MIT 6.256 - Algebraic techniques and semidefinite optimization February 24, 2010

Lecture 5

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

In this lecture we study univariate polynomials, particularly questions regarding the existence of real roots and nonnegativity conditions. For instance:

- When does a univariate polynomial have *only* real roots?
- What conditions must it satisfy for *all* roots to be real?
- When is a polynomial nonnegative, i.e., it satisfies $p(x) \geq 0$ for all $x \in \mathbb{R}$?

1 Univariate polynomials

A univariate polynomial $p(x) \in \mathbb{R}[x]$ of degree n has the form:

$$p(x) = p_n x^n + p_{n-1}x^{n-1} + \cdots + p_1 x + p_0, \tag{1}$$

where the coefficients p_k are real. We normally assume $p_n \neq 0$, and occasionally we will normalize it to $p_n = 1$, in which case we say that $p(x)$ is *monic.*

As we have seen, the field $\mathbb{C}$ of complex numbers is algebraically closed:

Theorem 1 (Fundamental theorem of algebra)**.** *Every nonzero univariate polynomial of degree n has exactly n complex roots (counted with multiplicity). Furthermore, we have the unique factorization*

$$p(x) = p_n \prod_{k=1}^{n} (x - x_k),$$

where $x_k \in \mathbb{C}$ are the roots *of $p(x)$.*

If all the coefficients p_k are real, if x_k is a root, then so its complex conjugate x_k^*. In other words, all complex roots appear in complex conjugate pairs.

2 Counting real roots

How many *real roots* does a polynomial have? There are many options, ranging from all roots being real (e.g., $(x-1)(x-2)\ldots(x-n)$), to all roots being complex (e.g., $x^{2d}+1$). We will give a couple of different characterizations of the location of the roots of a polynomial, both of them in terms of some associated symmetric matrices.

2.1 The companion matrix

A very well-known relationship between univariate polynomials and matrices is given through the so-called companion matrix.

Definition 2. *The* companion matrix $\mathcal{C}_p$ *associated with the polynomial $p(x)$ in (1) is the $n \times n$ real matrix*

$$\mathcal{C}_p := \begin{bmatrix} 0 & 0 & \cdots & 0 & -p_0/p_n \\ 1 & 0 & \cdots & 0 & -p_1/p_n \\ 0 & 1 & \cdots & 0 & -p_2/p_n \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -p_{n-1}/p_n \end{bmatrix}.$$

Lemma 3. *The characteristic polynomial of* $\mathcal{C}_p$ *is (up to a constant) equal to* $p(x)$*. Formally,* $\det(xI - \mathcal{C}_p) = \frac{1}{p_n} p(x)$*.*

From this lemma, it directly follows that the eigenvalues of $\mathcal{C}_p$ are exactly equal to the roots x_i of $p(x)$, including multiple roots the appropriate number of times. In other words, if we want to obtain the roots of a polynomial, we can do this by computing instead the eigenvalues of the associated (nonsymmetric) companion matrix. In fact, that is exactly the way that MATLAB computes roots of polynomials; see the source file `roots.m`.

Left and right eigenvectors The companion matrix $\mathcal{C}_p$ is diagonalizable if and only the polynomial $p(x)$ has no multiple roots. What are the corresponding diagonalizing matrices (equivalently, the right and left eigenvectors)?

Define the $n \times n$ *Vandermonde matrix*

$$V = \begin{bmatrix} 1 & x_1 & \dots & x_1^{n-1} \\ 1 & x_2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^{n-1} \end{bmatrix} \tag{2}$$

where $x_1, \dots, x_n \in \mathbb{C}$. It can be shown that the matrix V is nonsingular if and only if all the x_i are distinct. We have then the identity

$$V \cdot \mathcal{C}_p = \operatorname{diag}[x_1, \dots, x_n] \cdot V, \tag{3}$$

and thus the left eigenvectors of $\mathcal{C}_p$ are the rows of the Vandermonde matrix.

The right eigenvectors are of course given by the columns of V^{-1}, as can be easily seen by left- and right-multiplying (3) by this inverse. A natural interpretation of this dual basis (i.e., the columns of V^{-1}) is in terms of the *Lagrange interpolating polynomials* of the points x_i. These are a set of n univariate polynomials that satisfy the property $L_j(x_i) = \delta_{ij}$, where δ is the Kronecker delta. It is easy to verify that the columns of V^{-1} are the coefficients (in the monomial basis) of the corresponding Lagrange interpolating polynomials.

