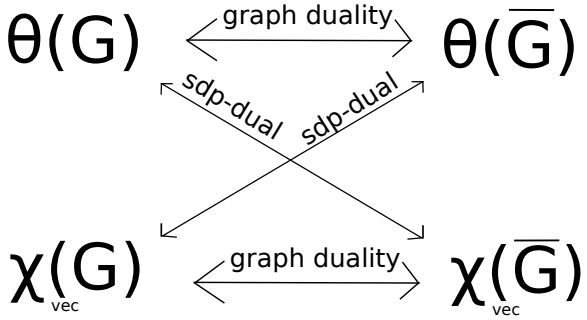
NOTES ON LOVASZ THETA

VICTOR BANKSTON

The purpose of this example is to illustrate some of the structure to $\vartheta(G)$.



Though the quantities above are scalars, they arise from arrangements of vectors. I will write down the vectors associated with each quantity above when G is the Petersen Graph.

Defining ϑ

First, we need a series of definitions to define ϑ[9]:

Definition 1. Given a graph G, an orthonormal representation of G is a mapping $r : V(G) \to \mathbb{S}^n \subset \mathbb{R}^{n+1}$ (for some $n \in \mathbb{N}$) such that if $i \neq j \in V(G)$, with $i \not\sim j$, then $r(i) \perp r(j)$. (Be careful: a vertex is not adjacent to itself).

Note that each graph has at least one representation, where v maps the verticies each to its own orthonormal vector.

Definition 2. A valuation of an orthonormal representation $val(r)$ is

$$\min_{\psi} \max_{v \in V(G)} \frac{1}{\left(\psi^T r(v)\right)^2}$$

where ψ ranges over all unit vectors (of the target space of r).

Given an orthonormal representation, its valuation is how tightly it can be embedded into a cone around some vector (ψ).

Definition 3. Define $\vartheta(G)$ to be the minimum valuation over all orthonormal representations of G.

We can show that this minimum is actually attained. We will use Bolzano-Weirstrass. To see this, fix n, and consider the orthonormal representations of the form: $v : V(G) \to \mathbb{S}^{n-1} \subset \mathbb{R}^n$. Observe that $\vartheta(G)$ remains unchanged if we require that $\psi = (1, 0, 0, 0 \dots)$: These valuations are defined by an inner product, which will not change if we apply a fixed unitary U to every vector. Choose U to send $\psi \mapsto (1, 0, 0, 0 \dots)$.

Fixing ψ, take a sequence of orthonormal representations whose values converge to $\vartheta(G)$. Observe that these orthonormal representations themselves can be considered as bounded vectors of dimension $n \cdot V(G)$, by concatenating all $V(G)$ vectors of dimension n. By the Bolzano-Weirstrass theorem, these have a convergent subsequence, so there is an accumulation point, r_∞, which we must show is an orthonormal representation.

Our convergent subsequence of orthonormal representations gives rise to $V(G)$ convergent sequences of vectors. We must show that each sequence of vectors goes to a unit vector, and that when $i \not\sim j$, with $i \neq j$, we have $r_\infty(i)^T r_\infty(j) = 0$. Both of these are consequences of the fact that dot products are continuous: $0 = \lim_{n\to\infty} r_n(i)^T r_n(j) = (\lim_{n\to\infty} r_n(i))^T (\lim_{n\to\infty} r_n(j))$.

There is no claim that such optimal representations are unique.

Graphs

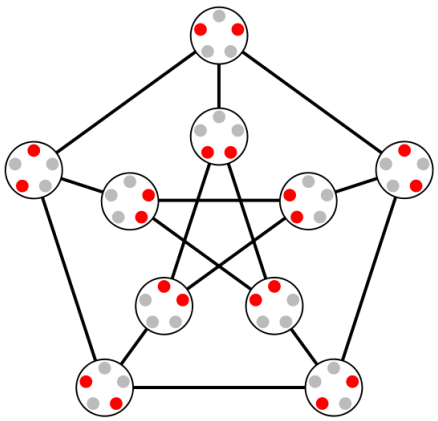
Definition 4. Define the Kneser Graph $k(n,r)$ to have $\binom{n}{r}$ verticies labeled by r-element subsets from a universe of size n. Two verticies are adjacent if their corresponding sets are disjoint. We assume that $n \geq 2r$

Kneser graphs are vertex and edge transitive. Given any two verticies (edges), there is an automorphism which sends one to the other.

Theorem 5. *If G is vertex and edge transitive, then $\vartheta(G)\vartheta\left(\overline{G}\right) = n$, and $\vartheta(G) = \frac{-n\lambda_n}{\lambda_1 - \lambda_n}$*

This powerful theorem was originally used to find $\vartheta(k(n,r))$. The proof of the theorem builds on the relations in the diagram.

Definition 6. The Petersen Graph, P, is the Kneser graph, $k(5,2)$.



We start with some graph properties.

Claim 7. The clique number of the kneser graph $\omega(k(n,r)) = \left\lfloor \frac{n}{r} \right\rfloor$, so $\omega(P) = 2$

A clique corresponds to a collection of disjoint sets.

Claim 8. [17]The coloring number $\chi(k(n,r)) = n - 2r + 2$, so $\chi(P) = 3$

This was a big open problem for many years. The optimal coloring is the following: Order the elements of the universe $u_1, \ldots, u_n$, and divide them into 3 pieces

with sizes $n-2r$, r and r. Let x be an r-set. If it intersects the first piece, color x with the color i, where $i = \min\{i \mid u_i \in x\}$. Otherwise, x is contained entirely in the last two pieces. These remaining verticies form a subgraph, where each vertex has a unique neighbor, and these can be colored with two colors.

Claim 9. The independence number of the kneser graph is $\alpha(k(n,r)) = \binom{n-1}{r-1}$, so $\alpha(P) = 4$

The collection of r-subsets which each contain u_1 is a set of this size. It isn't hard to show this is optimal.

Claim 10. [17]The clique covering number is $q(k(n,r)) = \left\lceil \frac{\binom{n}{k}}{\lfloor \frac{n}{k} \rfloor} \right\rceil$, $q(P) = 5$

$\vartheta(k(n,r)) = \binom{n-1}{r-1}$, and $\vartheta(k(n,r)) = \frac{n}{r}$, so $\vartheta(P) = 4, \vartheta\left(\overline{P}\right) = \frac{5}{2}$

This is proven by Theorem 5 and some tricky algebra, but this avoids (or at least obscures) creating explicit orthonormal representations, which is the point of this example.

Relations between Graph Constants

Theorem 11. $\alpha(G)\chi(G) \geq \mid V(G) \mid$

Each color is an independent set, and a proper coloring colors every vertex.

Theorem 12. *For any graph G, $\alpha(G) \leq \vartheta(G) \leq \chi\left(\overline{G}\right) = q(G)$*

In an orthonormal representation, an independent set, α, of G must be sent to a collection of pairwise independent vectors. For such vectors, it is easy to see that $\max_{v_i \in \alpha} \frac{1}{(\psi^T r(v_i))^2}$ is minimized when $\psi = \frac{\sum_{v_i \in \alpha} r(v_i)}{\sqrt{|V(G)|}}$ (when ψ is between all the vectors.) In this case, $\frac{1}{(\psi^T r(v_i))^2} = \mid \alpha \mid$, and this lower bound holds for all orthonormal representations. This shows $\alpha(G) \leq \vartheta(G)$.

Suppose we have clique cover of size $q(G)$. Define an orthonormal representation by choosing $q(G)$ pairwise orthonormal vectors. Send each clique to one of these vectors. This provides an explicit orthonormal representation with valuation $q(G)$. The minimum over all orthonormal representations may be less.

Orthonormal Representations

We start with the graph $\overline{k(n,r)}$ and construct an optimal orthonormal representation in dimension n, with orthonormal basis $u_1, \ldots, u_n$ (overloading the names of the basis elements with the elements of the universe) The choice is obvious: disjoint sets need to go to orthonormal vectors. Set $u_i^T r(v_j) = \frac{1}{\sqrt{r}}$ if $u_i \in v_j$, and 0 otherwise. Set $\psi = \frac{1}{\sqrt{n}}(1, 1, \ldots, 1)$. It is immediate that this is an orthonormal representation with valuation $\frac{1}{(\psi^T r(v_i))^2} = \frac{r \cdot n}{r^2} = \frac{n}{r}$. This O.R. spans a space of dimension 5.

Definition 13. Given an orthonormal representation, we can define the cost of a vertex to be $c(v) = \left(\psi_1^T (r_1(v_i))\right)^2$. This corresponds to the quantum-mechanical probability of measuring $r_1(v_1)$ when measuring from state ψ.

Theorem 14. [8]*(Certification of Orthonormal Representations): if we have two orthonormal representations r_1, r_2 of G and $\overline{G}$ and for all $i \in V(G)$ we have $c_1(v_i) = \frac{1}{\vartheta}$, and we also have $\sum_i c_1(v_i) c_2(v_i) = 1$, then $\vartheta = \vartheta(G)$*

Proof. We have the explicit orthonormal representation r_1, so $\vartheta(G) \leq \vartheta$. For the other direction, we use an alternate definition ϑ, $\vartheta(G) = \max_{Rep(\overline{G})} \sum_i c(v_i)$.

$$\min_{r \in O.R.(G)} \max_i \frac{1}{c_r(v_i)} = \vartheta(G) \leq \sum_i \vartheta c_1(v_i) c_2(v_i) = \sum_i c_2 v_i \leq \max_{r \in O.R.(\overline{G})} \sum_i c_r(v_i) = \vartheta(G)$$

□

The argument above also shows that certificates always exist.

The next definition is crucial, and describes the relationship between $\vartheta(G)$ and $\vartheta(\overline{G})$. From the physical perspective, this will relate bell inequalities of completely different experiments. Can this relation be found using the Sheaf Theory?

Definition 15. [7]Given a non-empty closed convex set $P \subset \mathbb{R}^n_+$ with the property that $x \in P$ and $0 \leq x' \leq x$ then $x' \in P$, the antiblocker of P is

$$AB(P) = \left\{x \in \mathbb{R}^n_+ : y^T x \leq 1 \text{ for all } y \in P\right\}$$

The condition that $0 \leq x' \leq x \implies x' \in P$ implies that $AB(AB(P)) = P$.

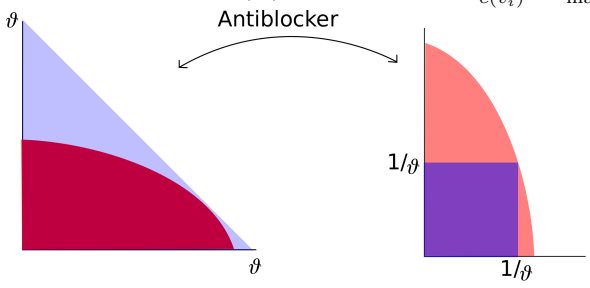
Example 16. Let $P_\vartheta = \left\{y \in \mathbb{R}^n_+ \mid \sum_{i=1}^n y_i \leq \vartheta\right\}$. Then $AB(P) = C_{\frac{1}{\vartheta}} = \left\{x \in \mathbb{R}^n_+ \mid \forall i, x_i \leq \frac{1}{\vartheta}\right\}$.

