Invariant Theory with Applications

Jan Draisma and Dion Gijswijt

September 17 2009

Contents

Chapter 1

Lecture1: introducing invariant theory

The first lecture gives some flavor of the theory of invariants. Basic notions such as (linear) group representation, the ring of regular functions on a vector space and the ring of invariant functions are defined, and some instructive examples are given.

1.1 Polynomial functions

Let V be a complex vector space. We denote by $V^* := \{f : V \to \mathbb{C} \text{ linear map}\}$ the dual vector space. Viewing the elements of V^* as functions on V, and taking the usual pointwise product of functions, we can consider the algebra of all $\mathbb{C}$-linear combinations of products of elements from V^*.

Definition 1.1.1. The *coordinate ring* $\mathcal{O}(V)$ of the vectorspace V is the algebra of functions $F : V \to \mathbb{C}$ generated by the elements of V^*. The elements of $\mathcal{O}(V)$ are called *polynomial* or *regular* functions on V.

If we fix a basis $e_1, \ldots, e_n$ of V, then a dual basis of V^* is given by the coordinate functions $x_1, \ldots, x_n$ defined by $x_i(c_1e_1 + \cdots + c_ne_n) := c_i$. For the coordinate ring we obtain $\mathcal{O}(V) = \mathbb{C}[x_1, \ldots, x_n]$. This is a polynomial ring in the x_i, since our base field $\mathbb{C}$ is infinite.

Exercise 1.1.2. Show that indeed $\mathbb{C}[x_1, \ldots, x_n]$ is a polynomial ring. In other words, show that the x_i are algebraically independent over $\mathbb{C}$: there is no nonzero polynomial $p \in \mathbb{C}[X_1, \ldots, X_n]$ in n variables $X_1, \ldots, X_n$, such that $p(x_1, \ldots, x_n) = 0$. Hint: this is easy for the case $n = 1$. Now use induction on n.

We call a regular function $f \in \mathcal{O}(V)$ *homogeneous of degree* d if $f(tv) = t^d f(v)$ for all $v \in V$ and $t \in \mathbb{C}$. Clearly, the elements of V^* are regular of degree

1, and the product of polynomials f, g homogeneous of degrees d, d' yields a homogeneous polynomial of degree $d+d'$. It follows that every regular function f can be written as a sum $f = c_0 + c_1 f_1 + \cdots + c_k f_k$ of regular functions f_i homogeneous of degree i. This decomposition is unique (disregarding the terms with zero coefficient). Hence we have a direct sum decomposition $\mathcal{O}(V) = \bigoplus_{d \in \mathbb{N}} \mathcal{O}(V)_d$, where $\mathcal{O}(V)_d := \{f \in \mathcal{O}(V) \mid f \text{ homogeneous of degree } d\}$, making $\mathcal{O}(V)$ into a *graded algebra*.

Exercise 1.1.3. Show that indeed the decomposition of a regular function f into its homogeneous parts is unique.

In terms of the basis $x_1, \ldots, x_n$, we have $\mathcal{O}(V)_d = \mathbb{C}[x_1, \ldots, x_n]_d$, where $\mathbb{C}[x_1, \ldots, x_n]_d$ consists of all polynomials of total degree d and has as basis the monomials $x_1^{d_1} x_2^{d_2} \cdots x_n^{d_n}$ for $d_1 + d_2 + \cdots + d_n = d$.

1.2 Representations

Central objects in this course are linear representations of groups. For any vector space V we write $\mathrm{GL}(V)$ for the group of all invertible linear maps from V to itself. When we have a fixed basis of V, we may identify V with $\mathbb{C}^n$ and $\mathrm{GL}(V)$ with the set of invertible matrices $n \times n$ matrices $\mathrm{GL}(\mathbb{C}^n) \subset \mathrm{Mat}_n(\mathbb{C})$.

Definition 1.2.1. Let G be a group and let X be a set. An *action of G on X* is a map $\alpha : G \times X \to X$ such that $\alpha(1, x) = x$ and $\alpha(g, \alpha(h, x)) = \alpha(gh, x)$ for all $g, h \in G$ and $x \in X$.

If α is clear from the context, we will usually write gx instead of $\alpha(g, x)$. What we have just defined is sometimes called a *left action* of G on X; *right actions* are defined similarly.

Definition 1.2.2. If G acts on two sets X and Y, then a map $\phi : X \to Y$ is called *G-equivariant* if $\phi(gx) = g\phi(x)$ for all $x \in X$ and $g \in G$. As a particular case of this, if X is a subset of Y satisfying $gx \in X$ for all $x \in X$ and $g \in G$, then X is called *G-stable*, and the inclusion map is G-equivariant.

