6.1 Notation

Let $\mathbb{N} = \{0, 1, \dots\}$ denote the set of non-negative integers. For $\alpha \in \mathbb{N}^n$, define the monomial $x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}$. For a polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$ with real coefficients on n variables – which we abbreviate by $\mathbb{R}[\mathbf{x}]$ – we express it as

$$p = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x^\alpha.$$

We define the support of p by $\text{supp}(p) = \{\alpha : \alpha \in \mathbb{N}^n, p_\alpha \neq 0\}$. For $\alpha \in \mathbb{N}^n$, define $|\alpha| = \sum_{i=1}^n \alpha_i$. We define the degree of p by $\max\{|\alpha| : \alpha \in \text{supp}(p)\}$.

Define $\mathbb{N}_d^n = \{\alpha \in \mathbb{N}^n : |\alpha| \leq d\}$ and $\mathbb{T}_d^n = \{x^\alpha : \alpha \in \mathbb{N}_d^n\}$, the set of monomials of degree at most d, and $\mathbb{R}[\mathbf{x}]_d$ to be all polynomials of degree at most d. More generally, for any $S \subseteq \mathbb{N}^n$, define $\mathbb{T}_S^n = \{x^\alpha : \alpha \in S\}$, the monomials indexed by S, and $\mathbb{R}_S[\mathbf{x}] = \left\{\sum_{\alpha \in S} c_\alpha x^\alpha : c \in \mathbb{R}^S\right\}$, all polynomials with support in S.

Given any probability distribution μ over $\mathbb{R}^n$ and $p \in \mathbb{R}^n[x]$, we write $\mathbb{E}_{x \sim \mu}[p(x)]$ to denote the expected value of p under μ. We will often abbreviate this $\mathbb{E}_\mu[p(x)]$ when the context is clear.

For any set $S \subseteq V$, where V is a real vector space, we define the linear span of V as $\text{span}(V) = \left\{\sum_{i=1}^k c_i v_i : k \in \mathbb{N}, c_i \in \mathbb{R}, v_i \in V, i \in [k]\right\}$, and the cone generated by V as $\text{cone}(V) = \left\{\sum_{i=1}^k \lambda_i v_i : k \in \mathbb{N}, \lambda_i \geq 0, v_i \in V, i \in [k]\right\}$.

6.2 The Lasserre Relaxation

Imagine we wish to solve the following general polynomial optimization problem

$$\begin{aligned} \sup \quad & p(x) \\ \text{subject to} \quad & f_i(x) \geq 0, \forall i \in [k] \\ & g_i(x) = 0, \forall i \in [l] \end{aligned} \tag{6.1}$$

where $f_1, \ldots, f_k, g_1, \ldots, g_l \in \mathbb{R}[\mathbf{x}]$. While this may seem a daunting task, using the existential theory of the reals, it is indeed possible to solve this type of problem in time exponential in the number of variables (see for example [Ren92]).

In this course however, we will be interested in tractable approximations for such problems. For this purpose, Lasserre and Parillo have developed a hierarchy of tractable semidefinite relaxations to polynomial optimization problems which are parameterized by degree. These relaxations give a powerful automated way to reason about polynomial optimization problems, and in interesting cases (which will see throughout the course) they will allow us to develop good approximation algorithms.

During the course, we will mostly be concerned with polynomial optimization problems where the feasible region will either be the hypercube (with possibly additional constraints) or the hypersphere. As the general theory behind the Lasserre [Las01] & Parrilo [Par03] relaxations is both illuminating and helpful, we will explain the relaxations in their full generality first and only later specialize them to the hypercube and sphere.

To begin let $F = \{f_1, \ldots, f_k\}$ and $G = \{g_1, \ldots, g_l\}$ and let

$$K_{F,G} = \{x \in \mathbb{R}^n : f(x) \geq 0, f \in F, g(x) = 0, \quad \forall g \in G\} \ , \tag{6.2}$$

denote the semi-algebraic set corresponding to the feasible region. Note that the most important set in combinatorial optimization, the hypercube $\{0,1\}^n$ is an algebraic variety, i.e. the zero set of a system of polynomials. In particular

$$\{0,1\}^n = \left\{x \in \mathbb{R}^n : x_i^2 - x_i = 0, \forall i \in [n]\right\} \ .$$

Furthermore, the sphere $S^{n-1} = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 - 1 = 0\}$ is also clearly an algebraic variety.

