

5. Properties of the Lasserre Relaxation.

Lecturer: Massimo Lauria

Disclaimer: this lecture note has not yet been reviewed by the main lecturer. It is released as it is for the convenience of the students.

In this lecture, we study the properties of the solution vectors of the Lasserre relaxation. We start with some basic properties of the solutions, and then continue to show that the solution vectors give us a family of locally consistent distributions over feasible integral local solutions. At the end of the lecture, we see an example of this property in graph coloring.

In this lecture we prove some properties of the solution vectors of the Lasserre hierarchy. (These notes follow closely the lecture notes made by Thomas Rothvoß¹.)

First, we recall the notation defined in previous lectures. Let

$$K = \{x \in \mathbb{R}^n \mid \forall i, h_i(x) \geq 0\} \quad (1)$$

be some relaxation of a binary feasibility problem. That is, K is the set of fractional points that satisfies all the constraints, where the h_i 's are polynomial constraints. In most of the previous examples, all h_i 's are linear and therefore K is a **polytope**. Otherwise, we usually call K a *semialgebraic set*. In the rest of the lecture, we assume that all constraints are linear.

The polytope $\mathcal{P}_I = \text{conv}(K \cap \{0, 1\}^n)$ is the convex hull of all integral solutions. Recall that the t -th level of the Lasserre hierarchy $L_t(K)$ is the set of vectors $y \in \mathbb{R}^{2^{\lfloor n \rfloor}}$ that satisfies the following constraints:

$$\begin{aligned} M_t(y) &\geq 0, \\ M_t(h_i \circ y) &\geq 0, \quad \forall i, \\ y_\emptyset &= 1. \end{aligned} \quad (2)$$

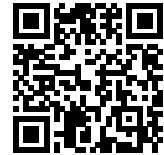
Let $Q_t(K)$ be the projection of $L_t(K)$ over $y_{\{i\}} =: x_i$. For simplicity of notation, we write y_i instead of $y_{\{i\}}$ in the rest of the note.

Basic properties of the Lasserre relaxation

We first discuss some of the basic properties of the Lasserre relaxation. For more details, see Lemma 1 and the corresponding proof in lecture notes².

Lemma 1. Define K and $L_t(K)$ as above for some $t \geq 0$, and let $y \in L_t(K)$. Then the following holds:

1. For all $|I| \leq t$, $0 \leq y_I \leq 1$.
2. For any $I \subseteq J$, $|J| \leq t$, $0 \leq y_J \leq y_I$.
3. For $|I| \leq t$, $|J| \leq t$, $|y_{I \cup J}| \leq \sqrt{y_I y_J}$.
4. For $|S| \leq t$, $y \in L_t(K)$, if there exists $i \in S$ such that $y_i = 0$, then $y_S = 0$. Similarly, if for all $i \in S$, $y_i = 1$, then $y_S = 1$.



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

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

² Thomas Rothvoß. The lasserre hierarchy in approximation algorithms. Lecture Notes for the MAPSP 2013 Tutorial (preliminary version), 2013

Proof. For proving the first two statements we make use of the fact that $M_t(y) \geq 0$ if and only if every principal minor of $M_t(y)$ is greater than or equal to zero.

1) Consider the principal minor

$$\begin{vmatrix} y_\emptyset & y_I \\ y_I & y_I \end{vmatrix} = y_I(1 - y_I) \geq 0. \quad (3)$$

Constraint (3) yields $0 \leq y_I \leq 1$.

2) Consider the principal minor

$$\begin{vmatrix} y_I & y_{I \cup J} \\ y_{I \cup J} & y_J \end{vmatrix} = y_I y_J - y_{I \cup J}^2 \geq 0. \quad (4)$$

Since $I \subseteq J$, we have that $y_{I \cup J} = y_J$. Therefore constraint (4) is equivalent to $y_I y_J - y_J^2 \geq 0$. From 1) we have $y_I, y_J \geq 0$, thus $y_I \geq y_J$.

3) Again, consider (4).

4) The first half of the statement follows by applying the first statement of Lemma 1 with $I := \{i\}$ and $J := S$.

