Notes on Lasserre hierarchies, SOS relaxations and pseudo-expectations

Amitabh Basu

December 9, 2019

1 Why these notes?

Lasserre hierarchies were introduced as convex relaxations of non-convex optimization problems with polynomial objective and constraints [5, 6]; see also [7]. Subsequently, these relaxations were studied from different perspectives by researchers in optimization, approximation algorithms, complexity theory, and control theory. Given all of these different ways of looking at it, the mathematical objects that go by the name of "Lasserre hierarchies" ostensibly look very different from each other. Each community has its own way of defining what a "Lasserre hierarchy" is and, to the uninitiated, it is not immediate why they define the same, or even related optimization problems. Adding to the (at least the author's) confusion is the "dual picture" of *Sum-Of-Squares (SOS)* relaxations.

This note is an attempt by the author to clarify for himself the precise relationships between the different definitions of Lasserre hierarchies and SOS relaxations. We first present the different ways these problems are defined in the literature, and in the subsequent sections try to achieve a synthesis.

Optimization with 0-1 variables. In combinatorial optimization, one is interested in a feasible set which is given by the 0-1 vertices of the standard hypercube that are contained in some polytope, or more generally, in some convex set. One can consider a generalization where instead of a polytope or convex set, one considers a set given by polynomial inequalities. Thus, one considers a feasible region of the form

$$\{\mathbf{x} \in \{0,1\}^n : q_i(\mathbf{x}) \geq 0,\ i = 1, \ldots, m\}, \tag{1.1}$$

where $q_1(\mathbf{x}), \ldots, q_m(\mathbf{x})$ are arbitrary multilinear polynomials (one may also consider general multivariate polynomials that are not multilinear, but this does not give any extra modeling power since we are restricting to the Boolean hypercube; see the discussion in Section 2; in particular, part 3 of Lemma 2.2.). As mentioned, the typical situation in combinatorial optimization uses degree 1 (linear) polynomials, i.e., we consider $0-1$ points in a polytope. The Lasserre hierarchy for (1.1) is defined as a sequence $L_1, L_2, \ldots, L_t, \ldots$ of convex sets in $\mathbb{R}^n$. First, we need to establish some notation. For any $z \in \mathbb{R}^{2^n}$ we consider its coordinates as indexed by all subsets of $\{1, \ldots, n\}$. Given any $z \in \mathbb{R}^{2^n}$ and $t \geq 0$, we define $M_t(z)$ as the matrix whose rows and columns are indexed by all subsets of $\{1, \ldots, n\}$ of size at most t, and the entry in row I and column J is given by $z_{I\cup J}$. Note that the coefficients of any multilinear polynomial $q(\mathbf{x})$ can also be indexed by subsets of $\{1, \ldots, n\}$, which we will refer to as q_I. Associated with each such polynomial q, we define a linear operator P_q on $\mathbb{R}^{2^n}$ as follows: $P_q(z)_I := \sum_J q_J z_{I\cup J}$.

Given multivariate polynomials $q_1, \ldots, q_m$, one first defines the sequence of convex sets $\tilde{L}_t, t \geq 1$ as follows:

$$\tilde{L}_t(q_1, \ldots, q_m) := \{y \in \mathbb{R}^{2^n} : y_\emptyset = 1,\ \ M_t(y), M_t(P_{q_1}(y)), \ldots, M_t(P_{q_m}(y)) \text{ are all PSD matrices}\}. \tag{1.2}$$

Each L_t is defined as the projection of $\tilde{L}_t$ on to the space of the n singletons, and is called the *t-th level of the Lasserre hierarchy*. The discussion above is based on [8] and (as far as the author can see) is the predominant view of the Lasserre hierarchy amongst the approximation algorithms community.

Sum-Of-Squares (SOS) relaxations in optimization/control. For the "dual" SOS picture, we will need the following definitions.

Definition 1.1. 1. Let $\mathbb{R}[\mathbf{x}]$ denote the vector space/algebra of all polynomials over n variables $\mathbf{x} = (x_1, \ldots, x_n)$. Given any $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_+$, we use the compact notation $\mathbf{x}^\alpha$ to denote the monomial $x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n}$.

2. A polynomial $u \in \mathbb{R}[\mathbf{x}]$ is called a *sum of squares (SOS)* polynomial if there exist polynomials $p_1, \ldots, p_k \in \mathbb{R}[\mathbf{x}]$ such that $u = \sum_i p_i^2$.

3. Let $D \geq 1$ be a natural number. Let $\mathbb{R}[\mathbf{x}]_D$ denote the set of polynomials with degree at most D. Let SOS_D denote the subset of SOS polynomials in $\mathbb{R}[\mathbf{x}]_D$.

Let $p, q_1, \ldots, q_m \in \mathbb{R}[\mathbf{x}]$. Consider the polynomial optimization problem

$$\begin{array}{rll} p^\star := & \inf_{\mathbf{x} \in \mathbb{R}^n} & p(\mathbf{x}) \\ & \text{s.t.} & q_i(\mathbf{x}) \geq 0 \quad i = 1, \ldots, m \end{array} \tag{1.3}$$

If we define $K = \{\mathbf{x} : q_i(\mathbf{x}) \geq 0\}$, then (1.3) is equivalent to

$$\begin{array}{ll} \sup_{\lambda \in \mathbb{R}} & \lambda \\ \text{s.t.} & p - \lambda \geq 0 \text{ on } K \end{array} \tag{1.4}$$

For any even natural number $d \geq 2$, let $D = \max\{\deg(p), d + \deg(q_1), \ldots, d + \deg(q_m)\}$, (1.4) can be relaxed to the following conic optimization problem over the vector space of polynomials $\mathbb{R}[\mathbf{x}]_D$, called the *degree-d SOS relaxation* of (1.4):

$$\begin{array}{rll} \gamma_d := & \sup_{\lambda \in \mathbb{R}, u_0, u_1, \ldots, u_m \in \mathbb{R}[\mathbf{x}]_d} & \lambda \\ & \text{s.t.} & p - \lambda = u_0 + u_1 q_1 + \ldots u_m q_m \\ & & u_0, u_1, \ldots, u_m \in SOS_d \end{array} \tag{1.5}$$

Thus, $\gamma_d \leq p^\star$ because any λ satisfying the constraints in (1.5) must satisfy the constraints in (1.4). Using Theorem 2.11, one can write a *semidefinite program (SDP)* that has the same optimal value as (1.5). Moreover, the size of the SDP (i.e., dimension or number of variables of the SDP) is $O(mn^D)$. The SOS picture as presented here is based on [7].

