MIT 6.7230 - Algebraic techniques and semidefinite optimization March 1, 2024

Lecture 6

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

Last week we learned about explicit conditions to determine the number of real roots of a univariate polynomial. Today we will expand on these themes, and study two mathematical objects of fundamental importance: the *resultant* of two polynomials, and the closely related *discriminant*.

The resultant will be used to decide whether two univariate polynomials have common roots, while the discriminant will give information about the existence of multiple roots. Furthermore, we will see the intimate connections between discriminants and the boundary of the cone of nonnegative polynomials.

Besides the properties described above, a direct consequence of their definitions, there are many other interesting applications of resultants and discriminant. We describe a few of them below, and we will encounter them again in later lectures, when studying elimination theory and the construction of cylindrical algebraic decompositions. For much more information about resultants and discriminants, particularly their generalizations to the sparse and multipolynomial case, we refer the reader to the very readable introductory article [Stu98] and the books [CLO97, GKZ94].

1 Resultants

Consider two polynomials $p(x)$ and $q(x)$, of degree n, m, respectively. We want to obtain an easily checkable criterion to determine whether they have a common root, that is, there exists an $x_0 \in \mathbb{C}$ for which $p(x_0) = q(x_0) = 0$. There are several approaches, seemingly different at first sight, for constructing such a criterion:

- **Sylvester matrix**: If $p(x_0) = q(x_0) = 0$, then we can write the following $(n+m) \times (n+m)$ linear system:

$$\begin{bmatrix} p_n & p_{n-1} & \cdots & p_1 & p_0 & & & & \\ & p_n & & \ddots & \ddots & \ddots & & & \\ \vdots & & & & & & & & \\ & & & & & & p_1 & p_0 & \\ & & & & & & p_2 & p_1 & p_0 \\ q_m & q_{m-1} & \cdots & q_0 & & & & & \\ & q_m & & \ddots & \ddots & & & & \\ & & \ddots & & & & & & \\ & & & & & & q_1 & q_0 & \\ & & & & & & q_2 & q_1 & q_0 \end{bmatrix} \begin{bmatrix} x_0^{n+m-1} \\ x_0^{n+m-2} \\ \vdots \\ x_0^n \\ x_0^{n-1} \\ \vdots \\ x_0 \\ 1 \end{bmatrix} = \begin{bmatrix} p(x_0)x_0^{m-1} \\ p(x_0)x_0^{m-2} \\ \vdots \\ p(x_0)x_0 \\ p(x_0) \\ q(x_0)x_0^{n-1} \\ q(x_0)x_0^{n-2} \\ \vdots \\ q(x_0)x_0 \\ q(x_0) \end{bmatrix} = 0.$$

  This implies that the matrix on the left-hand side, called the *Sylvester matrix* $\mathrm{Syl}_x(p,q)$ associated to p and q, is singular and thus its determinant must vanish. It is not too difficult to show that the converse is also true; if $\det \mathrm{Syl}_x(p,q) = 0$, then there exists a vector in the kernel of $\mathrm{Syl}_x(p,q)$ of the form shown in the equation above, and thus a common root x_0.

- **Root products and companion matrices**: Let α_j, β_k be the roots of $p(x)$ and $q(x)$, respectively. By construction, the expression

$$\prod_{j=1}^{n}\prod_{k=1}^{m}(\alpha_j - \beta_k)$$

vanishes if and only if there exists a root of p that is equal to a root of q. Although the computation of this product seems to require explicit access to the roots, this can be avoided. Multiplying by a convenient normalization factor, we have:

$$\begin{aligned} p_n^m q_m^n \prod_{j=1}^{n}\prod_{k=1}^{m}(\alpha_j - \beta_k) &= p_n^m \prod_{j=1}^{n} q(\alpha_j) = p_n^m \det q(\mathcal{C}_p) \\ &= (-1)^{nm} q_m^n \prod_{k=1}^{m} p(\beta_k) = (-1)^{nm} q_m^n \det p(\mathcal{C}_q) \end{aligned} \tag{1}$$

