CAS CS 599 B: Mathematical Methods for TCS

Lecturer: Mark Bun | Spring 2022

Lecture Notes 25:

Pseudoexpectations, Sum-of-Squares

Reading.

- Fleming, Kothari, and Pitassi, "Semialgebraic Proofs and Efficient Algorithm Design"

Last time, we introduced the Sherali-Adams hierarchy of LP relaxations. Given an integer linear program of the form

$$\begin{aligned} \max \quad & \langle \vec{c}, \vec{x} \rangle \\ \text{s.t.} \quad & \langle \vec{a}_1, \vec{x} \rangle \geq 0 \\ & \vdots \\ & \langle \vec{a}_m, \vec{x} \rangle \geq 0 \\ & x_i \in \{0, 1\} \ \forall i \in [n] \end{aligned}$$

we construct an LP on $\binom{n}{\leq d}$ variables and $O(m \cdot \binom{n}{\leq d})$ constraints via the process:

Extend: Include every constraint of the form $\prod_{i \in S} x_i \prod_{j \in T}(1 - x_i) \geq 0$ or $\langle \vec{a}_i, \vec{x} \rangle \prod_{i \in S} x_i \prod_{j \in T}(1 - x_i) \geq 0$ where the LHS has degree at most d.

Linearize: Multilinearize each constraint by replacing each occurrence of x_i^c with x_i, and then each occurrence of $\prod_{i \in S} x_i$ with a new variable y_S.

We motivated $SA(d)$ as a proof system that can be used to certify an upper bound on the objective of the original LP. That is, if our $SA(d)$ relaxation takes the form

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} c_i y_{\{i\}} \\ \text{s.t.} \quad & p_1(\vec{y}) \geq 0 \\ & \vdots \\ & p_N(\vec{y}) \geq 0, \end{aligned}$$

then if we can express the objective function as $\beta - \sum_{i=1}^{n} c_i y_{\{i\}} = \lambda_1 p_1(\vec{y}) + \cdots + \lambda_N p_N(\vec{y})$, we can certify an upper bound of OPT $\leq$ SA-OPT $\leq \beta$ for the original combinatorial optimization problem.

1 Pseudoexpectations

The LP dual interpretation of the SA relaxation gives rise to the certification interpretation above. But the primal LP itself has a natural interpretation in terms of approximation algorithms as well. Let $(\alpha_S)_{|S|\le d}$ be a feasible solution to $SA(d)$ Then for each monomial $\prod_{i\in S} x_i$ over the original variables, define the *pseudoexpectation* operator associated to this solution by

$$\widetilde{\mathbb{E}}\left[\prod_{i\in S} x_i\right] := \alpha_S.$$

We can also extend the pseudoexpectation operator to arbitrary polynomials by linearity:

$$\widetilde{\mathbb{E}}\left[\sum_{|S|\le d} c_S \prod_{i\in S} x_i\right] := \sum_{|S|\le d} \alpha_S.$$

Then the $SA(d)$ relaxation can be equivalently stated as

$$\begin{aligned} \max \quad & \widetilde{\mathbb{E}}[\langle \vec{c}, \vec{x}\rangle] \\ \text{s.t.} \quad & \widetilde{\mathbb{E}}[p_i(\vec{x})] \ge 0 \quad \forall i = 1, \ldots, N. \end{aligned}$$

where the maximum is taken over all possible pseudoexpectation operators $\widetilde{\mathbb{E}}$, or equivalently, over the variables $\widetilde{\mathbb{E}}[\prod_{i\in S} x_i]$ for all $|S| \le d$.

A pseudoexpectation is a linear functional that behaves like an expectation on polynomials of degree at most d. The way to think about a solution to this LP is as a pseudodistribution over solutions to the original ILP. Here, a pseudodistribution is an object that behaves like a distribution as far as polynomials of degree at most d are concerned. To see what I mean:

- Let $v_1, \ldots, v_n \in \{0, 1\}$ be a feasible solution to the original ILP. Then setting $\widetilde{\mathbb{E}}[p(\vec{x})] = p(\vec{v})$ for every degree-d polynomial p gives a solution to this relaxed LP with objective value $\langle \vec{c}, \vec{v}\rangle$. This sanity checks our assertion that the SA relaxation is actually a relaxation.

- Let D be a distribution over feasible solutions $\vec{v} \in \{0, 1\}^n$ to the ILP. Then if we set $\widetilde{\mathbb{E}}[p(\vec{x})] = \mathbb{E}_{\vec{v}\sim D}[p(\vec{v})]$, then we get a solution with objective value $\mathbb{E}_{\vec{v}\sim D}[\langle \vec{c}, \vec{v}\rangle]$.