To understand how to relax Problem (6.1), we start with the first simple observation

$$\sup\{p(x) : x \in K_{F,G}\} = \sup\{\mathbb{E}_\mu[p(x)] : \mu \text{ probability distribution supported on } K_{F,G}\} \ ,$$

namely, the value of the program stays the same when we optimize over probability distributions over our feasible region (since we can always consider distributions supported on a single point of our set). Note that any optimizing distribution must be supported only optimal solutions for the objective $p(x)$.

At this point, one should immediately wonder how we can even hope to specify such distributions, which certainly seems like a hopeless task in general. To make this perspective useful, we will perform the first "lossy" transformation of the problem.

Instead of trying to directly completely specify distributions, we will only try to keep track of their "aggregate statistics", namely, their **moments**. More precisely, we will only allow ourselves to inspect a distribution μ via measurements of the form $\mathbb{E}_\mu[q(x)]$, for a limited set of polynomials q.

6.2.1 Pseudo Expectation Operators

One highly successful approach here has been to restrict attention to only **low degree** moments of μ. That is, we will only try keep track of $\mathbb{E}_\mu[p(x)]$ for all $p \in \mathbb{R}[\mathbf{x}]_d$. Note that by linearity of expectation

$$\mathbb{E}_\mu[p(x)] = \sum_{|\alpha| \le d} p_\alpha \mathbb{E}_\mu[x^\alpha] \ ,$$

thus to be able to evaluate any degree d polynomial, we need only keep track of the moments $\mathbb{E}_\mu[x^\alpha]$ for $|\alpha| \le d$. In particular, since there are at most $\binom{n+d}{d} = n^{O(d)}$ monomials of degree at most d on n variables, for fixed degree d, the amount of space we need to store this information is polynomial in n.

Clearly, for this to make sense with respect to our problem 6.1, we will need that the degree d of the polynomials we consider to be at least as large as maximum degree of the polynomials $p, f_1, \dots, f_k, g_1, \dots, g_l$ defining the semi-algebraic optimization problem.

Now let $L : \mathbb{R}[\mathbf{x}]_d \to \mathbb{R}$ be a linear map that sends polynomials of degree d to real numbers. Note that the expectation operator $q(x) \to \mathbb{E}_\mu[q(x)]$ is such a linear map. The main idea for relaxing (6.1) will be to try and maximize $L(p(x))$ subject to L "looking like" the expectation operator of a real distribution over $K_{F,G}$. We shall formally define below what we mean by an operator L "looking like" an expectation operator, where we shall dub these objects "pseudo-expectation" operators. We are now lead to the following fundamental question:

Question 1. *Given a linear map* $L : \mathbb{R}[\mathbf{x}]_d \to \mathbb{R}$*, can one find a distribution* μ *supported on* $K_{F,G}$ *such that* $L[p(x)] = \mathbb{E}_\mu[p(x)]$*,* $\forall p \in \mathbb{R}[\mathbf{x}]_d$*?*

This question is an example of a **moment problem** and is unfortunately extremely difficult to resolve in general (i.e. at least NP-Hard). We note that even in the case that $K_{F,G} = \mathbb{R}$ the real line and degree d polynomials, the answer (as we will see), is quite non-trivial.

While we cannot hope to certify that L as above is consistent with the expectation operator of a true distribution over $K_{F,G}$, we can go along way towards verifying that L satisfies many of the properties of such an operator.

Somewhat surprisingly, one of the most powerful tests for distinguishing a real expectation operator from a "fake one" is derived from the most basic inequality on real numbers, namely, for any real number a we have that $a^2 \geq 0$. Applying this to expectation operators and polynomials, for any probability measure μ and any polynomial $p \in \mathbb{R}[\mathbf{x}]$, we have that $\mathbb{E}_\mu[p(x)^2] \geq 0$. Thus, for an operator $L : \mathbb{R}[\mathbf{x}]_d \to \mathbb{R}$ to be consistent with probability measure, we must have that $L[p(x)^2] \geq 0$ whenever L is defined on $p(x)^2$, namely when $\deg(p) \leq d/2$.

