MIT 6.972 Algebraic techniques and semidefinite optimization	March 14, 2006
Lecture 10	
Lecturer: Pablo A. Parrilo	*Scribe: ???*

In this lecture we begin our study of one of the main themes of the course, namely the relationships between polynomials that are sums of squares and semidefinite programming.

1 Nonegativity and sums of squares

Recall from a previous lecture the definition of a polynomial being a sum of squares.

Definition 1. *A univariate polynomial $p(x)$ is a* sum of squares *(SOS) if there exist $q_1, \ldots, q_m \in \mathbb{R}[x]$ such that*

$$p(x) = \sum_{k=1}^{m} q_k^2(x). \tag{1}$$

If a polynomial $p(x)$ is a sum of squares, then it obviously satisfies $p(x) \geq 0$ for all $x \in \mathbb{R}$. Thus, a SOS condition is a sufficient condition for global nonnegativity.

As we have seen, in the univariate case, the converse is also true:

Theorem 2. *A univariate polynomial is nonnegative if and only if it is a sum of squares.*

As we will see, there is a very direct link between sum of squares conditions on polynomials and semidefinite programming. We study first the univariate case.

2 Sums of squares and semidefinite programming

Consider a polynomial $p(x)$ of degree $2d$ that is a sum of squares, i.e., it can be written as in (1). Notice that the degree of the polynomials q_k is at most equal to d, since the highest term of each q_k^2 is positive, and thus there cannot be any cancellation in the highest power of x. Then, we can write

$$\begin{bmatrix} q_1(x) \\ q_2(x) \\ \vdots \\ q_m(x) \end{bmatrix} = V \begin{bmatrix} 1 \\ x \\ \vdots \\ x^d \end{bmatrix}, \tag{2}$$

where $V \in \mathbb{R}^{m \times (d+1)}$, and its kth row contains the coefficients of the polynomial q_k. For future reference, let $[x]_d$ be the vector in the right-hand side of (2). Consider now the matrix $Q = V^T V$. We then have $p(x) = \sum_{k=1}^{m} q_k^2(x) = (V[x]_d)^T (V[x]_d) = [x]_d^T V^T V [x]_d = [x]_d^T Q [x]_d$.

Conversely, assume there exists a symmetric positive definite Q, for which $p(x) = [x]_d^T Q [x]_d$. Then, by factorizing $Q = V^T V$ (e.g., via Choleski, or square root factorization), we arrive at a SOS decomposition of p.

We formally express this in the following lemma, that gives a direct relation between positive semidefinite matrices and a sum of squares condition.

Lemma 3. *Let $p(x)$ be a univariate polynomial of degree $2d$. Then, $p(x)$ is nonnegative (or SOS) if and only if there exists $Q \in \mathcal{S}_+^{d+1}$ that satisfies*

$$p(x) = [x]_d^T Q [x]_d.$$

Indexing the rows and columns of Q by $\{0,\ldots,d\}$, we have:

$$[x]_d^T Q [x]_d = \sum_{j=0}^{d}\sum_{k=0}^{d} Q_{jk} x^{j+k} = \sum_{i=0}^{2d}\left(\sum_{j+k=i} Q_{jk}\right) x^i$$

Thus, for this expression to be equal to $p(x)$, it should be the case that

$$p_i = \sum_{j+k=i} Q_{jk}, \qquad i = 0,\ldots,2d. \tag{3}$$

This is a system of $2d+1$ linear equations between the entries of Q and the coefficients of $p(x)$. Thus, since Q is simultaneously constrained to be positive semidefinite, and to belong to a particular affine subspace, a SOS condition is exactly equivalent to a semidefinite programming problem.

Lemma 4. *A polynomial $p(x) = \sum_{i=0}^{2d} p_i x^i$ is a sum of squares if and only if there exists $Q \in \mathcal{S}_+^{d+1}$ satisfying (3). This is a semidefinite programming problem.*

3 Applications and extensions

We discuss first a few applications of the SDP characterization of nonnegative polynomials, followed by several extensions.

