Advanced SDPs	**Lecture 7: March 23, 2017**
Lecturers: Nikhil Bansal and Daniel Dadush	*Scribe: Daniel Dadush*

In this lecture, we will try to understand the power of the Lasserre relaxation more deeply. In particular, we will prove a basic inequality for pseudo-expectation operators, i.e. Hölder's inequality, we will show that Lasserre solutions for unconstrained univariate systems essentially always correspond to measures, we will go through the basics of conditioning Lasserre solutions, and use it to prove convergence of Lasserre on the hypercube.

7.1 Hölder's Inequality

We now state one of the most powerful and basic inequalities for pseudo-expectation operators.

Lemma 1 (Hölder's Inequality)**.** *Let* $L \in \mathrm{Las}_d(F, \emptyset)$ *where* $F = \{f_1, \dots, f_k\} \subset \mathbb{R}[\mathbf{x}]$. *Then for* $f \in F \cup \{1\}$ *and* $p, q \in \mathbb{R}[\mathbf{x}]$ *such that* $\deg(f) + 2\max\{\deg(p), \deg(q)\} \leq d$, *we have that*

$$|L[f(x)p(x)q(x)]| \leq L[f(x)p(x)^2]^{1/2} L[f(x)q(x)^2]^{1/2} \ . \tag{7.1}$$

Proof. Note that by the degree restrictions of p, q and since $L \in \mathrm{Las}_d(F, \emptyset)$, all the above expressions of L are well-defined. In particular, both $L[f(x)p(x)^2], L[f(x)q(x)^2] \geq 0$, so their square roots are well-defined. By possibly replacing p by $-p$, we may without loss of generality assume that $|L[f(x)p(x)q(x)]| = L[f(x)p(x)q(x)] \geq 0$.

Since $L \in \mathrm{Las}_d(F, \emptyset)$, for any $c > 0$, we have that

$$0 \leq L[f(x)(cp(x) - q(x)/c)^2] = c^2 L[f(x)p(x)^2] - 2L[f(x)p(x)q(x)] + L[f(x)q(x)^2]/c^2 \Leftrightarrow$$

$$L[f(x)p(x)q(x)] \leq \frac{1}{2}\left(c^2 L[f(x)p(x)^2] + L[f(x)q(x)^2]/c^2\right) \ .$$

If $L[f(x)p(x)^2] = 0$, by letting $c \to \infty$, the right hand side tends to 0, and hence $L[f(x)p(x)q(x)] = 0$. Similarly, if $L[f(x)q(x)^2] = 0$, we get the same conclusion letting $c \to 0^+$. Hence, inequality (7.1) holds in both these cases.

From here, we may assume that both $L[f(x)p(x)^2], L[f(x)q(x)^2] > 0$. The desired inequality now follows by setting $c = (L[f(x)q(x)^2]/L[f(x)p(x)^2])^{\frac{1}{4}} > 0$. $\square$

7.2 The Hamburger Moment Problem

In this section, we show that in the unconstrained univariate case, Lasserre solutions generally correspond to the expectation operator of a distribution supported on the real line. This is known as the *truncated* Hamburger moment problem. Namely, given a list of moments $m_1, m_2, \ldots, m_{2d}$, we wish to know when is there a distribution μ supported on $\mathbb{R}$ such that $\mathbb{E}_\mu[x^i] = m_i$, $\forall i \in [2d]$.

From last class, we already know that the associated pseudo expectation operator $L : \mathbb{R}[x]_{2d} \to \mathbb{R}$ given by $L[p(x)] = p_0 + \sum_{i=1}^{2d} p_i m_i$ should satisfy $L[q(x)^2] \geq 0$ for all $q \in \mathbb{R}[x]_d$. What is perhaps surprising is that is essentially also sufficient:

Theorem 1. *Let $L \in \mathrm{Las}_{1,2d}$ be an unconstrained pseudo-expectation operator on the real line of degree $2d$. Then there exists a distribution μ over $\mathbb{R}$ such that $\mathbb{E}_\mu[p(x)] = L[p(x)]$, $\forall p \in \mathbb{R}[x]_{2d}$ iff for all $q \in \mathbb{R}[x]_{d-1}$, $L[q(x)^2] = 0$ implies that $L[x^{d+1} q(x)] = 0$.*