- **Kronecker products**: Using a well-known connection to Kronecker products, we can also write (1) as

$$p_n^m q_m^n \det(\mathcal{C}_p \otimes I_m - I_n \otimes \mathcal{C}_q).$$

- **Bézout matrix:** Given $p(x)$ and $q(x)$ as before, consider the bivariate function

$$B(s,t) := \frac{p(s)q(t) - p(t)q(s)}{s-t}.$$

It is easy to see that this is actually a polynomial in the variables s, t, and is invariant under the interchange $s \leftrightarrow t$. Let $d := \max(n, m)$, and $\mathrm{Bez}_x(p, q)$ be the symmetric $d \times d$ matrix that represents this polynomial in the standard monomial basis, i.e.,

$$B(s,t) = \begin{bmatrix} 1 \\ s \\ \vdots \\ s^{d-1} \end{bmatrix}^T \mathrm{Bez}_x(p,q) \begin{bmatrix} 1 \\ t \\ \vdots \\ t^{d-1} \end{bmatrix}.$$

The Bézout matrix is singular if and only p and q have a common root.

Notice the differences with the Sylvester matrix: while that approach requires a non-symmetric $(n+m) \times (n+m)$ matrix depending linearly on the coefficients, in the Bézout approach the matrix is smaller and symmetric, but with entries that depend bilinearly on the p_i, q_i.

If can be shown that all these constructions are equivalent. They define exactly the same polynomial, called the *resultant* of p and q, denoted as $\mathrm{Res}_x(p, q)$:

$$\begin{aligned} \mathrm{Res}_x(p,q) &= \det \mathrm{Syl}_x(p,q) \\ &= p_n^m \det q(\mathcal{C}_p) \\ &= (-1)^{nm} q_m^n \det p(\mathcal{C}_q) \\ &= p_n^m q_m^n \det(\mathcal{C}_p \otimes I_m - I_n \otimes \mathcal{C}_q) \\ &= \frac{(-1)^{\binom{n}{2}}}{p_n^{n-m}} \det \mathrm{Bez}_x(p,q). \end{aligned}$$

The resultant is a homogeneous multivariate polynomial, with integer coefficients, and of degree $n+m$ in the $n+m+2$ variables p_j, q_k. It vanishes if and only if the polynomials p and q have a common root. Notice that the definition is not symmetric in its two arguments, $\mathrm{Res}_x(p,q) = (-1)^{nm}\mathrm{Res}(q,p)$ (of course, this does not matter in checking whether it is zero).

Remark 1. *To compute the resultant of two polynomials $p(x)$ and $q(x)$ in Maple, you can use the command* `resultant(p,q,x)`. *In Mathematica, use instead* `Resultant[p,q,x]`.

2 Discriminants

As we have seen, the resultant allows us to write an easily checkable condition for the simultaneous vanishing of two univariate polynomials. Can we use the resultant to produce a condition for a polynomial to have a double root? Recall that if a polynomial $p(x)$ has a double root at x_0 (which can be real or complex), then its derivative $p'(x)$ also vanishes at x_0. Thus, we can check for the existence of a root of multiplicity two (or higher) by computing the resultant betweeen a polynomial and its derivative.

Definition 2. *The* discriminant *of a univariate polynomial $p(x)$ is defined as*

$$\mathrm{Dis}_x(p) := (-1)^{\binom{n}{2}} \frac{1}{p_n} \mathrm{Res}_x\left(p(x), \frac{dp(x)}{dx}\right).$$

Similarly to what we did in the resultant case, the discriminant can also be obtained by writing a natural condition in terms of the roots α_i of $p(x)$:

$$\mathrm{Dis}_x(p) = p_n^{2n-2} \prod_{j<k} (\alpha_j - \alpha_k)^2.$$

If $p(x)$ has degree n, its discriminant is a homogeneous polynomial of degree $2n-2$ in its $n+1$ coefficients $p_n, \ldots, p_0$.