3.1 Optimization

Our first application concerns the global optimization of a univariate polynomial $p(x)$. Rather than focusing on computing an $x_\star$ for which $p(x_\star)$ is as small as possible, we attempt first to obtain a good (or the best) lower bound on its optimal value. It is easy to see that a number γ is a global lower bound of a polynomial $p(x)$, if and only if the polynomial $p(x)-\gamma$ is nonnegative, i.e.,

$$p(x) \geq \gamma \quad \forall x \in \mathbb{R} \qquad \Longleftrightarrow \qquad p(x) - \gamma \geq 0 \quad \forall x \in \mathbb{R}.$$

Notice that the polynomial $p(x)-\gamma$ has coefficients that depend affinely on γ. Consider now the optimization problem defined by

$$\max \gamma \qquad \text{s.t.} \qquad p(x) - \gamma \text{ is SOS.}$$

It should be clear that this is a *convex* problem, since the feasible set is defined by an infinite number of linear inequalities. Its optimal solution $\gamma_\star$ is equal to the global minimum of the polynomial, $p(x_\star)$. Furthermore, using Lemma 4, we can easily write this as a semidefinite programming problem. We can thus obtain the global minimum of a univariate polynomial, by solving an SDP problem. Notice also that at optimality, we have $0 = p(x_\star) - \gamma_\star = \sum_{k=1}^{m} q_k^2(x_\star)$, and thus all the q_k simultaneously vanish at $x_\star$, which gives a way of computing the optimal solution $x_\star$. As we shall see later, we can also obtain this solution directly from the dual problem, by using complementary slackness.

Notice that even though $p(x)$ may be hightly nonconvex, we are nevertheless effectively computing its global minimum.

3.2 Nonnegativity on intervals

We have seen how to characterize a univariate polynomial that is nonnegative on $(-\infty,\infty)$ in terms of SDP conditions. But what if we are interested in polynomials that are nonnegative only in an interval (either finite, or semi-infinite)? As explained below, we can use very similar ideas, and two classical characterizations, usually associated to the names Pólya-Szegö, Fekete, or Markov-Lukacs. The basic results are the following:

Theorem 5. *The polynomial $p(x)$ is nonnegative on $[0, \infty)$, if and only if it can be written as*

$$p(x) = s(x) + x\, t(x),$$

where $s(x), t(x)$ are SOS. If $\deg(p) = 2d$, then we have $\deg(s) \leq 2d$, $\deg(t) \leq 2d - 2$, while if $\deg(p) = 2d + 1$, then $\deg(s) \leq 2d$, $\deg(t) \leq 2d$.

Theorem 6. *Let $a < b$. Then, $p(x)$ is nonnegative on $[a, b]$, if and only if it can be written as*

$$\begin{cases} p(x) = s(x) + (x-a)(b-x)t(x), & \text{if } \deg(p) \text{ is even} \\ p(x) = (x-a)s(x) + (b-x)t(x), & \text{if } \deg(p) \text{ is odd} \end{cases}$$

where $s(x), t(x)$ are SOS. In the first case, we have $\deg(p) = 2d$, and $\deg(s) \leq 2d$, $\deg(t) \leq 2d - 2$. In the second, $\deg(p) = 2d + 1$, and $\deg(s) \leq 2d$, $\deg(t) \leq 2d$.

Notice that in both of these results, one direction of the implication is evident.

3.3 Rational functions

What happens if we want to minimize a univariate rational function, rather than a polynomial? Consider a rational function given as a quotient of polynomials $p(x)/q(x)$, where $q(x)$ is strictly positive (why?). Then, we have

$$\frac{p(x)}{q(x)} \geq \gamma \qquad \Leftrightarrow \qquad p(x) - \gamma\, q(x) \geq 0,$$

and therefore we can find the global minimum of the rational function by solving

$$\max \gamma \qquad \text{s.t.} \quad p(x) - \gamma\, q(x) \text{ is SOS.}$$

The constrained case (i.e., over finite or semi-infinite intervals) are very similar, and can be formulated using the results in the Section 3.2. The details are left for the exercises.

