**Semidefinite programming and computational aspects of entanglement** **IHP Fall 2017**

# Lecture 4: November 17

*Lecturer: Anand Natarajan* *Scribe: Aram and Anand*

Today's lecture is about the sum-of-squares (SoS) hierarchy. This is a general method to find SDP relaxations to optimization problems with polynomial constraints and objective functions.

On the first day we saw that our problems of interest can be cast as polynomial optimization problems:

$$\begin{aligned} &\max_{x\in\mathbb{R}^n} f(x) \\ &\text{s.t. } g_i(x) = 0\,\forall i \\ &\qquad h_j(x) \geq 0\,\forall j, \end{aligned} \tag{4.1}$$

where $f(x), g_1(x), g_2(x), \ldots, h_1(x), \ldots$ are polynomials.

This is unfortunately not convex for polynomials with degree $> 1$. How can we make it convex? Let $S$ denote the feasible region

$$S = \{x \in \mathbb{R}^n : \forall_i g_i(x) = 0, \forall_j h_j(x) \geq 0\}. \tag{4.2}$$

Then we can replace the max over $x$ with a max over probability distributions $p$ supported on $S$. Our new optimization is

$$\max_{p:\operatorname{supp} p\subseteq S} \mathop{\mathbb{E}}_{x\sim p}[f(x)] \cong \langle \mathbb{E}, f\rangle. \tag{4.3}$$

This is now convex but infinite dimensional.

**Necessary conditions for $p$ supported on $S$.** We note that $\mathbb{E}$ must satisfy certain properties if it indeed arises from $p$ supported on $S$. Here is an (incomplete) list:

$$\mathbb{E}[\cdot] \text{ is a linear function} \tag{4.4a}$$
$$\mathbb{E}[1] = 1 \tag{4.4b}$$
$$\mathbb{E}[g_i(x)r(x)] = 0 \qquad \forall i, \forall r(x) \tag{4.4c}$$
$$\mathbb{E}[q^2(x)] \geq 0 \qquad \forall q(x) \tag{4.4d}$$
$$\mathbb{E}[h_j(x)q^2(x)] \geq 0 \qquad \forall j, \forall q(x) \tag{4.4e}$$

Observe also that $\mathbb{E}$ is a linear map from polynomials to real numbers. Our relaxation will replace $\mathbb{E}$ with a "pseudo-expectation" $\tilde{\mathbb{E}}$ that acts only on polynomials of degree $\leq d$, which we call $\mathbb{R}[x_1, \ldots, x_n]_{\leq d}$. The set of such $\tilde{\mathbb{E}}$ is a finite dimensional vector space, indeed it has dimension $O(n^d)$.

The level-$d$ SOS relaxation of (4.1) is then to maximize $\tilde{\mathbb{E}}[f]$ subject to $\tilde{\mathbb{E}}$ satisfying the conditions in (4.4). (Will explain the "SOS" name later.)

Why is this an SDP? If $u, v$ are monomials of degree $\leq d/2$ then we can define a matrix $X$ such that

$$X_{u,v} = \tilde{\mathbb{E}}[uv]. \tag{4.5}$$

For example, if $n = 2, d = 2$ then (using $x, y$ instead of $x_1, x_2$ as variables for readability)

$$X = \begin{array}{c c} & \begin{array}{ccc} 1 & x & y \end{array} \\ \begin{array}{c} 1 \\ x \\ y \end{array} & \begin{pmatrix} 1 & \tilde{\mathbb{E}}[x] & \tilde{\mathbb{E}}[y] \\ \tilde{\mathbb{E}}[x] & \tilde{\mathbb{E}}[x^2] & \tilde{\mathbb{E}}[xy] \\ \tilde{\mathbb{E}}[y] & \tilde{\mathbb{E}}[xy] & \tilde{\mathbb{E}}[y^2] \end{pmatrix} \end{array}. \tag{4.6}$$

What do the constraints look like in this picture? Linearity (4.4a) is automatic. The equality constraints (4.4b) and (4.4c) become simple equality constraints on linear combinations of matrix entries of $X$, e.g. $X_{1,1} = 1$. For (4.4c) note that we do not need to consider arbitrary $r(x)$ but it suffices to consider only all the monomials of degree $\leq d - \deg g_i$.

