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

Lecture 24

Instructor: Madhu Sudan *Scribe: Benjamin Edelman*

1 Plan

The focus of the lecture is using communication complexity to derive barriers to optimization. Specifically, we will study whether "extended formulation linear programs" are a viable approach to solve NP-complete problems when a naive linear program doesn't suffice.

The plan is to:

- Introduce linear programming (LP), the Max Cut problem, and extended formulations (EFs)
- Show that lower bounds for extended formulations follow from non-deterministic communication complexity lower bounds

The main references for this lecture are Chapter 5 of Tim Roughgarden's communication complexity (for algorithm designers) survey [1], Mihalis Yannakakis's pioneering 1991 paper [2] that introduced the connection between extended formulations and communication complexity, and finally Fiorini et al.'s 2015 paper that demonstrated that Yannakakis's technique could yield unconditional lower bounds for interesting linear programs like the Traveling Salesman and Max Cut polytopes.

2 Definitions

2.1 Linear programming

The Linear Programming problem (LP) is defined as follows:

Definition 1 (LP)**.** *Given the input $A \in \mathbb{R}^{n\times n}, b \in \mathbb{R}^n, c \in R^n$, output $x \in \mathbb{R}^n$ that maximizes $c^T x$ subject to the constraints $Ax \le b$.*

In other words, the inequalities $Ax \le b$ define a convex feasible set in the space $\mathbb{R}^n$, on which the linear function $c^T x$ is maximized. If $n = 2$, the feasible set is a polygon. In higher dimensions, we call it a "polytope".

Fact 2 (proved by Khachiyan, made practical by Karmarkar's interior point method)**.** *Linear programs can be optimized in polynomial time.*

2.2 The maximum cut problem

Let K_n be the complete graph on n vertices. K_n has $\binom{n}{2}$ edges. Let $S, \overline{S}$ be a cut of the vertices of K. The characteristic vector $\chi_S \in \{0,1\}^{\binom{n}{2}} \subseteq \mathbb{R}^{\binom{n}{2}}$ of S is defined as $\chi_S(e) = \mathbb{1}[e \text{ goes from } S \to \overline{S} \text{ or } \overline{S} \to S]$.

We now define the *cut polytope* as $P^n_{\text{cut}} = \text{Convex Hull}(\{\chi_S \mid S \subseteq [n]\})$.

Exercise 3. *Compute the cut polytope for $n = 2, 3$.*

Solution. K_2 has only one edge. When we cut K_2, we can either cut that edge or not cut it, so $P^2_{\text{cut}} = \text{Convex hull}(\{(0),(1)\})$, which is just the interval [0,1].

K_3 has three edges. Any cut either cuts two edges or zero edges, so

$$P^3_{\text{cut}} = \text{Convex hull}(\{(0,0,0),(1,1,0),(1,0,1),(0,1,1)\})$$

This is a tetrahedron with vertices at those four points. □

Definition 4 (Max Cut)**.** *Given the input $w \in \mathbb{R}^{\binom{n}{2}}, w \geq 0$, output $x \in \mathbb{R}^{\binom{n}{2}}$ that maximizes $w^T x$ subject to the constraint that $x \in P^n_{cut}$.*

We can think of w as an encoding of the edges of a graph $G \subseteq K_n$. The value of $w^T x$ achieved is the maximum number of edges that can go across a cut in G.[1]

Note that because P^n_{cut} is a polytope, it can be encoded in a system of linear constraints, so a Max Cut instance is a linear program. But while LP $\in$ P, Max Cut is NP-complete. How can this be? The answer is that the cut polytope P^n_{cut} may have exponentially many (in n) constraints, so the linear program for a Max Cut instance may be exponentially large. But might there be a clever way to reduce the number of constraints? That is the idea behind "extended formulations" of polytopes, which we now introduce.

3 Extended formulations

Suppose we have a polytope $P \in \mathbb{R}^n$ that, like the cut polytope, needs $\exp(n)$ inequalities in order to be specified. The idea of extended formulations is to find a $Q \in \mathbb{R}^{n+m}$ such that P is the "shadow" of Q, in the same way that a hexagon can be thought of as the shadow of a cube when light shines on it diagonally.

