CS286.2 Lecture 15: Tsirelson's characterization of XOR games

Scribe: Zeyu Guo

We first recall the notion of quantum multi-player games: a quantum k-player game involves a verifier V and k players $P_1, \ldots, P_k$. The verifier randomly picks an index i according to some distribution π over the set $\{1, \ldots, Q\}$ and sends a quantum state $|\varphi_i\rangle$ (the question) to the the players. Here $|\varphi_i\rangle$ consists of k quantum registers, $|\varphi_i\rangle = |\varphi_i\rangle_{P_1 \cdots P_k}$, but it is not necessarily a product state. Each player P_j applies a unitary operator $\mathcal{U}_j$ on his register P_j, as well as his ancilla qubits, and sends part of the resulting state back to the verifier. The verifier then performs a measurement $\mathcal{M}_i = \{\mathrm{Acc}_i, \mathrm{Rej}_i\}$ on the answers and decides to accept or reject according to the output.

The *value* $w(G)$ of a game G is defined to be the maximum probability that the verifier accepts, ranging over all possible strategies of players (i.e. all unitary operators $\mathcal{U}_j$). Here the ancilla that each player uses is initially set to $|0\rangle$. The *entangled value* $w^\star(G)$ is defined in the same way, except that the players have the freedom to use an arbitrary (and possibly entangled across all of them) initial ancilla state.

The classical PCP theorem can be interpreted as a statement that approximating $w(G)$ within some constant factor is NP-hard, even for 2-player games with a poly-size question set and constant-size answers. Similarly, the game variant of the quantum PCP conjecture asks: is it QMA-hard to approximate $w^\star(G)$ within a constant factor?

However, we note that even the NP-hardness of approximating $w^\star(G)$ is not obvious. This is because the usual reduction from CSPs to games breaks down when there is entanglement, as we have seen in the example of the Magic Square game. So we need to find another reduction.

In this lecture, we study *XOR games*, a simple class of games for which no such reduction exists (unless NP = P): for XOR games, $w(G)$ is NP-hard to approximate, but there exist polynomial time algorithms computing $w^\star(G)$!

Definition 1 (XOR game)**.** *An XOR game is a 2-player classical game (i.e. questions and answers are all classical) in which:*

- *questions are (s, t) chosen from $\{0, \ldots, n-1\}^2$ according to some distribution π;*
- *answers are bits $a, b \in \{0, 1\}$;*
- *the verifier's predicate $V(a, b|s, t) = f_{s,t}(a \oplus b)$ only depends on $a \oplus b$, where $\oplus$ denotes the XOR operation.*

We will see two famous examples of XOR games. The first example, the CHSH game, is one of the most famous games in quantum information theory.

Example 2 (CHSH game)**.** *Here $n = 2$ and $s, t \in \{0, 1\}$ are independent random bits. And we let $V(a, b|s, t) = 1$ iff $a \oplus b = s \wedge t$.*

It is not hard to show that $w(G) = 3/4$: both players just answer zero and this is the best they can do. However, CHSH (Clauser, Horne, Shimony, Holt) observed that $w^\star(G) \geq \cos^2(\frac{\pi}{8}) = 0.85\ldots$, achieved by the following strategy:

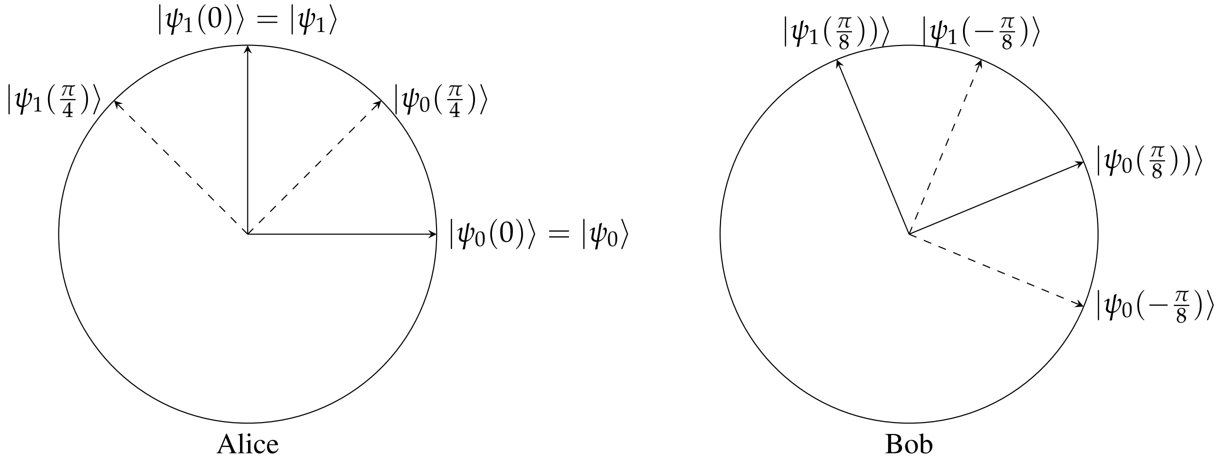
$|\psi_1(0)\rangle = |\psi_1\rangle$
$|\psi_1(\frac{\pi}{4})\rangle$
$|\psi_0(\frac{\pi}{4})\rangle$
$|\psi_0(0)\rangle = |\psi_0\rangle$
Alice
$|\psi_1(\frac{\pi}{8}))\rangle$
$|\psi_1(-\frac{\pi}{8})\rangle$
$|\psi_0(\frac{\pi}{8}))\rangle$
$|\psi_0(-\frac{\pi}{8})\rangle$
Bob

