Conventional notations.

November 1995

When we introduce a new symbol or definition we will use the convenient form := which means that the term introduced at its left is defined by the expression at its right. A typical example could be $P := \{x \in \mathbb{N} | 2 \text{ divides } x\}$ which stands for: *P is by definition the set of all natural numbers x such that 2 divides x.*

The symbol $\pi : A \to B$ denotes a mapping named π from the set A to the set B.

Most of our work will be for algebras over the field of real or complex numbers, sometimes we will take a more combinatorial point of view and analyze some properties over the integers. Associative algebras will implicitely be assumed to have a unit elment. When we discuss matrices over a ring A we always identify A with the scalar matrices.

We use the standard notations:

$$\mathbb{N},\ \mathbb{Z},\ \mathbb{Q},\ \mathbb{R},\ \mathbb{C}$$

for the natural numbers (including 0), the integers, rational, real and complex numbers.

We shall try to explain in detail all the constructions which belong to invariant theory and instead introduce and use, sending the reader to other texts, the notions of differential or algebraic geometry or of measure theory and functional analysis which might be necessary for the treatment. In general our point of view is that some of the interesting special objects under consideration may be treated by more direct and elementary methods, and we will try to do so whenever possible, since a direct approach often reveals some of the special features which are lost in a general theory.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

General methods and ideas

November 1995

Contents

Introduction In this chapter we will develop the formal language a nd some general methods and theorems, to some extent the reader is advised not to read it too systematically since most of the interesting examples will appear only in the next chapters. The exposition here is quite far from the classical point of view since we are forced to establish the language in a rather thin general setting. Hopefully this will be repaid in the chapters in which we will treat the interesting results of Invariant Theory.

1 Groups and their actions

1.1 In our treatment groups will appear always as transformation groups, the main point being that, given a set X the set of all bijective mappings of X into X is a group under composition. We will denote this group $S(X)$ and call it *the symmetric group* of X.

In practice the full symmetric group is used only for X finite, in this case it is usually more convenient to fix as X the set $[1, n]$ formed by the first n integers (for a given value of n); in this case the corresponding symmetric group has $n!$ elements and it is indicated by S_n, its elements are called *permutations.*

In general the groups wich appear are subgroups of the full symmetric group, defined by special properties of the set X arising from some extra structure (like a topology or the structure of a linear space etc.), the groups of our interest will usually be symmetry groups of the structure under consideration.

To illustrate this concept we start:

Definition. *A partition of a set X is a family of non empty disjoint subsets A_i with union X.*

A partition of a number n is a (non increasing) sequence of positive numbers:

$$m_1 \geq m_2 \geq \ldots \geq m_k > 0 \textit{ with } \sum_{j=1}^{k} m_j = n.$$

Remark. To a partition of the set $[1, 2, \ldots, n]$ we can associate the partition of n given by the cardinality of the sets.

We represent graphycally such a partition by a **Young diagram** e.g. (8,5,5,2):

If $X = \cup A_i$ is a partition, the set:

$$G := \{\sigma \in S_n | \sigma(A_i) = A_i, \ \forall i\},$$

is a subgroup of $S(X)$, isomorphic to the product $\prod S(A_i)$ of the symmetric groups on the sets A_i.

1.2 It is useful at this stage to procede in a formal way. We set:

Definition. *An action of a group G on a set X is a mapping $\pi : G \times X \to X$, denoted by $gx := \pi(g, x)$ satisfying the following conditions:*

$$1x = x, \quad h(kx) = (hk)x \tag{1.2.1}$$

for all $h, k \in G$ and $x \in X$.

The reader will note that the definition just given can be reformulated as follows:

i) The map $\varrho(h) := x \to hx$ from X to X is bijective for all $h \in G$.

ii) The map $\varrho : G \to S(X)$ is a *group homomorphism.*

In our theory we will usually fix our attention on a given group G and consider different actions of the group, it is then convenient to refer to a given action on a set X as to a *G-set*.

Examples. a) The action of G as left multiplications on itself.

b) The action of G on the set $G/H := \{gH | g \in G\}$ given by:

$$a(bH) := abH. \tag{1.2.2}$$

c) The action of $G \times G$ on G given by $(a, b)c = abc^{-1}$.

d) The action of G by conjugation on itself.

e) The action of a subgroup of G induced by restricting an action of G.

It is useful right from the start to use a *categorical* language:

Definition. *Given two $G-$sets X, Y, a $G-$equivariant mapping, or more simply a morphism, is a map $f : X \to Y$ such that for all $g \in G$ and $x \in X$ we have:*

$$f(gx) = gf(x).$$

In this case we also say thet f *intertwines* the two actions. Of course if f is bijective we speak of an *isomorphism* of the 2 actions.

The class of $G-$sets and equivariant maps is clearly a **Category**.

Example. The equivariant maps of the action of G on itself by left multiplication are the right multiplications. They form a group isomorphic to the *opposite* of G (but also to G).

More generally:

Proposition. *The equivariant maps of the action of G on G/H by left multiplication are induced by the right multiplications with elements of the normalizer $N_G(H)$ of H (cf. 1.4). They form a group Γ isomorphic to $N_G(H)/H$.*

Proof. Let $\sigma : G/H \to G/H$ be such a map, hence for all $a, b \in G$ we have $\sigma(a.bH) = a\sigma(bH)$. In particular if $\sigma(H) = uH$ we must have that:

$$\sigma(aH) = auH = \sigma(ahH) = ahuH, \ \forall h \in H,$$

hence $u \in N_G(H)$ and $uH = Hu$. Conversely if $u \in N_G(H)$ the map $\sigma(u) : aH \to auH = aHu$ is well defined and in Γ. The map $u \to \sigma(u^{-1})$ is clearly a surjective homomorphism from $N_G(H)$ to Γ with kernel H.

Exercise Describe the set of equivariant maps $G/H \to G/K$ for 2 subgroups.

2 Orbits, invariants and equivariant maps.

2.1 The first important notion in this setting is given by the following:

Proposition. *The binary relation R in X given by: xRy if and only if there exists $g \in G$ with $gx = y$, is an equivalence relation.*

Definition. *The equivalence classes under the previous equivalence are called G-orbits (or simply orbits), the orbit of a given element x is formed by the elements gx with $g \in G$ and is denoted Gx. The mapping $G \to Gx$ given by $g \to gx$ is called the orbit map.*

The orbit map is equivariant (with respect to the left action of G). The set X is partitioned in its orbits, and the set of all orbits (quotient set) is denoted by X/G.

In particular we say that the action of G is *transitive* or that X is a *homogeneous space* if there is a unique orbit.

More generally we say that a subset Y of X is G *stable* if it is a union of orbits. In this case G induces naturally an action on Y. Of course the complement $\mathcal{C}(Y)$ of Y in X is also G stable and X is decomposed as $Y \cup \mathcal{C}(Y)$ in 2 stable subsets.

The finest decomposition into stable subsets is the decomposition into orbits.

Basic examples

i Let $\sigma \in S_n$ be a permutation and A the cyclic group which it generates, then the orbits of A on the set $[1, n]$ are the *cycles* of the permutation.

ii Let G be a group and H, K be subgroups, we have the action of $H \times K$ on G induced by the left and right action. The orbits are the *double cosets*. In particular if either H or K is 1 we have left or right cosets.

iii Consider G/H, the set of left cosets gH, with the action given by 1.2.2. Given a subgroup K on G/H the K orbits in G/H are in bijictive correspondence with the double coset KgH.

iv The action of G on itself by conjugation $(g, h) \to ghg^{-1}$. Its orbits are the *conjugacy classes*.

v An action of the additive group $\mathbb{R}_+$ of real numbers on a set X is called a *1-parameter group of transformations* or in a more physical language a *reversible dynamical system*.

In this case the parameter t is thought as *time* and an orbit is seen as the time evolution of a physical state. The hypotheses of the group action mean that the evolution is reversible (i.e. all the group transformations are invertible) and the *forces* do not vary with time so that the evolution of a state depends only on the time lapse (group homomorphism property).

The previous examples lead to single out the following general fact:

Remark. Let G be a group and K a normal subgroup in G, if we have an action of G on a set X we see that G acts also on the set of K orbits X/K, since $gKx = Kgx$, moreover we have $(X/K)/G = X/G$.

2.2 The study of group actions should start with the elementary analysis of a single orbit. The next main concept is that of *stabilizer*:

Definition. *Given a point $x \in X$ we set $G_x := \{g \in G | gx = x\}$. G_x is called the* **stabilizer** *(or* **little group***) of x.*

Proposition. *G_x is a subgroup and the action of G on the orbit Gx is isomorphic to the action on the coset space G/G_x.*

Proof. The fact that G_x is a subgroup is clear. Given two elements $h, k \in G$ we have that $hx = kx$ if and only if $k^{-1}hx = x$ or $k^{-1}h \in G_x$.

The mapping between G/G_x and Gx which assigns to a coset hG_x the element hx is thus well defined and bijective, it is also clearly $G-$equivariant and so the claim follows.

Example. Consider the action of $G \times G$ on G by left right translation. G is a single orbit and the stabilizer of 1 is the subgroup $\Delta := \{(g, g) | g \in G\}$ isomorphic to G embedded in $G \times G$ diagonally.

Example. In the case of a 1-parameter subgroup acting continuously on a topological space, the stabilizer is a closed subgroup of $\mathbb{R}$. If it is not the full group it is the set of integral multiples $ma, m \in \mathbb{Z}$ of a positive number a. The number a is to be considered as the first time in which the orbit returns to the starting point. This is the case of a **periodic orbit**.

Remark. Given two different elements in the same orbit their stabilizers are conjugate, in fact $G_{hx} = hG_xh^{-1}$. In particular when we identify an orbit to a coset space G/H this implicitely means that we have made the choice of a point for which the stabilizer is H.

Remark. The orbit cycle decomposition of a permutation can be interpreted in the previous language. To give a perputation on a set S is equivalent to give an action of the group of integers $\mathbb{Z}$ on S.

If S is finite this induces an action of a finite cyclic group isomorphic to $Z/(n)$.

To study a single orbit we only remark that a subgroup will be of the form $Z/(m)$ with m a divisor of n. The corresponding coset space is $Z/(m)$ and the generator $\overline{1}$ of $Z/(n)$ acts on $Z/(m)$ as the cycle $\overline{x} \to \overline{x} + \overline{1}$.

Consider the set of all subgroups of a group G, on this set G acts by conjugation. The orbits of this action are the *conjugacy classes of subgroups*, let us denote by $[H]$ the conjugacy class of a subgroup H.

The stabilizer of a subgroup H under this action is called its *normalizer.* It should not be confused with the *centralizer* which for a given subset A of G is the stabilizer under conjugation of all the elements of A.