To prove this theorem, we will to make use of the following version of the spectral theorem that decomposes linear operators that are symmetric with respect to a general (possibly degenerate) inner product over a finite dimensional real vector space. Recall that over a real vector space V, an inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is a symmetric bilinear form satisfying $\langle y, y \rangle \geq 0$ for all $y \in V$. The inner product is non-degenerate if $\langle y, y \rangle \geq 0$ iff $y = 0$, however we will not require this. We denote the null space of the inner product by $N_{\langle \cdot, \cdot \rangle} = \{y \in V : \langle y, y \rangle = 0\}$. Using the Cauchy-Schwarz inequality, i.e. $|\langle v, w \rangle| \leq \langle v, v \rangle^{1/2} \langle w, w \rangle^{1/2}$, it is not hard to check that N is in fact a linear subspace. We denote the range of the inner product by $R_{\langle \cdot, \cdot \rangle} = \left\{ w \in V : \langle w, y \rangle = 0 \forall y \in N_{\langle \cdot, \cdot \rangle} \right\}$, i.e. the orthogonal complement of the null space. It is easy to see that the range always admits an orthogonal normal basis $b_1, \ldots, b_k \in R_{\langle \cdot, \cdot \rangle}$ (i.e. apply Gram schmidt orthogonalization to any basis of the range, and note that you'll never find length 0 vectors there), and that for any $y \in V$, $\langle y, y \rangle = \sum_{i=1}^k \langle b_i, y \rangle^2$ (the part of y in the null space doesn't contribute any length).

Theorem 2 (Spectral Theorem)**.** *Let V be a d dimensional real vector space and let $\langle \cdot, \cdot \rangle$ be an inner product over V with l-dimensional range $R \subseteq V$. Let $T : V \to V$ be a linear map that is symmetric with respect to the inner product, i.e. $\langle Tv, w \rangle = \langle v, Tw \rangle$ for all $v, w \in V$. Then there exists real numbers $\lambda_1, \ldots, \lambda_k$ and an orthonormal basis $z_1, \ldots, z_l$ of R, such that for all $v, w \in V$ and $k \geq 0$:*

$$\langle v, T^k w \rangle = \sum_{i=1}^{l} \lambda_i^k \langle v, z_i \rangle \langle w, z_i \rangle \ .$$

Proof of Theorem 1. We first prove necessity. Assume that L is consistent with the expectation operator of a probability measure μ supported on $\mathbb{R}$. Then, if $\mathbb{E}_\mu[p(x)^2] = 0$ for $p \in \mathbb{R}[x]_{d-1}$, by Hölder's inequality we know that

$$\mathbb{E}_\mu[p(x)x^{d+1}] \leq \mathbb{E}_\mu[p(x)^2]^{1/2}\mathbb{E}_\mu[x^{2d+2}]^{\frac{1}{2}} = 0 \ .$$

Thus, by consistency, $L[p(x)x^{d+1}] = \mathbb{E}_\mu[p(x)x^{d+1}] = 0$.

Now assume that the above condition holds, we will show that L is indeed consistent with a probability measure μ. Let $V = \mathbb{R}[x]_d$ be our real vector space and define the inner product $\langle p, q\rangle = L[pq]$ for $p, q \in \mathbb{R}[x]_d$. Note that this is clearly bilinear and symmetric and that $\langle p, p\rangle = L[p^2] \geq 0$ for all $p \in \mathbb{R}[x]_d$ since $L \in \mathrm{Las}_{1,2d}$.

We would now like to define a symmetric linear operator $T : \mathbb{R}[x]_d \to \mathbb{R}[x]_d$ given by relation $T(p) = xp$. Unfortunately, T is not defined on x^d since x^{d+1} does not exist in $\mathbb{R}[x]_d$. To define Tx^d we will use the conditions imposed by forcing T to be symmetric. In particular, $\langle Tx^d, x^i\rangle = \langle x^d, Tx^i\rangle = \langle x^d, x^{i+1}\rangle = L[x^{d+1+i}]$ for all $i \in \{0, \ldots, d-1\}$. Note that these conditions form a system of linear equations, i.e. we wish to solve for $p \in \mathbb{R}[x]_d$ such that

$$L[px^i] = L[x^{d+1+i}] \quad \forall i \in \{0, \ldots, d-1\} \, .$$

We claim that this system has a solution. Assume not, then by Farkas lemma there exists a combination $c_0, \ldots, c_{d-1} \in \mathbb{R}$ of the rows such that

$$L[(\sum_{i=0}^{d-1} c_i x^i)p] = 0 \quad \forall p \in \mathbb{R}[x]_d,$$

and

$$L[(\sum_{i=0}^{d-1} c_i x^i)x^{d+1}] = 1 \ .$$

Letting $q = \sum_{i=0}^{d-1} c_i x^i \in \mathbb{R}[x]_{d-1}$, note that the first condition implies that $L[q^2] = 0$ but $L[qx^{d+1}] \neq 0$, a contradiction to our initial assumption on L.