The pseudo-expectation picture. One can derive the dual conic program to (1.5) as done in Appendix A:

$$\begin{array}{rll} \rho_d := & \inf_{L : \mathbb{R}[\mathbf{x}]_D \to \mathbb{R} \text{ linear map}} & L(p) \\ & \text{s.t.} & L(\mathbf{1}) = 1 \\ & & L(u) \geq 0 \qquad \forall u \in SOS_d \\ & & L(uq_i) \geq 0 \quad \forall u \in SOS_d, \;\; \forall i = 1, \ldots, m \end{array} \tag{1.6}$$

Remark 1.2. By extending the linear map L in (1.6) trivially to all of $\mathbb{R}[\mathbf{x}]$ by setting all monomials of degree higher than D to 0, one can replace the condition in (1.6) that L is a linear functional on $\mathbb{R}[\mathbf{x}]_D$ to say that L is a linear functional on all of $\mathbb{R}[\mathbf{x}]$. This motivates the following definition.

Definition 1.3. For any $d \geq 1$, a *degree-d Lasserre map*, also known as a *degree-d pseudo-expectation operator* is a linear functional $L : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ such that $L(\mathbf{1}) = 1$, and $L(u) \geq 0$ for all SOS polynomials u of degree at most d (equivalently, $L(p^2) \geq 0$ for all polynomials p of degree at most $\frac{d}{2}$).

The problem (1.6) is also known in the literature as the *degree-d Lasserre relaxation for* (1.3). At first glance, it looks like the feasible region of (1.6) and the t-th levels of the Lasserre hierarchy L_t defined earlier have very little to do with each other. It will turn out that they define essentially the same thing when we consider $0-1$ problems.

The pseudo-expectation picture appears in both approximation algorithms literature, as well as complexity theory literature [2] and was also defined by Lasserre in his original papers; see [7]. However, in approximation algorithms, the L_t sets are perhaps the dominating definition for "Lasserre hierarchies". Moreover, in complexity theory, the predominant definition is in terms of what are known as "pseudo-distributions", which we explore next.

The pseudo-distribution picture. We will illustrate the complexity theory view of the Lasserre/SOS relaxations by considering a simpler question:

Given a polynomial $p \in \mathbb{R}[\mathbf{x}]$, decide if $p \geq 0$ for all $\mathbf{x} \in \{-1,1\}^n$.

Many of the interesting complexity theory questions can be phrased in this framework. The problem, as stated, is a special case of (1.3), since we can have p as the objective and the constraints as $x_i^2 \geq 1$ and $x_i^2 \leq 1$ for all $i = 1, \ldots, n$; if the minimum in (1.3) is greater than or equal to 0, we report "YES", else we report "NO". The following definition is central in the complexity theory view:

Definition 1.4. A *degree-d pseudo-distribution* is a function $D : \{-1,1\}^n \to \mathbb{R}$ such that $\sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} D(\bar{\mathbf{x}}) = 1$, and $\sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} D(\bar{\mathbf{x}})u(\bar{\mathbf{x}}) \geq 0$ for all SOS polynomials u of degree at most d.

The set of degree-d pseudo-distribution (viewed as a subset of $\mathbb{R}^{2^n}$) is easily seen to be a convex set, because it is the intersection of the (infinite many) halfspaces $\sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} D(\bar{\mathbf{x}})u(\bar{\mathbf{x}}) \geq 0$ for all SOS polynomials u of degree at most d, and the constraint $\sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} D(\bar{\mathbf{x}}) = 1$. The *degree-$d$ pseudo-distribution relaxation* for the nonnegativity question posed above is defined as the convex optimization problem

$$\inf\{ \sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} D(\bar{\mathbf{x}})p(\bar{\mathbf{x}}) \; : \; D \text{ is a degree-}d \text{ pseudo-distribution}\} \tag{1.7}$$

The optimal value of the above problem is clearly a lower bound on the smallest value p can take on $\{-1,1\}^n$ because the indicator functions of each vertex of the hypercube are valid solutions to the convex relaxation above. Below, we will show a simple connection between degree-d Lasserre maps (pseudo-expectation operators) and degree-d pseudo-distributions. Our presentation of this view of Lasserre relaxations is based on [3].

2 Definitions and Preliminaries

We will need to collect some facts about multivariate polynomials. In particular, we want to understand two things well: 1) the consequences of restricting the input of the polynomials to the Boolean hypercube, and 2) connections between SOS polynomials and semidefinite programs (SDPs).

All vector spaces in this note are over the field of reals, $\mathbb{R}$. The set of natural numbers will be denoted by $\mathbb{N} = \{1, 2, \ldots, \}$ and the integers will be denoted $\mathbb{Z}$ and the nonnegative integers by $\mathbb{Z}_+$.

2.1 Restricting the input of polynomials to the Boolean hypercube

Definition 2.1.

1. The *evaluation map or operator on* $\{-1,1\}^n$ is the map $E : \mathbb{R}[\mathbf{x}] \to \mathbb{R}^{2^n}$ defined by $E[p] = (p(\bar{\mathbf{x}}))_{\bar{\mathbf{x}} \in \{-1,1\}^n}$. Its kernel $\ker(E)$ is the set of all polynomials that evaluates to 0 on the hypercube $\{-1,1\}^n$; of course, $\ker(E)$ is a vector subspace of $\mathbb{R}[\mathbf{x}]$ (in fact $\ker(E)$ is an *ideal* for the ring $\mathbb{R}[\mathbf{x}]$, i.e., $q \cdot p \in \ker(E)$ for all $q \in \mathbb{R}[\mathbf{x}]$ and $p \in \ker(E)$ – but we will not exploit this algebraic geometry perspective below).

2. The *evaluation map or operator on* $\{0,1\}^n$ is the map $E^\dagger : \mathbb{R}[\mathbf{x}] \to \mathbb{R}^{2^n}$ defined by $E^\dagger[p] = (p(\bar{\mathbf{x}}))_{\bar{\mathbf{x}} \in \{0,1\}^n}$. Its kernel $\ker(E^\dagger)$ is the set of all polynomials that evaluates to 0 on the hypercube $\{0,1\}^n$; of course, $\ker(E^\dagger)$ is a vector subspace of $\mathbb{R}[\mathbf{x}]$ (in fact $\ker(E^\dagger)$ is an *ideal* for the ring $\mathbb{R}[\mathbf{x}]$, i.e., $q \cdot p \in \ker(E^\dagger)$ for all $q \in \mathbb{R}[\mathbf{x}]$ and $p \in \ker(E)$ – but we will not exploit this algebraic geometry perspective below).

