal{H}$, a unit vector $u \in \mathcal{H}$, and collections of projection operators $\{P_a^x : x \in X, a \in A\}$ and $\{Q_b^y : y \in Y, b \in B\}$ acting on $\mathcal{H}$, as described above, our interest is naturally with the numbers

$$M(a, b|x, y) = \langle u, P_a^x Q_b^y u \rangle, \tag{8.17}$$

ranging over all $x \in X$, $y \in Y$, $a \in A$, and $b \in B$.

These numbers arise when we consider the *Gram matrix* of the vectors

$$\{u\} \cup \{P_a^x u : x \in X, a \in A\} \cup \{Q_b^y u : y \in Y, b \in B\}. \tag{8.18}$$

Among the entries of this matrix, one finds all of the values

$$\langle P_a^x u, Q_b^y u \rangle = \langle u, P_a^x Q_b^y u \rangle = M(a, b|x, y) \tag{8.19}$$

that we ostensibly care about, as well as others, including (for instance)

$$\langle P_a^x u, P_c^z u \rangle, \quad \langle u, Q_b^y u \rangle, \quad \text{and} \quad \langle u, u \rangle, \tag{8.20}$$

for appropriate choices of x, y, z, a, b, and c.

We may now think about the various properties that must hold for such a Gram matrix, but before we do this let us discuss how we will index the entries of such matrices. Assuming that question and answer sets X, Y, A, and B have been fixed, we will introduce three alphabets:

$$\Sigma_A = X \times A, \quad \Sigma_B = Y \times B, \quad \text{and} \quad \Sigma = \Sigma_A \sqcup \Sigma_B. \tag{8.21}$$

Here, $\sqcup$ denotes the *disjoint union*, meaning that Σ_A and Σ_B are to be treated as disjoint sets when forming Σ. In words, there are $|X \times A| + |Y \times B|$ symbols in the alphabet Σ, one symbol for each pair $(x, a) \in X \times A$ and a separate symbol for each pair $(y, b) \in Y \times B$. The collection of vectors (8.18) may naturally be labeled by the set

$$\{\varepsilon\} \cup \Sigma = \Sigma^{\leq 1}, \tag{8.22}$$

and so we may consider that the Gram matrix of these vectors has rows and columns indexed by this set.

Now, supposing that

$$R \in \operatorname{Pos}\Big(\mathbb{C}^{\Sigma^{\leq 1}}\Big) \tag{8.23}$$

is such a Gram matrix—and naturally this is a positive semidefinite matrix, as all Gram matrices are—there are various things one may say, based on the conditions that commuting measurement strategies must meet. In particular, we observe these conditions:

1. $R(\varepsilon, \varepsilon) = 1$, given that u is a unit vector.
2. For every $x \in X$ we must have

$$\sum_{a \in A} R((x, a), s) = R(\varepsilon, s) \quad \text{and} \quad \sum_{a \in A} R(s, (x, a)) = R(s, \varepsilon), \tag{8.24}$$

   and likewise for every $y \in Y$ we must have

$$\sum_{b \in B} R((y, b), s) = R(\varepsilon, s) \quad \text{and} \quad \sum_{b \in B} R(s, (y, b)) = R(s, \varepsilon), \tag{8.25}$$

   for every $s \in \Sigma^{\leq 1}$ (in all four equalities). These conditions reflect the fact that summing over all operators in any given measurement yields the identity operator, as in (8.13).
3. For every $x \in X$ and $a, c \in A$ satisfying $a \neq c$, we have

$$R((x, a), (x, c)) = 0, \tag{8.26}$$

   because P_a^x and P_c^x must be orthogonal projection operators. Similarly, for every $y \in Y$ and $b, d \in B$ satisfying $b \neq d$, we have

$$R((y, b), (y, d)) = 0. \tag{8.27}$$

4. For every $(z,c) \in \Sigma$, we have the equality

$$R((z,c),(z,c)) = R(\varepsilon,(z,c)) = R((z,c),\varepsilon); \tag{8.28}$$

each P_a^x and Q_b^y is a projection operator, thus squaring to itself.

