CS 294 Sum of Squares **Fall 2018**

Lecture 4: Sum of Squares Semidefinite Program

Lecturer: Prasad Raghavendra Scribe: Antares Chen

In these notes, we discuss semidefinite programming duality and begin to characterize the strength of the SoS SDP relaxation. We begin by deriving the dual program for a standard SDP then discuss the broader context of conic programming. We then show certain conditions under which strong duality is met. Finally, we discuss the strength of the sum of squares SDP and compare it to the Sherali-Adams LP hierarchy.

4.1 Semidefinite Programming Duality

Recall that for $X, C, A_1, \ldots, A_m \in \mathbb{R}^{n \times n}$, the primal semidefinite program is defined as

$$\begin{aligned} \text{minimize} \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_1, X \rangle = b_1 \\ & \cdots \\ & \langle A_m, X \rangle = b_m \\ & X \succeq 0 \end{aligned} \tag{4.1}$$

Our goal is to derive the *dual program* to equation 4.1 (i.e. a program that computes the tightest lower-bound on the primal objective). As mentioned in previous lectures, it is sometimes convenient to treat equation 4.1 as an LP with infinite constraints to take the dual. Recall that $X \succeq 0$ if and only if $v^\top X v \geq 0$ for all $v \in \mathbb{R}^n$. We can rewrite the constraint $X \succeq 0$ as:

$$\begin{aligned} \text{minimize} \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_1, X \rangle = b_1 \\ & \cdots \\ & \langle A_m, X \rangle = b_m \\ & v^\top X v \geq 0 \qquad \forall v \in \mathbb{R}^n \end{aligned}$$

This is now a linear program in entries of X albeit with an infinite number of constraints. To take the dual of an LP, first apply a non-negative multiplier to each constraint.

$$y_i \ : \ \langle A_i, X \rangle y_i = b_i y_i \quad \forall i = 1, \ldots, m \qquad\qquad c_v \ : \ c_v v^\top X v \geq 0 \quad \forall v \in \mathbb{R}^n$$

As the multipliers are non-negative, the inequality directions from the original SDP are preserved. Summing over all inequalities derives

$$\sum_{i=1}^{n} \langle A_i, X \rangle y_i + \int_{v \in \mathbb{R}^n} c_v v^\top X v \, \mathrm{d}v \geq \sum_{i=1}^{n} b_i y_i$$

which further simplifies to

$$\left\langle A_i y_i + \int_v c_v v v^\top \, \mathrm{d}v, X \right\rangle \geq \sum_{i=1}^{n} b_i y_i$$

This is exactly the form that we want for computing the tightest lower-bound. If we constrain the left-hand side to be equal to C, then the right-hand side lower bounds the primal objective value. Finding the tightest lower-bound then amounts to maximizing the right-hand side of the inequality. The dual program is thus

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{n} b_i y_i \\ \text{subject to} \quad & \sum_{i=1}^{n} A_i y_i + \int_v c_v v v^\top \, \mathrm{d}v = C \end{aligned}$$

But this is in fact an SDP! Recall that a non-negative linear combination of matrices with form $vv^\top$ is PSD which means that we can introduce an additional constraint positing a positive semidefinite $Z \in \mathbb{R}^{n \times n}$ such that $Z = \int_v c_v v v^\top \, \mathrm{d}v$. The dual SDP to the primal SDP equation 4.1 is given by

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{n} b_i y_i \\ \text{subject to} \quad & \sum_{i=1}^{n} A_i y_i + Z = C \\ & Z \succeq 0 \end{aligned} \tag{4.2}$$

4.1.1 Conic Programming

It is possible to generalize semidefinite programming to a setting known as conic programming. Let us begin by defining a *convex cone* as a convex subset $\mathcal{S} \subseteq \mathbb{R}^n$ such that if $v \in \mathcal{S}$, then $\lambda v \in \mathcal{S}$ for any scaler $\lambda \geq 0$. Examples of convex cones include:

(1) Let $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i \geq 0 \; \forall i = 1, \ldots, n\}$ be the non-negative orthant, then $\mathbb{R}^n_+$ is a convex cone. Any non-negative multiplier λ preserves the sign of a vector x element wise. Hence $x \in \mathbb{R}^n_+$ if and only if $\lambda x \in \mathbb{R}^n_+$ for $\lambda \geq 0$.

