6. Upper bounds and approximation algorithms

Lecturer: Massimo Lauria

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

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 hierarchies. Additionally we introduce basic Fourier analysis in the boolean setting.

Example of local consistency

We start by showing how the solution to a Lasserre relaxation can be made locally consistent, i.e., made into a solutions that is 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 subset $S \subseteq [n]$ of size $|S| \leq t$, we can express y as a convex combination of points which S-coordinates are integer and expressible in a lower level of the hierarchy. Formally, there exists a probability distribution $\mathcal{D}(S)$ over $\{0,1\}^S$ such that $\Pr_{z \sim \mathcal{D}(S)}[\bigwedge_{i \in I} z_i = 1] = y_I$ for all $I \subseteq S$ (convex combination), $z_i \in \{0,1\}$ for $i \in S$ (integrality) and $z \in L_{t-|S|}(K)$ (lower level).

Our example is the 3-coloring problem: Given a graph $G(V,E)$ and a set $\{R,G,B\}$, we want to color the vertices such that no adjacent vertices share a color. We can relax the problem to a linear program. We have variables $x_{ic} \in [0,1]^{3|V|}$, meaning that vertex i is colored c if $x_{ic} = 1$, 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 color, and

$$x_{ic} + x_{jc} \leq 1 \quad \forall (i,j) \in E,$$

to ensure that adjacent vertices do not share a color.

Assume $y \in L_{3t}(K)$ 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.

This particular example would also work with a weaker proof system, such as Sherali-Adams. The benefit with Lasserre solutions is the ability to do global reasoning, since the constraint of positive semidefiniteness has a global structure.

Upper bound for Lasserre

We now describe an algorithm for the max-cut problem. The algorithm is formalizable in low levels of Lasserre but not in low levels of Sherali-Adams. We follow the presentation of Rothvoß[1].

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

In 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 the number of edges in the induced cut, $|E_G(S,\bar{S})|$, is maximized. Here, $\bar{S} = V \setminus S$.

We formalize the problem with a decision variable x_i for every vertex, where $x_i = 1$ if $i \in S$ and $x_i = 0$ otherwise, and a decision variable z_{ij}

for every edge, where $z_{ij} = 1$ if the edge belongs to the cut and $z_{ij} = 0$ otherwise. The integer program is then

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

The natural linear relaxation of Problem (1) is

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

If we set all x_i to $1/2$ and all z_{ij} to 1 we get a valid fractional solution of Problem (2). This solution can be very far from the integral one. For instance, when G is a complete graph, the fractional optimal value is approximately $n^2/2$ while the integral optimal value is approximately $n^2/4$. Thus, the integrality gap of this relaxation is roughly $1/2$.

We know from Charikar et al.[2] that even 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 do with linear relaxations.

However, there is an algorithm that gives an integrality gap of at least 0.878, and we will see that this algorithm can be formulated in terms of a Lasserre relaxation of low degree.[3]

Let us first do some observations about the space of solutions. We have seen in the previous lecture that when a moment matrix is positive semidefinite it can be expressed as an inner product of vectors $M^t(Y)_{IJ} = \langle v_I, v_J \rangle$. It follows that $y_I = \langle v_I, v_\emptyset \rangle = \langle v_I, v_I \rangle = |v_I|^2$ and $|v_\emptyset| = 1$. For any vector v_I, we have

$$|v_\emptyset/2 - v_I|^2 = |v_\emptyset/2|^2 - \langle v_\emptyset, v_I \rangle + |v_I|^2 = 1/4,$$

that is, the space of solutions is a sphere of radius $1/2$ centered at $v_\emptyset/2$. We can exploit this fact to build a cut $(S, \bar{S})b$ by separating the vectors that lie in this sphere. The separation will be obtained by sampling an hyperplane that splits the sphere in two halfspaces. The induced partition of the vectors will corresponds to the cut.

To simplify the process of sampling separators, we perform a vector transformation that maps the space of solutions to a unit sphere centered around the origin. The vector transformation is

$$u_i = v_\emptyset - 2v_i,$$

[2] 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

[3] 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

and the corresponding max-cut formulation becomes

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

We need to show that our original relaxation (2) implies the transformed formulation (3). We do that using the 5th level of the Lasserre hierarchy.

Let K be the polytope defined by the original relaxation (2) and let $y \in L_5(K)$. We can restrict the three variables z_{ij}, x_i, and x_j to be integers and obtain a distribution over solutions $\hat{y} \in L_2(K)$ which components $\hat{y}_{z_{ij}}$, $\hat{y}_{x_i}$, and $\hat{y}_{x_j}$ are integers. For these particular components, the constraints in Problem (2) and integrality 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 have

$$|\hat{y}_{x_i} - \hat{y}_{x_j}| = \hat{y}_{x_i} + \hat{y}_{x_j} - 2\hat{y}_{\{x_i,x_j\}}.$$

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.

If v is a solution in $L_5(K)$, then u defined by $u_i = v_\emptyset - 2v_i$ is a solution of Problem (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 Problem (3). Furthermore,

$$\begin{aligned} \langle u_i, u_j \rangle =& |v_\emptyset|^2 - 2\langle v_i, v_\emptyset \rangle - 2\langle v_\emptyset, v_j \rangle + 4\langle v_i, v_j \rangle \\ =& 1 - 2(x_i + x_j - 2x_{ij}) = 1 - 2z_{ij}, \end{aligned}$$

thus $z_{ij} = (1 - \langle u_i, u_j \rangle)/2$. We conclude that the objective functions which we are optimizing in Problems (2) and (3) have the same value.

Note that the 3rd level of Lasserre is enough[4] to perform the described relaxation, but we used the 5th level for the sake of simplicity.