Given a group G and an action on X it is useful to introduce the following notions.

For an orbit in X the conjugacy class of the stabilizers of its elements is well defined. We say that two orbits are of the same *orbit type* if the associated stabilizer class is the same. This is equivalent to say that the two orbits are isomorphic as $G-$spaces. It is often useful to partition the orbits according to the orbit types. The group of symmetries of the G action permutes transitively orbits of the same type.

Suppose that G and X are finite and assume that we have n_i orbits of type $[H_i]$ then we have, from the partition into orbits, the formula:

$$\frac{|X|}{|G|} = \sum_i \frac{n_i}{|H_i|}$$

we denote in general by $|A|$ the cardinality of a finite set A.

Exercise Let G be a group with $p^m n$ elements, p a prime number not dividing n. Deduce the theorems of Sylow by considering the action of G by left multiplication on the set of all subsets of G with p^m elements (Wielandt).

Exercise Given two subgroups H, K of G describe the orbits of H acting on G/K, in particular give a criterion for G/K to be a single H orbit.

Discuss the special case $[G : H] = 2$.

2.3 From all the elements of X we may single out the ones for which the stabilizer is the full group G.

These are the *fixed points* of the action or *invariant points*, i.e. the points whose orbit consists of the point alone. These points will be usually denoted by X^G.

$$X^G := \{\textit{fixed points} \text{ or } \textit{invariant points}\}.$$

We have thus introduced in a very general sense the notion of *Invariant* but its full meaning for the moment is completely obscure, we have first to procede with the formal theory.

2.4 One of the main features of set theory consists in the fact that it allows us to perform constructions, out of given sets we construct new ones.

This is also the case of $G-$sets. Let us point out at least 2 constructions:

(1) Given 2 $G-$sets X, Y we give the structure of a $G-$set to their disjoint sum $X \sqcup Y$ by acting separately on the two sets and to their product $X \times Y$ setting:

$$g(x, y) := (gx, gy), \tag{2.4.1}$$

(i.e. once the group acts on the elements it acts also on the pairs.)

(2) Consider now the set Y^X of all maps from X to Y, we can act with G (verify it) setting:

$$(gf)(x) := gf(g^{-1}x). \tag{2.4.2}$$

Notice that in the second definition we have used again twice the action of G, the particular formula given is justified by the fact thet it is really the only way to get a group action using the two actions.

It reflects a general fact well known in category theory, that maps between two objects X, Y are a covariant functor in Y and controvariant in X.

We want to explicit immediately a rather important consequence of our formalism:

Proposition. *A map $f : X \to Y$ between two $G-$sets is equivariant (cf. 1.2) if and only if it is a fixed point under the $G-$action on the maps.*

Proof. This statement is really a tautology, nevertheless deserves to be clearly understood. The proof is trivial following the definitions. Equivariance means that $f(gx) = gf(x)$. This, if we substitute x with $g^{-1}x$, reads $f(x) = gf(g^{-1}x)$ which in the functional language means that the function f equals the function gf, i.e. it is invariant.

Exercise

i) Show that the orbits of G on $G/H \times G/K$ are in canonical 1-1 correspondence with the double cosets HgK of G.

ii) Given a G equivariant map $\pi : X \to G/H$ show that:

a) $\pi^{-1}(H)$ is stable under the action of H.

b) The set of G orbits on X is in 1-1 correspondence with the H orbits on $\pi^{-1}(H)$.

c) Study the case in which $X = G/K$ is also homogeneous.

2.5 We will often consider a special case of the previous section, the case of the trivial action of G on Y. In this case of course the action of G on the functions is simply:

$$(gf)(x) = f(g^{-1}x) \tag{2.5.1}$$

A mapping is equivariant if and only if it is constant on the orbits. In this case we will always speak of *Invariant function.* In view of the particular role of this idea in our treatment we repeat the formal definition.

Definition. *A function f on a G set is called an invariant if $f(g^{-1}x) = f(x)$ for all $x \in X$ and $g \in G$.*

As we have just remarked a function is invariant if and only if it is constant on the orbits. Formally we may thus say that the quotient mapping $\pi := X \to X/G$ is an invariant map and any other invariant function factors as :

$$\begin{tikzcd} X \arrow[rr, "f"'] \arrow[dr, "\pi"'] & & Y \\ & X/G \arrow[ur, "\overline{f}"'] & \end{tikzcd}$$

We want to explicit the previous remark in a case of importance.

Let X be a finite G set. Consider a field F (a ring would suffice) and the set F^X of functions on X with values in F.

An element $x \in X$ can be identified to the characteristic function of $\{x\}$, in this way X becomes a basis of F^X as vector space.

The induced group action of G on F^X is by linear transformation and by permuting the basis elements.

Since a function is invariant if and only if it is constant on orbits we deduce:

Proposition. *The invariants of G on F^X form the subspace of F^X having as basis the characteristic functions of the orbits.*

In other words given an orbit $\mathcal{O}$ consider $u_{\mathcal{O}} := \sum_{x \in \mathcal{O}} x$. The elements $u_{\mathcal{O}}$ form a basis of $(F^X)^G$.

We finish this section with two examples which will be useful in the theory of symmetric functions.

Consider the set $[1, n] := \{1, 2, \ldots, n\}$ with its canonical action of the symmetric group.

The maps from $\{1, 2, \ldots, n\}$ to the field $\mathbb{R}$ of real numbers form the standard vector space $\mathbb{R}^n$. The symmentric group then acts by permuting the coordinates and in every orbit there is a unique vector $(a_1, a_2, \ldots, a_n)$ with $a_1 \geq a_2 \geq \ldots \geq a_n$.

The set of these vectors can thus be identified to the orbit space. It is a convex cone with boundary the elements in which at least two coordinates are equal.

Exercise Discuss the orbit types of the previous example.

Definition. *A function $M : \{1, 2, \ldots, n\} \to \mathbb{N}$ (to the natural numbers) is called a* **monomial**. *The set of monomials is a semigroup by addition of values and we indicate by x_i the monomial which is the characteristic function of $\{i\}$.*

Remark. It is customary to write the semigroup law of monomials multiplicatively. Given a monomial M if $M(i) = h_i$ we have that $M = x_1^{h_1} x_2^{h_2} \ldots x_n^{h_n}$. The number $\sum_i h_i$ is the **degree** of the monomial.

Representing a monomial as a vector $(h_1, h_2, \ldots, h_n)$ we see that every monomial is equivalent, under the symmetric group, to a unique vector in which the coordinates are non increasing. The non zero coordinates of such a vector form thus a partition, with at most n parts, of the degree of the monomial.

Exercise To a monomial M we can also associate a partition of the set $\{1, 2, \ldots, n\}$ by the equivalence $i \cong j$ iff $M(i) = M(j)$, show that the stabilizer of M is the group of permutations which preserve the sets of the partition (cf. 1.1) and determine a basis of invariant symmetric polynomials.

2.6 It is time to develop some other examples. First of all consider the set $[1, n]$ and a ring A (in most applications the integers or the real or complex numbers).

A function f from $[1, n]$ to A may be thought as a vector and displayed for instance as a row with the notation $(a_1, a_2, \ldots, a_n)$ where $a_i := f(i)$. The set of all functions is thus denoted by A^n. The symmetric group acts on such functions according to the general formula 2.5.1.

$$\sigma(a_1, a_2, \ldots, a_n) = (a_{\sigma^{-1}1}, a_{\sigma^{-1}2}, \ldots, a_{\sigma^{-1}n}).$$

In this simple example we already see that the group action is linear. We will refer to this action as the **standard permutation action**.

Remark that if $\underline{e}_i$ denote the canonical basis vector with coordinates 0 except 1 in the i^{th} position, we have $\sigma(\underline{e}_i) = \underline{e}_{\sigma(i)}$. This formula allows us to describe the matrix of σ in the given basis, it is the matrix $\delta_{\sigma^{-1}(j),i}$. These matrices are called *permutation matrices.*

This is a general fact, if we consider a $G-$set X and a ring A, the set of functions on X with values in A form also a ring under pointwise sum and multiplication and we have that:

Remark. The group G acts on the functions with values in A as a group of ring automorphisms.

In this particular example it is important to procede further. Once we have the action of S_n on A^n we may continue and act on the functions on A^n! In fact let us consider the *coordinate functions*: $x_i : (a_1, a_2, \ldots, a_n) \to a_i$, it is clear from the general formulas that the symmetric group permutes the cordinate functions and $\sigma(x_i) = x_{\sigma(i)}$. The reader may note the fact that the inverse has now disappeared.

If we have a ring R and an action of a group G on R as ring automorphisms it is clear that:

Proposition. *The invariant elements form a subring of R.*

Thus we can speak of *the ring of invariants R^G.*

2.7 We need another generality. Suppose that we have two group actions on the same set X i.e. assume that we have two groups G and H acting on the same set X.

We say that the two actions commute if $gh(x) = hg(x)$ for all $x \in X$, $g \in G$ and $h \in H$.

This means that every element of G gives rise to an H equivariant map (or we can reverse the roles of G and H) it also means that we really have an action of the product group $G \times H$ on X given by $(g, h)x = ghx$.

In this case we easily see that if a function f is G invariant and $h \in H$ then hf is also G invariant. Hence H acts on the set of G invariant functions.

More generally suppose that we are given a G action on X and a normal subgroup K of G, then it easily seen that the quotient group G/K acts on the set of K invariant functions and a function is G invariant if and only if it is K and G/K invariant.

Example. The right and left action of G on itself commute (Example 1.2 c).

3 Linear actions, groups of automorphisms, commuting groups.

3.1 We have seen how, given an action of a group G on a set X and a field F, we can deduce an action over the set F^X of functions from X to F, which is linear, i.e. given by linear operators.

In general the groups G and the sets X on which they act may have further structures, as in the case of e a topological or differentiable or algebraic action. In these cases it will be important to restrict the set of functions to the ones compatible with the structure under consideration, we will do it systematically.

If X is finite the vector space of functions on X with values in F has, as a possible basis, the characteristic functions of the elements. It is convenient to identify an element x with its characteristic function and thus say that our vector space has X as a basis (cf. 2.5).

A function f is thus written as $\sum_{x\in X} f(x)x$, the linear action of G on F^X induces on this basis the action from which we started, we call such an action a *permutation representation.*

In the algebraic theory we may in any case consider the set of all functions which are finite sums of the characteristic functions of points, i.e. the functions which are 0 outside a finite set.

These are usually called **functions with finite support**, we will often denote these functions by the symbol $F[X]$, which is supposed to remind us that its elements are linear combinations of elements of X.

In particular for the left action of G on itself we have the **algebraic regular representation** of G on $F[G]$. We shall see that this representation is particularly important.

Let us stress a feature of this representation.