(2) Let $\mathcal{S}^n = \{A \in \mathbb{R}^{n \times n} : A \succeq 0\}$, then $\mathcal{S}^n$ is a convex cone. Consider any $x \in \mathbb{R}^n$, $\lambda \geq 0$ and observe $x^\top(\lambda A)x = \lambda(x^\top A x)$. This implies that $\lambda A \succeq 0$ if and only if $A \succeq$.

A *conic program* is then a program which optimizes a linear objective with linear constraints over a convex cone. Specifically if $\mathcal{K}$ is a convex cone, then a conic program contains the constraint that $x \in \mathcal{K}$. Taking the convex cones from the examples above derives optimization settings that we have already encountered.

(1) Optimizing linear objectives with linear constraints over $\mathcal{K} = \mathbb{R}^n_+$ corresponds to linear programming.

(2) Choosing $\mathcal{K} = \mathcal{S}^n$ corresponds to semidefinite programming.

Taking the dual of a conic program replaces the conic constraint with $x \in \mathcal{K}^*$ where $\mathcal{K}^*$ is known as the dual cone of $\mathcal{K}$. The *dual cone* of a convex cone $\mathcal{K}$ is defined as the set

$$\mathcal{K}^* = \{v : \langle v, x \rangle \geq 0 \; \forall x \in \mathcal{K}\}$$

Geometrically, this is the set of all x which forms an acute angle with every point in $\mathcal{K}$. For example, the dual cone of $\mathcal{S}^n$ is itself. This follows immediately from the next fact.

Claim 4.1. *Let $A, X \in \mathbb{R}^{n \times n}$. Then $\langle A, X \rangle \geq 0$ for all $X \in \mathcal{S}_n$ if and only if $A \succeq 0$*

Proof. Suppose $\langle A, X \rangle \geq 0$ for all $X \in \mathcal{S}_n$ and consider any $x \in \mathbb{R}^n$. We have that $xx^\top \succeq 0$ thus $\langle A, xx^\top \rangle \geq 0$ or $x^\top A x \geq 0$ for any x.

Now suppose that $A \succeq 0$. For any $X \succeq 0$, X admits an eigendecomposition. This means

$$\langle A, X \rangle = \left\langle A, \sum_{i=1}^n \lambda_i v_i v_i^\top \right\rangle = \sum_{i=1}^n \lambda_i \langle A, v_i v_i^\top \rangle = \sum_{i=1}^n \lambda_i v_i^\top A v_i$$

Since $X \succeq 0$, we have that $\lambda_i \geq 0$ and because $A \succeq 0$, we have $v_i^\top A v_i \geq 0$. Consequently, $\langle A, X \rangle \geq 0$ satisfying the claim. □

In a similar vein, the dual cone for $\mathbb{R}^n_+$ is itself as well. Thus the dual programs LPs and SDPs are each LPs and SDPs respectively. This matches exactly what we have derived above.

4.1.2 Compactness of Feasible Set

There are certain issues with semidefinite programming that we need to take into account. For one, we can construct examples where the objective function does not achieve its optimal value within its feasible region.

$$\begin{aligned} \text{minimize} \quad & x_1 \\ \text{subject to} \quad & \begin{pmatrix} x_1 & 1 \\ 1 & x_2 \end{pmatrix} \succeq 0 \end{aligned}$$

This is equivalent to minimizing x_1 subject to $x_1, x_2 \geq 0$ and $x_1 x_2 \geq 1$. In particular, the solution

$$X_\epsilon = \begin{pmatrix} \epsilon & 1 \\ 1 & \frac{1}{\epsilon} \end{pmatrix}$$

is always feasible since $X_\epsilon \succeq 0$ for any $\epsilon > 0$, but $\epsilon = 0$ is not achieved. Cases like that above would force us to discuss the infimum / supremum of the feasible set. However,

Theorem 4.2. *If all feasible solutions are bounded by a PSD matrix, then the optimal value of the objective function is achieved.*

That is to say if the feasible set is bounded, then it is also compact and the minimum/maximum exist. Theorem 4.2 will often allow us to ignore this problem because feasible pseudo-expectations for problems we encounter will certainly be lower-bounded by 0 and often be upper-bounded by 1.