5. For every $(x,a) \in X \times A$ and $(y,b) \in Y \times B$ we have

$$R((x,a),(y,b)) = R((y,b),(x,a)), \tag{8.29}$$

as P_a^x and Q_b^y commute.

Remark 8.3. We can actually do a bit better in the case of the fifth item, and say that every entry $R((x,a),(y,b))$ must be a nonnegative real number, given that

$$R((x,a),(y,b)) = \langle P_a^x u, Q_b^y u \rangle = \left\| P_a^x Q_b^y u \right\|^2, \tag{8.30}$$

and likewise for $R((y,b),(x,a))$. This is a stronger condition than (8.29), as the entries $R((x,a),(y,b))$ and $R((y,b),(x,a))$ must be equal if they are real given that R is Hermitian. This stronger claim is not really needed though, and does not appear in the formal description of the NPA hierarchy coming up.

Let us now take $\mathcal{C}_1$ to be the subset of $\mathrm{L}(\mathbb{R}^X \otimes \mathbb{R}^Y, \mathbb{R}^A \otimes \mathbb{R}^B)$ containing all M for which there exists a positive semidefinite operator R satisfying items 1 through 5 above, as well as

$$M(a,b|x,y) = R((x,a),(y,b)) \tag{8.31}$$

for every $x \in X$, $y \in Y$, $a \in A$, and $b \in B$.

Observe that $\mathcal{C} \subseteq \mathcal{C}_1$, as every commuting measurement strategy defines a Gram matrix that satisfies the conditions of items 1 through 5. Thus, assuming once again that for a given nonlocal game $G = (X,Y,A,B,\pi,V)$ we have defined K as in (8.5), we see that

$$\omega^c(G) = \sup_{M \in \mathcal{C}} \langle K, M \rangle \leq \sup_{M \in \mathcal{C}_1} \langle K, M \rangle. \tag{8.32}$$

The inclusion $\mathcal{C} \subseteq \mathcal{C}_1$ is proper in general, but we can still use this relationship to obtain upper bounds on the commuting operator value (and therefore on the entangled value) of nonlocal games.

Now, notice that items 1 through 5 in the list above are all *affine linear constraints* on R. (With the exception of $R(\varepsilon,\varepsilon) = 1$ they are all linear constraints.) If we define a Hermitian operator $H \in \mathrm{Herm}\big(\mathbb{C}^{\Sigma^{\leq 1}}\big)$ as

$$H((x,a),(y,b)) = H((y,b),(x,a)) = \frac{1}{2}\pi(x,y)V(a,b|x,y) \tag{8.33}$$

for all $x \in X$, $y \in Y$, $a \in A$, and $b \in B$, and with all other entries equal to zero, we see that $\langle K, M \rangle = \langle H, R \rangle$. Thus, the optimization of $\langle K, M \rangle$ over all $M \in \mathcal{C}_1$ is represented by a semidefinite program, where we optimize $\langle H, R \rangle$ over all positive semidefinite R satisfying the affine linear constraints imposed by items 1 through 5 in the list above.

The semidefinite program just suggested is the *first* level of the NPA hierarchy. The idea behind subsequent levels is to consider not just the Gram matrix of the vectors (8.18), but of larger sets of vectors that include ones such as

$$P_a^x Q_b^y u, \; P_a^x Q_b^y P_c^z u, \; \text{etc.} \tag{8.34}$$

This will lead us to consider operators R whose rows and columns are indexed not by $\Sigma^{\leq 1}$, but by $\Sigma^{\leq k}$ for larger choices of k. (Larger choices for k yield higher levels in the hierarchy.) Although the inner products between most pairs of these vectors are not informative to the task of determining how well such a strategy performs in a given nonlocal game, the benefit comes from the introduction of additional *constraints* that reflect the same properties through which the five affine linear constraints above were derived. Indeed, the sequence of optimal values obtained from this hierarchy always converges to $\omega^c(G)$, as we will prove in the next section.

