

6. Upper bounds and approximation algorithms

Lecturer: Massimo Lauria

We show how to give a good approximation for the max-cut problem using only low levels of the Lasserre hierarchy. This implies a separation between Sherali-Adams and Lasserre. Additionally, we show how to obtain locally consistent solutions from intermediate polytopes and we introduce basic Fourier Analysis definitions and properties.



<http://www.csc.kth.se/~lauria/sos14/>

Example of local consistency

We start by showing how the Lasserre proof system can be made locally consistent. This is, we can find solutions that are integer over some suitable subset of the variables without needing to go up the whole hierarchy.

In the previous lecture we saw that if we have a point $y \in L_t(K)$ and a set of variables $|S| \leq t$ then we can express y as a convex combination of points whose S -coordinates are integer and are expressible in a lower level of the hierarchy. Formally, there exists a probability distribution $\mathcal{D}(S)$ such that $\Pr_{z \sim \mathcal{D}}[\bigwedge_{i \in I} z_i = 1] = y_I$ (convex combination), $z \cap S \in \{0, 1\}^{|S|}$ (integrality) and $z \in L_{t-|S|}$ (lower level).

Our example is the 3-colouring problem: given a graph $G(V, E)$ and a set $\{R, G, B\}$, we want to colour the vertices such that no adjacent vertices share a colour. We can model the problem as a linear program in the following way. We have variables x_{ic} meaning that the vertex i is coloured c , and we impose the restrictions $x_{iR} + x_{iG} + x_{iB} \geq 1 \quad \forall i \in V$ to ensure that every vertex has a colour and $x_{ic} + x_{jc} \leq 1 \quad \forall (i, j) \in E$ to ensure that adjacent vertices do not share a colour.

Assume $y \in L_{3t}(K)$ is a point in the $3t$ -th level of the Lasserre hierarchy and let $U \subseteq V$, $|U| \leq t$ be a subset of vertices. Then we can extract a distribution of points in $L_{3(t-|U|)}(K)$ that have integer values over $U \times \{R, G, B\}$, even though the solutions may be globally invalid.

This means that if we only want to satisfy local constraints, we do not need the full power of Lasserre but we can settle for going up to as many levels as variables we wish to satisfy.

This particular example would also work with a weaker proof system such as Sherali-Adams, what Lasserre buys you is the ability to do global reasoning since the SDP constraint has a global structure.

Upper bound for Lasserre

We now describe an algorithm for max-cut formalizable in low levels of Lasserre but not of Sherali-Adams. We follow the presentation of Rothvoß¹.

For the max-cut problem we are given a graph $G(V, E)$ and we want to find a subset $S \subseteq V$ of vertices such that its edge-neighbourhood $|E_G(S, \bar{S})|$ is maximum.

We formalize the problem with a decision variable x_i for every vertex, meaning whether i is in S , and a variable z_e for every edge, meaning whether

¹ Thomas Rothvoß. The lasserre hierarchy in approximation algorithms. Lecture Notes for the MAPSP 2013 Tutorial (preliminary version), 2013

e is in the cut. The integer problem is then

$$\begin{aligned} & \text{maximize} && \sum_{u \in V} z_{ij} \\ & \text{subject to} && x_i \oplus x_j = z_{ij} \quad \forall (i, j) \in E. \end{aligned} \quad (1)$$

The natural linear relaxation is

$$\begin{aligned} & \text{maximize} && \sum_{u \in V} z_{ij} \\ & \text{subject to} && x_i - x_j \leq z_{ij}, \\ & && x_j - x_i \leq z_{ij}, \\ & && z_{ij} \leq x_i + x_j, \\ & && z_{ij} \leq 2 - x_i - x_j \quad \forall (i, j) \in E. \end{aligned} \quad (2)$$

Observe that $x_i = 1/2$, $z_{ij} = 1$ is a valid fractional solution, but it can be very far from the feasible optimum. For instance, when G is the complete graph the fractional optimum is approximately $n^2/2$ while the real solution is approximately $n^2/4$. The integrality gap is roughly $1/2$.

At the $n^{\delta(\epsilon)}$ -th level of the Sherali-Adams hierarchy, the integrality gap is still $1/2 + \epsilon$, so this is the best we can do². However, there is an algorithm³ that gives an approximation ratio of at least 0.878, and we will see that this algorithm can be formulated in terms of a Lasserre program.

