CS369H: Hierarchies of Integer Programming Relaxations Spring 2016-2017

Lecture 2: The Lasserre Hierarchy

Professor Moses Charikar *Scribes: Andy Tsao*

Overview We introduce the Lasserre (equivalently Sum of Squares) Hierarchy. A good reference is Rothvoß's lecture notes [Rot13].

1 The Lasserre Hierarchy

1.1 Preliminaries and Notation

First, we recall the properties of a positive-semidefinite matrix.

Definition 2.1 (PSD). *A matrix $M \in \mathbb{R}^{n\times n}$ is positive-semidefinite if any of the following conditions hold:*

(a) For any vector $x \in \mathbb{R}^n$, $x^T M x \geq 0$.

(b) There exist vectors $v_1, \ldots, v_n \in \mathbb{R}^m$ such that $M_{ij} = \langle v_i, v_j\rangle$. Equivalently, $M = V^T V$, where V is the matrix with columns $v_1, \ldots, v_n$.

(c) Every principal submatrix U of M has nonnegative determinant.

For each n, we now define a vector $y \in \mathbb{R}^{2^n}$, indexed by subsets of $\{1, 2, \ldots, n\}$. The heuristic to keep in mind is that

$$y_I \approx \mathbb{P}(x_i = 1, \forall i \in I) = \mathbb{E}\left(\prod_{i\in I} x_i\right).$$

This vector y gives rise to the *moment matrix* $M_t(y) = (y_{I\cup J})_{|I|,|J|\leq t}$. Also, for a given constraint $\sum_{i=1}^{n} A_{li}x_i - b_l \geq 0$, we define the matrix $M_t^l(y) = \left(\sum_{i=1}^{n} A_{li}y_{I\cup J\cup\{i\}} - b_l y_{I\cup J}\right)$. These are known as the *moment matrices of slacks*.

Exercise 2.1. *Show that the constraint $M_t(y) \succeq 0$ is satisfied by integral solutions $y \in \{0,1\}^{2^n}$.*

1.2 Definition of Lasserre

Suppose we are given a polytope $K = \{x \in \mathbb{R}^n : Ax \geq b\}$ which is a linear relaxation of a binary optimization problem. Ideally, we would like to find a solution in conv$(K \cap \{0,1\}^n)$. This is equivalent to finding a probability distribution over the integer solutions. The moments of this probability distribution are given by the moment matrix as defined above. This motivates the following definition of the Lasserre Hierarchy:

Definition 2.2. *For a given relaxation K and integer t, we define the t^{th} level of the Lasserre Hierarchy $Las_t(K)$ as the set of vectors $y \in \mathbb{R}^{2^n}$ such that $y_\emptyset = 1$ and $M_t(y), M_t^l(y)$ are all positive-semidefinite. Furthermore, let $Las_t^{proj}(K)$ be the projection of $Las_t(K)$ onto the original variables.*

Remark: The PSD condition fits with the previously mentioned heuristic in the following manner: If we restrict to $x \in \{0,1\}^n$, then $M_t(y) = \hat{y}\hat{y}^T$, where $\hat{y}_I = \prod_{i \in I} x_i$. In this case, the moment matrix is clearly PSD. In the case where x is instead a distribution over integer-valued solutions, $\hat{y}$ would then be a matrix with columns equal to scaled integer solutions.

Recall now that in the Sherali-Adams Hierarchy, we symbolically multiplied the constraint $Ax - b \geq 0$ by

$$\left(\prod_{i \in I} x_i\right)\left(\prod_{i \in J}(1 - x_i)\right),$$

where I and J are subsets of $[n]$ of bounded size. In the Lasserre Hierarchy, we similarly multiply the above product by the moment matrix and constrain the product to be positive-semidefinite.

1.3 Properties

The following section will be devoted to proving several properties of the Lasserre Hierarchy.