Proof. Let $y \in P_\vartheta, x \in C_{\frac{1}{\vartheta}}$. Then $y^T x = \sum_{i=1}^n y_i x_i \leq \frac{1}{\vartheta} \sum_{i=1}^n y_i \leq 1$. This shows $C_{\frac{1}{\vartheta}} \subset AB(P_\vartheta)$. Conversely, if $x \notin C_{\frac{1}{\vartheta}}$ for some $i \in V(G)$ $x_i > \frac{1}{\vartheta}$. Choose y such that $y_j = 0$ when $i \neq j$, and $y_i = \vartheta$. Then $y \in P_\vartheta$, and $x^T y > 1$, so $x \notin AB(P_\vartheta)$ □

Definition 17. $TH(G) = \left\{(c(v_i), v_i \in V(G)) \in \mathbb{R}_+^{V(G)}\right\}$. These are assignable probabilities, which (claim) satisfy the hypotheses of definition 6. (Note, these probablities do not need to sum to 1. We allow that some experiments have outcomes which are disregarded. The problem is intractable otherwise.)

Theorem 18. $AB(TH(G)) = TH(\overline{G})$

These concepts provide a geometric description of two definitions of ϑ. ϑ is the maximal of a linear functional over $TH(G)$: $\vartheta(G) = \max_{\overline{O.R.}} \sum_{i=1}^{|V(G)|} c(v_i)$. This linear functional has hyperplanes as its level sets, and the optimal value corresponds to a level set which lies tangent to $TH(G)$. Thus, $\vartheta(G)$ is the smallest simplex S_ϑ such that $TH(G) \subset S_\vartheta$. If we take the antiblocker of this picture, we seek the reciprocal of the largest cube $C_{\frac{1}{\vartheta}}$ such that $C_{\frac{1}{\vartheta}} \subset AB(TH(G)) = TH(\overline{G})$. This explains the formula $\vartheta(G) = \min_{O.R.} \max_i \frac{1}{c(v_i)} = \frac{1}{\max_{O.R.} \min_i c(v_i)}$.



Next, we give an orthonormal representation of $\vartheta(P)$, which will certify the optimality of the orthonormal representation given at the begining of this section.

Assume a basis of size 10, $\{e_{s_1}, e_{s_2}, \ldots, e_{s_{10}}\}$ labeled by the $\binom{5}{2}$ subsets of the graph. Let $\psi = \frac{1}{\sqrt{10}}(1, 1, \ldots, 1)$. Finally, assume that we will have $e_{s_i}^T r(v_j) =$

$x_{|s_i \cap v_j|}$. This is a plausible assumption, because it will result in vectors whose orthongonality relations are invariant with respect to the automorphism group of P.

The fact that intersecting sets must be sent to orthonormal vectors translates into the contstraint

$$x_0^2 + 3x_1^2 + 4x_0x_1 + 2x_1x_2 = 0$$

At the same time, we would like to minimize $\frac{10\cdot\left(x_2^2+6x_1^2+3x_0^2\right)}{\left(x_2+6x_1+3x_0\right)^2}$. According to Wolfram Alpha the minimum is 4, when $(x_0, x_1, x_2) = \left(1, -4-\sqrt{15}, 6+\sqrt{15}\right)$, or when $(x_0, x_1, x_2) = (a, b, c) = \left(1, \sqrt{15}-4, 6-\sqrt{15}\right)$.

$$\begin{pmatrix}
c & b & b & b & b & b & b & a & a & a \\
b & c & b & b & b & a & a & b & b & a \\
b & b & c & b & a & b & a & b & a & b \\
b & b & b & c & a & a & b & a & b & b \\
b & b & a & a & c & b & b & b & b & a \\
b & a & b & a & b & c & b & b & a & b \\
b & a & a & b & b & b & c & a & b & b \\
a & b & b & a & b & b & a & c & b & b \\
a & b & a & b & b & a & b & b & c & b \\
a & a & b & b & a & b & b & b & b & c
\end{pmatrix}$$

These vectors span a space of dimension 6.

Remark 19. For any given graph G, it is not true that all optimal orthogonal representations have the same dimension. For example, there are two optimal orthonormal representations of C_4: $\{(0,1),(0,1),(1,0),(1,0)\}$ with handle $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, and $\{(a,a,0),(a,-a,0),(a,0,a),(a,0,-a)\}$ with handle $(1,0,0)$.

Finally, we apply the certification theorem. For the O.R. above, each cost is $\frac{1}{4}$, and $\sum_{i=1}^{10} \frac{1}{4}\cdot\frac{2}{5} = 1$. Hence, the O.R. above is optimal. Similarly, our $O.R.$ of $\overline{K(n,r)}$ can be seen to be optimal.

Vector Colorings of Graphs

Definition 20. Given a graph G, we assign a unit vector to each vertex. This time, we would like adjacent verticies to be sent to vectors whose dot product is as negative as possible. If $\chi(G) = k$, then we can associate each color with a vector in the regular k-simplex in $\mathbb{R}^{k+1}$. Such vectors have inner product $\frac{-1}{k-1}$. In light of this, we define $\chi_{vec}(G) = \min\left\{k \mid v_i^T v_j = \frac{-1}{k-1}\right\}$

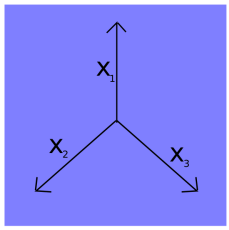
Theorem 21. $\chi_{vec}(G) = \vartheta\left(\overline{G}\right)$

This is proven by the fact that the two problems can be expressed as semidefinite-programming duals of one another. (The duality is between χ_{vec} and $\max_{\overline{O.R.}} \sum \left(\psi^T v_i\right)^2$.) Alternatively, there is a concrete way to move between optimal representations of the coloring problem and optimal representations for ϑ.

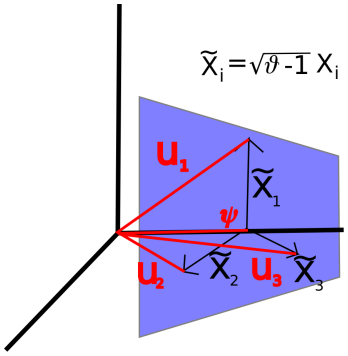
Proposition 22. *A vector coloring x_i is optimal (having value ϑ) iff $u_i = \frac{1}{\sqrt{\vartheta}}\left(\psi + \sqrt{\vartheta - 1}x_i\right)$ is an optimal Orthonormal representation (also having value ϑ).*

Proof. If x_i is an optimal vector-coloring for G with coloring number ϑ, and ψ is some unit vector orthogonal to each x_i, then we obtain an orthonormal representation $u_i = \frac{1}{\sqrt{\vartheta}}\left(\psi + \sqrt{\vartheta - 1}x_i\right)$. First, observe that these are all unit vectors. Secondly, let $u_i \neq u_j$ correspond to non-adjacent verticies in $\overline{G}$, so that $x_i^T x_j = \frac{-1}{\vartheta - 1}$. Then $u_i^T u_j = \frac{1}{\vartheta}\left(1 + (\vartheta - 1)\frac{-1}{\vartheta - 1}\right) = 0$, so u is an orthogonal representation of $\overline{G}$. Also, we have $\frac{1}{(\psi^T u_i)^2} = \vartheta$.

Conversely, if we start with the orthonormal representation with value ϑ (so $\vartheta = \frac{1}{\left(u_i^T \psi\right)^2}$ for all i. We have not yet shown that it's always possible to achieve equality, but it can be seen from the antiblocker picture.) we can recover the coloring by $x_i = \frac{\sqrt{\vartheta}u_i - \psi}{\sqrt{\vartheta - 1}}$. Now, if $x_i \sim x_j$ in G, then $x_i^T x_j = \frac{-\sqrt{\vartheta}u_i^T\psi - \sqrt{\vartheta}u_j^T\psi + 1}{\vartheta - 1} = \frac{-1}{\vartheta - 1}$ so the coloring has value ϑ. Also, $x_i^2 = \frac{(\vartheta + 1) - 2\sqrt{\vartheta}u_i \cdot \psi}{\vartheta - 1} = 1$. (There is a slight issue: if $u_i \cdot \psi = -\sqrt{\vartheta}$, we need to reassign $u_i \mapsto -u_i$.)



vector coloring of C_3



representation of $\overline{C}_3$(not normalized) □

Next, we will provide an optimal vector coloring of P. Assume a basis of size 5, and that we will map $\star\star\circ\circ\circ \mapsto (a, a, b, b, b)$, and extend this map by permutations of S_5. If x, y are two vector representations of intersecting sets, we would like to minimize

$$min_{a,b}\frac{x^T y}{\|x\|\,\|y\|} = \frac{4ab + b^2}{2a^2 + 3b^2}$$

The minumum occurs at $a = -3, b = 2$, and gives

$$\frac{x^T y}{\|x\|\,\|y\|} = \frac{-24 + 4}{18 + 12} = -\frac{2}{3} = \frac{-1}{\frac{5}{2} - 1}$$

Using numpy.linalg, we can find that these vectors span a space of dimension 4.

Finally, an optimal vector coloring of $\overline{P}$ can be found by assuming a basis of size 10 (the same basis we used for $\vartheta(P)$) and three variables, x_0, x_1, x_2. This gives us the optimization problem:

$$min_{x_0,x_1,x_2}\frac{x^T y}{\|x\|\,\|y\|} = \frac{x_0^2 + 3x_1^2 + 4x_0x_1 + 2x_1x_2}{3x_0^2 + 6x_1^2 + x_2^2}$$

The minimum (according to Wolfram) is found at $\left(-\frac{1}{\sqrt{18}}, \frac{1}{\sqrt{18}}, -\frac{1}{\sqrt{2}}\right)$ and gives

$$\frac{x^T y}{\|x\|\,\|y\|} = \frac{\frac{1}{18} + \frac{3}{18} - 4\frac{1}{18} - 2\frac{1}{6}}{3\frac{1}{18} + 6\frac{1}{18} + \frac{1}{2}} = \frac{-6}{18} = \frac{-1}{4-1}$$

This spans a space of dimension 5.

Problem 23. Since we have established that vector colorings correspond to orthonormal representations, we actually have 2 orthonormal representations of the Petersen Graph and 2 for its complement. Are these the same?

Quantum Measurements

We fix a finite-dimensional vector space, known as the state space. Pure quantum states will be given by unit vectors in this space.

> A projective measurement is described by an observable, M, a Hermitian operator on the state space of the system being observed. The observable has a spectral decomposition
>
> $$M = \sum_m m P_m$$
>
> where P_m is the projector onto the eigenspace of M with eigenvalue m. The possible outcomes of the measurement correspond to the eigenvalues, m, of the observable. Upon measuring the state $|\psi\rangle$, the probability of getting result m is given by
>
> $$\langle\psi| P_m |\psi\rangle$$
>
> Given that the outcome m occured, the state of the quantum system immediately after the measurement is $\frac{P_m|\psi\rangle}{\sqrt{p(m)}}$[12]

(!) Naimark's Dilation Theorem states that all quantum measurements can be viewed as projective measurements on a larger system. The Spectral Theorem states that a Hermitian operator M always has a decomposition of the form

$$M = \sum_m m P_m$$

where m ranges over real numbers, and the ranges of the P_m's are pairwise orthogonal. Conversely, if we start with a collection of pairwise orthogonal vectors $\{|m\rangle\}_{m=1}^n$which span the state space, we can create a Hermitian operator which has those vectors as its eigenvectors: $M = \sum_{m=1}^n m\,|m\rangle\langle m|$. Thus, when specifying a PVM, we only need to supply an Orthonormal Basis.

Example 24. Polarizing sunglasses are an example of a PVM.