3. Let $MULT_n$ be the subspace of $\mathbb{R}[\mathbf{x}]$ spanned by the 2^n multilinear monomials.

Lemma 2.2. The following are all true.

1. The set of 2^n vectors $\{E(m) : m \text{ multilinear monomial}\}$ is an orthogonal set in $\mathbb{R}^{2^n}$. The set of 2^n vectors $\{E^\dagger(m) : m \text{ multilinear monomial}\}$ is a linearly independent set in $\mathbb{R}^{2^n}$ (but not orthogonal).

2. The evaluation operator E is an isomorphism between the vector spaces $\mathbb{R}[\mathbf{x}]/\ker(E)$ and $\mathbb{R}^{2^n}$. The same holds when E is replaced by $E^\dagger$.

3. $MULT_n \cap \ker(E) = \{0\}$ and the quotient vector space $\mathbb{R}[\mathbf{x}]/\ker(E)$ is isomorphic to $MULT_n$. Therefore, $\mathbb{R}[\mathbf{x}] = MULT_n \oplus \ker(E)$, i.e., for any polynomial $p \in \mathbb{R}[\mathbf{x}]$, there exists a unique multilinear polynomial $\bar{p}$ and a polynomial $q \in \ker(E)$ such that $p = \bar{p} + q$.

   The same holds when E is replaced by $E^\dagger$ in the above. Of course, in this case, any $p \in \mathbb{R}[\mathbf{x}]$ is decomposed into $\tilde{p} + q$ where $q \in \ker(E^\dagger)$ and $\tilde{p}$ is different from $\bar{p}$ above.

Proof. 1. This is a well-known fact in Boolean complexity theory; we repeat the proof here for completeness. Let $m_1 = \mathbf{x}^{\alpha_1}, m_2 = \mathbf{x}^{\alpha_1}$ be two multilinear monomials with $\alpha_1, \alpha_2 \in \{0,1\}^n$. Consider the inner product

$$\langle E(m_1), E(m_2)\rangle = \sum_{\mathbf{y}\in\{-1,1\}^n} m_1(\mathbf{y})m_2(\mathbf{y}) = \sum_{\mathbf{y}\in\{-1,1\}^n} \mathbf{y}^{(\alpha_1 \ \ XOR \ \ \alpha_2)}.$$

The last term is 0 if $\alpha_1 \neq \alpha_2$, i.e., $m_1 \neq m_2$.

For $E^\dagger$, one observes that if $\{E^\dagger(m) : m \text{ multilinear monomial}\}$ is not linearly independent, then there exist scalars $\alpha_m \in \mathbb{R}$ for each multilinear monomial m, not all zero, such that $0 = \sum_m \alpha_m E^\dagger(m) = E^\dagger(\sum_m \alpha_m m)$. This implies that the polynomial $\sum_m \alpha_m m$ evaluates to 0 on all the vertices $\{0,1\}^n$. But this is a contradiction, because any multilinear polynomial with at least one nonzero coefficient evaluates to a nonzero zero value on some point in $\{0,1\}^n$ (one can prove this by induction on n).

2. From part 1. and the fact that $\mathbb{R}^{2^n}$ is a vector space with dimension 2^n, we observe that E is a surjective linear map and thus induces an isomorphism between $\mathbb{R}[\mathbf{x}]/\ker(E)$ and $\mathbb{R}^{2^n}$. The same argument holds for $E^\dagger$.

3. From part 1., we obtain that $MULT_n \cap \ker(E) = \{0\}$ because of the linear independence of $\{E(m) : m \text{ multilinear monomial}\}$. In fact, E induces an isomorphism between $MULT_n$ and $\mathbb{R}^{2^n}$. Combining with part 2., we realize that $MULT_n \cong \mathbb{R}^{2^n} \cong \mathbb{R}[\mathbf{x}]/\ker(E)$. Since $MULT_n$ is a subspace of $\mathbb{R}[\mathbf{x}]$, it follows that $\mathbb{R}[\mathbf{x}]$ is a direct sum of $MULT_n$ and $\ker(E)$. The same arguments hold for $E^\dagger$.

□

Restricting to $\{-1,1\}^n$. The cases of restricting to $\{-1,1\}^n$ and $\{0,1\}^n$ are slightly different and we first consider the $\{-1,1\}^n$ case.

Lemma 2.3. The ideal $\ker(E)$ is generated by the polynomials $x_i^2 - 1$, $i = 1,\ldots,n$. In other words, $p \in \mathbb{R}[\mathbf{x}]$ vanishes on $\{-1,1\}^n$ if and only if there exist polynomials $g_i \in \mathbb{R}[\mathbf{x}]$, $i = 1,\ldots,n$ such that $p = \sum_{i=1}^n g_i(x_i^2 - 1)$. Further, if $p \in \mathbb{R}[\mathbf{x}]_D$ for some natural number D, then the g_i's can be chosen in $\mathbb{R}[\mathbf{x}]_D$.

Proof. Let us denote the ideal generated by the polynomials $x_i^2 - 1$ by

$$I = \langle x_1^2 - 1, x_2^2 - 1, \ldots, x_n^2 - 1\rangle = \{p \in \mathbb{R}[x] : \exists g_1,\ldots,g_n \in \mathbb{R}[\mathbf{x}] \text{ such that } p = g_1(x_1^2 - 1) + \ldots + g_n(x_n^2 - 1)\}$$

Clearly, $I \subseteq \ker(E)$. We want to show the reverse containment, i.e., $\ker(E) \subseteq I$. We first observe the following relations:

$$\begin{aligned} x_i^{2k} &= (x_i^{2k-2} + x^{2k-4} + \ldots + 1)(x_i^2 - 1) + 1 \quad \forall i = 1,\ldots,n, \ \forall k \geq 1 \\ x_i^{2k+1} &= (x_i^{2k-1} + x^{2k-3} + \ldots + x_i)(x_i^2 - 1) + x_i \quad \forall i = 1,\ldots,n, \ \forall k \geq 1 \end{aligned} \tag{2.1}$$