Example 3. *Consider the quadratic univariate polynomial $p(x) = ax^2 + bx + c$. Its discriminant is:*

$$\mathrm{Dis}_x(p) = -\frac{1}{a}\mathrm{Res}_x(ax^2+bx+c, 2ax+b) = b^2 - 4ac.$$

For the cubic polynomial $p(x) = ax^3 + bx^2 + cx + d$ we have

$$\mathrm{Dis}_x(p) = -27a^2d^2 + 18adcb + b^2c^2 - 4b^3d - 4ac^3.$$

3 Applications

3.1 Polynomial equations

One of the most natural applications of resultants is in the solution of polynomial equations in two variables. For this, consider a polynomial system

$$p(x,y) = 0, \qquad q(x,y) = 0, \tag{2}$$

with only a finite number of solutions (which is generically the case). Consider a fixed value of y_0, and the two univariate polynomials $p(x,y_0), q(x,y_0)$. If y_0 corresponds to the y-component of

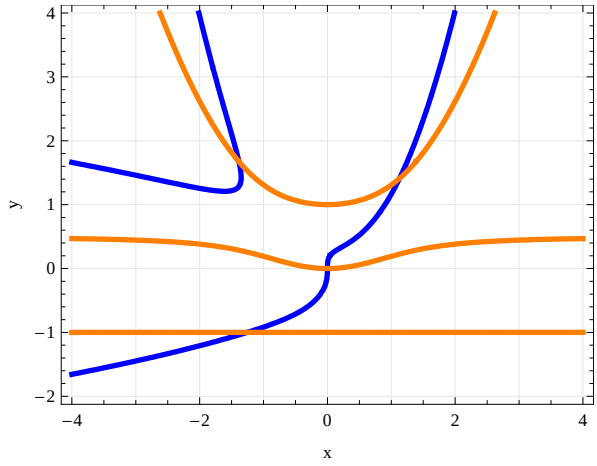

Figure 1: Zero sets of the polynomials $p(x, y)$ and $q(x, y)$ in Example 4.

a root, then these two univariate polynomials clearly have a common root, hence their resultant vanishes.

Therefore, to solve (2), we can compute $\mathrm{Res}_x(p, q)$, which is a univariate polynomial in y. Solving this univariate polynomial, we obtain a finite number of points y_i. Backsubstituting in p (or q), we obtain the corresponding values of x_i. Naturally, the same construction can be used by computing first the univariate polynomial in x given by $\mathrm{Res}_y(p, q)$.

Example 4. *Let* $p(x, y) = 2xy + 3y^3 - 2x^3 - x - 3x^2y^2$, *and* $q(x, y) = 2x^2y^2 - 4y^3 - x^2 + 4y + x^2y$. *The corresponding zero sets are shown in Figure 1. The resultant (with respect to the* x *variable) is*

$$\mathrm{Res}_x(p, q) = y(y + 1)^3(72y^8 - 252y^7 + 270y^6 - 145y^5 + 192y^4 - 160y^3 + 28y + 4).$$

One particular root of this polynomial is $y_\star \approx 1.6727$, *with the corresponding value of* $x_\star \approx -1.3853$.

3.2 Implicitization of plane rational curves

Consider a plane curve parametrized by rational functions, i.e.,

$$x(t) = \frac{p_1(t)}{q_1(t)}, \qquad y(t) = \frac{p_2(t)}{q_2(t)}.$$

What is the implicit equation of the curve, i.e., what constraint $h(x, y) = 0$ must the points $(x, y) \in \mathbb{R}^2$ that lie on the curve satisfy? The corresponding equation can be easily obtained by computing a resultant to eliminate the parametrizing variable t, i.e.,

$$h(x, y) = \mathrm{Res}_t(q_1(t) \cdot x - p_1(t), q_2(t) \cdot y - p_2(t)).$$

Example 5. *Consider the curve described by the parametrization.*

$$x(t) = \frac{t(1 + t^2)}{1 + t^4}, \qquad y(t) = \frac{t(1 - t^2)}{1 + t^4}. \tag{3}$$

Its implicit equation can be computed by the resultant:

$$\mathrm{Res}_t((1 + t^4)x - t(1 + t^2), (1 + t^4)y - t(1 - t^2)) = 4y^4 + 8y^2x^2 + 4x^4 + 4y^2 - 4x^2.$$

Remark 6. *The inverse problem (given an implicit polynomial equation for a curve, find a rational parametrization) is not always solvable. In fact, there is a full characterization of when this is possible, in terms of a topological invariant of the curve called the* genus *(the rationally parametrizable curves are exactly those of genus zero). For example, the curve* $x^4 + y^4 = 1$ *does* not *admit a polynomial parametrization, since it has genus 3.*