4.1.3 Conditions for Weak and Strong Duality

Another property is convenient to have is strong duality. Recall that an optimization problem is *primal-feasible* and *dual-feasible* if its primal and dual formulations have feasible solutions. The duality theorem can then be stated as

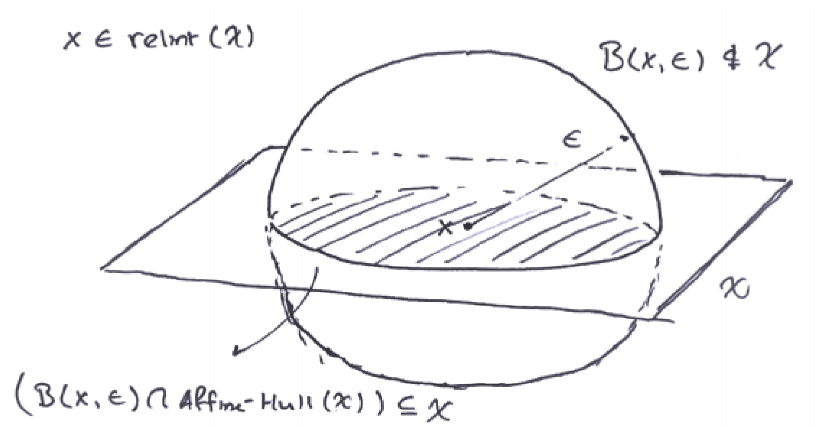
Theorem 4.3. *Suppose equation 4.1 is both primal and dual feasible. Let v_{primal} and v_{dual} denote the optimal primal and dual solution. The following statements hold*

(1) Weak duality: $v_{\text{primal}} \leq v_{\text{dual}}$

(2) Strong duality: If equation 4.1 is strictly feasible, then the dual optimal solution is achieved and $v_{\text{primal}} = v_{\text{dual}}$.

What does it mean for an optimization problem to be *strictly feasible*? Suppose that a program has a feasible set of points $\mathcal{X}$. We would like strictly feasible to inuitively imply the existence of a point that is “deep” inside $\mathcal{X}$.

It is not enough for us to say that a point $x \in \mathcal{X}$ is strictly feasible if some ball centered at x contained within $\mathcal{X}$ because the feasible set may have lower dimension than the ambient space. Consequently, a fully dimensional ball will never be a subset of $\mathcal{X}$ like that pictured below. For example, an LP optimizing over $\mathbb{R}^n$ constraining x to a hyperplane will never have an n-dimensional ball centered at x as a subset.



To fix this, we instead ask if the intersection of the ball and the affine hull of $\mathcal{X}$ is contained within $\mathcal{X}$, that is to say, if there exists a non-empty relative interior of $\mathcal{X}$. The *relative interior* of $\mathcal{X}$ is defined as:

$$\text{relint}(\mathcal{X}) = \{x \in \mathcal{X} : \exists \epsilon \geq 0 \text{ s.t. } \mathcal{B}(x, \epsilon) \cap \text{Affine-Hull}(\mathcal{X}) \subseteq \mathcal{X}\}$$

where $\mathcal{B}(x, \epsilon)$ denotes the ball centered at x with radius $\epsilon \geq 0$. The statement for strong duality is thus $v_{\text{primal}} = v_{\text{dual}}$ if $\text{relint}(\mathcal{X}) \neq \emptyset$. We will often assume strong duality because we can usually exhibit a point in $\text{relint}(\mathcal{X})$. Indeed for SoS SDPs over the boolean hypercube, we have the following theorem.

Theorem 4.4. *For a degree-d SoS SDP over $\{-1,+1\}^n$, there exists a point in the relative interior of the set containing all* $\tilde{\mathbf{E}}$ *such that*

$$\begin{aligned}
&\tilde{\mathbf{E}}\{(x_i^2-1)p(x)\}=0 && \forall \text{ polynomials } p(x): \deg((x_i^2-1)p(x))\le d\\
&\tilde{\mathbf{E}}\{q^2(x)\}\ge 0 && \forall \text{ polynomials } q(x): \deg(q^2(x))\le d\\
&\tilde{\mathbf{E}}\{1\}=1
\end{aligned}$$

Proof. sketch Consider the pseudo-expectation corresponding to the uniform distribution $\mathcal{U}$ over $\{-1,+1\}^n$

$$\tilde{\mathbf{E}}\{p(x)\}=\mathop{\mathbf{E}}_{x\sim\mathcal{U}}\{p(x)\}$$

and observe that for any monomial x^σ, we have that

$$\tilde{\mathbf{E}}\{x^\sigma\}=\mathop{\mathbf{E}}_{x\sim\mathcal{U}}\{x^\sigma\}=\begin{cases}1 & \text{if } x^\sigma \text{ is a perfect square}\\ 0 & \text{otherwise}\end{cases}$$