Formal description of the hierarchy

We are now ready to formally define the NPA hierarchy. We will begin by defining an equivalence relation $\sim$ on strings over the alphabet $\Sigma = \Sigma_A \sqcup \Sigma_B$ defined above. This equivalence relation will be used to equate various entries in matrices that generalize items 4 and 5 in the list described above.

Specifically, we take this equivalence relation to be the one generated by these rules holding for every $s, t \in \Sigma^*$, $(x, a) \in \Sigma_A$, and $(y, b) \in \Sigma_B$:

1. $s(x,a)t \sim s(x,a)(x,a)t$ and $s(y,b)t \sim s(y,b)(y,b)t$.
2. $s(x,a)(y,b)t \sim s(y,b)(x,a)t$.

That is, two strings are equivalent with respect to the relation $\sim$ if and only if one can be obtained from the other by any number of applications of the above rules. These equivalences reflect the fact that projections square to themselves (the first rule) and Alice's measurements commute with Bob's measurements (the second rule).

Next, notice that the values that appear in any Gram matrix R of the form suggested above always take the form

$$\left\langle u, \Pi_{c_1}^{z_1} \cdots \Pi_{c_n}^{z_n} u \right\rangle \tag{8.35}$$

for some finite sequence of projection operators $\Pi_{c_1}^{z_1}, \ldots, \Pi_{c_n}^{z_n}$ selected from Alice and Bob's projections.

With this observation in mind, we say that a function of the form $\phi : \Sigma^* \rightarrow \mathbb{C}$ is *admissible* if it satisfies these properties:

1. $\phi(\varepsilon) = 1$.
2. For all strings $s, t \in \Sigma^*$ we have
$$\sum_{a \in A} \phi(s(x,a)t) = \phi(st) \quad \text{and} \quad \sum_{b \in B} \phi(s(y,b)t) = \phi(st) \tag{8.36}$$
for every $x \in X$ and $y \in Y$.
3. For all strings $s, t \in \Sigma^*$ we have
$$\phi(s(x,a)(x,c)t) = 0 \quad \text{and} \quad \phi(s(y,b)(y,d)t) = 0 \tag{8.37}$$
for every $x \in X$ and $a, c \in A$ satisfying $a \neq c$, and every $y \in Y$ and $b, d \in B$ satisfying $b \neq d$, respectively.
4. For all strings $s, t \in \Sigma^*$ satisfying $s \sim t$ we have $\phi(s) = \phi(t)$.

We will also consider a restriction of this notion to functions of the form
$$\phi : \Sigma^{\leq k} \rightarrow \mathbb{C}, \tag{8.38}$$
which we call *admissible* if and only if the same conditions listed above hold for those strings s and t that are sufficiently short so that ϕ is defined on the arguments indicated within each condition.

Thus, if ϕ is a function that is defined from an actual commuting measurement strategy as
$$\phi((z_1, c_1) \cdots (z_n, c_n)) = \langle u, \Pi_{c_1}^{z_1} \cdots \Pi_{c_n}^{z_n} u \rangle \tag{8.39}$$
for every string $(z_1, c_1) \cdots (z_n, c_n)$, where each Π_c^z denotes P_c^z or Q_c^z depending on whether $(z, c) \in \Sigma_A$ or $(z, c) \in \Sigma_B$, respectively, then ϕ is necessarily admissible. This is true for functions of both forms $\phi : \Sigma^* \rightarrow \mathbb{C}$ and $\phi : \Sigma^{\leq k} \rightarrow \mathbb{C}$.

Finally, a positive semidefinite operator
$$R \in \text{Pos}\left(\mathbb{C}^{\Sigma^{\leq k}}\right) \tag{8.40}$$
is said to be a *k-th order admissible operator* if there exists an admissible function $\phi : \Sigma^{\leq 2k} \rightarrow \mathbb{C}$ such that
$$R(s, t) = \phi(s^{\mathsf{R}} t) \tag{8.41}$$
for every choice of strings $s, t \in \Sigma^{\leq k}$.