Thus, we now formally define T by its action on the basis $1, \ldots, x^d$, where $Tx^i = x^{i+1}$ for $i \in \{0, \ldots, d-1\}$, and Tx^d equals the solution p to the above system of equations. Note that for $i, j \in \{0, \ldots, d-1\}$, $\langle Tx^i, x^j\rangle = \langle x^i, Tx^j\rangle = L[x^{i+j+1}]$, for $i \in \{0, \ldots, d-1\}$, $\langle Tx^d, x^i\rangle = \langle p, x^i\rangle = \langle x^d, x^{i+1}\rangle = \langle x^d, Tx^i\rangle$ by our choice of p. Lastly $\langle Tx^d, x^d\rangle = \langle x^d, Tx^d\rangle$ by symmetry of the inner product. Since it suffices to check symmetry with respect to a basis, we see that T is indeed symmetric with respect to our chosen inner product.

Note that by symmetry, for any $k \leq 2d$, we have that

$$\langle 1, T^k 1 \rangle = \langle T^{\lceil k/2 \rceil} 1, T^{\lfloor k/2 \rfloor} 1 \rangle = \langle x^{\lceil k/2 \rceil}, x^{\lfloor k/2 \rfloor} \rangle = L[x^k] \ ,$$

since $\lceil k/2 \rceil \leq d$. Next, by the spectral theorem

$$L[x^k] = \langle 1, T^k 1 \rangle = \sum_{i=1}^{l} \lambda_i^k \langle z_i, 1 \rangle^2 \ ,$$

where $z_1, \ldots, z_l \in \mathbb{R}[x]_d$ form an orthonormal basis of the range of $\langle \cdot, \cdot \rangle$ and $\lambda_1, \ldots, \lambda_l \in \mathbb{R}$.

Let μ denote the probability measure with takes value λ_i w.p. $\langle z_i, 1 \rangle^2$, for all $i \in [l]$. We note that μ is indeed a probability measure since $\langle z_i, 1 \rangle^2 \geq 0$, $i \in [l]$, and $\sum_{i=1}^{l} \langle z_i, 1 \rangle^2 = \langle 1, 1 \rangle = L[1] = 1$. Given the above expression, we now clearly have that

$$L[x^k] = \mathbb{E}_\mu[x^k],$$

for all $k \in [2d]$, and hence L is consistent with the expectation operator of μ as needed. $\square$

7.3 Conditioning Lasserre Solutions

In the previous section, we saw an algebraic method for converting a one dimensional Lasserre solution into a probability measure. In the next sections, we will explore a different strategy for (generally approximately) recovering an underlying distribution via *conditioning*.

From the perspective of optimization, one of our fundamental goals will be to recover nearly optimal solutions from the Lasserre relaxation, i.e. we would like ways to "round" Lasserre solutions. As noted in the previous lecture, even if the Lasserre solution corresponds to the moments of a true distribution over optimal solutions, recovering a true optimal solution can still be challenging because the moments only contain "averaged" information about the solutions. The problem here is in fact completely analoguous to that of recovering an explicit description of the underlying distribution as described in the previous section.

Conditioning will give us a general technique for "pushing" the purported underlying distribution towards being supported on a single solution, after which recovering it becomes easy (or at least easier).

Let μ be a distribution supported on a set K. Assume we would like to condition μ on an event E. The main question is how does the expectation operator of the measure μ conditioned on E change compared to μ? This can be stated very easily in terms of the indicator function $1_E : K \to \{0,1\}$ of E, where $1_E(x) = 1$ if $x \in E \cap K$ and 0 otherwise. Precisely, given any function $f : K \to \mathbb{R}$ we have that

$$\mathbb{E}_\mu[f(x)|E] = \frac{\int_{K\cap E} f(x)\mathrm{d}\mu(x)}{\int_{K\cap E} 1\mathrm{d}\mu(x)} = \frac{\int_K f(x)1_E(x)\mathrm{d}\mu(x)}{\int_K 1_E(x)\mathrm{d}\mu(x)} = \frac{\mathbb{E}_\mu[f(x)1_E(x)]}{\mathbb{E}_\mu[1_E(x)]} \,. \tag{7.2}$$

In particular, the conditioned expectation operator can be written as a simple function of the original expectation operator. To make sense of this in the context of pseudo-expectation operators, we need only restrict our attention to indicator functions 1_E that can be expressed as *low degree polynomials* over the feasible region.

Such examples come easily in the context of the hypercube $\{0,1\}^n$. For example, the indicator the event $E = \{x \in \{0,1\}^n : x_1 = 1, x_2 = 0\}$ can be written as $x_1(1-x_2)$, a degree 2 polynomial. Note that it is crucial that we restrict to the domain of the indicator to points in the hypercube, since evaluating this function at say the point $(1,2,0,\ldots,0)$ yields the value $1(1-2) = -1$, which is non-sensical for an indicator function.