Figure 1: Alice (resp. Bob) measures using the solid arrows if $s = 0$ (resp. $t = 0$), and the dashed arrows if $s = 1$ (resp. $t = 1$). She/he answers bit i if the outcome corresponds to $|\psi_i(\theta)\rangle$ for the appropriate angle θ.

Consider an EPR pair $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B)$. For $\theta \in [-\pi, \pi]$, define

$$|\psi_0(\theta)\rangle = \cos(\theta)|0\rangle + \sin(\theta)|1\rangle$$
$$|\psi_1(\theta)\rangle = -\sin(\theta)|0\rangle + \cos(\theta)|1\rangle.$$

That is, $|\psi_0(\theta)\rangle$ (resp. $|\psi_1(\theta)\rangle$) is the rotation of $|0\rangle$ (resp. $|1\rangle$) by an angle of θ in the real plane spanned by $|0\rangle$ and $|1\rangle$.

Alice measures her qubit of $|\psi\rangle$ in the basis $\{|\psi_0(0)\rangle, |\psi_1(0)\rangle\}$ if $s = 0$, and in $\{|\psi_0(\frac{\pi}{4})\rangle, |\psi_1(\frac{\pi}{4})\rangle\}$ if $s = 1$. She answers the bit i if the outcome corresponds to $|\psi_i(0)\rangle$ (resp. $|\psi_i(\frac{\pi}{4})\rangle$). Bob measures his qubit in the basis $\{|\psi_0(\frac{\pi}{8})\rangle, |\psi_1(\frac{\pi}{8})\rangle\}$ if $s = 0$, and in $\{|\psi_0(-\frac{\pi}{8})\rangle, |\psi_1(-\frac{\pi}{8})\rangle\}$ if $s = 1$. He answers the bit i if the outcome corresponds to $|\psi_i(\frac{\pi}{8})\rangle$ (resp. $|\psi_i(-\frac{\pi}{8}\rangle)$). See Figure 1 for an illustration.

Next we calculate the winning probability. We could assume that Alice measures first, as the two measurements commute. Consider the case $s = 0$ and $t = 0$ where Alice and Bob are supposed to give the same answer. Alice measures her qubit and collapses $|\psi\rangle$ to $|\psi'\rangle_A|\psi'\rangle_B$, where $|\psi'\rangle = c|\psi_i(0)\rangle$ for some scalar c, $|c| = 1$, $i \in \{0,1\}$, and Alice answers bit i. Then the probability that Bob gets the correct outcome i is $|\langle\psi_i(\frac{\pi}{4})|\psi'\rangle|^2 = |\langle\psi_i(\frac{\pi}{4})|\psi_i(0)\rangle|^2 = \cos^2(\frac{\pi}{8})$. It is easy to check in the same way that Alice and Bob win with probability exactly $\cos^2(\frac{\pi}{8})$ as well in all the other cases. Overall the winning probability is $\cos^2(\frac{\pi}{8})$.

Tsirelson showed that this is indeed the best strategy, and hence $w^\star(G)$ is exactly $\cos^2(\frac{\pi}{8})$. We will see the proof later.

The next example is the MAXCUT game, one of the most famous games in CS.

