**Rounding Techniques in Approximation Algorithms**

# Lecture 22: Pseudodistributions and Global Correlation

*Lecturer: Nathan Klein*

## 1 The Dual of SoS

Last time we talked about the sum of squares hierarchy. We showed that a polynomial $p$ over the cube has a degree $d$ SoS certificate if and only if there is a matrix $A \in \mathbb{R}^{(n+1)^{d/2} \times (n+1)^{d/2}}$ so that $A \succeq 0$ and $p(x) = ((1,x)^{\otimes d/2})^T A (1,x)^{\otimes d/2}$ on all $x \in \{-1,1\}^n$. We also showed a degree 4 SoS certificate for the triangle graph, and in particular the polynomial $2 - \frac{1}{2}((1 - x_u x_v) + (1 - x_u x_w) + (1 - x_v x_w)) = \frac{1}{2}(1 + x_u x_v + x_u x_w + x_v x_w)$.

### 1.1 Computing a Degree $d$ SoS Proof

How do we actually find such proofs? Using the above, given a polynomial $p$, to find a degree $d$ SoS proof it's enough to find a matrix $A \in \mathbb{R}^{(n+1)^{d/2} \times (n+1)^{d/2}}$ so that $A \succeq 0$ and $p(x) = g_A(x)$ for all $x \in \mathbb{R}^n$, where $g_A(x) = ((1,x)^{\otimes d/2})^T A (1,x)^{\otimes d/2}$.

Let's try to find this matrix, then:

$$\begin{aligned} & A \succeq 0 \\ & g_A(x) = p(x) \quad \forall x \in \{-1,1\}^n \end{aligned}$$

We already know we can handle the PSD constraint using the ellipsoid method. But the second family of constraints looks a bit scary, since there are exponentially many of them. However, it's not hard to see that this can be simplified down to $(n+1)^d$ constraints. Why? Well, $g_A = ((1,x)^{\otimes d/2})^T A (1,x)^{\otimes d/2}$ has at most $(n+1)^d$ terms, as it has maximum degree $d$. So it is enough that all of the coefficients on these terms are equal after multilinearization, leading to $(n+1)^d$ constraints. (Notice we are assuming here that $p$ has degree at most $d$, but this is necessary to have any hope of solving this program.) Finally, notice that this is indeed a set of linear constraints: we can write the coefficient of $x^S = \prod_{i \in S} x_i$ by simply summing the entries of the matrix which contribute to the term $x^S$. So, we can use the ellipsoid method to find degree $d$ SoS certificates in time polynomial in $n^d$.

Now suppose we want to find the best possible degree $d$ SoS proof, say for Max Cut. Here, let $p_M = \sum_{\{u,v\} \in E} \frac{1}{2}(1 - x_u x_v)$. If we could prove that $\alpha - p_M$ was SoS for some $d$, this would demonstrate that $OPT \leq \alpha$. The goal then is to find the minimum $\alpha$ so that $\alpha - p_M$ has a degree $d$ SoS certificate. We could do this by trying the above feasibility problem for all $\alpha$, but we could also optimize directly, i.e. solve:

$$\begin{aligned} \min \ & \alpha \\ \text{s.t.} \ & A \succeq 0 \\ & g_A(x) \equiv \alpha - p_M(x) \end{aligned}$$

What about the dual of this problem?

## 1.2 Pseudodistributions

The dual of finding SoS certificates is finding pseudodistributions, which we will define shortly. First notice the following:

**Fact 1.1.** *The set of polynomials which have degree $d$ SoS certificates is a closed convex cone.*

*Proof.* A set $S$ is a convex cone if for any $f, g \in S$, we have $\alpha f + \beta g \in S$ whenever $\alpha, \beta \geq 0$. This holds here because if $f = \sum f_i^2$ and $g = \sum g_i^2$ then $f + g = \sum(\sqrt{\alpha} f_i)^2 + \sum(\sqrt{\beta} g_i)^2$. We leave the fact that it is closed as an exercise. □

We will think of functions as being specified by all $2^n$ inputs of the cube. So the set $SoS_d$ of functions with degree $d$ SOS certificates will live in $\mathbb{R}^{2^n}$. But now, since this is a convex cone, if a polynomial $p$ is outside of it there must be a hyperplane going through the origin separating it from this cone.