Light consists of an electric wave inducing a magnetic wave and vice versa. The direction of the electric wave determines the polarization of the light. A photon can be polarized in any 2-dimensional direction. Polarized lenses in sunglasses will let a photon through if it is polarized vertically, and will block it if it is polarized horizontally. The photon hitting the lens is a measurement. The state space is 2-dimensional, corresponding to the polarization directions. We are lucky in this case that polarization between horizontal and vertical directions has a direct physical meaning- that the light is polarized in a diagonal direction. Usually, this is not the case. Let $|v\rangle$ be the state of light which is polarized in the vertical direction and $|h\rangle$be the state of light polarized in the horizontal direction. When a vertically polarized photon hits a vertically polarized lens, it will surely pass through. We

represent the observable for this lens by $L = 1 \cdot |v\rangle \langle v| + 0 \cdot |h\rangle \langle h|$. Suppose we have light that is polarized diagonally,

$$|\psi\rangle = \frac{1}{\sqrt{2}} |v\rangle + \frac{1}{\sqrt{2}} |h\rangle$$

The probability of the light passing through the lense is $(\langle v| + \langle h|) |v\rangle \langle v| (|v\rangle + |h\rangle) = \frac{1}{2}$, and if it does so, its state is $\frac{|v\rangle\langle v| \frac{1}{\sqrt{2}}(|v\rangle + |h\rangle))}{\sqrt{\frac{1}{2}}} = |v\rangle$. Observe that the probability that the light passes through the lens is the square of the inner product and the resultant state we're interested in. This is true in general.

Forward Example: The CHSH Inequality

There are two conceivable ways to use the machinery of ϑ to investigate contextuality. Usually we start with a collection of measurements and a linear functional on the outcomes of these measurements. By drawing the exclusion graph of the measurements, we can identify bounds (Bell Inequalities) on these linear functionals. We will see that these bounds are given by ϑ and α. Alternatively, we can go backwards and search for graphs which have a large gap between ϑ and α, then find orthonormal representations to realize these graphs as collections of quantum experiments. Ultimately, this is the direction that I would like to pursue, but we should start with the standard technique[5] first.

As an example of the forward method, we will investigate the CHSH inequality. Imagine Alice and Bob, separated in time and space and unable to communicate. However, they share qubits which make up a quantum state. That is, these particles may be entangled, so the total state space is 4-dimensional. Alice may choose between two measurements A and A', and Bob may also choose between two measurements B and B', so there are 4 total measurements which may occur: $\{AB, A'B, AB', A'B'\}$. Each of Alice and Bob's local measurements yield an outcome of 0 or 1, so each total measurement has 4 possible outcomes.

(One formulation of) the CHSH inequality begins with such a scenario and, if Alice and Bob both choose their experiments uniformly at random, and assigns a value of 1 whenever the their outcomes agree and A, B is not chosen, and a value of -1 whenever their outcomes disagree and A, B is not chosen. If A, B is chosen, then we reverse the valuations, so that 1 is assigned when their measurements disagree and -1 when the agree. The CHSH inequality concerns the expectation of such a valuation. To formalize this, let $E\left(A^{(\prime)}, B^{(\prime)}\right)$ be the probability that the outcomes agree, given that $A^{(\prime)}, B^{(\prime)}$ is measured. Then we have a value

$$S = -E(A, B) + E(A', B) + E(A, B') + E(A', B')$$

The CHSH inequality states that under classical assumptions

$$S \leq 2$$

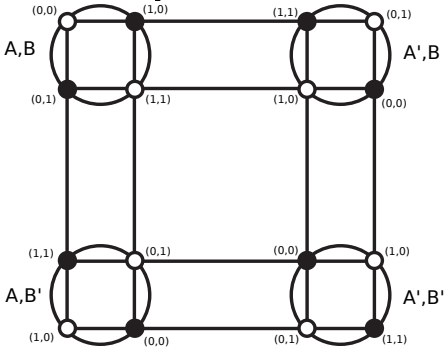
We will see where this bound comes from. Yet using quantum mechanics, we may achive a value of

$$S = 2\sqrt{2}$$

Concretely, the measurements which realize this bound are: Alice may either measure in the standard basis (A), or in the basis $\left\{\frac{|0\rangle + |1\rangle}{\sqrt{2}}, \frac{|0\rangle - |1\rangle}{\sqrt{2}}\right\}$ (A'). Bob may

either measure in the basis $\left\{\frac{\cos\left(\frac{\pi}{8}\right)|0\rangle-\sin\left(\frac{\pi}{8}\right)|1\rangle}{\sqrt{2}}, \frac{\sin\left(\frac{\pi}{8}\right)|0\rangle+\cos\left(\frac{\pi}{8}\right)|1\rangle}{\sqrt{2}}\right\}(B)$ and in the basis $\left\{\frac{\cos\left(\frac{\pi}{8}\right)|0\rangle+\sin\left(\frac{\pi}{8}\right)|1\rangle}{\sqrt{2}}, \frac{-\sin\left(\frac{\pi}{8}\right)|0\rangle+\cos\left(\frac{\pi}{8}\right)|1\rangle}{\sqrt{2}}\right\}$ (B'). (!) (TODO: check that this is accurate.)

To derive these bounds we first draw the exclusion graph for this scenario. The verticies of the graph are the 16 total outcomes which may occur (or, if you like, pairs of measurements and outcomes). Edges are drawn between vertices which cannot cooccur. For example, there is an edge between $(0,0 \mid AB)$ and $(1,0 \mid AB')$ because it is impossible that the local measurement A gives both 1 and 0.



Once the exclusion graph, H, is drawn, we obtain a weighting on the vertices which is determined by our valuation. For example, whenever the outcome $(1,1 \mid AB')$ occurs, we add 1 to our counter, and whenever $(0,0 \mid AB)$ occurs, we subtract 1 from our counter. Thus, each vertex gets a weight $w : V(G) \to \mathbb{R}$. We can insist that these weights be positive by adding 1 to all the experiments so that our final weighting corresponds to giving $w(v) = 2$ for the black vertices and $w(v) = 0$ for the white vertices. Typically, and in this case, all of the weights will be the same, except for those which are 0. Our weighting therefore identifies a subgraph, H'

To derive the classical bounds, observe that any classical state of the system must assign outcomes to all measurements, even those which have not occured. This means that a classical state will correspond to an independent set in the graph. It is intuitive (and easy to show, since S is a linear functional) S is maximized at a particular classical state rather than a mixture of them. The maximum value for S, classically, will correspond to an independent set and this independent set will restrict to an independent set of H'. Since $\alpha(H') = 3$, we can recover the classical bound on S by undoing our manipulations to the weighting.

$$S \leq_{classical} 2 \cdot 3 - 4 = 2$$

With respect to quantum mechanics, each of our outcomes corresponds to an eigenvector of an observable. Thus, each vertex in H' receives a vector. If the two outcomes associated with these vectors are exclusive, then the vectors must be orthogonal. Otherwise, it would be possible to measure the system, obtain one outcome, then measure the system again and obtain the other, exclusive outcome.

Therefore, the eigenvectors of the observables form an orthogonal representation of $\overline{H'}$. Recall our dual definition for $\vartheta(G) = \max_{\overline{OR}} \sum_i c(v_i)$ is exactly the

maximum expected value in the quantum setting. Thus,

$$S \leq 2 \cdot \vartheta(G) - 4 = 2 \cdot \left(2 + \sqrt{2}\right) - 4 = 2\sqrt{2}$$

In this case, the bound is tight. This is not always the case, because the dimension of the optimal orthonormal representation may be too large.

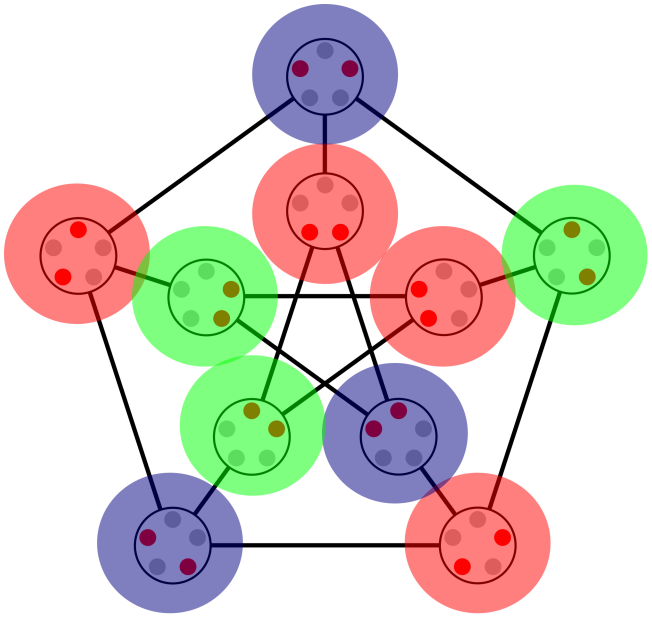
From Orthonormal Representations to Quantum Circuits

Next, we would like to sketch how we might find collections of quantum circuits which exhibit large amounts of contextuality. This procedure comes with many unanswered questions.

(1) Choose an optimal coloring of the graph.
(2) Produce an optimal orthonormal representation for this graph.
(3) Each color corresponds to a collection of pairwise orthonormal vectors. Extend each collection to an orthonormal basis.
(4) Find the unitary transformations from the given bases to the standard (computational) basis.
(5) All unitary transformations can be implemented by quantum gates.
(6) Our circuits consist of these unitary transformations, followed by measurement in the computational basis.

For a quantum circuit which takes some input state, ψ, applies a unitary transformation, U, then measures in the computational basis, the possible outcomes will be the elements of the computational basis. If x is one such basis element, the probability of measuring x is $\langle U\psi, x\rangle^2 = \langle \psi, U^\star x\rangle^2$. It is easier to think of the unitary transformations as moving the computational basis than moving the state. We will think of these quantum circuits as collections of bases under which to measure a particular state.

Since any two orthogonal vectors can be extended to a basis and we can perform a measurement (which will reveal one outcome) in that basis, two orthogonal vectors represent incompatible outcomes.



If we perform these steps, we will arive at $\chi(G)$ quantum circuits whose statistics on the input state ψ cannot be explained classically. Specifically, the results corresponding to vectors in our original Orthonormal Representation (prior to extending to a basis) will occur more often than is possible classically. The dimension of the orthonormal representation, d, is exactly the dimension of the hilbert space in which the quantum circuit lives, so we will need $\lceil \log_2 d \rceil$ qubits to implement such a circuit.

Some questions must be raised:

(1) Does the initial coloring matter?
(2) Can we effectively find the O.R. in step 2?
(3) How can we control the dimension of the optimal orthogonal representation? Can we use a non-optimal representation in a smaller dimension? (I have a specific construction, due to Lovasz in mind.)
(4) How can we control the circuit complexity of our resultant unitaries?
(5) Does it matter how extend our colored sets to bases?
(6) Can any of these circuits be achieved using only Clifford gates? Can we find a result linking the contextual resources (such as magic states) used to form the unitaries, and the final contextuality?

Graph Theory and Optimization

If we have an exclusivity graph G, we can consider the probabilities which may be assigned to vectors $\mathbb{R}_+^{V(G)}$. This identification also maps subgraphs of G to their incidence vectors. Recall that independent sets of our graph correspond to the classical, deterministic probablility distribution. The collection of classical distributions is therefore described by convex sums of our independent sets. The set of classical states is given by the "vertex polytope," VP

$$VP \coloneqq CvxHull(\{vec_\alpha \mid \alpha \text{ is independent}\})$$

We may express α (we will omit the argument G in the notation) as an optimization problem over this set. Specifically, α may be defined by an integer linear program. We begin with variables $\{x_i \in \{0,1\}\}_{i=1}^{V(G)}$ which are intepreted as: $x_i = 1$ when v_i is in the indendent set in question. Our program is then:

$$\alpha = \max_x \left(\sum_{i=1}^{V(G)} x_i \right)$$

such that $x_i \in \{0,1\}$ and

$$x_i + x_j \leq 1, \text{ for } v_i \sim v_j$$

Equivalently, we could write our constraint as $x \in VP$.