Thus, if we have a pseudo-expectation operator L on a system (F, G) and we have an event we wish to condition on whose indicator function is a low degree polynomial h, we may hope to define the operator L conditioned on h by the analoguous formula to (7.2), namely

$$L[ph]/L[h] \quad \text{for all } p \in \mathbb{R}[\mathbf{x}]$$

where the expression is defined. Note that nothing stops us from defining the conditional operator with respect to a polynomial h that does not correspond to an indicator. Namely, h need not be a $\{0,1\}$ over the feasible region. In this more general case, which will indeed be useful later in the course, it is more appropriate to think of h as a reweighting of the operator. Regardless, we must still understand when such a reweighting makes sense. In general, when reweighting a measure we should insure that the reweighting function is non-negative over the feasible region. Crucially, such a reweighting insures that the expectation with respect to any non-negative function over the feasible region remains non-negative. In the context of Lasserre, we will want the reweighted pseudo-expectation to preserve nonnegative expectations with respect to the "obviously" nonnegative polynomials, up to a small drop in degree. As is shown in the next lemma, a very simple case where such non-negativity is preserved is when the reweighting polynomial h is a sum of squares. We note that indicator functions over the feasible region are trivially squares (though not necessarily low degree), since the square of a zero one function is the function itself.

Lemma 2 (Reweighted Pseudo-Expectation)**.** *Let* $L \in \mathrm{Las}_d(F, G)$ *for* $F = \{f_1, \ldots, f_k\}$*,* $G = \{g_1, \ldots, g_l\} \subset \mathbb{R}[\mathbf{x}]$*. Take* $h \in \Sigma^2_{d_1}$*,* $d_1 \leq d$*, such that* $L[h] > 0$*. Then the reweighted pseudo-expectation operator* $L[\cdot\ ; h] : \mathbb{R}[\mathbf{x}]_{d-d_1} \to \mathbb{R}$*, defined by*

$$L[p\ ; h] = \frac{L[ph]}{L[h]} \quad \forall p \in \mathbb{R}[x]_{d-d_1}$$

satisfies $L[\cdot\ ; h] \in \mathrm{Las}_{d-d_1}(F, G)$*.*

In the above lemma, we use the notation $L[p\ ; h]$ instead of $L[p|h]$, since we will sometimes reweight with respect to h's that are not indicator funtions.

Proof of Lemma 2. Firstly, note that $L[p\ ; h]$ is well defined for $p \in \mathbb{R}_{d-d_1}[\mathbf{x}]$ since $\deg(ph) \leq d$ and $L[h] > 0$. Furthemore, clearly $L[1\ ; h] = L[h]/L[h] = 1$.

Next, for any $g \in G$, $q \in \mathbb{R}[x]$, $\deg(g) + \deg(q) \leq d - d_1$, we see that

$$L[gq\ ; h] = \frac{L[gqh]}{L[h]} = 0,$$

since $\deg(g) + \deg(qh) \leq \deg(g) + \deg(q) + \deg(h) \leq d$. Lastly, for $f \in F \cup \{1\}$, $q \in \mathbb{R}[x]$, $\deg(f) + 2\deg(q) \leq d - d_1$, we have that

$$L[fq^2\ ; h] = \frac{L[fq^2h]}{L[h]} \geq 0,$$

since $\deg(f) + \deg(q^2h) \leq \deg(f) + 2\deg(q) + \deg(h) \leq d$ and since $q^2h \in \Sigma^2$.

Thus, $L[\cdot\ ; h] \in \mathrm{Las}_{d-d_1}(F, G)$. $\square$

We remark that one can also build a theory of conditioning which allows one to condition on more general functions than simply sums of squares. In particular, one may wish to condition on functions in the truncated quadratic module $Q_{d_1}(F \cup \{1\})$. This is indeed possible, however in this case, to recover an analoguous statement to the lemma above, one must enforce that a Lasserre pseudo-expectation operator sends polynomials corresponding to products of the polynomials in F to non-negative numbers. While such a theory seems quite clean (though some additional layers of notation are needed), it has as of yet been poorly explored, and so we will not dwell on it during the rest of the course. The reader is however encouraged to find applications of this type of conditioning, as such results would be quite interesting.

Definition 1 (Restriction of Pseudo-expectation Operator). *As we lose degrees after conditioning, it will often be useful to restrict a pseudo-expectation operator $L : \mathbb{R}[\mathbf{x}]_d \to \mathbb{R}$ to its evaluation on lower degree polynomials.*

For this purpose, we shall use the notation $L_{\downarrow d_1}$ to denote the restriction of L to polynomials of degree at most $d_1 \leq d$.

7.4 Lasserre on the Hypercube

In this section, we will specialize the Lasserre relaxation a bit more in the context of the hypercube. In particular, we will factor in the effect of the so-called "vanishing ideal" of the hypercube, i.e. the ideal of polynomials which evaluate to 0 on the hypercube, more directly into the relaxation.