Example 1.2.3. The symmetric group S_4 acts on the set $\binom{[4]}{2}$ of unordered pairs of distinct numbers in $[4] := \{1, 2, 3, 4\}$ by $g\{i, j\} = \{g(i), g(j)\}$. Think of the edges in a tetrahedron to visualise this action. The group S_4 also acts on the set $X := \{(i, j) \mid i, j \in [4] \text{ distinct}\}$ of all ordered pairs by $g(i, j) = (g(i), g(j))$—think of directed edges—and the map $X \to \binom{[4]}{2}$ sending (i, j) to $\{i, j\}$ is S_4-equivariant.

Definition 1.2.4. Let G be a group and let V be a vector space. A *(linear) representation of G on V* is a group homomorphism $\rho : G \to \mathrm{GL}(V)$.

If ρ is a representation of G, then the map $(g, v) \mapsto \rho(g)v$ is an action of G on V. Conversely, if we have an action α of G on V such that $\alpha(g, .) : V \to V$ is a linear map for all $g \in G$, then the map $g \mapsto \alpha(g, .)$ is a linear representation.

As with actions, instead of $\rho(g)v$ we will often write gv. A vector space with an action of G by linear maps is also called a *G-module*.

Given a linear representation $\rho : G \to \mathrm{GL}(V)$, we obtain a linear representation $\rho^* : G \to \mathrm{GL}(V^*)$ on the dual space V^*, called the *dual representation* or *contragredient representation* and defined by

$$(\rho^*(g)x)(v) := x(\rho(g)^{-1}v) \text{ for all } g \in G,\ x \in V^* \text{ and } v \in V. \tag{1.1}$$

Exercise 1.2.5. Let $\rho : G \to \mathrm{GL}_n(\mathbb{C})$ be a representation of G on $\mathbb{C}^n$. Show that with respect to the dual basis, ρ^* is given by $\rho^*(g) = (\rho(g)^{-1})^\mathsf{T}$, where A^T denotes the transpose of the matrix A.

1.3 Invariant functions

Definition 1.3.1. Given a representation of a group G on a vector space V, a regular function $f \in \mathcal{O}(V)$ is called *G-invariant* or simply *invariant* if $f(v) = f(gv)$ for all $g \in G, v \in V$. We denote by $\mathcal{O}(V)^G \subseteq \mathcal{O}(V)$ the subalgebra of invariant functions. The actual representation of G is assumed to be clear from the context.

Observe that $f \in \mathcal{O}(V)$ is invariant, precisely when it is constant on the orbits of V under the action of G. In particular, the constant functions are invariant.

The representation of G on V induces an action on the (regular) functions on V by defining $(gf)(v) := f(g^{-1}v)$ for all $g \in G, v \in V$. This way the invariant ring can be discribed as the set of regular functions fixed by the action of G: $\mathcal{O}(V)^G = \{f \in \mathcal{O}(V) \mid gf = f \text{ for all } g \in G\}$. Observe that when restricted to $V^* \subset \mathcal{O}(V)$, this action coincides with the action corresponding to the dual representation. In terms of a basis $x_1, \ldots, x_n$ of V^*, the regular functions are polynomials in the x_i and the action of G is given by $gp(x_1, \ldots, x_n) = p(gx_1, \ldots, gx_n)$ for any polynomial p. Since for every d, G maps the set of polynomials homogeneous of degree d to itself, it follows that the homogeneous parts of an invariant are invariant as well. This shows that $\mathcal{O}(V)^G = \bigoplus_d \mathcal{O}(V)^G_d$, where $\mathcal{O}(V)^G_d := \mathcal{O}(V)_d \cap \mathcal{O}(V)^G$.

Example 1.3.2. Consider the representation $\rho : \mathbb{Z}/3\mathbb{Z} \to \mathrm{GL}_2(\mathbb{C})$ defined by mapping 1 to the matrix $\left(\begin{smallmatrix} 0 & -1 \\ 1 & -1 \end{smallmatrix}\right)$ (and mapping 2 to $\left(\begin{smallmatrix} -1 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ and 0 to the identity matrix). With respect to the dual basis x_1, x_2, the dual representation is given by:

$$\rho^*(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \rho^*(1) = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}, \qquad \rho^*(2) = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}. \tag{1.2}$$

The polynomial $f = x_1^2 - x_1x_2 + x_2^2$ is an invariant:

$$\rho^*(1)f = (-x_1 + x_2)^2 - (-x_1 + x_2)(-x_1) + (-x_1)^2 = x_1^2 - x_1x_2 + x_2^2 = f, \tag{1.3}$$

and since 1 is a generator of the group, f is invariant under all elements of the group. Other invariants are $x_1^2x_2 - x_1x_2^2$ and $x_1^3 - 3x_1x_2^2 + x_2^3$. These three invariants generate the ring of invariants, althought it requires some work to show that.

A simpler example in which the complete ring of invariants can be computed is the following.

