2

Symmetric Functions

Summary. Our aim is to alternate elements of the general theory with significant examples. We deal now with symmetric functions.

In this chapter we will develop some of the very basic theorems on symmetric functions, in part as a way to give a look into 19^{th} century invariant theory, but as well to establish some useful formulas which will show their full meaning only after developing the representation theory of the linear and symmetric groups.

1 Symmetric Functions

1.1 Elementary Symmetric Functions

The theory of symmetric functions is a classical theory developed (by Lagrange, Ruffini, Galois, and others) in connection with the theory of algebraic equations in one variable and the classical question of resolution by radicals.

The main link is given by the formulas expressing the coefficients of a polynomial through its roots. A formal approach is the following.

Consider polynomials in variables $x_1, x_2, \dots, x_n$ and an extra variable t over the ring of integers. The *elementary symmetric functions* $e_i := e_i(x_1, x_2, \dots, x_n)$ are implicitly defined by the formula

$$(1.1.1) \qquad p(t) := \prod_{i=1}^{n}(1 + tx_i) := 1 + \sum_{i=1}^{n} e_i t^i .$$

More explicitly, $e_i(x_1, x_2, \dots, x_n)$ is the sum of $\binom{n}{i}$ terms: the products, over all subsets of $\{1, 2, \dots, n\}$ with i elements of the variables with indices in that subset. That is,

$$(1.1.2) \qquad e_i = \sum_{1 \le a_1 < a_2 < \cdots < a_i \le n} x_{a_1} x_{a_2} \cdots x_{a_i} .$$

If σ is a permutation of the indices, we obviously have

$$\prod_{i=1}^{n}(1+tx_i)=\prod_{i=1}^{n}(1+tx_{\sigma i}).$$

Thus the elements e_i are invariant under permutation of the variables.

Of course the polynomial $t^n p\big(-\frac{1}{t}\big)$ has the elements x_i as its roots.

Definition. A polynomial in the variables $(x_1, x_2, \dots, x_n)$, invariant under permutation of these variables, is called a symmetric function.

The functions e_i are called *elementary symmetric functions.*

There are several obviously symmetric functions, e.g., the *power sums* $\psi_k := \sum_{i=1}^{n} x_i^k$ and the functions S_k defined as the sum of all monomials of degree k. These are particular cases of the following general construction.

Consider the basis of the ring of polynomials given by the monomials. This basis is permuted by the symmetric group. By Proposition 2.5 of Chapter 1 we have:

A basis of the space of symmetric functions is given by the sums of monomials in the same orbit, for all orbits.

Orbits correspond to non-increasing vectors $\lambda := (h_1 \geq h_2 \geq \cdots \geq h_n), h_i \in \mathbb{N}$, and we have set m_λ to be the sum of monomials in the corresponding orbit.

As we will soon see there are also some subtler symmetric functions (the Schur functions) indexed by partitions, and this will play an important role in the sequel. We can start with a first important fact, the explicit connection between the functions e_i and ψ_k. To see this connection, we will perform the next computations in the ring of formal power series, although the series that we will consider also have meaning as convergent series.

Start from the identity $\prod_{i=1}^{n}(tx_i+1)=\sum_{i=0}^{n} e_i t^i$ and take the logarithmic derivative (relative to the variable t) of both sides. We use the fact that such an operator transforms products into sums to get

$$\sum_{i=1}^{n}\frac{x_i}{(tx_i+1)}=\frac{\sum_{i=1}^{n} ie_it^{i-1}}{\sum_{i=0}^{n} e_it^i}.$$

The left-hand side of this formula can be developed as

$$\sum_{i=1}^{n} x_i\sum_{h=0}^{\infty}(-tx_i)^h=\sum_{h=0}^{\infty}(-t)^h\psi_{h+1}.$$

From this we get the identity

$$\left(\sum_{h=0}^{\infty}(-t)^h\psi_{h+1}\right)\left(\sum_{i=0}^{n} e_it^i\right)=\sum_{i=1}^{n} ie_it^{i-1}$$

which gives, equating coefficients:

$$(1.1.3)\quad (-1)^m\psi_{m+1}+\sum_{i=1}^{m}(-1)^i\psi_i e_{m+1-i}=\sum_{i+j=m}(-1)^i\psi_{i+1}e_j=(m+1)e_{m+1}$$

where we take $e_i=0$ if $i>n$.

It is clear that these formulas give recursive ways of expressing the ψ_i in terms of the e_j with integral coefficients. On the other hand, they can also be used to express the e_i in terms of the ψ_j, but in this case it is necessary to perform some division operations; the coefficients are rational and usually not integers.[6]

It is useful to give a second proof. Consider the map:

$$\pi_n : \mathbb{Z}[x_1, x_2, \dots, x_n] \to \mathbb{Z}[x_1, x_2, \dots, x_{n-1}]$$

given by evaluating x_n at 0.

Lemma. *The intersection of* $\mathrm{Ker}(\pi_n)$ *with the space of symmetric functions of degree* $< n$ *is reduced to* 0.

Proof. Consider $m_{(h_1,h_2,\dots,h_n)}$, a sum of monomials in an orbit. If the degree is less than n, we have $h_n = 0$. Under π_n we get $\pi_n(m_{(h_1,h_2,\dots,h_{n-1},0)}) = m_{(h_1,h_2,\dots,h_{n-1})}$. Thus if the degree is less than n, the map π_n maps these basis elements into distinct basis elements.

Now we give the second proof of 1.1.3. In the identity $\prod_{i=1}^{n}(t - x_i) := \sum_{i=0}^{n}(-1)^i e_i t^{n-i}$, substitute t with x_i, and then summing over all i we get (remark that $\psi_0 = n$):

$$0 = \sum_{i=0}^{n}(-1)^i e_i \psi_{n-i}, \text{ or } \psi_n = \sum_{i=1}^{n}(-1)^{i-1} e_i \psi_{n-i}.$$

By the previous lemma this identity also remains valid for symmetric functions in more than n variables and gives the required recursion.

1.2 Symmetric Polynomials

It is actually a general fact that symmetric functions can be expressed as polynomials in the elementary symmetric functions. We will now discuss an algorithmic proof.