Let $H = \{x_i^2 - x_i : i \in [n]\}$ denote the defining polynomials for the hypercube. Define $\mathbb{R}[\mathbf{x}]_H = \mathbb{R}[\mathbf{x}]/I(H)$, i.e. the ring of polynomials quotiented by the ideal generated by H. Note that the main effect of H to allow us to replace any x_i^2 term by an x_i term. Any polynomial $p \in \mathbb{R}[\mathbf{x}]$ thus becomes equivalent under $I(H)$ to a polynomial where every monomial x^α has $\alpha \in \{0,1\}^n$. Precisely,

$$\sum_{\alpha \in \mathbb{N}^n} p_\alpha x^\beta \equiv \sum_{\alpha \in \{0,1\}^n} (\sum_{\beta \in \mathbb{N}^n} p_{\alpha+\beta}) x^\alpha \pmod{I(H)} .$$

The polynomials on the right hand side are known as **multilinear** polynomials. Thus, every polynomial on the hypercube is equivalent under $I(H)$ to a multilinear one. Since multilinear monomials have only $\{0,1\}$ degrees, one can associate each such monomial with a subset $\alpha \subseteq [n]$. Slightly abusing notation, for $\alpha \subseteq [n]$, we will write $x^\alpha := \prod_{i \in \alpha} x_i$. We note that throughout these notes and the course, when working over the hypercube, we shall often be implicitly be identifying a polynomial with its equivalence class in $\mathbb{R}[\mathbf{x}]_H$. We will endeavor to make this explicit whenever it can lead to confusion.

In the above paragraph, we saw how every polynomial over the hypercube is equivalent to a multilinear one. In fact, more generally, every real value function on the hypercube is *uniquely* expressible as a multilinear polynomial. This is proven in the following lemma:

Lemma 3. *The set of monomial $(x^\alpha : \alpha \subseteq [n])$ form a basis of the set of functions from $\{0,1\}^n$ to $\mathbb{R}$.*

Proof. We first show any function $f : \{0,1\}^n \to \mathbb{R}$ can be expressed as a multilinear polynomial. We claim that it suffices to show that for any $y \in \{0,1\}^n$, the indicator function $I_y : \{0,1\}^n \to \{0,1\}$, where $I_y(x) = 1$ if $x = y$ and 0 otherwise, that I_y is a multilinear polynomial. This follows since $f = \sum_{y\in\{0,1\}^n} f(y)I_y$, i.e. f is a linear combination of indicator functions. To show that I_y is a multilinear polynomial, note that for $x \in \{0,1\}^n$

$$I_y(x) = \prod_{i:y_i=1} x_i \cdot \prod_{i:y_i=0} (1 - x_i) \ ,$$

which is clearly multilinear in x.

To show that $(x^\alpha : \alpha \subseteq [n])$ form a basis of the set of real valued functions on the hypercube, we need to show that they are linearly independent as functions over the hypercube. Since they span this set of functions and because there are 2^n such monomials, this follows from the fact that the set of real valued functions on the hypercube is 2^n dimensional. $\square$

The above justifies that when thinking about functions on the hypercube, we might as well work with multilinear polynomials. However, it is important to treat these polynomials as elements of $\mathbb{R}[\mathbf{x}]_H$, since this encodes the rule by which we should multiply such polynomials. In particular, given $p, q \in \mathbb{R}[\mathbf{x}]_H$, with multilinear representatives $\sum_{\alpha\subseteq[n]} p_\alpha x^\alpha$ and $\sum_{\beta\subseteq[n]} q_\beta x^\beta$ respectively, then the multilinear representative of their product is

$$pq \equiv \sum_{\alpha,\beta\subseteq[n]} p_\alpha q_\beta x^{\alpha\cup\beta} \pmod{I(H)} \ .$$

Notice that we use $x^{\alpha\cup\beta}$ instead of $x^{\alpha+\beta}$ to model the effect of modding out by $I(H)$.

We now come to the standard notion of degree for polynomials in $\mathbb{R}[\mathbf{x}]_H$, which will be crucial for our specialization of Lasserre.

Definition 2 (Degree over the Hypercube)**.** *For $p \in \mathbb{R}[\mathbf{x}]_H$, we define the degree of p to be the degree of its multilinear representation. That is, if p is equivalent to $\sum_{\alpha\subseteq[n]} p_\alpha x^\alpha$ under $I(H)$, we define the degree* $\deg(p)$ *of p as* $\max\{|\alpha| : \alpha \subseteq [n], p_\alpha \neq 0\}$*. Here, it is easy to check that the multilinear representative of p is in fact the representative of minimum degree as a polynomial in $\mathbb{R}[\mathbf{x}]$. For $d \in \{0,\ldots,n\}$, we define $\mathbb{R}[\mathbf{x}]_{H,d}$ to be the polynomials in $\mathbb{R}[\mathbf{x}]_H$ of degree at most d.*

Note that unlike in the ring $\mathbb{R}[\mathbf{x}]$, degree in $\mathbb{R}[\mathbf{x}]_H$ is not **additive***. In particular, every polynomial in $\mathbb{R}[\mathbf{x}]_H$ has degree at most n, and hence multiplying two degree n polynomials yields a polynomial of degree* at most *n instead of* exactly *$2n$. Degree however remains* **subadditive***, namely* $\deg(pq) \leq \deg(p) + \deg(q)$ *for $p, q \in \mathbb{R}[\mathbf{x}]_H$.*

We now come to the definition of Lasserre over the hypercube.