Formally, suppose P is given by $Ax \leq b$. And suppose Q is given by $A'\binom{x}{y} \leq \binom{b'}{c'}$. Then P is the *shadow* of Q if $P = \{x \mid \exists y \text{ s.t. } (x, y) \in Q\}$. We say Q is an *extended formulation* of P.

Amazingly, there are polytopes that require $\exp(n)$ inequalities to specify, but are the shadows of poly(n)-dimensional polytopes with poly(n) constraints.

Exercise 5 (Non-trivial but worthwhile)**.** *Consider the polytope P defined by the inequalities $\sum_{i \in S} x_i \geq B$ $\forall S \subseteq [n]$ s.t. $|S| \geq k$, and $0 \leq x_i \leq 1$. Give a* poly(n, k)*-sized Q that is an extended formulation for P.*

Solution. First, we can note that that inequalities $\sum_{i \in S} x_i \geq B$ can all be captured by the property that the k smallest x_i's sum up to at least B. Thus, if we can introduce variables w_i for $i \leq k$, where w_i is (upper bounded by) the ith smallest element of $\{x_i\}_{i \in [n]}$, then we are done. This condition on w_i, in turn, can be captured by the property that there exists a permutation π of $[n]$ such that $w_i \leq x_{\pi(j)} \forall j \in [k, n]$. We can introduce n^2 further auxiliary variables y_{ij} to encode permutation matrices (or rather their convex hull, the stochastic matrices), add n^2 variables z_{ij} that are upper-bounded by the corresponding y_{ij} and by the corresponding x_j and that satisfy the property that $\sum_{ij} z_{ij} \leq \sum_i x_i$, in order to avoid the problem that we are dealing with stochastic matrices and not just permutation matrices. Then w_i can be $\sum_{j=1}^n z_{ij}$ and we are done. □

Extended formulations are useful because maximizing $c^T x$ subject to $x \in P$ is equivalent to maximizing $c^T x$ subject to $(x, y) \in Q$. So even if P can't be specified succinctly, as long as Q is small we can solve the LP given by (P, c). In particular, if we gave a small extended formulation for Max Cut, we would have proved Max Cut $\leq LP$, implying $P = NP$. Is there a chance that such an extended formulation exists?

Theorem 6 (Yannakakis [2], Fiorini et al. [3])**.** *No. There is no extended formulation for* Max Cut[2] *of size* poly(n).

The reason we are studying this in an information theory course is that Yannakakis proved that if we have a succinct extended formulation, this gives us a solution to a certain (rather esoteric) communication problem. Fiorini et al. proved that the communication problem corresponding to the cut polytope (well, actually a different polytope that is even easier to optimize over) is in fact hard.

[1] Because the optimization function is convex, the fact that P^n_{cut} is a convex hull of the actual set we care about doesn't affect the maximum attained.

[2] Or for the polytope corresponding to the traveling salesman problem.

4 Yannakakis's lemma

4.1 The Face-Vertex communication problem

A *face* of a polytope P is, roughly speaking, a bounding surface of P of any dimension. Consider a cube. It has eight 0-dimensional faces, which we call *vertices*; twelve 1-dimensional faces; and six 2-dimensional faces, which we call *facets*. In general, vertices are 0-dimensional faces and facets are $(n-1)$-dimensional faces. When we refer to the number of inequalities needed to specify P, we mean the number of facets of P.

Here are formal definitions of faces and facets. A *supporting hyperplane* h of P is a hyperplane $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a} \cdot \mathbf{x} = b\}$ such that all of P lies on one side of h (for all $\mathbf{x} \in P, \mathbf{a} \cdot \mathbf{x} \leq b$ and h has nonempty intersection with P. Intuitively, h is "tangent" to P. A face of P is the intersection of P with some supporting hyperplane h. A facet is a maximal face: a face that is not strictly contained in another face. In the non-degenerate setting, facets are equivalently the $(n-1)$-dimensional faces, as we said earlier.