Example 1.3.3. Let D_4 be the symmetry group of the square, generated by a rotation r, a reflection s and the relations $r^4 = e, s^2 = e$ and $srs = r^3$, where e is the identity. The representation ρ of D_4 on $\mathbb{C}^2$ is given by

$$\rho(r) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad \rho(s) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{1.4}$$

the dual representation is given by the same matrices:

$$\rho^*(r) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad \rho^*(s) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{1.5}$$

It is easy to check that $x_1^2+x_2^2$ and $x_1^2x_2^2$ are invariants, and so are all polynomial expressions in these two invariants. We will show that in fact $\mathcal{O}(\mathbb{C}^2)^{D_4} = \mathbb{C}[x_1^2 + x_2^2, x_1^2x_2^2] =: R$. It suffices to show that all homogeneous invariants belong to R.

Let $p \in \mathbb{C}[x_1, x_2]$ be a homogeneous invariant. Since $sp = p$, only monomials having even exponents for x_1 can occur in p. Since r^2s exchanges x_1 and x_2, for every monomial $x_1^ax_2^b$ in p, the monomial $x_1^bx_2^a$ must occur with the same exponent. This proves the claim since every polynomial of the form $x_1^{2n}x_2^{2m} + x_1^{2m}x_2^{2n}$ is an element of R. Indeed, we may assume that $n \leq m$ and proceed by induction on $n + m$, the case $n + m = 0$ being trivial. If $n > 0$ we have $q = (x_1^2x_2^2)^n(x_2^{2m-2n} + x_1^{2m-2n})$ and we are done. If $n = 0$ we have $2q = 2(x_1^{2m} + x_2^{2m}) = 2(x_1^2+x_2^2)^m - \sum_{i=1}^{m-1} \binom{m}{i}(x_1^{2i}x_2^{2m-2i})$ and we are done by induction again.

1.4 Conjugacy classes of matrices

In this section we discuss the polynomial functions on the square matrices, invariant under conjugation of the matrix variable by elements of $\mathrm{GL}_n(\mathbb{C})$. This example shows some tricks that are useful when proving that certain invariants are equal. Denote by $M_n(\mathbb{C})$ the vectorspace of complex $n \times n$ matrices. We consider the action of $G = \mathrm{GL}_n(\mathbb{C})$ on $M_n(\mathbb{C})$ by conjugation: $(g, A) \mapsto gAg^{-1}$ for $g \in \mathrm{GL}_n(\mathbb{C})$ and $A \in M_n(\mathbb{C})$. We are interested in finding all polynomials in the entries of $n \times n$ matrices that are invariant under G. Two invariants are given by the functions $A \mapsto \det A$ and $A \mapsto \mathrm{tr}A$.

Let

$$\chi_A(t) := \det(tI - A) = t^n - s_1(A)t^{n-1} + s_2(A)t^{n-2} - \cdots + (-1)^n s_n(A) \tag{1.6}$$

be the characteristic polynomial of A. Here the s_i are polynomials in the entries of A. Clearly,

$$\chi_{gAg^{-1}}(t) = \det(g(tI - A)g^{-1}) = \det(tI - A) = \chi_A(t) \tag{1.7}$$

holds for all $t \in \mathbb{C}$. It follows that the functions $s_1, \dots, s_n$ are G-invariant. Observe that $s_1(A) = \mathrm{tr}A$ and $s_n(A) = \det A$.

Proposition 1.4.1. *The functions $s_1, \dots, s_n$ generate $\mathcal{O}(\mathrm{Mat}_n(\mathbb{C}))^{\mathrm{GL}_n(\mathbb{C})}$ and are algebraically independent.*

Proof. To each $c = (c_1, \dots, c_n \in \mathbb{C}^n$ we associate the so-called *companion matrix*

$$A_c := \begin{pmatrix} 0 & \cdots & \cdots & 0 & -c_n \\ 1 & \ddots & & \vdots & -c_{n-1} \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & c_2 \\ 0 & \cdots & 0 & 1 & c_1 \end{pmatrix} \in M_n(\mathbb{C}). \tag{1.8}$$

A simple calculation shows that $\chi_{A_c}(t) = t^n + c_{n-1}t^{n-1} + \cdots + c_1 t + c_0$.

Exercise 1.4.2. Verify that $\chi_{A_c}(t) = t^n + c_{n-1}t^{n-1} + \cdots + c_1 t + c_0$.

This implies that $s_i(A_c) = (-1)^i c_i$ and therefore

$$\{(s_1(A_c), s_2(A_c), \dots, s_n(A_c) \mid A \in M_n(\mathbb{C})\} = \mathbb{C}^n. \tag{1.9}$$

It follows that the s_i are algebraically independent over $\mathbb{C}$. Indeed, suppose that $p(s_1, \dots, s_n) = 0$ for some polynomial p in n variables. Then

$$0 = p(s_1, \dots, s_n)(A) = p(s_1(A), \dots, s_n(A)) \tag{1.10}$$

for all A and hence $p(c_1, \dots, c_n) = 0$ for all $c \in \mathbb{C}^n$. But this implies that p itself is the zero polynomial.