Example 3 (MAXCUT game)**.** *Let $\mathcal{G} = \mathcal{G}(V, E)$ be an undirected graph. The verifier picks $(i, j) \in E$ at random, chooses (s, t) among (i, j), (j, i), (i, i), (j, j) with equal probability $1/4$, and sends s and t to Alice and Bob respectively. The verifier accepts the answer (a, b) from Alice and Bob iff one of the following is true:*

- *(s, t) is chosen to be (i, j) or (j, i), and $a \oplus b = 1$;*
- *(s, t) is chosen to be (i, i) or (j, j), and $a \oplus b = 0$.*

We first consider the classical game value $w(G)$. For $a = a(s), b = b(t) \in \{0,1\}$, we write $x_s = (-1)^a, y_t = (-1)^b \in \{-1,1\}$ and use $(x_s)_{s\in V}, (y_t)_{t\in V}$ to represent Alice and Bob's strategy. Then

$$\begin{aligned} w(G) &= \max_{(x_s)_{s\in V},(y_t)_{t\in V}} \mathop{\mathbb{E}}_{(i,j)\in E} \left[\frac{1}{4}\left(\frac{1-x_iy_j}{2} + \frac{1-x_jy_i}{2} + \frac{1+x_iy_i}{2} + \frac{1+x_jy_j}{2}\right)\right] \\ &= \frac{1}{2} + \max_{(x_s)_{s\in V},(y_t)_{t\in V}} \mathop{\mathbb{E}}_{(i,j)\in E} \left[\frac{(x_i-x_j)(y_i-y_j)}{8}\right] \\ &\le \frac{1}{2} + \frac{1}{2}\max_{(x_s)_{s\in V}} \mathop{\mathbb{E}}_{(i,j)\in E} \left[\left(\frac{x_i-x_j}{2}\right)^2\right] && \text{(Cauchy-Schwarz)} \\ &= \frac{1}{2} + \frac{1}{2}\cdot\frac{\#\text{MAXCUT}}{\#\text{EDGES}}. \end{aligned}$$

The last step follows by noting that $\left(\frac{x_i-x_j}{2}\right)^2$ contributes 1 if $x_i \neq x_j$, and 0 otherwise, and hence it counts the number of edges in the cut defined by $(x_s)_{s\in V}$. Also note that the inequality in the second to last step is in fact an equality. We conclude that

$$w(G) = \frac{1}{2} + \frac{1}{2}\cdot\frac{\#\text{MAXCUT}}{\#\text{EDGES}}.$$

But we know that MAXCUT is NP-hard. And the PCP theorem implies that even approximating MAXCUT is NP-hard. So we have

Corollary 4. *$w(G)$ is NP-hard to approximate within a factor $1+c$ for some constant $c > 0$.*

What about $w^\star(G)$? We will see that it is easy to *exactly* compute $w^\star(G)$ by solving semi-definite programs! For this we need to first establish an important characterization of $w^\star(G)$ for general XOR games.

Tsirelson's characterization of $w^\star(G)$

The quantum strategy of Alice and Bob in an XOR game can be described by

- a quantum state $|\psi\rangle_{AB} \in \mathbb{C}^d \otimes \mathbb{C}^d$,
- for every question s sent to Alice, a POVM $\{A_s^0, A_s^1\}$,
- for every question t sent to Bob, a POVM $\{B_t^0, B_t^1\}$.

Then the probability of answering (a,b) to questions (s,t) is given by $\langle\psi|A_s^a \otimes B_t^b|\psi\rangle$.

Write $A_s = A_s^0 - A_s^1$ and $B_t = B_t^0 - B_t^1$. Then

$$A_s^a = \frac{\mathbb{I} + (-1)^a A_s}{2}, \quad B_t^b = \frac{\mathbb{I} + (-1)^b B_t}{2}, \qquad a, b \in \{0,1\}.$$

Moreover, POVMs $\{A_s^0, A_s^1\}$ and $\{B_t^0, B_t^1\}$ exactly correspond to Hermitian A_s, B_t with $\|A_s\|, \|B_t\| \le 1$.