We can rephrase our guiding question now as: given P with $\exp(n)$ vertices and facets, is it the shadow of a higher-dimensional polytope with $\text{poly}(n)$ facets?

Definition 7 (The Face-Vertex(P) problem)**.** *Alice is given a vertex $v \in P$. Bob is given a face $f \in P$ in the form of the hyperplane f lies on. The face-vertex function $FV(f, v)$ is defined as $\mathbb{1}[v \notin f]$. Alice and Bob need to compute $FV(f, v)$.*[3]

It is natural to ask about the deterministic communication complexity of the Face-Vertex(P) problem (we'll denote this $\text{CC}(FV)$). This is what we are familiar with in this class. However, we can also ask about the *non-deterministic* communication complexity of the problem. Here is how non-deterministic communication works: Alice and Bob don't communicate with each other directly. Instead, they each receive identical messages m (which we can think of as an advice string or proof) from a third character, Merlin, who is also on their team and knows both v and f. After receiving m, Alice and Bob each output a bit, and they win if the AND of their bits is the correct answer.

Formally, $\text{NCC}(FV) \leq k$ if there is a protocol for Alice and Bob such that whenever $v \notin f$, there exists a message m, $|m| \leq k$, such that Alice and Bob both output 1 when given m, and whenever $v \in F$ then for all messages $|m| \leq k$ either Alice or Bob outputs 0.

Nondeterministic communication complexity is a much-studied notion, and there is a remarkable result that connects it to deterministic complexity:

Fact 8 (Aho et al. [4])**.** *Let $F : X \times Y \to \{0, 1\}$ be any function, and let $\overline{F}$ be $1 - F$. Then*

$$CC(F) \leq NCC(F) \cdot NCC(\overline{F})$$

In other words, if a function has high deterministic complexity, than either it or its complement must have high non-deterministic complexity. For many classic hard communication problems like Disjointness, it is easy to come up with an efficient nondeterministic protocol for the problem (or its complement), so consequently the complement (or the original problem) must have high non-deterministic communication complexity.

Now we're ready to see Yannakakis's key lemma.

4.2 The lemma

Lemma 9 (Yannakakis [2])**.** *If the polytope P has an extended formulation Q with r facets, then*

$$NCC(\text{FaceVertex}(P)) \leq \log r$$

Thus, if we can prove that Face-Vertex(P) has $\Omega(\text{poly}(n))$ non-deterministic communication complexity, it will immediately follow that P does not have any extended formulation of size $o(\exp(n))$.

[3]Note that both Alice and Bob know P to begin with.

Proof sketch. We are given P with an extended formulation Q that has r facets, and we want to describe a protocol such that Alice and Bob can solve the Face-Vertex(P) problem after receiving a message from Merlin of length $\leq \log r$.

Let v^* be the "lifting" of Alice's vertex v from P to Q, and let f^* be the lifting of Bob's face f:

$$v^* = \{(v, y) \mid (v, y) \in Q\}$$

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

The basic idea of the protocol is for Merlin's message m to be the index of a facet $\tilde{f}$ of Q. Remember: not all faces are facets. Merlin should, if possible, choose $\tilde{f}$ such that $f^* \subseteq \tilde{f}$ but $v^* \nsubseteq \tilde{f}$. It turns out that Merlin can only find such an $\tilde{f}$ if $v \notin f$. Thus, Alice's output should be $\mathbb{1}[v^* \nsubseteq \tilde{f}]$ and Bob's output should be $\mathbb{1}[f^* \subseteq \tilde{f}]$, and the protocol is complete. □

Exercise 10. *Prove that there exists a facet $\tilde{f} \subseteq Q$ satisfying $f^* \subseteq \tilde{f}$ and $v^* \nsubseteq \tilde{f}$ if and only if $v \notin f$. (You may use the fact that a face is the intersection of the facets that contain it.)*

Solution. If $f^* \subseteq \tilde{f}$ and $v^* \nsubseteq \tilde{f}$, then $v^* \in Q - f^*$, so there must be some $(v, y) \in v^*$ such that $(v, y) \in Q - f^*$. Suppose now that $v \in f$; but then $(v, y) \in f^*$ because $(v, y) \in Q$, so we have a contradiction, so $v \notin f$.