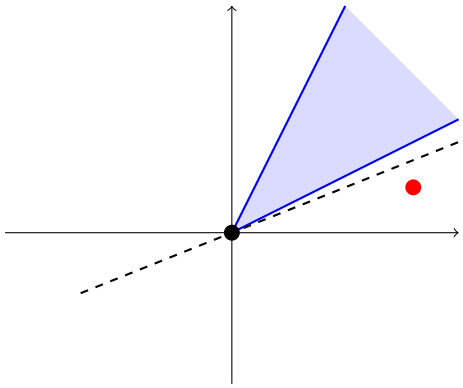

Figure 1: A separating hyperplane through the origin for a convex cone and a point.

We define $\langle f, g \rangle = \sum_{x \in \{-1,1\}^n} f(x)g(x)$ for two functions on the cube. Now, there is a halfspace

$$H = \{f \in \mathbb{R}^{2^n} \mid \langle \mu, f \rangle \geq 0\}$$

which contains all polynomials in $SoS_d$[1] but not $p$ (where $\mu \in \mathbb{R}^{2^n}$). Furthermore, without loss of generality, we can scale $\mu$ so that its entries sum to 1 (since it's just an inequality). So, it's kind of like a distribution, although it's really not, since it can be supported on negative numbers. That's why we call this a pseudodistribution, which has the following properties:

**Definition 1.2** (Pseudodistribution)**.** *A degree $d$ pseudodistribution for a polynomial $p$ is a function $\mu : \{-1,1\}^n \to \mathbb{R}$ so that the expectation $\tilde{\mathbb{E}}_\mu$ obeys:*

1. $\tilde{\mathbb{E}}_\mu[1] = 1$
2. *For all polynomials $g$ of maximum degree $d/2$, $\tilde{\mathbb{E}}_\mu[g^2] \geq 0$.*

Where we define $\tilde{\mathbb{E}}_\mu[f(x)] = \langle \mu, f \rangle = \sum_{x \in \{-1,1\}^n} \mu(x) f(x)$, similar to a normal expectation.

[1] Remember that every function on the cube is a polynomial.

**Fact 1.3.** *Suppose $p \notin SoS_d$ and $\mu$ is the normal vector of a hyperplane going through the origin separating $p$ from $SoS_d$, scaled WLOG so that $\sum_{x\in\{-1,1\}^n} \mu(x) = 1$. Then $\mu$ is a degree $d$ pseudodistribution.*

*Proof.* To show (1), note: $\tilde{\mathbb{E}}_\mu[1] = \sum_{x\in\{-1,1\}^n} \mu(x) = 1$ since we scaled $\mu$. (2) follows because the SoS cone is contained in $H$. In particular, $g^2 \in SoS_d \subseteq H$ for any $g$ of degree at most $d/2$, so $\tilde{\mathbb{E}}_\mu[g^2] = \sum_{x\in\{-1,1\}^n} \mu(x)g^2(x) \geq 0$ as desired. □

To gain some more intuition for this object, notice that a real distribution is always a pseudodistribution. And an easy fact is that every pseudodistribution of degree at least $2n$ is a real distribution. This is because the indicator of any point in the cube is a polynomial of degree $n$, which says that $\mu$ must be non-negative everywhere.

But, of course, when we move below degree $2n$, there are pseudodistributions with negative probabilities that can masquerade as real distributions. Let's see an example of this on our favorite instance: max cut for the triangle. Our polynomial $p$ is $\frac{1}{2}(1 + x_u x_v + x_u x_w + x_v x_w)$: we would like to prove that the max cut is at most 2. This is degree 4 SoS but not degree 2. There is a reason that degree 2 SoS does not work: the following pseudodistribution over $x \in \{-1,1\}^3$. This is a dual certificate for the polynomial, i.e. a proof that it is not degree 2 SoS.

1. Assign weight $-\frac{1}{16}$ to the points $(-1,-1,-1)$ and $(1,1,1)$ and assign weight $\frac{3}{16}$ to the remaining points. By summing the weights we can see $\tilde{\mathbb{E}}_\mu[1] = 1$.
2. Where $Y$ is the "covariance matrix" of this distribution with $Y_{uv} = \tilde{\mathbb{E}}_\mu[x_u x_v]$, we have:
$$\tilde{\mathbb{E}}_\mu[(\sum c_v x_v)^2] = c^T Y c \geq 0$$
since one can compute that $Y = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix}$, the same covariance matrix we saw for max cut, which is PSD. So, all polynomials $p$ of degree 1 obey $\tilde{\mathbb{E}}_\mu[p^2] \geq 0$.