Now let $f \in \mathcal{O}(\mathrm{Mat}_n(\mathbb{C}))^G$ be an invariant function. Define the polynomial p in n variables by $p(c_1, \dots, c_n) := f(A_c)$, and $P \in \mathcal{O}(\mathrm{Mat}_n(\mathbb{C}))^G$ by $P(A) := p(-s_1(A), s_2(A), \dots, (-1)^n s_n(A))$. By definition, P and f agree on all companion matrices, and since they are both G-invariant they agree on $W := \{gA_cg^{-1} \mid g \in G, c \in \mathbb{C}^n\}$. To finish the proof, it suffices to show that W is dense in $\mathrm{Mat}_n(\mathbb{C})$ since $f - P$ is continuous and zero on W. To show that W is dense in $\mathcal{O}(\mathrm{Mat}_n(\mathbb{C}))$, it suffices to show that the set of matrices with n distinct nonzero eigenvalues is a subset of W and is itself dense in $\mathcal{O}(\mathrm{Mat}_n(\mathbb{C}))$. This we leave as an exercise.

Exercise 1.4.3. Let $A \in \mathrm{Mat}_n(\mathbb{C})$ have n distinct nonzero eigenvalues. Show that A is conjugate to A_c for some $c \in \mathbb{C}^n$. Hint: find $v \in \mathbb{C}^n$ such that

$v, Av, A^2v, \ldots, A^{n-1}v$ is a basis for $\mathbb{C}^n$. You might want to use the fact that the Vandermonde determinant

$$\det \begin{pmatrix} 1 & \ldots & 1 \\ c_1 & \ldots & c_n \\ c_1^2 & \ldots & c_n^2 \\ \vdots & \ddots & \vdots \\ c_1^{n-1} & \ldots & c_n^{n-1} \end{pmatrix} \tag{1.11}$$

is nonzero if $c_1, \ldots, c_n$ are distinct and nonzero.

Exercise 1.4.4. Show that the set of matrices with n distinct nonzero eigenvalues is dense in the set of all complex $n \times n$ matrices. Hint: every matrix is conjugate to an upper triangular matrix.

□

1.5 Exercises

Exercise 1.5.1. Let G be a finite group acting on $V = \mathbb{C}^n$, $n \geq 1$. Show that $\mathcal{O}(V)^G$ contains a nontrivial invariant. That is, $\mathcal{O}(V)^G \neq \mathbb{C}$. Give an example of an action of an infinite group G on V with the property that only the constant functions are invariant.

Exercise 1.5.2. Let $\rho : \mathbb{Z}/2\mathbb{Z} \to \mathrm{GL}_2(\mathbb{C})$ be the representation given by $\rho(1) := \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. Compute the invariant ring. That is, give a minimal set of generators for $\mathcal{O}(\mathbb{C}^2)^{\mathbb{Z}/2\mathbb{Z}}$.

Exercise 1.5.3. Let $U := \{\left(\begin{smallmatrix} 1 & a \\ 0 & 1 \end{smallmatrix}\right) \mid a \in \mathbb{C}\}$ act on $\mathbb{C}^2$ in the obvious way. Denote the coordinate functions by x_1, x_2. Show that $\mathcal{O}(\mathbb{C}^2)^U = \mathbb{C}[x_2]$.

Exercise 1.5.4. Let $\rho : \mathbb{C}^* \to \mathrm{GL}_3(\mathbb{C})$ be the representation given by $\rho(t) = \left(\begin{smallmatrix} t^{-2} & 0 & 0 \\ 0 & t^{-3} & 0 \\ 0 & 0 & t^4 \end{smallmatrix}\right)$. Find a minimal system of generators for the invariant ring.

Exercise 1.5.5. Let $\pi : \mathrm{Mat}_n(\mathbb{C}) \to \mathbb{C}^n$ be given by $\pi(A) := (s_1(A), \ldots, s_n(A))$. Show that for every $c \in \mathbb{C}^n$ the fiber $\{A \mid \pi(A) = c\}$ contains a unique conjugacy class $\{gAg^{-1} \mid g \in \mathrm{GL}_n(\mathbb{C})\}$ of a diagonalizable (semisimple) matrix A.

Chapter 2

Lecture2: Symmetric polynomials

In this chapter, we consider the natural action of the symmetric group S_n on the ring of polynomials in the variables $x_1, \ldots, x_n$. The fundamental theorem of symmetric polynomials states that the elementary symmetric polynomials generate the ring of invariants. As an application we prove a theorem of Sylvester that characterizes when a univariate polynomial with real coefficients has only real roots.

2.1 Symmetric polynomials

Let the group S_n act on the polynomial ring $\mathbb{C}[x_1, \ldots, x_n]$ by permuting the variables:

$$\sigma p(x_1, \ldots, x_n) := p(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \text{ for all } \sigma \in S_n. \tag{2.1}$$

The polynomials invariant under this action of S_n are called *symmetric polynomials.* As an example, for $n = 3$ the polynomial $x_1^2x_2 + x_1^2x_3 + x_1x_2^2 + x_1x_3^2 + x_2^2x_3 + x_2x_3^2 + 7x_1 + 7x_2 + 7x_3$ is symmetric, but $x_1^2x_2 + x_1x_3^2 + x_2^2x_3$ is not symmetric (although it is invariant under the alternating group).

