CS369H: Hierarchies of Integer Programming Relaxations Spring 2016-2017

Lecture 4: Polynomial Optimization

Professor Moses Charikar *Scribes: Mona Azadkia*

1 Introduction

Non-negativity over the hypercube. Given a low degree polynomial $f : \{0,1\}^n \to \mathbb{R}$, we want to decide whether $f \geq 0$ over the hypercube or there exists a point $x \in \{0,1\}^n$ such that $f(x) < 0$.

Now let's see how we can formulate the **Max-cut** problem in this language. Remember that we have a graph $G = (V, E)$ and we are looking for a bipartition of the set of vertices V such that the number of edges between these two parts (size of the cut) will be maximized. For $|V| = n$ we encode a bipartition of the vertex set of G by a vector $x \in \{0,1\}^n$ and we let $f_G(x)$ be the number of edges cut by the bipartition x. This function is a degree-2 polynomial,

$$f_G(x) = \sum_{\{ij\} \in E(G)} (x_i - x_j)^2.$$

Therefore deciding if the polynomial $c - f_G(x)$ takes negative value over the hypercube is equivalent to deciding if the maximum cut in G is larger than c.

2 Sum-of-Squares Algorithm

The *Sum-of-Squares* algorithm gets a polynomial $f : \{0,1\}^n \to \mathbb{R}$ as input and outputs

1. Either a proof that $f(x) \geq 0$ for all $x \in \{0,1\}^n$,
2. or an object that "pretends to be" a point $x \in \{0,1\}^n$ with $f(x) < 0$.

2.1 Sum-of-Squares Certificate

Definition 4.1 (SOS certificate)**.** *A degree$-d$ SOS certificate for a function $f : \{0,1\}^n \to \mathbb{R}$ consists of polynomials $g_1, \cdots, g_r : \{0,1\}^n \to \mathbb{R}$ of degree at most $d/2$ for some $r \in \mathbb{N}$ such that*

$$f(x) = \sum_{i=1}^{r} g_i^2(x)$$

for every $\underline{x \in \{0,1\}^n}$.

We will refer to degree$-d$ SOS certificate for f also as a degree$-d$ SOS proof of the inequality $f \geq 0$.

Now one question is that how big is r? We will answer this question later.

2.2 Verifying Certificates

How do we show that $f - \sum_{i=1}^{r} g_i^2(x)$ vanishes for every $x \in \{0,1\}^n$? Since $g_1, \cdots, g_r$ have degree at most $d/2$, we can represent each polynomial g_i by $n^{O(d)}$ coefficients (say in a monomial basis). Thus in $n^{O(d)}$ time we can verify whether all the coefficients of $f - \sum_{i=1}^{r} g_i^2(x)$ are equal to zero or not by reducing it to a multilinear polynomial by repeatedly applying $x_i^2 = x_i$ (which holds for $x_i \in \{0,1\}$).

Let's consider $f : \{0,1\}^n \to \mathbb{R}$ which $\forall x \in \{0,1\}^n$ we have $f(x) \geq 0$. Is there always a certificate for f?

Proposition 4.1. *For non-negative $f : \{0,1\}^n \to \mathbb{R}$ there exists a degree-$2n$ SOS certificate.*

Proof. Using the fact that $\mathbb{I}\{x = y\} = \prod_{i \in Ones(y)} x_i^2 \prod_{y \in Zeros(y)} (1 - x_i)^2$ for $x, y \in \{0,1\}^n$, we may write

$$f(x) = \sum_{y \in \{0,1\}^n} (f(y) \cdot \mathbb{I}\{x = y\}) = \sum_{y \in \{0,1\}^n} f(y) \prod_{i \in Ones(y)} x_i^2 \prod_{y \in Zeros(y)} (1 - x_i)^2$$

then by construction we have found a certificate for f where $g_y(x) = \sqrt{f(y) \prod_{i \in Ones(y)} x_i^2 \prod_{y \in Zeros(y)} (1 - x_i)^2}$. $\square$

2.3 Finding certificates

We saw that we can check sos certificates efficiently. Also the following theorem shows that we can also find them in an efficient way. This *sum-of-squares algorithm* is based on semidefinite programming and has first been proposed by Naum Shor in 1987, later refined by Pablo Parrilo in 2000, and Jean Lasserre in 2001.