Observe that for each positive integer k, the condition that a positive semidefinite operator is k-th order admissible is an affine linear constraint, as there are a finite number of affine linear constraints imposed by the equation (8.41) and the condition that ϕ is admissible. Thus, the optimization over all k-th order admissible operators can be represented by a semidefinite program. This is the NPA hierarchy, where different choices of k correspond to different levels of the hierarchy. (Any linear objective function may naturally be considered, but often the objective function reflects the probability to win a given nonlocal game. Alternatively, one can set up a semidefinite program that tests whether a given M agrees with some k-th order admissible operator.)

8.4 Convergence of the NPA hierarchy

Define $\mathcal{C}_k$ to be the subset of $\mathrm{L}(\mathbb{R}^X \otimes \mathbb{R}^Y, \mathbb{R}^A \otimes \mathbb{R}^B)$ containing all M for which there exists a k-th order admissible operator $R \in \mathrm{Pos}\big(\mathbb{C}^{\Sigma^{\leq k}}\big)$ satisfying (8.31) for every $x \in X$, $y \in Y$, $a \in A$, and $b \in B$. One might call any such M a *k-th order pseudo-commuting measurement strategy*.

A moment's thought reveals that

$$\mathcal{C}_1 \supseteq \mathcal{C}_2 \supseteq \mathcal{C}_3 \supseteq \cdots \tag{8.42}$$

as any $(k+1)$-st order admissible operator must yield a k-th order admissible operator when its rows and columns corresponding to strings longer than k are deleted. It is also the case that $\mathcal{C} \subseteq \mathcal{C}_k$ for every positive integer k; just like in the $k = 1$ case, an actual commuting measurement strategy defines a Gram matrix R that is k-th order admissible for every choice of k.

The remainder of the lecture is devoted to proving that the sequence (8.42) converges to $\mathcal{C}$ in the sense made precise by the following theorem.

Theorem 8.4. *Let X, Y, A, and B be finite and nonempty sets and let $\mathcal{C}$ and $\mathcal{C}_k$, for every positive integer k, be as defined above. It is the case that*

$$\mathcal{C} = \bigcap_{k=1}^{\infty} \mathcal{C}_k. \tag{8.43}$$

Equivalently, for every $M \in \mathrm{L}(\mathbb{R}^X \otimes \mathbb{R}^Y, \mathbb{R}^A \otimes \mathbb{R}^B)$ the following two statements are equivalent:

1. *M is a commuting measurement strategy.*
2. *M is a k-th order pseudo-commuting measurement strategy for every positive integer k.*

The easier implication

We have already observed the implication that statement 1 implies statement 2. In more detail, under the assumption that statement 1 holds, it must be that M is defined by a commuting measurement strategy in which Alice and Bob's projective measurements are described by $\{P_a^x : a \in A\}$ for Alice and $\{Q_b^y : b \in B\}$ for Bob, with all of these projections acting on a Hilbert space $\mathcal{H}$, along with a unit vector $u \in \mathcal{H}$. As before, let Π_c^z denote P_c^z if $z \in X$ and $c \in A$, or Q_c^z if $z \in Y$ and $c \in B$. With respect to this notation, one may consider the k-th order admissible operator R defined by

$$R(s,t) = \phi(s^{\mathsf{R}}t), \tag{8.44}$$

where the function ϕ is defined as

$$\phi\big((z_1,c_1)\cdots(z_j,c_j)\big) = \big\langle u, \Pi_{c_1}^{z_1}\cdots\Pi_{c_j}^{z_j}u\big\rangle \tag{8.45}$$

for every string $(z_1,c_1)\cdots(z_j,c_j) \in \Sigma^{\leq 2k}$. A straightforward verification reveals that this operator is consistent with M, in the sense of (8.31), and therefore M is a k-th order pseudo-commuting measurement strategy.

A bound on entries of any k-th order admissible operator

Before approaching the more challenging implication of Theorem 8.4, which is that statement 2 implies statement 1, we will prove that every entry of a k-th order admissible operator is bounded by 1 in absolute value.