Therefore, any monomial $\mathbf{x}^\alpha$ where $\alpha \in \mathbb{Z}_+^n$ in $\mathbb{R}[\mathbf{x}]$ can be written as $\mathbf{x}^\alpha = m+q$ where m is a multilinear monomial and $q \in I$, simply by replacing each $x_i^{\alpha_i}$ using the relations above (depending on if α_i is even or odd) – in fact, m is multilinear monomial $\prod_{i=1}^n x_i^{(\alpha_i \mod 2)}$. Consequently, any polynomial $p \in \mathbb{R}[\mathbf{x}]$ can be written as $p = \bar{p} + q$ where $q \in I$ and $\bar{p}$ is a linear combination of multilinear monomials. Now if $p \in \ker(I)$, this means that $\bar{p} \in \ker(E)$ (since $q \in I \subseteq \ker(E)$). But a linear combination of multilinear monomials is in $\ker(E)$ if and only if all the coefficients are 0, by Lemma 2.2, part 3. Thus, $\bar{p} = 0$. Therefore, $p = q$ and so $p \in I$.

Since the polynomials used in the relations (2.1) were of the same degree as the exponent of the variable x_i, the second part of the lemma also follows. □

Definition 2.4. For any monomial $\mathbf{x}^\alpha$, we say that $\prod_{i=1}^n x_i^{(\alpha_i \mod 2)}$ is the *equivalent multilinear monomial with respect to* $\{-1,1\}^n$. We similarly define the *multilinear polynomial equivalent to polynomial* p *with respect to* $\{-1,1\}^n$, as the polynomial $\bar{p}$ obtained by changing each monomial of p to its equivalent multilinear monomial.

Corollary 2.5. For any polynomial $p \in \mathbb{R}[x]$, there exists a multilinear polynomial $\hat{p}$ and polynomial q in the ideal generated by $x_i^2 - 1$ such that $p = \hat{p} + q$, and this decomposition is unique.

More explicitly, there exist $g_i \in \mathbb{R}[\mathbf{x}]$, $i = 1, \dots, n$ such that $p = \bar{p} + g_1(x_1^2 - 1) + \dots + g_n(x_n^2 - 1)$, where $\bar{p}$ is the equivalent polynomial of p with respect to $\{-1,1\}^n$.

Proof. By Lemma 2.2, part 3., there exists a unique decomposition $p = \hat{p} + q$ where $\hat{p}$ is a multilinear polynomial and $q \in \ker(E)$. By Lemma 2.3, $\ker(E)$ is equal to the ideal generated by the polynomials $x_i^2 - 1$, and therefore there exist polynomials g_i such that $q = g_1(x_1^2 - 1) + \dots + g_n(x_n^2 - 1)$.

Finally observe that if $\bar{p}$ is the equivalent multilinear polynomial to p, then $E(\bar{p}) = E(p)$ and so $p - \bar{p} \in \ker(E)$. By the uniqueness of the decomposition $p = \hat{p} + q$, we must have $\hat{p} = \bar{p}$. □

Restricting to $\{0,1\}^n$. We now consider the $\{0,1\}^n$ case.

Lemma 2.6. The ideal $\ker(E^\dagger)$ is generated by the polynomials $x_i^2 - x_i$, $i = 1, \dots, n$. In other words, $p \in \mathbb{R}[\mathbf{x}]$ vanishes on $\{0,1\}^n$ if and only if there exist polynomials $g_i \in \mathbb{R}[\mathbf{x}]$, $i = 1, \dots, n$ such that $p = \sum_{i=1}^n g_i(x_i^2 - x_i)$. Further, if $p \in \mathbb{R}[\mathbf{x}]_D$ for some natural number D, then the g_i's can be chosen in $\mathbb{R}[\mathbf{x}]_D$.

Proof. The proof is the very similar to the proof of Lemma 2.6, where instead of the relations (2.1), one uses

$$x_i^k = (x_i^{k-2} + x^{k-1} + \dots + x_i + 1)(x_i^2 - x_i) + x_i \quad \forall i = 1, \dots, n, \ \forall k \geq 1$$

□

Definition 2.7. For any monomial $\mathbf{x}^\alpha$, we say that $\prod_{i=1}^n x_i^{\mathbf{1}_{\alpha_i > 0}}$ is the *equivalent multilinear monomial with respect to* $\{0,1\}^n$. We similarly define the *multilinear polynomial equivalent to polynomial* p *with respect to* $\{0,1\}^n$, as the polynomial $\tilde{p}$ obtained by changing each monomial of p to its equivalent multilinear monomial.

Corollary 2.8. For any polynomial $p \in \mathbb{R}[x]$, there exists a multilinear polynomial $\hat{p}$ and polynomial q in the ideal generated by $x_i^2 - 1$ such that $p = \hat{p} + q$, and this decomposition is unique.

More explicitly, there exist $g_i \in \mathbb{R}[\mathbf{x}]$, $i = 1, \dots, n$ such that $p = \tilde{p} + g_1(x_1^2 - 1) + \dots + g_n(x_n^2 - 1)$, where $\tilde{p}$ is the equivalent polynomial of p with respect to $\{0,1\}^n$.

Proof. The proof is identical to that of Corollary 2.5. □

2.2 Relationship between SOS polynomials, semidefinite matrices and $\mathbb{R}[\mathbf{x}]$.

Definition 2.9. Let $D \geq 1$ be a natural number. Let $N_D := \{(\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}_+^n : \alpha_1 + \ldots + \alpha_n \leq D\}$. This set corresponds to the possible exponents of all monomials of degree at most D. Note that $|N_D| = \binom{n+D}{D}$. Let X_D denote the column vector of monomials of degree at most D, so X_D is a vector of length N_D. For a polynomial $p \in \mathbb{R}[\mathbf{x}]_D$, i.e., $p = \sum_{\alpha \in N_D} p_\alpha \mathbf{x}^\alpha$, we denote by $C(p)$ the column vector of its coefficients that is $C(p)_\alpha = p_\alpha$. Then we can write

$$p = C(p)^T X_D.$$

Proposition 2.10. Let d be an even natural number. For any polynomial $p \in \mathbb{R}[\mathbf{x}]_d$, i.e., $p = \sum_{\alpha \in N_d} p_\alpha \mathbf{x}^\alpha$, there exists a symmetric matrix $Q \in \mathbb{R}^{|N_{d/2}| \times |N_{d/2}|}$ such that

$$p_\alpha = \sum_{\beta, \gamma \in \mathbb{Z}_+^n : \beta + \gamma = \alpha} Q_{\beta\gamma} \quad \forall \alpha \in N_d,$$

and therefore,

$$p = \sum_{\alpha \in N_d} p_\alpha \mathbf{x}^\alpha = C(p)^T X_d = (X_{d/2})^T Q X_{d/2}.$$

Proof. One can take $Q_{\beta\gamma} = \frac{p_\alpha}{|\{(\beta,\gamma) \in \mathbb{Z}_+^n \times \mathbb{Z}_+^n : \beta + \gamma = \alpha\}|}$. □

The following theorem makes the connection between SOS polynomials and PSD matrices.