For the other direction, suppose $v \notin f$. Then $v^* \nsubseteq f^*$ (in fact, $v^* \cap f^* = \emptyset$). Because f^* is the intersection of the facets that contain it, there must be one such facet $\tilde{f}$ that doesn't contain v^* (otherwise we would have $v^* \subseteq f^*$). □

[Aside: Suppose we wanted to modify the above argument to show that if P has an extended formulation Q with r facets then $NCC(\overline{\text{FaceVertex}(P)}) \leq \log r$. Merlin, perhaps, could choose $\tilde{f}$ such that $f^* \subseteq \tilde{f}$ and $v^* \subseteq \tilde{f}$. But this argument doesn't work, because it doesn't follow that $v^* \subseteq f^*$.]

5 Applying the lemma

It took a few decades until Fiorini et al. proved that Yannakakis's lemma could be applied to prove lower bounds on the sizes of extended formulations for interesting polytopes like P^n_{cut}.

Fiorini et al. didn't actually work directly with the cut polytope; instead, they focused on the *correlation polytope*, defined as follows:

Definition 11 (correlation polytope)**.** $P^n_{cor} = \textit{Convex Hull}(\{xx^T \in \mathbb{R}^n \mid x \in \{0,1\}^n\})$

Note that P^n_{cor} is in the space $\mathbb{R}^{n \times n}$ of $n \times n$ matrices. It is the convex hull of the set of all rank-1 binary symmetric matrices. We can relate the cut polytope and the correlation polytope with a neat theorem of De Simone:

Fact 12 (De Simone, 1990 [5])**.** *P^{n+1}_{cut} and P^n_{cor} are linearly isomorphic.*

In other words, there is an invertible linear map between P^n_{cut} and P^n_{cor}. It is not too difficult to verify that extension complexity is preserved under linear isomorphism, so a bound on the extension complexity of the correlation polytope will imply a bound for the cut polytope.

The core of Fiorini et al.'s argument is the following lemma:

Lemma 13. *For all subsets $S \in [n]$, P^n_{cor} has a face f_S such that for all subsets $R \in [n]$, letting $x_R \in [0,1]^n$ be the characteristic vector of R and $v_R = x_R {x_R}^T$, then $v_R \in f_S$ if and only if $|S \cap R| = 1$.*

In other words, for every subset S the face f_S in some sense encodes something that looks like the Face-Vertex problem into something that looks like Unique Disjointness, which is the variation of Disjointness where we are promised that Alice's set and Bob's set have an intersection of size ≤ 1. This reduction can be made formal, to show that in fact NCC(FaceVertex(P)) $\geq$ NCC(Unique Disjointness).

The final step of the proof is to show that NCC(Unique Disjointness) $= \Omega(n)$.[4]

[4] The "yes" case of disjointness (and unique disjointness) is that the sets are disjoint. The non-deterministic communication complexity of the opposite problem is only $\log n$, since Merlin would only need to specify an index on which the sets overlap.

References

[1] Roughgarden, T. (2016). Communication complexity (for algorithm designers). Foundations and Trends in Theoretical Computer Science, 11(3–4), 217-404.

[2] Yannakakis, M. (1991). Expressing combinatorial optimization problems by linear programs. Journal of Computer and System Sciences, 43(3), 441-466.

[3] Fiorini, S., Massar, S., Pokutta, S., Tiwary, H. R., & Wolf, R. D. (2015). Exponential lower bounds for polytopes in combinatorial optimization. Journal of the ACM (JACM), 62(2), 17.

[4] Aho, A. V., Ullman, J. D., & Yannakakis, M. (1983, December). On notions of information transfer in VLSI circuits. In Proceedings of the fifteenth annual ACM symposium on Theory of computing (pp. 133-139). ACM.

[5] De Simone, C. (1990). The cut polytope and the Boolean quadric polytope. Discrete Mathematics, 79(1), 71-75.