Lecture 8: Spectrahedral lifts and positive semi-definite rank
CSE 599S: Entropy optimality, Winter 2016
Instructor: James R. Lee **Last updated:** February 15, 2016

1 PSD rank and sums-of-squares degree

We have previously explored whether the cut polytope can be expressed as the linear projection of a polytope with a small number of facets (i.e., whether it has a small linear programming extended formulation).

For many cut problems, semi-definite programs (SDPs) are able to achieve better approximation ratios than LPs. The most famous example is the Goemans-Williamson 0.878-approximation for MAX-CUT. The techniques we have seen so far (see [Chan-Lee-Raghavendra-Steurer 2013]) are able to show that no polynomial-size LP can achieve better than factor 1/2.

The goal now is to give an indication, following [Lee-Raghavendra-Steurer 2015], of how one can prove similarly that small SDPs cannot capture the cut polytope.

1.1 Spectrahedral lifts

The feasible regions of LPs are polyhedra. Up to linear isomorphism, every polyhedron P can be represented as $P = \mathbb{R}^n_+ \cap V$ where $\mathbb{R}^n_+$ is the positive orthant and $V \subseteq \mathbb{R}^n$ is an affine subspace.

In this context, it makes sense to study other cones that can be optimized over efficiently. A prominent example is the positive semi-definite cone. Let us define $\mathcal{S}^n_{\mathrm{sym}} \subseteq \mathbb{R}^{n^2}$ to be the set of real, symmetric $n \times n$ matrices. By the spectral theorem, every $A \in \mathcal{S}^n_{\mathrm{sym}}$ can be written $A = P^T DP$ where D is a diagonal matrix containing the eigenvalues of A (which are all real), and P is an orthogonal matrix. If A has any repeated eigenvalues, this representation will not be unique.

We let $\mathcal{S}^n_+ \subseteq \mathbb{R}^{n^2}$ denote the subset of *positive semi-definite* matrices, i.e. those with all non-negative eigenvalues. A *spectrahedron* is the intersection $\mathcal{S}^n_+ \cap V$ with an affine subspace V.

In analogy with the γ parameter we defined for polyhedral lifts, let us define $\bar{\gamma}_{\mathsf{sdp}}(P)$ for a polytope P to be the minimal dimension of a spectrahedron that linearly projects to P.

Exericse (0.5 points) 1.1. Show that $\bar{\gamma}_{\mathsf{sdp}}(P) \leqslant \bar{\gamma}(P)$ for every polytope P. In other words, spectahedral lifts are at least as powerful as polyhedral lifts in this model.

In fact, spectrahedral lifts can be strictly more powerful. Certainly there are many examples of this in the setting of approximation (like the Goemans-Williamson SDP mentioned earlier), but there are also recent gaps between $\bar{\gamma}$ and $\bar{\gamma}_{\mathsf{sdp}}$ for exact characterizations of polytopes; see the work of Fawzi, Saunderson, and Parrilo (2015).

Nevertheless, we are now capable of proving strong lower bounds on the dimension of such lifts. Let us consider the cut polytope CUT_n as in previous posts.

Theorem 1.2 (Lee-Raghavendra-Steurer 2015)**.** *There is a constant $c > 0$ such that for every $n \geqslant 1$, $\bar{\gamma}_{\mathsf{sdp}}(\mathrm{CUT}_n) \geqslant e^{cn^{2/11}}$.*

Our goal now is to understand how the general framework we have seen for LP lower bounds extends to the SDP setting.

1.2 PSD rank and factorizations

Just as in the setting of polyhedra, there is a notion of "factorization through a cone" that characterizes the parameter $\bar{\gamma}_{\mathsf{sdp}}(P)$. Let $M \in \mathbb{R}_+^{m\times n}$ be a non-negative matrix. One defines the *psd rank* of M as the quantity

$$\mathrm{rank}_{\mathsf{psd}}(M) = \min\left\{r : M_{ij} = \mathrm{Tr}(A_i B_j) \text{ for some } A_1,\ldots,A_m,B_1,\ldots,B_n \in \mathcal{S}_+^r\right\} .$$

The following theorem was independently proved by Fiorini-Massar-Pokutta-Tiwari-de Wolf and Gouveia-Parrilo-Thomas. The proof is a direct analog of Yannakakis' proof for non-negative rank.