For the second half of the statement, we can use the fact that the Sherali-Adams constraints are implied by the Lasserre constraints, and use

$$\sum_{H \subseteq S'} (-1)^{|H|} y_H \geq 0$$

inductively for all sets $S' \subseteq S$. □

Consistent local distributions

Next, we show that the vectors in $L_t(K)$ can be written as a convex combination of fractional solutions that are locally integral, see for example Section 2.3 in lecture notes³. We have the following theorem:

Theorem 2. *Let $y \in L_t(K)$ and $S \subseteq [n]$, $|S| \leq t$. Then $y \in \text{conv}(z \mid z \in L_{t-|S|}(K); \forall i \in S, z_i \in \{0, 1\})$.*

To prove Theorem 2, we first prove the following claim:

Claim 3. *Let $y \in L_t(K)$, and $i \in [n]$ be some variable where $y_i \notin \{0, 1\}$. Then there exist $z^0, z^1 \in L_{t-1}(K)$, and $\alpha, \beta \in [0, 1]$, such that $\alpha + \beta = 1$, $z_i^0 = 0$, $z_i^1 = 1$, and $y = \alpha z^0 + \beta z^1$.*

Proof. Since $y_i \notin \{0, 1\}$, we can define the elements of z^0 and z^1 as

$$z_I^1 = \frac{y_{I \cup \{i\}}}{y_i},$$

$$z_I^0 = \frac{y_I - y_{I \cup \{i\}}}{1 - y_i}.$$

Clearly $z_i^0 = 0$ and $z_i^1 = 1$, and with $\alpha = 1 - y_i$ and $\beta = y_i$, we get $y = \alpha z^0 + \beta z^1$. Now we verify that $z^0, z^1 \in L_t(K)$.

First, we check that $M_t(y) \geq 0$ implies that $M_{t-1}(z^1) \geq 0$. Recall that since the moment matrix M_t is positive semidefinite, there exists a v_S for

³ Thomas Rothvoß. The lasserre hierarchy in approximation algorithms. Lecture Notes for the MAPSP 2013 Tutorial (preliminary version), 2013

each $S \subseteq [n]$, $|S| \leq t$, such that $M_t(\mathbf{y})_{I,J} = \langle v_I, v_J \rangle$, for all $|I| \leq t$, $|J| \leq t$.

Define

$$u_I^1 = \frac{1}{\sqrt{y_i}} v_{I \cup \{i\}}.$$

Then the matrix $M = (\langle u_I^1, u_J^1 \rangle)_{|I| \leq t-1, |J| \leq t-1}$ is positive semidefinite by definition. We need to show that $M_{I,J} = z_{I \cup J}^1$ for $|I|, |J| \leq t-1$. Note that

$$M_{I,J} = \frac{1}{y_i} \langle v_{I \cup \{i\}}, v_{J \cup \{i\}} \rangle = \frac{1}{y_i} y_{I \cup J \cup \{i\}} = z_{I \cup J}^1,$$

thus, $M_{t-1}(z^1) \geq 0$.

Second, we check that $M_t(\mathbf{y}) \geq 0$ implies that $M_{t-1}(z^0) \geq 0$. Similarly to above, we can define

$$u_I^0 = \frac{v_I - v_{I \cup \{i\}}}{\sqrt{1 - y_i}},$$

and conclude that $M_{t-1}(z^0) \geq 0$.

The final thing we need to show is that $M_t(h \circ \mathbf{y}) \geq 0$ implies that $M_{t-1}(h \circ z^0) \geq 0$ and $M_{t-1}(h \circ z^1) \geq 0$. The proof idea is similar to the above argument. Remember that if we assume $h = \sum_S \alpha_S \prod_{i \in S} x_i$, then

$$M_t(h \circ \mathbf{y})_{I,J} = \langle \tilde{v}_I, \tilde{v}_J \rangle = \sum_S \alpha_S y_{I \cup J \cup S}.$$

Now define

$$\tilde{v}_I^1 = \frac{\tilde{v}_{I \cup \{i\}}}{y_i},$$

and we can verify that indeed $M_{t-1}(h \circ z^1) \geq 0$. Similarly we can show that $M_{t-1}(h \circ z^0) \geq 0$.