To make the proof transparent, let us also stress in our formulas the number of variables and denote by $e_i^{(k)}$ the i^{th} elementary symmetric function in the variables $x_1, \dots, x_k$. Since

$$\left(\sum_{i=0}^{n-1} e_i^{(n-1)} t^i\right)(tx_n + 1) = \sum_{i=0}^{n} e_i^{(n)} t^i,$$

we have

$$e_i^{(n)} = e_{i-1}^{(n-1)} x_n + e_i^{(n-1)} \quad \text{or} \quad e_i^{(n-1)} = e_i^{(n)} - e_{i-1}^{(n-1)} x_n.$$

In particular, in the homomorphism $\pi : \mathbb{Z}[x_1, \dots, x_n] \to \mathbb{Z}[x_1, \dots, x_{n-1}]$ given by evaluating x_n at 0, we have that symmetric functions map to symmetric functions and

$$\pi(e_i^{(n)}) = e_i^{(n-1)}, \; i < n, \quad \pi(e_n^{(n)}) = 0.$$

[6] These formulas were found by Newton, hence the name *Newton functions* for the ψ_k.

Given a symmetric polynomial $f(x_1, \dots, x_n)$ we evaluate it at $x_n = 0$. If the resulting polynomial $f(x_1, \dots, x_{n-1}, 0)$ is 0, then f is divisible by x_n.

If so, by symmetry it is divisible by all of the variables and hence by the function e_n. We perform the division and move on to another symmetric function of lower degree.

Otherwise, by recursive induction one can construct a polynomial p in $n-1$ variables which, evaluated in the $n-1$ elementary symmetric functions of $x_1, \dots, x_{n-1}$, gives $f(x_1, \dots, x_{n-1}, 0)$. Thus $f - p(e_1, e_2, \dots, e_{n-1})$ is a symmetric function vanishing at $x_n = 0$.

We are back to the previous step.

The uniqueness is implicit in the algorithm which can be used to express any symmetric polynomial as a unique polynomial in the elementary symmetric functions.

Theorem 1. *A symmetric polynomial $f \in \mathbb{Z}[x_1, \dots, x_n]$ is a polynomial, in a unique way and with coefficients in $\mathbb{Z}$, in the elementary symmetric functions.*

It is quite useful, in view of the previous lemma and theorem, to apply the same ideas to symmetric functions in larger and larger sets of variables. One then constructs a *limit ring*, which one calls just the *formal ring of symmetric functions* $\mathbb{Z}[e_1, \dots, e_i, \dots]$. It can be thought of as the polynomial ring in infinitely many variables e_i, where formally we give degree (or weight) i to e_i. The ring of symmetric functions in n variables is obtained by setting $e_i = 0, \ \forall i > n$. One often develops formal identities in this ring with the idea that, in order to verify an identity which is homogeneous of some degree m, it is enough to do it for symmetric functions in m variables.

In the same way the reader may understand the following fact. Consider the $n!$ monomials

$$x_1^{h_1} \cdots x_{n-1}^{h_{n-1}}, \ 0 \le h_i \le n - i.$$

Theorem 2. *The above monomials are a basis of $\mathbb{Z}[x_1, \dots, x_n]$ over $\mathbb{Z}[e_1, \dots, e_n]$.*

Remark. The same theorem is clearly true if we replace the coefficient ring $\mathbb{Z}$ by any commutative ring A. In particular, we will use it when A is itself a polynomial ring.

2 Resultant, Discriminant, Bézoutiant

2.1 Polynomials and Roots

In order to understand the importance of Theorem 1 of 1.2 on elementary symmetric functions and also the classical point of view, let us develop a geometric picture.

Consider the space $\mathbb{C}^n$ and the space $P_n := \{t^n + b_1 t^{n-1} + \cdots + b_n\}$ of monic polynomials (which can be identified with $\mathbb{C}^n$ by the use of the coefficients).

Consider next the map $\pi : \mathbb{C}^n \to P_n$ given by

$$\pi(\alpha_1, \dots, \alpha_n) := \prod_{i=1}^{n}(t - \alpha_i).$$

We thus obtain a polynomial $t^n - a_1 t^{n-1} + a_2 t^{n-2} + \cdots + (-1)^n a_n = 0$ with roots $\alpha_1, \dots, \alpha_n$ (and the coefficients a_i are the elementary symmetric functions in the roots). Any monic polynomial is obtained in this way (Fundamental Theorem of Algebra).

Two points in $\mathbb{C}^n$ project to the same point in P_n if and only if they are in the same orbit under the symmetric group, i.e., P_n parameterizes the S_n-orbits.

Suppose we want to study a property of the roots which can be verified by evaluating some symmetric polynomials in the roots (this will usually be the case for any condition on the set of all roots). Then one can perform the computation without computing the roots, since one has only to study the formal symmetric polynomial expression and, using the alogrithm discussed in §1.2 (or any equivalent algorithm), express the value of a symmetric function of the roots through the coefficients.

In other words, a symmetric polynomial function f on $\mathbb{C}^n$ factors through the map π giving rise to an effectively computable[7] polynomial function $\overline{f}$ on P_n such that $f = \overline{f}\pi$.

A classical example is given by the *discriminant.*

The condition that the roots be distinct is clearly that $\prod_{i<j}(\alpha_i - \alpha_j) \neq 0$. The polynomial $V(x) := \prod_{i<j}(x_i - x_j)$ is in fact not symmetric. It is the value of the Vandermonde determinant, i.e., the determinant of the matrix:

$$A := \begin{pmatrix} x_1^{n-1} & x_2^{n-1} & \dots & x_n^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ x_1 & x_2 & \dots & x_n \\ 1 & 1 & \dots & 1 \end{pmatrix}. \tag{2.1.1}$$

Proposition 1. *$V(x)$ is antisymmetric, i.e., a permutation of the variables results in the multiplication of $V(x)$ by the sign of the permutation.*

Remark. The theory of the sign of permutations can be deduced by analyzing the Vandermonde determinant. In fact, since for a transposition τ it is clear that $V(x)^\tau = -V(x)$, it follows that $V(x)^\sigma = V(x)$ or $-V(x)$ according to whether σ is a product of an even or an odd number of transpositions. The sign is then clearly a homomorphism.

