12 Nonnegative univariate polynomials (continued)

SDP representation of P_{2d}^*: Recall that we have derived in Lecture 10 a semidefinite programming representation of P_{2d}. We are now going to derive a semidefinite representation of the dual cone P_{2d}^*. To do this let us go back to our setting where we have a random variable X on $\mathbb{R}$. Since nonnegative polynomials are sums of squares, saying that $\mathbb{E}[p(X)] \geq 0$ for all nonnegative polynomials of degree $\leq 2d$ is the same as saying that $\mathbb{E}[q(X)^2] \geq 0$ for all polynomials q of degree at most d. If $q(X) = \sum_{k=0}^d q_k X^k$ then

$$\mathbb{E}[q(X)^2] = \sum_{0 \leq i,j \leq d} q_i q_j \mathbb{E}[X^{i+j}] = \sum_{0 \leq i,j \leq d} q_i q_j y_{i+j} = q^T H(y) q$$

where $H(y) = [y_{i+j}]_{0 \leq i,j \leq d}$ is the *Hankel* matrix associated to y:

$$H(y) = \begin{bmatrix} y_0 & y_1 & \cdots & y_d \\ y_1 & \cdots & y_d & y_{d+1} \\ \vdots & & & \\ \vdots & \cdots & \cdots & y_{2d-1} \\ y_d & y_{d+1} & \cdots & y_{2d} \end{bmatrix} = [y_{i+j}]_{0 \leq i,j \leq d}. \tag{1}$$

Thus saying that $\mathbb{E}[q(X)^2] \geq 0$ for all polynomial q of degree at most d is the same as saying that $q^T H(y) q \geq 0$ for all $q \in \mathbb{R}^{d+1}$ which is equivalent to saying $H(y) \succeq 0$. We thus get the following semidefinite programming description of P_{2d}^*:

Theorem 12.1. *$P_{2d}^* = \left\{(y_0, \ldots, y_{2d}) \in \mathbb{R}^{2d+1} : H(y) \succeq 0\right\}$ where $H(y) \in \mathbf{S}^{d+1}$ is defined as in (1).*

Proof. We write a formal proof which captures the argument we just gave. Since P_{2d} coincides with polynomials that are sums of squares, we have $y \in P_{2d}^*$ if and only if $\langle p, y \rangle \geq 0$ for all polynomials p of the form $p = q^2$ where q is an arbitrary polynomial of degree $\leq d$. If $q(x) = \sum_{k=0}^d q_k x^k$ then the coefficients of the polynomial $p = q^2$ are $p_k = \sum_{0 \leq i,j \leq d : i+j=k} q_i q_j$. Thus

$$\langle q^2, y \rangle \geq 0 \iff \sum_{0 \leq i,j \leq d} q_i q_j y_{i+j} = q^T H(y) q \geq 0.$$

Thus having $\langle q^2, y \rangle \geq 0$ for all q of degree at most d is equivalent to having $H(y) \succeq 0$. This completes the proof. $\square$

Remark 1. *We can also prove Theorem 12.1 using the semidefinite representation of P_{2d} proved in Theorem 10.3. In Theorem 10.3 we showed that $P_{2d} = \pi(\mathbf{S}_+^{d+1})$ where $\pi : \mathbf{S}^{d+1} \to \mathbb{R}^{2d+1}$ is the linear map:*

$$\pi(M) = \left(\sum_{\substack{0 \leq i,j \leq d \\ i+j=k}} M_{ij} \right)_{k=0,\ldots,2d} \tag{2}$$

It is easy to prove that if π is a linear map and K any arbitrary set then $(\pi(K))^* = \{y : \pi^*(y) \in K^*\}$ where π^* is the adjoint of π (we leave it as an exercise). The adjoint of the map π defined in (2) turns out to be the Hankel map defined in (1). This gives us another proof of Theorem 12.1.

Finding an atomic measure associated to a sequence of moments, and connections with quadrature formulas From Theorem 12.1 we know that if $H(y) \succ 0$ then y is in the interior of P_{2d}^* and so (from Lecture 11) it can be written as a conic combination of elements from the moment curve $M_{2d} = \{y_x : x \in \mathbb{R}\}$, i.e.,

$$y = \sum_{i=1}^{r} \lambda_i y_{x_i} \tag{3}$$

where $\lambda_i \geq 0$, $x_i \in \mathbb{R}$. One question is: how can we find such a conic combination? The purpose of this paragraph is to draw a connection with quadrature rules for integration. In a typical quadrature problem we are given a measure μ and we are looking for points $x_1, \ldots, x_r \in \mathbb{R}$ and weights $\lambda_1, \ldots, \lambda_r > 0$ such that

$$\int p(x) d\mu(x) = \sum_{i=1}^{r} \lambda_i p(x_i) \tag{4}$$

holds for all polynomials p up to some degree, say $2d$. The Gaussian quadrature approach to this problem is to define an inner product in the space of polynomials given by $\langle p|q \rangle = \int pq d\mu$ and consider the sequence of orthogonal polynomials with respect to this inner product; the Gaussian quadrature nodes x_i in (4) then correspond to the roots of the polynomial of degree d of the sequence of orthogonal polynomials (the quadrature rule will have $r = d+1$ nodes). Observe that the requirement (4) is the same as (3) where $y_k = \int x^k d\mu(x)$. Using this identification, the condition $H(y) \succ 0$ has a natural interpretation in terms of the inner product $\langle p|q \rangle$ we just defined: it simply means that this inner product is a valid (positive definite) inner product on the space of polynomials of degree at most d. Indeed for any p, q of degree at most d we have $\langle p|q \rangle = \int (\sum_{i=0}^d p_i x^i)(\sum_{j=0}^d q_j x^j) d\mu(x) = \sum_{i,j=0}^d p_i q_j \int x^{i+j} d\mu(x) = p^T H(y) q$ (where we identified p and q with their coefficient vector in the last equality). The inner product $\langle \cdot|\cdot \rangle$ is thus a valid one provided $H(y) \succ 0$.

