

## Lecture 12

Lecturer: Pablo A. Parrilo

Scribe: ???

## 1 Recovering a measure from its moments

We review next a classical method for producing a univariate atomic measure with a given set of moments. Essentially similar variations of this method are often used in signal processing, e.g., Pisarenko's harmonic decomposition method, where we are interested in producing a superposition of sinusoids with a given covariance matrix.

Consider the set of moments  $(\mu_0, \mu_1, \dots, \mu_{2n-1})$  for which we want to find an associated nonnegative measure, supported on the real line. The resulting measure will be discrete, of the form  $\sum_{i=1}^n w_i \delta(x - x_i)$ . For this, consider the linear system

$$\begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_{n-1} \\ \mu_1 & \mu_2 & \cdots & \mu_n \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n-1} & \mu_n & \cdots & \mu_{2n-2} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix} = - \begin{bmatrix} \mu_n \\ \mu_{n+1} \\ \vdots \\ \mu_{2n-1} \end{bmatrix}. \quad (1)$$

The Hankel matrix on the left-hand side of this equation is the one that appeared earlier as a sufficient condition for the moments to represent a nonnegative measure. The linear system in (1) has a unique solution if the matrix is positive definite. In this case, we let  $x_i$  be the roots of the univariate polynomial

$$x^n + c_{n-1}x^{n-1} + \cdots + c_1x + c_0 = 0,$$

which are all real and distinct (why?). We can then obtain the corresponding weights  $w_i$  by solving the nonsingular Vandermonde system given by

$$\sum_{i=1}^n w_i x_i^j = \mu_j \quad (0 \leq j \leq n-1).$$

In the exercises, you will have to prove that this method actually works (i.e., the  $x_i$  are real and distinct, the  $w_i$  are nonnegative, and the moments are the correct ones).

**Example 1.** Let's find a nonnegative measure whose first six moments are given by  $(1, 1, 2, 1, 6, 1)$ . The solution of the linear system (1) yields the polynomial

$$x^3 - 4x^2 - 9x + 16 = 0,$$

whose roots are  $-2.4265$ ,  $1.2816$ , and  $5.1449$ . The corresponding weights are  $0.0772$ ,  $0.9216$ , and  $0.0012$ , respectively.

## 2 A probabilistic interpretation

We also mention here an appealing probabilistic interpretation of the dual (2), commonly used in integer and quadratic programming, and developed by Lasserre in the polynomial case [Las01]. Consider as before the problem of minimizing a polynomial. Now, rather than looking for a minimizer  $x$  in  $\mathbb{R}^n$ , let's "relax" our notion of solution to allow for probabilities densities  $\mu$  on  $\mathbb{R}^n$ , and replace the objective function by its natural generalization  $\int p(x) d\mu$ . It clearly holds that the new objective is never larger than the original one, since we are making the feasible set bigger.

This change makes the problem trivially convex, although infinite-dimensional. To produce a finite dimensional approximation (which may or may not be exact), we rewrite the objective function in terms of the moments of the measure  $\mu$ , and write valid semidefinite constraints for the moments  $\mu_k$ .

### 3 Duality and complementary slackness

What is the relationship between this classical method and semidefinite programming duality? Recall our approach to minimizing a polynomial  $p(x)$  by computing

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

If this relaxation is exact (i.e., the optimal  $\gamma$  is equal to the optimal value of the polynomial) then at optimality, we necessarily have  $p(x_*) - \gamma_* = \sum_i g_i^2(x_*)$ . This implies that all the  $g_i$  vanish at the optimal point. We can thus obtain the optimal value by looking at the roots of the polynomials  $g_i(x)$ .

However, it turns out that if we are simultaneously solving the primal and the dual SDPs (as most modern interior point solvers) this is unnecessary, since from complementary slackness we can extract almost all the information needed. In particular, notice that if we have

$$p(x) - \gamma = [x]_d^T Q [x]_d = 0$$

then necessarily  $Q \cdot [x]_d = 0$ .