Let us now demonstrate that this pseudo-expectation is valid. Recalling that coordinates are sampled independently and the above fact, we have the following.

$$\tilde{\mathbf{E}}\{(x_i^2-1)p(x)\}=\mathop{\mathbf{E}}_{x\sim\mathcal{U}}\{x_i^2p(x)\}-\mathop{\mathbf{E}}_{x\sim\mathcal{U}}\{p(x)\}=\mathop{\mathbf{E}}_{x\sim\mathcal{U}}\{p(x)\}-\mathop{\mathbf{E}}_{x\sim\mathcal{U}}\{p(x)\}=0$$

Futhermore, $\tilde{\mathbf{E}}\{q^2(x)\}\ge 0$ since the expectation is 1 for squared monomials. Finally, $\tilde{\mathbf{E}}\{1\}=1$ readily holds as 1 is constant. Now to see why this pseudo-expectation is in the relative interior, consider $\tilde{\mathbf{E}}$'s PSD matrix realization. For multilinear monomials $x_{\mathcal{S}}=\prod_{i\in\mathcal{S}}x_i$ and $x_{\mathcal{T}}=\prod_{i\in\mathcal{T}}x_i$ where $\mathcal{S},\mathcal{T}\subseteq[n]$, we have that

$$\tilde{\mathbf{E}}\{x_{\mathcal{S}}x_{\mathcal{T}}\}=\begin{cases}1 & \text{if } \mathcal{S}=\mathcal{T}\\ 0 & \text{otherwise}\end{cases}$$

since $x_{\mathcal{S}}x_{\mathcal{T}}$ will always have an odd power if $\mathcal{S}\neq\mathcal{T}$ otherwise it is a perfect square. If we index the moment matrix by multilinear monomials, we have

$$\tilde{\mathbf{E}}=\begin{array}{c c}
 & \begin{array}{cccc} 1 & x_1 & & \prod_i x_i \end{array}\\
\begin{array}{c} 1\\ x_1\\ \\ \prod_i x_i \end{array} & \begin{pmatrix} 1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1 \end{pmatrix}
\end{array}=\mathrm{Id}$$

The identity matrix is strictly positive definite. Consequently, $\tilde{\mathbf{E}}$ is in the relative interior of the feasible set for a degree-d SoS SDP. $\square$

This proof is just a sketch. To complete the proof, we would formally construct the feasible set for a degree-d SoS SDP and argue that there is some ϵ for which the ball around the pseudo-expectation corresponding to the uniform distribution is contained inside the feasible region.

However, the mechanics that we worked through are indicative of a standard approach to claim strong duality. Regardless of whether or not we are working over the boolean hypercube, it remains a decent idea to check the pseudo-expectation corresponding to the uniform distribution as it often lies inside the relative interior of feasible sets corresponding to different SoS SDP relaxations.

4.2 Power of Sum of Squares

We now provide two theorems that characterize the power of the sum of squares relaxation hierarchy. The first theorem states that a sufficiently high degree pseudo-expectation can capture any true fact we wish to prove about a polynomial system.

Theorem 4.5. *Suppose that $\tilde{\mathbf{E}}$ is a degree-$2n$ pseudo-expectation over $\{-1,+1\}^n$ then*

(1) There exists a probability distribution μ over $\{-1,+1\}^n$ such that $\tilde{\mathbf{E}}\{p(x)\} = \mathop{\mathbf{E}}_{x\sim\mu}\{p(x)\}$.

(2) For all non-negative polynomials $f : \{-1,+1\}^n \to \mathbb{R}_{\geq 0}$, we have $\tilde{\mathbf{E}}\{f\} \geq 0$.