Lemma 2.1. *Let $K = \{x \in \mathbb{R}^n : Ax \geq b\}$ and $y \in Las_t(K)$. Then,*

(a) $K \cap \{0,1\}^n \subseteq Las_t^{proj}(K)$.

(b) $0 \leq y_I \leq 1$ *for all* I *such that* $|I| \leq t$.

(c) $0 \leq y_J \leq y_I \leq 1$ *for all* $I \subseteq J \subseteq [n]$.

(d) $|y_{I \cup J}| \leq \sqrt{y_I y_J}$.

(e) $Las_t^{proj}(K) \subseteq K$.

(f) $Las_0(K) \supseteq Las_1(K) \supseteq \ldots \supseteq Las_n(K)$.

Proof. (a) Let $x \in K \cap \{0,1\}^n$ be a feasible integral solution, and let $y_I = \prod_{i \in I} x_i$. Then, the moment matrix $M_t(y)$ is a submatrix of the PSD matrix yy^T, and is therefore also PSD. Similarly, the moment matrix of slacks is submatrix of $(Ax - b)yy^T$, which is the product of a nonnegative quantity and a PSD matrix. Thus, it follows that $y \in \text{Las}_t(K)$, and the claim follows.

(b) Let U be the 2×2 submatrix indexed by $\{\emptyset, I\}$. Then, the determinant of U is $y_I(1 - y_I)$. Since the moment matrix is PSD, this determinant must be nonnegative, yielding the desired result.

(c) The proof is the same as in part (b), except we index U by $\{I, J\}$.

(d) The proof is the same as in part (b), except we index U by $\{I, J\}$.

(e) The $(\emptyset, \emptyset)$ entry of $M_t^l(y)$ is

$$\sum_{i=1}^{n} A_{li} y_{\{i\}} - b_l,$$

which is nonnegative since $M_t^l(y) \succeq 0$.

(f) Observe that $M_t(y)$ is a submatrix of $M_{t+1}(y)$. So $M_{t+1}(y) \succeq 0$ automatically implies that $M_t(y) \succeq 0$. The same thing also holds for $M_t^l(y)$.

$\square$

1.4 A convex combination of integer solutions

As each solution y ideally represents a probability distribution over integer solutions, it is desirable to write y as a convex combination of integral vectors in K. This is possible on $K \cap \{0,1\}^n$, but for a solution $y \in \text{Las}_t(K)$, we can make the weaker claim that y can be written as a convex combination of solutions that are integral on a particular subset $S \subseteq [n]$.

Lemma 2.2. *For $y \in \text{Las}_t(K)$, there exist $z^{(0)}, z^{(1)} \in \text{Las}_{t-1}(K)$ such that y is a convex combination of $z^{(0)}, z^{(1)}$. More precisely, let*

$$\begin{aligned} z_I^{(1)} &= \frac{y_{I \cup \{i\}}}{y_{\{i\}}} \\ z_I^{(0)} &= \frac{y_I - y_{I \cup \{i\}}}{1 - y_{\{i\}}}. \end{aligned}$$

Then, $y = y_{\{i\}} z^{(1)} + (1 - y_{\{i\}}) z^{(0)}$ and $z^{(0)}, z^{(1)} \in \text{Las}_{t-1}(K)$. Furthermore,

$$\begin{aligned} z_{\{i\}}^{(0)} &= 0 \\ z_{\{i\}}^{(1)} &= 1. \end{aligned}$$

Before we prove this lemma, we remark that $z^{(0)}$ and $z^{(1)}$ can be thought of as the solutions corresponding to conditioning on $x_i = 0$ and $x_i = 1$, respectively.

Proof. We focus on the claim that $z^{(0)}, z^{(1)} \in \text{Las}_{t-1}(K)$. The rest of the lemma is easy to verify.

We present two proofs of the above claim.