Definition 3 (Hypercube Lasserre)**.** *For a polynomial system $F = \{f_1, \ldots, f_k\}, G = \{g_1, \ldots, g_l\} \subseteq \mathbb{R}[\mathbf{x}]_H$ and $d \geq 0$, we define the degree d Lasserre relaxation over the hypercube as*

$$\begin{aligned}\mathrm{Las}_d^H(F, G) = \big\{L : \quad & L : \mathbb{R}[\mathbf{x}]_{H,d} \to \mathbb{R} \text{ linear },\\ & L[1] = 1,\\ & L[fq^2] \geq 0, \quad \forall f \in F \cup \{1\}, q \in \mathbb{R}[\mathbf{x}]_H, \deg(f) + 2\deg(q) \leq d,\\ & L[gq] = 0, \quad \forall g \in G, q \in \mathbb{R}[\mathbf{x}]_H, \deg(g) + \deg(q) \leq d\big\} .\end{aligned} \tag{7.3}$$

Since L is linear and the polynomials of degree d in $\mathbb{R}[\mathbf{x}]_H$ are spanned by the monomials $\{x^\alpha : \alpha \subseteq [n], |\alpha| \leq d\}$, to define L we need only keep track of the moments $L[x^\alpha]$ with respect to these monomials. Thus, L can be associated with a vector of dimension $\sum_{i=0}^d \binom{n}{i} = n^{O(d)}$.

We now specialize the notion of degree bounded sums of squares, quadratic module and ideal to polynomial systems in $\mathbb{R}[\mathbf{x}]_H$. For $d \in \{0, \ldots, n\}$, we define the degree d sum of squares polynomials by

$$\Sigma^2_{H,d} = \mathrm{cone}(q^2 : q \in \mathbb{R}[\mathbf{x}]_H, 2\deg(q) \leq d) .$$

For $F = \{f_1, \ldots, f_k\} \subseteq \mathbb{R}[\mathbf{x}]_H$, the define the degree d quadratic module

$$Q_d(F) = \mathrm{cone}(fq^2 : q \in \mathbb{R}[\mathbf{x}]_H, f \in F, \deg(f) + 2\deg(q) \leq d) .$$

For $G = \{g_1, \ldots, g_l\} \subseteq \mathbb{R}[\mathbf{x}]_H$, we define the degree d ideal

$$I_d(G) = \mathrm{span}(gq : g \in G, q \in \mathbb{R}[\mathbf{x}]_H, \deg(g) + \deg(q) \leq d) .$$

As before, we let $\Sigma^2_H, Q(F), I(G)$ denote the same objects without degree constraints. We now rephrase the degree d Lasserre relaxation for (F, G) over the hypercube:

$$\begin{aligned} L \quad & : \mathbb{R}[\mathbf{x}]_{H,d} \to \mathbb{R} \text{ linear },\\ L[1] \quad & = 1,\\ L[v] \quad & \geq 0, \quad \forall v \in Q_d(F \cup \{1\})\\ L[w] \quad & = 0, \quad \forall w \in I_d(G) .\end{aligned} \tag{7.4}$$

Some quick comments are in order with these definitions, which may not be immediately clear. Firstly, note that the notion of degree within the definitions of $\Sigma^2_{H,d}, Q_d(F), I_d(G)$ are *pessimistic*, i.e. we ask for the sum of degrees to be small instead of the degree itself. The reason for this is that degree cancellation over $\mathbb{R}[\mathbf{x}]_H$

is very difficult to predict. For example, a low degree polynomial may in fact be the square of a very high degree polynomial, which is not easy to verify. For a simple example, note that while the polynomial x_1 is a square since $x_1 \equiv x_1^2$, it is a degree 2 square and NOT a degree 1 square. Similarly, $1 - x_1 \equiv (1 - x_1)^2$ is also a degree 2 square but NOT degree 1.

Another important fact, is that while over $\mathbb{R}^n$ not every non-negative polynomial is a sum of squares, over the hypercube, every *non-negative function* is in fact a degree $2n$ sum of squares.

Lemma 4. *Let $f : \{0,1\}^n \to \mathbb{R}_+$. Then there exists $p \in \Sigma^2_{H,2n}$ such that p agrees with f on $\{0,1\}^n$.*

Proof. Firstly, note that since f is non-negative, the pointwise square root $\sqrt{f}$ of f is well-defined. By lemma 3, there exists a multilinear polynomial q which agrees with $\sqrt{f}$ on $\{0,1\}^n$.