Theorem 2.11. [SOS polynomials and PSD matrices] Let d be an even natural number. Let $p \in \mathbb{R}[\mathbf{x}]_d$, i.e., $p = \sum_{\alpha \in N_d} p_\alpha \mathbf{x}^\alpha$. Then $p \in SOS_d$ if and only if there exists a PSD matrix $Q \in \mathbb{R}^{|N_{d/2}| \times |N_{d/2}|}$ such that

$$p_\alpha = \sum_{\beta, \gamma \in \mathbb{Z}_+^n : \beta + \gamma = \alpha} Q_{\beta\gamma} \quad \forall \alpha \in N_d$$

Proof. We observe that there exist polynomials $q_1, \ldots, q_k$ such that $p = \sum_{i=1}^k q_i^2$ if and only if there exist coefficient vectors $C(q_i) \in \mathbb{R}^{N_{d/2}}$ such that

$$p = C(p)^T X_d = \sum_i (C(q_i)^T X_{d/2})^2 = \sum_i X_{d/2}^T C(q_i) \cdot C(q_i)^T X_{d/2} = X_{d/2}^T (\sum_i C(q_i) \cdot C(q_i)^T) X_{d/2},$$

which happens if and only if there exists a PSD matrix $Q \in \mathbb{R}^{|N_{d/2}| \times |N_{d/2}|}$ such that

$$p_\alpha = C(p)_\alpha = \sum_{\beta, \gamma \in \mathbb{Z}_+^n : \beta + \gamma = \alpha} Q_{\beta\gamma}.$$

□

Lemma 2.12. For any natural number D, the span of the set SOS_D is equal to $\mathbb{R}[\mathbf{x}]_D$. More concretely, for any $p \in \mathbb{R}[\mathbf{x}]_D$, there exists two SOS polynomials $u, v \in \mathbb{R}[x]_D$ such that $p = u - v$.

Proof. Using Proposition 2.10, we find a symmetric matrix Q with the property that

$$p = \sum_{\alpha \in N_d} p_\alpha \mathbf{x}^\alpha = C(p)^T X_d = (X_{d/2})^T Q X_{d/2}.$$

Doing an eigen-decomposition of Q, and the coleecting the positive eigenvalues into one group and the negative eigenvalues into another group, and using Theorem 2.11, we obtain the desired u, v. □

3 The relationships between the different problems introduced in Section 1

We now show the connections between the sets $\tilde{L}_t$, $t \geq 1$, and the optimization problems (1.5), (1.6) and (1.7). We first establish some structural results.

Definition 3.1. Given any $d \in \mathbb{N}$, any L linear on $\mathbb{R}[\mathbf{x}]$ and any polynomial $q \in \mathbb{R}[\mathbf{x}]$, the *moment matrix* $M(L,q,d)$ *of order* d *associated with* L *and* q is a $\binom{n+d-1}{d} \times \binom{n+d-1}{d}$ matrix whose rows and columns are indexed by $\alpha, \beta \in \mathbb{Z}_+^n$ such that $|\alpha|, |\beta| \leq d$, and $M(L,q,d)_{\alpha,\beta} := L(q(\mathbf{x})\mathbf{x}^\alpha \mathbf{x}^\beta) = L(q(\mathbf{x})\mathbf{x}^{(\alpha+\beta)})$. If q is the constant polynomial 1, then we use the simpler notation $M(L,d)$.

Lemma 3.2. Let d be an even natural number d, $q \in \mathbb{R}[\mathbf{x}]$ any polynomial and $L : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ be any linear map. The condition $L(qu) \geq 0$ for all $u \in SOS_d$ is equivalent to the condition that the matrix $M(L,q,d/2)$ from Definition 3.1 is positive semidefinite. In particular, the condition $L(u) \geq 0$ for all $u \in SOS_d$ is equivalent to the condition that the matrix $M(L,d/2)$ is positive semidefinite.

Proof. $L(qu) \geq 0$ for all $u \in SOS_d$ is equivalent to saying that $L(qp^2) \geq 0$ for all polynomials p of degree at most $d/2$. Polynomials p of degree at most $d/2$ can be written as $C(p)^T X_{d/2}$ – see Definition 2.9. Conversely, any vector $\mathbf{a} \in \mathbb{R}^{N_{d/2}}$ gives a polynomial $p = \mathbf{a}^T X_{d/2} = \sum_{|\alpha| \leq d/2} \mathbf{a}_\alpha \mathbf{x}^\alpha$. It can now be checked that

$$L(q(\mathbf{x})(\sum_{|\alpha| \leq d/2} \mathbf{a}_\alpha \mathbf{x}^\alpha)^2) = L(q(\mathbf{x})\mathbf{a}^T X_{d/2} X_{d/2}^T \mathbf{a}) = L(\mathbf{a}^T q(\mathbf{x}) X_{d/2} X_{d/2}^T \mathbf{a}) = \mathbf{a}^T M(L,q,d/2)\mathbf{a},$$

Thus, the condition that $L(qp^2) \geq 0$ for all polynomials p of degree at most $d/2$ is equivalent to the condition that $\mathbf{a}^T M(L,q,d/2)\mathbf{a} \geq 0$ for all $\mathbf{a} \in \mathbb{R}^{N_{d/2}}$, which is equivalent to the positive semidefiniteness of $M(L,q,d/2)$. Setting $q = 1$ gives the special case stated. ☐

3.1 Equivalence of $\tilde{L}_t$ and the pseudo-expectation problem (1.6)

Theorem 3.3. Consider problem (1.3) such that $x_j^2 - x_j \geq 0$ and $x_j - x_j^2 \geq 0$ for each $j = 1, \ldots, n$ are both amongst the constraints $q_i(\mathbf{x})$, and the rest of the constraints are of degree at most 2, i.e., quadratic and the objective p is linear. Fix some $t \geq 1$.

