CS 229r Information Theory in Computer Science April 25, 2019

Lecture 24: "Barriers" to Optimization

Instructor: Madhu Sudan *Scribe: Shuran Zheng*

1 Overview

Today we're going to talk about a line of works started a little bit before 1985, when some people tried to show that the Travel Salesman Problem could be solved by LP in polynomial time. Finally Yannakakis [1991] came up with the beautiful paper that killed the entire approach. We'll talk about how the approach was supposed to be and how it was killed. Below is the outline for today's class.

- LP, Max Cut, *extended formulations*.
- Lower bound for extended formulation via nondeterministic communication complexity.
- Non-Det-CC(Unique Disj) $= \Omega(n)$.
- Max Cut extended formulation lower bound $= 2^{\Omega(n)}$.

Next time we'll talk about Quantum Information.

2 Extended Formulations

2.1 Linear Programming

We'll start with the *linear programming*. A linear program can be defined as

$$\begin{array}{cc} \max_{x\in\mathbb{R}^n} & c^T x \\ \text{s.t.} & Ax \le b \end{array}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$. Geometrically, each constraint restricts the feasible solutions to a half-space, and the feasible region of a LP is the convex set defined by the intersection of these half-spaces. Our goal is to maximize a linear function within this convex set. It is known that linear programs can be optimized in polynomial time.

2.2 Cut Polytope

We now show how LP can possibly be used to solve Max Cut, which is also a NP-hard problem as TSP. Consider a complete graph on n nodes with $n(n-1)/2$ edges. The characteristic vector χ_S of a cut $S, \overline{S}$ is a vector with length equal to the number of edges, each coordinate of which represents whether an edge is in the cut-set. More specifically,

$$\chi_S(e) = \begin{cases} 1, & \text{if } e \text{ goes from one of } S, \overline{S} \text{ to the other} \\ 0, & \text{otherwise} \end{cases}$$

The cut polytope is defined as the convex hull of all possible characteristic vectors of cuts,

$$\mathcal{P}_{cut} = ConvexHull(\chi_S | S \subseteq [n]). \tag{1}$$

Then the Max Cut problem can be defined as a LP. Given a vector $w \in \mathbb{R}^{n(n-1)/2}$ that represents the weights of the edges of a given graph G, the Max Cut problem is just

$$\begin{array}{ll} \max & w^T x \\ \text{s.t.} & x \in \mathcal{P}_{cut} \end{array}$$

But we know that the Max Cut problem is NP-hard, so it cannot be solved efficiently by a LP assuming P $\neq$ NP. Where does the gap come from? One guess is that the solution needs to be integral. This is actually not a problem as we can always find an optimal solution that is an extreme point, which should automatically be integral. The other guess is that this formulation may take exponentially many constraints to define $\mathcal{P}_{cut}$. In other words, $\mathcal{P}_{cut}$ may have exponentially many facets. But it turns out that this is also not enough for the problem to be hard.

2.3 Definition

We introduce a method that can possibly give a polynomial-size description of a polytope P with exponentially many facets. The approach is to add some auxiliary decision variables and define constraints in a higher-dimensional space. The idea is that this complicated polytope P may be a "shadow" of some simple polytope Q from a higher-dimensional space, as projecting onto a subset of variables can blow up the number of facets. Below is the formal definition.

Definition 1 (Extended formulations)**.** *For a polytope $P = \{x|Ax \leq b\} \subseteq \mathbb{R}^n$, polytope $Q = \{z|A'z \leq b'\} \subseteq \mathbb{R}^{n+m}$ is an extended formulation of P if $m = poly(n)$, and*

$$P = \{x \mid \exists y \in \mathbb{R}^m \text{ s.t. } (x, y) \in Q\}$$

One may ask if there really exists large P with small Q. The answer is yes. One example is as follows.

Exercise 2. *Consider the polytope $P \subseteq \{(x_1, \ldots, x_n)\} = \mathbb{R}^n$ defined by the exponentially many constraints*

$$\begin{aligned} &\sum_{i \in S} x_i \geq B, \text{ for all } S \subseteq [n], |S| \geq k \\ &0 \leq x_i \leq 1. \end{aligned}$$

Give a small (polynomial-size) extended formulation Q for P.