Example 4. *Consider the polynomial* $p(x) = (x-1)(x-2)(x-5)$*. Its companion matrix is*

$$\mathcal{C}_p = \begin{bmatrix} 0 & 0 & 10 \\ 1 & 0 & -17 \\ 0 & 1 & 8 \end{bmatrix},$$

and it is diagonalizable since p *has simple roots. Ordering the roots as* $\{1, 2, 5\}$*, the corresponding Vandermonde matrix and its inverse are:*

$$V = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 5 & 25 \end{bmatrix}, \qquad V^{-1} = \frac{1}{12} \begin{bmatrix} 30 & -20 & 2 \\ -21 & 24 & -3 \\ 3 & -4 & 1 \end{bmatrix}.$$

From the columns of V^{-1}*, we can read the coefficients of the Lagrange interpolating polynomials; e.g.,* $L_1(x) = (30 - 21x + 3x^2)/12 = (x-2)(x-5)/4$*.*

Symmetric functions of roots For any $A \in \mathbb{C}^{n \times n}$, we always have $\operatorname{Tr} A = \sum_{i=1}^n \lambda_i(A)$, and $\lambda_i(A^k) = \lambda_i(A)^k$. Therefore, it follows that $\sum_{i=1}^n x_i^k = \operatorname{Tr}[\mathcal{C}_p^k]$. As a consequence of linearity, we have that if $q(x) = \sum_{j=0}^m q_j x^j$ is a univariate polynomial,

$$\sum_{i=1}^n q(x_i) = \sum_{i=1}^n \sum_{j=0}^m q_j x_i^j = \sum_{j=0}^m q_j \operatorname{Tr}[\mathcal{C}_p^j] = \operatorname{Tr}[\sum_{j=0}^m q_j \mathcal{C}_p^j] = \operatorname{Tr}[q(\mathcal{C}_p)], \tag{4}$$

where the expression $q(\mathcal{C}_p)$ indicates the evaluation of the polynomial $q(x)$ on the companion matrix of $p(x)$. Note that if p is monic, then the final expression in (4) is a polynomial in the coefficients of p. This is an identity that we will use several times in the sequel.

Remark 5. *Our presentation of the companion matrix has been somewhat unmotivated, other than noticing that "it just works." After presenting some additional material on Gröbner bases, we will revisit this construction, where we will give a natural interpretation of $\mathcal{C}_p$ as representing a well-defined linear operator in the quotient ring $\mathbb{R}[x]/\langle p(x)\rangle$. This will enable a very appealing extension of many results about companion matrices to multivariate polynomials, in the case where the underlying system has only a finite number of solutions (i.e., a "zero dimensional ideal"). For instance, the generalization of the diagonalizability of the companion matrix $\mathcal{C}_p$ when $p(x)$ has only simple roots will be the fact that the multiplication algebra associated with a zero-dimensional ideal is semisimple if and only if the ideal is radical.*

2.2 Inertia and signature

Definition 6. *Consider a symmetric matrix A. The* inertia *of A, denoted $\mathcal{I}(A)$, is the integer triple (n_+, n_0, n_-), where n_+, n_0, n_- are the number of positive, zero, and negative eigenvalues, respectively. The* signature *of A is equal to the number of positive eigenvalues minus the number of negative eigenvalues, i.e., the integer $n_+ - n_-$.*

Notice that, with the notation above, the rank of A is equal to $n_+ + n_-$. A symmetric positive definite $n \times n$ matrix has inertia $(n, 0, 0)$, while a positive semidefinite one has $(n-k, k, 0)$ for some $k \geq 0$. The inertia is an important invariant of a quadratic form, since it holds that $\mathcal{I}(A) = \mathcal{I}(T^*AT)$, where T is nonsingular. This invariance of the inertia of a matrix under congruence transformations is known as *Sylvester's law of inertia*; see for instance [HJ95].

2.3 The Hermite form

While the companion matrix is quite useful, we will present now a different characterization of the roots of a polynomial. Among others, an advantage of this formulation is the fact that we will be using *symmetric* matrices.

Let $q(x)$ be a fixed auxiliary polynomial. Consider the following $n \times n$ symmetric Hankel matrix $H_q(p)$ with entries defined by

$$[H_q(p)]_{jk} = \sum_{i=1}^{n} q(x_i) x_i^{j+k-2}. \tag{5}$$

Like every symmetric matrix, $H_q(p)$ defines an associated quadratic form via

$$\begin{aligned} f^T H_q(p) f &= \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_{n-1} \end{bmatrix}^T \begin{bmatrix} \sum_{i=1}^n q(x_i) & \sum_{i=1}^n q(x_i)x_i & \cdots & \sum_{i=1}^n q(x_i)x_i^{n-1} \\ \sum_{i=1}^n q(x_i)x_i & \sum_{i=1}^n q(x_i)x_i^2 & \cdots & \sum_{i=1}^n q(x_i)x_i^{n} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n q(x_i)x_i^{n-1} & \sum_{i=1}^n q(x_i)x_i^{n} & \cdots & \sum_{i=1}^n q(x_i)x_i^{2n-2} \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_{n-1} \end{bmatrix} \\ &= \sum_{i=1}^{n} q(x_i)(f_0 + f_1 x_i + \cdots + f_{n-1}x_i^{n-1})^2 \\ &= \mathrm{Tr}[(qf^2)(\mathcal{C}_p)]. \end{aligned}$$