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

We now follow the Goemans-Williamson strategy to sample solutions to Problem (3). We use the intuition that vectors with a small angle between should corresponds to vertices in the same side of the cut, while vectors with a large angle between them should corresponds to verticed in different sides. We cut the sphere with a hyperplane and let the partition be each of the half-spaces. That 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 coordinate of h independently according to a normal distribution $N(0, 1)$. The normal probability density function of a coordinate is

$$d(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2},$$

and the corresponding density function of a vector is

$$d((v_1, \ldots, v_m)) = \frac{1}{(\sqrt{2\pi})^m} e^{-\sum_i v_i^2/2} = \frac{1}{(\sqrt{2\pi})^m} e^{-|v|^2/2},$$

which depends on the length of $(v_1, \ldots, v_m)$ but not on its direction.

We derive a lower bound for the approximation ratio $\mathsf{ROUND}/\mathsf{OPT}$ and the integrality gap $\mathsf{OPT}/\mathsf{FRAC}$, where OPT is the optimal value of Problem (1), FRAC is the optimal value of Problem (3), and ROUND is the expected number of edges in the cut when using the Goemans-Williamson strategy to round the solution of Problem (3). We have that

$$\mathsf{ROUND} \leq \mathsf{OPT} \leq \mathsf{FRAC},$$

and therefore

$$\frac{\mathsf{ROUND}}{\mathsf{FRAC}} \leq \frac{\mathsf{OPT}}{\mathsf{FRAC}} \text{ and } \frac{\mathsf{ROUND}}{\mathsf{FRAC}} \leq \frac{\mathsf{ROUND}}{\mathsf{OPT}}.$$

Note that each factor is at most 1 since we are solving a maximization problem. The solution obtained with the Goemans-Williamson strategy is

$$\mathsf{ROUND} = \mathbb{E}|E(S,\bar{S})| = \sum_{(i,j)\in E} \Pr[(i,j) \in E(S,\bar{S})] = \sum_{(i,j)\in E} \theta_{ij}/\pi,$$

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 (Problem (3)) is

$$\mathsf{FRAC} = \sum_{(i,j)\in E} z_{ij}.$$

Therefore, the lower bound of the approximation ratio and the integrality gap becomes

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

A quick summary about Fourier analysis for the hypercube

And now for something completely different, we introduce definitions and notation for Fourier analysis of Boolean functions. We only give the basic definitions and properties, despite the fact that the theory is so rich and with so many applications throughout theoretical computer science would deserve more space. I suggest you to look at the book by Ryan O'Donnel.[5]

[5] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014

Given a boolean function $f : \{0,1\}^n \longrightarrow \mathbb{R}$, we want to express it as a linear combination of simple functions that represent parity functions. The character function of a set S, χ_S, counts the parity of the numbers of ones in the bit vector $x \wedge S$, that is $\chi_S(x) = (-1)^{\sum_{i\in S} x_i}$. We aim to represent f as $\sum_{S\subseteq[n]} \alpha_S \chi_S$, namely to find the coefficients α_S.

We use the so-called Fourier variables $y_i = 1 - 2x_i = (-1)^{x_i}$. This (affine) change of variable is invertible and allows to express the characters as square-free monomials $\chi_S(x) = \prod_{i\in S} y_i$. The consequence is that the coefficients α_S are just the coefficients of the representation of f as a multilinear polynomial in variables y_i.

We list some useful properties of the Fourier basis on the hypercube.

Lemma 1.

$$\sum_{x\in\{0,1\}^n} \frac{\chi_S(x)}{2^n} = \begin{cases} 1 & \textit{if } S = \emptyset \\ 0 & \textit{otherwise} \end{cases} \tag{4}$$

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

$$\begin{aligned} \sum_{x\in\{0,1\}^n} \chi_S(x) &= \sum_{x:x_i=0} \chi_S(x) + \sum_{x:x_i=1} \chi_S(x) \\ &= \sum_{x:x_i=0} \chi_S(x) - \sum_{x:x_i=0} \chi_S(x) = 0 \end{aligned}$$

□

Lemma 2. $\{\chi_S\}_S$ *is an orthonormal basis for the vector space of boolean real functions on the hypercube, with respect to the inner product*

$$\langle f, g\rangle = \frac{1}{2^n} \sum_{x\in\{0,1\}^n} f(x)g(x) \; .$$

Proof. The vector space has size 2^n and the character functions are 2^n. To prove that it is a basis we will prove orthonormality (and therefore linear independence). Observe that $\chi_S(x)\chi_T(x) = \chi_{S\triangle T}(x)$, where the symmetric union of two sets is denoted $S\triangle T = S \cup T \setminus (S \cap T)$. The inner product

$$\langle \chi_S, \chi_T\rangle = \frac{1}{2^n} \sum_{y\in\{-1,1\}^n} \chi_{S\triangle T}$$

is therefore 1 when $S = T$ and is zero otherwise. □

We denote the coefficients α_S of f in the Fourier basis as $\hat{f}(S)$, and because of orthonormality 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

$$\langle f, g\rangle_{\{0,1\}\to\mathbb{R}} = \langle \hat{f}, \hat{g}\rangle_{\mathbb{R}^{2^n}} \tag{5}$$

and hence the Parseval identity

$$\|f\|^2_{\{0,1\}\to\mathbb{R}} = \|\hat{f}\|^2_{\mathbb{R}^{2^n}} \tag{6}$$

hold. Note that there are two inner products are involved, one about boolean functions and one about the vectors of coefficients.

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.

[O'D14] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[Rot13] Thomas Rothvoß. The lasserre hierarchy in approximation algorithms. Lecture Notes for the MAPSP 2013 Tutorial (preliminary version), 2013.