We have two actions of G on G the left and the right action which commute with each other, or in other words we have an action of $G \times G$ on G, given by $(h,k)g = hgk^{-1}$ (for which $G = G \times G/\Delta$ where $\Delta = G$ embedded diagonally cf. 1.2c and 2.2).

Thus we have the corresponding two actions on $F[G]$ by $(h,k)f(g) = f(h^{-1}gk)$ and we may view the right action as symmetries of the left action and conversely.

Sometimes it is convenient to denote by ${}^hf^k = (h,k)f$ to stress the left and right actions.

After these basic examples we give a general definition:

Definition. *Given a vector space V over a field F (or more generally a module) we say that an action of a group G on V is linear if every element of G induces a linear transformation on V, a linear action of a group is also called a representation.*

In a different language let us consider the set of all linear invertible transformations of V, this is a group under composition (i.e. it is a subgroup of the group of all invertible transformations) and will be called the:

General linear group of V, indicated with the symbol $GL(V)$.

In case we take $V = F^n$ (or equivalently in case V is finite dimensional and we identify V with F^n by choosing a basis) we can identify $GL(V)$ with the group of $n \times n$ invertible matrices, denoted $GL(n,F)$.

According to our general principles a linear action is thus a homomorphism ϱ of G in $GL(V)$ (or in $GL(n,F)$).

When we are dealing with linear representations we usually consider also equivariant linear maps between them, thus obtaining a category.

Exercise Two linear representations $\rho_1, \rho_2 : G \to GL(n, F)$ are (linearly) isomorphic if and only if there is an invertible matrix X such that $X\rho_1(g)X^{-1} = \rho_2(g)$ for all $g \in G$.

Before we procede any further we should remark an important feature of the theory.

Given 2 linear representations U, V we can form their direct sum $U \oplus V$ which is a representation by setting $g(u, v) = (gu, gv)$. If $X = A \cup B$ is a G set, disjoint union of 2 G stable subsets, we clearly have $F^{A\cup B} = F^A \oplus F^B$ thus the decomposition in direct sum is a generalization of the decomposition of a space in G stable sets.

If X is an orbit it cannot be further decomposed as set while F^X might be decomposable. The simplest example is $G = \{1, \tau = (12)\}$ the group with 2 elements of permutations of $[1, 2]$, the space F^X decomposes, setting:

$$u_1 := \frac{e_1 + e_2}{2}, \; u_2 := \frac{e_1 - e_2}{2}$$

we have $\tau e_1 = e_1, \; \tau(e_2) = -e_2$.

We have implicitely used the following ideas:

Definition. *i) Given a linear representation V a subspace U of V is a* **subrepresentation** *if it is stable under G.*

ii) V is a **decomposable representation** *if we can find a decomposition $V = U_1 \oplus U_2$ with the U_i proper subrepresentations, otherwise it is called* **indecomposable**.

iii) V is an **irreducible** *representation if the only subrepresentations of V are V and 0.*

We will study in detail some of the deep connections between these notions.

First 2 basic examples:

Example. Let A, B be the algebra of all, resp. of upper triangular (i.e. 0 below the diagonal) $n \times n$ matrices over a field F.

Exercise The vector space F^n is irreducible as an A module, indecomposable but not irreducible as a B module.

Given 2 linear representations U, V of a group G, the space of G equivariant linear maps is denoted $Hom_G(U, V)$ and called:

Definition. *Space of intertwining operators.*

In this book we will almost always treat finite dimensional representations, thus unless specified otherwise our vector spaces will always be assumed to be finite dimensional.

It is quite useful to rephrase the theory of linear representations in a different way:

Consider the space $F[G]$:

Proposition. *i) The group multiplication extends to a bilinear product on $F[G]$ for which $F[G]$ is an associative algebra with 1, called* **the group algebra**.

ii) Linear representations of G are the same as $F[G]$ modules.

Proof. The first part is immediate. As for the second given a linear representation of G we have the module action $(\sum_{g\in G} a_g g)v := \sum_{g\in G} a_g(gv)$. The converse is clear.

Remark. 1) Consider the left and right action on the functions $F[G]$.

Let $h, k, g \in G$ and identify g with the characteristic function of the set $\{g\}$ then ${}^h g^k = hgk^{-1}$ (as functions).

The space $F[G]$ as $G \times G$ module is the permutation representation associated to $G = G \times G/\Delta$ with its $G \times G$ action (3.1).

Thus a space of functions on G is stable under left (res. right) action if and only if it is a left (resp. right) ideal of the group algebra $F[G]$.

2) Notice that the direct sum of representations is the same as the direct sum as modules, also a G linear map between two representations is the same as a module homomorphism.

Example. Let us consider a finite group G, a subgroup K and the linear space $F[G/K]$, which as we have seen is a permutation representation.

We can identify the functions on G/K as the functions on G which are invariant under the right action of K, in this way the element $gK \in G/K$ is identified to the characteristic function of the coset gK and $F[G/K]$ is identified to a subrepresentation of the group algebra $F[G]$.

$$F[G/K] = \{a \in F[G] | ah = a, \ \forall h \in K\}. \tag{3.1.1}$$

If we denote by u the characteristic function of the subgroup H we see that u generates this module over $F[G]$.

Given 2 subgroups H, K and the linear spaces $F[G/H]$, $F[G/K] \subset F[G]$ we want to determine their intertwiners.

For an intertwiner f, and $u := \chi_H$ as before, let $f(u) = a \in F[G/K]$. We have $hu = u$, $\forall h \in H$ and so, since f is an intertwiner $a = f(u) = f(hu) = ha$. Thus we must have that a is also left invariant under H. Conversely given such an a the map $b \to \frac{ba}{|H|}$ is an intertwiner mapping u to a. Since u generates $F[G/H]$ as a module we see that:

Proposition. *The space $Hom_G(F[G/H], F[G/K])$ of intertwiners can be identified with the H invariants of $F[G/K]$, or to the $H - K$ invariants ${}^H F[G]^K$ of $F[G]$. It has as basis the characteristic funtions of the double cosets HgK.*

In particular for $H = K$ we have that the functions which are biinvariants under H form under convolution the endomorphism algebra of $F[G/H]$.

These functions have as basis the characteristic funtions of the double cosets HgH, one usually indicates by $T_g = T_{HgH}$ the corresponding operator. In this way we have the *Hecke* *algebra* and *Hecke operators*, the multiplication rule between such operators depends on the multiplication on cosets $HgHHkH = \cup Hh_iH$ and each double coset appearing in this product appears with a positive integer multiplicity so that $T_gT_h = \sum n_iT_{h_i}$.[1]

Similar results when we have 3 subgroups H, K, L and compose:

$$Hom_G(F[G/H],\, F[G/K])Hom_G(F[G/K],\, F[G/L]) \to Hom_G(F[G/H],\, F[G/L])$$

The notion of permutation representation is a special case of that of **induced representation**, if M is a representation of a subgroup H of a group G we consider the space

$$Ind_H^G M := \{f : G \longrightarrow M | f(gh) = h^{-1}f(g),\ \forall h \in H,\, g \in G\}.$$

On this space of functions define a G action by $(gf)(x) := f(gx)$. It is easy to see that this is a well defined action. Moreover we can identify $m \in M$ with the function f such that $f(x) = 0$ if $x \notin H$ and $f(h) = h^{-1}m$ if $h \in H$.

Exercise Verify that, choosing a set of representatives of the cosets G/H we have as vector space the decomposition

$$Ind_H^G M := \oplus_{g\in G/H} gM.$$

3.2 Suppose we are now given a linear function $f \in V^*$ on V, by definition the function gf is given by $(gf)(v) = f(g^{-1}v)$ and hence it is again a linear function.

Thus G acts dually on the space V^* of linear functions and it is clear that this is a linear action which is called the *contragredient* action.

In matrix notations, if we use as basis of the dual space the dual of a given basis, the contragredient action of an operator T is given by the inverse transpose fo the matrix of T.

We will use the notation $< \varphi|v >$ for the value of a linear form on a vector and thus have the identity:

(3.2.1) $$< g\varphi|v > = < \varphi|g^{-1}v > .$$

Alternatively it may be convenient to define on V^* a *right action* by the more symmetric formula:

(3.2.2) $$< \varphi g|v > = < \varphi|gv > .$$

Exercise. Prove that the dual of a permutation representation is isomorphic to the same permutation representation. In particular one can apply this to the dual of the group algebra.

[1] It is important in fact to use these concepts in a much more general way as done by Hecke in the theory of modular forms. Hecke studies the action of $Sl(2, Z)$ on $M_2(\mathcal{Q})$ the 2×2 rational matrices. In this case one has also double cosets, a product structure on $M_2(\mathcal{Q})$ and the fact that a double coset is a finite union of right or left cosets. These properties suffice to develop the Hecke algebra. In this case this algebra acts on a different space of functions, the modular forms (cf. Ogg).

In the set of all functions on a finite dimensional vector space V a special role play the polynomial functions. By definition a polynomial fuction is an element of the subalgebra (of the algebra of all functions with values in F) generated by the linear functions.

If we choose a basis and consider the coordinate functions $x_1, x_2, \dots, x_n$ with respect to the chosen basis, a polynomial function is a usual polynomial in the x_i. If F is infinite thie expression as a polynomial is unique and we can consider the x_i as given variables.

The ring of polynomial functions on V will be denoted by $P[V]$ the ring of fromal polymomials by $F[x_1, x_2, \dots, x_n]$.

Choosing a basis we have always a surjective homomorphism $F[x_1, x_2, \dots, x_n] \to P[V]$ which is an isomorphism if F is infinite.

Exercise If F is a finite field with q elements prove that $P[V]$ has q^n elements and that the kernel of the map $F[x_1, x_2, \dots, x_n] \to P[V]$ is the ideal generated by the elements $x_i^q - x_i$.

Since the linear funtions are preserved under a given group action we have:

Proposition. *Given a linear action of a group G on a vector space V, G acts on the polynomial functions $P[V]$ by the rule $(gf)(v) = f(g^{-1}v)$ as a group of ring automorphisms.*

Of course the full linear group acts on the polynomial functions. In the language of coordinates we may view the action as linear changes of coordinates.

Exercise Show that we always have a linear action of $GL(n, F)$ on the formal polynomial ring $F[x_1, x_2, \dots, x_n]$.

3.3 We assume the base field infinite for simplicity although the reader can see easily what happens for finite fields. One trivial but important remark is that the group action on $P[V]$ preserves the degree.

Recall that a function f is homogeneous of degree k if $f(\alpha v) = \alpha^k f(v)$ for all $\alpha's$ and $v's$.

The set $P[V]_q$ of homogeneous polynomials of degree q is a subspace, called in classical language the space of *quantics*. If $dim(V) = n$ one speaks of $n-ary$ quantics.