For a different proof formulation see Proof 1 of Lemma 2 in lecture notes⁴. □

To prove Theorem 2, we simply apply Claim 3 iteratively to all variables in S .

The following statement shows that we need at most level n of the Lasserre hierarchy to obtain \mathcal{P}_I :

Corollary 4. $Q_n(K) = \mathcal{P}_I$.

Suppose now that we have some $\mathbf{y} \in L_n(K)$. Then, summarizing the above, we can see that there exists some distribution \mathcal{D} over $\{0, 1\}^n$, such that for any $I \subseteq [n]$, we have

$$y_I = \Pr_{X \sim \mathcal{D}} \left[\bigwedge_{i \in I} (X_i = 1) \right].$$

Now consider the sets of indices $J_0, J_1 \subseteq [n]$. By the inclusion-exclusion principle, we have

$$\Pr_{X \sim \mathcal{D}} \left[\bigvee_{i \in J_0} (X_i = 1) \right] = \sum_{\emptyset \subset H \subseteq J_0} (-1)^{|H|+1} \Pr_{X \sim \mathcal{D}} \left[\bigwedge_{i \in H} (X_i = 1) \right].$$

Negating this gives

$$\Pr_{X \sim \mathcal{D}} \left[\bigwedge_{i \in J_0} (X_i = 0) \right] = \sum_{H \subseteq J_0} (-1)^{|H|} \Pr_{X \sim \mathcal{D}} \left[\bigwedge_{i \in H} (X_i = 1) \right]. \quad (5)$$

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

Equation (5) yields the generalized inclusion-exclusion formula

$$\begin{aligned} & \Pr_{X \sim \mathcal{D}} \left[\bigwedge_{i \in J_1} (X_i = 1) \wedge \bigwedge_{i \in J_0} (X_i = 0) \right] \\ &= \sum_{H \subseteq J_0} (-1)^{|H|} \Pr_{X \sim \mathcal{D}} \left[\bigwedge_{i \in J_1 \cup H} (X_i = 1) \right]. \end{aligned} \quad (6)$$

This motivates the following definition:

Definition 5. For $y \in L_n(K)$, define

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

From this, we have

$$\begin{aligned} y_{I \cup \{i\}}^{J_0, J_1} &= y_I^{J_0, J_1 \cup \{i\}} \\ y_I^{J_0 \cup \{i\}, J_1} &= y_I^{J_0, J_1} - y_{I \cup \{i\}}^{J_0, J_1}. \end{aligned} \quad (7)$$

Also, corresponding to (6), we have

$$\begin{aligned} y_{\emptyset}^{J_0, J_1} &= \Pr_{X \sim \mathcal{D}} \left[\bigwedge_{i \in J_0} (X_i = 0) \wedge \bigwedge_{i \in J_1} (X_i = 1) \right] \\ y_I^{J_0, J_1} &= \Pr_{X \sim \mathcal{D}} \left[\bigwedge_{i \in I} (X_i = 1) \wedge \bigwedge_{i \in J_0} (X_i = 0) \wedge \bigwedge_{i \in J_1} (X_i = 1) \right]. \end{aligned}$$

Definition 6. Assume that $y_{\emptyset}^{J_0, J_1} > 0$, define

$$z_I^{J_0, J_1} = \frac{y_I^{J_0, J_1}}{y_{\emptyset}^{J_0, J_1}}. \quad (8)$$

If $y \in L_n(K)$, we have, by Bayes' formula, that

$$z_I^{J_0, J_1} = \Pr_{X \sim \mathcal{D}} \left[\bigwedge_{i \in I} (X_i = 1) \mid \bigwedge_{i \in J_0} (X_i = 0) \wedge \bigwedge_{i \in J_1} (X_i = 1) \right].$$

This is the intuition of the following lemma:

Lemma 7. Let $y \in L_t(K)$, $S \subseteq [n]$, $|S| \leq t$. Define z^{J_0, J_1} as in (8). Then we have