More explicitly, if $R \in \operatorname{Pos}\big(\mathbb{C}^{\Sigma^{\leq k}}\big)$ is k-th order admissible, then

$$\big|R(s,t)\big| \leq 1 \tag{8.46}$$

for every $s,t \in \Sigma^{\leq k}$. To see that this is so, observe first that

$$\big|R(s,t)\big| \leq \sqrt{R(s,s)}\sqrt{R(t,t)} \tag{8.47}$$

for each $s,t \in \Sigma^*$, which is a consequence of the fact that each 2×2 principal submatrix

$$\begin{pmatrix} R(s,s) & R(s,t) \\ R(t,s) & R(t,t) \end{pmatrix} \tag{8.48}$$

must be positive semidefinite. Noting that $R(s,s)$ is real and nonnegative, it therefore suffices to prove that

$$R(s,s) \leq 1 \tag{8.49}$$

for every $s \in \Sigma^{\leq k}$.

The bound (8.49) may be proved by induction on the length of s. For the base case, one has that $R(\varepsilon, \varepsilon) = 1$. For the general case, one has that for any string $s \in \Sigma^*$ and any choice of $(z, c) \in \Sigma$, it holds that

$$\begin{aligned} R((z,c)s,(z,c)s) \leq \sum_d R((z,d)s,(z,d)s) &= \sum_d \phi(s^{\mathsf{R}}(z,d)(z,d)s) \\ &= \sum_d \phi(s^{\mathsf{R}}(z,d)s) = \phi(s^{\mathsf{R}}s) = R(s,s), \end{aligned} \tag{8.50}$$

where $\phi : \Sigma^{\leq 2k} \to \mathbb{C}$ is the admissible function from which R is defined, and the sums are over all $d \in A$ or $d \in B$, depending on whether $z \in X$ or $z \in Y$, respectively. By the hypothesis of induction the required bound (8.49) follows.

Entry-wise convergence to an admissible matrix

Now assume statement 2 of Theorem 8.4 holds. For every $k \geq 1$, let

$$R_k \in \text{Pos}\left(\mathbb{C}^{\Sigma^{\leq k}}\right) \tag{8.51}$$

be a k-th order admissible operator satisfying

$$M(a,b|x,y) = R_k((x,a),(y,b)) \tag{8.52}$$

for every $x \in X$, $y \in Y$, $a \in A$, and $b \in B$, and let $\phi_k : \Sigma^{\leq 2k} \to \mathbb{C}$ be the admissible function that defines R_k.

We will begin with the observation that there exists an infinite matrix

$$R : \Sigma^* \times \Sigma^* \to \mathbb{C} \tag{8.53}$$

having the following properties:

1. Every finite principal submatrix of R is positive semidefinite.
2. $M(a,b|x,y) = R((x,a),(y,b))$ for every $x \in X$, $y \in Y$, $a \in A$, and $b \in B$.
3. There exists an admissible function $\phi : \Sigma^* \to \mathbb{C}$ such that $R(s,t) = \phi(s^{\mathsf{R}}t)$ for all $s, t \in \Sigma^*$.

Note here that we must draw a distinction between an infinite matrix of the form (8.53) and a linear operator, as these concepts are no longer equivalent in infinite dimensions.

Such an infinite matrix R can, in fact, be obtained in a fairly straightforward fashion. Observe first that for every string $s \in \Sigma^*$ and every infinite, strictly increasing sequence of positive integers $(k_1, k_2, k_3, \ldots)$, the sequence

$$(\phi_{k_1}(s),\ \phi_{k_2}(s),\ \phi_{k_3}(s),\ \ldots) \tag{8.54}$$

must have at least one limit point.[3] This is because each value $\phi_k(s)$ agrees with an entry of R_k, and the entries of each operator R_k are bounded by 1 in absolute value.

Beginning with the string $s = \varepsilon$ and the sequence $(k_1, k_2, \ldots) = (1, 2, \ldots)$, we consider the function $\phi : \Sigma^* \to \mathbb{C}$ defined by the following process:

1. Define $\phi(s)$ to be any limit point of the sequence (8.54).
2. Restrict the sequence $(k_1, k_2, \ldots)$ to any infinite subsequence for which (8.54) converges to the chosen limit point, and rename the indices forming this subsequence as $(k_1, k_2, k_3, \ldots)$.
3. Increment s and return to step 1.