In general a direct sum of vector spaces $U = \oplus_{k=0}^{\infty} U_k$ is called a *graded vector space*. A subspace W of U is called *homogeneous*, if, setting $W_i := W \cap U_i$, we have $W = \oplus_{k=0}^{\infty} W_k$.

The space of polynomials is thus a graded vector space $P[V] = \oplus_{k=0}^{\infty} P[V]_k$. One has immediately $(gf)(\alpha v) = f(\alpha g^{-1} v) = \alpha^k (gf)(v)$ which has an important consequence:

Theorem. *If a polynomial f is an invariant (under some linear group action) then also its homogeneous components are invariant.*

Proof. Let $f = \sum f_i$ be the decomposition of f in homogeneous components, $gf = \sum gf_i$ is the decomposition in homogeneous components of gf. If f is invariant $f = gf$ and then $f_i = gf_i$ for each i since the decomposition into homogeneous components is unique.

In order to summarize the analysis done up to now let us also recall that an algebra A is called a *graded algebra* if it is a graded vector space, $A = \oplus_{k=0}^{\infty} A_k$ and, for all h, k we have $A_h A_k \subset A_{h+k}$.

Proposition. *The spaces $P[V]_k$ are subrepresentations. The set $P[V]^G$ of invariant polynomials is a graded subalgebra.*

3.4 To some extent the previous theorem may be viewed as a special case of the more general setting of commuting actions.

Let thus be given two representations $\varrho_i : G \to GL(V_i)$, $i = (1, 2)$, consider the linear transformations between V_1 and V_2 which are G equivariant, it is clear that they form a linear subspace of the space of all linear maps between V_1 and V_2.

The space of all linear maps will be denoted by $hom(V_1, V_2)$ while the space of equivariant maps will be denoted $hom_G(V_1, V_2)$. In particular when the two spaces coincide we write $End(V)$ or $End_G(V)$ instead of $hom(V, V)$ or $hom_G(V, V)$.

These spaces are in fact now algebras, under composition of operators. Choosing bases we have that $End_G(V)$ is the set of all matrices which commute with all the matrices coming from the group G.

Consider now the set of invertible elements of $End_G(V)$, i.e. the group H of all linear operators which commute with G.

By the remarks of 3.3, H preserves the degrees of the polynomials and maps the algebra of G invariant functions in itself thus:

Remark. H induces a group of automorphisms of the graded algebra $P[V]^G$.

We view this remark as a generalization of Proposition 3.3 since the group of scalar multiplications commutes (by definition of linear transformation) with all linear operators. Moreover it is easy to prove:

Exercise Given a graded vector space $U = \oplus_{k=0}^{\infty} U_k$ define an action ϱ of the multiplicative group F^* of F setting $\varrho(\alpha)(v) := \alpha^k v$ if $v \in U_k$. Prove that a subspace is stable under this action if and only if it is a graded subspace (F is assumed to be infinite).

4 Symmetric functions

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

The theory of symmetric functions is a classical theory developed (by Lagrange, 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 are 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 implicitely defined by the formula:

$$p(t) := \prod_{i=1}^{n}(1 + tx_i) := \sum_{i=0}^{n} e_i t^i. \tag{4.1.1}$$

More explicitely $e_i(x_1, x_2, \dots, x_n)$ 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 indeces in that subset.

$$e_i = \sum_{1 \le a_1 < a_2 < \dots < a_i \le n} x_{a_1} x_{a_2} \dots x_{a_i}. \tag{4.1.2}$$

If σ is a permutation of the indeces we obviously have

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

and thus the elements e_i are invariant under permutation of the variables.

Of course the polynomial $t^n p(-\frac{1}{t})$ 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. The power sums $\psi_k := \sum_{i=1}^{n} x_i^k$ and the funtions 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 which is permuted by the symmetric group.

By Proposition 2.4 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 $(h_1 \ge h_2 \ge \dots \ge h_n)$, $h_i \in \mathbb{N}$ and we may set $\Sigma_{(h_1, h_2, \dots, h_n)}$ to be the sum of monomials in the corresponding orbit.

As we will see soon there are also some subtler symmetric functions (the Schur functions) that 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 the ψ_k.

To do this we will perform the next computations in the ring of formal power series, although the series that we will consider have also a meanining 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 logaritmic derivative (relative to the variable t) of both sides. We use the fact that such an operator transforms products into sums to get

$$\sum_{i=1}^n \frac{x_i}{(tx_i + 1)} = \frac{\sum_{i=1}^n ie_i t^{i-1}}{\sum_{i=0}^n e_i t^i}.$$

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

$$(\sum_{h=0}^\infty (-t)^h \psi_{h+1})(\sum_{i=0}^n e_i t^i) = (\sum_{i=1}^n ie_i t^{i-1})$$

which gives, equating coefficients:

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

where we intend $e_i = 0$ 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 divisions and the coefficients are rational and usually not integers.

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 in 0.

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

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

Now the second proof. 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 sum over all i we get:

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

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

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

To make the proof transparent let us stress in our formulas also 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:

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

we have:

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

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 in 0 we have that symmetric functions map to symmetric functions and

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

Given a symmetric polynomial $f(x_1, \dots, x_n)$ we evaluate it at $x_n = 0$, if the resulting polynomial $\overline{f}(x_1, \dots, x_{n-1})$ 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 pass to another symmetric function of lower degree.

Otherwise by induction there exists 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. *A symmetric polynomial is a polynomial, in a unique way, in the elementary symmetric functions.*

4.2 In the same way the reader may discuss the following fact.

Consider the $n!$ monomials

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

Theorem. *The previous 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.

5 Resultant, discriminant, Bezoutiante

5.1 In order to understand the importance of theorem 4.1 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 $P_n := \{t^n + b_1 t^{n-1} + \ldots + b_n\}$ of monic polynomials (which can be identified to $\mathbb{C}^n$ by the use of the coefficients).

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

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

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 $\alpha_1, \ldots, \alpha_n$ (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 parametrizes 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 expliciting the roots, since one has only to study the formal symmetric polynomial expression and, using the previous or another algorithm express the value of a symmetric function of the roots through the coefficients.

In other words a polynomial function f on $\mathbb{C}^n$ which is symmetric, factors through the map π giving rise to an effectively computable[2] 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:

$$A := \begin{pmatrix} x_1^{n-1} & x_2^{n-1} & \ldots & x_n^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^2 & x_2^2 & \ldots & x_n^2 \\ x_1 & x_2 & \ldots & x_n \\ 1 & 1 & \ldots & 1 \end{pmatrix} \tag{5.1.1}$$

Proposition. *$V(x)$ is antisymmetric, i.e. permuting the variables it is multiplied by the sign of the permutation.*

Remark. The theory of the sign of permutations can be deduced by analyzing the Vandermonde. In fact since for a transposition τ it is clear that $V(x)^\tau = -V(x)$ it follows

[2] i.e. computable without solving the equation

that $V(x)^\sigma = V(x),$ or $- V(x)$ according to whether σ is a product of an even or an odd number of permutations. 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 ij position of B we find the symmetric function $\psi_{2n-(i+j)}$ and its determinant is V^2.

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

We write V^2 as a polynomial $D(e_1, e_2, \dots, e_n)$ in the elementary symmetric functions.

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, $f(t)$ a monic polynomial (of degree n) with coefficients in F and let $R := F[t]/(f(t))$. 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 (and also a right one). We define $tr(a) := tr(L_a)$, the trace of the operator L_a.

We consider next the bilinear form $(a, b) := tr(ab)$ 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 $tr(t^k) = 0$ if $k > 0$ and so the trace form has rank 1 with kernel the ideal generated by t.

To compute it 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 $1, (t - \alpha_i), (t - \alpha_i)^2, \dots, (t - \alpha_i)^{h_i - 1}$ (mod $(t - \alpha_i)^{h_i}$) the matrix of t is upper triangular with constant eigenvalue α_i on the diagonal and so the claim follows.

As a consequence we see that the matrix of the trace form, in the basis $1, t, \dots, t^{n-1}$ is the Bezoutiante 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. *The rank of the Bezoutiante 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. In characteristic zero this polynomial is obtained dividing $f(t)$ by the G.C.D. between $f(t)$ and its derivative.

Let us consider now 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 $\alpha_1, \alpha_2, \ldots, \alpha_k$ and the complex ones $\beta_1, \overline{\beta}_1, \beta_2, \overline{\beta}_2, \ldots, \beta_h, \overline{\beta}_h$.

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. On $\mathbb{R}$ the trace form is just x^2 but on $\mathbb{C}$ we have $tr((x+iy)^2) = 2(x^2 - y^2)$. We deduce that:

Theorem. *The number of real roots of $f(t)$ equals the signature of its Bezoutiante.*

There are simple variations on this theme, for instance if we consider the quadratic form $Q(x) := tr(tx^2)$ we see that its matrix is again easily computed in terms of the ψ_k and its signature equals the number of real positive 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 which now bears his name.

5.2 Let us go back to the roots, if $x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_m$ 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 which has this property is a multiple of A.

By the general theory A can be expressed as a polynomial R in elementary symmetric funtions.

Let us denote by $a_1, a_2, \ldots, a_n$ the elementary symmetric functions in the x_i's and $b_1, \ldots, b_m$ the ones in the y_j's. Thus $A(x,y) = R(a_1, \ldots, a_n, b_1, \ldots, b_m)$.

The polynomial R is called the *resultant.*

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

There is a classical expression as determinant, the theory is the following.

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

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

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

$$\begin{pmatrix}
a_n & 0 & 0 & \dots & 0 & b_m & 0 & \dots & 0 & 0 \\
a_{n-1} & a_n & 0 & \dots & 0 & b_{m-1} & b_m & \dots & \ddots & 0 \\
a_{n-2} & a_{n-1} & a_n & 0 & \dots & 0 & b_{m-1} & b_m \ddots & & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
a_1 & a_2 & a_3 & & & & & & \ddots & \vdots \\
a_0 & a_1 & a_2 & & & & & & & \\
0 & a_0 & a_1 & & & & & & \ddots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & \vdots & & & & b_0 & b_1 & b_2 & \ddots & \vdots \\
0 & 0 & \dots & & & 0 & b_0 & b_1 & \ddots & \vdots \\
0 & 0 & & & & 0 & 0 & b_0 & \ddots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & b_0 & \vdots \\
0 & 0 & 0 & \dots & a_0 & 0 & \dots & \dots & 0 & b_0
\end{pmatrix} \tag{5.2.1}$$

Proposition. *If $a_0b_0 \neq 0$, the rank of $T_{f,g}$ ($n \leq m$) equals $m+n-d$ where d is the degree of $h := G.C.D(f,g)$.*

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 of pairs $(sg', -sf')$ where $f = hf', g = hg'$, hence the claim.

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.