3.3 Eigenvalue distribution of random matrices

This section is based on the results in [RE08]. The eigenvalues of a random symmetric matrix belonging to a given ensemble can be characterized in terms of the asymptotic eigenvalue distribution $F(x)$ (e.g., the semi-circle law, Marčenko-Pastur, etc). Often, rather than the actual distribution, it is more convenient to use instead some other equivalent object, such as its moment generating function, Stieltjes transform, R-transform, etc. For many ensembles of interest, these auxiliary transforms $\tilde{F}(z)$ are algebraic functions, in the sense that they satisfy an equation of the form $\psi(\tilde{F}(z), z) = 0$, where $\psi(s,t)$ is a bivariate polynomial, and furthermore they can all be derived from each other. As a consequence, to each given random ensemble of this class we can associate a bivariate polynomial that uniquely describes the limiting eigenvalue distribution.

A natural question arises: given two matrices M_1, M_2, belonging to random ensembles with associated polynomials $\psi_1(s,t)$ and $\psi_2(s,t)$, what can be said about the eigenvalue distribution of the sum $M_1 + M_2$ (or the product $M_1 M_2$)? Voiculescu has shown that under a certain natural independence condition ("freeness"), the R-transform of the sum is the sum of the individual transforms (this is somewhat akin to the well-known fact that the pdf of the sum of independent random variables is the convolution of the individual pdfs, or the additivity of the moment generating function). Under the freeness condition, the bivariate polynomial associated with the ensemble $M_3 = M_1 + M_2$ can be computed from the individual polynomials ψ_1, ψ_2 via:

$$\psi_3(s,t) = \mathrm{Res}_u(\psi_1(s-u,t), \psi_2(u,t)).$$

Similar expressions are also possible for the product $M_1 M_2$, also in terms of resultants. This allows the computation of the spectra of arbitrary random ensembles, that can be built from individual "building blocks" with known eigenvalue distributions.

We cannot provide a full description here of this area, and the very interesting connections with "free probability." We refer the reader to [RE08] for a more complete account.

4 The set of nonnegative polynomials

One of the main reasons why nonnegativity conditions about polynomials are difficult is because these sets can have a quite complicated structure, even though they are always convex.

Recall from last lecture that we have defined $P_n \subset \mathbb{R}^{n+1}$ as the set of nonnegative polynomials of degree n. It is easy to see that if $p(x)$ lies on the boundary of the set P_n, then it must have a real root, of multiplicity at least two. Indeed, if there is no real root, then $p(x)$ is in the strict interior of P (small enough perturbations will not create a root), and if it has a simple real root it clearly cannot be nonnegative.

Thus, on the boundary of P_n, the discriminant of $p(x)$ must necessarily vanish. However, it turns out that $\mathrm{Dis}_x(p)$ does not vanish *only* on the boundary, but it also vanishes at points inside the set. Why is this?

Example 7. *Consider the univariate polynomial* $p(x) = x^4 + 2ax^2 + b$*. For what values of* a, b *does it hold that* $p(x) \geq 0 \, \forall x \in \mathbb{R}$*? Since the leading term* x^4 *has even degree and is strictly positive,*

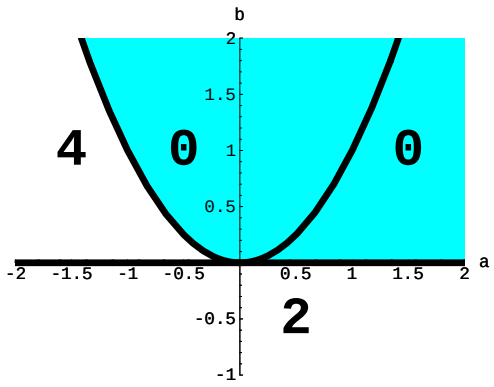

Figure 2: The shaded region corresponds to the values of (a, b) for which the polynomial $x^4 + 2ax^2 + b$ is nonnegative. The numbers indicate how many real roots the polynomial has.