We have

$$\begin{aligned}
w^\star(G) &= \sup_{|\psi\rangle,\{A_s^a\},\{B_t^b\}} \mathop{\mathbb{E}}_{(s,t)\sim\pi} \sum_{a,b} V(a,b|s,t)\langle\psi|A_s^a\otimes B_t^b|\psi\rangle \\
&= \sup_{\substack{|\psi\rangle,\{A_s\},\{B_t\}\\ A_s,B_t \text{ Hermitian}\\ \|A_s\|,\|B_t\|\le 1}} \mathop{\mathbb{E}}_{(s,t)\sim\pi} \sum_{a,b} f_{s,t}(a\oplus b)\left\langle\psi\left|\frac{\mathbb{I}+(-1)^a A_s}{2}\otimes\frac{\mathbb{I}+(-1)^b B_t}{2}\right|\psi\right\rangle \\
&= \mathop{\mathbb{E}}_{(s,t)\sim\pi} \sum_{a,b}\frac{f_{s,t}(a\oplus b)}{4} + \frac{1}{2} \sup_{\substack{|\psi\rangle,\{A_s\},\{B_t\}\\ A_s,B_t \text{ Hermitian}\\ \|A_s\|,\|B_t\|\le 1}} \mathop{\mathbb{E}}_{(s,t)\sim\pi} g_{s,t}\langle\psi|A_s\otimes B_t|\psi\rangle
\end{aligned}$$

where $g_{s,t} := f_{s,t}(0) - f_{s,t}(1)$. Note that the first term in the last line is the expected value for random answers, and does not depend on the players' strategy. So we introduce the *bias*

$$\beta^\star(G) := \sup_{\substack{|\psi\rangle,\{A_s\},\{B_t\}\\ A_s,B_t \text{ Hermitian}\\ \|A_s\|,\|B_t\|\le 1}} \mathop{\mathbb{E}}_{(s,t)\sim\pi} g_{s,t}\langle\psi|A_s\otimes B_t|\psi\rangle$$

as the advantage the players can possibly have over the random strategy:

$$w^\star(G) = \mathop{\mathbb{E}}_{(s,t)\sim\pi} \sum_{a,b}\frac{f_{s,t}(a\oplus b)}{4} + \frac{1}{2}\beta^\star(G).$$

A further simplification: while in general A_s and B_t are Hermitians with norm at most one, for the best strategy we can further assume w.l.o.g. that they are actually *observables*, i.e., the eigenvalues of A_s and B_t are ± 1, or equivalently $A_s^2 = B_t^2 = \mathbb{I}$. To see this, note that for fixed $|\psi\rangle$ and B_t, the value $w^\star(G)$ is linear in the matrices A_s, and by convexity, the best choice can be assumed to be at an extreme point of the unit ball of Hermitian matrices, i.e., all A_s (and similarly all B_t) are observables.

Example 5. *In the CHSH game, Alice's measurements are* $\{A_0^0 = |0\rangle\langle 0|, A_0^1 = |1\rangle\langle 1|\}$ *and* $\{A_1^0 = |+\rangle\langle +|, A_1^1 = |-\rangle\langle -|\}$, *and* $A_0 = |0\rangle\langle 0| - |1\rangle\langle 1|$ *and* $A_1 = |+\rangle\langle +| - |-\rangle\langle -|$ *are observables. The case for Bob is similar.*

Theorem 6 (Tsirelson)**.** *Given an* $n\times n$ *complex matrix* $C = (C_{s,t})$, *the following are equivalent:*

i) there exist $d\in\mathbb{N}$, $|\psi\rangle\in\mathbb{C}^d\otimes\mathbb{C}^d$, $A_s, B_t \in \mathrm{Herm}(\mathbb{C}^d)$, $A_s^2 = B_t^2 = \mathbb{I}$, *such that* $C_{s,t} = \langle\psi|A_s\otimes B_t|\psi\rangle$;

ii) there exist real unit vectors $\mathbf{x}_s, \mathbf{y}_t \in \mathbb{R}^{n+2}$ *for* $1\le s,t\le n$ *such that* $C_{s,t} = \mathbf{x}_s\cdot\mathbf{y}_t$ *;*

iii) the same as (i) but $d \le 2^{\lceil\frac{n+2}{2}\rceil}$.

Proof. (iii) $\Longrightarrow$ (i): trivial.

(i) $\Longrightarrow$ (ii): For each s,t, define $|x_s\rangle = (A_s\otimes\mathbb{I})|\psi\rangle$, $|y_t\rangle = (\mathbb{I}\otimes B_t)|\psi\rangle \in \mathbb{C}^d$. Then $\langle x_s|y_t\rangle = C_{s,t}$. Write $|x_s\rangle = \sum_i u_i|e_i\rangle$ and $|y_t\rangle = \sum_i v_i|e_i\rangle$ in some fixed orthonormal basis $\{|e_i\rangle\}$ of $\mathbb{C}^d$. Define real vectors $\mathbf{x}_s, \mathbf{y}_t \in \mathbb{R}^{2d}$:

$$\begin{aligned}
\mathbf{x}_s &= (\mathrm{Re}(u_1), \mathrm{Im}(u_1), \ldots, \mathrm{Re}(u_d), \mathrm{Im}(u_d)), \\
\mathbf{y}_t &= (\mathrm{Re}(v_1), \mathrm{Im}(v_1), \ldots, \mathrm{Re}(v_d), \mathrm{Im}(v_d)).
\end{aligned}$$