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 and we see that developing it along the first row we get the desired identity.

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 substitute in R to the variables a_i the element $(-1)^i a_0 e_i(x_1, \dots, x_n)$ and to b_i the element $(-1)^i b_0 e_i(y_1, \dots, y_m)$. The polynomial we obtain is $a_0^m b_0^n A(x, y)$.

5.3 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 $R(f, f')$ to the variables a_i the element $(-1)^i a_0 e_i(x_1, \dots, x_n)$ we have a polynomial in the x with coefficients involving a_0 which vanishes whenever two x's coincides.

Thus it is divisible by the discriminant 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 since the given substitution in the matrix $R_{f,f'}$ gives that every row of the resulting matrix is a multiple of a_0 thus a_0 will appear with exponent $2n-1$ and $c = c_0 a_0^{2n-1}$ with c_0 an integer, by specializing to the polynomial $x^n - 1$ we see that $c_0 = \pm n^n$.

6 Schur functions

6.1 It is important to discuss along symmetric, also alternating functions, we assume to work on integral polynomials.

Definition. *A polynomial f in the variables $(x_1, x_2, \dots, x_n)$, is called an alternating function, if given a 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 as a basic alternating polynomial, $V(x) := \prod_{i<j}(x_i - x_j)$.

The main remark on alternating functions is the following.

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

Proof. Substitute, in an alternating polynomial f to 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 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 $A = V(x)S$ or A is a free rank 1 module over the ring S generated by $V(x)$.

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 $Z[x_1, x_2, \dots, x_n]$ 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 integers. To $m_1 \geq m_2 \geq m_3 \cdots \geq m_n \geq 0$ corresponds the orbit of the monomial $x_1^{m_1} x_2^{m_2} x_3^{m_3} \dots x_n^{m_n}$.

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

It follows that all the monomials which have non 0 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 $x_1^{m_1} x_2^{m_2} x_3^{m_3} \dots x_n^{m_n}$ and of $x_{\sigma(1)}^{m_1} x_{\sigma(2)}^{m_2} x_{\sigma(3)}^{m_3} \dots x_{\sigma(n)}^{m_n}$ differ by the sign of σ.

It follows that:

Theorem. *The functions:*

$$\sum_{\sigma \in S_n} \epsilon_\sigma x_{\sigma(1)}^{m_1} x_{\sigma(2)}^{m_2} \dots x_{\sigma(n)}^{m_n}, \ m_1 > m_2 > m_3 \cdots > m_n \geq 0$$

are an integral basis of the space of antisymmetric functions.

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

6.2 It is convenient to use the following conventions. Consider the sequence

$$\varrho := (n-1, n-2, \dots, 2, 1, 0),$$

Lemma. *The map* $\lambda = (m_1, m_2, m_3, \dots, m_n) \longrightarrow \lambda + \varrho == (m_1 + n - 1, m_2 + n - 2, m_3 + n - 3, \dots, m_n)$ *is a a bijiective correspondence between decreasing and strictly decreasing sequences.*

We thus indicate by A_λ the corresponding antisymmetric function. We can express it also as a determinant of the matrix M_λ having in the i, j position the element $x_j^{m_i+n-i}$ and remark that $A_0 = V(x)$.[3]

We next set $S_\lambda(x) := A_\lambda / V(x)$ the *Schur function* associated to λ, when there is no ambiguity we will drop the variables symbol and speak of S_λ.

Sometimes we refer to λ as a *partition* of the integer $\sum m_i$ and write $\lambda \vdash \sum_i m_i$.

We call **heigth** of λ and denote it by $ht(\lambda)$ the number of non 0 elements in the sequence m_i.

We also graphically represent it by a *Young diagram.* The numbers m_i appear then as the lengths of the rows (cf. Chapter 3).

We can also consider the columns of the diagram which will be thought as rows of the **dual partition**.

We have thus that:

Theorem. *The functions* S_λ *are an integral basis of the ring of symmetric functions. If* $\lambda \vdash m$ *the degree of* S_λ *is* m.

Notice that the Vandermonde determinant is the alternating function A_0 and $S_0 = 1$.

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 3.

If $\lambda = (m_1, m_2, m_3, \dots, m_n)$ is a partition and a a positive integer let us denote by $\underline{a}$ the partition $(a, a, a, \dots, a)$ then from 6.1.1 follows that

$$A_{\lambda+\underline{a}} = (x_1 x_2 \dots x_n)^a A_\lambda, \ \ S_{\lambda+\underline{a}} = (x_1 x_2 \dots x_n)^a S_\lambda. \tag{6.2.1}$$

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

Let $\lambda := h_1 \geq h_2 \geq \dots \geq h_{m+1} \geq 0$, we have seen that, if $h_{m+1} > 0$ then $S_\lambda(x_1, \dots, x_{m+1}) = \prod_{i=1}^{m+1} x_i S_{\overline{\lambda}}(x_1, \dots, x_{m+1})$ where $\overline{\lambda} := h_1 - 1 \geq h_2 - 1 \geq \dots \geq h_{m+1} - 1$.

In this case, clearly $S_\lambda(x_1, \dots, x_m, 0) = 0$.

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

Assume now $h_{m+1} = 0$ and denote by the same symbol λ the sequence $h_1 \geq h_2 \geq \cdots \geq h_m$. Let us start from the Vandermonde determinant $V(x_1, \ldots, x_m, x_{m+1}) = \prod_{i<j\leq m+1}(x_i - x_j)$ and set $x_{m+1} = 0$ getting

$$V(x_1, \ldots, x_m, 0) = \prod_{i=1}^{m} x_i \prod_{i<j\leq m}(x_i - x_j) = \prod_{i=1}^{m} x_i V(x_1, \ldots, x_m).$$

Now consider the alternating function $A_\lambda(x_1, \ldots, x_m, x_{m+1})$.

Set $\ell_i := h_i + m + 1 - i$ so that $\ell_{m+1} = 0$ and

$$A_\lambda(x_1, \ldots, x_m, x_{m+1}) = \sum_{\sigma \in S_{m+1}} \epsilon_\sigma x_1^{\ell_{\sigma(1)}} \ldots x_{m+1}^{\ell_{\sigma(m+1)}},$$

setting $x_{m+1} = 0$ we get the sum restricted only on the terms for which $\sigma(m+1) = m+1$ or

$$A_\lambda(x_1, \ldots, x_m, 0) = \sum_{\sigma \in S_m} \epsilon_\sigma x_1^{\ell_{\sigma(1)}} \ldots x_m^{\ell_{\sigma(m)}}$$

now in $m-$ variables the partition λ corresponds to the decreasing sequence $h_i + m + 1 - i$ hence

$$A_\lambda(x_1, \ldots, x_m, 0) = \prod_{i=1}^{m} A_\lambda(x_1, \ldots, x_m), \quad S_\lambda(x_1, \ldots, x_m, 0) = S_\lambda(x_1, \ldots, x_m).$$

Thus we see that, under the evaluation of x_{m+1} to 0 the Schur functions S_λ vanish, if $heigth(\lambda) = m+1$ otherwise they map to the corresponding Schur functions in $m-$variables.

One uses these remarks as follows. Consider a fixed degree n, 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 $S_\lambda(x_1, \ldots, x_m)$ where $\lambda \vdash n$ has heigth $\leq m$. Under the evaluation $x_m \to 0$ we have a map $S_m^n \to S_{m-1}^n$. We have proved that this map is an isomorphism as soon as $m > n$ hence all identities which we prove for symmetric functions in n variables of degree n are valid in any number of variables.[4]

We want to prove now that for the elementary symmetric functions we have

$$e_h = S_{1^h}. \tag{6.2.2}$$

According to our previous discussion we can set all the variables $x_i,\ i > h$ to 0. Then e_h reduces to $\prod_{i=1}^{h} x_i$ as well as S_{1^h} from 6.2.1.

[4] One way of formalizing this is to pass formally to a ring of symmetric functions in infinitely many variables which has as basis all Schur functions without restriction to the heigth and is a polynomial ring in infinitely many variables corresponding to all possible *elementary symmetric functions*.

6.3 Next we want to discuss the value of $S_\lambda(1/x_1, 1/x_2, \ldots, 1/x_n)$. We see that substituting x_i with $1/x_i$ in the matrix M_λ 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 $\lambda' := m'_1, m'_2, \ldots, m'_n$ where $m_i + m'_{n-i+1} = m_1$. Up to a sign thus:

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

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

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.

7 Cauchy formulas

7.1 The formulas we want to discuss have important applications in representation theory, for the moment we wish to present them as purely combinatorial identities.

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

the right hand side is the sum over all partitions.

$$\prod_{i \le j=1,n} \frac{1}{1 - x_i x_j} = \sum_{\lambda \in \Lambda_{ec}} S_\lambda(x),$$

if n is even

$$\prod_{i < j=1,n} \frac{1}{1 - x_i x_j} = \sum_{\lambda \in \Lambda_{er}} S_\lambda(x),$$

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

Let us start from the first one. It can be deduced considering the determinant of the $n \times n$ matrix:

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

Claim

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

Subtracting the first row to the i^{th} one has a new matrix (b_{ij}) where:

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

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} columnn the factor $\frac{1}{1-x_1y_j}$.

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

$$\begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ \frac{y_1}{1-x_2y_1} & \frac{y_2}{1-x_2y_2} & \dots & \dots & & \frac{y_n}{1-x_2y_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{y_1}{1-x_ny_1} & \frac{y_2}{1-x_ny_2} & \dots & \dots & & \frac{y_n}{1-x_ny_n} \end{pmatrix} \tag{7.1.1}$$

subtracting the first column to the i^{th} we get the terms $\frac{y_i-y_1}{(1-x_jy_1)(1-x_jy_i)}$ thus we end extracting the product $\prod_{i=2}^n \frac{(y_i-y_1)}{(1-x_iy_1)}$ and we are left with the determinant of the same type of matrix but without the variables x_1, y_1, we can thus finish by induction.

Now we can develop the determinant by developing each element $\frac{1}{1-x_iy_j} = \sum_{k=0}^\infty x_i^k y_j^k$ 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:

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

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 so to be decreasing at at the same time we must introduce a sign, collecting all the terms in which the k_i are a permutation of a given sequence $\lambda + \rho$ we get the term $A_\lambda(x)A_\lambda(y)$. Finally:

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

From this the required identity follows.

8 The conjugation action

8.1 We inspect now a representation strictly 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}.

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, by definition $\sigma_i(A)$ is the i^{th} elementary symmetric function computed in the eigenvalues of A.

Recall that S_n can be viewed as a subroup 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 the ones which come from the algebraic structure of these sets (as affine varieties) and namely, on $M_n(\mathbb{C})$ 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 plynomial 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. 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 theorem follows therefore from:

Exercise. *The set of diagonalizable matrices is dense.*

Hint A matrix with distinct eigenvalues is diagonalizable, these matrices are characterized by the fact that on them the discriminant is non 0.