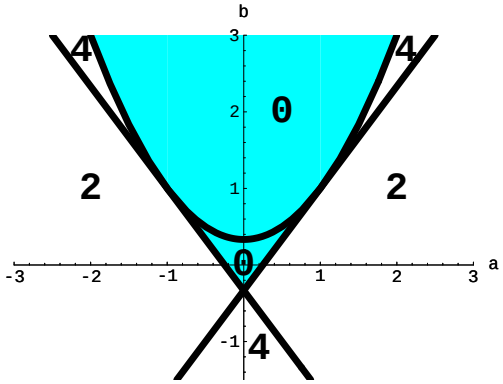

Figure 3: Region of nonnegativity of the polynomial $x^4 + 4ax^3 + 6bx^2 + 4ax + 1$, and number of real roots.

$p(x)$ is strictly positive if and only if it has no real roots. The discriminant of $p(x)$ is equal to $256\, b\, (a^2 - b)^2$. The set of (a, b) for which $p(x)$ is nonnegative is shown in Figure 2.

Here is a slightly different example, showing the same phenomenon.

Example 8. *Consider now the polynomial $p(x) = x^4 + 4ax^3 + 6bx^2 + 4ax + 1$. Its discriminant, in factored form, is equal to $256(1 + 3b + 4a)(1 + 3b - 4a)(1 + 2a^2 - 3b)^2$. The corresponding nonnegativity region and number of real roots are presented in Figure 3.*

As we can see, this creates some difficulties. For instance, even though we have a perfectly valid analytic expression for the boundary of the set, we cannot get a good sense of "how far we are" from the boundary by looking at the absolute value of the discriminant.

From the mathematical viewpoint, there are a couple of (unrelated?) reasons with these sets cannot be directly handled by "standard" optimization, at least if we want to keep the polynomial structure. One has to do with its algebraic structure, and the other one with convexity, and in particular its facial structure.

Lemma 9 (e.g., [And03])**.** *The set described in Figure 2 is not basic closed semialgebraic.*

Remark 10. *Notice that the convex sets described in Figures 2 and 3 both have an uncommon feature. They both have proper faces that are not* exposed*, i.e., they cannot be isolated by a supporting*

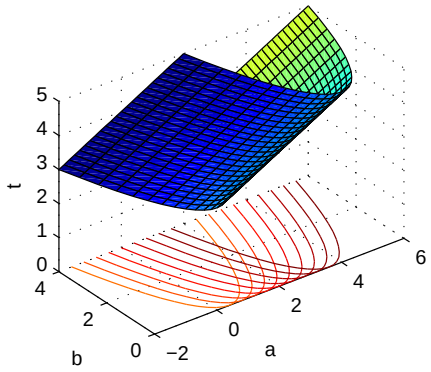

Figure 4: A three-dimensional convex set, described by one quadratic and one linear inequality, whose projection on the (a, b) plane is equal to the set in Figure 2.

hyperplane[1]*. Indeed, in Figure 2 the origin* $(0, 0)$ *is a non-exposed zero-dimensional face, while in Figure 3 the point* $(1, 1)$ *has the same property. A non-exposed face is a known obstruction for a convex set to be the feasible set of a semidefinite program, see [RG95].*

Even though these sets have these complicating features, it turns out that we can often provide some "good" representations. These are normally given as a *projection* from higher dimensional spaces, where the object "upstairs" is much more smooth and well-behaved. For instance, as a graphical illustration, in Figure 4 we can see the three-dimensional convex set $\{(a, b, t) \in \mathbb{R}^3 : b \geq (a - t)^2,\ t \geq 0\}$, whose projection on the plane (a, b) is exactly the set discussed in Example 7 and Figure 2.

The presence of "extraneous" components of the discriminant inside the feasible set is an important roadblock for the availability of "easily computable" barrier functions. Indeed, every polynomial that vanishes on the boundary of the set P_n must necessarily have the discriminant as a factor. This is an striking difference with the case of the case of the nonnegative orthant or the PSD cone, where the standard barriers are given (up to a logarithm) by products of the linear constraints or a determinant (which are polynomials). The way out of this problem is to produce non-polynomial barrier functions, either by partial minimization from a higher-dimensional barrier (i.e., projection) or other constructions such as the "universal" barrier function introduced by Nesterov and Nemirovski [NN94].

In this direction, there have been several research efforts that aim at directly characterizing barrier functions for the set of nonnegative polynomials (or related modifications). Among them, we mention the work of Kao and Megretski [KM02] and Faybusovich [Fay02], both of which produce barriers that rely on the computation of one or more integral expressions. Given the fact that these integrals must be computed numerically, there is no clear consensus yet on how useful this approach is in practical optimization problems.