$\mu$ forms a hyperplane so that all polynomials $g$ that *are* degree 2 SoS obey $\tilde{\mathbb{E}}_\mu[g] \geq 0$: that's what (2) implies, since any such $g$ can be written as a sum of squares of polynomials of degree at most 1. But this isn't true for $p$! When $x = (-1,-1,-1)$ or $(1,1,1)$ we get 2 for a contribution of $-\frac{1}{4}$. In the other cases $p(x) = 0$. So, $\tilde{\mathbb{E}}_\mu[p] = -\frac{1}{4} < 0$.

## 1.3 Representing Pseudodistributions: Pseudoexpectations

You may worry that pseudodistributions aren't very useful objects since they have exponential size. Thankfully, the following is true:

**Lemma 1.4.** *Let $\mu$ be a degree $d$ pseudodistribution. Then, there is a polynomial $\mu'$ of degree at most $d$ such that*
$$\tilde{\mathbb{E}}_\mu[p] = \tilde{\mathbb{E}}_{\mu'}[p]$$
*for every polynomial $p$ of degree at most $d$.*

Why is this useful? Well, $\mu'$ has only $(n+1)^d$ coefficients, so it has polynomial size for constant $d$. This means that even if we cannot exactly query $\mu$, we can at least efficiently evaluate the expectation of low degree polynomials.[2] Let's prove it:

[2]Multiply the polynomials and expand to see all the at most $n^{O(d)}$ terms. Now using linearity of expectation we can easily compute the expectation overall.

*Proof.* Consider the subspace $S$ of all polynomials on the cube of degree at most $d$. We can write $\mu = \mu' + \mu_\perp$, where $\mu'$ is in $S$ and $\mu_\perp$ is orthogonal to it, i.e. $\langle g, \mu_\perp \rangle = 0$ for all $g \in S$. But now,

$$\tilde{\mathbb{E}}_\mu[g] = \langle \mu' + \mu_\perp, g \rangle = \langle \mu', g \rangle = \tilde{\mathbb{E}}_{\mu'}[p]$$

as desired. □

Notice we didn't use anything about $\mu$, this holds for any distribution. The result is we have an efficiently computable pseudoexpectation operator $\tilde{\mathbb{E}}_\mu$ which has $\tilde{\mathbb{E}}_\mu[g^2] \geq 0$ for polynomials $g$ of degree at most $d/2$.

One final useful fact:

**Fact 1.5.** *$\mu$ is a degree $d$ pseudodistribution if and only if $\tilde{\mathbb{E}}_\mu[1] = 1$ and*

$$\tilde{\mathbb{E}}_\mu[((1,x)^{\otimes d/2})((1,x)^{\otimes d/2})^T] \succeq 0.$$

*This matrix is called the pseudomoment matrix.*

**SoS Algorithm**

For all $d$ and all polynomials $p$ of degree at most $d$, exactly one of the following holds:

1. $p$ has a degree $d$ SoS certificate
2. There is a degree $d$ pseudoexpectation $\tilde{\mathbb{E}}_\mu$ such that $\tilde{\mathbb{E}}_\mu[p] < 0$.

And we can determine this in polynomial time for constant $d$ (up to exponentially small additive error), as well as obtain the certificate in case (1) or all the pseudomoments in case (2).

## 1.4 Back to Max Cut

We can now rephrase the max cut algorithm in the language of pseudoexpectations.

**Theorem 1.6.** *Given a degree-2 pseudodistribution $\mu$, there is a probability distribution $\nu$ over the cube so that*

$$\mathbb{E}_\nu[p_M] \geq 0.878 \cdot \tilde{\mathbb{E}}_\mu[p_M]$$

*where $p_M$ is the max cut polynomial $\sum_{\{u,v\} \in E} \frac{1}{2}(1 - x_u x_v)$.*

The proof is exactly the same. $\tilde{\mathbb{E}}_\mu$ allows us to determine $\tilde{\mathbb{E}}_\mu[x_u x_v]$ for all $u, v \in V$. This lets us build a covariance matrix and proceed as before. In fact, all we have done is rename the variables in some sense. $\tilde{\mathbb{E}}_\mu[x_u x_v]$ is really just $y_{uv}$.