Since q has degree at most n, we have that $q^2 \in \Sigma^2_{H,2n}$. Since by construction q^2 agrees with f on $\{0,1\}^n$, the statement is proven. $\square$

To conclude, we state the pseudo-expectation reweighting lemma for the hypercube. Its proof is identical to that of Lemma 2, so we leave it to the reader.

Lemma 5 (Hypercube Reweighting Lemma)**.** *Let $L \in \mathrm{Las}^H_d(F,G)$ for $F = \{f_1,\dots,f_k\}$, $G = \{g_1,\dots,g_l\} \subset \mathbb{R}[\mathbf{x}]_H$. Take $h \in \Sigma^2_{H,d_1}$, $d_1 \le d$, such that $L[h] > 0$. Then the reweighted pseudo-expectation operator $L[\cdot\ ; h] : \mathbb{R}[\mathbf{x}]_{H,d-d_1} \to \mathbb{R}$, defined by*

$$L[p\ ; h] = \frac{L[ph]}{L[h]} \quad \forall p \in \mathbb{R}[x]_{H,d-d_1}$$

satisfies $L[\cdot\ ; h] \in \mathrm{Las}_{d-d_1}(F,G)$.

In particular, for $h = x^I(1_n - x)^J$, the indicator of the event

$$\{x \in \{0,1\}^n : x_i = 1, \forall i \in I, x_j = 0, \forall j \in J\},$$

if $L[h] > 0$, then $L[\cdot\ ; h] \in \mathrm{Las}_{d-d_1}(F,G)$ where $2|I \cup J| = d_1$.

As in the previous section, for a pseudo-expectation operator $L : \mathbb{R}[\mathbf{x}]_{H,d} \to \mathbb{R}$, we define $L_{\downarrow d_1}$ to be its restriction to the polynomials $\mathbb{R}[\mathbf{x}]_{H,d_1}$ of degree at most $d_1 \le d$.

7.5 Convergence of Lasserre on the Hypercube

In this section, we prove for any polynomial system (F, G) over the hypercube we get convergence for $O(n)$ levels of Lasserre. That is, pseudo-expectation operators of this level are consistent with a true distribution over feasible hypercube points.

For notational convenience, we will associate points in the hypercube with subsets of $[n]$. That is, for $S \subseteq [n]$, we will use the notation $1_S \in \{0,1\}^n$, where $(1_S)_i = 1$ if $i \in S$ and 0 otherwise.

The main theorem is stated below.

Theorem 3. *Let* $F = \{f_1, \dots, f_l\}$, $G = \{g_1, \dots, g_k\} \subseteq \mathbb{R}[\mathbf{x}]_H$. *Then for* $d \geq \max\{2n + \deg(f)\}$, $f \in F \cup \{1\}$, *for any* $L \in \mathrm{Las}_d^H(F, G)$ *there exists a probability distribution* μ *supported on*

$$\{x \in \{0,1\}^n : f(x) \geq 0, \forall f \in F, g(x) = 0, \forall g \in G\}$$

such that

$$L[p(x)] = \mathbb{E}_\mu[p(x)] \quad \forall p \in \mathbb{R}[\mathbf{x}]_H \ .$$

To prove the theorem we will need the following key lemma. It essentially states that conditioning a pseudo-expectation operator with respect to the indicator function of a point on the hypercube, corresponds to evaluating at that point.

Lemma 6. *For* $L \in \mathrm{Las}_n^H := \mathrm{Las}_n^H(\emptyset, \emptyset)$, $p \in \mathbb{R}[\mathbf{x}]$, *and* $S \subseteq [n]$, *we have that*

$$L[p(x)x^S(1_n - x)^{[n]\setminus S}] = p(1_S)L[x^S(1_n - x)^{[n]\setminus S}] \ .$$

Proof. Since $p \equiv \sum_{\alpha \subseteq [n]} p_\alpha x^\alpha$, by linearity of L it suffices to prove that above statement when $p = x^I$ for some $I \subseteq [n]$. Note that $x^I(1_S) = 1$ if $I \subseteq S$ and 0 otherwise.