Theorem 1.3. *For every polytope P, it holds that $\bar{\gamma}_{\mathsf{sdp}}(P) = \mathrm{rank}_{\mathsf{psd}}(M)$ for any slack matrix M of P.*

Recall the class QML_n^+ of non-negative quadratic multi-linear functions that are positive on $\{0,1\}^n$ and the matrix $\mathcal{M}_n : \mathsf{QML}_n^+ \times \{0,1\}^n \to \mathbb{R}_+$ given by

$$\mathcal{M}_n(f,x) = f(x)\,.$$

We saw previously that $\mathcal{M}_n$ is a submatrix of some slack matrix of CUT_n. Thus our goal is to prove a lower bound on $\mathrm{rank}_{\mathsf{psd}}(\mathcal{M}_n)$.

1.3 Sum-of-squares certificates

Just as in the setting of non-negative matrix factorization, we can think of a low psd rank factorization of $\mathcal{M}_n$ as a small set of "axioms" that can prove the non-negativity of every function in QML_n^+. But now our proof system is considerably more powerful.

For a subspace of functions $\mathcal{U} \subseteq L^2(\{0,1\}^n)$, let us define the cone

$$\mathsf{sos}(\mathcal{U}) = \mathrm{cone}\left(q^2 : q \in \mathcal{U}\right) .$$

This is the cone of squares of functions in $\mathcal{U}$. We will think of $\mathcal{U}$ as a set of axioms of size $\dim(\mathcal{U})$ that is able to assert non-negativity of every $f \in \mathrm{sos}(\mathcal{U})$ by writing

$$f = \sum_{i=1}^{k} q_i^2$$

for some $q_1,\ldots,q_k \in \mathsf{sos}(\mathcal{U})$.

Fix a subspace $\mathcal{U}$ and let $r = \dim(\mathcal{U})$. Fix also a basis $q_1,\ldots,q_r : \{0,1\}^n \to \mathbb{R}$ for $\mathcal{U}$.

Define $B : \{0,1\}^n \to \mathcal{S}_+^r$ by setting $B(x)_{ij} = q_i(x)q_j(x)$. Note that $B(x)$ is PSD for every x because $B(x) = \vec{q}(x)\vec{q}(x)^T$ where $\vec{q}(x) = (q_1(x),\ldots,q_r(x))$.

We can write every $p \in \mathcal{U}$ as $p = \sum_{i=1}^r \lambda_i q_i$. Defining $\Lambda(p^2) \in \mathcal{S}_+^r$ by $\Lambda(p^2)_{ij} = \lambda_i\lambda_j$, we see that

$$\mathrm{Tr}(\Lambda(p^2)Q(x)) = \sum_{i,j} \lambda_i\lambda_j q_i(x)q_j(x) = p(x)^2\,.$$

Now every $f \in \mathsf{sos}(\mathcal{U})$ can be written as $\sum_{i=1}^k c_i p_i^2$ for some $k \geqslant 0$ and $\{c_i \geqslant 0\}$. Therefore if we define $\Lambda(f) = \sum_{i=1}^k c_i\Lambda(p_i^2)$, we arrive at the representation

$$f(x) = \mathrm{Tr}(\Lambda(f)Q(x))\,.$$

In conclusion, if $\mathsf{QML}_+^n \subseteq \mathsf{sos}(\mathcal{U})$, then $\mathrm{rank}_{\mathsf{psd}}(\mathcal{M}_n) \leqslant \dim(\mathsf{sos}(\mathcal{U}))$.

Exercise (1 point) 1.4. First, show that an approximate converse holds: $\dim(\mathsf{sos}(\mathcal{U})) \leqslant \mathrm{rank}_{\mathsf{psd}}(\mathcal{M}_n)^2$.