Proof. Let us begin with claim (1). Suppose $\tilde{\mathbf{E}}$ is a degree-$2n$ pseudo-expectation and define $\mu : \{-1,+1\} \to \mathbb{R}$ such that for $\alpha \in \{-1,+1\}^n$, we have

$$\mu(\alpha) = \tilde{\mathbf{E}}\big\{\mathbb{I}[x=\alpha]\big\} = \tilde{\mathbf{E}}\Big\{\prod_{i=1}^{n} \frac{1+\alpha_i x_i}{2}\Big\}$$

To demonstrate that μ is a valid distribution, we first show $\mu(\alpha) \geq 0$ for any α. Consider for a value of α, we have the identity:

$$\begin{aligned} \alpha_i = +1 \quad &: \quad \frac{1+x_i}{2} = \left(\frac{1+x_i}{2}\right)^2 + \frac{1-x_i^2}{4} = \left(\frac{1+\alpha_i x_i}{2}\right)^2 + \frac{1-x_i^2}{4} \\ \alpha_i = -1 \quad &: \quad \frac{1-x_i}{2} = \left(\frac{1-x_i}{2}\right)^2 + \frac{1-x_i^2}{4} = \left(\frac{1+\alpha_i x_i}{2}\right)^2 + \frac{1-x_i^2}{4} \end{aligned}$$

Thus $\mathbb{I}[x=\alpha]$ is equivalent to

$$\prod_{i=1}^{n} \frac{1+\alpha_i x_i}{2} = \prod_{i=1}^{n} \left(\frac{1+\alpha_i x_i}{2}\right)^2 + \sum_{i=1}^{n} (1-x_i^2) r_i(x)$$

for polynomials $r_1, \ldots, r_n$ where $\deg((1-x_i^2)r_i(x)) \leq 2n$. Recalling that $\tilde{\mathbf{E}}\{(x_i^2-1)p(x)\} = 0$ for polynomials p where $\deg((x_i^2-1)p(x)) \leq 2n$, we have

$$\tilde{\mathbf{E}}\Big\{\prod_{i=1}^{n} \frac{1+\alpha_i x_i}{2}\Big\} = \tilde{\mathbf{E}}\Big\{\prod_{i=1}^{n} \left(\frac{1+\alpha_i x_i}{2}\right)^2\Big\} + \tilde{\mathbf{E}}\Big\{\sum_{i=1}^{n} (1-x_i^2) r_i(x)\Big\} = \tilde{\mathbf{E}}\Big\{\prod_{i=1}^{n} \left(\frac{1+\alpha_i x_i}{2}\right)^2\Big\}$$

The pseudo-expectation of any squared polynomial is non-negative thus $\mu(\alpha) \geq 0$.

Next we show $\sum_\alpha \mu(\alpha) = 1$. Consider

$$\sum_{\alpha\in\{-1,+1\}^n} \mathbb{I}[x=\alpha] = \sum_{\alpha\in\{-1,+1\}^n} \prod_{i=1}^{n} \frac{1+\alpha_i x_i}{2} = \sum_{\alpha\in\{-1,+1\}^n} \frac{1}{2^n} = 1$$

This follows as terms with $\alpha_i x_i$ will cancel with each-other for appropriately chosen pairs of $\alpha \in \{-1,+1\}^n$. Consequently $\sum_\alpha \mu(\alpha) = \tilde{\mathbf{E}}\{1\} = 1$ implying μ is a valid distribution. Finally, consider any multilinear monomial $x_\mathcal{S}$ for $\mathcal{S} \subseteq [n]$ and define $\alpha_\mathcal{S} = \prod_{i\in\mathcal{S}} \alpha_i$. Observe that for any $\mathcal{S} \subseteq [n]$, the following holds.

$$x_\mathcal{S} = \sum_{\alpha\in\{-1,+1\}^n} \alpha_\mathcal{S} \cdot \left(\prod_{i=1}^{n} \frac{1+\alpha_i x_i}{2} \right)$$

Thus the pseudo-expectation of $x_\mathcal{S}$ is given by

$$\begin{aligned}
\tilde{\mathbf{E}}\{x_\mathcal{S}\} &= \tilde{\mathbf{E}}\Bigg\{ \sum_{\alpha\in\{-1,+1\}^n} \alpha_\mathcal{S} \cdot \left(\prod_{i=1}^{n} \frac{1+\alpha_i x_i}{2} \right) \Bigg\} \\
&= \sum_{\alpha\in\{-1,+1\}^n} \alpha_\mathcal{S} \cdot \tilde{\mathbf{E}}\Bigg\{ \prod_{i=1}^{n} \frac{1+\alpha_i x_i}{2} \Bigg\} \\
&= \sum_{\alpha\in\{-1,+1\}^n} \alpha_\mathcal{S} \mu(\alpha) \\
&= \mathop{\mathbf{E}}_{x\sim\mu} \{x_\mathcal{S}\}
\end{aligned}$$

Since any polynomial p can be under the basis of multilinear monomials, we have that claim (1) holds. Claim (2) now follows immediately as the expectation under μ of any non-negative polynomial is itself non-negative. $\square$

In some sense, we can construct this proof because $\{-1,+1\}^n$ has an impulse function that can be decomposed nicely into a finite polynomial (i.e. in this case, we decompose the impulse function using the Fourier basis). For more general spaces such as $\mathbb{R}^n$, this proof does not hold since such impulse functions do not exist (i.e. there does not exist a finite polynomial decomposition of the Diract delta function).