Nonnegativity on intervals So far we have looked at polynomials nonnegative on the real line. What if we are interested in polynomials nonnegative on an interval? The following result (which is left as an exercise to the reader) gives necessary and sufficient conditions for a polynomial to be nonnegative on $[-1, 1]$.

Theorem 12.2 (Nonnegative polynomials on $[-1, 1]$)**.** *A polynomial p of even degree $2d$ is nonnegative on $[-1, 1]$ if and only if there exist $s_1 \in P_{2d}$ and $s_2 \in P_{2d-2}$ such that $p(x) = s_1(x) + (1 - x^2)s_2(x)$.*

A polynomial p of odd degree $2d+1$ is nonnegative on $[-1, 1]$ if and only if there exist $s_1 \in P_{2d}$ and $s_2 \in P_{2d}$ such that $p(x) = (1-x)s_1(x) + (1+x)s_2(x)$.

Let $P_{2d}[-1, 1]$ be the cone of polynomials of degree $2d$ nonnegative on $[-1, 1]$. The previous theorem shows that

$$P_{2d}[-1, 1] = \left\{ p = (p_0, \ldots, p_{2d}) \; : \; \exists s_1 \in P_{2d}, s_2 \in P_{2d-2} \text{ s.t. } p = s_1 + (1 - x^2)s_2 \right\}.$$

It is important to note that the constraint $p = s_1 + (1-x^2)s_2$ is linear in (p, s_1, s_2) (x is just an indeterminate). Using the semidefinite representation of P_{2d} we can get a semidefinite representation of $P_{2d}[-1, 1]$. For example the problem

$$\min_{x \in [-1,1]} p(x) = \max \gamma \; : \; p - \gamma \in P_{2d}[-1, 1] \tag{5}$$

can be expressed as a semidefinite program. The same is true for $P_{2d+1}[-1,1]$ (polynomials of degree $2d+1$ nonnegative on $[-1,1]$).

Example: The following code implements the problem (5) on CVX (we use CVX's built-in function `nonneg_poly_coeffs(2*d)` which internally represents the cone P_{2d} using the semidefinite representation of Theorem 10.2).

```
% Find the minimum of p(x) on [-1,1]
% p(x) = 4x^4 + 3x^3 - 2*x^2 + 2
p = [4 3 -2 0 2]';
d = (length(p)-1)/2;
cvx_begin
    variable g % gamma
    variable s1(2*d+1)      % polynomial of degree 2d
    variable s2(2*d-1)      % polynomial of degree 2d-2
    maximize g
    subject to
        % p(x) - gamma = s_1(x) + (1-x^2)*s_2(x)
        p - [zeros(2*d,1); g] == s1 + conv( [-1 ; 0 ; 1] , s2 );
        s1 == nonneg_poly_coeffs(2*d);    % s_1 \in P_{2d}
        s2 == nonneg_poly_coeffs(2*d-2);  % s_2 \in P_{2d-2}
cvx_end
```

Application: probability inequalities [BP05] We now briefly explain an application of nonnegative univariate polynomials for probability inequalities, due to [BP05]. Assume we have a random variable X of which we know only its first $2d$ moments $(y_0,\dots,y_{2d})$. We want to use these moments to derive an upper bound on the probability of an event, say $\Pr[X\in A]$ where A is a subset of $\mathbb{R}$. An example of such an upper bound is Chebyshev's inequality which says that $\Pr[|X|\geq t]\leq \mathbb{E}[X^2]/t^2$ for any parameter $t>0$. Let $y_k=\mathbb{E}[X^k]$ $(k=0,\dots,2d)$ be the moments of X which we assume are given and consider the optimization problem:

$$\underset{p\in\mathbb{R}^{2d+1}}{\text{minimise}} \sum_{k=0}^{2d} p_k y_k \quad \text{subject to} \quad \sum_{k=0}^{2d} p_k x^k \geq \mathbf{1}_A(x) \quad \forall x\in\mathbb{R} \tag{6}$$

where $\mathbf{1}_A$ is the indicator function of A:

$$\mathbf{1}_A(x)=\begin{cases}1 & \text{if } x\in A\\ 0 & \text{else.}\end{cases}$$

It is not difficult to show that (6) gives an upper bound on $\Pr[X\in A]$. Indeed note that $\Pr[X\in A]=\mathbb{E}[\mathbf{1}_A(X)]\leq\mathbb{E}[p(X)]=\sum_{k=0}^{2d}p_k y_k$ where we used the constraint that $p\geq\mathbf{1}_A$ in (6), where $p(x)=\sum_{k=0}^{2d}p_k x^k$. Now if, for example $A=[-1,1]$, then the constraint $p\geq\mathbf{1}_A$ is equivalent to the following two constraints: $p-1\in P_{2d}[-1,1]$ and $p\in P_{2d}$. Since $P_{2d}[-1,1]$ and P_{2d} have semidefinite representation, this allows us to formulate (6) as a semidefinite program when $A=[-1,1]$. A similar formulation can be obtained more generally when A is a finite union of intervals.

References

[BP05] Dimitris Bertsimas and Ioana Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization*, 15(3):780–804, 2005. 3