Let $L : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ be any linear map satisfying the constraints of the degree-$2t$ Lasserre relaxation as defined in (1.6) (see Remark 1.2) and define the vector $z := (L(\mathbf{x}^I))_{I \subseteq \{1,\ldots,n\}} \in \mathbb{R}^{2^n}$, where $\mathbf{x}^I$ denotes the multilinear polynomial involving the variables indexed by I. Then $z \in \tilde{L}_t$ and $L(p) = \sum_{I \subseteq \{1,\ldots,n\}} z_{\{i\}} p_i$, where p_i are the coefficients of the linear objective.

Conversely, let $y \in \tilde{L}_t \subseteq \mathbb{R}^{2^n}$. Define $L_y : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ as $L_y(\mathbf{x}^I) = y_I$, where $I \subseteq \{1,\ldots,n\}$ and for all other monomials m, define $L_y(m) := L(\tilde{m})$ where $\tilde{m}$ is the equivalent monomial of m with respect to $\{0,1\}^n$ (see Definition 2.7). Then L_y satisfies the degree-$2t$ Lasserre relaxation as defined in (1.6).

Proof. Let us start with a linear map $L : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ satisfying the constraints of the degree-$2t$ Lasserre relaxation as defined in (1.6).

We will first show that $L(q) = 0$ for all $q \in \ker(E^\dagger)$. Observe that by Remark 1.2, we may assume $L(m) = 0$ for any monomial with degree $2t+2$ or more. Hence, we consider an arbitrary $q \in \ker(E^\dagger)$ with degree of q at most $2t+2$. Since $x_j^2 - x_j \geq 0$ and $x_j - x_j^2 \geq 0$ for each $j = 1, \ldots, n$ are amongst the constraints, by setting $u = 1$, we obtain from the last set of constraints in (1.6) that $L(x_j^2 - x_j) = 0$ for all $j = 1, \ldots, n$. Using Lemma 2.3 and Lemma 2.12,we can express $q = (\sum_j u_j(x_j^2 - x_j)) - (\sum_j v_j(x_j^2 - x_j))$ for some polynomials $u_j, v_j \in SOS_{2t}$, $j = 1, \ldots, n$. Therefore, by linearity of L, $L(q) = \sum_j L(u_j(x_j^2 - x_j)) - \sum_j L(v_j(x_j^2 - x_j)) = 0$ because of the last set of constraints in (1.6).

Using Lemma 3.2, we know that the matrices $M(L,t)$ and $M(L,q_i,t)$ are all PSD matrices. Since $L(q) = 0$ for all $q \in \ker(E^\dagger)$, for any monomial m, $L(m) = L(\tilde{m})$, where $\tilde{m}$ is the equivalent monomial of m with respect to $\{0,1\}^n$ (see Definition 2.7). This implies that

$$M(L,t) = \begin{bmatrix} M_t(z) & M_t(z) & \cdots \\ M_t(z) & M_t(z) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \qquad M(L,q_i,t) = \begin{bmatrix} M_t(P_{q_i}(z)) & M_t(P_{q_i}(z)) & \cdots \\ M_t(P_{q_i}(z)) & M_t(P_{q_i}(z)) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

where z is the vector defined in the statement of this theorem, and $M_t(z), M_t(P_{q_i}(z))$ are the matrices defined above (1.2) (and also used in (1.2)). Because of this structure, $M(L,t), M(L,q_i,t)$ are PSD if and only if $M_t(z), M_t(P_{q_i}(z))$ are PSD.

Finally, $L(\mathbf{1}) = 1$ implies that $z_\emptyset = 1$. This completes the first part of the theorem.

The converse follows using essentially the same argument in reverse. □

3.2 Connecting with the SOS relaxation (1.5)

By the argument following (1.3), we know that $\gamma_d \leq p^\star$, i.e., (1.5) is a relaxation of problem (1.3). From weak duality of conic optimization, it is also clear that $\gamma_d \leq \rho_d$ for all even natural numbers d. But what about ρ_d and $p^\star$? Is there a relationship between them? This is clarified by the following theorem.

Proposition 3.4. For all even natural numbers $d \in \mathbb{N}$, $\gamma_d \leq \rho_d \leq p^\star$.

Proof. The first inequality is just stating weak duality of conic optimization, so it suffices to show the second inequality. For this, we observe that if ν is any probability measure defined on K, then $p^\star \leq \int p d\nu$. On the other hand, for any $\mathbf{x} \in K$ if we consider the Dirac delta measure $\delta_{\mathbf{x}}$, then we obtain that $p(\mathbf{x}) = \int p d\delta_{\mathbf{x}}$. Thus, we obtain that

$$\begin{array}{rll} p^\star = & \inf_\nu & \int p d\nu \\ & \text{s.t.} & \nu \text{ is a probability measure on } K \end{array} \tag{3.1}$$

We finally observe that if ν is a probability measure defined on K, then one can associate a natural linear functional $L_\nu : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ with it, defined by $L_\nu(p) = \int p d\nu$. Moreover, since $u \geq 0$ for all $u \in SOS_d$, $L_\nu(u) = \int u d\nu \geq 0$. Also, since $q_i(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in K$, we similarly have $L_\nu(uq_i) = \int uq_i d\nu \geq 0$. Of course, $L(\mathbf{1}) = \int d\nu = 1$ because ν is a probability measure. Thus, for all probability measures ν, L_ν satisfies the constraints of (1.6). Therefore, ρ_d is less than or equal to the right hand side of (3.1) and we obtain that $\rho_d \leq p^\star$. □

This justifies the use of "relaxation" in (1.6). In light of Theorem 3.3, this also shows that optimizing a linear function over the t-th level of the Lasserre hierarchy is indeed a lower bound on optimizing the same linear function over the set defined in (1.1). In fact, it is not hard to see that something stronger holds: the set defined in (1.1) is contained in every $\tilde{L}_t$, $t \geq 1$.

3.3 Equivalence of degree-d Lasserre maps (pseudo-expectations) and degree-d pseudo-distributions

Theorem 3.5. For any linear functional $L : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ (not necessarily a pseudo-expectation operator) such that $L(p) = 0$ for all $p \in \ker(E)$, there exists a function $D_L : \{-1,1\}^n \to \mathbb{R}$ (not necessarily a pseudo-distribution) such that

$$L(p) = \sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} D_L(\bar{\mathbf{x}}) p(\bar{\mathbf{x}})$$

for all polynomials $p \in \mathbb{R}[\mathbf{x}]$.