This problem is intractible, so we might formulate a linear program by allowing $x_i \in [0,1]$ to take on a continum of values. The resulting problem is tractible, but is too relaxed. For example, $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ is a solution to the triangle. Thus, we choose a different generalization of the constraints:

$$\alpha^\star = \max_x \left(\sum_{i=1}^{V(G)} x_i \right)$$

such that $x_i \in [0,1]$ and

$$\sum_{i \in K} x_i \leq 1 \text{ for all cliques } K$$

In this problem, our feasible set is the "fractional vertex packing polytope"

$$FVP = \left\{ x \mid \sum_{i \in K} x_i \leq 1, \forall \text{ cliques } K \right\}$$

This program is important enough to get a name: $\alpha^\star$, the fractional packing number. Moreover, the $\alpha^\star \neq \alpha$, since, for the 5-cycle we have the solution $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$.

The fractional packing number has a nice interpretation: If our variables x_i correspond to probabilities, then our constraint enforces that the sum of probabilities of pairwise exclusive events be bounded by 1, which is implied by one of Kolmogorov's Axioms for probability. General probabilistic models have their Bell Inequalities bounded by $\alpha^\star$ rather than ϑ.

This problem is NP-hard because there may be exponentially many cliques. For perfect graphs, however, $\alpha = \alpha^\star$ and both are polynomial-time computable.

Theorem 25. $FVP\left(\overline{G}\right) = AB\left(VP\left(G\right)\right)$.

Suppose $x \in FVP\left(\overline{G}\right)$. If we show that $x^T v_\alpha \leq 1$, for all indepdent sets α, we will have shown that $FVP\left(\overline{G}\right) \subset AB\left(VP\left(G\right)\right)$. Since α is an independent set of G, it is also a clique of $\overline{G}$, so $\sum_{a \in \alpha} x_a + \sum_{b \in G - \alpha} x_b \cdot 0 \leq 1 + 0 = 1$. The argument works conversely as well: If $x^T y \leq 1$ for all $y \in VP\left(G\right)$, then $x^T v_\alpha \leq 1$ for all independent α. Again, since these independent α in G become cliques ω in $\overline{G}$, our constraints precisely name FVP.

Linear Programming Duality

Every linear program comes with a dual program with the same optimum. The linear program

$$\text{Max}_x \left\{ c^T x \mid Ax \leq b, x \geq 0 \right\}$$

comes with the dual

$$\text{Min}_y \left\{ b^T y \mid A^T y \geq c, y \geq 0 \right\}$$

If the primal problem asks: "Given my ingredients and recipies for baking various cakes and the profits for each type of cake, how much of each cake should I make in order to maximize my profit? And what is my final profit?" Then the dual problem asks "What is the minimum amount of money I would accept to sell my ingredients, given that I could use them to make cakes and later sell these cakes? And how much money would I make from doing this." In this formulation, it is intuitive that the primal and dual problems will have the same optimal.

Another way to think about the dual program that it provides upper bounds for the primal problem. In this perspective, the new variables y refer to different ways to linearly combine our constraint equation to achive(on the left side) a linear functionals $A^T y$ which is an upper bound on our primal functional, c. Meanwhile, on the right side of the inequality, we get $b^T y$, so this is an upper bound[3].

A third way to think about the dual problem is via Lagrange Multipliers. When optimizing a function over a feasible set, at the optimal solution, the direction in which the function increases the fastest increase must be parallel to the normal of the feasible set. The level sets of our objective function are hyperplanes, and our feasible set is a polytope. Clearly, maxima will occur at a vertex of the polytope. At this point, we will be able to combine the normals (which define the facets adjacent to the vertex) linearly, with non-negative coefficients to arrive at the normal of the objective function.

$$L(x, \lambda) = c^T x - \sum_i \lambda_i a_i^T x$$

In the dual formulation of $\alpha^\star$, we want to minimize the sum of cliques, such that the sum of cliques over any single vertex is at least one. Let $\{y_i \in [0,1]\}_N$, where N is the total number of cliques in the graph, and each y_i corresponds to a clique.

$$\alpha^\star = q^\star = \min_y \left(\sum_{i=1}^{N} y_i \right)$$

under the constraints $y_i \in [0,1]$. For every vertex v,

$$\sum_{y_i \ni v} y_i \geq 1$$

Observe that if we form the associated integer linear program by restricting $y_i \in \{0,1\}$, then we will have the clique covering problem, so $\alpha^\star$ may also be also called $q^\star$. Thus, α and q are an example of dual integer linear programs which do not agree.

ϑ as a Semidefinite Program

Semidefinite programs are "vector relaxations." Instead of integers or real numbers, our variables range over vectors. In the case of ϑ, these are our Orthonormal Representations. These collections of vectors are in perfect correspondence with positive semidefinite matricies. A symmetric matrix $m \in \mathbb{R}^n \times \mathbb{R}^n$ is called positive semidefinite if $\langle x, mx \rangle \geq 0$ for all $x \in \mathbb{R}^n$. If this is the case, then we can define a new inner product as $(x, y) = \langle x, my \rangle$, sans positive-definiteness. To look at the entries of m, we can set x and y to be members of the basis. The ij^{th} element of m will be $m_{ij} = \langle e_i, me_j \rangle = (e_i, e_j)$. In other words, the matrix m tells us how our new inner product behaves on our basis elements. The argument also works conversely, so positive semidefinite matricies are in $1-1$ correspondence with gram matricies.

A semidefinite program is of the form

$$Min_X (C \cdot X) = Min_X (TR(CX))$$

such that X satisfies

$$X \succeq 0, \{D_i \cdot X = p_i\}_{i=1}^k$$

In such a problem, we will be supplied with the k $n \times n$ symmetric matricies C and $\{D_i\}$, along with real numbers $\{p_i\}_{i=1}^k$.

The notation $X \succeq 0$ states that X be positive semidefinite. As in the linear programming case, our objective function is some linear function of the variables. We are allowed linear equality constraints on the entries of X, which are in effect constraints of linear combinations of dot products of our vector assignments.

Rather than supplying the linear equality constraints $\{D_i \cdot X = p_i\}_{i=1}^k$, we could express the feasible set directly, in the form $X(y) = G_0 + \sum_{i=1}^m (y_i G_i)$ ard requiring that $X(y) \succeq 0$. Set $d_i = C \cdot G_i$. We obtain one equivalent formulation:

$$Min_y d^T y$$

such that

$$X(y) \succeq 0$$

Theorem 26. *We can express ϑ as a semidefinite program.*

Proof. Consider the following program, which computes the vector chromatic number:

$$\frac{-1}{\vartheta - 1} = \min_m (\alpha)$$

such that $m \succeq 0$ and

$$m_{ij} = \alpha \text{ if } v_i \sim v_j$$

$$m_{ii} = 1$$

It's clear that the program computes the vector chromatic number, but we have to justify that it is a semidefinite optimization problem, as defined above. In particular, the constraint $m_{ij} = \alpha$ is problematic, since α is our objective function. We can replace these constraints with a set of constraints which requires that all m_{ij}

where $v_i \sim v_j$ be equal. To do this, order the edges $e_1, e_2, \ldots, e_{E(G)}$ arbitrarily, and require that $m_{e_i} = m_{e_{i+1}}$. Rather than minimizing α, we can minimize m_{e_1}. Then we arrive at an equivalent problem, which is clearly semidefinite. Let $y \in \mathbb{R}^{k+1}$, and $d = (1, 0, 0, \ldots)$ Then

$$\frac{-1}{\vartheta - 1} = \min\left(d^T y\right)$$

subject to

$$I + y_1 A + \sum_{i=2}^{k} y_i D_i = M(y) \succeq 0$$

Here, J is the matrix of all 1's, $A \in \mathbb{R}^{n \times n}$ is the adjacency matrix of the graph, which which has the i, j^{th} entry 1 when $v_i \sim v_j$, and 0 otherwise. D_i's are the collection of matricies with 1 entry of 1 in the i, j^{th} entry where $v_i \not\sim v_j$.

Observe that if we scale the solution to the vector coloring problem so that each vector has length $\sqrt{\vartheta - 1}$, then this will imply that the dot products of adjacent vectors are -1. And conversely, if we have a solution to the scaled problem we can recover a solution to the original problem. Hence,

$$\vartheta - 1 = \min\left(\frac{1}{n} I \cdot Z\right)$$

subject to the constraints that

$$Z_{11} = Z_{22} = \cdots = Z_{n,n};\ Z_{ij} = -1 \text{ if } v_i \sim v_j; Z \succeq 0$$

□

Definition 27. Suppose we have semidefinite program in the second formulation. Its dual is defined[16] as

$$Max_Z\left(-G_0 \cdot Z\right)$$

such that

$$Z \succeq 0, \{G_i \cdot Z = d_i\}_{i=1}^{m}$$

Just as the linear programming duality arises from finding bounds on the optimal values for the primal program, the duality in semidefinite programming also arises in this way, since

$$d^T y + Z \cdot G_0 = \sum_{i=1}^{m} (Z \cdot G_i) y_i + Z \cdot G_0 = Z \cdot G(y) \geq 0$$

so

$$d^T y \geq -G_0 \cdot Z$$

In fact, we have equality, assuming that there is a positive definite feasible matrix in either the primal or dual problem. We can see that this occurs exactly when $Z \cdot G(y) = 0$. (!) This is the antiblocker theorem.

Theorem 28. *We can express ϑ as a semidefinite program*[9]*:*

$$\vartheta = \max (J \cdot B)$$

subject to

$$b_{ij} = 0 \text{ if } v_i \sim v_j; Tr(B) = 1$$

Where J is the 1's matrix (so $J \cdot B$) is the sum of entries in B.

Proof. We will take the dual of the semidefinite program is Theorem 27.

$$\frac{-1}{\vartheta - 1} = \max_Z (-I \cdot Z) = \max_Z (-Tr(Z))$$

subject to the constraints

$$A \cdot Z = 1, \{D_i Z = 0\}_{i=1}^k$$

The second set of constraints implies that Z_{ij} is 0 whenever $v_i \sim v_j$. It follows that $J \cdot Z = (I + A) \cdot Z = Tr(Z) + A \cdot Z$. Thus, for the optimal $Z^\star$, $J \cdot Z = 1 + \frac{1}{\vartheta - 1} = \frac{\vartheta}{\vartheta - 1}$. Thus, if we scale up all the entries in Z by $\vartheta - 1$, we arrive at the desired semidefinite program. □

Remark 29. The feasible set B is related to the set of Orthonormal Representations in the *max* definition of ϑ. Given an O.R., one can compute its gram matrix and obtain a PSD matrix C such that $Tr(C) = n$. It is not obvious how to interpret the objective function, $J \cdot B$ as relating to $\sum_{i \in V(G)} \left(\psi^T v_i\right)^2$. Lovasz proves[9] our max definition from the SDP. In the course of his proof, he appears(!) to show that the optimal handle is always $\psi^\star = \frac{\sum v_i}{|\sum v_i|}$.

A consequence of the 'max' semidefinite program (!)