Since we interested in achievable moments of distributions supported on $K_{F,G}$, we should also impose the "obvious" inequalities implied by our polynomial system. We thus finally arrive at what is known as the Lasserre relaxation of $K_{F,G}$:

Definition 1 (Level d Lasserre relaxation)**.** *For* $F = \{f_1, \ldots, f_l\}, G = \{g_1, \ldots, g_k\} \subseteq \mathbb{R}[\mathbf{x}]$*, yielding the system*

$$K_{F,G} = \{x \in \mathbb{R}^n : f_i(x) \geq 0, \forall i \in [l], g_j(x) = 0, \forall j \in [k]\}$$

we define the degree $d \in \mathbb{N}$ *Lasserre relaxation over* (F, G) *by*

$$\begin{aligned} \mathrm{Las}_d(F,G) = \big\{ L : \mathbb{R}[\mathbf{x}]_d \to \mathbb{R} \textit{ linear } : \\ & L[1] = 1, \\ & L[q(x)^2] \geq 0 \quad \forall q \in \mathbb{R}[\mathbf{x}]_{d/2}, \\ & L[f(x)q(x)^2] \geq 0 \quad \forall q \in \mathbb{R}[\mathbf{x}]_{(d-\deg(f))/2}, f \in F, \\ & L[g(x)q(x)] = 0 \quad \forall q \in \mathbb{R}[\mathbf{x}]_{d-\deg(g)}, g \in G \big\} \ . \end{aligned} \tag{6.3}$$

A solution $L \in \mathrm{Las}_d(F, G)$ *is called a degree* d *pseudo-expectation operator over the system* (F, G) *(we shall omit the reference to* (F, G) *when the context is clear).*

For the optimization problem $\sup\{p(x) : x \in K_{F,G}\}$*, the corresponding degree* d *Lasserre relaxation is*

$$\sup\{L[p(x)] : L \in \mathrm{Las}_d(F, G)\} \ . \tag{6.4}$$

All the above desired inequalities for L in the system (6.3) are basic inequalities that clearly hold for the expectation operator of any true distribution over $K_{F,G}$. We note that in the relaxation, we restrict the degree of the q "test polynomials" to guarantee that the resulting check can be evaluated by L. That is, we can only test the inequality $L[f_i(x)q^2(x)] \geq 0$ if $\deg(f_i q^2) \leq d$, and similarly $L[g_j(x)q(x)] = 0$ if $\deg(g_j q) \leq d$, since otherwise the expressions are not defined for L.

We remark that we not yet in fact completely exhausted the list of "obvious" inequalities for $K_{F,G}$. In particular, for any subset $I \subseteq [l]$, it should also hold that $L[\prod_{i \in I} f_i(x) q(x)^2] \geq 0$, namely we can include inequalities depending on products of the non-negativity constraints. However, it turns out that for most "well-conditioned" systems such products are unnecessary (which will hold for the hypercube and sphere), if we are willing to let the degree d tend to infinity (as we will see in the next subsection), and hence we will focus on the "simple" Lasserre relaxation.

An issue we have not yet discussed, is that even if we are given an optimal L^* maximizing $L[p(x)]$ in the program (6.4), and we are furthermore guaranteed that L^* coincides with the moments of a true distribution, it can still be difficult to recover an actual optimal (or even near optimal) solution. If the distribution underlying L^* was supported on just a single point, then rounding would be easy, namely $x^* = (L^*[x_1], \dots, L^*[x_n])$ would be an optimal solution. However, in general the underlying distribution can be a mixture over many optimal solutions, which can easily foil this strategy. As a simple example, an optimal distribution with respect to maximizing x^2 over the interval $[-1, 1]$ is to take μ uniform $\{-1, 1\}$, but this satisfies $\mathbb{E}_\mu[x] = 0$, and hence naive "expectation" rounding doesn't work. Recovering good solutions from the Lasserre relaxation, i.e. Lasserre rounding strategies, will be a major theme of the rest of the course.

6.2.2 The Sum of Squares Dual

In this section, we will relate the Lasserre relaxation to fundamental objects in real algebraic geometry as well as conic optimization. Once the basic definitions have been established, we will show how to write the sums of squares dual for the Lasserre relaxation and give a simple example on the hypercube.

We start with important cones and ideals that will give us a new language with respect to which we can express the Lasserre relaxation.