The inequality constraints are more interesting. Start with (4.4d). Let $q = \sum_i \alpha_i q_i$ where $\alpha_i \in \mathbb{R}$ and each $q_i$ is a monomial of degree $\leq d/2$. Then

$$0 \leq \tilde{\mathbb{E}}[q^2] = \tilde{\mathbb{E}}\left[\sum_{i,j} \alpha_i \alpha_j q_i q_j\right] = \alpha^T X \alpha, \tag{4.7}$$

where in the last step we have defined the vector $\alpha$ to have components $\alpha_i$. (4.4d) holds for all $q$ iff (4.7) holds for all $\alpha$ iff $X \succeq 0$. Thus (4.4d) is equivalent to a psd constraint.

Likewise we can define $Y^{(j)}_{u,v} = \tilde{\mathbb{E}}[u h_j v]$ and it turns out that $Y^{(j)} \succeq 0$ iff (4.4e) holds for all $q(x)$.

Thus we have an SDP with $O(n^d)$ variables which can be solved in time $n^{O(d)}$.

**Dual picture.** The dual is

$$\min\{\gamma : \gamma - f(x) = \sum_i s_i^2(x) + \sum_j r_j(x) g_j(x) + \sum_{k,l} q_{k,l}^2(x) h_k(x)\}, \tag{4.8}$$

where all polynomials have degree $\leq d$. Note that dual feasible points are "proofs" that $f(x) \leq \gamma$ for all $x \in S$. We can see this by applying $\tilde{\mathbb{E}}$ to both sides of (4.8). We refer to terms such as $\sum_i s_i^2(x)$ as "sums of squares" of polynomials, aka "SOS polynomials". It is not immediately obvious that (4.8) is an SDP. To see this, we observe that there is an SDP to search over SOS polynomials. The argument is similar to the one in (4.7). Let $p(x) = \sum_{u,v} A_{u,v} uv$, where $u, v$ are monomials and $A_{u,v}$ depends only on the product $uv$. If $A \succeq 0$ then $A = B^T B$ for some matrix $B$. This means that $p(x) = \sum_{u,v,i} B_{i,u} B_{i,v} uv = \sum_i b_i(x)^2$, where $b_i(x) = \sum_u B_{i,u} u$. Thus $p$ is SOS. The argument can run in the other direction too: there is a 1-1 correspondence between polynomials $p$ and matrices $A$, and between SOS polynomials $p$ and psd matrices $A$.

We can think of the dual picture as searching for a "SOS proof" that $\gamma - f(x) \geq 0$. An SOS proof is just a way of proving a polynomial is nonnegative by writing it as a sum of squares, plus terms that are forced to be zero or nonnegative by the constraints.

The convergence of the dual is established by the following theorem.

**Putinar's Positivstellensatz.** Given $S$ suppose there is an SOS proof that $\|x\|_2^2 \leq C$ for some constant $C$. This is known as the "Archimedean condition." Given a function $f$ that is strictly positive on $S$, there is an SOS proof that $f$ is nonnegative on $S$, i.e.

$$f = \sum_i s_i^2 + \sum_j r_j g_j + \sum_k q_{k,l}^2 h_k, \tag{4.9}$$

for appropriate polynomials $s_i, r_j, q_{k,l}$. In other words, the SOS hierarchy converges to the correct answer as $d \to \infty$.

**Example: MAX-CUT SDP (Goemans-Williamson).** The polynomial optimization problem is $\min \sum_{i,j} A_{i,j} x_i x_j$ s.t. $x_i^2 = 1, \forall i \in [n]$. The $d = 2$ SOS relaxation is

$$\begin{aligned} &\min_X \operatorname{Tr}[AX] \\ &\text{s.t. } X_{i,i} = 1, \forall i \\ &\quad X \succeq 0 \end{aligned} \tag{4.10}$$

This equivalent to taking $\tilde{\mathbb{E}}$ as our variables and computing

$$\begin{aligned} &\min \tilde{\mathbb{E}}[\sum_{i,j} A_{i,j} x_i x_j] \\ &\text{s.t. } \tilde{\mathbb{E}}[x_i^2] = 1, \forall i \end{aligned} \tag{4.11}$$

**Example:** $h_{\text{Sep}}$ Applying it to $h_{\text{Sep}}$:

$$\max_{x,y} \sum_{ijk\ell} M_{ijk\ell} x_i^* y_j^* x_k y_\ell \tag{4.12a}$$

$$\text{s.t.} \sum_i x_i^* x_i = 1 \tag{4.12b}$$

$$\text{s.t.} \sum_i y_i^* y_i = 1. \tag{4.12c}$$

We now have complex variables. So we we need to allow polynomials in $x, y, x^*, y^*$.