We also see immediately that V^2 is a symmetric polynomial. We can compute it in terms of the functions ψ_i as follows. Consider the matrix $B := AA^t$. Clearly in the i, j entry of B we find the symmetric function $\psi_{2n-(i+j)}$, and the determinant of B is V^2.

[7] I.e., computable without solving the equation, usually by polynomial expressions in the coefficients.

The matrix B (or rather the one reordered with ψ_{i+j-2} in the i, j position) is classically known as the *Bézoutiant*, and it carries some further information about the roots. We shall see that there is a different determinant formula for the determinant of B directly involving the elementary symmetric functions.

Let $D(e_1, e_2, \ldots, e_n)$ be the expression for V^2 as a polynomial in the elementary symmetric functions (e.g., $n = 2$, $\quad D = e_1^2 - 4e_2$).

Definition. The polynomial D is called the discriminant.

Since this is an interesting example we will pursue it a bit further.

Let us assume that F is a field, and $f(t)$ is a monic polynomial (of degree n) with coefficients in F, and let $R := F[t]/(f(t))$. We have that R is an algebra over F of dimension n.

For any finite-dimensional algebra A over a field F we can perform the following construction.

Any element a of A induces a linear transformation $L_a : x \to ax$ on A by left multiplication (and also one by right multiplication). We define $\mathrm{tr}(a) := \mathrm{tr}(L_a)$, the trace of the operator L_a.

We consider next the bilinear form $(a, b) := \mathrm{tr}(ab)$. This is the *trace form* of A. It is symmetric and *associative* in the sense that $(ab, c) = (a, bc)$.

We compute it first for $R := F[t]/(t^n)$. Using the fact that t is nilpotent we see that $\mathrm{tr}(t^k) = 0$ if $k > 0$. Thus the trace form has rank 1 with kernel the ideal generated by t.

To compute for the algebra $R := F[t]/(f(t))$ we pass to the algebraic closure $\overline{F}$ and compute in $\overline{F}[t]/(f(t))$.

We split the polynomial with respect to its distinct roots, $f(t) = \prod_{i=1}^{k}(t - \alpha_i)^{h_i}$, and $\overline{F}[t]/(f(t)) = \oplus_{i=1}^{k}\overline{F}[t]/(t - \alpha_i)^{h_i}$. Thus the trace of an element mod $f(t)$ is the sum of its traces mod $(t - \alpha_i)^{h_i}$.

Let us compute the trace of t^k mod $(t - \alpha_i)^{h_i}$. We claim that it is $h_i\alpha_i^k$. In fact in the basis $1, (t - \alpha_i), (t - \alpha_i)^2, \ldots, (t - \alpha_i)^{h_i - 1}$ (mod $(t - \alpha_i)^{h_i}$) the matrix of t is lower triangular with constant eigenvalue α_i on the diagonal, and so the claim follows.

Adding all of the contributions, we see that in $F[t]/(f(t))$, the trace of multiplication by t^k is $\sum_i h_i\alpha_i^k$, the k^{th} Newton function of the roots.

As a consequence we see that the matrix of the trace form, in the basis $1, t, \ldots, t^{n-1}$, is the Bézoutiant of the roots. Since for a given block $\overline{F}[t]/(t - \alpha_i)^{h_i}$ the ideal generated by $(t - \alpha_i)$ is nilpotent of codimension 1, we see that it is exactly the radical of the block, and the kernel of its trace form. It follows that:

Proposition 2. *The rank of the Bézoutiant equals the number of distinct roots.*

Given a polynomial $f(t)$ let $\overline{f}(t)$ denote the polynomial with the same roots as $f(t)$ but all distinct. It is the generator of the radical of the ideal generated by $f(t)$. In characteristic zero this polynomial is obtained dividing $f(t)$ by the GCD between $f(t)$ and its derivative $f'(t)$.

Let us consider next the algebra $R := F[t]/(f(t))$, its radical N and $\overline{R} := R/N$. By the previous analysis it is clear that $\overline{R} = F[t]/(\overline{f}(t))$.

Consider now the special case in which $F = \mathbb{R}$ is the field of real numbers. Then we can divide the distinct roots into the real roots $\alpha_1, \alpha_2, \dots, \alpha_k$ and the complex ones $\beta_1, \overline{\beta}_1, \beta_2, \overline{\beta}_2, \dots, \beta_h, \overline{\beta}_h$.

The algebra $\overline{R}$ is isomorphic to the direct sum of k copies of $\mathbb{R}$ and h copies of $\mathbb{C}$. Its trace form is the orthogonal sum of the corresponding trace forms. Over $\mathbb{R}$ the trace form is just x^2 but over $\mathbb{C}$ we have $\operatorname{tr}((x+iy)^2) = 2(x^2 - y^2)$. We deduce:

Theorem. *The number of real roots of $f(t)$ equals the signature*[8] *of its Bézoutiant.*

As a simple but important corollary we have:

Corollary. *A real polynomial has all its roots real and distinct if and only if the Bézoutiant is positive definite.*

There are simple variations on this theme. For instance, if we consider the quadratic form $Q(x) := \operatorname{tr}(tx^2)$ we see that its matrix is again easily computed in terms of the ψ_k and its signature equals the number of real positive roots minus the number of real negative roots. In this way one can also determine the number of real roots in any interval.

These results are Sylvester's variations on Sturm's theorem. They can be found in the paper in which he discusses the Law of Inertia that now bears his name (cf. [Si]).

2.2 Resultant

Let us go back to the roots. If $x_1, x_2, \dots, x_n; y_1, y_2, \dots, y_m$ are two sets of variables, consider the polynomial

$$A(x, y) := \prod_{i=1}^{n}\prod_{j=1}^{m}(x_i - y_j).$$

This is clearly symmetric, separately in the variables x and y. If we evaluate it in numbers, it vanishes if and only if one of the values of the x's coincides with a value of the y's. Conversely, any polynomial in these two sets of variables that has this property is divisible by all the factors $x_i - y_j$, and hence it is a multiple of A.