References

[And03] C. Andradas. Characterization and description of basic semialgebraic sets. In *Algorithmic and quantitative real algebraic geometry (Piscataway, NJ, 2001)*, volume 60 of *DIMACS*

[1] A *face* of a convex set S is a convex subset $F \subseteq S$, with the property that $x, y \in S, \frac{1}{2}(x + y) \in F \Rightarrow x, y \in F$. A face F is *exposed* if it can be written as $F = S \cap H$, where H is a supporting hyperplane of S.

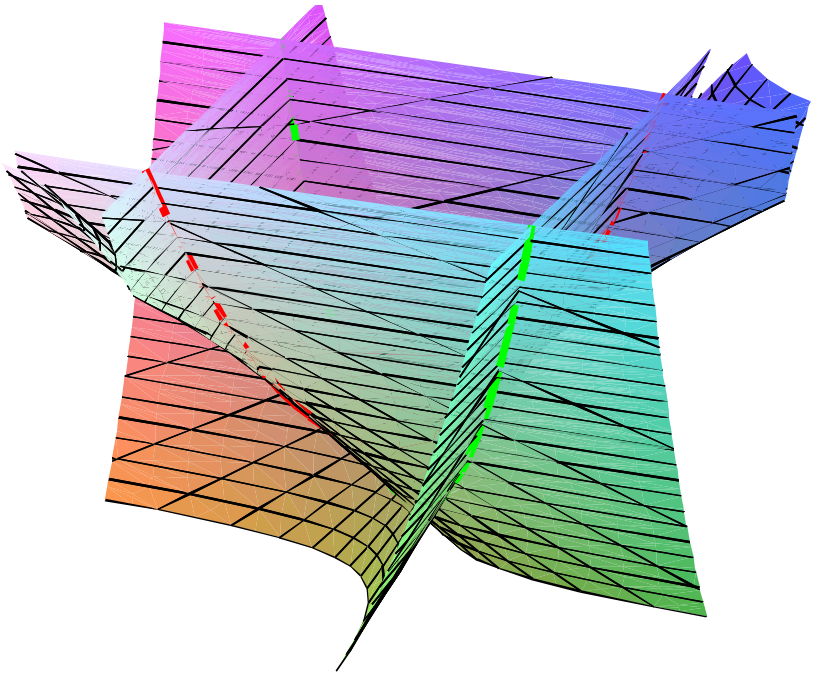

Figure 5: The discriminant of the polynomial $x^4 + 4ax^3 + 6bx^2 + 4cx + 1$. The convex set inside the "bowl" corresponds to the region of nonnegativity. There is an additional one-dimensional component inside the set.

Ser. Discrete Math. Theoret. Comput. Sci., pages 1–12. Amer. Math. Soc., Providence, RI, 2003.

[CLO97] D. A. Cox, J. B. Little, and D. O'Shea. *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Springer, 1997.

[Fay02] L. Faybusovich. Self-concordant barriers for cones generated by Chebyshev systems. *SIAM J. Optim.*, 12(3):770–781, 2002.

[GKZ94] I. M. Gel'fand, M. Kapranov, and A. Zelevinsky. *Discriminants, Resultants, and Multidimensional Determinants*. Birkhäuser, 1994.

[KM02] C. Y. Kao and A. Megretski. A new barrier function for IQC optimization problems. In *American Control Conference*, 2002.

[NN94] Y. E. Nesterov and A. Nemirovski. *Interior point polynomial methods in convex programming*, volume 13 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, 1994.

[RE08] N. Raj Rao and Alan Edelman. The polynomial method for random matrices. *Found. Comput. Math.*, 8(6):649–702, 2008.

[RG95] M. Ramana and A. J. Goldman. Some geometric results in semidefinite programming. *J. Global Optim.*, 7(1):33–50, 1995.

[Stu98] B. Sturmfels. Introduction to resultants. In *Applications of computational algebraic geometry (San Diego, CA, 1997)*, volume 53 of *Proc. Sympos. Appl. Math.*, pages 25–39. Amer. Math. Soc., Providence, RI, 1998.