Definition 2 (Sum of Squares Cone). *Define* $\Sigma^2_{n,d} = \text{cone}(q^2 : q \in \mathbb{R}[\mathbf{x}], \deg(q) \leq d/2)$ *and* $\Sigma^2_n := \Sigma^2_{n,\infty}$, *to be the cone of polynomials on n variables of degree at most d and unbounded degree respectively which are sums of squares. We will often write* Σ^2_d *and* Σ^2 *when the context is clear.*

Definition 3 (Quadratic Module). *For a polynomial system* $F = \{f_1, \dots, f_k\} \in \mathbb{R}[\mathbf{x}]$, *define the quadratic module of F by* $Q(F) = \text{cone}(fq^2 : f \in F, q \in \mathbb{R}[\mathbf{x}])$ *and the degree d (truncated) quadratic module* $Q_d(F) = \text{cone}(fq^2 : f \in F, \deg(fq^2) \leq d)$. *Note that with this notation* $Q(1) = \Sigma^2_n$ *and* $Q_d(1) = \Sigma^2_{n,d}$.

Definition 4 (Ideal). *For a polynomial system $G = \{g_1, \ldots, g_l\} \subseteq \mathbb{R}[\mathbf{x}]$, we define the ideal generated by G by $I(G) = \mathrm{span}(gq : \forall g \in G, q \in \mathbb{R}[\mathbf{x}])$, and the degree d (truncated) ideal $I_d(G) = \mathrm{span}(gq : \forall g \in G, q \in \mathbb{R}[\mathbf{x}], \deg(gq) \leq d)$. Note that $I_d(G)$ is a* linear subspace *of* $\mathbb{R}[\mathbf{x}]_d$.

From here, it is now direct to see that degree d Lasserre relaxation of $\sup \{p(x) : x \in K_{F,G}\}$ can be restated as

$$\begin{array}{rl} \sup L[p(x)] & \\ \text{subject to} & L[1] = 1 \\ & L[v(x)] \geq 0 \ \forall\ v \in Q_d(F \cup \{1\}) \\ & L[w(x)] = 0 \ \forall\ w \in I_d(G) \\ & L : \mathbb{R}[\mathbf{x}]_d \to \mathbb{R} \text{ linear} . \end{array} \tag{6.5}$$

Compared to (6.3), the main difference is that in the above we explicitly add all non-negative combinations of the $L[f_i(x)q(x)^2] \geq 0$ type constraints as well as arbitrary combinations of the $L[g_j(x)q(x)] = 0$ constraints, however both formulations are clearly equivalent.

A principle strength of convex programming is that one can use convexity to get both upper and lower bounds on the value of an optimization problem. Thus, one may ask how do we derive good upper bounds on the value of the Lasserre relaxation (6.5).

As one might expect with duality, the answer is to combine the information we get from the constraints of the program. In the case of Lasserre relaxations, the dual corresponds to finding "good" sums of squares decompositions of p:

$$\begin{array}{rl} \inf \lambda & \\ \text{subject to} & p = \lambda - v + w \\ & \lambda \in \mathbb{R} \\ & w \in Q_d(F \cup \{1\}) \\ & v \in I_d(G) \ , \end{array} \tag{6.6}$$

We call the above program the degree d sums of squares relaxation for the optimization problem $\sup \{p(x) : x \in K_{F,G}\}$.

To show that formulation is meaningful we show the following simple duality relation between the sums of squares and Lasserre relaxation:

Lemma 1 (Weak Duality). *Let $L \in \mathrm{Las}_d(F, G)$ and $p = \lambda - v + w$, $\lambda \in \mathbb{R}$, $v \in Q_d(F)$, $w \in I_d(G)$. Then $L[p(x)] \leq \lambda$.*

Proof.

$$\begin{aligned} L[p(x)] &= L[\lambda + w - v] = L[\lambda] + L[w] - L[v] \\ &= \lambda + 0 - L[v] \leq \lambda \ , \end{aligned}$$

since $L[1] = 1$, $L[w] = 0 \ \ \forall\, w \in I_d(G)$, $L[v] \geq 0 \ \ \forall\, v \in Q_d(F \cup \{1\})$. □

6.2.3 Hypercube Example

We now illustrate the primal and dual viewpoints using a simple example on the hypercube. Assume we wish to maximize

$$\max \sum_{i=1}^{n} x_i^2, x \in \{0,1\}^n \ .$$

Clearly, the optimal solution is $x = 1_n$, the all ones vector, which has value n. Our goal is to show that the degree 2 Lasserre relaxation indeed recovers this solution, and that we can witness the optimality of this solutions using a degree 2 sums of square certificate.