Recall the SDP formulation is given by

$$\max \gamma \quad \text{s.t.} \quad \begin{cases} p_0 - \gamma = Q_{00} \\ p_i = \sum_{j+k=i} Q_{jk} \quad i = 1, \dots, 2d \\ Q \succeq 0 \end{cases}$$

and its dual

$$\min \sum_{i=0}^{2d} p_i \mu_i \quad \text{s.t.} \quad M(\mu) := \begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_d \\ \mu_1 & \mu_2 & \cdots & \mu_{d+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_d & \mu_{d+1} & \cdots & \mu_{2d} \end{bmatrix} \succeq 0, \quad \mu_0 = 1. \quad (2)$$

At optimality, complementarity slackness holds, i.e., the product of the primal and dual matrices vanishes. We have then  $M(\mu) \cdot Q = 0$ . Assume that the leading  $k \times k$  submatrix of  $M(\mu)$  is nonsingular. Then, the procedure described in Section 1 gives a  $k$ -atomic measure, with support in the minimizers of  $p(x)$ . Generically, this matrix  $M(\mu)$  will be rank one, which will correspond to the case of a unique optimal solution.

**Remark 2.** *Unlike the univariate case, a multivariate polynomial that is bounded below may not achieve its minimum. A well-known example is  $p(x, y) = x^2 + (1 - xy)^2$ , which clearly satisfies  $p(x, y) \geq 0$ . Since  $p(x, y) = 0$  would imply  $x = 0$  and  $1 - xy = 0$  (which is impossible), this value cannot be achieved. However, we can get arbitrarily close, since  $p(\epsilon, 1/\epsilon) = \epsilon^2$ , for any  $\epsilon > 0$ .*

### 4 Multivariate case

We have seen previously that in the multivariate case, it is no longer the case that nonnegative polynomials are always sums of squares. The corresponding result on the dual side is that the set of valid moments is no longer described by the “obvious” semidefinite constraints, obtained by considering the expected value of squares (even if we require strict positivity).

**Example 3** (“Dual Motzkin”). *Consider the existence of a probability measure on  $\mathbb{R}^2$ , that satisfies the moment constraints:*

$$\begin{aligned} E[1] &= E[X^4 Y^2] = E[X^2 Y^4] = 1, \\ E[X^2 Y^2] &= 2, \\ E[XY] &= E[XY^2] = E[X^2 Y] = E[X^2 Y^3] = E[X^3 Y^2] = E[X^3 Y^3] = 0. \end{aligned} \quad (3)$$

The “obvious” nonnegativity constraints are satisfied, since

$$E[(a + bXY + cXY^2 + dX^2Y)^2] = a^2 + 2b^2 + c^2 + d^2 \geq 0.$$

However, it turns out that these conditions are only necessary, but not sufficient. This can be seen by computing the expectation of the Motzkin polynomial (which is nonnegative), since in this case we have

$$E[X^4Y^2 + X^2Y^4 + 1 - 3X^2Y^2] = 1 + 1 + 1 - 6 = -3,$$

thus proving that no measure with the given moments can exist.

## 5 Density results

Recent results by Blekherman [Ble04] give quantitative bounds on the relative density of the cone of sum of squares versus the cone of nonnegative polynomials. Concretely, in [Ble04] it is proved that a suitably normalized section of the cone of positive polynomials  $\tilde{P}_{n,2d}$  satisfies

$$c_1 n^{-\frac{1}{2}} \leq \left( \frac{\text{Vol } \tilde{P}_{n,2d}}{\text{Vol } B_M} \right)^{\frac{1}{D_M}} \leq c_2 n^{-\frac{1}{2}},$$

while the corresponding expression for the section of the cone of sum of squares  $\tilde{\Sigma}_{n,2d}$  is

$$c_3 n^{-\frac{d}{2}} \leq \left( \frac{\text{Vol } \tilde{\Sigma}_{n,2d}}{\text{Vol } B_M} \right)^{\frac{1}{D_M}} \leq c_4 n^{-\frac{d}{2}},$$

where  $c_1, c_2, c_3, c_4$  depend on  $d$  only (explicit expressions are available),  $D_M = \binom{n+2d}{2d} - 1$ , and  $B_M$  is the unit ball in  $\mathbb{R}^{D_M}$ .

These expressions show that for fixed  $d$ , as  $n \rightarrow \infty$  the volume of the set of sum of squares becomes vanishingly small when compared to the nonnegative polynomials.

Show the values of the actual bounds, for reasonable dimensions

ToDo

## References

- [Ble04] G. Blekherman. There are significantly more nonnegative polynomials than sums of squares. Preprint, 2004.
- [Las01] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817, 2001.