Our first, 'min' semidefinite program resulted from our ability to phrase the min definition of ϑ in a way that got rid of ψ and resulted in a linear objective function. This was the vector coloring problem. It would be nice to be able to remove ψ from the max definition and arrive at the max semidefinite program through an arrangement of vectors, but there are difficulties. The following inequality comes from [9], and is a nice application of Cauchy Schwarz. Let w_i be some rescaling of v_i such that $\sum_i w_i^2 = 1$.

$$\left(\sum_{i=1}^n \left(\psi^T v_i\right)^2\right) = \left(\sum_{i=1}^n w_i^2\right)\left(\sum_{i=1}^n \left(\psi^T v_i\right)^2\right) \geq \left(\sum_{i=1}^n |w_i| \psi^T v_i\right)^2 = \left(\sum_{i=1}^n \psi^T w_i\right)^2 \leq \left(\sum_{i=1}^n w_i\right)^2 = \sum_{i,j} w_i^T w_j$$

The first inequality can be made to hold if we select our scaling such that $w_i^2 \propto \left(\psi^T v_i\right)^2$. The second inequality can be made into equality if we select ψ to be a unit vector in the direction of $\sum_{i=1}^n w_i$. Mysteriously, we can make both of these true at the same time.

The Ellipsoid Method: Convex Optimization is Polynomial Time

There is a general technique to optimize any linear objective function over any convex feasible set. This technique is slow in practice, but runs in polynomial time. In particular, suggests that semidefinite optimization problems may be solved efficiently. A correct analysis of the technique is complicated because we need to take rounding errors into account. Since the optimal may not have a terminating decimal representation, we need to allow solutions which only approximate that optimal value up to ϵ, and we need to show that this approximation is polynomial in $\log\left(\frac{1}{\epsilon}\right)$. Furthermore, we need to take into account rounding errors in the representation convex set itself. The following description comes from [14]

In the following, a convex body will be modeled as an oracle which, given a point, will either assert that the point is in the convex body, or else it will return a separating hyperplane between the point and the body.

First, suppose that we have some convex body C and we wish to optimize $\max_{x\in C}\left(a^T x\right)$ with $\|a\| = 1$ Suppose we have a function $sample\left(C\right)$ which takes a convex body and returns a point in that set or correctly asserts that C is empty. We must also assume that C contains a ball $b_{0,r}$ and is contained in a ball $B_{0,R}$. Then we can solve the optimization problem by means of a binary search. We know that $0 \leq \max_x \left\{a^T x \mid x \in C\right\} \leq R$. We can sample x_0 and calculate $a^T x_0$. Then, we consider $C_1 = C \cap \left\{x \mid a^T x \geq \frac{a^T x_0 + R}{2}\right\}$. We can sample again, and if we find a point, x_1, we known that $\max_{x\in C}\left(a^T x\right) \geq a^T x_1$. Otherwise, we know $\max_{x\in C}\left(a^T x\right) \leq a^T x_1$. In either case, we have halved the range of possible values for $\max_{x\in C}\left(a^T x\right)$, so it only takes one more iteration of the algorithm to calculate one more bit of the output. In other words, it is polynomial-time.

Thus, if we can find points inside a convex body, C, we can optimize over C in polynomial time. The ellipsoid method is a technique to find a point inside a convex body. If we know nothing about C, we are left guessing at points and the problem is hopeless. We must assume we are told our convex body is contained in a ball $S\left(0, B\right)$.

The ellipsoid method works via ellipsoids whose volumes shrink by constant factors. The first ellipsoid is the ball $S\left(0, B\right)$. Then, we guess a point at the center of the ellipsoid, 0. If this fails, the oracle will return a separating hyperplane to us. Call the halfspace containing C, H. We construct a new ellipsoid, which is the smallest ellipsoid containing $S\left(0, B\right) \cap H$. Generally, we guess at the center of the current ellipsoid, and if we're wrong, use the resulting hyperplane to construct an new ellipsoid with less volume by a constant (depending on the dimension) factor. These ellipsoids do not work in practice, but a simple (if you ignore precision) way to argue that optimization over a convex body can be achieved in polynomial time.

The ellipsoid method requires a separation oracle. For a semidefinite programming problem, we must be able to decide whether or not a point is in the feasible set, and if not, then we should be able to compute a separating hyperplane. Specifically, we need to be able to determine whether a symmetric matrix is positive semidefinite or not. First, we diagonalize the matrix is $O\left(n^3\right)$ time, thereby writing $M = UDU^T$ where U is orthonormal and D is diagonal. M is positive semidefinite exactly when D has non-negative entries. If D has a negative entry, then we can find a vector c such that $U^T c$ is the basis vector associated with that entry, and

$c^T UDU^T c = c^T Mc < 0$. Since $c^T Mc = \sum c_i c_j m_{ij}$, this is a linear functional on M which separates M from the positive cone.

Interior Point Methods

Linear programming problems are usually solved by the simplex method, where the vertices of the feasible polytope are examined systematically. Our semidefinite programs do not have polytopes as feasible sets, so there would be infinitely many vertices to check. Instead, we use interior point methods, where we search the interior of the feasible set. But first we rephrase the problem to contain both the primal and dual problems.

$$min\left(\left(c^T y\right) + TR\left(F_0 Z\right)\right) = F\left(y\right) \cdot Z$$

subject to the fact that $F\left(y\right)$ and Z are primal and dual feasible respectively.

We know by the strong duality theorem for linear programming that this minimum is actually 0. By finding the location where this occurs, we can find the optimal values $F\left(y\right)$ and Z.

To do this, we assign a barrier function which is infinite at the boundary of both feasible sets.

$$\phi\left(Z\right) = \begin{cases} \log\left(\det Z^{-1}\right) & \text{if } Z > 0 \\ \infty & else \end{cases}$$

This function is convex (why?), so it has a unique minimum, known as the analytic center. We can find this analytic center using Newton's Method[15]:

(1) Start with an arbitrary point, x. (in our case, this is some feasible positive definite matrix Z.)
(2) Calculate the tangent line for which the second order approximation for f is minimized (it happens to be$-\nabla^2 f\left(x\right)^{-1} \nabla f\left(x\right)$)
(3) Find the point x' along this tangent line for which $f\left(x'\right)$. This is just a 1-dimensional search and can be performed efficiently.

Newton's Method has good convergence results when f is self-concordant: $\left|f'''\left(x\right)\right| \leq 2f''\left(x\right)^{\frac{3}{2}}$, and our barrier function is such a function.

The analytic center of the primal and dual feasible sets can be extended to a "central path" to the optimal points[16]. Specifically, for each $\gamma > 0$, we consider the set $F\left(y\right) > 0$, and $d^T y = \gamma$. This is a subset of the feasible set, and for γ near the optimum, it is nonempty. Therefore, we can calculate its analytic center. Doing so, we obtain a curve parameterized by γ of feasible matrices $F\left(y\right)$ which leads to the optimal feasible matrix.

Proposition 30. *If $F\left(y\right)$ is on the central path, there will be a corresponding dual feasible matrix Z on the dual central path.*

Proof. [16] If $F\left(y\right)$ is on the central path, with parameter γ, then $F\left(y\right)$ is the solution to the optimization

$$y^\star = argmin\left(\log\left(\det F\left(y\right)\right)^{-1}\right)$$

such that $F\left(y\right) \succeq 0$ and $c^T y = \gamma$. We can ignore the first condition, because our objective function is ∞ when it is violated. The method of Lagrange Multipliers

says that a convex function can only be minimized when the gradient of the objective function is parallel to the constraint. To perform partial derivative $\frac{\partial}{\partial y_y}$ on the objective function, we need to use Jacobi's Formula. (! wikipedia) In particular, for invertible A we have $\frac{d}{dt}\det A(t) = Tr\left(\det(A) A^{-1}\frac{dA(t)}{dt}\right)$. In our case, this yeilds

$$\frac{1}{\det F(y^\star)^{-1}} Tr\left(\det F(y^\star)^{-1} F(y^\star)\frac{dF(y)^{-1}}{dy}\right) = Tr\left(F(y^\star)\frac{dF(y)^{-1}}{dy_i}\right) = Tr\left(F(y^\star)^{-1} F_i\right) = \lambda c_i$$

Or, in other words, $\frac{F(y)^{-1}}{\lambda}$ is dual-feasible. It's clear that the duality gap between $Tr\left(\frac{F(y^\star)^{-1}}{\lambda}\cdot F(y^\star)\right) = \frac{n}{\lambda}$. Also, its claimed (! the proof appears to be missing from [16]) that $\frac{F(y^\star)^{-1}}{\lambda}$ is on the central path. Thus, the two central paths can be paremerized by the dual gap, and the corresponding points are near-inverses. □

Next, we introduce a distance function to serve as a penalty for deviating from the central path.

$$\psi(y, Z) \coloneqq -\log\det(F(y) Z) + \log\det F(y^\star) Z^\star = -\log\det F(y) Z + n\log Tr(F(y) Z) - n\log n$$

This can actually be rewritten independently of $y^\star$ by using the fact that $F(y^\star) Z^\star = \frac{\eta}{n} I$.

Finally, we introduce a potential which takes into account ψ, the deviation from the central path and η, the duality gap.

$$\phi \coloneqq \nu\sqrt{n}\log(F(y)\cdot Z) + \psi(y, Z)$$

We can apply a modified Newton's Method to minimize ϕ. First, we need a direction to search. This is the time-limiting step.

Kochen-Specker Sets

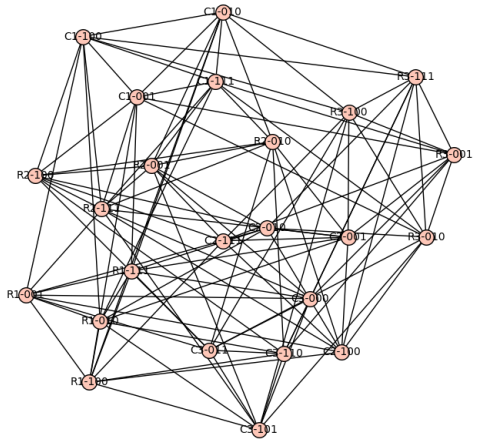
It is possible to establish contextuality without probabilities, and in a way that is independent of the quantum state.

IZ	ZI	ZZ	I
XI	IX	XX	I
XZ	ZX	YY	I
I	I	I	

Mermin's square is a special arrangement of observables on pairs of qubits which provides an example of a state-independent proof of contextuality. In fact, there are no satisfying assignments to the observables at all. In the diagram, the pauli operators X, Y, Z satisfy $X^2 = Y^2 = Z^2 = I$ and $XY = iZ$. Each row and column consists of 3 commuting observables, and the products of each row and column are I, except for the last column, where the product is $-I$. This means that any row

and any column can be measured and that the product of the outcomes is 1 for each row/column, except -1 for the last column.

We should draw the exclusivity graph for this well-studied example.



Here, we have 6 contexts, $R1, R2, R3, C1, C2, C3$ and each context has 4 possible outcomes. For example, for the node $R1 - 001$, the '$R1$' indicates that we chose to measure the first row, $\{IZ, ZI, ZZ\}$, simultaneously. The '001' means that IZ measured -1, ZI measured -1 and ZZ measured 1.

This graph was created by adding edges where experiments disagree. For example, there is an edge between $R3-010$ and $C2-100$ because the measurement ZX is $+1$ in the first node and -1 in the second. After a measurement, the state vectors will be orthogonal to one another, because the first will be in the positve eigenspace of ZX and the second will be in the negative. Hence, two 'incompatible' vectors will be orthogonal. The graph should really be drawn according to orthogonalities, so we need to be careful and consider the reverse: Do all orthogonalities come from observables? Generally, it's true that there is some observable to distinguish between any two orthogonal states: We can define an observable by its positive and negative eigenvectors- which are the ones we want to distinguish. In the Mermin Square example, it happens that there are no more orthogonality relations than the ones found by disagreements in measurement, but this is not necessarily the case. If we dropped YY from the measurements, so that $C3 = \{ZZ, XX\}$ and $R3 = \{XZ, ZX\}$, then the outcome $C3 - 11$ would be orthogonal to $R3 - 11$, we would have no measurement in our list on which these two disagree. Therefore, I stress that one must actually compute eigenvectors and their orthogonalities to draw these graphs properly. The script in grenerateGraphFromObservables.py does this.