Here, when we say "increment s," we are referring to any fixed total ordering of Σ^* for which ε is the first string. For example, the strings in Σ^* may be ordered according to their length, with strings of equal length being ordered according to the natural "dictionary ordering" induced by a fixed ordering of Σ (which is the so-called *lexicographic ordering* of Σ^*). It must be that any function ϕ obtained through this process is admissible, by virtue of the fact that every ϕ_k is admissible.

Now we may define

$$R(s,t) = \phi(s^{\mathsf{R}}t). \tag{8.55}$$

The three properties required of R follow, either directly or from the recognition that every finite submatrix of R is equal to the limit of the corresponding submatrices of some convergent subsequence of the sequence

$$(R_1, R_2, R_3, \ldots) \tag{8.56}$$

together with the fact that the positive semidefinite cone is closed.

Construction of a commuting measurement strategy

We will now make use of a fact concerning countably infinite matrices for which all finite principal submatrices are positive semidefinite—and that is that any such matrix must be the Gram matrix of a countably infinite set of vectors chosen from some Hilbert space. This is not a trivial fact to prove, but we will take it as given. In the case at hand, this implies that there must exist a Hilbert space $\mathcal{K}$ and a collection of vectors

$$\{u_s : s \in \Sigma^*\} \subset \mathcal{K}, \tag{8.57}$$

such that

$$R(s,t) = \langle u_s, u_t \rangle \tag{8.58}$$

[3] When considering the sequence (8.54), we ignore those indices k for which $\phi_k(s)$ is not defined, of which there are at most finitely many.

for every $s, t \in \Sigma^*$. (The inner product is, naturally, the inner product on $\mathcal{K}$.)

Now let us take $\mathcal{H}$ to be the closure of the span of the set $\{u_s : s \in \Sigma^*\}$, which is a Hilbert space having countable dimension (and is therefore a separable Hilbert space). We will define a commuting measurement strategy for Alice and Bob, with $\mathcal{H}$ being the Hilbert space for this strategy. The unit vector associated with this strategy will be $u_\varepsilon \in \mathcal{H}$, which is indeed a unit vector given that

$$\langle u_\varepsilon, u_\varepsilon \rangle = R(\varepsilon, \varepsilon) = 1. \tag{8.59}$$

Next we define a collection of projections on $\mathcal{H}$. For each choice of $x \in X$ and $a \in A$, define P_a^x to be the projection onto the orthogonal complement of the set

$$\{u_{(x,c)s} : c \in A \backslash \{a\},\ s \in \Sigma^*\}, \tag{8.60}$$

and along similar lines, for each choice of $y \in Y$ and $b \in B$, define Q_b^y to be the projection onto the orthogonal complement of the set

$$\{u_{(y,d)s} : d \in B \backslash \{b\},\ s \in \Sigma^*\}. \tag{8.61}$$

(The orthogonal complement of any collection of vectors is closed, so these are well-defined projection operators.)

In order to verify that the objects just defined induce a valid commuting measurement strategy that agrees with M, we will first prove the following useful fact. For all $(x, a) \in X \times A$, $(y, b) \in Y \times B$, and $s \in \Sigma^*$, it is the case that

$$P_a^x u_s = u_{(x,a)s} \quad \text{and} \quad Q_b^y u_s = u_{(y,b)s}. \tag{8.62}$$

Note first that for any choice of $x \in X$, $a, c \in A$ with $a \neq c$, and $s, t \in \Sigma^*$, we have

$$\langle u_{(x,a)s}, u_{(x,c)t} \rangle = R((x,a)s, (x,c)t) = \phi(s^{\mathsf{R}}(x,a)(x,c)t) = 0, \tag{8.63}$$

and similarly for any choice of $y \in Y$, $b, d \in B$ with $b \neq d$, and $s, t \in \Sigma^*$, we have

$$\langle u_{(y,b)s}, u_{(y,d)t} \rangle = 0. \tag{8.64}$$

This implies that

$$P_a^x u_{(x,a)s} = u_{(x,a)s} \quad \text{and} \quad Q_b^y u_{(y,b)s} = u_{(y,b)s}. \tag{8.65}$$