Now let us describe how to get an exact characterization of the same form. Let $L^2(\{0,1\}^n; \ell_2)$ denote the Hilbert space of functions $f : \{0,1\}^n \to \ell_2$, where ℓ_2 denotes the usual space of infinite sequences (x_i) of real numbers equipped with the norm $\|(x_i)\|_2 = \sqrt{\sum_{i=1}^{\infty} x_i^2}$.

For a subspace $\mathcal{U} \subseteq L^2(\{0,1\}^n, \ell_2)$, define

$$\mathsf{sos}(\mathcal{U}) = \mathrm{cone}\left(\|q\|_{L^2}^2 : q \in \mathcal{U}\right) ,$$

where we use the norm given by

$$\|q\|_{L^2}^2 = \mathbb{E}_{x \in \{0,1\}^n} \|q(x)\|_2^2 .$$

Prove that

$$\mathrm{rank}_{\mathsf{psd}}(\mathcal{M}_n) = \min\left\{\dim(\mathsf{sos}(\mathcal{U})) : \mathsf{QML}_+^n \subseteq \mathsf{sos}(\mathcal{U})\right\} .$$

1.4 The canonical axioms

And just as d-juntas were the canonical axioms for our "non-negative matrix factorization" proof system, there is a similar canonical family in the SDP setting: Let Q_d be the subspace of all degree-d multi-linear polynomials on $\mathbb{R}^n$. We have

$$\dim(Q_d) \leqslant \sum_{k=0}^{d} \binom{n}{k} \leqslant 1 + n^d . \tag{1.1}$$

For a function $f : \{0,1\}^n \to \mathbb{R}_+$, one defines

$$\deg_{\mathsf{sos}}(f) = \min\{d : f \in \mathsf{sos}(Q_d)\} .$$

(One could debate whether the definition of sum-of-squares degree should have $d/2$ or d.)

On the other hand, our choice has the following nice property.

Lemma 1.5. *For every $f : \{0,1\}^n \to \mathbb{R}$, we have $\deg_{\mathsf{sos}}(f) \leqslant \deg_J(f)$.*

Proof. If q is a non-negative d-junta, then $\sqrt{q}$ is also a non-negative d-junta. It is elementary to see that every d-junta on $\{0,1\}^n$ has a multi-linear polynomial representation of degree at most d, thus q is the square of a multi-linear polynomial of degree at most d. □

1.5 The dual cone

As with junta-degree, there is a simple characterization of sos-degree in terms of separating functionals. Say that a functional $\varphi : \{0,1\}^n \to \mathbb{R}$ is *degree-d pseudo-positive* if

$$\langle \varphi, q^2 \rangle = \mathop{\mathbb{E}}_{x \in \{0,1\}^n} \varphi(x) q(x)^2 \geqslant 0$$

whenever $q : \{0,1\}^n \to \mathbb{R}$ satisfies $\deg(q) \leqslant d$ (and by deg here, we mean degree as a multi-linear polynomial on $\{0,1\}^n$).

Again, since $\mathsf{sos}(Q_d)$ is a closed convex cone, these separating functionals are the only way of exhibiting non-membership.

Exercise (1 point) 1.6. For every $f : \{0,1\}^n \to \mathbb{R}_+$, it holds that $\deg_{\mathsf{sos}}(f) > d$ if and only if there is a degree-d pseudo-positive functional $\varphi : \{0,1\}^n \to \mathbb{R}$ such that $\langle \varphi, f \rangle < 0$.

1.6 The connection to psd rank

Following the analogy with non-negative rank, we have two objectives left: (1) to exhibit a function $f \in \mathsf{QML}_n^+$ with $\deg_{\mathsf{sos}}(f)$ large, and (ii) to give a connection between the sum-of-squares degree of f and the psd rank of an associated matrix.

Notice that the function $g(x) = (1 - \sum_{i=1}^m x_i)^2$ we used for junta-degree has $\deg_{\mathsf{sos}}(g) = 1$, making it a poor candidate. In fact, this implies that $\mathrm{rank}_{\mathsf{psd}}(M_n^g) \leqslant O(n)$, while we have seen that $\mathrm{rank}_+(M_n^g) \geqslant \Omega((n/\log n)^m)$ as $n \to \infty$.