The computer (see the file MerminSquareGraph.py) indicates that

$$\vartheta(G) = 6.0, \alpha(G) = 5$$

The interpretation of this is slightly different than in the CHSH example. In that example, we did not have a concrete set of observables, and were instead trying to find a bound over a class of them which satisfied certain relations. Here, the Mermin square defines a specific set of observables. In the CHSH example, we had to take a subset which corresponds to a weighting of the nodes. Here, our graph is the vectors of the experiment themselves.

The fact that $\alpha(G) < q(G) = 6$, indicates that there is no independent set which can meet every clique. We've said that classical states (global assignments of outcomes) correspond to independent sets. Each context corresponds to a clique. Thus, there is no assignment to the observables which fits into every context. In other words, we have strong contextuality.

In fact, it appears that you can always detect (strong) contextuality in this way: A section corresponds to a vertex. A global section corresponds to an independent set that meets each clique. If the vertex of a section section can be extended to an independent set which meets ever clique, that section can be extended to a global section, and conversely as well.

Given a measurement scenario, our graphs are precisely the complement of the graph whose vertices are the sections over each context, and whose edges are compatibility relations. If we draw these graphs, and calculate independence numbers, it seems that we can completely solve the problem of detecting contextuality. This is implemented in generateGraphFromObservables.py

Since we no longer have an optimization problem, ϑ is less useful here. The value $\vartheta(G) = 6$ indicates this configuration of vectors in 4 dimensions is optimal. Recall that $\vartheta(G)$ can be regarded as the maximum sum of probabilities at each vertex over all possible orthonormal representations. Our observables' eigenvectors form an orthonormal representation whose sum of probabilities is 6, since we are guaranteed a successful measurement in each of the 6 contexts.

It would be interesting to see what solution the semidefinite program returns. This program should return the positive semidefinite matrix B whose normalized Cholesky decomposition gives the orthonormal representation, and the handle is the weighted sum of cholesky vectors. The maximum of a linear functional over a convex set occurs at the boundary. Thus, I would expect that $det(B) = 0$, which implies that the cholesky vectors should never be full dimension.

The Relationship Between Orthogonality Graphs and Possibilistic Empirical Models

Contextuality has also been studied via cohomology[2]. In this setup, we begin with a measurement scenario: $\langle X, M, O\rangle$, where X is a collection of abstract measurements (not necessarily maximal), $M \subset 2^X$, a collection of contexts, which define those measurements which are assumed to be co-measureable, and O, a collection of outcomes. We assume that M has a simplicial structure (a sub-collection of co-measureable measurements must again be co-measureable). This setup is quite general and does not assume that the measurements X have any particular structure, such as that they are quantum measurements defined by hermitian matricies.

We then define a sheaf of events: $\mathscr{E} : \left(2^X\right)^{op} \to SET$, where $\mathscr{E}(U) \coloneqq O^U$[6]. For every collection of measurements, the sheaf of events describes the possible outcomes for these measurements. There is no restriction to these outcomes, and we only define the presheaf of events so that we can define the possibilistic empirical models. These define the possible outcomes for our model in the same way that the orthogonality graph describes which outcomes may occur.

A possibilistic empirical model is a sub presheaf of the sheaf of events satisfying

(1) $\mathscr{S}(C) \neq \emptyset \forall C \in M$

(2) The restriction map $\rho_U^{U'} \coloneqq \mathscr{S}(U') \to \mathscr{S}(U') :: s \mapsto s\mid_U$ is surjective when $U \subset U' \subset C$ for some $C \in M$.
(3) Every family $\{s_C \in \mathscr{S}(C)\}_{C \in M}$ such that $s_C \mid_{C \cap C'} = s_{C'} \mid_{C \cap C'}$ for all $C, C' \in M$ induces a global section in $\mathscr{S}(X)$.

The fact that $\mathscr{S}$ is a subpresheaf of $\mathscr{E}$ indicates that the possible assignments for experiments defined by $\mathscr{S}$ are indeed assignments to the set of outcomes. Condition 1) ensures that every context has at least one satisfying assignment. 2) ensures that an assignment within a context can be restricted to a sub-assignment. We only assume that this is possible within a context, because a context consists of a collection of measurements which can be performed simultaneously without affecting their outcomes. Therefore, an assignment to a collection $U \subset C$ can always be extended to an assignment of outcomes on C by performing the rest of the measurements. 3) insists that every assignment to contexts which agrees on their overlap can be extended to a global section. In other words, any assignments to all measurements X which agree on each context is assumed to actually provide a global section. This section must be unique, since $\mathscr{S} \subset \mathscr{E}$.

The question of logical contextuality is "When is a section over a context extendible to a global section?" By assumption 3), this is equivalent to asking whether or not there are compatible sections for every other context. We may also ask for strong logical contextuality: "Are there any global assignments?" There are various results relating to these questions which rely on cohomological invariants, but there appears to be a much simpler solution.

Definition 31. Given a measurement scenario $\langle X, M, O\rangle$ and a possibilistic empirical model for that measurement scenario, define the graph $\mathscr{G}$ to have vertices $\{s_C \in \mathscr{S}(C)\}_{C \in M}$. Let $a \sim b$ when $\forall C$, $a \mid_C = b \mid_C$ (in other words, the sections are compatible.)

Each vertex of $\mathscr{G}$ is defined to provide a an assignment of outcomes to a particular context. Each of the vertices in our orthogonality graph, G, also provides an assignment of outcomes to a particular context. The vertices of $\mathscr{G}$ are adjacent according to compatibility conditions, whereas the vertices in our graph are adjacent according to incompatibility conditions. Due to the technicality described in the previous section, it is possible that there will be more orthogonality conditions than incompatibility conditions, but by adding additional measurements we can always ensure that the two agree. Thus, if our measurement scenario is provided with explicit measurements X in terms of Hermitian matrices, we can calculate G and conclude that $\overline{G} = \mathscr{G}$. If our measurement scenario is not given according to explicit measurements, we have still defined $\mathscr{G}$.

Proposition 32. *Our empirical model will be strongly contextual iff* $\omega(\mathscr{G}) \neq |M|$.

Proof. Since no two sections over a particular context C are compatible, $\omega(\mathscr{G}) \leq |M|$. Suppose that we have a global section, S. Then $\forall C$, $S \mid_C$ is a section over C, and thus corresponds to a vertex of $\mathscr{G}$. Moreover, for any two contexts C, C', we must have $(S \mid_C) \mid_{C'} = (S \mid_{C'}) \mid_C = S \mid_{C \cap C'}$, which means that the corresponding vertices are adjacent. Hence, S induces a clique of size M, and $\omega(\mathscr{G}) = |M|$.

Conversely, if $\omega(\mathscr{G}) = |M|$, then we have a clique of size $|M|$. The vertices of this clique correspond to sections for each context, and since every pair of vertices in the clique is adjacent, every pair of sections is compatible. Due to assumption 3) of empirical models, there is a global section. □

Similarly, we can detect logical contextuality with respect to a section by determining whether or not the vertex in $\mathscr{G}$ associated with that section is part of a clique of size $|M|$.

State-Independent Contextuality

The Peres-Mermin square does not have any satisfying assignments. Thus, its classical set trivially obeys any Bell Inequality. A weaker form of state-independent contextuality (henceforth state-independent contextuality) occurs when a single Bell Inequality separates all quantum states from all classical ones. This stands in contrast to logical contextuality. Since the quantum set and classical sets are convex bodies, there is a separating hyperplane between the two. This hyperplane corresponds to a Bell Inequality. Thus, we do not gain a more general notion of state-independent contextuality by allowing the inequality to vary with the state.

This weaker form of contextuality can be recognized.

Theorem 33. [4] *State Independent Contextuality is equivalent to: We can assign non-negative numbers $w = (w_1, w_2, \dots)$ to the vertices of our orthogonality graph, such that*

$$\sum_{j \in \alpha} w_j < 1 \text{ for all } \alpha \text{ and } \sum_v w_v \Pi_v \geq I$$

where Π_i is the projector associated with the vertex i. α runs over the set of independent sets which meet every context. If this is the case, the functional in our inequality is given by the weightings w_i.

Proof. Suppose we have such an assignment of the weights. Then any classical assignment of outcomes corresponds to an independent set, where the sum of weights is strictly bounded by 1 by fiat. For any state vector ψ, the weighted sum of probabilities is lower bounded

$$\sum_v w_v \langle\psi| \Pi_v |\psi\rangle = \langle\psi| \left(\sum_v w_v \Pi_v \right) |\psi\rangle \geq \langle\psi| I |\psi\rangle = 1$$

Thus, we have state-independent contextuality.

Conversely, suppose we have some functional which separates all classical states from all quantum states. Such a functional corresponds to a corresponds to a weighting of the vertices w_i of non-negative numbers such that for some $A \in \mathbb{R}$

$$\sum_{j \in \alpha} w_j < A \text{ and } \sum_v w_v \langle\psi| \Pi_v |\psi\rangle \geq A$$

If we rescale our weights, we can set $A = 1$

$$\sum_{j \in \alpha} w_j < 1 \text{ and } \sum_v w_v \langle\psi| \Pi_v |\psi\rangle \geq 1 = \langle\psi| I |\psi\rangle$$

Now we re-arrange the second condition.

$$0 \leq \sum_v w_v \langle\psi| \Pi_v |\psi\rangle - \langle\psi| I |\psi\rangle = \langle\psi| \left(\sum_v w_v \Pi_v - I \right) |\psi\rangle$$

and holds occurs for all ψ, in particular for any eigenvectors of $(\sum_v w_v \Pi_v - I)$. Thus, $(\sum_v w_v \Pi_v - I)$ is positive semidefinite and we have $\sum_v w_v \Pi_v \geq I$. □

It is claimed[4] that this necessary and sufficient condition can be verified by a semidefinite program. However, it's a little unsatisfying because it is relies on eigenvectors and is not fully combinatorial. There is a simple necessary combinatorial condition[13][4].

Proposition 34. *(!!)In order to be state-independent contextual, the underlying orthogonality graph must satisfy*

$$\omega^{\star}(G) > d$$

where $\omega^{\star}(G) = \alpha^{\star}\left(\overline{G}\right)$ *is the fractional chromatic number and* d *is the dimension.*

Proof. If we have state-independent contextuality, then such contextuality must hold for the maximally mixed state, which assigns a probability of $\frac{1}{d}$ to every vector. Thus, we need this probability distribution to be outside of the classical set before we can hope to have state-independent contextuality. The condition above is a clever way to check if this distribution is classical or not.

The antiblocker theorem (Theorem 25) between $VP(G)$ and $FVP(G)$ states that a point $x \in VP(G)$ iff $\forall y \in FVP\left(\overline{G}\right)$, $x^T y \leq 1$, thus, we would like to calculate $\max_y x^T y$, where y ranges over $FVP\left(\overline{G}\right) = VP(G)$. This exactly defines $\alpha^{\star}\left(\overline{G}; x\right) = \omega^{\star}(G; x)$, where x is now the weighting. In particular, we can choose x to be the constant weighting where each vertex is weighted with $\frac{1}{d}$. Since everything is linear, we can conclude that $\frac{1}{d}\omega^{\star}(G) \leq 1$ iff the constant probability $\frac{1}{d}$ is in $VP(G)$. □

The proof above contains a big hole: it shows that the maximally mixed state provides a probability distribution which is outside of $VP(G)$ iff $\omega^{\star}(G) > d$. However, $VP(G)$ is defined as the convex hull of independent sets, not the convex hull of independent sets which meet every context. Thus, it is possible that the maximally mixed state might be part of $VP(G)$, yet outside the set of classical assignments. It's plausible that when the maximally mixed state is part of $VP(G)$, the quantum set and the classical set will necessarily interesct at some point, but this needs proof. Before any kind of proof, it is clear that our framework needs refinement. Since we have been requiring that our probabilities (both classical and quantum) sum to exactly 1 in each context (rather than being bounded by 1), we need tools to explicitly represent the contexts which are considered. Such tools are provided in the hypergraph approach[1].