By the general theory A, a symmetric polynomial in the x_i's, can be expressed as a polynomial R in the elementary symmetric functions $e_i(x)$ with coefficients that are polynomials symmetric in the y_j. These coefficients are thus in turn polynomials in the elementary symmetric functions of the y_j's.

Let us denote by $a_1, \dots, a_n$ the elementary symmetric functions in the x_i's and by $b_1, \dots, b_m$ the ones in the y_j's. Thus $A(x, y) = R(a_1, \dots, a_n, b_1, \dots, b_m)$ for some explicit polynomial R.

[8] The Bézoutiant is a real symmetric matrix; for such a matrix the notion of *signature* is explained in Chapter 5, 3.3. There are effective algorithms to compute the signature.

The polynomial R is called the *resultant.*

When we evaluate the variables x and y to be the roots of two monic polynomials $f(t),\ g(t)$ of degrees n, m, respectively, we see that the value of A can be computed by evaluating R in the coefficients (with some signs) of these polynomials. Thus the resultant is a polynomial in their coefficients, vanishing when the two polynomials have a common root.

There is a more general classical expression for the resultant as a determinant, and we drop the condition that the polynomials be monic. The theory is the following.

Let $f(t) := a_0t^n + a_1t^{n-1} + \cdots + a_n$, $g(t) := b_0t^m + b_1t^{m-1} + \cdots + b_m$ and let us denote by P_h the $h+1$-dimensional space of all polynomials of degree $\leq h$.

Consider the linear transformation:

$$T_{f,g} : P_{m-1} \oplus P_{n-1} \to P_{m+n-1} \text{ given by } T_{f,g}(a,b) := fa + gb.$$

This is a transformation between two $n+m$-dimensional spaces and, in the bases $(1,0), (t,0), \dots, (t^{m-1},0), (0,1), (0,t), \dots, (0,t^{n-1})$ and $1, t, t^2, \dots, t^{n+m-1}$, it is quite easy to write down its square matrix $R_{f,g}$:

$$(2.2.1) \qquad \begin{pmatrix}
a_n & 0 & 0 & \dots & 0 & b_m & 0 & \dots & 0 & 0 & 0 \\
a_{n-1} & a_n & 0 & \dots & 0 & b_{m-1} & b_m & \dots & \dots & \ddots & \\
a_{n-2} & a_{n-1} & a_n & 0 & \dots & b_{m-2} & b_{m-1} & b_m & \ddots & \ddots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\
a_1 & a_2 & a_3 & & & & & & \ddots & \ddots & \vdots \\
a_0 & a_1 & a_2 & & & & & & & & \\
0 & a_0 & a_1 & & & & & & \ddots & \ddots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\
0 & \vdots & & & & b_0 & b_1 & b_2 & \ddots & \vdots & \\
0 & 0 & \dots & & & 0 & b_0 & b_1 & \ddots & \ddots & \vdots \\
0 & 0 & & & & 0 & 0 & b_0 & \ddots & \ddots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & b_0 & \vdots \\
0 & 0 & 0 & \dots & a_0 & 0 & \dots & \dots & 0 & 0 & b_0
\end{pmatrix}.$$

Proposition. *If $a_0b_0 \neq 0$, the rank of $T_{f,g}$ equals $m+n-d$ where d is the degree of $h := \mathrm{GCD}(f,g)$*[9].

Proof. By Euclid's algorithm the image of $T_{f,g}$ consists of all polynomials of degree $\leq n+m-1$ and multiples of h. Its kernel consists of pairs $(sg', -sf')$ where $f = hf'$, $g = hg'$. The claim follows.

[9] $\mathrm{GCD}(f,g)$ is the greatest common divisor of f, g.

As a corollary we have that the determinant $R(f, g)$ of $R_{f,g}$ vanishes exactly when the two polynomials have a common root. This gives us a second definition of resultant.

Definition. The polynomial $R(f, g)$ is called the *resultant* of the two polynomials $f(t), g(t)$.

If we consider the coefficients of f and g as variables, we can still think of $T_{f,g}$ as a map of vector spaces, except that the base field is the field of rational functions in the given variables. Then we can solve the equation $fa + gb = 1$ by Cramer's rule and we see that the coefficients of the polynomials a, b are given by the cofactors of the first row of the matrix $R_{f,g}$ divided by the resultant. In particular, we can write $R = Af(t) + Bg(t)$ where A, B are polynomials in t of degrees $m-1, n-1$, respectively, and with coefficients polynomials in the variables $(a_0, a_1, \dots, a_n, b_0, b_1, \dots, b_m)$.

This can also be understood as follows. In the matrix $R_{f,g}$ we add to the first row the second multiplied by t, the third multiplied by t^2, and so on. We see that the first row becomes $(f(t), f(t)t, f(t)t^2, \dots, f(t)t^{m-1}, g(t), g(t)t, g(t)t^2, \dots, g(t)t^{n-1})$. Under these operations of course the determinant does not change. Then developing it along the first row we get the desired identity.

We have given two different definitions of resultant, which we need to compare:

Exercise. Consider the two polynomials as $a_0 \prod_{i=1}^{n}(t - x_i), b_0 \prod_{j=1}^{m}(t - y_j)$ and thus, in R, substitute the element $(-1)^i a_0 e_i(x_1, \dots, x_n)$ for the variables a_i and the element $(-1)^i b_0 e_i(y_1, \dots, y_m)$ for b_i. The polynomial we obtain is $a_0^m b_0^n A(x, y)$.

2.3 Discriminant

In the special case when we take $g(t) = f'(t)$, the derivative of $f(t)$, we have that the vanishing of the resultant is equivalent to the existence of multiple roots. We have already seen that the vanishing of the discriminant implies the existence of multiple roots. It is now easy to connect the two approaches.

The resultant $R(f, f')$ is considered as a polynomial in the variables $(a_0, a_1, \dots, a_n)$. If we substitute in $R(f, f')$ the element $(-1)^i a_0 e_i(x_1, \dots, x_n)$ for the variables a_i we have a polynomial in the x with coefficients involving a_0 that vanishes whenever two x's coincide.