Second: WLOG, only nonzero moments are those that where $x, x^*$ and $y, y*$ are balanced (let $x, y$ have independent, uniformly chosen phases).

Third: to specify such a moment matrix, sufficient to specify the highest degree moments only.

$$\rho_{ijk\ell} := \tilde{\mathbb{E}}[x_i y_j x_k^* y_\ell^*].$$

This is because of the normalization constraints which effectively allow us to carry about partial traces and obtain lower degree moments. Indeed $\tilde{\mathbb{E}}[r(x)] = \tilde{\mathbb{E}}[r(x)\|x\|^2]$.

These moments can be viewed as a quantum state! The normalization conditions (4.12b), (4.12c) imply that $\mathrm{tr}[\rho] = 1$. And the usual psd conditions from the SOS hierarchy yield that $\rho \succeq 0$.

Let us generalize the degree-$d$ condition slightly and allow degree $d_x$ in the $x$ variables and degree $d_y$ in the $y$ variables. We obtain

$$\rho \in D(\mathcal{H}_X^{\otimes d_x/2} \otimes \mathcal{H}_Y^{\otimes d_y/2}), \tag{4.13}$$

where $\mathrm{supp}\, \rho$ is contained in $\mathrm{Sym}^{d_x/2} \mathcal{H}_X \otimes \mathrm{Sym}^{d_y/2} \mathcal{H}_Y$. If we take the marginal on one of the copies of $\mathcal{H}_X$ and one of the copies of $\mathcal{H}_Y$ then we obtain a state that is in the set called $\mathrm{DPS}_k$. (This is named after Doherty, Parrilo, Spedalieri, who proposed this relaxation.)

It includes the $k$-extendible condition which states that a separable state $\sigma^{XY}$ can be found as the $XY_1$ marginal of some state $\rho^{XY_1 \ldots Y_k}$ that is symmetric under permutation of the $Y_1, \ldots, Y_k$. (Other versions also exist, e.g. we could have $d_x/2$ copies of $\mathcal{H}_X$ and $d_y/2$ copies of $\mathcal{H}_Y$.)

It also includes the PPT condition. Indeed take $d_x = d_y = 2$ and consider some polynomial $q = \sum_{i,j} q_{i,j} x_i y_j^*$. Then $\tilde{\mathbb{E}}[|q|^2] = \langle q| \rho^\Gamma |q\rangle$ and so we obtain the condition that $\rho^\Gamma \succeq 0$. This is called the PPT, or positive partial transpose condition.

**Example: $\omega^*$.** Recall that

$$\begin{aligned} \omega* = \max\ & \langle\psi| \sum_{a,b,x,y} G(a,b,x,y) A_a^x B_b^y |\psi\rangle \\ \text{s.t.}\ & A_a^x \geq 0,\ \forall a, x \\ & B_b^y \geq 0,\ \forall b, y \\ & \sum_x A_a^x = I,\ \forall a \\ & \sum_y B_b^y = I,\ \forall b \\ & [A_a^x, B_b^y] = 0\, \forall a, x, b, y \end{aligned} \tag{4.14}$$

We will need a new SDP hierarchy for this. But we can make use of:

**The noncommutative Positivstellensatz (Helton, McCullough).** Let $S := \{X : p_i(X) \geq 0 \forall i\}$ where $X$ is a list of Hermitian operators, and the $p_i$ are expectations of polynomials of $X$. Suppose $f(X) > 0$ for all $X \in S$ and we assume some analogue of the Archmedean condition. Then

$$f = \sum_k r_k^\dagger r_k + \sum_{i,j} s_{i,j}^\dagger p_i s_{i,j}, \tag{4.15}$$

where all these terms are noncommutative polynomials in $X$.