The Hypergraph Approach

Our orthogonality graphs only keep track of outcomes and compatibilities of those outcomes. There is no mention of which outcomes actually form a context and need to have probabilities summing to 1. In most of our cases, the 'outcomes' can be considered to be 1-dimensional projectors corresponding to the simultaneous eigenvectors of the collections of commuting observables.

An alternative approach[1] makes explicit use of the fact that probabilities sum to exactly 1. In this hypergraph approach, our measurements correspond to hyperedges whose vertices are outcomes.

Definition 35. A hypergraph consists of a finite set of vertices, V (which correspond to outcomes), together with a collection of hyperedges E, $H \subset 2^V$, which

correspond to our contexts. We assume that these hyperedges form an antichain: for any two $h_1, h_2 \in H$, we have $h_1 \not\subset h_2$ and that $\bigcup_{e \in E} = V$. (These are also called clutters)

If our measurements consist of 1-d projectors, then these hyperedges will each contain d vertices, where d is the dimension of the space. However, we do not assume any particular structure for the measurements so that the framework is general.

First, observe that a graph can be recovered from any hypergraph by defining

Definition 36. Given a hypergraph H, $Or(H)$ is the graph with vertices $V(H)$ and edges $w \sim v$ iff $w, v \subset e \in E(H)$ for some hyperedge e.

The graph defined in this way is an orthogonality graph is our usual sense: if two vertices are contained within the same hyperedge, then this means that these two vertices correspond to different outcomes for the same context. If we perform a quantum measurement corresponding to this hyperedge, then the outcomes of that measurement will correspond to the eigenspaces of our measurment, and different outcomes will be in orthogonal subspaces. If we are given specific observables to serve as the hyperedges then we might find more orthogonalities. The type of orthogonality that comes automatically from disagreeing tests is called "local orthogonality,"

Just as we have defined $FVP(G)$, $TH(G)$, and $VP(G)$ for orthogonality graphs which capture the general, quantum, and classical probability distributions on outcomes, we can also define three classes of probability distributions for hypergraphs, along with charaterizations of these bodies which follow from the Antiblocker property.

Our original use of the convex bodies FVP, TH, VP was to derive Bell Inequalities. The graphs given as input were actually subgraphs of the graph of the system we were interested in. Thus, we allowed that the probabilities be subnormalized, because the vertices which weren't part of our subgraph could hide some of the probability. In the hypergraph approach, we require that each measurement be normalized.

Definition 37. A probabilistic model on a hypergraph H consists of an assignment $p : V(H) \to [0,1]$ such that $\sum_{v \in e} p(v) = 1$.

One tradeoff is that there are now hypergraphs which do not admit any probabilistic models.

Proposition 38. *For a hypergraph H, probabilistic model p is extremal exactly when there is a subhypergraph $W \subset H$ such that W admits a unique probabilistic model p_W, which extends to p by setting $p(v) = 0$ when $v \notin W$.*

Proof. By induction. The base case is when H itself has a unique probabilistic model, and in this case there is nothing to prove.

Otherwise, probabilistic models, by definition, satisfy

$$p(v) \geq 0, \text{ and } \sum_{v \in e} p(v) = 1 \ \forall e \in E(H)$$

This is a polytope, so extreme points will be extreme points of facets. A facet will consist of those probability distributions where one of the constraining inequalities

is met. Hence, there is some vertex v such that $p(v) = 0$. Define $W^1 = H - \{v\}$ and recurse. □

We do not assume that these models obey the exclusivity principle. Their summations are only normalized over hyperedges, and $Or(H)$ may include cliques which do not correspond to edges. The omission of the exclusivity principle in[1] was deliberate so as to allow two hypergraphs H_1 and H_2 to be tensored and so that this tensor product would admit the tensor of the probabilistic models on each of H_1 and H_2. The scheme for tensoring these two hypergraphs is complicated and is omitted here. The point of doing this is to explore non-locality, which we might revisit later.

Proposition 39. *Given a hypergraph H, a probabilistic model $p : V(H) \to [0,1]$ satisfies the exclusivity principle iff $\omega(Or(H); p) = 1$.*

Proof. First, suppose that we have a probabilistic model, p which satisfies the exclusivity principle. The exclusivity principle states $\omega(Or(H); p) \leq 1$. Any hyperedge, e, will yield a clique in $Or(H)$ for which $\sum_{v \in e} p(v) = 1$, so $\omega(Or(H); p) = 1$.

Conversely, suppose that $\omega(Or(H); p) = 1$. Then $\omega(Or(H); p) \leq 1$, and the exclusivity principle is satisfied. □

For us, satisfying the exclusivity principle defines what is meant by a general probabilistic model. The probabilities p for which $\omega(Or(H); p)$ can be interpreted with the antiblocker property. These are the points such that scaling up by any amount would remove p from $FVP(Or(H))$. These are the points on the boundary of $FVP(Or(H))$ and are also probabilistic models. It is not clear whether or not these 'general probabilistic models' in this sense are convex.

Definition 40. A probabilistic model is deterministic if $p(v) \in \{0,1\}$ for all $v \in V(H)$. A probabilistic model is classical if it is a convex combination of deterministic models.

In constrast to the deterministic models for graphs, our definition of probabilistic models requires that the probabilities within every context sum to 1. This captures that the deterministic models are not only independent sets of $Or(H)$ but also that these independent sets must meet every clique.

Proposition 41. *A probabilistic model p is classical iff $\omega^\star(Or(H); p) = 1$.*

Proof. Since p is a probabilistic model, we have $\omega^\star(Or(H); p) \geq \omega(Or(H); p) \geq 1$, since any hyperedge determines a clique whose sum of probabilities is 1.

Recall that the feasible set of $\omega^\star(G)$ is $FVP(\overline{G})$. The antiblocker property says that for any graph G, $FVP(G) = AB(VP(\overline{G}))$. If $p \in VP(Or(H))$, so for any $q \in FVP\left(\overline{Or(H)}\right)$, $p \cdot q \leq 1$. Since $\omega^\star(Or(H), p)$ is defined to be $\max_{q \in FVP\left(\overline{Or(H)}\right)} q \cdot p$, we have $\omega^\star(Or(H), p) \leq 1$, so $\omega^\star(Or(H), p) = 1$.

Conversely, suppose that $\omega^\star(Or(H); p) = 1$. Then, $\forall q \in FVP\left(\overline{Or(H)}\right)$, $q \cdot p \leq 1$, so $p \in AB\left(FVP\left(\overline{Or(H)}\right)\right) = VP(Or(H))$. In other words, p can be written as a convex combination of independent sets. Each of these independent sets must be maximal. If $p = x + p'$, where x is an independent set which does not meet an edge e_x and p' consists of the remaining terms of a decomposition of p into

independent sets, then $\sum_{v\in e_x} p(v) = \sum_{v\in e_x} p'(v) < 1$, which contradicts that p was a probabilistic model. □

The condition described by the previous proposition indicates that the set of classical probability models consists of a particular facet of $VP(G)$, bounded by the vertices which are independent sets which meet every context.

The case of $TH(G)$ is interesting. It would be natural to define the quantum models by analogy with the previous two propositions:

Definition 42. A probabilistic model $p \in Q_1(H)$ if $\vartheta\left(\left(\overline{Or(H)}, p\right)\right) = 1$[1].

However, we will se that $Q_1(H)$ is larger than the quantum set:

Definition 43. The quantum set, Q, is the collection of probabilities $p : V(H) \to [0,1]$ such that there exists a Hilbert space $\mathscr{H}$, a quantum state $\rho \in \mathscr{B}_{+,1}(\mathscr{H})$ and a projection operator $P_v \in \mathscr{B}(\mathscr{H})$ associated to every $v \in V$ which are projective measurments:

$$\sum_{v\in e} P_v = I_{\mathscr{H}}$$

and reproduce the given probabilities $p(v)$ according to $p(v) = Tr(\rho P_v)$.

There are several distinctions between Q_1 and the quantum set. In Q_1, we require that our projectors be 1-d, and we also require that the Hilbert Space be finite dimensional. These are both conditions which might suggest that $Q_1 \subset Q$, but in fact $Q \subset Q_1$. In the quantum set, we require that the projectors sum to $I_{\mathscr{H}}$. For $p \in Q_1$, we have $\sum_{v\in e} p(v) = 1$ (since p is a probabilistic model) for some particular state ψ (the handle), but the projectors will not sum to the identity unless the dimension of the Orthonormal Representation is the size of each context. It might be possible to extend the 1-d projectors to higher-dimensional projectors which agree with the original projectors for ψ, but these projections will need to be compatible with one another. Thus, the quantum set has a more stringent requirement.

Unlike the set of classical and general probability sets, the quantum set cannot be defined by the underlying orthogonality graph.

Proposition 44. *There exists a hypergraph H for which $Q(H) \neq Q_1(H)$*

Proof. (Not really a proof) [10] We can characterize the quantum set of $B_{2,2,2}$ by the angle between the two choices of measurement for one of the parties. There is a particular functional which separates Q_1 from Q by direct computation. □

It would be nice to have a good explanation for why the "almost quantum set" Q_1 differs from Q. However, according to [10] Q_1 satisfies most information principles which have been conjectured to characterize Q. This makes it difficult to find a high-level reason for the discrepancy between Q_1 and Q. It would be nice to have some sort of general recipe for generating probabilities which lie in $Q_1 - Q$.

Corollary 45. *There are two hypergraphs, H_1 and H_2 with the same underlying orthogonality graph, but different quantum sets. That is, the quantum set is not determined by the underlying orthogonality graph.*

TODO: collect the proof from [1].

The NPA Hierarchy

The Quantum Set is mysterious- its definition does not suggest an algorithm for determining whether or not a probabilistic model is quantum. However, it is simple to certify that a probabilistic model is quantum by providing the projectors which give rise to the desired probabilities. While it is conjectured that membership in the quantum is not decidable[1], there is a collection of semidefinite programs which completely determines it[11].

Suppose that we have a probabilistic model which is actually quantum. According to Definition 43, there are projectors which can be associated to each vertex, such that projectors P_v in any hyperedge are mutually orthogonal and sum to I, and a state $|\psi\rangle$ which produces the desired probabilities. Consider the vectors of the form $\{(\prod_{i=1}^{n} P_{v_i})|\psi\rangle \cup |\psi\rangle\}$ where n is finite. We can form a semidefinite matrix, Γ^n by taking the Gram matrix for these vectors. By its construction Γ^n is positive semidefinite. Also, if the Hilbert Space has dimension $N < n$, then no more than N of our vectors are linearly independent. Hence, the rank of Γ^n will be at most N.

Any edge of the hypergraph must give its vectors a decomposition of I into projectors. This enforces constraints on our projectors P_v, which, in turn, are constraints on the entries of Γ^n. If $v_1 \sim v_2$, then they must correspond to orthogonal projectors. Hence, for any product of projectors S, we will have $P_{v_1} S|\psi\rangle \perp P_{v_2} S|\psi\rangle$.

Any the projectors associated with any edge in the hypergraph must sum to I. This translates to the fact that for any product of projector S, we have $\sum_{v\in E} P_v P_S|\psi\rangle = P_S|\psi\rangle$.

