13 The moment problem

We consider in this lecture the following question, called the *moment problem*: given numbers $(y_0, y_1, \ldots, y_{2d}) \in \mathbb{R}^{2d+1}$, does there exist a random variable X on $\mathbb{R}$ such that $\mathbb{E}[X^k] = y_k$ for all $k = 0, \ldots, 2d$? If the answer is true we will say that y is a *valid moment vector*.

It is clear that not any vector $y \in \mathbb{R}^{2d+1}$ is a valid moment vector. For example we must have $y_{2p} \geq 0$ for any $p = 1, \ldots, d$. Also we have $\operatorname{var}(X) = \mathbb{E}[(X - y_1)^2] = y_2 - y_1^2$ must be nonnegative. So $y_2 - y_1^2 \geq 0$. What other inequalities must be true? If p is any polynomial nonnegative on $\mathbb{R}$ then we must have $\mathbb{E}[p(X)] \geq 0$. If we let $p = (p_0, \ldots, p_{2d})$ be the coefficients of this polynomial this means we must have:

$$0 \leq \mathbb{E}[p(X)] = \mathbb{E}\left[\sum_{k=0}^{2d} p_k X^k\right] = \sum_{k=0}^{2d} p_k \mathbb{E}[X^k] = \sum_{k=0}^{2d} p_k y_k.$$

In other words if y is a valid moment then we must have

$$\langle p, y\rangle \geq 0 \quad \forall p \in P_{2d}.$$

This means, by definition of dual cone, that $y \in P_{2d}^*$.

Let $\mathcal{M}_{2d}$ be the set of valid moment vectors of nonnegative measures on $\mathbb{R}$. Then we have just shown that $\mathcal{M}_{2d} \subseteq P_{2d}^*$. In fact since P_{2d}^* is closed we know that

$$\operatorname{cl} \mathcal{M}_{2d} \subseteq P_{2d}^*.$$

For any $x \in \mathbb{R}$ we introduced the notation $y_x = (1, x, \ldots, x^{2d})$ in last lecture. Note that y_x is the moment vector associated to the Dirac measure δ_x that puts all its mass at $\{x\}$. Any conic combination of these vectors is a valid moment vector. Indeed if $y = \sum_{i=1}^N p_i y_{x_i}$ where $p_1, \ldots, p_N \geq 0$, then y is the moment vector of the *atomic measure* $\sum_{i=1}^N p_i \delta_{x_i}$. It thus follows that $\operatorname{cl}\operatorname{cone}(y_x : x \in \mathbb{R}) \subseteq \operatorname{cl} \mathcal{M}_{2d} \subseteq P_{2d}^*$. By Theorem 12.3 from previous lecture we thus have

$$\operatorname{cl}\operatorname{cone}(y_x : x \in \mathbb{R}) = \operatorname{cl} \mathcal{M}_{2d} = P_{2d}^*.$$

To summarise we have the following duality picture:

nonnegative polynomials of degree $\leq 2d$	$\overset{\textit{duality}}{\longleftrightarrow}$	moment vectors $(y_0, \ldots, y_{2d})$ of nonnegative measures (up to closure)

SDP representation of P_{2d}^*: Recall that we have derived in Lecture 11 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 $H(y) = [y_{i+j}]_{0\le i,j\le d}$ is the *Hankel* matrix associated to y:

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

Thus saying that $\mathbb{E}[q(X)^2] \ge 0$ for all polynomial q of degree at most d is the same as saying that $q^T H(y) q \ge 0$ 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 13.1. *$P_{2d}^* = \left\{(y_0,\dots,y_{2d}) \in \mathbb{R}^{2d+1} : H(y) \succeq 0\right\}$ 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} coincide with polynomials that are sums of squares, we have $y \in P_{2d}^*$ if and only if $\langle p, y\rangle \ge 0$ for all polynomials p of the form $p = q^2$ where q is an arbitrary polynomial of degree $\le d$. If $q(x) = \sum_{k=0}^d q_k x^k$ then the coefficients of the polynomial $p = q^2$ are $p_k = \sum_{0\le i,j\le d : i+j=k} q_i q_j$. Thus