Solution. We add variables $p_1, \ldots, p_n$ and r. The extended formulation Q is defined as follows

$$\begin{gathered} kr + \textstyle\sum_{i=1}^n p_i \geq B, \\ x_i \geq p_i + r, \\ p_i \leq 0, \\ 0 \leq x_i \leq 1. \end{gathered}$$

Firstly for any subset $S \subseteq [n]$ with $|S| = k$, the above constraints should guarantee that

$$\sum_{i \in S} x_i \geq \sum_{i \in S} (p_i + r) = \sum_{i \in S} p_i + kr \geq \sum_{i=1}^n p_i + kr \geq B.$$

Then we show that for any $x \in P$ there exist corresponding $p_1, \ldots, p_n$ and r that satisfy the above constraints. For a point $(x_1, \ldots, x_n)$, let's order its coordinates from smallest to largest to get $(x_{(1)}, \ldots, x_{(n)})$. It suffices to define

$$\begin{gathered} r = x_{(k)} \\ p_i = \min\{x_i - r, 0\}. \end{gathered}$$

□

3 Max Cut Extended Formulation Lower Bound

So we want to prove that there exists *no* small extended formulation of the cut polytope $\mathcal{P}_{cut}$. We will create an artificial hard communication problem Face-Vertex(P) so that if there exists a nice extended formulation for polytope P, then there will exist an associated (very esoteric) communication protocol related to P that is "easy".

We first define *faces* of a polytope. Basically faces = all "boundary" surfaces. For example, a cube has

- eight 0-dimensional faces, which are the vertices of the cube,
- twelve 1-dimensional faces, which are the edges,
- six 2-dimensional faces, which are the facets,
- one 3-dimensional face, which is the cube.

3.1 Face-Vertex Problem

The communication task Face-Vertex(P) is as follows.

- Alice knows a vertex v of P.
- Bob knows a face f of P.
- The goal is to design a communication protocol, so that Alice and Bob output 1 if $v \notin f$, and output 0 if $v \in f$.

The minimum worst-case number of bits used by any protocol Π that gives the correct output is the *deterministic communication complexity of the problem*, denoted by Det-CC(Face-Vertex(P)),

$$\text{Det-CC(Face-Vertex}(P)) = \min_{\Pi} \max_{v \in P, f \subseteq P} \#\text{ of bits used by } \Pi \text{ to decide if } v \in f.$$

3.2 Nondeterministic Communication Complexity

Our proof will involve another CC notion, *nondeterministic communication complexity.* Suppose now we have Merlin, who knows both v and f. Merlin can send a message m to both Alice and Bob, and Alice and Bob never communicate but just output a bit according to Merlin's message. The goal is to design the protocol so that

- If $v \notin f$, then there exists a message m such that both Alice and Bob output 1.
- If $v \in f$, then for all message m, at least one of Alice and Bob outputs 0.

The minimum worst-case number of bits required for m is just the nondeterministic communication complexity of the problem, denoted by Non-Det-CC(Face-Vertex(P)). In this definition, Alice and Bob are given a "proof" from Merlin. They do not communicate with each other, but just verify the proof m.

A digression on the nondeterministic communication complexity: a well-known result for Non-Det-CC is that for any problem Π,

$$\text{Det-CC}(\Pi) \leq \text{Non-Det-CC}(\Pi) \cdot \text{Non-Det-CC}(\overline{\Pi}),$$

where $\overline{\Pi}$ is the complement of Π by flipping 0s and 1s. For a lot of problems (e.g. Set Disjointness, Equality) with high Det-CC, they have very low Non-Det-CC for one of Π, $\overline{\Pi}$. So by the above inequality, we can argue that the Non-Det-CC of their complements is high.

3.3 Yannakakis's Lemma

We're going to use the following lemma without a formal proof.

Lemma 3 (Yannakakis's Lemma)**.** *If a polytope P has an extended formulation Q with r facets, then Face-Vertex(P) has Non-Det-CC $\leq \log r$.*

We outline the idea of the proof. Suppose a polytope P has an extended formulation Q with r facets. For a vertex $v \in P$ and a face f of P, we define convex sets $v^*, f^* \subseteq Q$,

$$v^* = \{(v, y) | (v, y) \in Q\}$$
$$f^* = \{(x, y) | (x, y) \in Q, x \in f\}$$

So if $v \notin f$, $v^* \cap f^* = \emptyset$. It can be proved that (we're not going to prove this)

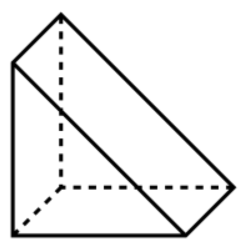
$$v \notin f \text{ if and only if there exists a facet } \widetilde{f} \text{ of } Q \text{ s.t. } f^* \subseteq \widetilde{f} \text{ and } v^* \nsubseteq \widetilde{f}.$$

Therefore Merlin can just send the index of $\tilde{f}$ with length $O(\log r)$, and Alice and Bob can compute v^* and f^* by themselves and test the above conditions $f^* \subseteq \widetilde{f}$ and $v^* \nsubseteq \widetilde{f}$ respectively.