Putting our observations in terms of Γ^n,

(1) $\Gamma^n \succeq 0$
(2) $\Gamma^n_{\emptyset,\emptyset} = 1$
(3) $(\forall s, w) \sum_{v\in E} \Gamma^n_{sv,w} = \Gamma^n_{s,w}$
(4) If $v = v_1 \dots v_k$ and $w = w_1 \dots w_m$, with $v_k \sim w_m$, then $\Gamma^n_{v,w} = 0$
(5) $\Gamma^n_{\emptyset,v} = Pr(v)$

Thus, if our probabilistic model, p, is actually quantum, then for any n, we can find a certificate, Γ^n, which satisfies properties 1-5. The NPA hierarchy at level n, Q_n, is the set of probabilistic models such that there exists a certificate Γ^n which satisfies 1-5. We will see that the notation agrees with definition 42, and that $Q = Q_\infty$. First, we note that the certificates Γ^n really do form a hierarchy.

Proposition 46. *Given a hypergraph H, $Q(H) \subset \dots \subset Q_n(H) \subset \dots \subset Q_1(H)$.*

Proof. We have already argued that $\forall n$, $Q(H) \subset Q_n(H)$, since if our model is really quantum, then we can construct Γ^n as the Gram matrix of vectors of the form $\{(\prod_{i=1}^{n} P_{v_i})|\psi\rangle \cup |\psi\rangle\}$. Suppose that we have a probabilistic model $p \in Q_n(H)$. Then we have a certificate $\overline{\Gamma^n}$ which certifies p. By restricting this matrix to entries involving only $n-1$ projectors, we obtain a Q_{n-1} certificate for Γ^{n-1}. □

Theorem 47. *The NPA hierarchy exactly characterizes the quantum set. $\bigcap_{n=1}^{\infty} Q_n(H) = Q(H)$.*

Proof. [sketch] The previous proposition implies that $Q(H) \subset \bigcap_{n=1}^{\infty} Q_n(H)$. To show the other direction, suppose that p is a probabilistic model in $\bigcap_{n=1}^{\infty} Q_n(H)$.

Thus, p posesses certificates $\overline{\Gamma^n}$ for each n. The entries of each certificate are bounded, since they come from inner products of projections of unit vectors. Each certificate can be viewed as a vector in the l_∞ norm, if we complete the non-existant entries to 0. The Banach-Alaoglu theorem states that the unit ball is compact in the weak-$\star$ topology, so our sequence of certificates has a convergent subsequence. A convergent subsequence in the weak-$\star$ topology will converge when any functional $\rho : v \to \mathbb{R}$ is applied. In particular, if ρ is returns a single component, we see that convergence in the weak-$\star$ topology implies pointwise convergence. Hence, we have an infinite certificate Γ^∞ which satisfies 2-5, and any submatrix formed by restriction is positive semidefinite. (!) By taking limits, we can see that Γ^∞ itself is positive semidefinite.

If the behavior of p is quantum, then the vertices of H form a $C^\star$-algebra. The infinite certificate, Γ^∞ defines a state on this $C^\star$ algebra, and the GNS construction provides a way of converting the $C^\star$-algebra into operators on a Hilbert Space, and the state into a vector in this Hilbert Space.

In order to use the GNS construction, we must argue that Γ^∞ really does define a state on the $C^\star$-algebra. Suppose that X is in this $C^\star$-algebra, so that X can be written as a linear combination of products of vertices in the hypergraph. The state, ψ_{Γ^∞}, must be shown to be a positive linear functional of norm 1. We define $\psi_{\Gamma^\infty}\left(P_{v_1}P_{v_2}\dots P_{v_n}\right)$ by looking up $\Gamma^\infty_{\emptyset,P_{v_1}\dots P_{v_n}}$ and we extend this by linearity. The norm is 1, $\psi_{\Gamma^\infty}\left(I\right) = \Gamma^\infty_{\emptyset,\emptyset} = 1$, by 2). To see that ψ_{Γ^∞} is positive, we must show that $\psi_{\Gamma^\infty}\left(X^\star X\right) \geq 0$. This follows from the fact that Γ^∞ is positive semidefinite, since if we write X as a linear combination of the vertices (projectors, or generators of our $C^\star$-algebra), $X = \sum_{i=1}^n \left(x_k \prod_i v_{i\in I_k}\right)$, then $\psi_{\Gamma^\infty}\left(X^\star X\right) = \left\langle x\right| \Gamma^\infty \left|x\right\rangle$, where $\left|x\right\rangle = \sum_{k=1}^n x_k \left|\prod_{i\in I_k} v_i\right\rangle$ □

The certificates Γ^n are not a semidefinite program, since they do not involve an optimization. They are convex bodies defined by a semidefinite condition, so they are feasible sets of a semidefinite program. Checking whether a particular probabilistic model lies in the n^{th} level of the hierarchy amounts to checking whether or not this feasible set is empty, which can be achieved by the semidefinite program:

$$\max\left(\lambda\right)$$

subject to

$$\Gamma^n - \lambda I \succeq 0$$

where Γ^n is now assumed to satisfy 2-5. If it can satisfy 1) as well, then $\lambda = 0$ is a solution. Otherwise, the feasible set is empty.

Lemma 48. *The rank of a gram matrix is the dimension of the span of the vectors in its decomposition.*

Proof. We use the fact that the rank of a gram matrix is the dimension of the space spanned by the vectors of this gram matrix. First, suppose that the vectors are all linearly independent. Then the Cholesky decomposition $\Gamma = L^T L$ shows that the Γ is non-singular when L is non-singular, and L will be non-singular if its vectors are all linearly independent.

Now, suppose that our vectors consist of $\{a_1, \dots, a_n, b_1, \dots, b_m\}$, where the a_i's are all linearly independent and the b_i's may be written as a linear combination of the a_i's. The first row of the matrix will be $(a_1 \cdot a_1, a_1 \cdot a_2, \dots a_1 \cdot a_n, a_1 \cdot b_1, \dots a_1 \cdot b_m)$,

and the $n+1^{th}$ row will be $(b_1 \cdot a_1, b_1 \cdot a_2, b_1, \ldots b_1 \cdot a_n, b_1 \cdot b_1 \ldots b_1 \cdot b_m)$. By assumption, we can write b_1 as a linear combination of the a_i's, and thus we can write the $n+1^{th}$ row as a linear combination of the previous rows. Thus, the rows greater than n do not contribute to the rank of Γ. □

Lemma 49. *Suppose we have a finite collection of projectors $\{P_i\}_{i=1}^n$ such that $\sum_{i=1}^n P_i = I$. Then for $i \neq j$, $P_i P_j = 0$.*

Proof. Since $\sum_{i=1}^n P_i = I$, conjugation by P_j gives $P_j \left(\sum_{i=1}^n P_i\right) P_j = I$. Hence, $0 = P_j \left(\sum_{i\neq j}^n P_i\right) P_j = \sum_{i\neq j}^n P_j P_i P_j = \sum_{i\neq j} (P_i P_j)^\star (P_i P_j)$. The sum of positive operators can only be 0 if each of them is 0. Hence, $(P_i P_j)^\star (P_i P_j) = 0$ for all $i \neq j$. So $P_i P_j = 0$. □

Theorem 50. *(Rank-Loop condition) Suppose we have a probabilistic model p. There exists two certificates Γ^n and Γ^{n+1} such that Γ^{n+1} restricts to Γ and Rank$\left(\Gamma^{n+1}\right)$ =Rank(Γ^n) iff p is the behavior of a finite-dimensional quantum system. In this case, the dimension of the system exceeds this rank.*

Proof. Suppose first that our system is finite dimensional of dimension d. Then the vectors from which we form our Gram matrix span a space of at most d. The lemma implies that the gram matrix formed by these vectors has rank at most d.

Conversely, suppose that we have two certificates Γ^n and Γ^{n+1} which agree on the $n \times n$ submatrix, and which have the same rank. Our Hilbert Space will be the finite dimensional space spanned by $H = Span\left(\prod_{j=1}^n P_j |\psi\rangle\right)$, and we will define projectors $P_i \coloneqq Proj\left(Span\left(\left\{P_i S \mid S = \prod_{j=1}^n P_j |\psi\rangle\right\}\right)\right)$. Clearly, the P_i's are projectors, and the dimension of the space is Rank(Γ^n). To complete the proof, we must verify Definition 43. We must establish that for any edge e, $\sum_{v\in e} P_v = I$. Our axioms for certificates (namely 3.)) allows us to conclude something very similar: $\sum_{v\in e} S^\dagger P_v^\dagger W = S^\dagger W$ when W, S are of the form $\prod_{j=1}^n P_j |\psi\rangle$. And we have defined these vectors to span our Hilbert Space. By linearlity, we see that for any vectors $\phi_1, \phi_2 \in H$, we have $\sum_{v\in e} \langle\phi_1| P_v^\dagger |\phi_2\rangle = \langle\phi_1| |\phi_2\rangle$. This shows that $\sum_{v\in e} P_v^\dagger = I$.

Finally, I'd like to point out that we will recover the desired probabilities. Let P_v be any projector as defined above. Choose any edge e so $\sum_{r\in e} P_r = I$ by the previous argument. Then, denoting by e^n the collection of sequences of length n of projectors in e, we obtain

$$\langle\psi| P_v |\psi\rangle = \langle\psi| P_v I^n |\psi\rangle = \langle\psi| P_v \left(\sum_{r\in e} P_r\right)^n |\psi\rangle = \sum_{r\in e^n} \langle\psi| P_v \left(\prod_{i=1}^n P_{r_i}\right) |\psi\rangle = \sum_{r\in e^n} \Gamma^{n+1}_{vr,\emptyset} = \Gamma^{n+1}_{v,\emptyset}$$

Here, we have used axiom 3.) n times in the last equality. Finally, axiom 5.) completes the argument. □

Remark 51. A finer condition for stopping is given in [11].

References

[1] A. Leverrier A. Antonio, T. Fritz and A. Sainz. A combinatorial approach to nonlocality and contextuality. 2015.

[2] S. Abramsky, R. Barbosa, K. Kishida, L. Raymond, and S. Mansfield. Contextuality, cohomology and paradox. 1998.

[3] A. Ben-Tal and A. Nemirovski. Lectures on modern convex optimization.

[4] A. Cabello. Simple explanation of the quantum limits of genuine n-body nonlocality. 2010.

[5] A. Cabello, S. Severini, and A. Winter. Graph-theoretic approach to quantum correlations. 2014.
[6] Giovanni Caru. On the cohomology of contextuality. 2017.
[7] M. Grotschel, L. Lovasz, and A. Schrijver. Relaxations of vertex packing. *Journal of Combinatorial Theory*, SeriesB 40:330–343, 1986.
[8] D. Knuth. The sandwich theorem. *The Electronic Journal of Combinatorics*, 1994.
[9] L. Lovasz. On the shannon capacity of a graph. 1978.
[10] M. Hoban M. Navascues, Y. Guryanova and A. Acin. Almost quantum correlations. 2014.
[11] S. Pironio M. Navascues and A. Acin. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. 2008.
[12] M. Nielsen and I Chuang. *Quantum Computation and Quantum Information*. 2010.
[13] R. Ramanathan and H. Pawel. Cecessary and sufficient condition for state-independent contextual measurment scenarios. 2014.
[14] Bruce Reed. A gentle introduction to semi-definite programming. In J. Alfonsin and B. Reed, editors, *Perfect Graphs*, chapter 11, pages 233–259. 2001.
[15] L. Vandenberghe S. Boyd. *Convex Optimization*. 2004.
[16] L. Vandenberghe and S. Body. Semidefinite programming. 1996.
[17] WolframMathworld. Kneser graph. Accessed 2018-15-10.