Conversely, given any function $D : \{-1,1\}^n \to \mathbb{R}$, defining the map $L_D(p) := \sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} D(\bar{\mathbf{x}}) p(\bar{\mathbf{x}})$ makes L_D a linear functional on $\mathbb{R}[\mathbf{x}]$ such that $L_D(p) = 0$ for all $p \in \ker(E)$.

Proof. For the first part of the theorem, we observe that since $L(p) = 0$ for all $p \in \ker(E)$, the linear map $L \circ E^{-1}$ is a well-defined linear map on $\mathbb{R}^{2^n}$. By the standard finite-dimensional Riesz representation theorem, there exists a vector $\mathbf{a} \in \mathbb{R}^{2^n}$ such that $L \circ E^{-1}(\mathbf{v}) = \langle \mathbf{a}, \mathbf{v} \rangle$ for all $\mathbf{v} \in \mathbb{R}^{2^n}$. Define $D_L(\bar{\mathbf{x}}) = \mathbf{a}_{\bar{\mathbf{x}}}$ for all $\bar{\mathbf{x}} \in \{-1,1\}^n$, where $\mathbf{a}_{\bar{\mathbf{x}}}$ is the coordinate of $\mathbf{a}$ corresponding to $\bar{\mathbf{x}}$. Thus for any $p \in \mathbb{R}[\mathbf{x}]$ with $\mathbf{v} = E(p)$, we have that

$$L(p) = L \circ E^{-1}(\mathbf{v}) = \langle \mathbf{a}, \mathbf{v} \rangle = \sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} \mathbf{a}_{\bar{\mathbf{x}}} \mathbf{v}_{\bar{\mathbf{x}}} = \sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} D_L(\bar{\mathbf{x}}) \mathbf{v}_{\bar{\mathbf{x}}} = \sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} D_L(\bar{\mathbf{x}}) p(\bar{\mathbf{x}}).$$

The second part of the theorem is simply the observation that the map $L_D(p) := \sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} D(\bar{\mathbf{x}}) p(\bar{\mathbf{x}})$ is linear and $L_D(p) = \sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} D(\bar{\mathbf{x}}) p(\bar{\mathbf{x}}) = 0$ for all $p \in \ker(E)$. ☐

Remark 3.6. Note that if $L : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ is a linear functional such that there exists a function $D : \{-1,1\}^n \to \mathbb{R}$ such that $L(p) = \sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} D(\bar{\mathbf{x}}) p(\bar{x})$ for all polynomials $p \in \mathbb{R}[\mathbf{x}]$, then it must satisfy the condition $L(p) = 0$ for all $p \in \ker(E)$ because for such a p, the RHS is always 0. **Thus, the kernel condition in Theorem 3.5 cannot be removed.**

Corollary 3.7. For any degree-d pseudo-expectation operator L that satisfies $L(p) = 0$ for all $p \in \ker(E)$, there exists a degree d pseudo-distribution D_L such that $L(p) = \sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} D_L(\bar{\mathbf{x}}) p(\bar{\mathbf{x}})$ for all polynomials $p \in \mathbb{R}[\mathbf{x}]$.

Conversely, for any degree d pseudo-distribution D, there exists a degree d pseudo-expectation operator such that $L_D(p) = \sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} D(\bar{\mathbf{x}}) p(\bar{\mathbf{x}})$.

Proof. One simply observes that $E(\mathbf{1}) = (1, 1, \ldots, 1)$ (the all one's vector). The rest follows from Theorem 3.5 and the definitions. ☐

An alternate proof of Corollary 3.7. Given a linear functional L such that $L(p) = 0$ for all $p \in \ker(E)$, we can think of this as a linear map on the quotient space $\mathbb{R}[\mathbf{x}]/\ker(E)$. We now consider the following system of 2^n linear equations in 2^n variables $D(\bar{\mathbf{x}})$, $\bar{\mathbf{x}} \in \{-1,1\}^n$:

$$L(m) = \sum_{\bar{\mathbf{x}} \in \{-1,1\}^n} D(\bar{\mathbf{x}}) m(\bar{\mathbf{x}}) \qquad \text{for all multilinear monomials } m.$$

By Lemma 2.2, part 1., the above system always has a solution $D(\bar{\mathbf{x}})$, $\bar{\mathbf{x}} \in \{-1,1\}^n$ since that constraint matrix is full rank. Using parts 2. and 3. of Lemma 2.2, we obtain another proof of Theorem 3.5. ☐

Theorem 3.8. Let $p \in \mathbb{R}[\mathbf{x}]$ and let d be a natural number such that $d + 2$ is at least degree of p. Then the degree-d pseudo-distribution relaxation (1.7) is equivalent to the degree-d Lasserre relaxation (1.6) with only the constraints $x_j^2 - 1 \geq 0$ and $1 - x_j^2 \geq 0$.

Proof. Using Remark 1.2 and similar argument as in the first part of the proof of Theorem 3.3, any feasible solution to the degree-d Lasserre relaxation (1.6) evaluates all polynomials in $\ker(E)$ to 0. Corollary 3.7 then completes the proof. ☐

Acknowledgment

The author is very grateful for long and detailed discussions with Mike Dinitz, Xin Li, Anirbit Mukherjee and Yasamin Nazari. Their insights and inputs were invaluable in putting together these notes.

References

[1] E.J. Anderson and P. Nash. *Linear Programming in Infinite-Dimensional Spaces: Theory and Applications*. Wiley, 1987.

[2] Nikhil Bansal and Daniel Dadush. https://homepages.cwi.nl/~dadush/teaching/sos-2017/.

[3] Boaz Barak and David Steurer. http://www.sumofsquares.org/public/index.html.

[4] A. Barvinok. *A course in convexity*. American Mathematical Society, 2002.

[5] Jean Lasserre. An explicit exact sdp relaxation for nonlinear 0-1 programs. *Integer programming and combinatorial optimization*, pages 293–303, 2001.

[6] Jean B Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[7] Jean Bernard Lasserre. *Moments, positive polynomials and their applications*. World Scientific, 2009.

[8] Monique Laurent. A comparison of the sherali-adams, lovász-schrijver, and lasserre relaxations for 0–1 programming. *Mathematics of Operations Research*, 28(3):470–496, 2003.