On any $\mathbb{C}^k$ the set of points where a (non identically zero) polynomial $u(x)$ is non zero 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..

The statements for the two groups are simiar and we leave them to the reader.

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 leave 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 $A^i v$, $i = 0, \ldots, n-1$ 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*

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

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 matrix conjugate to A the one of the same linear transformation in the basis $A^i v$, $i = 0, \ldots, n-1$.

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

As for 3 we leave it here since we have not yet developed enough geometry of orbits.

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

One can prove easily 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 a new proof of the theorem on invariants for the conjugation representation.

There is a deeper theory of all sheets due to Peterson [P].

9 The Aronhold method, polarization.

9.1 Before proceding let us recall in a language suitable for our purposes the usual Taylor Maclaurin expansion.

Consider a function $F(x)$ of a vector variable $x \in V$, under various types of assumptions we have a development for the function $F(x+y)$ of two vector variables.

In our case we may restrict to polynomials and develop $F(x+y) := \sum_{i=0}^{\infty} F_i(x,y)$ where by definition $F_i(x,y)$ is homogeneous of degree i in y (of course for polynomials the sum is really finite).

Therefore, for any value of a parameter λ, we have $F(x+\lambda y) := \sum_{i=0}^{\infty} \lambda^i F_i(x,y)$.

If F is also homogeneous of degre k we have:

$$\sum_{i=0}^{\infty} \lambda^k F_i(x,y) = \lambda^k F(x+y) = F(\lambda(x+y)) = F(\lambda x + \lambda y) = \sum_{i=0}^{\infty} \lambda^i F_i(\lambda x, y)$$

and we deduce that $F_i(x,y)$ is also homogeneous of degree $k-i$ in x.

Given 2 functions F, G we clearly have that

$$F(x+y)G(x+y) = \sum_{i=0}^{\infty} \sum_{a+b=i} F_a(x,y)G_b(x,y)$$

is the decomposition in homogeneous components relative to y.

The operator $D = D_{y,x}$ defined by the formula $D_{y,x}F(x) := F_1(x,y)$ is clearly linear and also by the previous formula we have $D(FG) = D(F)G + FD(G)$, these are the defining conditions of a *derivation.*

If we use coordinates $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ we have that $D_{y,x} = \sum_{i=1}^{n} y_i \frac{\partial}{\partial x_i}$.

Definition. *The operator* $D_{y,x} = \sum_{i=1}^{n} y_i \frac{\partial}{\partial x_i}$ *is called a polarization operator.*

So the effect of applying it to a bihomogeneous function of 2 variables x, y is to decrease by 1 the degree of the function in x and raise by 1 the degree in y.

Assume we are now in characteristic 0, we have then the standard theorem of calculus:

Theorem. $F(x+y) = \sum_{i=0}^{\infty} \frac{1}{i!} D_{y,x}^i F(x)$.

Proof. We reduce to the one variable theorem and deduce that

$$F(x+\lambda y) := \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \frac{d^i}{d\lambda^i} F(x+\lambda y)_{\lambda=0}$$

then

$$F_i(x,y) = \frac{1}{i!} \frac{d^i}{d\lambda^i} F(x+\lambda y)_{\lambda=0}$$

and this is by the chain rule $\frac{1}{i!}D^i_{y,x}F(x)$.

9.2 Suppose now that we consider the action of an invertible linear transformation on functions, we have $(gF)(x+y) = F(g^{-1}x + g^{-1}y)$ hence we deduce that the polarization operator commutes with the action of the linear group. The main consequence is the:

Proposition. *If $F(x)$ is an invariant of a group G so are the polarized forms $F_i(x,y)$.*

Of course implicitely we are using the (direct sum) linear action of G on pairs of variables.

Let us further push this idea, consider now any number of vector variables and, for a polynomial function F, homogeneous of degree m the expansion:

$$F(x_1 + x_2 \cdots + x_n) = \sum_{h_1,h_2,\dots,h_n} F_{h_1,h_2,\dots,h_n}(x_1,x_2,\dots,x_n)$$

where $\sum h_i = m$, the indeces h_i represent the degrees of homogeneity in the variables x_i. A repeated application of the Taylor Maclaurin expansion gives:

$$(9.2.1) \qquad F_{h_1,h_2,\dots,h_n}(x_1,x_2,\dots,x_n) = \frac{1}{h_1!h_2!\dots h_n!} D^{h_1}_{x_1x} D^{h_2}_{x_2x} \dots D^{h_n}_{x_nx} F(x)$$

In particular in the expansion of $F(x_1 + x_2 \cdots + x_m)$ there will be a term which is linear in all the variables x_i, this is called the *full polarization of the form F.*

Let us indicate with $PF := F_{1,1,\dots,1}(x_1,\dots,x_m)$, to stress the fact that this is a linear operator. It is clear that, if $\sigma \in S_m$ is a permutation

$$F(x_1 + x_2 \cdots + x_n) = F(x_{\sigma 1} + x_{\sigma 2} \cdots + x_{\sigma n})$$

hence we deduce that the polarized form satisfies the *symmetry* property:

$$PF(x_1,\dots,x_m) = PF(x_{\sigma 1},\dots,x_{\sigma m}),$$

we have thus found that:

Lemma. *The full polarization is a linear map from the space of homogeneous forms of degree m to the space of symmetric multilinear functions in m (vector) variables.*

Now let us substitute for each variable x_i the variable $\lambda_i x$ (the λ_i's being distinct numbers), we obtain:

$$(\lambda_1 + \lambda_2 \cdots + \lambda_m)^m F(x) = F((\lambda_1 + \lambda_2 \cdots + \lambda_m)x) =$$

$$= F(\lambda_1 x + \lambda_2 x \cdots + \lambda_m x) = \sum_{h_1,h_2,\dots,h_m} F_{h_1,h_2,\dots,h_m}(\lambda_1 x, \lambda_2 x, \dots, \lambda_m x) =$$

$$= \sum_{h_1,h_2,\dots,h_m} \lambda_1^{h_1}\lambda_2^{h2}\dots\lambda_m^{h_m} F_{h_1,h_2,\dots,h_m}(x,x,\dots,x),$$

comparing the coefficients of the same monomials on the two sides we get

$$\binom{m}{h_1h_2\dots h_m} F(x) = F_{h_1,h_2,\dots,h_m}(x,x,\dots,x),$$

in particular:

$$m!F(x) = PF(x, x, \dots, x)$$

Since we are working in characteristic zero we can also rewrite this identity as:

$$F(x) = \frac{1}{m!}PF(x, x, \dots, x)$$

The linear operator

$$R : G(x_1, x_2, \dots, x_m) \to \frac{1}{m!}G(x, x, \dots, x)$$

is called in the classical literature the *restitution.* We have:

Theorem. *The maps P, R are inverse isomorphisms, equivariant for the group of all linear transformations, between the space of homogeneous forms of degree m and the space of symmetric multilinear functions in m variables.*

Proof. We have already proved that $RPF = F$, let now $G(x_1, x_2, \dots, x_m)$ be a symmetric multilinear function. In order to compute PRG we must determine the multilinear part of $\frac{1}{m!}G(\sum x_i, \sum x_i, \dots, \sum x_i)$.

By the multilinearity of G we have that

$$G(\sum x_i, \sum x_i, \dots, \sum x_i) = \sum G(x_{i_1}, x_{i_2}, \dots, x_{i_m})$$

where the right sum is over all possible sequences of indeces $i_1 i_2 \dots i_m$ out of the numbers $1, \dots, m$. But the multilinear part is exactly the sum over all the sequences without repetitions, i.e. the permutations. Thus

$$PRG = \frac{1}{m!}\sum_{\sigma \in S_m} G(x_{\sigma 1}, x_{\sigma 2}, \dots, x_{\sigma m}),$$

since G is symmetric this sum is in fact G. □

Remark. The main use that we will make of the previous theorem will be to reduce the computation of invariants to the multilinear ones. At this point it is not yet clear why this should be simpler, in fact we will see that in several interesting cases this turns out to be true and we will succeed to compute all the invariants by this method. This sequence of ideas is sometimes referred to as *Arhonold's method.*

9.3 In order to formalize the previous method consider an infinite sequence of $n-$dimensional vector variables $x_1, x_2, \dots, x_k, \dots$, each x_i being a vector $x_{1i}, x_{2i}, \dots, x_{ni}$, in other words we cosider the x_{ij} as the coordinates of the space of $n \times \infty$ matrices or of the space of sequences of (column) vectors.

Let $A = F[x_{ij}]$ be the polynomial ring in the variables x_{ij}, for the elements of A we have the notion of being homogeneous with respect to one of the vector variables x_i and A is a *multigraded ring.*

With $A_{h_1, h_2, \dots, h_i, \dots}$ we denote the multihomogeneous part relative to the degrees

$h_1, h_2, \ldots, h_i, \ldots, .$

We have of course the notions of multihomogeneous subspace or subalgebra. For each pair i, j of indeces we consider the corresponding polarization operator $D_{ij} = \sum_{h=1}^{n} x_{hi} \frac{\partial}{\partial x_{hj}}$, we view these operators as acting as derivations on the ring A.

Given a function F homogeneous in the vector variables $x_1, x_2, \ldots, x_m$ of degrees $h_1, h_2, \ldots, h_m$, we can perform the process of polarization on each of the variables x_i as follows.

Choose out of the infinite list of vector variables m disjoint sets X_i of variables each with h_i elements and we fully polarize the variable x_i with respect to the chosen set X_i.

The result is multilinear and symmetric in each of the sets X_i, the function F is recovered from the polarized form by a sequence of restitutions.

We should remark that a restitution is a particular form of polarization since, if a function F is linear in the varable x_i the effect of the operator D_{ji} on F is that of substituting in F the variable x_i with x_j.

Definition. *A subspace V of the ring A is stable under polarization if it is stable under all polarization operators.*

Remark. Given a polynomial F , F is homogeneous of degree m with respect to the vector variable x_i if and only if $D_{ii}F = mF$.

From this remark one can easily prove the following:

Lemma. *A subspace V of A is stable under the polarizations D_{ii} if and only if it is multihomogeneous.*

9.4 In this section we will use the term *multilinear function* in the following sense:

Definition. *We say that a polynomial $F \in A$ is multilinear if it is homogeneous of degree 0 or 1 in each of the variables x_i.*

In particular we can list the indeces of the variables $i_1, \ldots, i_k$ in which F is linear (the variables which appear in the polynomial) and say that F is multilinear in the x_{i_j}.

Given a subspace V of A we will denote by V_m the set of multilinear elements of V.