Firstly, note for the hypercube $F = \emptyset$ and that $G = \{x_i^2 - x_i : \forall i \in [n]\}$. Therefore

$$Q_2(F \cup \{1\}) = \Sigma^2_{n,2} = \left\{ \sum_{j=1}^{k} (c_{j,0} + \sum_{i=1}^{n} c_{j,i} x_i)^2 : k \in \mathbb{N}, c_1, \ldots, c_k \in \mathbb{R}^{n+1} \right\} \ ,$$

i.e. sums of squares of linear polynomials, and

$$I_2(G) = \left\{ \sum_{i=1}^{n} \gamma_i (x_i^2 - x_i) : \gamma \in \mathbb{R}^n \right\} \ .$$

Note that the polynomials in G already have degree 2, so to keep degree 2 in $I_2(G)$ we can only take scalar combinations of them.

Given the above, the level 2 Lasserre relaxation can be written as

$$\begin{aligned} \max L[\sum_{i=1}^{n} x_i^2] & \\ \text{subject to } & L : \mathbb{R}_2[\mathbf{x}] \to \mathbb{R} \text{ linear} \\ & L[1] = 1 \\ & L[(a_0 + \sum_{i=1}^{n} a_i x_i)^2] \geq 0 \quad \forall a \in \mathbb{R}^{n+1} \\ & L[x_i^2 - x_i] = 0 \quad \forall\, i \in [n] \end{aligned}$$

Given we know that the optimal solution is the all ones vector for the real program, let us simply set $L[q(x)] = q(1_n)$, $\forall q \in \mathbb{R}_2[x]$, i.e. the evaluation of q at the all ones vector. Clearly $L[\sum_{i=1}^{n} x_i^2] = n$ as we would expect. Now the goal is to show that this is indeed the optimal Lasserre solution. Thus we must show that the exists a degree 2 sums of squares relaxation solution with the same value.

In particular, we must show that we can express

$$\sum_{i=1}^{n} x_i^2 = n - v(x) + w(x)$$

where $v \in \Sigma_2^2$ and $w \in I_2(G)$ as above. Note that it suffices to show that $x_i^2 = 1 - v_i(x) + w_i(x)$, $v_i \in \Sigma_2^2, w_i \in I_2(G)$, since we can then just add up these solutions together. From here, we see that

$$\begin{aligned} x_i^2 = 1 - 1 + x_i^2 &= 1 - (1 - 2x_i + x_i^2) + 2(x_i^2 - x_i^2) \\ &= 1 - (1 - x_i)^2 + 2(x_i^2 - x_i)\ . \end{aligned} \tag{6.7}$$

Since $(1 - x_i)^2 \in \Sigma_2^2$ and $2(x_i^2 - x_i) \in I_2(G)$, this is the desired decomposition. Hence, the optimal degree 2 Lasserre value is n is coincides with the value of the real program.

Building Dual Solutions Step by Step While the above proof of optimality is simple, it is perhaps a bit difficult to come up with a the dual decomposition "all at once". Let us know show that $L[\sum_{i=1}^n x_i] \leq n$ in a more "step by step" manner:

$$\begin{aligned} L[\sum_{i=1}^n x_i^2] &=_{(L[1]=1)} n - L[\sum_{i=1}^n 1 - x_i^2] \\ &=_{(x_i = x_i^2)} n - L[\sum_{i=1}^n 1 - x_i] \\ &=_{(x_i = x_i^2)} n - L[\sum_{i=1}^n (1 - x_i)^2] \\ &\leq_{(1-x_i)^2 \geq 0} n \ . \end{aligned}$$

Here it is easy to check that one can derive the certificate of the form (6.7) by "unfolding" the above proof. Regardless, it is in general easier to derive such proofs in the above manner.

6.3 Convergence Results

Now that we have developed the basic language for both the Lasserre relaxation and its sums of squares dual, we can now state a theorem of Lasserre, which is based on the work of Putinar, which tells us that under quite general conditions one can expect the value of these relaxations to converge to the true value of the optimization problem.

The main thing we will require that implies convergence is the system (F, G) needs to have a simple "algebraic proof of boundness".

Definition 5. *A polynomial system* (F, G) *is* **Archemedean** *if for some* $R > 0$*, the polynomial* $R^2 - \sum_{i=1}^n x_i^2 \in Q(F \cup \{1\}) + I(G)$*.*

Notice that for an Archemedian system, the feasible region

$$K_{F,G} = \{x \in \mathbb{R}^n : f_i(x) \geq 0, g_j(x) = 0, \forall i \in [k], j \in [l]\} \subseteq R\mathbb{B}_2^n$$

is contained inside a Euclidean ball of radius R around the origin. However, there are polynomial systems whose feasible region is indeed bounded, but where the system itself is not Archemedian.