We also see that for any choice of $x \in X$ and $s, t \in \Sigma^*$,

$$\begin{aligned} \sum_{a \in A} \langle u_t, u_{(x,a)s} \rangle &= \sum_{a \in A} R(t, (x,a)s) \\ &= \sum_{a \in A} \phi(t^{\mathsf{R}}(x,a)s) = \phi(t^{\mathsf{R}}s) = R(t,s) = \langle u_t, u_s \rangle, \end{aligned} \tag{8.66}$$

from which it follows that

$$u_s = \sum_{a\in A} u_{(x,a)s}. \tag{8.67}$$

Similarly, for any choice of $y \in Y$ and $s \in \Sigma^*$,

$$u_s = \sum_{b\in B} u_{(y,b)s}. \tag{8.68}$$

Consequently,

$$P_a^x u_s = \sum_{c\in A} P_a^x u_{(x,c)s} = P_a^x u_{(x,a)s} = u_{(x,a)s}, \tag{8.69}$$

and by similar reasoning

$$Q_b^y u_s = u_{(y,b)s}, \tag{8.70}$$

as claimed.

With the formulas (8.62) in hand, we can verify the required properties of $\mathcal{H}$, u_ε, $\{P_a^x\}$, and $\{Q_b^y\}$. First, see that

$$\begin{aligned}\big\langle u_s, P_a^x Q_b^y u_t\big\rangle &= \big\langle u_s, u_{(x,a)(y,b)t}\big\rangle = \phi\big(s^{\mathsf{R}}(x,a)(y,b)t\big)\\ &= \phi\big(s^{\mathsf{R}}(y,b)(x,a)t\big) = \big\langle u_s, u_{(y,b)(x,a)t}\big\rangle = \big\langle u_s, Q_b^y P_a^x u_t\big\rangle\end{aligned} \tag{8.71}$$

for every $s, t \in \Sigma^*$. This implies that P_a^x and Q_b^y commute on the span of the vectors $\{u_s : s \in \Sigma^*\}$, and it follows that they commute on all of $\mathcal{H}$ by continuity. Second, for every $x \in X$ and $s \in \Sigma^*$ we find that

$$\sum_{a\in A} P_a^x u_s = \sum_{a\in A} u_{(x,a)s} = u_s, \tag{8.72}$$

from which it follows (again by continuity) that

$$\sum_{a\in A} P_a^x = \mathbb{1}_{\mathcal{H}} \tag{8.73}$$

and similarly

$$\sum_{b\in B} Q_b^y = \mathbb{1}_{\mathcal{H}}. \tag{8.74}$$

To complete the proof, it remains to observe that the strategy represented by the unit vector u_ε and the projections $\{P_a^x\}$ and $\{Q_b^y\}$ yields the strategy M. This is also evident from the formulas (8.62), as one has

$$R\big((x,a),(y,b)\big) = \big\langle u_{(x,a)}, u_{(y,b)}\big\rangle = \big\langle u_\varepsilon, P_a^x Q_b^y u_\varepsilon\big\rangle \tag{8.75}$$

and therefore

$$M(a,b|x,y) = \big\langle u_\varepsilon, P_a^x Q_b^y u_\varepsilon\big\rangle \tag{8.76}$$

for every choice of $x \in X$, $y \in Y$, $a \in A$, and $b \in B$.

Two implications

We will conclude the lecture by briefly observing two facts that follow from Theorem 8.4. The first is that the set $\mathcal{C}$ of commuting measurement strategies is closed, as it is the intersection of the closed sets $\mathcal{C}_1, \mathcal{C}_2, \ldots$ (and, by similar reasoning, it is convex). The second fact is that there is no loss of generality in restricting one's attention to separable Hilbert spaces in the definition of commuting measurement strategies, for this is so for the strategy constructed in the proof.