Although not immediately obvious from the definition (5), the expression above shows that when $p(x)$ is monic, the entries of $H_q(p)$ are actually polynomials in the coefficients of $p(x)$. Notice that we have used (4) in the derivation of the last step.

Recall that a Vandermonde matrix defines a linear transformation mapping the coefficients of a degree $n-1$ polynomial f to its values $\big[f(x_1), \dots, f(x_n)\big]$. Since this transformation is invertible, given any $y \in \mathbb{R}^n$ there always exists an f of degree $n-1$ such that $f(x_i) = y_i$ (i.e., there is always an interpolating polynomial). From expression (5), we have the factorization

$$H_q(p) = V^T \operatorname{diag}[q(x_1), \dots, q(x_n)]\, V.$$

This is almost a congruence transformation, except that there are complex entries in V if some of the x_i are complex. However, this can be easily resolved, to obtain the theorem below.

Theorem 7. *The signature of $H_q(p)$ is equal to the number of real roots x_j of p for which $q(x_j) > 0$, minus the number of real roots for which $q(x_j) < 0$.*

Proof. For simplicity, we assume all roots are distinct (this is easy to change, at the expense of slightly more complicated notation). We have then

$$\begin{aligned} f^T H_q(p) f &= \sum_{j=1}^{n} q(x_j)(f_0 + f_1 x_j + \cdots + f_{n-1} x_j^{n-1})^2 \\ &= \sum_{x_j \in \mathbb{R}} q(x_j) f(x_j)^2 + \sum_{x_j, x_j^* \in \mathbb{C}\setminus\mathbb{R}} q(x_j) f(x_j)^2 + q(x_j^*) f(x_j^*)^2 \\ &= \sum_{x_j \in \mathbb{R}} q(x_j) f(x_j)^2 + 2 \sum_{x_j, x_j^* \in \mathbb{C}\setminus\mathbb{R}} \begin{bmatrix} \Re f(x_j) \\ \Im f(x_j) \end{bmatrix}^T \begin{bmatrix} \Re q(x_j) & -\Im q(x_j) \\ -\Im q(x_j) & -\Re q(x_j) \end{bmatrix} \begin{bmatrix} \Re f(x_j) \\ \Im f(x_j) \end{bmatrix}. \end{aligned}$$

Notice that an expression of the type $f(x_i)$ is a linear form in $[f_0, \dots, f_{n-1}]$. Because of the assumption that all the roots x_j are distinct, the linear forms $\{f(x_j)\}_{j=1,\dots,n}$ are linearly independent (the corresponding Vandermonde matrix is nonsingular), and thus so are $\{f(x_j)\}_{x_j \in \mathbb{R}} \cup \{\Re f(x_j), \Im f(x_j)\}_{x_j \in \mathbb{C}\setminus\mathbb{R}}$. Therefore, the expression above gives a congruence transformation of $H_q(p)$, and we can obtain its signature by adding the signatures of the scalar elements $q(x_j)$ and the 2×2 blocks. The signature of the 2×2 blocks is always zero (they have zero trace), and thus the result follows. □

In particular, notice that if we want to count the number of real roots, we can just use $q(x) = 1$. The matrix corresponding to this quadratic form (called the *Hermite form*) is:

$$H_1(p) = V^T V = \begin{bmatrix} s_0 & s_1 & \cdots & s_{n-1} \\ s_1 & s_2 & \cdots & s_n \\ \vdots & \vdots & \ddots & \vdots \\ s_{n-1} & s_n & \cdots & s_{2n-2} \end{bmatrix}, \qquad s_k = \sum_{j=1}^{n} x_j^k.$$

The s_k are known as the *power sums* and can be computed using (4) (although there are much more efficients ways, such as the Newton identities). When $p(x)$ is monic, the s_k are polynomials of degree k in the coefficients of $p(x)$.