We now state Lasserre's main convergence theorem[Las01]:

Theorem 1. *Assume that* (F,G), $F = \{f_1,\dots,f_k\}, G = \{g_1,\dots,g_l\} \subset \mathbb{R}[\mathbf{x}]$, *is an Archemedean system. Then*

$$\begin{aligned}\max\{p(x) : x \in K_{F,G}\} &= \lim_{d\to\infty} \sup\{L[p(x)] : L \in \mathrm{Las}_d(F,G)\} \\ &= \lim_{d\to\infty} \inf\{\lambda : \exists v \in Q_d(F\cup\{1\}), w \in I_d(G) \text{ s.t. } p = \lambda - v + w\} \ .\end{aligned}$$

Furthermore, for all d *large enough, the supremum and infimum values for the degree* d *Lasserre and sum of squares relaxations are attained and are equal to each other.*

6.4 A Semidefinite Programming Formulation

In this section, we show that both the degree d Lasserre and Sum of Squares relaxations for $\sup\{p(x) : x \in K_{F,G}\}$, where $F = \{f_1,\dots,f_k\}, G = \{g_1,\dots,g_l\} \subset \mathbb{R}[\mathbf{x}]$, are efficiently solvable. More precisely, we will show that they can be expressed as semidefinite programs of size $(k+l)n^{O(d)}$.

6.4.1 The Lasserre Relaxation

The goal is to show that

$$\begin{aligned}&\max L[p(x)] \\ &\text{subject to} \quad L[1] = 1, \\ &\qquad L[f(x)q(x)^2] \geq 0 \quad \forall q \in \mathbb{R}[\mathbf{x}]_{(d-\deg(f))/2}, f \in F\cup\{1\}, \\ &\qquad L[g(x)q(x)] = 0 \quad \forall q \in \mathbb{R}[\mathbf{x}]_{d-\deg(g)}, g \in G, \\ &\qquad L : \mathbb{R}[\mathbf{x}]_d \to \mathbb{R} \text{ linear} \ .\end{aligned}$$

is a semidefinite program of size $(k+l)n^{O(d)}$. Recall that to represent L it suffices to know the evaluation of L on all monomials $(L[x^\alpha])_{|\alpha|\leq d}$, which will correspond to the $\binom{n+d}{d}$ variables of the SDP representation. We will now go through the constraints one by one and show how to express them as linear matrix inequalities. Clearly, the constraint $L[1] = 1$ is a linear equation on L. For the constraints $L[g(x)q(x)] = 0$, for $g \in G$, $\deg(gq) \leq k$, it clearly suffices to check that

$$L[g(x)x^\alpha] = 0 \quad \forall |\alpha| \leq d - \deg(g) \ .$$

Expanding out, we get the constraints

$$0 = L[g(x)x^\alpha] = \sum_\beta g_\beta L[x^{\alpha+\beta}], \quad \forall |\alpha| \leq d - \deg(g) \ ,$$

which is simply a linear constraint in L. Thus, all the G constraints, can be expressed using at most $k\binom{n+d}{d} = kn^{O(d)}$ different homogeneous linear equations.

Next, for $f \in F \cup \{1\}$, we need to check that

$$L[f(x)q(x)^2] \geq 0 \quad \forall q \in \mathbb{R}[x], \deg(q) \leq (d - \deg(f))/2 \ .$$

Letting $r_f = \lfloor (d - \deg(f))/2 \rfloor$, any such q can be written as $q = \sum_{|\alpha| \leq r_f} q_\alpha x^\alpha$. Again, expanding out, we get that

$$\begin{aligned} L[f(x)q(x)^2] &= \sum_{|\alpha|,|\beta| \leq r_f} q_\alpha q_\beta L[f(x)x^{\alpha+\beta}] \\ &= \sum_{|\alpha|,|\beta| \leq r_f} q_\alpha q_\beta (\sum_\gamma f_\gamma L[x^{\alpha+\beta+\gamma}]) \ . \end{aligned}$$

Let $M^f_{\alpha,\beta} = \sum_\gamma f_\gamma L[x^{\alpha+\beta+\gamma}]$, the above requirement for all q of degree at most r_f is equivalent to

$$\sum_{\alpha,\beta} M^f_{\alpha,\beta} c_\alpha c_\beta \geq 0, \quad \forall c \in \mathbb{R}^{\mathbb{N}^n_{r_f}} \Leftrightarrow M^f \succeq 0 \ .$$

Since the entries of M^f are linear functions of L, the above is a linear matrix inequality, and hence semidefinite representable. Thus, we can represent all the $F \cup \{1\}$ constraints using at most $l + 1$ semidefinite constraints on matrices of size at most $\binom{n+d}{d} = n^{O(d)}$ as needed.