Theorem 4.1 (sum-of-squares algorithm-certificate version)**.** *There exists an algorithm that for a given function $f : \{0,1\}^n \to \mathbb{R}$, $k \in \mathbb{N}$ it outputs a degree-k sos certificate for $f + 2^{-n}$ in time $n^{O(k)}$ if f has a degree-k sos certificate.*

Theorem 4.2. *f has a degree-d sos certificate $\Leftrightarrow$ $\exists$ p.s.d matrix A such that $\forall x \in \{0,1\}^n$,*

$$f(x) = \left\langle (1,x)^{\otimes d/2}, A(1,x)^{\otimes d/2} \right\rangle.$$

Proof. First let's prove the $(\Leftarrow)$. If $A \succeq 0$, then we can find the following representation of A

$$A = \sum_i V_i V_i^T.$$

Then note that by $(1,x)^{\otimes d/2}$ we mean the vector $(1, x_i, \cdots, x_i x_j, \cdots)$ therfore we will have

$$\begin{aligned} \left\langle (1,x)^{\otimes d/2}, A(1,x)^{\otimes d/2} \right\rangle &= \left\langle (1,x)^{\otimes d/2}, \sum_i V_i V_i^T (1,x)^{\otimes d/2} \right\rangle \\ &= \sum_i \left\langle (1,x)^{\otimes d/2}, V_i V_i^T (1,x)^{\otimes d/2} \right\rangle \end{aligned}$$

Then note that

$$\begin{aligned}\left\langle (1,x)^{\otimes d/2}, V_i V_i^T (1,x)^{\otimes d/2} \right\rangle &= [1, x_i, \cdots, x_i x_j, \cdots] \begin{bmatrix} V_i \end{bmatrix} [V_i^T][1, x_i, \cdots, x_i x_j, \cdots]^T \\ &= \underbrace{([1, x_i, \cdots, x_i x_j, \cdots] \begin{bmatrix} V_i \end{bmatrix})}_{g_i(x)} ([1, x_i, \cdots, x_i x_j, \cdots] \begin{bmatrix} V_i \end{bmatrix})^T\end{aligned}$$

So defining $g_i(x) = V_i^T (1,x)^{\otimes d/2}$ will give us $f(x) = \sum g_i^2(x)$.

To prove ($\Rightarrow$) going in the backward direction of the previous argument will give us the desired A. $\square$

Note that based on this proof we can conclude if f has a sos certificate then we can find a degree-r sos certificate such that $r \leq n^{d/2}$.

Exercise 4.1. *For a graph $G = (V, E)$ consider its Laplacian $L_G = \sum_{(i,j)} (e_i - e_j)(e_i - e_j)^T$. Show that $\lambda_{max}(L_G)\frac{n}{2} - f_G$ has degree-2 sos certificate where f_G is the max cut polynomial.*

Exercise 4.2. *$\forall f : \{0,1\}^n \to \mathbb{R}$ with degree at most d for even $d \in \mathbb{N}$ there exists $M \in \mathbb{R}_{\geq 0}$ such that $M - f$ has degree-d sos certificate. Also M can be chosen $n^{O(d)}$ times the largest coefficient of f in the monomial basis.*

2.4 Degree-k sos certificate

First of all note that the the functions with degree-k sos certificate form a closed convex cone. Then by *hyperplane separation theorem* for convex cones, for every function $f : \{0,1\}^n \to \mathbb{R}$ without degree-k sos certificate there exists a hyperplane through the origin that separates f from the cone of functions with degree-k sos certificate in the sense that the halfspace H above that hyperplane contains the degree-k sos certificate cone but not f.

Now let's see how does such a halfspace look like? We can represent that halfspace by its normal function $\mu : \{0,1\}^n \to \mathbb{R}$ so that

$$H = \left\{ g \in \{0,1\}^n \to \mathbb{R} | \sum_{x \in \{0,1\}^n} \mu(x) g(x) \geq 0 \right\}$$

And by scaling without loss of generality we can assume that $\sum_{x\{0,1\}^n} \mu(x) = 1$. If we had $\mu(x) \geq 0$ for all $x \in \{0,1\}^n$ this μ would correspond to a probability distribution on the hypercube. In that case, the hyperplane H is simply the set of all the functions with nonnegative expectation with respect to the measure μ. Therefore for $f \notin H$ the expectation of f with respect to μ is negative which means that there exists at least one point x on the hypercube such that $f(x) < 0$.

The point is that μ is not necessarily nonnegative, there is no guarantee that $\mu(x) \geq 0$ for all $x \in \{0,1\}^n$. But still it behaves like a probability distribution in many ways. So let's formalize this idea. We are going to define a *pseudo-distribution* and *pseudo-expectation*

Definition 4.2 (Degree-d pseudo distribution)**.** *over the hypercube is a function $\mu : \{0,1\}^n \to \mathbb{R}$ such that for every polynomial f of degree at most $d/2$ we have $\tilde{\mathbb{E}}_\mu f^2 \geq 0$ and $\tilde{\mathbb{E}}_\mu 1 = 1$, where*

$$\tilde{\mathbb{E}}_\mu f = \sum_{x \in \{0,1\}^n} \mu(x) f(x)$$

*and we call it **pseudo expectation** of f.*

Lemma 4.1. *Suppose μ is a degree-ℓ pseudo distribution. Then there exists a multilinear polynomial μ' of degree at most ℓ such that*

$$\tilde{\mathbb{E}}_\mu p = \tilde{\mathbb{E}}_{\mu'} p \qquad \forall p \text{ of degree } \ell$$