Example 8. *Consider the monic cubic polynomial*

$$p(x) = x^3 + p_2 x_2 + p_1 x + p_0.$$

Then, the first five power sums are:

$$\begin{aligned} s_0 &= 3 \\ s_1 &= -p_2 \\ s_2 &= p_2^2 - 2p_1 \\ s_3 &= -p_2^3 + 3p_1 p_2 - 3p_0 \\ s_4 &= p_2^4 - 4p_1 p_2^2 + 2p_1^2 + 4p_0 p_2. \end{aligned}$$

Lemma 9. *The signature of $H_1(p)$ is equal to the number of real roots. The rank of $H_1(p)$ is equal to the number of distinct complex roots of $p(x)$.*

Corollary 10. *If $p(x)$ has odd degree, there is always at least one real root.*

Example 11. *Consider $p(x) = x^3 + 2x^2 + 3x + 4$. The corresponding Hermite matrix is:*

$$H(p) = \begin{bmatrix} 3 & -2 & -2 \\ -2 & -2 & -2 \\ -2 & -2 & 18 \end{bmatrix}$$

This matrix has one negative and two positive eigenvalues, all distinct (i.e., its inertia is $(2, 0, 1)$). Thus, $p(x)$ has three simple roots, and exactly one of them is real.

Sylvester's law of inertia guarantees that this result is actually coordinate independent.

3 Nonnegativity

An important property of a polynomial is whether it only takes nonnegative values. As we will see, this is of interest in a wide variety of applications.

Definition 12. *A univariate polynomial $p(x)$ is* positive semidefinite *or* nonnegative *if $p(x) \geq 0$ for all real values of x.*

Clearly, if $p(x)$ is nonnegative, then its degree must be an even number. The set of nonnegative polynomials has very interesting properties. Perhaps the most appealing one for our purposes is the following:

Theorem 13. *Consider the set P_n of nonnegative univariate polynomials of degree less than or equal to n (n is even). Then, identifying a polynomial with its $n + 1$ coefficients $(p_n, \dots, p_0)$, the set P_n is a proper cone (i.e., closed, convex, pointed, solid) in $\mathbb{R}^{n+1}$.*

An equivalent condition for the (nonconstant) univariate polynomial (1) to be strictly positive, is that $p(x_0) > 0$ for some x_0, and it that has no real roots. Thus, we can use Theorem 7 to write explicit conditions for a polynomial $p(x)$ to be nonnegative in terms of the signature of the associated Hermite matrix $H_1(p)$.

4 Sum of squares

Definition 14. *A univariate polynomial $p(x)$ is a* sum of squares *(SOS) if there exist $q_1, \dots, q_m \in \mathbb{R}[x]$ such that*

$$p(x) = \sum_{k=1}^{m} q_k^2(x).$$

If a polynomial $p(x)$ is a sum of squares, then it obviously satisfies $p(x) \geq 0$ for all $x \in \mathbb{R}$. Thus, a SOS condition is a sufficient condition for global nonnegativity.

Interestingly, in the univariate case, the converse is also true:

Theorem 15. *A univariate polynomial is nonnegative if and only if it is a sum of squares.*

Proof. (⇐) Obvious. If $p(x) = \sum_k q_k^2(x)$ then $p(x) \geq 0$.

(⇒) Since $p(x)$ is univariate, we can factorize it as

$$p(x) = p_n \prod_j (x - r_j)^{n_j} \prod_k (x - a_k + ib_k)^{m_k} (x - a_k - ib_k)^{m_k},$$

where r_j and $a_k \pm ib_k$ are the real and complex roots, respectively, of multiplicities n_j and m_k. Because $p(x)$ is nonnegative, then $p_n > 0$ and the multiplicies of the real roots are even, i.e., $n_j = 2s_j$.

Notice that $(x - a + ib)(x - a - ib) = (x - a)^2 + b^2$. Then, we can write

$$p(x) = p_n \prod_j (x - r_j)^{2s_j} \prod_k \left((x - a_k)^2 + b_k^2\right)^{m_k},$$

Since products of sums of squares are sums of squares, and all the factors in the expression above are SOS, it follows that $p(x)$ is SOS.

Furthermore, the two-squares identity $(\alpha^2 + \beta^2)(\gamma^2 + \delta^2) = (\alpha\gamma - \beta\delta)^2 + (\alpha\delta + \beta\gamma)^2$ allows us to combine every partial product as a sum of only two squares. □

Notice that the proof shows that if $p(x)$ is SOS, then there exists a representation $p(x) = q_1^2(x) + q_2^2(x)$.

As we will see very soon, we can decide whether a univariate polynomial is a sum of squares (equivalently, if it is nonnegative) by solving a semidefinite optimization problem.