Let us first do some observations about the space of solutions. We had seen in the previous lecture that the moment matrix is positive semidefinite and so it can be expressed as an inner product of vectors, this is $M^t(Y)_{IJ} = \langle v_i, v_j \rangle$. It follows that $y_I = \langle v_I, v_\emptyset \rangle = \langle v_I, v_I \rangle = |v_I|^2$. For any vector v_I it holds that $|v_\emptyset/2 - v_I|^2 = |v_\emptyset/4|^2 - \langle v_\emptyset, v_I \rangle + |v_I|^2 = 1/4$, so the space of solutions is a sphere of radius $1/2$ centered at $v_\emptyset/2$. We can try to exploit this fact by separating vectors that lie far away in this sphere into the two components S, \bar{S} .

To simplify the process of sampling separators, we will use another formulation where the space of solutions is a unit sphere.

$$\begin{aligned} & \text{maximize} && \frac{1 - \langle u_i, u_j \rangle}{2} \\ & \text{subject to} && |u_i|^2 = 1 \quad \forall i \in V. \end{aligned} \quad (3)$$

So first of all we need to show that our original relaxation (2) implies the transformed formulation (3) and in fact we will show that this is the case using the 5th level of the Lasserre hierarchy.

Let K be the polytope defined by the original relaxation (2) and let $y \in L_5(K)$ be a solution in the 5th Lasserre level. We can restrict the three variables z_{ij} , x_i , and x_j to be integers and we obtain a distribution over solutions $\hat{y} \in L_2(K)$ whose components $\hat{y}_{z_{ij}}$, \hat{y}_{x_i} , and \hat{y}_{x_j} are integers. For these particular components, the constraints in (2) plus the integrality restriction imply the stronger relation $\hat{y}_{z_{ij}} = |\hat{y}_{x_i} - \hat{y}_{x_j}|$. Since \hat{y} is on the 2nd level of the Lasserre hierarchy, we can rewrite $|\hat{y}_{x_i} - \hat{y}_{x_j}| = \hat{y}_{x_i} + \hat{y}_{x_j} - 2\hat{y}_{\{x_i, x_j\}}$.

² Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality gaps for sherali-adams relaxations. In *Proceedings of the Annual ACM Symposium on Theory of Computing*, pages 283–292. ACM, 2009.
³ Michael Grigni and Avrim Blum. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.

This relation also holds for any (even fractional) solution in the 5th level of the Lasserre hierarchy because we can see it as a convex combination of integer solutions in the 2nd level. We repeat the argument for every choice of i and j .

So if v is a solution in $L_5(K)$, then u defined by $u_i = v_\emptyset - 2v_i$ is a solution of (3). Indeed, $|u_i|^2 = |v_\emptyset - 2v_i|^2 = |v_\emptyset|^2 - 4\langle v_\emptyset, v_i \rangle + 4|v_i|^2 = 1$, which is the only constraint of (3).

Furthermore, $\langle u_i, u_j \rangle = |v_\emptyset|^2 - 2\langle v_i, v_\emptyset \rangle - 2\langle v_\emptyset, v_j \rangle + 4\langle v_i, v_j \rangle = 1 - 2(x_i + x_j - 2z_{ij}) = z_{ij}$, from where $z_{ij} = (1 - \langle u_i, u_j \rangle)/2$. We conclude that the functions we are optimizing have the same value.

Observe that we are crucially using positive semidefiniteness to rewrite the problem with inner products. In fact the 3rd level of Lasserre is enough⁴, but we used the 5th level for the sake of simplicity.

We now follow the Goemans-Williamson strategy to sample solutions to (3). We use the intuition that vectors with close angles are in the same part of the graph, while vectors with distant angles are in different parts. We cut the sphere with a hyperplane and let the partition be each of the halfspaces. This is, if we choose a hyperplane by sampling its normal vector h , then $i \in S$ if and only if $\langle u_i, h \rangle > 0$.

We want to sample h uniformly over all directions, which we can do by sampling each component independently according to a normal distribution $N(0, 1)$. Indeed, the probability of sampling a specific coordinate x_i is $\Pr[h_i = x_i] = \frac{1}{\sqrt{2\pi}}e^{-x_i^2/2}$. Then the probability of sampling a specific vector x is $\Pr[h = x] = \frac{1}{\sqrt{2\pi}}e^{-\sum_i x_i^2/2} = \frac{1}{\sqrt{2\pi}}e^{-|x|^2/2}$, which depends on the length of x but not on its direction.