Thus $R(f, f')$ is divisible by the discriminant D of these variables. A degree computation shows in fact that it is a constant (with respect to the x) multiple cD. The constant c can be evaluated easily, for instance specializing to the polynomial $x^n - 1$. This polynomial has as roots the n^{th} roots $e^{2\pi ik/n}$, $0 \le k < n$ of 1. The Newton functions

$$\psi_h := \sum_{i=0}^{n-1} e^{\frac{2\pi ihk}{n}} = \begin{cases} 0 & \text{if } \quad h \not| n \\ n & \text{if } \quad h \mid n; \end{cases}$$

hence the Bézoutiant is $-(-n)^n$ and the computation of the resultant is n^n, so the constant is $(-1)^{n-1}$.

3 Schur Functions

3.1 Alternating Functions

Along with symmetric functions, it is also important to discuss alternating (or skew-symmetric, or antisymmetric) functions. We restrict our considerations to integral polynomials.

Definition. A polynomial f in the variables $(x_1, x_2, \dots, x_n)$ is called an alternating function if, for every permutation σ of these variables,

$$f^\sigma = f(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) = \epsilon_\sigma f(x_1, x_2, \dots, x_n),$$

ϵ_σ being the sign of the permutation.

We have seen the Vandermonde determinant $V(x) := \prod_{i<j}(x_i - x_j)$ as a basic alternating polynomial. The main remark on alternating functions is the following.

Proposition 1. *A polynomial $f(x)$, in the variables x, is alternating if and only if it is of the form $f(x) = V(x)g(x)$, with $g(x)$ a symmetric polynomial.*

Proof. Substitute, in an alternating polynomial f, for a variable x_j a variable x_i for $i \neq j$. We get the same polynomial if we first exchange x_i and x_j in f. Since this changes the sign, it means that under this substitution f becomes 0.

This means in turn that f is divisible by $x_i - x_j$; since i, j are arbitrary, f is divisible by $V(x)$. Writing $f = V(x)g$, it is clear that g is symmetric.

Let us be more formal. Let A, S denote the sets of *antisymmetric* and *symmetric* polynomials. We have seen that:

Proposition 2. *The space A of antisymmetric polynomials is a free rank* 1 *module over the ring S of symmetric polynomials generated by $V(x)$ or $A = V(x)S$.*

In particular, any integral basis of A gives, dividing by $V(x)$, an integral basis of S. In this way we will presently obtain the *Schur functions.*

To understand the construction, let us make a fairly general discussion. In the ring of polynomials $\mathbb{Z}[x_1, x_2, \dots, x_n]$, let us consider the basis given by the monomials (which are permuted by S_n).

Recall that the orbits of monomials are indexed by non-increasing sequences of nonnegative integers. To $m_1 \geq m_2 \geq m_3 \cdots \geq m_n \geq 0$ corresponds the orbit of the monomial $x_1^{m_1} x_2^{m_2} x_3^{m_3} \cdots x_n^{m_n}$.

Let f be an antisymmetric polynomial and (ij) a transposition. Applying this transposition to f changes the sign of f, while the transposition fixes all monomials in which x_i, x_j have the same exponent.

It follows that all of the monomials which have nonzero coefficient in f must have distinct exponents. Given a sequence of exponents $m_1 > m_2 > m_3 > \cdots > m_n \geq 0$ the coefficients of the monomial $x_1^{m_1} x_2^{m_2} x_3^{m_3} \cdots x_n^{m_n}$ and of $x_{\sigma(1)}^{m_1} x_{\sigma(2)}^{m_2} x_{\sigma(3)}^{m_3} \cdots x_{\sigma(n)}^{m_n}$ differ only by the sign of σ.

It follows that:

Theorem. *The functions*

$$A_{m_1>m_2>m_3>\cdots>m_n\geq 0}(x) := \sum_{\sigma\in S_n} \epsilon_\sigma x_{\sigma(1)}^{m_1} x_{\sigma(2)}^{m_2} \cdots x_{\sigma(n)}^{m_n} \tag{3.1.1}$$

are an integral basis of the space of antisymmetric polynomials.

It is often useful when making computations with alternating functions to use a simple device. Consider the subspace SM spanned by the set of *standard monomials* $x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}$ with $k_1 > k_2 > k_3 \cdots > k_n$ and the linear map L from the space of polynomials to SM which is 0 on the nonstandard monomials and the identity on SM. Then $L\big(\sum_{\sigma\in S_n} \epsilon_\sigma x_{\sigma(1)}^{m_1} x_{\sigma(2)}^{m_2} \cdots x_{\sigma(n)}\big) = x_1^{m_1}x_2^{m_2}\cdots x_n^{m_n}$, and thus L establishes a linear isomorphism between the space of alternating polynomials and SM which maps the basis of the theorem to the standard monomials.

3.2 Schur Functions

It is convenient to use the following conventions. Consider the sequence $\varrho := (n-1, n-2, \ldots, 2, 1, 0)$. We clearly have:

Lemma. *The map*

$$\begin{aligned}\lambda = (p_1, p_2, p_3, \ldots, p_n) &\mapsto \lambda + \varrho \\ &= (p_1 + n - 1, p_2 + n - 2, p_3 + n - 3, \ldots, p_n)\end{aligned}$$

is a bijective correspondence between decreasing and strictly decreasing sequences.

We thus indicate by $A_{\lambda+\varrho}$ the corresponding antisymmetric function. We can express it also as a determinant of the matrix M_λ having the element $x_j^{p_i+n-i}$ in the i, j position.[10]

Definition. The symmetric function $S_\lambda(x) := A_{\lambda+\varrho}/V(x)$ is called the *Schur function* associated to λ.

When there is no ambiguity we will drop the symbol of the variables x and use S_λ.

We can identify λ with a *partition*, with at most n parts, of the integer $\sum p_i$ and write $\lambda \vdash \sum_i p_i$.

Thus we have (with the notations of Chapter 1, 1.1) the following:

Theorem 1. *The functions S_λ, with $\lambda \vdash m$ and $ht(\lambda) \leq n$, are an integral basis of the part of degree m of the ring of symmetric functions in n variables.*

Notice that the Vandermonde determinant is the alternating function A_ϱ and $S_0 = 1$.

[10] It is conventional to drop the numbers equal to 0 in a decreasing sequence.