In terms of linear representations of a group, we have a linear representation $\rho : S_n \to \mathrm{GL}_n(\mathbb{C})$ given by $\rho(\sigma)e_i := e_{\sigma(i)}$, where $e_1, \ldots, e_n$ is the standard basis of $\mathbb{C}^n$. On the dual basis $x_1, \ldots, x_n$ the dual representation is given by $\rho^*(\sigma)x_i = x_{\sigma(i)}$, as can be easily checked. The invariant polynomial functions on $\mathbb{C}^n$ are precisely the symmetric polynomials.

There are some obvious examples of symmetric polynomials:

$$s_1 := x_1 + x_2 + \ldots + x_n \text{ and } s_2 := x_1x_2 + x_1x_3 + \ldots + x_{n-1}x_n \tag{2.2}$$

are invariant. More generally, for every $k = 1, \ldots, n$, the *k-th elementary symmetric polynomial*

$$s_k := \sum_{i_1 < \ldots < i_k} x_{i_1} \cdots x_{i_k} \tag{2.3}$$

is invariant. Recall that these polynomials express the coefficients of a univariate polynomial in terms of its roots:

$$\prod_{i=1}^{n}(t-x_i) = x^n + \sum_{k=1}^{n}(-1)^k s_k t^{n-k}. \tag{2.4}$$

Moreover, if g is any polynomial in n variables $y_1, \ldots, y_n$, then $g(s_1, \ldots, s_n)$ is again a polynomial in the x_i which is invariant under all coordinate permutations. A natural question is: which symmetric polynomials are expressible as a polynomial in the elementary symmetric polynomials. For example $x_1^2+\cdots+x_n^2$ is clearly symmetric and it can be expressed in terms of the s_i:

$$x_1^2 + \cdots + x_n^2 = s_1^2 - 2s_2. \tag{2.5}$$

It is a beautiful fact that the elementary symmetric polynomials generate *all* symmetric polynomials.

Theorem 2.1.1 (Fundamental theorem of symmetric polynomials)**.** *Every S_n-invariant polynomial $f(x_1, \ldots, x_n)$ in the x_i can be written as $g(s_1, \ldots, s_n)$, where $g = g(y_1, \ldots, y_n)$ is a polynomial in n variables. Moreover, given f the polynomial g is unique.*

The proof of this result uses the *lexicographic order* on monomials in the variables $\underline{x} = (x_1, \ldots, x_n)$. We say that $\underline{x}^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ is (lexicographically) larger than $\underline{x}^\beta$ if there is a k such that $\alpha_k > \beta_k$ and $\alpha_i = \beta_i$ for all $i < k$. So for instance $x_1^2 > x_1x_2^4 > x_1x_2^3 > x_1x_2x_3^5$, etc. The *leading monomial* $\mathrm{lm}(f)$ of a non-zero polynomial f in the x_i is the largest monomial (with respect to this ordering) that has non-zero coefficient in f.

Exercise 2.1.2. Check that $\mathrm{lm}(fg) = \mathrm{lm}(f)\mathrm{lm}(g)$ and that $\mathrm{lm}(s_k) = x_1 \cdots x_k$.

Exercise 2.1.3. Show that there are no infinite lexicographically strictly decreasing chains of monomials.

Since every decreasing chain of monomial is finite, we can use this order to do induction on monomials, as we do in the following proof.

Proof of Theorem 2.1.1. Let f be any S_n-invariant polynomial in the x_i. Let $\underline{x}^\alpha$ be the leading monomial of f. Then $\alpha_1 \geq \ldots \geq \alpha_n$ because otherwise a suitable permutation applied to $\underline{x}^\alpha$ would yield a lexicographically larger monomial, which has the same non-zero coefficient in f as $\underline{x}^\alpha$ by invariance of f. Now consider the expression

$$s_n^{\alpha_n} s_{n-1}^{\alpha_{n-1}-\alpha_n} \cdots s_1^{\alpha_1-\alpha_2} \tag{2.6}$$

The leading monomial of this polynomial equals

$$(x_1 \cdots x_n)^{\alpha_n}(x_1 \cdots x_{n-1})^{\alpha_{n-1}-\alpha_n} \cdots x_1^{\alpha_1-\alpha_2}, \tag{2.7}$$

which is just $\underline{x}^\alpha$. Subtracting a scalar multiple of the expression from f therefore cancels the term with monomial $\underline{x}^\alpha$, and leaves an S_n-invariant polynomial with a strictly smaller leading monomial. After repeating this step finitely many times, we have expressed f as a polynomial in the s_k.