5 Positive semidefinite matrices

Recall from Lecture 2 the (apparent) disparity between the stated conditions for a matrix to be positive definite versus the semidefinite case. In the former, we could use a test (Sylvester's criterion) that required the calculation of only n minors, while for the semidefinite case apparently we needed a much larger number, $2^n - 1$.

If the matrix X is positive definite, Sylvester's criterion requires the positivity of the leading principal minors, i.e.,

$$\det X_{1,1} > 0, \quad \det X_{12,12} > 0, \quad \ldots, \quad \det X > 0.$$

For positive semidefiniteness, it is not enough to replace strict positivity with the nonstrict inequality; a simple counterexample is the matrix

$$\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix},$$

for which the leading minors vanish, but is not PSD. As mentioned, an alternative approach is given by the following classical result:

Lemma 16. *Let $A \in \mathcal{S}^n$ be a symmetric matrix. Then $A \succeq 0$ if and only if all $2^n - 1$ principal minors of A are nonnegative.*

Although the condition above requires the nonnegativity of $2^n - 1$ expressions, it is possible to do the same by checking only n inequalities:

Theorem 17 (e.g. [HJ95, p. 403])**.** *A real $n \times n$ symmetric matrix A is positive semidefinite if and only if all the coefficients c_i of its characteristic polynomial $p(\lambda) = \det(\lambda I - A) = \lambda^n + p_{n-1}\lambda^{n-1} + \cdots + p_1\lambda + p_0$ alternate in sign, i.e., they satisfy $p_i(-1)^{n-i} \geq 0$.*

We prove this below, since we will use a slightly more general version of this result when discussing hyperbolic polynomials. Note that in the $n = 2$ case, Theorem 17 is the familiar result that $A \in \mathcal{S}^2$ is positive semidefinite if and only if $\det A \geq 0$ and $\mathrm{Tr} A \geq 0$.

Lemma 18. *Consider a monic univariate polynomial* $p(t) = t^n + \sum_{k=0}^{n-1} p_k t^k$*, that has only real roots. Then, all roots are nonpositive if and only if all coefficients are nonnegative (i.e.,* $p_k \geq 0, k = 0, \ldots, n-1$*).*

Proof. Since all roots of $p(t)$ are real, this can be obtained from a direct application of Descartes' rules of signs; see e.g. [BPR03]. For completeness, we present here a direct proof.

If all roots t_i are nonpositive ($t_i \leq 0$), from the factorization

$$p(t) = \prod_{k=1}^{n} (t - t_i)$$

it follows directly that all coefficients p_k are nonnegative.

For the other direction, from the nonnegativity of the coefficients it follows that $p(0) \geq 0$ and $p(t)$ is nondecreasing. If there exists a $t_i > 0$ such that $p(t_i) = 0$, then the polynomial must vanish in the interval $[0, t_i]$, which is impossible since it is monic and hence nonzero. □

Definition 19. *A set* $S \subset \mathbb{R}^n$ *is* basic closed semialgebraic *if it can be written as*

$$S = \{x \in \mathbb{R}^n \mid f_i(x) \geq 0, \quad h_j(x) = 0\}$$

for some finite set of polynomials $\{f_i, h_j\}$*.*

Theorem 20. *Both the primal and dual feasible sets of a semidefinite program are basic closed semialgebraic.*

Proof. The condition $X \succeq 0$ is equivalent to n nonstrict polynomial inequalities in the entries of X. This can be conveniently shown applying Lemma 18 to the characteristic polynomial of $-X$, i.e.,

$$p(\lambda) = \det(\lambda I + X) = \lambda^n + \sum_{k=0}^{n-1} p_k(X) \lambda^k.$$

where the $p_k(X)$ are homogeneous polynomials of degree $n - k$ in the entries of X. For instance, we have $p_0(X) = \det X$, and $p_{n-1}(X) = \mathrm{Tr} X$.

Since X is symmetric, all its eigenvalues are real, and thus $p(\lambda)$ has only real roots. Positive semidefiniteness of X is equivalent to $p(\lambda)$ having no roots that are strictly positive. It then follows than the two following statements are equivalent:

$$X \succeq 0 \quad \Leftrightarrow \quad p_k(X) \geq 0 \quad k = 0, \ldots, n-1.$$

□

Remark 21. *These inequalities correspond to the elementary symmetric functions* e_k *evaluated at the eigenvalues of the matrix* X*.*

As we will see in subsequent lectures, the same inequalities will reappear when we consider a class of optimization problems known as *hyperbolic programs.*

References

[BPR03] S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in real algebraic geometry*, volume 10 of *Algorithms and Computation in Mathematics.* Springer-Verlag, Berlin, 2003.

[HJ95] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, 1995.