Then $\mathbf{x}_s \cdot \mathbf{y}_t = \mathrm{Re}(\sum_i \overline{u}_i v_i) = \mathrm{Re}(\langle x_s | y_t \rangle) = \mathrm{Re}(C_{s,t})$. But A_s, B_t being Hermitian implies that $C_{s,t}$ is real. So $\mathbf{x}_s \cdot \mathbf{y}_t = C_{s,t}$. We also have $\|\mathbf{x}_s\|^2 = \mathrm{Re}(\langle x_s | x_s \rangle) = 1$, so all $\mathbf{x}_s$, and similarly all $\mathbf{y}_t$, are unit vectors.

The only problem is that the dimension is $2d$ rather than $n+2$. But note that replacing each $\mathbf{x}_s$ by its orthogonal projection onto the subspace spanned by all $\mathbf{y}_t$ does not affect any $\mathbf{x}_s \cdot \mathbf{y}_t$. So we may assume all $\mathbf{x}_s$ and $\mathbf{y}_t$ lie in an n-dimensional subspace. By representing $\mathbf{x}_s, \mathbf{y}_t$ in an orthonormal basis of this subspace, we have $\mathbf{x}_s, \mathbf{y}_t \in \mathbb{R}^n$ and $\mathbf{x}_s \cdot \mathbf{y}_t = C_{s,t}$.

Finally, note that projecting $\mathbf{x}_s$ and $\mathbf{y}_t$ shrinks their norms. But we can fix it by enlarging the dimension by two. For each $\mathbf{x}_s$ we add two coordinates $(\sqrt{1 - \|\mathbf{x}_s\|^2}, 0)$. And for each $\mathbf{y}_t$ the new coordinates are $(0, \sqrt{1 - \|\mathbf{y}_t\|^2})$. It does not affect $\mathbf{x}_s \cdot \mathbf{y}_t$ and hence all properties are satisfied.

(ii) $\Longrightarrow$ (iii): Choose $d = 2^{\lceil \frac{n+2}{2} \rceil}$ and consider the maximal entangled state $|\psi\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle |i\rangle$. This state has the property that for any $A = (A_{ij})$, $B = (B_{ij})$,

$$\langle \psi | A \otimes B | \psi \rangle = \frac{1}{d} \sum_{i,j} \langle i | A | j \rangle \langle i | B | j \rangle = \frac{1}{d} \sum_{i,j} A_{ij} B^{\mathsf{T}}_{ji} = \frac{1}{d} \sum_{i} (AB^{\mathsf{T}})_{ii} = \frac{1}{d} \mathrm{Tr}(AB^{\mathsf{T}}).$$

Next we need a construction of anti-commuting matrices. This is a “representation of the Clifford algebra”. Recall the Pauli matrices

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \; Y = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \; Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

They anti-commute and all square to the identity. For $1 \le i \le \lceil \frac{n+2}{2} \rceil$, define the $d \times d$ matrices

$$T_{2i-1} = \underbrace{\mathbb{I} \otimes \cdots \otimes \mathbb{I}}_{i-1} \otimes X \otimes \underbrace{Y \otimes \cdots \otimes Y}_{\lceil \frac{n+2}{2} \rceil - i},$$

$$T_{2i} = \underbrace{\mathbb{I} \otimes \cdots \otimes \mathbb{I}}_{i-1} \otimes Z \otimes \underbrace{Y \otimes \cdots \otimes Y}_{\lceil \frac{n+2}{2} \rceil - i}.$$

All of them are Hermitian, square to the identity, and mutually anti-commute.