This yields the ncSOS hierarchy for nonlocal games [1, 2].

$$\begin{aligned}
\max_{\tilde{\mathbb{E}}} \ & \tilde{\mathbb{E}}\left[\sum_{a,b,x,y} G(a,b,x,y) A_a^x B_b^y\right] \\
\text{s.t. } & \tilde{\mathbb{E}}[q^\dagger A_a^x q] \geq 0 && \forall q, a, x \\
& \tilde{\mathbb{E}}[q^\dagger B_b^y q] \geq 0 && \forall q, b, y \\
& \tilde{\mathbb{E}}[p[A_a^x, B_b^y]q] = 0 && \forall p, q, a, b, x, y \\
& \tilde{\mathbb{E}}[p(\sum_x A_a^x - I)q] = 0 && \forall p, q, a \\
& \tilde{\mathbb{E}}[p(\sum_y B_b^y - I)q] = 0 && \forall p, q, b \\
& \tilde{\mathbb{E}}[q^\dagger q] \geq 0 && \forall q \\
& \tilde{\mathbb{E}}[I] = 1
\end{aligned} \tag{4.16}$$

**Tsirelson's characterization of 2-player XOR games.** For any XOR game, $x, y \in \{0,1\}$ and $V(x,y|a,b) = 1$ if $x \oplus y = s_{a,b}$ or 0 if not. We can define $A_a = A_a^0 - A_a^1$ and $B_b = B_b^0 - B_b^1$. Then $A_a^2 = B_b^2 = I$, $[A_a, B_b] = 0$, and the XOR condition states that

$$\omega^* = \max \sum_{a,b} \underbrace{\pi(a,b)(-1)^{s_{a,b}}}_{G(a,b)} \langle\psi| A_a B_b |\psi\rangle + \text{ constant.} \tag{4.17}$$

Tsirelson proved that level 2 of the ncSOS hierarchy gives the exact answer in this case. This remarkable since the commutative analogue is a generalization of MAX-CUT where the SOS hierarchy does not give an exact answer at $O(1)$ levels.

The proof will take a solution to the ncSOS relaxation and construct an entangled strategy achieving the same value. Since the ncSOS relaxation gives a value that is $\geq \omega^*$ and any given strategy achieves a value $\leq \omega^*$, this proves that we have actually found $\omega^*$.

The ncSOS in this case is

$$\begin{aligned}
\max \ & \sum_{a,b} G(a,b) X_{a,b} \\
\text{s.t. } & X \succeq 0 \\
& X_{1,1} = 1 \\
& X_{a,a} = 1 && \forall a \\
& X_{b,b} = 1 && \forall b \\
& X_{a,b} = X_{b,a} && \forall a, b
\end{aligned} \tag{4.18}$$

Since this is psd, it is a Gram matrix and we have $X_{a,b} = \langle u_a, u_b \rangle$.

Now define the Clifford observables, $C_1, \ldots, C_n$ to be operators satisfying

$$\{C_i, C_j\} = 2\delta_{i,j} I, \forall i, j \in [n] \tag{4.19}$$

One way to achieve this is to take $C_i = \sigma_x^{\otimes i-1} \otimes \sigma_z \otimes I^{\otimes n-i}$.

Given a vector $u \in \mathbb{R}^N$, let $C(u) = \sum_i u_i C_i$. Observe that

$$\{C(u), C(v)\} = 2\langle u, v \rangle I. \tag{4.20}$$

Now we use the shared state $|\Phi\rangle = \frac{1}{\sqrt{2^N}} \sum_{i=1}^{2^N} |i\rangle \otimes |i\rangle$ and take $A_a = C(u_a)$, $B_b = C(u_b)$. We find that the payoff from this strategy is

$$\begin{aligned}
&\sum_{a,b} G(a,b) \langle\Phi| A_a B_b |\Phi\rangle \\
=&\sum_{a,b} G(a,b) \frac{\operatorname{tr} A_a B_b}{2^N} \\
=&\sum_{a,b} G(a,b) \frac{\operatorname{tr}\{A_a, B_b\}}{2 \cdot 2^N} \\
=&\sum_{a,b} G(a,b) \frac{\operatorname{tr}\{C(u_a), C(u_b)\}}{2 \cdot 2^N} \\
=&\sum_{a,b} G(a,b) \langle u_a, u_b\rangle \frac{\operatorname{tr} I}{2^N} \\
=&\sum_{a,b} G(a,b) X_{a,b}.
\end{aligned} \tag{4.21}$$

## References

[1] A. C. Doherty, Y.-C. Liang, B. Toner, and S. Wehner. The quantum moment problem and bounds on entangled multi-prover games. In *CCC '08*, pages 199–210, 2008, arXiv:0803.4373.

[2] M. Navascués, S. Pironio, and A. Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New J. Phys.*, 10(7):073013, 2008, arXiv:0803.4290.