MIT 6.7230 - Algebraic techniques and semidefinite optimization March 20, 2024

Lecture 10

Lecturer: Pablo A. Parrilo *Scribe: ???*

In previous lectures, we described necessary conditions for the existence of a nonnegative measure with given moments. In the univariate case, these conditions were also sufficient. We revisit first a classical algorithm to effectively obtain this measure.

1 Recovering a measure from moments

We review next a classical method for producing a univariate atomic measure with a given set of moments (e.g., [ST43, Dev86]). Other similar variations of this method are commonly 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. This technique (or essentially similar ones) is known under a variety of names, such as Prony's method, or the Vandermonde decomposition of a Hankel 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. In general, there are infinitely many measures that will exact match those moments.

One possible approach, on which we will not elaborate here, is to pick the measure with *maximum entropy* that matches the given moments. This corresponds to the Gaussian case (if moments up to second order are given) or exponential families (in the general case).

The approach we follow here is to compute a *discrete* (finitely supported) measure, of the form $\sum_{i=1}^n w_i \delta(x - x_i)$. For this, consider the linear system

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

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

$$x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0 = 0.$$

These roots are all real and distinct (why?). Given the locations x_i, we can then obtain the corresponding weights w_i by solving the nonsingular Vandermonde system given by

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

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.*

Example 2. *We outline here a "stylized" application of these results. Consider a time-domain signal that is the sum of k Dirac functions, i.e.,* $f(x) := \sum_{i=1}^{k} w_i \delta(x - x_i)$, *where the* $2k$ *parameters* w_i, x_i *are unknown. By the results above, it is enough to obtain* $2k$ *linear functionals on the signal (namely, the moments* $\mu_i := \int x^i f(x) dx$*) to fully recover it from the measurements. Indeed, the signal can always be exactly reconstructed from these* $2k$ *moments, by using the algorithm described above. Notice that the nonnegativity assumption on the weights* w_i *is not critical, and can easily be removed.*

More realistic, but essentially similar results can be obtained by considering signals that are sums of (possibly damped) sinusoids of different frequencies. This viewpoint has a number of interesting connections with error-correcting codes (in particular, interpolation-based codes such as Reed-Solomon), as well as the recent "compressed sensing" results.

Remark 3. *As described, the measure recovery method described always works correctly, provided the computations are done in exact arithmetic. In most practical applications, it is necessary or convenient to use floating-point computations. Furthermore, in many settings such as optimization the moment information may be noisy, and therefore the matrices may contain some (hopefully small) perturbations from their nominal values. For these reasons, it is of interest to understand sensitivity issues, both at the level of what is intrinsic about the problem (conditioning), and about the specific algorithm used (numerical stability).*

As described, the technique described above can run into numerical difficulties. On the conditioning side, it is well-known that from the numerical viewpoint, the monomial basis (with respect to which we are taking moments) is a "bad" basis for the space of polynomials. On the numerical stability side, the algorithm above does a number of inefficient calculations, such as explicitly computing the coefficients c_i *of the polynomial corresponding to the support of the measure. A better approach involves directly computing the nodes* x_i *as the generalized eigenvalues of a matrix pencil. Some of these issues will be explored in more detail in the exercises.*

2 A probabilistic interpretation

We also mention here an appealing probabilistic interpretation of the dual (5), commonly used in integer and quadratic programming or game theory, and developed by Lasserre in the polynomial optimization case [Las01]. Consider as before the problem of minimizing a polynomial. Now, rather than looking directly for the minimizer $x_\star$ in $\mathbb{R}^n$, let's "relax" our notion of solution to allow for probabilities densities μ on $\mathbb{R}^n$, and replace the objective function by its natural generalization $\mathbf{E}_\mu[p(x)] = \int p(x) d\mu$. Since

$$\min_{\mu} \mathbf{E}_\mu[p(x)] \leq \mathbf{E}_{\delta_{x_\star}}[p(x)] = p(x_\star),$$

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 convex (trivially so), 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 μ, and write valid semidefinite contraints for the moments μ_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

$$\max \gamma \qquad \text{s.t.} \qquad p(x) - \gamma \quad \text{is SOS.} \tag{2}$$

The corresponding dual is

$$\min_{L} L[p] \qquad \text{s.t.} \qquad \begin{cases} L[q^2] \geq 0 \quad \forall q \in \mathbb{R}[x]_d, \\ L[1] = 1, \end{cases} \tag{3}$$

where $L[\cdot]$ is an element of the dual space $\mathbb{R}[x]^*_{2d}$ (a "Riesz functional" or "pseudoexpectation"). Here, L is constrained to be nonnegative on squares, and normalized. By choosing appropriate bases of $\mathbb{R}[x]$ and $\mathbb{R}[x]^*$, the constraints can be expressed as a standard SDP; see below for details.

On the primal side, the SDP formulation of the univariate optimization problem is given by

$$\max \gamma \qquad \text{s.t.} \qquad \begin{cases} Q_{00} + \gamma = p_0 \\ \displaystyle\sum_{j,k:j+k=i} Q_{jk} = p_i \qquad i = 1, \ldots, 2d \\ Q \succeq 0 \end{cases} \tag{4}$$

and its dual

$$\min \sum_{i=0}^{2d} p_i \mu_i \qquad \text{s.t.} \quad M(\mu) := \begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_d \\ \mu_1 & \mu_2 & \cdots & \mu_{d+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_d & \mu_{d+1} & \cdots & \mu_{2d} \end{bmatrix} \succeq 0, \quad \mu_0 = 1. \tag{5}$$

Notice the direct relationship between equations (2)-(4) and (3)-(5). Indeed, to obtain the corresponding SDPs, on the primal side we use a monomial basis and the representation $p(x) - \gamma = [x]_d^T Q [x]_d$, while on the dual we have $L[q] = \sum_{k=0}^{2d} c_k \mu_k$, where $q(x) = \sum_{k=0}^{2d} c_k x^k$.

When is the relaxation exact? If this relaxation is exact (i.e., the optimal γ is equal to the optimal value of the polynomial) then at optimality, we necessarily have $p(x_\star) - \gamma_\star = \sum_i g_i^2(x_\star)$. This implies that all the g_i must 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 $Q \cdot [x]_d = 0$.

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 4. *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 5 ("Dual Motzkin"). *Consider the existence of a probability measure on* $\mathbb{R}^2$*, that satisfies the moment constraints:*

$$\begin{aligned} E[1] = E[X^4Y^2] = E[X^2Y^4] = 1, \\ E[X^2Y^2] = 2, \\ E[XY] = E[XY^2] = E[X^2Y] = E[X^2Y^3] = E[X^3Y^2] = E[X^3Y^3] = 0. \end{aligned} \tag{6}$$

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 nonnegative measure with the given moments can exist.

5 Density results

Recent results by Blekherman [Ble06] give quantitative bounds on the relative density of the cone of sum of squares versus the cone of nonnegative polynomials. Concretely, in [Ble06] it is proved that a suitably normalized section of the cone of positive polynomials $\tilde{P}_{n,2d}$ satisfies

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

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

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

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 $n \to \infty$ 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

[Ble06] G. Blekherman. There are significantly more nonnegative polynomials than sums of squares. *Israel Journal of Mathematics*, 153(1):355–380, 2006.

[Dev86] L. Devroye. *Nonuniform random variate generation.* Springer-Verlag, New York, 1986.

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

[ST43] J.A. Shohat and J.D. Tamarkin. *The Problem of Moments.* American Mathematical Society Mathematical surveys, vol. II. American Mathematical Society, New York, 1943.