One may wonder whether the condition $v^* \nsubseteq \widetilde{f}$ can be replaced by a simpler condition $v^* \cap \widetilde{f} = \emptyset$. The answer is no.

Exercise 4. *Find an example of v, f, P, Q with $v \notin f$, but the only facets $\widetilde{f}$ of Q containing f^* intersect v^*.*

Solution. We can choose P to be a square and Q to be a triangular-prism in $\mathbb{R}^3$ above the square.



If we choose v to be one of the vertices on the left and let f be the edge on the right, the only facets $\widetilde{f}$ of Q containing f^* intersect v^*. □

Note that the above proof sketch does not imply that the Non-Det-CC of the complement of Face-Vertex(P) is $\leq \log r$ (in the complement of Face-Vertex(P), Merlin wants to prove that $v \in f$).

Exercise 5. *Show that the existence of a single facet $\widetilde{f}$ with $f^* \subseteq \widetilde{f}$ and $v^* \subseteq \widetilde{f}$ does not imply $v^* \subseteq f^*$.*

Solution. Again consider the triangular-prism example above where P is the square and Q is triangular-prism. Let v be the right front vertex and let f be the right back vertex. Then the upper facet of Q contains both v^* and f^* but $v^* \nsubseteq f^*$. □

3.4 Extended Formulation Lower Bound

We now prove the lower bound. Lemma 3 reduces the task of proving lower bounds on the size of extended formulations of the cut-polytope $\mathcal{P}_{cut}$ to proving lower bounds on the nondeterministic communication complexity of Face-Vertex($\mathcal{P}_{cut}$).

We will then prove a lower bound of Non-Det-CC(Face-Vertex($\mathcal{P}_{cut}$)) by reduction from the Unique Disjointness problem, which is known to be a high-complexity problem.

Theorem 6. *Non-Det-CC(Unique Disjointness) $= \Omega(n)$.*

The proof of the theorem can be found in Section 5.4.5 of [Roughgarden et al., 2016]. To reduce the Unique Disjointness problem to Face-Vertex, for each Disjointness problem instance $\text{Disj}(S, R)$, we will construct a communication problem where Alice has a vertex associated with S, denoted by χ_S (with a little abuse of notation), and Bob has a facet associated with R, denoted by H_R, so that

$$\chi_S \in H_R, \text{ if and only if } |S \cap R| = 1.$$

To find such χ_S and H_R, we first transform the cut polytope $\mathcal{P}_{cut}$ into a linearly isomorphic polytope, the *correlation polytope* $\mathcal{P}_{cor}$, so that it will be easier to find χ_S and H_R in $\mathcal{P}_{cor}$. Recall that the vertices of $\mathcal{P}_{cut}$ are the characteristic vectors of all the cuts, χ_S. Here it would be more convenient if we represent χ_S by matrices. We abuse the notation a little bit by writing the characteristic vector of cut $S, \overline{S}$ as

$$\chi_S \cdot \chi_{\overline{S}}^T,$$

where $\chi_S \in \mathbb{R}^n$ is the characteristic vector of set S. The *correlation polytope* $\mathcal{P}_{cor}$ is then the polytope we get by mapping vertices $\chi_S \cdot \chi_{\overline{S}}^T$ to $\chi_S \cdot \chi_S^T$.

Notice that $|S \cap R| = 1$ if and only if

$$\left(\sum_{i \in R} \chi_S(i) - 1 \right)^2 = 0,$$

where $\chi_S(i)$ indicates the i-th coordinate of χ_S, so by definition $\chi_S(i) = 1$ iff $i \in S$. The LHS of the above equality can be expanded and then becomes a linear function of the elements in $\chi_S \cdot \chi_S^T$. Let H_R be the coefficients of this linear function, then for vertices $\chi_S \cdot \chi_S^T$ in the correlation polytope, we find hyperplanes H_R so that

$$H_R(\chi_S \cdot \chi_S^T) \geq 0, \forall S, R$$
$$H_R(\chi_S \cdot \chi_S^T) = 0, \text{ iff } |S \cap R| = 1.$$

More details can be found in Section 5.4 of [Roughgarden et al., 2016].

References

Tim Roughgarden et al. Communication complexity (for algorithm designers). *Foundations and Trends® in Theoretical Computer Science*, 11(3–4):217–404, 2016.

Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences*, 43(3):441–466, 1991.