Finally, though theorem 4.5 holds for degree-$2n$ pseudo-expectation functions, it is possible to prove the same for a degree-n pseudo-expectations. This is because, in some sense, there does not exist moments of degree $\geq n$ over the boolean hypercube. It would suffice to show that for every polynomial of degree $\geq n$ on $\{-1,+1\}^n$, there exists a polynomial p' such that $\deg(p') \leq n$ and p' agrees with p on every point.

4.3 Sum of Squares and Sherali-Adams

Our second theorem states that sum of squares captures the Sherali-Adams LP hierarchy, another relaxation hierarchy that represents polynomial optimization problems as linear programs. When we derived the relaxation for max-cut, we chose to replace the distribution μ over possible solutions with its degree-2 moments as a way to decrease the size of our representation.

Instead of maintaining moments, the Sherali-Adams hierarchy can be derived by maintaining marginal distributions for μ and consistency constraints on marginals with intersecting support. For example, the derivation of the Sherali-Adams relaxation for max-cut would be the following.

$$\underset{x\in\{-1,+1\}^n}{\text{maximize}}\, G(x)$$

$$\begin{array}{lll} \text{maximize} & \underset{x\sim\mu}{\mathbf{E}}\{G(x)\} & \\ \text{subject to} & \mu(x) \geq 0 & \forall x \in \{+1,-1\}^n \\ & \displaystyle\sum_{x\in\{+1,-1\}^n} \mu(x) = 1 & \end{array}$$

$$\begin{array}{lll} \text{maximize} & \displaystyle\sum_{(i,j\in E)} \underset{x\sim\mu_{\{i,j\}}}{\mathbf{E}}\left\{(x_i - x_j)^2\right\} & \\ \text{subject to} & \mu_{\mathcal{S}} : \text{prob. distribution over } \{-1,+1\}^{|\mathcal{S}|} & \forall \mathcal{S} \subseteq [n] \text{ and } |\mathcal{S}| \leq d \\ & \mu_{\mathcal{S}}\big|_{\mathcal{S}\cap\mathcal{S}'} = \mu_{\mathcal{S}'}\big|_{\mathcal{S}\cap\mathcal{S}'} & \forall \mathcal{S},\mathcal{S}' \subseteq [n] \text{ and } |\mathcal{S}|, |\mathcal{S}'| \leq d \end{array}$$

The final program in this flow-chart is called the *d-round Sherali-Adams LP relaxation* where the last constraint states that two marginal distributions restricted to $\mathcal{S}, \mathcal{S}'$ should agree on variables in the intersection $\mathcal{S} \cap \mathcal{S}'$. Additionally, it is good to note that setting $d = n$ recovers the program in the second box.

The manner in which SoS captures Sherali-Adams is given by the following theorem.

Theorem 4.6. *Suppose there exists a degree-d pseudo-expectation function $\tilde{\mathbf{E}}$ over $\{-1,+1\}^n$ for some given d. Then for all $\mathcal{S} \subseteq [n]$ where $|\mathcal{S}| \leq d$, there exists a probability distribution $\mu_{\mathcal{S}} : \{-1,+1\}^{|\mathcal{S}|} \to \mathbb{R}^n$ such that for every polynomial $p(x_{\mathcal{S}}) \in \mathbb{R}[x_{\mathcal{S}}]$*

$$\tilde{\mathbf{E}}\{p(x_{\mathcal{S}})\} = \underset{x_{\mathcal{S}}\sim\mu_{\mathcal{S}}}{\mathbf{E}}\{p(x)\}$$

which states that a degree-d pseudo-expectation corresponds to a true expectation evaluated over a distribution restricted on $\leq d$ variables. This theorem follows immediately by applying theorem 4.5 to pseudo-expectation $\tilde{\mathbf{E}}$ restricted to $\mathbb{R}[x_{\mathcal{S}}]$ for given $\mathcal{S} \subseteq [n]$.