This shows existence of g in the theorem. For uniqueness, let $g \in \mathbb{C}[y_1, \ldots, y_n]$ be a nonzero polynomial in n variables. It suffices to show that $g(s_1, \ldots, s_n) \in \mathbb{C}[x_1, \ldots, x_n]$ is not the zero polynomial. Observe that

$$\operatorname{lm}(s_1^{\alpha_1} \cdots s_n^{\alpha_n}) = x_1^{\alpha_1+\cdots+\alpha_n} x_2^{\alpha_2+\cdots+\alpha_n} \cdots x_n^{\alpha_n}. \tag{2.8}$$

It follows that the leading monomials of the terms $s_1^{\alpha_1} \cdots s_n^{\alpha_n}$, corresponding to the monomials occuring with nonzero coefficient in $g = \sum_\alpha \underline{y}^\alpha$, are pairwise distinct. In particular, the largest such leading monomial will not be cancelled in the sum and is the leading monomial of $g(s_1, \ldots, s_n)$. □

Remark 2.1.4. The proof shows that in fact the coefficients of the polynomial g lie in the ring generated by the coefficients of f. In particular, when f has real coefficients, also g has real coefficients.

Exercise 2.1.5. Let $\pi : \mathbb{C}^n \to \mathbb{C}^n$ be given by

$$\pi(x_1, \ldots, x_n) = (s_1(x_1, \ldots, x_n), \ldots, s_n(x_1, \ldots, x_n)). \tag{2.9}$$

Use the fact that every univariate polynomial over the complex numbers can be factorised into linear factors to show that π is surjective. Use this to show that $s_1, \ldots, s_n$ are algebraically independent over $\mathbb{C}$. Describe for $b \in \mathbb{C}^n$ the fiber $\pi^{-1}(b)$.

Remark 2.1.6. The above proof of the fundamental theorem of symmetric polynomials gives an algorithm to write a given symmetric polynomial as a polynomial in the elementary symmetric polynomials. In each step the initial monomial of the residual symmetric polynomial is decreased, ending with the zero polynomial after a finite number of steps. Instead of using the described lexicographic order on the monomials, other linear orders can be used. An example would be the *degree lexicographic order*, where we set $\underline{x}^\alpha > \underline{x}^\beta$ if either $\alpha_1 + \cdots + \alpha_n > \beta_1 + \cdots + \beta_n$ or equality holds and there is a k such that $\alpha_k > \beta_k$ and $\alpha_i = \beta_i$ for all $i < k$.

Example 2.1.7. We write $x_1^3 + x_2^3 + x_3^3$ as a polynomial in the s_i. Since the leading monomial is $x_1^3 x_2^0 x_3^0$ we subtract $s_3^0 s_2^0 s_1^3$ and are left with $-3(x_1^2 x_2 + x_1^2 x_3 + x_1 x_2^2 + x_1 x_3^2 + x_2^2 x_3 + x_2 x_3^2) - 6(x_1 x_2 x_3)$. The leading monomial is now $x_1^2 x_2$, so we add $3 s_3^0 s_2^1 s_1^{2-1}$. This leaves $3 x_1 x_2 x_3 = 3 s_3^1 s_2^{1-1} s_1^{1-1}$, which is reduced to zero in the next step.

This way we obtain $x_1^3 + x_2^3 + x_3^3 = s_1^3 - 3 s_1 s_2 + 3 s_3$.

Exercise 2.1.8. Give an upper bound on the number of steps of the algorithm in terms of the number of variables n and the (total) degree of the input polynomial f.

2.2 Counting real roots

Given a (monic) polynomial $f(t) = t^n - a_1 t^{n-1} + \cdots + (-1)^n a_n$, the coefficients are elementary symmetric functions in the roots of f. Therefore, any property that can be expressed as a symmetric polynomial in the roots of f, can also be expressed as a polynomial in the coefficients of f. This way we can determine properties of the roots by just looking at the coefficients of f. For example: when are all roots of f distinct?

Definition 2.2.1. For a (monic) polynomial $f = (t - x_1)\cdots(t - x_n)$, define the *discriminant* $\Delta(f)$ *of* f by $\Delta(f) := \prod_{1\le i<j\le n}(x_i - x_j)^2$.

Clearly, $\Delta(f) = 0$ if and only if all roots of f are distinct. It is not hard to see that $\Delta(f)$ is a symmetric polynomial in the roots of f. We will see later how f can be expressed in terms of the coefficients of f.

Exercise 2.2.2. Let $f(t) = t^2 - at + b$. Write $\Delta(f)$ as a polynomial in a and b.