4 Multivariate polynomials

For polynomials in more than one variable, it is no longer true that nonnegativity is equivalent to a sum of squares condition. In fact, for polynomials of degree greater than or equal to four, deciding polynomial nonnegativity is an NP-hard problem (as a function of the number of variables).

More than a century ago, David Hilbert showed that equality between the set of nonnegative and SOS polynomials holds only in the following three cases:

- Univariate polynomials (i.e., $n = 1$)
- Quadratic polynomials ($2d = 2$)
- Bivariate quartics ($n = 2$, $2d = 4$)

For all other cases, there always exist nonnegative polynomials that are *not* sums of squares. A classical counterexample is the bivariate sextic ($n = 2$, $2d = 6$) due to Motzkin, given by (in dehomogenized form)

$$M(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2.$$

This polynomial is nonnegative, but is not a sum of squares. We will prove both facts later. An excellent account of much of the classical work in this area has been provided by Bruce Reznick [Rez00].

4.1 SDP formulation

Essentially the same construction we have seen in Lemma 4 applies to the multivariate case. In this case, we consider polynomials of degree $2d$ in n variables. In the dense case, i.e., when the polynomial is not sparse, the number of coefficients is equal to $\binom{n+2d}{2d}$. If we let $p(x) = \sum_\alpha p_\alpha x^\alpha$, and indexing the matrix Q by the $\binom{n+d}{d}$ monomials in n variables of degree d, we have the SDP conditions on $Q \in \mathcal{S}_+^{\binom{n+d}{d}}$:

$$Q \succeq 0, \qquad p_\alpha = \sum_{\beta+\gamma=\alpha} Q_{\beta\gamma}. \tag{4}$$

We have exactly $\binom{n+2d}{2d}$ linear equations, one per each coefficient of $p(x)$. As before, these conditions are affine conditions relating the entries of Q and the coefficients of $p(x)$. Thus, we can decide membership to, or optimize over, the set of SOS polynomials by solving a semidefinite programming problem.

4.2 Using the Newton polytope

Recall that we have defined in a previous lecture the *Newton polytope* of a polynomial $p(x) \in \mathbb{R}[x_1, \ldots, x_n]$ as the convex hull of the set of exponents appearing in p. This allowed us to introduce a notion of sparseness for a polynomial, related to the size of its Newton polytope. Sparsity (in this algebraic sense) allows a notable reduction in the computational cost of checking sum of squares conditions of multivariate polynomials. The reason is the following theorem due to Reznick:

Theorem 7 ([Rez78], Theorem 1)**.** *If $p(x) = \sum q_i(x)^2$, then $New(q_i) \subseteq \frac{1}{2} New(p)$.*

In other words, this theorem allows us, without loss of generality, to restrict the set of monomials appearing in the representation (4) to those in the Newton polytope of p, scaled by a factor of $\frac{1}{2}$. This reduces the size of the corresponding matrix Q, thus simplifying the SDP problem.

Example 8. *Consider the following polynomial:*

$$p = (w^4 + 1)(x^4 + 1)(y^4 + 1)(z^4 + 1) + 2w + 3x + 4y + 5z.$$

The polynomial p has degree $2d = 16$, and four independent variables ($n = 4$). A naive approach, along the lines described earlier, would require a matrix Q of size $\binom{n+d}{d} = 495$. However, the Newton polytope of p is easily seen to be the four dimensional hypercube with vertices in $(0, 0, 0, 0)$ and $(4, 4, 4, 4)$. Therefore, the polynomials q_i in the SOS decomposition of p will have at most $3^4 = 81$ distinct monomials, and as a consequence the full decomposition can be computed by solving a much smaller SDP.

5 Duality and density

In the next lecture, we will revisit the sum of squares construction, but emphasizing this time the dual side, and its appealing measure-theoretic interpretation. We will also review some recent results on the relative density of the cones of nonnegative polynomials and SOS.

References

[Rez78] B. Reznick. Extremal PSD forms with few terms. *Duke Mathematical Journal*, 45(2):363–374, 1978.

[Rez00] B. Reznick. Some concrete aspects of Hilbert's 17th problem. In *Contemporary Mathematics*, volume 253, pages 251–272. American Mathematical Society, 2000.