- A degree-d *pseudodistribution* is a collection $(D_S)_{|S|\le d}$ of distributions D_S over assignments to the variables in S such that for all $T \subseteq S$ with $|S| \le d$, the marginal distribution of D_S over variables in T is equal to D_T. Setting $\widetilde{\mathbb{E}}[\prod_{i\in S} x_i] = \mathbb{E}_{\vec{v}\sim D_S}[\prod_{i\in S} v_i]$ also gives a feasible solution to the SA relaxation. One can show that expectations over degree-d pseudodistributions exactly characterize degree-d pseudoexpectations.

2 Sum-of-Squares (Lasserre) Hierarchy

Just as Sherali-Adams is a hierarchy of progressively tighter LP relaxations for combinatorial optimization problems, we can define an analogous hierarchy of SDP relaxations. Recall that the idea behind Sherali-Adams was to augment our constraints by using the fact that non-negative juntas $\prod_{i\in S} x_i \prod_{j\in T}(1 - x_j)$ of degree at most d are non-negative over $\{0,1\}$. Sum-of-squares takes this a step farther by adding a constraint for all degree-d polynomials $p(\vec{x})$ that can be expressed as the square of a polynomial. To see why this is at least as powerful as Sherali-Adams, we observe the fact that:

Fact 1. *A multilinear polynomial $p : \{0,1\}^d \to \mathbb{R}$ is non-negative iff $p = q^2$ for a multilinear polynomial $q : \{0,1\}^d \to \mathbb{R}$.*

Proof. Let q be the (unique) multilinear polynomial such that $q(x) = \sqrt{p(x)}$ for every $x \in \{0,1\}^d$. □

Thus, an equivalent way to view $SA(d)$ is as including all constraints of the form $(q(x))^2 \geq 0$ where q is a polynomial depending on at most d variables. Meanwhile, $SOS(2d)$ includes all constraints of the form $(q(x))^2$ where q is any polynomial of degree at most d.

Goemans-Williamson Re-revisited. Recall our formulation of the Max-Cut problem as a quadratic integer program.

$$\begin{aligned} \max \qquad & \sum_{(i,j)\in E} \frac{1}{2} - \frac{1}{2}x_i x_j \\ \text{s.t. } & x_i^2 = 1 \qquad \forall i \in [n]. \end{aligned}$$

To get the $SOS(2)$ relaxation of this program, we first extend the constraints to enforce $(q(\vec{x}))^2 \geq 0$ for every linear polynomial q. Since a linear polynomial can be written as $\langle \vec{a}, \vec{x} \rangle$[1], we get that this is equivalent to:

$$0 \leq \langle \vec{a}, \vec{x} \rangle^2 = \langle \vec{a}, \vec{x}\vec{x}^T \vec{a} \rangle$$

for every $\vec{a}$, or in other words, $\vec{x}\vec{x}^T$ is PSD.

Next, we multilinearize and then linearize by replacing each occurrence of $x_i x_j$ with $y_{\{i,j\}}$. This gives us

$$\begin{aligned} \max \qquad & \sum_{(i,j)\in E} \frac{1}{2} - \frac{1}{2}y_{\{i,j\}} \\ \text{s.t. } & y_{\{i\}} = 1 \qquad \forall i \in [n] \\ & Y = (y_{\{i,j\}}) \succeq 0. \end{aligned}$$

This is exactly the SDP relaxation we saw before.

Another classic application was given by Arora, Rao, and Vazirani, who showed that degree-4 SOS can be used to estimate the conductance of a graph. This is sometimes called the "sparsest cut" problem and is known to be NP-hard. Recall Cheeger's inequality, which says $\nu_2/2 \leq \phi(G) \leq \sqrt{2\nu_2(G)}$. It implies that by computing the second eigenvalue of the normalized Laplacian, one gets an $(O(\sqrt{\beta}), \beta)$-certification algorithm for the sparsest cut, i.e., if $\phi(G) \leq \beta$ one can certify that $\phi(G) \leq O(\sqrt{\beta})$. It turns out that more is true; one can interpret the proof of Cheeger's inequality as a degree-2 SOS certificate, as well as recover a cut with value at most $O(\sqrt{\beta})$.

Improving on an $(O(\log n)\beta, \beta)$-approximation algorithm of Leighton and Rao, the ARV result gave an $(O(\sqrt{\log n})\beta, \beta)$-approximation algorithm.

[1]Lying a bit, since these are only linear homogeneous polynomials. Those with constant terms turn out not to add anything.

Other SOS Facts The degree-$2d$ SOS relaxation gives rise to an SDP over $n^{O(d)}$ variables with (most of the time) can be solved in time $n^{O(d)}$. The interpretation of Sherali-Adams in terms of pseudoexpectations and pseudodistributions also holds.