Theorem. *Given 2 subspaces V, W of A stable under polarization and such that $V_m \subset W_m$ then $V \subset W$.*

Proof. Since V is multihomogeneous it is enough to prove that, given a multihomogeneous function F in V we have $F \in W$. We know that F can be obtained by restitution fron its fully polarized form $F = RPF$. The hypotheses imply that $PF \in V$ and hence $PF \in W$. Since the restitution is a composition of polarization operators and W is assumed to be stable under polarization we deduce that $F \in W$.

Corollary. *If two subspaces V, W of A are stable under polarization and $V_m = W_m$ then $V = W$.*

This corollary is what we shall use often to compute invariants, the strategy is as follows. We want to compute the space W of invariants in A under some group G of linear transformations in $n-$dimensional space. We produce some list of invariants (which are more or less obvious) forming a subspace V closed under polarization and we hope to have found all invariants and try to prove $V = W$, then if we can do it for the multilinear invariants we are done.

10 Lie Algebras and Lie groups

10.1 The polarization operators are special types of *derivations*. Let us recall the general definitions. First of all, given an associative algebra A, we define the *Lie product*

$$[a, b] := ab - ba,$$

and verify immediately that it satisfies the properties:

$[a, b] = -[b, a]$ (antisymmetry), and $[a, [b, c]] + [b, [c, a]] + [c, [a, b]] = 0$ (Jacobi identity).

More generally an algebra (i.e. a vector space with a bilinear product denoted by $[a, b]$) satisfying the antisymmetry and the Jacobi identity is called a *Lie algebra.*

The first class of Lie algebras to be considered are the algebras $gl(U)$, the Lie algebra associated to the associative algebra $End(U)$ of linear operators on a vector space U.

Given any algebra A (not necessarily associative), with product denoted ab, we define:

Definition. *A derivation of A is a linear mapping $D : A \to A$ satisfying $D(ab) = D(a)b + aD(b)$ for every $a, b \in A$.*

The main remarks are:

Proposition.

i) In a Lie algebra L the Jacobi identity expresses the fact that the map

$$Ad(a) := b \to [a, b]$$

is a derivation.[5]

ii) The derivations of any algebra A form a Lie subalgebra of the space of linear operators.

Proof. By direct verification.

The main reason why Lie algebras and derivations are important is that they express the infinitesimal analogues of groups and symmetries.

The main geometric example is given by:

[5] *Ad* stands for *adjoint* action.

Definition. *A derivation X of the algebra $C^\infty(M)$ of C^∞ functions on a manifold M is called a vector field.*

The first formal properties of vector fields is that they are *local*, this means that, given a function $f \in C^\infty(M)$ and an open set U the value of $X(f)$ on U depends only on the value of f on U, in other words:

Lemma. *If $f \cong 0$ on U also $X(f) \cong f$ on U.*

Proof. Let $p \in U$ and V a small neighborhood of p in U so that we can find a C^∞ function u on M which is 1 outside U and 0 in V.

Hence $f = uf$ and $X(f) = X(u)f + uX(f)$ which is manifestly 0 in p.

The previous property can be easily interpreted by saying that we can consider either a derivation of the algebra of C^∞ functions as a section of the tangent bundle, in local coordinates x_i we have $X = \sum_{i=1}^n f_i(x_1, \dots, x_n)\dfrac{\partial}{\partial x_i}$.

It is important at this moment to start to introduce and use more systematically the language of Lie groups and Lie algebras, we will not start in the full generality.

Consider the finite dimensional vector space F^n where F is either $\mathbb{C}$ or $\mathbb{R}$ (complex or real field), with its standard Hilbert norm. Given a matrix A we define its norm:

$$\{|A| := max(\frac{|A(v)|}{|v|}),\ v \neq 0\},\ \text{ or } |A| =\ max(|A(v)|), |v| = 1.$$

Of course this extends to infinite dimensional Hilbert spaces and bounded operators, i.e. linear operators A with

$$sup_{|v|=1}(|A(v)|) := |A| < \infty$$

$|A|$ is a norm on the space of bounded operators, i.e.:

(1) $|A| \geq 0,\ |A| = 0$ if and only if $A = 0$.
(2) $|\alpha A| = |\alpha||A|,\ \forall \alpha \in F, \forall A$.
(3) $|A + B| \leq |A| + |B|$.

With respect to the multiplicative structure the following facts can be easily verified:

Proposition. *i) Given two operators A, B we have $|AB| \leq |A||B|$.*
ii) The series $e^A := \sum_{k=0}^\infty \frac{A^k}{k!}$ is totally convergent in any bounded subset of the space of operators.
iii) The series $log(1 + A) := \sum_{k=1}^\infty (-1)^{k+1}\frac{A^k}{k}$ is totally convergent for $|A| \leq 1 - \epsilon$ any $\varepsilon > 0$.
iv) The functions e^A and $logA$ are inverse of each other in suitable neighborhoods of 0 and 1.

Remark. For matrices (a_{ij}) we can also take the equivalent norm $max(|a_{ij}|)$.

10.2 The following properties of the exponential map are easily verified:

Proposition.

i) If A, B are two commuting operators (i.e. $AB = BA$) we have $e^A e^B = e^{A+B}$ and also $log(AB) = log(A) + log(B)$ if A, B are sufficiently close to 1.

ii) $e^{-A}e^A = 1$.

iii) $\dfrac{de^{tA}}{dt} = Ae^{tA}$

iv) $Be^A B^{-1} = e^{BAB^{-1}}$.

Furthermore for matrices we have also:

v) If $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the eigenvalues of A, the eigenvalues of e^A are $e^{\alpha_1}, e^{\alpha_2}, \ldots, e^{\alpha_n}$.

vi) $det(e^A) = e^{Tr(A)}$.

vii) $e^{A^t} = (e^A)^t$.

In particular the mapping $t \to e^{tA}$ is a homomorphism from the additive group of real (or complex) numbers to the multiplicative group of matrices (real or complex).

This homomorphism is called *the 1 parameter subgroup generated by A*. Given a vector v_0 the function $v(t) := e^{tA}v_0$ is the solution to the differential equation $v'(t) = Av(t)$ with initial condition $v(0) = v_0$.

10.3 It is not restrictive to consider such 1-parameter subgroups, in fact we have:

Proposition. *A continuous homomorphism $\varphi : \mathbb{R} \to Gl(n, F)$ (from the additive group of real numbers to $Gl(n, F)$) is of the form e^{tA} for a unique matrix A, called the infinitesimal generator of the group φ. We also have $A = \dfrac{d\varphi(t)}{dt}_{t=0}$.*

Proof. Since the map φ is continuous there is a neighborhood of 0 such that the values $\varphi(t)$ lie in a neighborhood of 1 where the logarithm is defined.

From the group homomorphism hypotheses and the properties of the logarithm one immediately has that $log(\varphi(t + s)) = log(\varphi(t)) + log(\varphi(s))$ for s, t sufficiently small.

Now it is an easy exercise to prove that any continuous mapping $\lambda(t)$ from a neighborhood of 0 in $\mathbb{R}$ to a finite dimensional vector space F^n which satisfies $\lambda(t+s) = \lambda(t)+\lambda(s)$, for t, s close to zero is, in a neighborhood of 0, of the form $\lambda(t) = tv$ for a unique vector v.

Thus, at least for t small we have $log(\varphi(t)) = tA$ for some matrix A and so $\varphi(t) = e^{tA}$.

The group homomorphism hypotheses imply immediately that this eqality is valid for all values of t. In fact any t it is of the form $t = nt^0$ with n a positive integer and t^0 arbitrarily small, and so $\varphi(t) = \varphi(t^0)^n = (e^{t^0 A})^n = e^{tA}$.

In the proof we had in mind the following simple facts.

Remark. If two actions of a group G coincide on a set of generators of G then they are equal.

A connected topological group is generated by the elements of any given non empty open set.

10.4 Let us develop a few basic properties of 1 parameter groups. First of all:

Proposition. *A vector v is fixed under a group e^{tA} if and only if $Av = 0$.*

Proof. If $Av = 0$ then $t^k A^k v = 0$ for $k > 0$ hence $e^{tA}v = v$, conversely if $e^{tA}v$ is constant its derivative is 0, but its derivative at 0 is in fact Av.

Remark. Suppose that e^{tA} leaves a subspace U stable, then, if $v \in U$ we have that $Av \in U$ since this is the derivative of $e^{tA}v$ at 0. Conversely if A leaves U stable it is clear from the definition that e^{tA} leaves U stable and on U it is the 1 parameter subgroup generated by the restriction of A.

We stress again that the main application of the exponential is to systems of ordinary linear differential equations. Precisely let us consider the system $\frac{dy(t)}{dt} = Ay(t)$, where A is a constant $n \times n$ matrix and $y(t)$ a vector of unkown functions, if we fix the initial condition $y(0)$ we have the global solution $y(t) = e^{tA}y(0)$.

Let us now consider a function $f(y)$ on an n-dimensional vector space, we can follow its evolution under a 1 parameter subgroup and set $\varphi(t)(f) := f(t, y) := f(e^{-tA}y)$, we have thus a 1 parameter group of transformations on the space of functions, induced from the action on the space.

Of course this is not a finite dimensional space and so we cannot apply directly the results of the previous paragraph. If we restrict for a moment to homogeneous polynomials we are in the finite dimensional case, thus for this group we have $\varphi(t)(f) = e^{tD_A}f$ where D_A is the linear operator defined by:

$$D_A(f) = \frac{df(t,y)}{dt}_{t=0}.$$

We have $\frac{df(t,y)}{dt} = \sum_{i=1}^n \frac{\partial f}{\partial y_i}\frac{dy_i(t)}{dt}$ and since $\frac{dy(t)}{dt} = -Ay(t)$, at $t = 0$ $\frac{dy_i(t)}{dt}_{t=0} = -\sum_{j=1}^n a_{ji}y_j$ hence $\frac{df(t,y)}{dt}_{t=0} = -\sum_{i=1}^n \sum_{j=1}^n a_{ji}y_j \frac{\partial f}{\partial y_i}$.

Thus we have found that D_A is the differential operator:

$$D_A := -\sum_{i=1}^n \sum_{j=1}^n a_{ji}y_j \frac{\partial}{\partial y_i}.$$

We deduce that the formula $\varphi(t)(f) = e^{tD_A}f$ is just the Taylor series:

$$f(t,y) = \sum_{k=0}^{\infty} \frac{(tD_A)^k}{k!} f(y).$$

In order to understand better the operators D_A let us compute

$$D_A y_i = \sum_{j=1}^n a_{ji}y_j.$$

We see that, on the linear space spanned by the functions y_i's this is just the linear operator given by the matrix A.