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

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

Exercise 13.1. *Let $y = (y_0,\dots,y_{2d}) \in \mathbb{R}^{2d+1}$. Show that the solution to the following problem is either $-\infty$ or 0, and that the solution is 0 precisely when $y \in P_{2d}^*$:*

$$\underset{p\in\mathbb{R}^{2d+1}, M\in\mathbf{S}^{d+1}}{\text{minimise}} \langle p, y\rangle \quad \text{s.t.} \quad \sum_{\substack{0\le i,j\le d \\ i+j=k}} M_{ij} = p_k, M \succeq 0.$$

Using strong duality show that $y \in P_{2d}^*$ if and only if $H(y) \succeq 0$.

Exercise 13.2. *Let p be a polynomial of degree $2d$. Consider the problem*

$$\max \gamma \;\; \text{s.t.} \;\; p - \gamma \in P_{2d}. \tag{2}$$

We saw that the solution of this problem is equal to $\min_{x\in\mathbb{R}} p(x)$. Write the dual of (2) and compare it with $\min_{x\in\mathbb{R}} p(x)$. Give a simple argument why the dual problem you get is equal to $\min_{x\in\mathbb{R}} p(x)$.

Exercise 13.3 (Probability inequalities [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 $\mathbb{R}$.*

1. *Show that the following optimisation problem gives an* upper bound *on* $\Pr[X \in A]$*:*

$$\underset{p \in \mathbb{R}^{2d+1}}{\text{minimise}} \sum_{k=0}^{2d} p_k y_k \quad \text{subject to} \quad \sum_{k=0}^{2d} p_k x^k \geq \mathbf{1}_A(x) \quad \forall x \in \mathbb{R} \tag{3}$$

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

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

2. *Explain why* (3) *can be cast as a semidefinite program when* A *is an interval, or the complement of an interval.*

3. *Recall the inequality* $\Pr[|X| \geq t] \leq \mathbb{E}[X^2]/t^2$*. Show that the optimal value of* (3) *(assuming* $d \geq 1$*) for* $A = \mathbb{R} \setminus [-t, t]$ *will be smaller or equal than* $y_2/t^2 = \mathbb{E}[X^2]/t^2$*.*

Finding a measure associated to a sequence of moments We have shown that if $H(y) \succeq 0$ then y is the moment vector of a nonnegative measure on $\mathbb{R}$ (up to closure). Our proof however was not constructive: indeed in our proof of Theorem 12.3 we showed that if y is not a valid moment vector, then it is not in P_{2d}^*, via a separating hyperplane argument. A natural question is to know if there is an *algorithm* to construct a measure satisfying the moment constraints.

In general there will be many measures that satisfy the moment constraints. In the case where we want our measure to be atomic (i.e., a finite combination of Dirac masses) this problem is related to the *quadrature* problem in numerical integration. A typical quadrature problem asks for points $x_1, \ldots, x_N \in \mathbb{R}$ and weights $w_1, \ldots, w_N > 0$ such that

$$\int f(x) d\mu(x) = \sum_{i=1}^{N} w_i f(x_i)$$

holds for all polynomials f of degree at most D. Note that this is the same as saying that the moments of μ up to degree D agree with the moments of the atomic measure $\sum_{i=1}^{N} w_i \delta_{x_i}$. A quadrature algorithm which takes the moments of μ up to degree D and outputs the pairs (w_i, x_i) for $i = 1, \ldots, N$ would then solve our problem. We refer to [BPT12, Section 3.5.5] for more details on an algorithm to perform this.

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. 2

[BPT12] Grigoriy Blekherman, Pablo A. Parrilo, and Rekha R. Thomas. *Semidefinite optimization and convex algebraic geometry*. SIAM, 2012. 3