For $1 \le s, t \le n$, define $A_s = \sum_{i=1}^{n+2} (\mathbf{x}_s)_i T_i$ and $B_t = \sum_{i=1}^{n+2} (\mathbf{y}_t)_i T_i^{\mathsf{T}}$. Then for all $1 \le s \le n$,

$$A_s^2 = \sum_i (\mathbf{x}_s)_i^2 T_i^2 + \sum_{i<j} (\mathbf{x}_s)_i (\mathbf{x}_s)_j \underbrace{(T_i T_j + T_j T_i)}_{=0} = \sum_i (\mathbf{x}_s)_i^2 T_i^2 = \|\mathbf{x}_s\|^2 \cdot \mathbb{I} = \mathbb{I}.$$

Similarly $B_t^2 = \mathbb{I}$ for all $1 \le t \le n$. Finally,

$$\begin{aligned} \langle \psi | A_s \otimes B_t | \psi \rangle = \frac{1}{d} \mathrm{Tr}(A_s B_t^{\mathsf{T}}) &= \frac{1}{d} \sum_{i,j} (\mathbf{x}_s)_i (\mathbf{y}_t)_j \mathrm{Tr}(T_i T_j) \\ &= \frac{1}{d} \sum_i (\mathbf{x}_s)_i (\mathbf{y}_t)_i \mathrm{Tr}(T_i^2) + \frac{1}{d} \sum_{i<j} (\mathbf{x}_s)_i (\mathbf{y}_t)_j \mathrm{Tr}\underbrace{(T_i T_j + T_j T_i)}_{=0} \\ &= \frac{1}{d} \sum_i (\mathbf{x}_s)_i (\mathbf{y}_t)_i \mathrm{Tr}(\mathbb{I}) \\ &= \mathbf{x}_s \cdot \mathbf{y}_t \end{aligned}$$

which equals $C_{s,t}$, as desired. □

This theorem is very powerful. From (i) $\iff$ (ii), we have an alternative characterization of the bias $\beta^\star(G)$:

$$\beta^\star(G) = \sup_{\mathbf{x}_s,\mathbf{y}_t \in \mathbb{R}^{n+2}, \|\mathbf{x}_s\|=\|\mathbf{y}_t\|=1} \mathop{\mathbb{E}}_{(s,t)\sim\pi} g_{s,t}\mathbf{x}_s \cdot \mathbf{y}_t. \tag{$*$}$$

Note that the supremum is now over a compact set of bounded dimension $n+2$, and hence is actually a maximum. And from (i) $\iff$ (iii), we have a bound $2^{\lceil \frac{n+2}{2} \rceil}$ for the dimension of an optimal strategy $|\psi\rangle$.

Moreover, the characterization $(*)$ is a semi-definite program for which polynomial-time algorithms exist. So we could compute $\beta^\star(G)$, and hence $w^\star(G)$, in polynomial time.

Example 7. *In the CHSH game, we can check by hand that $\beta^\star(G) \le \frac{\sqrt{2}}{2}$:*

$$\begin{aligned}
\beta^\star(G) &= \sup_{\mathbf{x}_s,\mathbf{y}_t} \frac{1}{4}(\mathbf{x}_0\cdot\mathbf{y}_0 + \mathbf{x}_0\cdot\mathbf{y}_1 + \mathbf{x}_1\cdot\mathbf{y}_0 - \mathbf{x}_1\cdot\mathbf{y}_1) \\
&= \sup_{\mathbf{x}_s,\mathbf{y}_t} \frac{1}{4}(\mathbf{x}_0\cdot(\mathbf{y}_0+\mathbf{y}_1) + \mathbf{x}_1\cdot(\mathbf{y}_0-\mathbf{y}_1)) \\
&\le \sup_{\mathbf{x}_s,\mathbf{y}_t} \frac{1}{4}(\|\mathbf{x}_0\|\|\mathbf{y}_0+\mathbf{y}_1\| + \|\mathbf{x}_1\|\|\mathbf{y}_0-\mathbf{y}_1\|) \\
&\le \sup_{\mathbf{x}_s,\mathbf{y}_t} \frac{1}{4}(\|\mathbf{y}_0+\mathbf{y}_1\| + \|\mathbf{y}_0-\mathbf{y}_1\|) \\
&\le \sup_{\mathbf{x}_s,\mathbf{y}_t} \frac{\sqrt{2}}{4}\sqrt{\|\mathbf{y}_0+\mathbf{y}_1\|^2 + \|\mathbf{y}_0-\mathbf{y}_1\|^2} \\
&= \sup_{\mathbf{x}_s,\mathbf{y}_t} \frac{\sqrt{2}}{4}\sqrt{2\|\mathbf{y}_0\|^2 + 2\|\mathbf{y}_1\|^2} \\
&= \frac{\sqrt{2}}{2}.
\end{aligned}$$

So $w^\star(G) \le \frac{1}{2} + \frac{\sqrt{2}}{4} = \cos^2(\frac{\pi}{8})$: The strategy we have seen before is optimal!