Fortunately, Grigoriev has shown that the *knapsack polynomial* has large sos-degree.

Theorem 1.7. *For every odd $m \geqslant 1$, the function*

$$f(x) = \left(\frac{m}{2} - \sum_{i=1}^{m} x_i\right)^2 - \frac{1}{4}$$

has $\deg_{\mathsf{sos}}(f) \geqslant \lfloor m/2 \rfloor$.

Observe that this f is non-negative over $\{0,1\}^m$ (because m is odd), but it is manifestly *not* non-negative on $\mathbb{R}^m$.

Finally, we recall the submatrices of $\mathcal{M}_n$ defined as follows. Fix some integer $m \geqslant 1$ and a function $g : \{0,1\}^m \to \mathbb{R}_+$. Then $M_n^g : \binom{[n]}{m} \times \{0,1\}^n \to \mathbb{R}_+$ is given by

$$M_n^g(S, x) = g(x|_S)\,.$$

Our goal in the coming lecture is to sketch the following analog of Theorem **??**.

Theorem 1.8 (Lee-Raghavendra-Steurer 2015)**.** *For every $m \geqslant 1$ and $g : \{0,1\}^m \to \mathbb{R}_+$, there exists a constant $C(g)$ such that the following holds. For every $n \geqslant 2m$,*

$$\mathrm{rank}_{\mathsf{psd}}(M_n^g) \geqslant C\left(\frac{n}{\log n}\right)^{d/2},$$

where $d = \deg_{\mathsf{sos}}(g)$.

1.7 Exercise (2 points): Proving a lower bound on $\deg_{\mathsf{sos}}$

[This exercise follows an elegant argument of J. Kaniewski, T. Lee, and R. de Wolf (2014).]

You will prove a lower bound on the sum-of-squares degree of the function $f : \{0,1\}^n \to \mathbb{R}_+$ given by

$$f(x) = (|x| - 1)(|x| - 2)\,, \tag{1.2}$$

where we use $|x| = \sum_{i=1}^n x_i$ for the hamming weight of $x \in \{0,1\}^n$.

Suppose that we can write

$$f(x) = \sum_{i=1}^{N} p_i(x)^2\,,$$

where $\deg(p_i) \leqslant d$ for every $i \in [N]$. Define the function $q_i : [n] \to \mathbb{R}$ by

$$q_i(k) = \mathop{\mathbb{E}}_{x \in \{0,1\}^n : |x|=k} [p_i(x)] .$$

The first step can be accomplished using the Fourier representation of functions on $\{0,1\}^n$, or using an appropriate averaging procedure.

(a) Show that there is a function $\tilde{q}_i : \mathbb{R} \to \mathbb{R}$ that agrees with q_i on $[n]$ and such that $\deg(\tilde{q}_i) \leqslant d$.

Now let us define $Q(t) = \sum_{i=1}^{N} \tilde{q}_i(t)^2$, which is a polynomial of degree at most $2d$. We also have $Q(1) = Q(2) = 0$ since $f(x) = 0$ for $|x| \in \{1,2\}$. The zeroes of a non-negative real polynomial must have multiplicity at least 2, thus we can write

$$Q(t) = (t-1)^2(t-2)^2 q(t)$$

for some polynomial q with $\deg(q) \leqslant 2d - 4$.

(b) Your goal now is to prove a lower bound $\deg(q) \geqslant \Omega(\sqrt{n})$, implying that $\deg_{\mathsf{sos}}(f) \geqslant \Omega(\sqrt{n})$. [Note that plugging this into Theorem 1.8 is enough to show that $\bar{\gamma}_{\mathsf{sdp}}(\mathrm{CUT}_n)$ must grow faster than any polynomial.]

You should be able to do this using the following oft-employed lemma of A. A. Markov.

Lemma 1.9. *If $q : \mathbb{R} \to \mathbb{R}$ is a polynomial, then for every $T \geqslant 0$,*

$$\deg(q) \geqslant \sqrt{\frac{T}{2} \frac{\max_{x \in [0,T]} |q'(x)|}{\max_{x \in [0,T]} |q(x)|}} .$$