A Deriving the duality between (1.5) and (1.6)

We first introduce standard conic optimization duality. Our presentation is based on [1] and [4].

Definition A.1. The *algebraic dual space* of any vector space V is the set of all linear functionals $\ell : V \to \mathbb{R}$. The algebraic dual is another vector space and will be denoted by V'. [If $V = \mathbb{R}^n$, then it can be shown that V' is isomorphic to $\mathbb{R}^n$.]

Definition A.2. Let X, Y be two vector spaces, and let $A : X \to Y$ be any linear map. Then the *adjoint* of A, denoted by $A' : Y' \to X'$ is a linear map from the algebraic dual of Y to the algebraic dual of X defined as follows: $A'(\psi) \in X'$ is the linear functional defined by $A'(\psi)(x) = \psi(A(x))$ for all $x \in X$. [If $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$ and A is represented by an $m \times n$ matrix, then the adjoint A' is represented by the transpose of this matrix.]

Definition A.3. Let V be any vector space and let $C \subseteq Y$ be a convex cone in Y. The *algebraic dual cone* of C is

$$C' = \{\psi \in V' : \psi(y) \geq 0 \text{ for all } y \in C\}.$$

Elements of C' are called *positive linear functionals* on V.

Definition A.4. A general conic optimization problem is given by the following data: X, Y are two vector spaces, $A : X \to Y$ is a linear map, $C \subseteq X$ is a convex cone in X, $b \in Y$ is any vector in Y, and $\phi : X \to \mathbb{R}$ is any linear functional. The optimization problem is then the following:

$$\begin{array}{rrcl} \sup_{x \in X} & \phi(x) & & \\ \text{s.t.} & A(x) & = & b \\ & x & \in & C. \end{array} \tag{ConLP}$$

The algebraic dual optimization problem is defined as:

$$\begin{array}{rl} \inf_{\psi \in Y'} & \psi(b) \\ \text{s.t.} & A'(\psi) - \phi \in C' \end{array} \tag{ConDLP}$$

The relation (ConDLP) $\leq$ (ConLP) is known as *weak duality* and follows from the definitions. If (ConDLP) $\leq$ (ConLP) holds then we say one has *zero duality gap*, which need not always hold.

We now show that (1.5) is a special case of (ConLP). Using the notation introduced for (1.5), this is done by setting

$$
\begin{aligned}
&X = \underbrace{\mathbb{R}[\mathbf{x}]_d \times \mathbb{R}[\mathbf{x}]_d \times \ldots \times \mathbb{R}[\mathbf{x}]_d}_{m+1 \text{ times}} \times \mathbb{R} \\
&Y = \mathbb{R}[\mathbf{x}]_D \\
&A : X \to Y \text{ defined by } A(u_0, u_1, \ldots u_m, \lambda) = u_o + q_1 u_1 + q_2 u_2 + \ldots + q_m u_m + \lambda \\
&C = \underbrace{SOS_d \times SOS_d \times \ldots \times SOS_d}_{m+1 \text{ times}} \times \mathbb{R} \\
&b = p \\
&\phi : X \to \mathbb{R} \text{ defined by } \phi(u_0, u_1, \ldots u_m, \lambda) = \lambda
\end{aligned}
$$

To see that (ConDLP) gives (1.6), we observe the following:

1. Y' is the set of linear functionals from $\mathbb{R}[\mathbf{x}]_D$ to $\mathbb{R}$, which is the space over which we are optimizing in both (ConDLP) and (1.6).

2. The dual cone C' is the set of all linear maps that map $(u_0, u_1, \ldots, u_m, \lambda) \in \underbrace{SOS_d \times SOS_d \times \ldots \times SOS_d}_{m+1 \text{ times}} \times \mathbb{R}$ to nonnegative values.

3. Given any $L \in Y'$ (i.e., a linear map from $\mathbb{R}[\mathbf{x}]_D$ to $\mathbb{R}$), $A'(L)$ is the linear functional mapping $X \to \mathbb{R}$ as $A'(L)(u_0, u_1, \ldots u_m, \lambda) = L(A(u_0, u_1, \ldots u_m, \lambda)) = L(u_o + q_1 u_1 + q_2 u_2 + \ldots + q_m u_m + \lambda)$. Therefore, $\eta_L : A'(L) - \phi$ is the linear map from X to $\mathbb{R}$ given by

$$\eta_L(u_0, u_1, \ldots u_m, \lambda) = L(u_o + q_1 u_1 + q_2 u_2 + \ldots + q_m u_m + \lambda) - \lambda.$$

4. Using points 2. and 3. above to write the constraint in (ConDLP), we see that the constraint is equivalent to searching over all linear maps $L : \mathbb{R}[\mathbf{x}]_D \to \mathbb{R}$ such that $L(u_o + q_1 u_1 + q_2 u_2 + \ldots + q_m u_m + \lambda) - \lambda \geq 0$ for all $(u_0, u_1, \ldots, u_m, \lambda) \in \underbrace{SOS_d \times SOS_d \times \ldots \times SOS_d}_{m+1 \text{ times}} \times \mathbb{R}$.

   Selecting $(u_0, u_1, \ldots, u_m, \lambda) = (u, 0, \ldots, 0, 0)$ for $u \in SOS_d$, we obtain that $L(u) \geq 0$ for all $u \in SOS_d$, which is the second constraint in (1.6).

   Selecting $(u_0, u_1, \ldots, u_m, \lambda) = (0, 0, \ldots, u, \ldots, 0, 0)$ for $u \in SOS_d$, where u appears in the i-th position for $i = 1, \ldots, m$, we obtain that $L(q_i u) \geq 0$ for all $u \in SOS_d$, which is the third constraint in (1.6).

   Selecting $(u_0, u_1, \ldots, u_m, \lambda) = (0, 0, \ldots, 0, \lambda)$ for $\lambda \in \mathbb{R}$, we obtain that $L(\lambda) - \lambda \geq 0$ for all $\lambda \in \mathbb{R}$. In other words, for all constant polynomials λ, $L(\lambda) \geq \lambda$ which implies that $L(\lambda) = \lambda$ (by replacing λ with $-\lambda$.) Since L is a linear map, this is equivalent to saying that $L(\mathbf{1}) = 1$, where $\mathbf{1}$ is the constant polynomial that evaluates to 1. This gives the first constraint in (1.6).