6.4.2 The Sum of Squares Relaxation

The goal is to show that

$$\begin{aligned} &\min \lambda \\ \text{subject to} \quad & p = \lambda - v + w \\ & \lambda \in \mathbb{R} \\ & v \in Q_d(F \cup \{1\}) \\ & w \in I_d(G) \end{aligned}$$

is a semidefinite program of size $(k + l)n^{O(d)}$.

The claim will follow directly from the following two lemmas.

Lemma 2. *Let $p \in \mathbb{R}[\mathbf{x}]_d$, $f \in \mathbb{R}[\mathbf{x}]$, $\deg(f) = d - 2r$. Then*

$$p \in Q_d(f) \Leftrightarrow$$

$$\exists M \succeq 0 \text{ such that } p_\eta = \sum_{\substack{|\alpha|,|\beta| \leq r, |\gamma| \leq d-2r \\ \alpha+\beta+\gamma=\eta}} f_\gamma M_{\alpha,\beta}, \quad \forall |\eta| \leq d \ .$$

In particular, $Q_d(f)$ has a semidefinite representation using a PSD constraint of size $\binom{n+r}{r}$ and $\binom{n+d}{d}$ linear equations.

Proof. Since $\deg(f) = d - 2r$, we see that $Q_d(f) = \text{cone}\{fq^2 : q \in \mathbb{R}[\mathbf{x}]_r\}$. Now assume that $p \in \text{cone}(fq^2 : q \in \mathbb{R}[\mathbf{x}]_r)$. Then we can write

$$p = \sum_{i=1}^{k} f q_i^2 \ ,$$

for $k \in \mathbb{N}$, $q_i \in \mathbb{R}[\mathbf{x}]_r$, $i \in [k]$. Writing

$$q_i = \sum_{|\alpha| \le r} c_{i,\alpha} x^\alpha = \mathbf{c}_i^\mathsf{T} (x^\alpha)_{|\alpha| \le r} \ ,$$

We can express

$$\begin{aligned}
\sum_{i=1}^{k} f q_i^2 &= f \cdot \sum_{i=1}^{k} (\mathbf{c}_i^\mathsf{T} (x^\alpha)_{|\alpha| \le r})^2 \\
&= f \cdot \sum_{i=1}^{k} (x^\alpha)_{|\alpha| \le r}^\mathsf{T} \mathbf{c}_i \mathbf{c}_i^\mathsf{T} (x^\alpha)_{|\alpha| \le r} \\
&= f \cdot (x^\alpha)_{|\alpha| \le r}^\mathsf{T} \left(\sum_{i=1}^{k} \mathbf{c}_i \mathbf{c}_i^\mathsf{T} \right) (x^\alpha)_{|\alpha| \le r}
\end{aligned}$$

Defining $M = \sum_{i=1}^{k} \mathbf{c}_i \mathbf{c}_i^\mathsf{T}$, we have that $M \succeq 0$ and

$$\begin{aligned}
f \cdot (x^\alpha)_{|\alpha| \le r}^\mathsf{T} M (x^\alpha)_{|\alpha| \le r} &= f \cdot \sum_{|\alpha|, |\beta| \le r} M_{\alpha,\beta} x^{\alpha+\beta} \\
&= \sum_{|\alpha|, |\beta| \le r, |\gamma| \le d-2r} f_\gamma M_{\alpha,\beta} x^{\alpha+\beta+\gamma}.
\end{aligned}$$

Putting everything together, we get the equality

$$p = \sum_{|\alpha|, |\beta| \le r, |\gamma| \le d-2r} f_\gamma M_{\alpha,\beta} x^{\alpha+\beta+\gamma} \ .$$

Since the polynomials are equal, they must have the same coefficients, and hence

$$p_\eta = \sum_{\substack{|\alpha|, |\beta| \le r, |\gamma| \le d-2r \\ \alpha+\beta+\gamma = \eta}} f_\gamma M_{\alpha,\beta} \quad \forall |\eta| \le d \ ,$$

where we note that the right hand side equal 0 for $|\eta| > d$ since $\alpha + \beta + \gamma \le d$.