Several interesting combinatorial facts are associated to these functions; we will see some of them in the next section. The main significance of the Schur functions is in the representation theory of the linear group, as we will see later in Chapter 9.

If a is a positive integer let us denote by $\underline{a}$ the partition $(a, a, a, \dots, a)$. If $\lambda = (p_1, p_2, p_3, \dots, p_n)$ is a partition from 3.1.1, it follows that

(3.2.1) $$A_{\lambda+\varrho+\underline{a}} = (x_1 x_2 \cdots x_n)^a A_{\lambda+\varrho}, \quad S_{\lambda+\underline{a}} = (x_1 x_2 \cdots x_n)^a S_\lambda.$$

We let n be the number of variables and want to understand given a Schur function $S_\lambda(x_1, \dots, x_n)$ the form of $S_\lambda(x_1, \dots, x_{n-1}, 0)$ as symmetric function in $n-1$ variables.

Let $\lambda := h_1 \geq h_2 \geq \cdots \geq h_n \geq 0$. We have seen that, if $h_n > 0$, then $S_\lambda(x_1, \dots, x_n) = \prod_{i=1}^n x_i S_{\overline{\lambda}}(x_1, \dots, x_n)$ where $\overline{\lambda} := h_1 - 1 \geq h_2 - 1 \geq \cdots \geq h_n - 1$. In this case, clearly $S_\lambda(x_1, \dots, x_{n-1}, 0) = 0$.

Assume now $h_n = 0$ and denote the sequence $h_1 \geq h_2 \geq \cdots \geq h_{n-1}$ by the same symbol λ. Let us start from the Vandermonde determinant $V(x_1, \dots, x_{n-1}, x_n) = \prod_{i<j\leq n}(x_i - x_j)$ and set $x_n = 0$ to obtain

$$V(x_1, \dots, x_{n-1}, 0) = \prod_{i=1}^{n-1} x_i \prod_{i<j\leq n-1} (x_i - x_j) = \prod_{i=1}^{n-1} x_i V(x_1, \dots, x_{n-1}).$$

Now consider the alternating function $A_{\lambda+\varrho}(x_1, \dots, x_{n-1}, x_n)$.

Set $\ell_i := h_i + n - i$ so that $\ell_n = 0$ and

$$A_{\lambda+\varrho}(x_1, \dots, x_{n-1}, x_n) = \sum_{\sigma \in S_n} \epsilon_\sigma x_1^{\ell_{\sigma(1)}} \cdots x_n^{\ell_{\sigma(n)}}.$$

Setting $x_n = 0$ we get the sum restricted only to the terms for which $\sigma(n) = n$ or

$$A_{\lambda+\varrho}(x_1, \dots, x_{n-1}, 0) = \sum_{\sigma \in S_{n-1}} \epsilon_\sigma x_1^{\ell_{\sigma(1)}} \cdots x_{n-1}^{\ell_{\sigma(n-1)}}.$$

Now $\ell_i = h_i + n - i = (h_i + 1) + (n-1) - i$, and so in $(n-1)$ variables,

$$A_{\lambda+\varrho}(x_1, \dots, x_{n-1}, 0) = A_{\lambda+\varrho+\underline{1}}(x_1, \dots, x_{n-1}) = \prod_{i=1}^{n-1} x_i A_\lambda(x_1, \dots, x_{n-1}).$$

It follows that $S_\lambda(x_1, \dots, x_{n-1}, 0) = S_\lambda(x_1, \dots, x_{n-1})$. Thus we see that:

Proposition. *Under the evaluation of x_n at 0, the Schur function S_λ vanishes when $ht(\lambda) = n$. Otherwise it maps to the corresponding Schur function in $(n-1)$ variables.*

One uses these remarks as follows. Consider a fixed degree n, and for any m let S_m^n be the space of symmetric functions of degree n in m variables.

From the theory of Schur functions the space S_m^n has as basis the functions $S_\lambda(x_1, \dots, x_m)$ where $\lambda \vdash n$ has height $\leq m$. Under the evaluation $x_m \mapsto 0$, we have a map $S^n_m \to S^n_{m-1}$. We have proved that this map is an isomorphism as soon as $m > n$.

We recover the lemma of Section 1.1 of this chapter and the consequence that all identities which we prove for symmetric functions in n variables of degree n are valid in any number of variables.

Theorem 2. *The formal ring of symmetric functions in infinitely many variables has as basis all Schur functions S_λ. Restriction to symmetric functions in m variables sets to 0 all S_λ with height $> m$.*

When using partitions it is often more useful to describe a partition by specifying the number of parts with 1 element, the number of parts with 2 elements, and so on. Thus one writes a partition as $1^{a_1} 2^{a_2} \dots i^{a_i} \dots$.

Proposition. *For the elementary symmetric functions we have*

$$e_h = S_{1^h}. \tag{3.2.2}$$

Proof. According to our previous discussion we can set all the variables x_i, $i > h$ equal to 0. Then e_h reduces to $\prod_{i=1}^h x_i$ as well as S_{1^h} from 3.2.1. □

3.3 Duality

Next we want to discuss the value of $S_\lambda(1/x_1, 1/x_2, \dots, 1/x_n)$.

We see that substituting x_i with $1/x_i$ in the matrix M_λ (cf. §3.2) and multiplying the j^{th} column by $x_j^{m_1+n-1}$, we obtain a matrix which equals, up to rearranging the rows, that of the partition $\lambda' := m'_1, m'_2, \dots, m'_n$ where $m_i + m'_{n-i+1} = m_1$. Thus, up to a sign,

$$(x_1 x_2 \cdots x_n)^{m_1+n-1} A_{\lambda+\varrho}(1/x_1, \dots, 1/x_n) = A_{\lambda'+\varrho}.$$

For the Schur function we have to apply the procedure to both numerator and denominator so that the signs cancel, and we get $S_\lambda(1/x_1, 1/x_2, \dots, 1/x_n) = (x_1 x_2 \cdots x_n)^{-m_1} S_{\lambda'}$.

If we use the diagram notation for partitions we easily visualize λ' by inserting λ in a rectangle of base m_1 and then taking its complement.