Definition 2.2.3. Given n complex numbers $x_1, \ldots, x_n$, the *Vandermonde matrix* A for these numbers is given by

$$A := \begin{pmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{pmatrix}. \tag{2.10}$$

Lemma 2.2.4. *Given numbers $x_1, \ldots, x_n$, the Vandermonde matrix A has nonzero determinant if and only if the $x_1, \ldots, x_n$ are distinct.*

Proof. View the determinant of the Vandermonde matrix (called the *Vandermonde determinant*) as a polynomial p in the variables $x_1, \ldots, x_n$. For any $i < j$, $p(x_1, \ldots, x_n) = 0$ when $x_i = x_j$ and hence p is divisible by $(x_j - x_i)$. Expanding the determinant, we see that p is homogeneous of degree $\binom{n}{2}$, with lowest monomial $x_1^0 x_2^1 \cdots x_n^{n-1}$ having coefficient 1. It follows that $p = \prod_{1\le i<j\le n}(x_j - x_i)$, since the right-hand side divides p, and the two polynomials have the same degree and the same nonzero coefficient for $x_1^0 x_2^1 \cdots x_n^{n-1}$. □

Exercise 2.2.5. Show that the Vandermonde matrix A of numbers $x_1, \ldots, x_n$ satisfies $\det A = \prod_{1\le i<j\le n}(x_j - x_i)$ by doing row- and column-operations on A and applying induction on n.

Definition 2.2.6. Let $f = (t - \alpha_1)(t - \alpha_2)\cdots(t - \alpha_n) \in \mathbb{C}[t]$ be a monic polynomial of degree n in the variable t. We define the *Bezoutiant matrix* $\mathrm{Bez}(f)$ *of* f by

$$\mathrm{Bez}(f) = (p_{i+j-2}(\alpha_1, \ldots, \alpha_n))_{i,j=1}^{n}, \tag{2.11}$$

where $p_k(x_1, \ldots, x_n) := x_1^k + \cdots + x_n^k$ for $k = 0, 1, \ldots$ is the k-th *Newton polynomial.*

Since the entries of $\mathrm{Bez}(f)$ are symmetric polynomials in the roots of f, it follows by the fundamental theorem of symmetric polynomials that the entries are polynomials (with integer coefficients) in the elementary symmetric functions and hence in the coefficients of f. In particular, when f has real coefficients, $\mathrm{Bez}(f)$ is a real matrix. Another useful fact is that $\mathrm{Bez}(f) = A^{\mathsf{T}}A$, where A is the Vandermonde matrix for the roots $\alpha_1, \ldots, \alpha_n$ of f.

Exercise 2.2.7. Show that the discriminant of f satisfies: $\Delta(f) = \det \mathrm{Bez}(f)$.

Example 2.2.8. Let $f = t^2 - at + b$ have roots α and β. So $a = \alpha + \beta$ and $b = \alpha\beta$. We compute $\mathrm{Bez}(f)$. We have $p_0 = 2$, $p_1 = a$, $p_2 = a^2 - 2b$ so $\mathrm{Bez}(f) = \left(\begin{smallmatrix} 2 & a \\ a & a^2-2b \end{smallmatrix}\right)$. The determinant equals $a^2 - 4b$ and the trace equals $a^2 - 2b + 2$. There are three cases for the eigenvalues $\lambda_1 \geq \lambda_2$ of $\mathrm{Bez}(f)$:

- If $a^2 - 4b > 0$, we have $\lambda_1, \lambda_2 > 0$ and α, β are distinct real roots.
- If $a^2 - 4b = 0$, we have $\lambda_1 > 0, \lambda_2 = 0$ and $\alpha = \beta$.
- If $a^2 - 4b < 0$, we have $\lambda_1 > 0, \lambda_2 < 0$ and α and β are complex conjugate (nonreal) roots.

The determinant of $\mathrm{Bez}(f)$ determines whether f has double roots. The matrix $\mathrm{Bez}(f)$ can give us much more information about the roots of f. In particular, it describes when a polynomial with real coefficients has only real roots!

Theorem 2.2.9 (Sylverster)**.** *Let $f \in \mathbb{C}[t]$ be a polynomial in the variable t with r distinct roots in $\mathbb{R}$ and $2k$ distinct roots in $\mathbb{C} \setminus \mathbb{R}$. Then the Bezoutiant matrix* $\mathrm{Bez}(f)$ *has rank $r + 2k$, with $r + k$ positive eigenvalues and k negative eigenvalues.*

proof of Theorem 2.2.9. Number the roots $\alpha_1, \ldots, \alpha_n$ of f in such a way that $\alpha_1, \ldots, \alpha_{2k+r}$ are distinct. We write m_i for the multiplicity of the root α_i, $i = 1, \ldots, 2k+r$. Let A be the Vandermonde matrix for the numbers $\alpha_1, \ldots, \alpha_n$, so that $\mathrm{Bez}(f) = A^{\mathsf{T}}A$. We start by computing the rank of $\mathrm{Bez}(f)$.