Matrix Proof: Define

$$\begin{aligned} \mathcal{P}_{-i} &= \{I : |I| < t, i \notin I\} \\ \mathcal{P}_{+i} &= \{I \cup \{i\} : |I| < t, i \notin I\} \end{aligned}$$

and let M_{-i}, M_{+i} be the principal submatrices of M corresponding to $\mathcal{P}_{-i}$ and $\mathcal{P}_{+i}$, respectively. Then the principal submatrix of $M_t(y)$ indexed by $\mathcal{P}_{-i}, \mathcal{P}_{+i}$ can be written as

$$\begin{aligned} M_t(y) &= \begin{pmatrix} M_{-i} & M_{+i} \\ M_{+i} & M_{+i} \end{pmatrix} \\ &= y_i \begin{pmatrix} \frac{1}{y_i} M_{+i} & \frac{1}{y_i} M_{+i} \\ \frac{1}{y_i} M_{+i} & \frac{1}{y_i} M_{+i} \end{pmatrix} + (1 - y_i) \begin{pmatrix} \frac{1}{1 - y_i}(M_{+i} - M_{-i}) & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned}$$

where the block matrices correspond to the subsets $\mathcal{P}_{-i}$ and $\mathcal{P}_{+i}$, respectively. It is easy to see that $M_{t-1}(z^{(1)})$ and $M_{t-1}(z^{(0)})$ are principal submatrices of the two terms on the right hand side, so it suffices to show that the two matrices on the right hand side are both positive semidefinite. The first is clear since the submatrix M_{+i} is PSD. For the second, we have for any vector w that

$$w^T(M_{-i} - M_{+i})w = \begin{pmatrix} w & -w \end{pmatrix} \begin{pmatrix} M_{-i} & M_{+i} \\ M_{+i} & M_{+i} \end{pmatrix} \begin{pmatrix} w \\ -w \end{pmatrix} \geq 0.$$

A similar argument can be used to show that $M^l_{t-1}(z^{(1)})$ and $M^l_{t-1}(z^{(0)})$ are both PSD. We can therefore conclude that $z^{(1)}$ and $z^{(0)}$ both belong to $\mathrm{Las}_{t-1}(K)$.

Vector Proof: Since $M_t(y) \succeq 0$, there exist vectors $v_I, I \subseteq [n]$ such that $\langle v_I, v_J \rangle = y_{I \cup J}$ for all $|I|, |J| \leq t$. It suffices to find vectors $v^{(0)}, v^{(1)}$ so that

$$\begin{aligned} \langle v_I^{(1)}, v_J^{(1)} \rangle &= \frac{y_{I\cup J\cup\{i\}}}{y_{\{i\}}} \\ \langle v_I^{(0)}, v_J^{(0)} \rangle &= \frac{y_{I\cup J} - y_{I\cup J\cup\{i\}}}{1 - y_{\{i\}}} \end{aligned}$$

These equations are satisfied by

$$\begin{aligned} v_I^{(1)} &= \frac{v_{I\cup\{i\}}}{\sqrt{y_{\{i\}}}} \\ v_I^{(0)} &= \frac{v_I - v_{I\cup\{i\}}}{\sqrt{1 - y_{\{i\}}}} \end{aligned}$$

$\square$

The previous lemma shows that $y = \mathrm{conv}(z \in \mathrm{Las}_{t-1}(K) : z_i \in \{0,1\})$. We can therefore iterate this process to get the following:

Corollary 2.1. *If $y \in Las_t(K)$ and $S \subseteq [n], |S| \leq t$, then*

$$y \in conv\left\{Z \in Las_{t-|S|}(K) : z_i \in \{0,1\} \forall i \in S\right\}.$$

In particular, this shows that we recover the convex hull of integer solutions after iterating the Lasserre SDP at most n times. The following lemma explicitly describes the above decomposition.