4 Cauchy Formulas

4.1 Cauchy Formulas

The formulas we want to discuss have important applications in representation theory. For now, we wish to present them as purely combinatorial identities.

$$\prod_{i,j=1} \frac{1}{1 - x_i y_j} = \sum_\lambda S_\lambda(x) S_\lambda(y), \tag{C1}$$

where the right-hand side is the sum over all partitions.

(C2)
$$\prod_{1\le i\le j\le 2m} \frac{1}{1-x_ix_j} = \sum_{\lambda\in\Lambda_{ec}} S_\lambda(x),$$

if $n = 2m$ is even.

For all n,

(C3)
$$\prod_{1\le i<j\le n} \frac{1}{1-x_ix_j} = \sum_{\lambda\in\Lambda_{er}} S_\lambda(x).$$

Here Λ_{ec}, resp. Λ_{er}, indicates the set of diagrams with columns (resp. rows) of even length.

(C4)
$$\prod_{i=1,\ j=1}^{n,\ m} (1+x_iy_j) = \sum_\lambda S_\lambda(x)S_{\tilde\lambda}(y),$$

where $\tilde\lambda$ denotes the dual partition (Chapter 1, 1.1) obtained by exchanging rows and columns.

We prove the first one and leave the others to Chapter 9 and 11, where they are interpreted as character formulas. We offer two proofs:

First proof of C1. It can be deduced (in a way similar to the computation of the Vandermonde determinant) considering the determinant of the $n \times n$ matrix:

$$A := (a_{ij}),\ \text{with } a_{ij} = \frac{1}{1-x_iy_j}.$$

We first prove that we have

(4.1.1)
$$\frac{V(x)V(y)}{\prod_{i,j=1}^{n}(1-x_iy_j)} = \det(A).$$

Subtracting the first row from the i^{th}, $i > 1$, one has a new matrix (b_{ij}) where

$$b_{1j} = a_{1j},\ \text{and for } i>1,\ b_{ij} = \frac{1}{1-x_iy_j} - \frac{1}{1-x_1y_j} = \frac{(x_i-x_1)y_j}{(1-x_iy_j)(1-x_1y_j)}.$$

Thus from the i^{th} row, $i > 1$, one can extract from the determinant the factor $x_i - x_1$ and from the j^{th} column the factor $\frac{1}{1-x_1y_j}$.

Thus the given determinant is the product $\frac{1}{(1-x_1y_1)}\prod_{i=2}^{n}\frac{(x_i-x_1)}{(1-x_1y_i)}$ with the determinant

(4.1.2)
$$\begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ \frac{y_1}{1-x_2y_1} & \frac{y_2}{1-x_2y_2} & \dots & \dots & & \frac{y_n}{1-x_2y_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{y_1}{1-x_ny_1} & \frac{y_2}{1-x_ny_2} & \dots & \dots & & \frac{y_n}{1-x_ny_n} \end{pmatrix}.$$

Subtracting the first column from the i^{th} we get the terms $\frac{y_i - y_1}{(1-x_j y_1)(1-x_j y_i)}$. Thus, after extracting the product $\prod_{i=2}^{n} \frac{(y_i - y_1)}{(1-x_i y_1)}$, we are left with the determinant of the same type of matrix but without the variables x_1, y_1. The claim follows by induction.

Now we can develop the determinant of A by developing each element $\frac{1}{1-x_i y_j} = \sum_{k=0}^{\infty} x_i^k y_j^k$, or in matrix form, each row (resp. column) as a sum of infinitely many rows (or columns).

By multilinearity in the rows, the determinant is a sum of determinants of matrices:

$$\sum_{k_1=0}^{\infty} \cdots \sum_{k_n=0}^{\infty} \det(A_{k_1,k_2,\dots,k_n}), \quad A_{k_1,k_2,\dots,k_n} := ((x_i y_j)^{k_i}).$$

Clearly $\det(A_{k_1,k_2,\dots,k_n}) = \prod_i x_i^{k_i} \det(y_j^{k_i})$. This is zero if the k_i are not distinct; otherwise we reorder the sequence k_i to be decreasing. At the same time we must introduce a sign. Collecting all of the terms in which the k_i are a permutation of a given sequence $\lambda + \rho$, we get the term $A_{\lambda+\varrho}(x)A_{\lambda+\varrho}(y)$. Finally,

$$\frac{V(x)V(y)}{\prod_{i,j=1,n}(1 - x_i y_j)} = \sum_{\lambda} A_{\lambda+\varrho}(x)A_{\lambda+\varrho}(y).$$

From this the required identity follows. □

Second proof of C1. Change the matrix to $\frac{1}{x_i - y_j}$ using the fact that

$$V(x_1^{-1}, \dots, x_n^{-1}) = \left(-\prod_i x_i\right)^{1-n} V(x_1, \dots, x_n)$$

and develop the determinant as the sum of fractions $\frac{1}{\prod(x_i - y_{\sigma(i)})}$. Writing it as a rational function $\frac{f(x,y)}{\prod_{i,j=1,n}(x_i - y_j)}$, we see immediately that $f(x, y)$ is alternating in both x, y of total degree $n^2 - n$. Hence $f(x, y) = cV(x)V(y)$ for some constant c, which will appear in the formula $C1$. Comparing in degree 0 we see that $C1$ holds. □

Let us remark that Cauchy formula C1 also holds when $m \leq n$, since $\prod_{i=1}^{n}\prod_{j=1}^{m} \frac{1}{1-x_i y_j}$ is obtained from $\prod_{i=1}^{n}\prod_{j=1}^{m} \frac{1}{1-x_i y_j}$ by setting $y_j = 0, \forall m < j \leq n$.

From Proposition 3.2 we get

$$\prod_{i=1}^{n}\prod_{j=1}^{m} \frac{1}{1 - x_i y_j} = \sum_{\lambda \vdash n,\, ht(\lambda) \leq m} S_\lambda(x_1, \dots, x_n) S_\lambda(y_1, \dots, y_m).$$

Remark. The theory of symmetric functions is in fact a rather large chapter in mathematics with many applications to algebra, combinatorics, probability theory, etc. The reader is referred to the book of I.G. Macdonald [Mac] for a more extensive treatment.

5 The Conjugation Action

5.1 Conjugation

Here we study a representation closely connected to the theory of symmetric functions.