1. $\sum_{J_0 \cup J_1 = S} y_{\emptyset}^{J_0, J_1} = 1$.
2. for all $i \in J_0$, $z_i^{J_0, J_1} = 0$, and for all $i \in J_1$, $z_i^{J_0, J_1} = 1$.
3. for all J_0, J_1 , $z^{J_0, J_1} \in L_{t-|S|}(K)$.
4. y expressed as the convex combination

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

Proof. We prove the statements inductively on the size of S . Suppose $S = S' \cup \{i\}$, and we have proved the lemma for S' . Then

$$\begin{aligned} y_I &= \sum_{J_0 \sqcup J_1 = S'} y_I^{J_0, J_1} \\ &= \sum_{J_0 \sqcup J_1 = S'} \left((y_{I \cup \{i\}}^{J_0, J_1}) + (y_I^{J_0, J_1} - y_{I \cup \{i\}}^{J_0, J_1}) \right) \\ &= \sum_{J_0 \sqcup J_1 = S'} y_I^{J_0, J_1 \cup \{i\}} + \sum_{J_0 \sqcup J_1 = S'} y_I^{J_0 \cup \{i\}, J_1} \\ &= \sum_{J_0 \sqcup J_1 = S} y_I^{J_0, J_1}. \end{aligned}$$

This, along with Definition 6, gives the RHS of the equality in 4). Setting $I = \emptyset$ and noticing that $y_\emptyset = 1$ give us the equality in 1). The proof of 2) and 3) is similar to that of Claim 3. \square

From this, it is easy to define locally consistent probability distributions from vectors in $L_t(K)$.

Theorem 8. *Let $y \in L_t(K)$. Then for any $S \subseteq [n]$, $|S| \leq t$, there is a distribution $\mathcal{D}(S)$ over $\{0, 1\}^S$, such that for all $I \subseteq S$*

$$\Pr_{z \sim \mathcal{D}(S)} \left[\bigwedge_{i \in I} z_i = 1 \right] = y_I.$$

Proof. Using Lemma 7, we can write y as a combination of vectors,

$$y = \sum_{\substack{J_0 \sqcup J_1 = S \\ y_\emptyset^{J_0, J_1} > 0}} y_\emptyset^{J_0, J_1} z^{J_0, J_1}.$$

In the distribution $\mathcal{D}(S)$, with probability $y_\emptyset^{J_0, J_1}$, we set all bits in J_1 to 1 and all bits in J_0 to 0. Note that for $I \subseteq S$,

$$\begin{aligned} \Pr_{z \sim \mathcal{D}(S)} \left[\bigwedge_{i \in I} (z_i = 1) \right] &= \sum_{\substack{I \subseteq J_1 \subseteq S \\ J_0 = S \setminus J_1}} y_\emptyset^{J_0, J_1} \\ &= \sum_{J_0 \sqcup J_1 = S \setminus I} y_I^{J_0, J_1} = y_I. \end{aligned}$$

\square

For a concrete example of this property, we look at GRAPH 3-COLORING. Let $G = (V, E)$ be an undirected graph and suppose it is 3-colorable, i.e., the nodes can be colored with red, green and blue such that adjacent nodes get assigned different colors. Then we get the following relaxation:

$$\begin{cases} x \in \{0, 1\}^n \\ x_{i,c} + x_{j,c} \leq 1 & \forall c \in \{R, G, B\}, \{i, j\} \in E \\ \sum_c x_{i,c} \geq 1 & \forall i \in V. \end{cases}$$

Lemma 9. *Let $y \in L_{3t}(K)$ be a solution for the $3t$ -th level Lasserre relaxation of the above program. Then there is a family of distributions $\{\mathcal{D}(S)\}_{S \subseteq V, |S| \leq t}$, such that for any $S \subseteq V$ with $|S| \leq t$, we have that*

1. any χ in the support of $\mathcal{D}(S)$ is a legal coloring for the induced subgraph $G[S]$.
2. the probability $\Pr_{\chi \sim \mathcal{D}(S)} [\chi(i_1) = c_1, \forall i = 1, \dots, k] = y_{\{(i_1, c_1), (i_2, c_2), \dots, (i_k, c_k)\}}$, for all $i_1, \dots, i_k \in S$, $c_1, \dots, c_k \in \{R, G, B\}$.

References

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