Since a derivation is determined by its action on the variables y_i's we have:

Proposition. *The differential operators D_A are a Lie algebra and:*

$$[D_A, D_B] = D_{[A,B]}.$$

Proof. This is true on the space spanned by the variables (cf. later comments on the contragredient action).

Example. $Sl(2,\mathbb{C})$.

We want to study the case of polynomials in 2 variables x, y. The Lie algebra of 2×2 matrices decomposes as the direct sum of the 1 dimensional algebra generated by $D = x\dfrac{\partial}{\partial x} + y\dfrac{\partial}{\partial y}$ and the 3 dimensional algebra $sl(2,\mathcal{C})$ with basis the operators

$$H = -x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y},\ E = -y\frac{\partial}{\partial x},\ F = -x\frac{\partial}{\partial y},$$

these operators correspond to the matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

We can see how these operators act on homogeneous polynomials of degree n, this is an $n+1$ dimensional space spanned by the monomials $u_i := (-1)^i y^{n-i} x^i$ on which D acts by multiplication by n . We have:

$$Hu_i = (n-2i)u_i \quad Fu_i = (n-i)u_{i+1} \quad Eu_i = iu_{i-1}.$$

The reader who has seen these operators before will recognize the standard irreducible representations of the Lie algebra $sl(2,\mathcal{C})$.

We will return to this point later remarking only that for larger numbers of variables the situation is much more complex.

10.5 Of course these ideas have much more general range of validity, for instance the main facts about exponential and logarithm are sufficiently general to hold for any Banach algebra, i.e. an algebra with a norm for which it is complete. Thus one can apply these results also to bounded operators on a Hilbert space.

Moreover the linearity of the transformations A is really not essential, for instance if we consider a C^∞ differentable manifold M we can discuss dynamical systems (in this case also called *flows*) in 2 ways.

Definition. *A C^∞ flow on a manifold M is a C^∞ map:*

$$\phi(t,x) : \mathbb{R} \times M \to M$$

which defines a group action of $\mathbb{R}^+$ on M.

It is also usual to think of the diffeomorphisms:

$$\phi_s : M \to M, \qquad \phi_s(m) := \phi(s,m); \qquad \phi_0 = 1_M, \quad \phi_{s+t} = \phi_s \circ \phi_t.$$

To a flow is associated a vector field X, called the *infinitesimal generator of the flow.*

The vector field X associated to a flow $\phi(t,x)$ can be defined in each point p as:

$$X_p := d\phi_{(0,p)}(\frac{d}{dt}).$$

In other words, given a function f on M we have:

$$X(f)(p) := \frac{d}{dt}f(\phi(t,p))_{t=0}.$$

In other words, given a point p the map $t \to \phi(t,p)$ describes a curve in M which is the evolution of p or the orbit under the flow and X_p is the *velocity* of this evolution in p, which by the group action depends only on the position and not on the time:

$$X_{\phi(s,p)} := d\phi_{(s,p)}(\frac{d}{dt}).$$

In the previous paragraph thus we were studying a special class of vector fields on $n-$dimensional space:

$$D_A := -\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ji}x_j\frac{\partial}{\partial x_i}.$$

A vector field gives rise at least locally, to a flow which one obtains solving a linear system of ordinary differential equations

$$\frac{dx_i(t)}{dt} = f_i(x_1,\dots,x_n)$$

thus one has local solutions $\varphi(t)(x) = F(t,x)$, with $F(0,x) = x$ defined for small values of t in a neighborhhod of a given point x^0, which by the uniqueness of solutions of ordinary differential equations satisfy the properties of local 1 parameter groups, again a point is fixed under the flow if and only if the vector field vanishes in it.

For the vector field D_A the flow is the 1-parameter group of linear operators e^{tA}.

We can again define, at least locally, the evolution of a function f by the formula $f(t,x) := f(F(t,x))$, by restricting to an orbit we again have the Taylor series for $f(t,x)$.

But, by the definition of the orbit, the derivative with respect to t in a point of the orbit is the same as the derivative with respect to the vector given by X hence the Taylor series $\sum_{k=0}^{\infty} t^k\frac{X^k}{k!}f(x)$.

In this sense the flow becomes a linear flow on the space of functions with infinitesimal generator X and equals e^{tX}.

Of course in order to make this equality strict we need some hypotheses, as the fact that the flow exists globally and also that the functions under consideration are analytic.

The special case of linear flows has the characteristic that one can find global coordinates on the manifold so that the evolution of these coordinates is by a linear group of ttransformations of the finite dimensional vector space spanned by the coordinates! In general of course the evolution of coordinates develops non linear terms.

10.6 Let us return to the point *derivations* and *automorphisms*.

Consider thus an algebra A and a linear operator D on A, assume that there are sufficient convergence properties to insure the existence of e^{tD} as convergent power series (like for Banach algebras) then:

Proposition. *D is a derivation if and only if e^{tD} is a group of automorphisms.*

Proof. This is again a variation of the fact that a vector v is fixed under e^{tD} if and only if $Dv = 0$, in fact to say that e^{tD} are automorphisms means that:

$$a, b \in A,\ e^{tD}(ab) - e^{tD}(a)e^{tD}(b) = 0.$$

Writing in power series and taking the coefficient of the linear term we get

$$D(ab) - D(a)b - aD(b) = 0$$

the condition for a derivation.

Conversely given a derivation we see by an easy induction that, for any positive integer k,

$$D^k(ab) = \sum_{i=0}^{k} \binom{k}{i} D^{k-i}(a)D^i(b),$$

hence:

$$e^{tD}(ab) = \sum_{k=0}^{\infty} \frac{t^k D^k(ab)}{k!} = \sum_{k=0}^{\infty} t^k \sum_{i=0}^{k} \frac{\binom{k}{i}}{k!} D^{k-i}(a)D^i(b) =$$

$$= \sum_{k=0}^{\infty} \sum_{i=0}^{k} \frac{1}{(k-i)!i!} t^{k-i} D^{k-i}(a) t^i D^i(b) = e^{tD}(a)e^{tD}(b).$$

§11 Lie groups

11.1 As we have already mentioned it is quite interesting to analyze group actions subject to special structural requirements.

First of all in a group G we always have the two basic maps which describe its structure, the multiplication $m : G \times G \to G,\ m(a, b) := ab$ and the inverse $i : G \to G,\ i(g) := g^{-1}$. If G has an extra structure it is natural to consider the compatibility of these maps with the structure thus we will say that:

Definition. *A group G is a:*

1) topological group,

2) Lie group,

3) complex analytic group,
4) algebraic group,
5) affine group,

if G is also a

1) topological space,
2) differentiable manifold,
3) complex analytic manifold,
4) algebraic variety,
5) affine variety,

and if the two maps m, i are compatible with the given strucure, i.e. are continuous, differentiable, complex analytic or regular algebraic.

When speaking of Lie groups we have not discussed the precise differentiability hypotheses. In fact there is a rather general theorem (solution of Hilbert's inserire problem) which insures that a topological group which is locally homeomorphic to euclidean space can be naturally given a real analytic structure, thus Lie groups are in fact real analytic manifolds.

The group $GL(n, \mathbb{C})$ is clearly an affine algebraic group, acting on $\mathbb{C}^n$ by linear and hence algebraic transformations. A group G is called a linear group if it can be embedded in $GL(n, \mathbb{C})$ (of course one shoud more gerally consider as linear groups the subgroups of $GL(n, F)$ for an arbitrary field F).

As we shall see a closed subgroup of a Lie group is automatically a real analytic submanifold and hence a Lie subgroup.

For an action of G on a set X we can also have the same type of analysis.

Continuous action of a topological group on a topological space, differentiable actions of Lie groups on manifolds etc..

We shall meet many very interesting examples of these actions in the course of our treatment.

Let us consider manifolds and Lie groups.

11.2 In order to understand the previous considerations in a more general setting let $H \subset GL(n, F)$ be a closed linear subgroup (F the field of real or complex numbers). What we are going to discuss would be valid for any closed subgroup of a Lie group.

Set $L := \{A \in M_n(F) | e^{tA} \in H, \ \forall t \in \mathbb{R}.\}$

Theorem. *a) L is a Lie subalgebra of $M_n(F)$ (called $Lie(H)$).*

b) There are neighborhoods A, B of 0,1 in $M_n(F)$ so that exp, log are inverse isomorphisms between A, B and the set $log(B \cap H)$ is the intersection of A with the Lie subalgebra L.

In particular we will have that $exp(L) \subset H$ and, if H is connected it is generated by $exp(L)$.

We want to apply the previous analysis to invariant theory, let us give a linear action of a connected Lie group G on a space U, thus G acts as a group of automorphisms on the polynomial ring $P[U]$ and its Lie algebra as derivations .

Theorem. *A polynomial f is invariant under G if and only if it satisfies the differential equations $Af = 0$ for all $A \in Lie(G)$.*

Proof. Since G is connected it is generated by its 1 parameter subgroups $exp(tA)$, $A \in Lie(G)$. Hence a polynomial f is fixed under G if and only if it is fixed under these 1 parameter groups. We have seen that f is fixed under $exp(tA)$ if and only if $Af = 0$.

11.3 We finish by making the connection with polarizations, let us consider as in section 1.3.3 $m-$tuples of vector variables $x_1, x_2, \ldots, x_m$ each x_i being a vector $x_{1i}, x_{2i}, \ldots, x_{ni}$, in other words we consider the x_{ij} as the coordinates of the space $M_{n,m}$ of $n \times m$ matrices.

Let $A = F[x_{ij}]$ be the polynomial ring in the variables x_{ij}, which we also think as polynomials in the vector variables x_i given by the columns; on $M_{n,m}$ we want to consider some special 1 parameter subgroups (induced by left or right multiplications).

For any $m \times m$ matrix A we consider the 1 parameter group $X \to Xe^{-tA}$.

In particular for the elementary matrix e_{ij} (with 1 in the ij position and 0 elsewhere), the matrix $Xe^{-te_{ij}}$ is obtained from X adding to its j^{th} column its i^{th} column multiplied by $-t$.

We act dually on the functions in A and the 1 parameter group acts substituting x_j with $x_j + tx_i$. By the results of the previous sections we see that:

Proposition. *The infinitesimal generator of the transformation of functions induced by $X \to Xe^{-te_{ij}}$ is the polarization operator D_{ji}.*

We should summarize these ideas.

On the space of $n \times m$ matrices acts the group $GL(m, F)$ by the action $(A, X) \to XA^{-1}$.

The infinitesimal action is then $X \to -XA$.

If we denote by r_A this operator, we have therefore $[r_A, r_B] = r_{[A,B]}$. In other words the map $A \to r_A$ is a *Lie algebras homomorphism* associated to the given action.

The derivation operators induced on polynomials are the linear span of the polarization operators which correspond to elementary matrices.

A space of functions is stable under polarization if and only if it is stable under the action of $GL(m, F)$.