Let us consider the space $M_n(\mathbb{C})$ of $n \times n$ matrices over the field $\mathbb{C}$ of complex numbers. We view it as a representation of the group $G := GL(n, \mathbb{C})$ of invertible matrices by conjugation: XAX^{-1}; its orbits are thus the conjugacy classes of matrices.

Remark. The scalar matrices $\mathbb{C}^*$ act trivially, hence we have a representation of the quotient group (the *projective linear group*):

$$PGL(n, \mathbb{C}) := GL(n, \mathbb{C})/\mathbb{C}^*.$$

Given a matrix A consider its characteristic polynomial:

$$\det(t - A) := \sum_{i=0}^{n} (-1)^i \sigma_i(A) t^{n-i}.$$

The coefficients $\sigma_i(A)$ are polynomial functions on $M_n(\mathbb{C})$ which are clearly conjugation invariant. Since the eigenvalues are the roots of the characteristic polynomial, $\sigma_i(A)$ is the i^{th} elementary symmetric function computed in the eigenvalues of A.

Recall that S_n can be viewed as a subgroup of $GL(n, \mathbb{C})$ (the permutation matrices). Consider the subspace D of diagonal matrices. Setting $a_{ii} = a_i$ we identify such a matrix with the vector $(a_1, \dots, a_n)$. The following is clear.

Lemma. *D is stable under conjugation by S_n. The induced action is the standard permutation action (2.6). The function $\sigma_i(A)$, restricted to D, becomes the i^{th} elementary symmetric function.*

We want to consider the conjugation action on $M_n(\mathbb{C})$, $GL(n, \mathbb{C})$, $SL(n, \mathbb{C})$ and compute the invariant functions. As functions we will take those which come from the algebraic structure of these sets (as affine varieties, cf. Chapter 7). Namely, on $M_n(\mathbb{C})$ we take the polynomial functions: On $SL(n, \mathbb{C})$ the restriction of the polynomial functions, and on $GL(n, \mathbb{C})$ the regular functions, i.e., the quotients f/d^k where f is a polynomial on $M_n(\mathbb{C})$ and d is the determinant function.

Theorem. *Any polynomial invariant for the conjugation action on $M_n(\mathbb{C})$ is a polynomial in the functions $\sigma_i(A)$, $i = 1, \dots, n$.*

Any invariant for the conjugation action on $SL(n, \mathbb{C})$ is a polynomial in the functions $\sigma_i(A)$, $i = 1, \dots, n - 1$.

Any invariant for the conjugation action on $GL(n, \mathbb{C})$ is a polynomial in the functions $\sigma_i(A)$, $i = 1, \dots, n$ and in $\sigma_n(A)^{-1}$.

Proof. We prove the first statement of the theorem. The proofs of the other two statements are similar and we leave them to the reader. Let $f(A)$ be such a polynomial. Restrict f to D. By the previous remark, it becomes a symmetric polynomial which can then be expressed as a polynomial in the elementary symmetric functions. Thus we can find a polynomial $p(A) = p(\sigma_1(A), \dots, \sigma_n(A))$ which coincides with $f(A)$ upon restriction to D. Since both $f(A)$, $p(A)$ are invariant under conjugation, they must coincide also on the set of all diagonalizable matrices. The statement follows therefore from:

Exercise. The set of diagonalizable matrices is dense.

Hint.

(i) A matrix with distinct eigenvalues is diagonalizable, and these matrices are characterized by the fact that the discriminant is nonzero on them.
(ii) For every integer k, the set of points in $\mathbb{C}^k$ where a (non-identically zero) polynomial $u(x)$ is nonzero is dense. (Take any point P and a P_0 with $g(P_0) \neq 0$, on the line connecting P, P_0 the polynomial g is not identically 0, etc.). □

Remark. The map $M_n(\mathbb{C}) \to \mathbb{C}^n$ given by the functions $\sigma_i(A)$ is constant on orbits, but a fiber is not necessarily a conjugacy class. In fact when the characteristic polynomial has a multiple root, there are several types of Jordan canonical forms corresponding to the same eigenvalues.

There is a second approach to the theorem which is also very interesting and leads to some generalizations. We omit the details.

Proposition. *For an $n \times n$ matrix A the following conditions are equivalent:*

(1) There is a vector v such that the n vectors $A^i v$, $i = 0, \dots, n-1$, are linearly independent.
(2) The minimal polynomial of A equals its characteristic polynomial.
(3) The conjugacy class of A has maximal dimension $n^2 - n$.
(4) A is conjugate to a companion matrix

$$\begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & a_n \\ 1 & 0 & 0 & \dots & 0 & 0 & a_{n-1} \\ 0 & 1 & 0 & \dots & 0 & 0 & a_{n-2} \\ 0 & 0 & 1 & \dots & 0 & 0 & a_{n-3} \\ \dots & & & & \dots & & \dots \\ \dots & & & & \dots & & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 & a_2 \\ 0 & 0 & 0 & \dots & 0 & 1 & a_1 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}$$

with characteristic polynomial $t^n + \sum_{i=1}^n a_i t^{n-i}$.
(5) In a Jordan canonical form distinct blocks belong to different eigenvalues.

Proof. (1) and (4) are clearly equivalent, taking as the matrix conjugate to A the one of the same linear transformation in the basis $A^i v$, $i = 0, \dots, n-1$.

(2) and (5) are easily seen to be equivalent and also (5) and (1).

We do not prove (3) since we have not yet developed enough geometry of orbits. One needs the theory of Chapter 4, 3.7 showing that the dimension of an orbit equals the dimension of the group minus the dimension of the stabilizer and then one has to compute the centralizer of a regular matrix and prove that it has dimension n.

Definition. The matrices satisfying the previous conditions are called *regular*, and their set is the *regular set* or *regular sheet*.

One can easily prove that the regular sheet is open dense, and it follows again that every invariant function is determined by the value it takes on the set of companion matrices; hence we have a new proof of the theorem on invariants for the conjugation representation.

With this example we have given a glance at a set of algebro-geometric phenomena which have been studied in depth by several authors. The representations for which the same type of ideas apply are particularly simple and interesting (cf. [DK]).