Proof. Let $\mathcal{U}_\ell \subset \mathbb{R}^{\{0,1\}^n}$ be the linear subspace of multilinear polynomials of degree at most ℓ. Then this space contains all polynomials of degree at most ℓ. Decompose the function μ as $\mu = \mu' + \mu''$ such that $\mu' \in \mathcal{U}_\ell$ and $\mu'' \perp \mathcal{U}_\ell$. Then for every $p \in \mathcal{U}_\ell$ we have

$$\tilde{\mathbb{E}}_\mu p =< \mu' + \mu'', p >=< \mu', p >= \tilde{\mathbb{E}}_{\mu'} p$$

□

The notion of *pseudo expectation* can be easily extended to *vector valued* functions, in which case this denotes the vector obtained by taking expectation of every coordinate of f. Using this notion we can write the conclusion of last lemma more succinctly as

$$\tilde{\mathbb{E}}_\mu (1,x)^{\otimes \ell} = \tilde{\mathbb{E}}_{\mu'} (1,x)^{\otimes \ell}.$$

Exercise 4.3. *If μ has degree bigger than ℓ what is the projection of μ onto $\mathcal{U}_\ell$?*

Exercise 4.4. *Show that if μ is a degree-$2n$ pseudo distribution, then $\mu(x) \geq 0$ for all x.*

Exercise 4.5. *Show that $\mu : \{0,1\}^n \to \mathbb{R}$ is a pseudo distribution if and only if*

$$\tilde{\mathbb{E}}_\mu 1 = 1,$$

and

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

Exercise 4.6. *For all d and all pseudo distributions μ of degree d there exists a degree-d pseudo distribution μ' with the same pseudo moments up to degree d as μ such that*

$$|\mu'(x)| \leq 2^{-n} \sum_{d'=0}^{d} \binom{n}{d'}$$

Hint: Fourier Analysis.

Exercise 4.7. *Show that the set of degree-d pseudo distributions over $\{0,1\}^n$ admits a separation algorithm with running time $n^{O(d)}$. Concretely show that there exists an $n^{O(d)}$-time algorithm that given a vector $N \in (\mathbb{R}^n)^{\otimes d}$ outside of the following set χ_d outputs a halfspace that separates N from χ_d. Here χ_d is the set that consists of all coefficient vectors $M \in (\mathbb{R}^{n+1})^{\otimes d}$ such that the function $\mu : \{0,1\}^n \to \mathbb{R}$ with $\mu(x) =< M, (1,x)^{\otimes d} >$ is a degree-d pseudo distribution over $\{0,1\}^n$.*

Exercise 4.8. *Show that for every $d \in \mathbb{N}$, the following set of pseudo moments admits a separation algorithm with running time $n^{O(d)}$,*

$$\mathcal{M}_d = \Big\{ \tilde{\mathbb{E}}_\mu (1,x)^{\otimes d} \Big| \mu \text{ is deg-}d \text{ pseudo distribution over } \{0,1\}^n \Big\}.$$

3 Duality

Now we show that there is a dual relationship between sos certificates and pseudo distributions.

Theorem 4.3. *For all functions $f : \{0,1\}^n \to \mathbb{R}$ and every $d \in \mathbb{N}$, there exists a degree-d sos certificate for the non-negativity of f if and only if every degree-d pseudo distribution μ over $\{0,1\}^n$ satisfies $\tilde{\mathbb{E}}_\mu f \geq 0$.*

Proof. ($\Rightarrow$) If f has a degree-d sos certificate then we have $f(x) = \sum_i g_i^2(x)$ where g_i's are of degree at most $d/2$. Then for every degree-d pseudo distribution μ we have $\tilde{\mathbb{E}}_\mu g_i^2 \geq 0$ which will give $\tilde{\mathbb{E}}_\mu \sum g_i^2 = \tilde{\mathbb{E}}_\mu f \geq 0$.

($\Leftarrow$) If there is no degree-d sos certificate, then we want to show that there exists a pseudo distribution μ such that $\tilde{\mathbb{E}}_\mu f < 0$. Now by hyperplane separation theorem, there exists a halfspace H through the origin such that contains the cone but not f. Let $\mu : \{0,1\}^n \to \mathbb{R}$ be the normal of H so that

$$H = \left\{ g : \{0,1\}^n \middle| \tilde{\mathbb{R}}_\mu g \geq 0 \right\}.$$

Since $f \notin H$ we know that $\tilde{\mathbb{E}}_\mu f < 0$. Since H contains the degree-d sos cone, every polynomial g of degree at most $d/2$ satisfies $\tilde{\mathbb{E}}_\mu g^2 \geq 0$. It remains to argue that $\tilde{\mathbb{E}}_\mu 1 = 1$, which means that we can rescale μ by a nonnegative factor to ensure that $\tilde{\mathbb{E}}_\mu 1 = 1$. In fact by one of the exercises we know that there exists $M \in \mathbb{R}_{\geq 0}$ such that $M + f$ has a degree-d sos certificate, which means that

$$\tilde{\mathbb{E}}_\mu 1 = \frac{1}{M} \left(\tilde{\mathbb{E}}_\mu M + f - \tilde{\mathbb{E}}_\mu f \right) > 0,$$

as desired. $\square$