For the converse, one may follow the above proof in reverse, using the fact that a matrix $M \succeq 0$ if and only if $M = \sum_{i=1}^{k} \mathbf{c}_i \mathbf{c}_i^{\mathsf{T}}$ for an appropriate set of vectors (i.e. it admits a Cholesky factorization).

For the semidefinite representation, note that the description given in the statement of the lemma is such a representation. ☐

Lemma 3. *Let $p \in \mathbb{R}[\mathbf{x}]_d$, $g \in \mathbb{R}[\mathbf{x}]$, $\deg(g) = d - r$. Then*

$$p \in I_d(g) \Leftrightarrow \exists c \in \mathbb{R}^{\mathbb{N}^n_r} \text{ such that } p_\eta = \sum_{\substack{|\alpha| \le r \\ \alpha + \gamma = \eta}} g_\gamma c_\alpha, \quad \forall |\eta| \le d \ .$$

In particular, $I_d(g)$ has a linear representation of using $\binom{n+r}{r}$ variables and $\binom{n+d}{d}$ linear equations.

Proof. First, since $\deg(g) = d - r$, we have that $I_d(g) = \{gq : q \in \mathbb{R}[\mathbf{x}]_r\}$. Now assume that $p = gq$ for $q \in \mathbb{R}[\mathbf{x}_r]$. Since we can express $p = \sum_{|\alpha| \le r} c_\alpha x^\alpha$, expanding out, we get that

$$p = gq = g \cdot \sum_{|\alpha| \le r} c_\alpha x^\alpha = \sum_{|\alpha| \le r, |\gamma| \le d - r} g_\gamma c_\alpha x^{\alpha + \gamma} \ .$$

Since the polynomial p equals the polynomial on the right, the coefficients must be equal, and hence

$$p_\eta = \sum_{\substack{|\alpha| \le r \\ \alpha + \gamma = \eta}} g_\gamma c_\alpha, \quad \forall |\eta| \le d,$$

where we note the right hand side equals 0 for $|\eta| > d$ since $\alpha + \gamma \le d$.

For the linear representation, note that the expression for $I_d(g)$ in the theorem statement is such a representation. ☐

We now give the final SDP representation for the sums of squares relaxation. Firstly, it is easy to check that

$$Q_d(F \cup \{1\}) = \Sigma_d^2 + \sum_{i=1}^{k} Q_d(f_i) \quad \text{and} \quad I_d(G) = \sum_{j=1}^{l} I_d(g_j) \ .$$

Given this, we can write the sum of squares relaxation as

$$\begin{array}{rl} \min \lambda & \\ \text{subject to} & p = \lambda - (v_0 + \sum_{i=1}^{k} v_i) + \sum_{j=1}^{l} w_l \\ & \lambda \in \mathbb{R} \\ & v_0 \in \Sigma_d^2 \\ & v_i \in Q_d(f_i) \quad \forall i \in [k] \\ & w_j \in I_d(g_j) \quad \forall j \in [l] \; . \end{array} \tag{6.8}$$

Given $\lambda, v_0, \ldots, v_l, w_1, \ldots, w_k$, checking

$$p = \lambda - (v_0 + \sum_{i=1}^{k} v_i) + \sum_{j=1}^{l} w_l$$

corresponds to checking $\deg(p) \leq d$ (generally assumed) and that the coefficients match

$$p_0 = \lambda - (v_{0,0} + \sum_{i=1}^{k} v_{i,0}) + \sum_{j=1}^{l} w_{l,0} \; ,$$

$$p_\alpha = -(v_{0,\alpha} + \sum_{i=1}^{k} v_{i,\alpha}) + \sum_{j=1}^{l} w_{l,\alpha} \quad \forall |\alpha| \leq d, \alpha \neq 0 \; .$$

Since all the sets $\Sigma_d^2 := Q_d(1), Q_d(f_1), \ldots, Q_d(f_l), I_d(g_1), \ldots, I_d(g_l)$ have semidefinite representations of size $n^{O(d)}$, the program (6.8) yields a semidefinite representation of size $(k+l)n^{O(d)}$ for the sum of squares relaxation.

References

[Las01] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[Par03] Pablo A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming*, 96(2):293–320, 2003.

[Ren92] James Renegar. On the computational complexity and geometry of the first-order theory of the reals. part i: Introduction. preliminaries. the geometry of semi-algebraic sets. the decision problem for the existential theory of the reals. *Journal of Symbolic Computation*, 13(3):255 – 299, 1992.