Lemma 2.3. *Given a set $S \subseteq [n]$ which is a disjoint union of J_0 and J_1, we define*

$$y_I^{J_0J_1} = \sum_{H\subseteq J_0} (-1)^{|H|} Y_{I\cup J_1\cup H}.$$

Then,

$$y = \sum_{J_0\cup J_1 = S, y_\emptyset^{J_0J_1} > 0} y_\emptyset^{J_0J_1} z^{J_0J_1},$$

where $z^{J_0J_1} = \dfrac{y^{J_0J_1}}{y_\emptyset^{J_0J_1}}$ is in $Las_{t-|S|}(K)$, with $z_i^{J_0J_1} = \begin{cases} 1 & i \in J_1 \\ 0 & i \in J_0 \end{cases}$.

Proof. The proof is by induction. Suppose the claim is shown for S and we wish to show it for $S \cup \{i\}$. By definition,

$$y_{I\cup\{i\}}^{J_0 J_1} = \sum_{H \subseteq J_0} (-1)^{|H|} y_{I\cup J_1 \cup H \cup \{i\}} = y_I^{J_0, J_1\cup\{i\}}.$$

Furthermore,

$$\begin{aligned} y_I^{J_0 J_1} - y_{I\cup\{i\}}^{J_0 J_1} &= \sum_{H \subseteq J_0} (-1)^{|H|} y_{I\cup J_1 \cup H} - \sum_{H \subseteq J_0} (-1)^{|H|} y_{I\cup J_1 \cup H \cup \{i\}} \\ &= \sum_{H \subseteq J_0 \cup \{i\}} (-1)^{|H|} y_{I\cup J_1 \cup H} \\ &= y_I^{J_0\cup\{i\}, J_1}. \end{aligned}$$

The result then follows using Lemma 2.2. □

2 Application

Consider the set cover problem. Given sets $S_1, \ldots, S_m$ with associated costs $c_1, \ldots, c_m$ and elements $\{1, \ldots, n\}$, we wish to cover all elements with the minimum cost. It is known that the greedy algorithm has a $1 + \ln(n)$ approximation ratio (or an approximation gap of $\ln(n)$); that is, it finds a covering that may be $1 + \ln(n)$ times as large as the minimum one. A natural question is whether or not we can find an efficient algorithm with a smaller approximation gap. However, finding a polynomial time algorithm that achieves a $(1 - \epsilon)\ln(n)$ gap is known to be hard. What we now present is an algorithm that achieves a $(1 - \epsilon)\ln(n)$ gap in time 2^{n^ϵ}.

Given a set cover instance, our goal is to recover $y \in \text{Las}_{n^\epsilon}(K)$. It turns out that we must first guess the optimal solution. We will skip over the technical details of this step and assume that we have a guess OPT. Our variable of interest will be the vector x, where $x_i = \mathbb{I}\{S_i \text{ is picked}\}$. Our optimization problem will be as follows:

$$\begin{aligned} \text{minimize} \ & \sum_i c_i x_i \\ \text{s.t.} \ & \sum_{i: j \in S_i} \geq 1, \\ & \sum_i c_i x_i \leq OPT. \end{aligned}$$

Remark: We will note without proof that if we constrain the maximum set size to be k, then we can get a $\ln(k)$ approximation gap.

The main idea of the algorithm is to reduce to a solution where the maximum set size is $n^{1-\epsilon}$. To do this, we start with the solution $y \in \text{Las}_{n^\epsilon}(K)$ and repeat the following for $n^\epsilon - 1$ iterations:

1. Pick the largest set that has a positive weight under the current Lasserre solution $y_t \in \text{Las}_t(K)$.

2. Condition on the weight of that set being 1. This gives us a solution $y_{t-1} \in \text{Las}_{t-1}(K)$.

Observe that the third condition still holds after each iteration. Furthermore, after $n^{\epsilon} - 1$ iterations, no remaining set has more than $n^{1-\epsilon}$ uncovered elements. An application of the earlier remark then gives the desired result.

References

[Rot13] Thomas Rothvoß. The lasserre hierarchy in approximation algorithms, June 2013.