We show a lower for the approximation ratio ROUND/OPT and the integrality gap OPT/FAC by computing

$$\frac{\text{ROUND}}{\text{FAC}} = \frac{\text{ROUND}}{\text{OPT}} \cdot \frac{\text{OPT}}{\text{FAC}} \quad (4)$$

and noting that each factor is at most 1 because we are solving a maximization problem.

The solution we obtain by the GW approximation technique is $\text{ROUND} = \mathbb{E}|E(S, \bar{S})| = \sum_{(i,j) \in E} \Pr[(i,j) \in E(S, \bar{S})] = \sum_{(i,j) \in E} \theta_{ij}/\pi$, where θ_{ij} is the angle between u_i and u_j . Then $\theta_{ij} = \arccos(\langle u_i, u_j \rangle) = \arccos(1 - 2z_{ij})$.

On the other hand, the solution we obtain by the SDP technique is $\text{FAC} = \sum_{(i,j) \in E} z_{ij}$. We obtain a lower bound for the approximation ratio and the integrality gap of

$$\frac{\text{ROUND}}{\text{FAC}} = \frac{\sum_{(i,j) \in E} \arccos(1 - 2z_{ij})/\pi}{\sum_{(i,j) \in E} z_{ij}} \geq \inf_{0 < x < 1} \frac{\arccos(1 - 2x)}{\pi x} \geq 0.878. \quad (5)$$

A reminder of Fourier Analysis

And now for something completely different we introduce definitions and notation for Fourier analysis of Boolean functions so we will have them fresh for the following lectures.

⁴ Thomas Rothvoß. The lasserre hierarchy in approximation algorithms. Lecture Notes for the MAPSP 2013 Tutorial (preliminary version), 2013

Given a Boolean function $f : \{0, 1\} \rightarrow \mathbb{R}$, we want to express it as a linear combination of simple functions, this is $f = \sum_{S \subseteq [n]} \alpha_S \chi_S$. The character function of a set χ_S counts the parity of $x \wedge S$, this is $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$.

If we use the so-called Fourier variables $y_i = 1 - 2x_i = (-1)^{x_i}$, then we can also express the characters as square-free monomials $\chi_S(x) = \prod_{i \in S} y_i$.

The standard properties of Fourier basis hold, as it is the case with real functions, but with simpler proofs since we only need to deal with finite domains.

Lemma 1. $\mathbb{E} \chi_S = [S = \emptyset]$.

Proof. If $S = \emptyset$ we are done. Otherwise pick any component $i \in S$.

$$2^n \mathbb{E} \chi_S = \sum_{x \in \{0,1\}^n} \chi_S(x) = \sum_{x: x_i=0} \chi_S(x) + \sum_{x: x_i=1} \chi_S(x) \quad (6)$$

$$= \sum_{x: x_i=0} \chi_S(x) - \sum_{x: x_i=0} \chi_S(x) = 0 \quad (7)$$

□

Lemma 2. $\{\chi_S\}_S$ is an orthonormal basis.

Proof. Define the inner product by $\langle \chi_S, \chi_T \rangle = \mathbb{E} \chi_{S \Delta T}$ ⁵. □

⁵ The symmetric union of two sets is $S \Delta T = S \cup T \setminus (S \cap T)$, the set of elements in exactly one of the sets

We usually write the coefficients α_S of f in the Fourier basis as $\hat{f}(S)$, and it holds that $\hat{f}(S) = \langle f, \chi_S \rangle$.

If we denote the vector of coefficients of f by \hat{f} , then the Plancherel identity

$$\langle f, g \rangle_{\{0,1\} \rightarrow \mathbb{R}} = \langle \hat{f}, \hat{g} \rangle_{\mathbb{R}^{2^n}} \quad (8)$$

and the Parseval identity

$$\|f\|_{\{0,1\} \rightarrow \mathbb{R}}^2 = \|\hat{f}\|_{\mathbb{R}^{2^n}}^2 \quad (9)$$

hold. Note that two inner products are involved, the product as functions and the product of the coefficients as vectors.

References

- [CMM09] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality gaps for sherali-adams relaxations. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 283–292. ACM, 2009.
- [GW95] Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.
- [Rot13] Thomas Rothvoß. The lasserre hierarchy in approximation algorithms. Lecture Notes for the MAPSP 2013 Tutorial (preliminary version), 2013.