Denote by $\tilde{A}$ the $(2k + r) \times n$ submatrix of A consisting of the first $2k + r$ rows of A. An easy computation shows that

$$\mathrm{Bez}(f) = A^{\mathsf{T}}A = \tilde{A}^{\mathsf{T}} \operatorname{diag}(m_1, \ldots, m_{2k+r})\tilde{A}, \tag{2.12}$$

where $\operatorname{diag}(m_1, \ldots, m_{2k+r})$ is the diagonal matrix with the multiplicities of the roots on the diagonal. Since, $\tilde{A}$ contains a submatrix equal to the Vandermonde matrix for the distinct roots $\alpha_1, \ldots, \alpha_{2k+r}$, it follows by Lemma 2.2.4 that the rows of $\tilde{A}$ are linearly independent. Since the diagonal matrix has full rank, it follows that $\mathrm{Bez}(f)$ has rank $2k + r$.

To finish the proof, we write $A = B + iC$, where B and C are real matrices and i denotes a square root of -1. Since f has real coefficients, $\mathrm{Bez}(f)$ is a real matrix and hence

$$\mathrm{Bez}(f) = B^{\mathsf{T}}B - C^{\mathsf{T}}C + i(C^{\mathsf{T}}B + B^{\mathsf{T}}C) = B^{\mathsf{T}}B - C^{\mathsf{T}}C. \tag{2.13}$$

We have

$$\operatorname{rank}(B) \leq r + k, \qquad \operatorname{rank}(C) \leq k. \tag{2.14}$$

Indeed, for any pair $\alpha, \overline{\alpha}$ of complex conjugate numbers, the real parts of α^j and $\overline{\alpha}^j$ are equal and the imaginary parts are opposite. Hence B has at most $r + k$ different rows and C has (up to a factor -1) at most k different nonzero rows.

Denote the kernel of $\operatorname{Bez}(f), B$ and C by N, N_B and N_C respectively. Clearly $N_B \cap N_C \subseteq N$. This gives

$$\begin{aligned} \dim N \geq \dim(N_B \cap N_C) &\geq \dim N_B + \dim N_C - n && (2.15)\\ &\geq (n - r - k) + (n - k) - n && (2.16)\\ &= n - 2 - 2k = \dim N. && (2.17) \end{aligned}$$

Hence we have equality throughout, showing that $\dim N_B = n-r-k, \dim N_C = n - k$ and $N_B \cap N_C = N$.

Write $N_B = N \oplus N'_B$ and $N_C = N \oplus N'_C$ as a direct sum of vector spaces. For all nonzero $u \in N'_C$, we have $u^{\mathsf{T}}C^{\mathsf{T}}Cu = 0$ and $u^{\mathsf{T}}B^{\mathsf{T}}Bu > 0$ and so $u^{\mathsf{T}}\operatorname{Bez}(f)u > 0$. This shows that $\operatorname{Bez}(f)$ has at least $\dim N'_C = r + k$ positive eigenvalues (see exercises). Similarly, $u^{\mathsf{T}}\operatorname{Bez}(f)u < 0$ for all nonzero $u \in N'_B$ so that $\operatorname{Bez}(f)$ has at least $\dim N'_B = k$ negative eigenvalues. Since $\operatorname{Bez}(f)$ has $n - r - 2k$ zero eigenvalues, it has exactly $r + k$ positive eigenvalues and exactly k negative eigenvalues. □

Exercise 2.2.10. Let B be a real $n \times n$ matrix and $x \in \mathbb{R}^n$. Show that $x^{\mathsf{T}}B^{\mathsf{T}}Bx \geq 0$ and equality holds if and only if $Bx = 0$.

Exercise 2.2.11. Let A be a real symmetric $n \times n$ matrix. Show that the following are equivalent:

- there exists a linear subspace $V \subseteq \mathbb{R}^n$ of dimension k such that $x^{\mathsf{T}}Ax > 0$ for all nonzero $x \in V$,
- A has at least k positive eigenvalues.

Exercise 2.2.12. Use the previous exercise to show Sylvesters law of inertia: Given a real symmetric $n \times n$ matrix A and an invertible real matrix S, the two matrices A and $S^{\mathsf{T}}AS$ have the same number of positive, negative and zero eigenvalues. This implies that the signature of A can be easily determined by bringing it into diagonal form using simultaneous row and column operations.

2.3 Exercises

Exercise 2.3.1. Let $f(t) := t^3 + at + b$, where a, b are real numbers.

- Compute $\operatorname{Bez}(f)$.
- Show that $\Delta(f) = -4a^3 - 27b^2$.

- Determine, in terms of a and b, when f has only real roots.

Exercise 2.3.2. Prove the following formulas due to Newton:

$$p_k - s_1 p_{k-1} + \cdots + (-1)^{k-1} s_{k-1} p_1 + (-1)^k k s_k = 0 \tag{2.18}$$

for all $k = 1, \ldots, n$.

Show that for $k > n$ the following similar relation holds:

$$p_k - s_1 p_{k-1} + \cdots + (-1)^n s_n p_{k-n} = 0. \tag{2.19}$$

Hint: Let $f(t) = (1 - tx_1) \cdots (1 - tx_n)$ and compute $f'(t)/f(t)$ in two ways.