Assume that $I \subseteq S$. Then clearly

$$\begin{aligned} x^I x^S (1_n - x)^{[n]\setminus S} &\equiv (\prod_{i \in I \cap S} x_i^2) x^{S \setminus I}(1_n - x)^{[n]\setminus S} \\ &\equiv x^S(1_n - x)^{[n]\setminus S} \pmod{I(H)} \ , \end{aligned}$$

and hence

$$L[x^I x^S (1_n - x)^{[n]\setminus S}] = 1 \cdot L[x^S (1_n - x)^{[n]\setminus S}] \ .$$

as needed. Now assume that $I \setminus S \neq \emptyset$. Pick $j \in I \setminus S$. Then

$$\begin{aligned} x^I x^S (1_n - x)^{[n]\setminus S} &\equiv x^{I\setminus\{j\}} x^S (1_n - x)^{[n]\setminus S \cup \{j\}} x_j (1 - x_j) \\ &\equiv x^{I\setminus\{j\}} x^S (1_n - x)^{[n]\setminus S \cup \{j\}} (x_j^2 - x_j) \equiv 0 \pmod{I(H)} \ , \end{aligned}$$

and hence

$$L[x^I x^S (1_n - x)^{[n]\setminus S}] = 0 \cdot L[x^S (1_n - x)^{[n]\setminus S}] = 0 \ ,$$

as needed. Note that we used the condition that $L \in \mathrm{Las}_n^H$ simply to make sure that L is defined on all polynomials in $\mathbb{R}[\mathbf{x}]_H$. □

Proof of Theorem 3. Take $L \in \mathrm{Las}_d^H(F, G)$. We now define the probability measure μ on $\{0,1\}^n$ to take on value $1_S \subseteq [n]$ with probability $L[x^S (1_n - x)^{[n]\setminus S}]$.

We first show that μ is well-defined, i.e. that is indeed a probability measure of $\{0,1\}^n$. For this purpose, we will use the formal identity

$$1 = \sum_{S \subseteq [n]} x^S (1_n - x)^{[n]\setminus S} \ . \tag{7.5}$$

To see this, note that as functions on the hypercube, both the left and right hand side are equal to the constant function 1. Since both the left and right side are multilinear polynomials and are equal as functions on the hypercube, by Lemma 3 their coefficients must also be equal. From this, we get that

$$\sum_{S \subseteq [n]} L[x^S (1_n - x)^{[n]\setminus S}] = L[1] = 1 \ .$$

It now remains to show that the purported probabilities are non-negative. Since $d \geq 2n$, for any $S \subseteq [n]$, we have that

$$L[x^S (1_n - x)^{[n]\setminus S}] = L[(x^S (1_n - x)^{[n]\setminus S})^2] \geq 0 \ ,$$

where the first equality holds since $x_i^2 \equiv x_i \pmod{I(H)}$ and $(1 - x_i)^2 \equiv 1 - x_i \pmod{I(H)}$. Thus, μ is probability distribution as claimed.

Next, we show that L is consistent with the expectation operator of μ. Take $p \in \mathbb{R}[\mathbf{x}]_H$. Then by identity (7.5) and Lemma 6 (note that $L \in \mathrm{Las}_n^H$ since $d \geq n$), we have that

$$L[p(x)] = \sum_{S \subseteq [n]} L[x^S (1_n - x)^{[n]\setminus S} p(x)] = \sum_{S \subseteq [n]} p(1_S) L[x^S (1_n - x)^{[n]\setminus S}] = \mathbb{E}_\mu[p(x)] \ ,$$

where the last equality is by construction of μ.

To conclude, we check that μ is supported only on the hypercube points that are feasible for the system (F, G). Namely, for any $S \subseteq [n]$ for which $L[x^S(1_n - x)^{[n]\setminus S}] > 0$, we have that $f(1_S) \geq 0$, $\forall f \in F$, and that $g(1_S) = 0$, $\forall g \in G$.

Since $d \geq n$, by Lemma 6, we have that

$$\begin{aligned} L[f(x)x^S(1_n - x)^{[n]\setminus S}] &= f(1_S)L[x^S(1_n - x)^{[n]\setminus S}], \quad \forall f \in F, \\ L[g(x)x^S(1_n - x)^{[n]\setminus S}] &= g(1_S)L[x^S(1_n - x)^{[n]\setminus S}], \quad \forall g \in G \ . \end{aligned} \tag{7.6}$$

Restricting our attention to $f \in F$, note that since $\deg(f) + 2n \leq d$, we have that

$$L[f(x)x^S(1_n - x)^{[n]\setminus S}] = L[f(x)(x^S(1_n - x)^{[n]\setminus S})^2] \geq 0 \ . \tag{7.7}$$

Combining (7.6), (7.7), we see that $L[x^S(1_n - s)^{[n]\setminus S}] > 0 \Rightarrow f(1_S) \geq 0, \forall f \in F$ as needed. Next for $g \in G$, since $\deg(g) + n \leq 2n \leq d$, we have that

$$L[g(x)x^S(1_n - x)^{[n]\setminus S}] = 0 \ . \tag{7.8}$$

Combining (7.6), (7.8), we see that $L[x^S(1_n - s)^{[n]\setminus S}] > 0 \Rightarrow g(1_S) = 0, \forall g \in G$. The theorem now follows. $\square$