Introduction and preliminaries

1. Introduction

The purpose of these notes is to give a unified, complete and self contained exposition of the main algebraic theorems of invariant theory for matrices in a characteristic free approach.

We only treat the general symbolic theory and do not discuss the several special theorems and computations which can be found in a comprehensive paper of Vesselin Drensky, [**23**], which also contains a rather complete literature. Nor do we treat the applications to non-commutative algebra which will appear in a forthcoming book *Rings with polynomial identities and finite dimensional representations of algebras* by E. Aljadeff, A. Giambruno, C. Procesi, A. Regev ([**1**]).

In particular we will avoid the use of algebraic geometry, and results on the cohomology of line bundles over the flag variety which are the standard approach to the theory, see [**34**]. Instead, we shall use the theory of quasi-hereditary algebras and of standard bitableaux, an algebraic and combinatorial approach.

If G is a group acting linearly on a finite dimensional vector space W, then G acts on the algebra $S[W^*]$ of polynomial functions on W by $(gf)(w) := f(g^{-1}w)$ and one defines.

DEFINITION 1.1. The algebra of invariants of G acting on W is:

$$S[W^*]^G := \{f \in S[W^*] \,|\, gf = f, \ \forall g \in G\}.$$

Invariant theory has a long and complicated history whose discussion goes far beyond the purpose of these notes.

The prototype theory is the theory of symmetric functions, largely studied in the last two centuries.

This can be formulated in an *arithmetic way* by taking the ring $\mathbb{Z}[x_1, \ldots, x_d]$ of polynomials in d variables over the integers and the action of the symmetric group S_d permuting the variables. In this case the ring of invariants $\mathbb{Z}[x_1, \ldots, x_d]^{S_d}$ is a polynomial ring $\mathbb{Z}[\sigma_1(x), \ldots, \sigma_d(x)]$ where the polynomials $\sigma_i(x_1, \ldots, x_d)$, called *elementary symmetric functions* are defined by the formula

$$\prod_{i=1}^{d}(\lambda - x_i) = \lambda^d - \sigma_1(x)\lambda^{d-1} + \sigma_2(x)\lambda^{d-2} + \ldots + (-1)^d\sigma_d(x). \tag{1}$$

For a modern treatment see [**41**]. It is useful to pass to infinitely many variables as follows, consider the map $\pi_{d+1} : \mathbb{Z}[x_1, \ldots, x_d, x_{d+1}] \to \mathbb{Z}[x_1, \ldots, x_d]$ given by $\pi_{d+1} : x_{d+1} = 0$.

Then the restriction of π_{d+1} to $\mathbb{Z}[x_1, \ldots, x_d, x_{d+1}]^{S_{d+1}} = \mathbb{Z}[\sigma_1, \ldots, \sigma_d, \sigma_{d+1}]$ maps it to $\mathbb{Z}[x_1, \ldots, x_d]^{S_d}$ and is given by $\pi_{d+1} : \sigma_{d+1} = 0$.

Thus one can define a limit algebra, called the *ring of symmetric functions* which is the polynomial ring in the infinitely many variables $\sigma_i(x), \ i \in \mathbb{N}$. The ring $\mathbb{Z}[x_1, \ldots, x_d]^{S_d}$ is obtained from this ring by setting $\sigma_i(x) = 0, \ \forall i > d$. Identities among different types of symmetric functions are usually described in the limit ring.

In the last century, due to work of Brauer, Schur and Weyl invariant theory is tied to representation theory. We will sketch some of the theory in §3.1. In particular in the book of H. Weyl *The classical groups* [**72**], it is developed a first fundamental theorem, *FFT* and a second fundamental theorem, *SFT* for invariants of several copies of the defining representation V of $GL(V)$ and its dual V^* and also for the other classical groups, in all these cases *tensor symmetry* is the main tool.

These notes deal with a more difficult case in which a first and a second fundamental theorems can be proved, the case of several copies of $End(V)$ under conjugation action by $GL(V)$. Moreover contrary to Weyl's approach which is valid only in characteristic 0 we will develop a characteristic free approach.

We will not discuss the theory of several copies of $End(V)$ under conjugation action by the classical groups $O(V)$ or $Sp(V)$ when V has a non degenerate symmetric, resp. skew symmetric bilinear form.

Such a theory exists in characteristic 0, see [**45**] or [**48**], and in a characteristic free approach one can see [**77**] and [**39**].

Thus the object of study is the following. Let F be an infinite field or the integers $\mathbb{Z}$, we will denote by $M_m(F)$ the algebra of $m \times m$ matrices with coefficients in F. Given two integers n, m consider the ring

$$S_m[\xi_1, \dots, \xi_n] = F[x^h_{i,j}],\ i, j = 1, \dots, m;\ h = 1, \dots, n \tag{2}$$

the (commutative) algebra of polynomial functions (in $m^2 n$ variables) on the vector space $M_m(F)^n$ of n-tuples of $m \times m$ matrices. The symbols ξ_h refer to the *generic matrices* $\xi_h := (x^h_{i,j})$. The linear group $GL(m, F)$ acts on the space $M_m(F)^n$ by simultaneous conjugation. So the algebra $S_m[\xi_1, \dots, \xi_n]^{GL(m,F)}$ of invariants of n-tuples of $m \times m$ matrices, is the subring of polynomials f such that $f(g\xi_1 g^{-1}, \dots, g\xi_n g^{-1}) = f(\xi_1, \dots, \xi_n)$, $\forall g \in GL(m, F)$.

Remark 1.2. As we shall see this ring of invariants is in fact *defined* over the prime field, and obtained by change of scalars from the invariants defined over $\mathbb{Z}$.

The question is to describe $S_m[\xi_1, \dots, \xi_n]^{GL(m,F)}$. In classical language (as in Weyl's book [**72**]) one describes the invariants of a given representation by a *First and second fundamental theorem*

- The first fundamental theorem, *FFT* consists in describing a set of generators for the invariants.
- The second fundamental theorem, *SFT* consists in describing the relations between the given set of generators for the invariants.

In principle one may also try to study the *syzygies* but these are usually too hard to describe (but one has in this direction, at least one important result due to Lascoux [**37**] in characteristic 0).

Let us start from just one (generic) matrix $\xi = (x_{i,j}),\ i, j = 1, \dots m$. In the algebra $S[\xi] = F[x_{i,j}]$, of polynomial functions on $M_m(F)$, consider the characteristic polynomial of ξ:

$$\det(\lambda - \xi) = \lambda^m - \sigma_1(\xi)\lambda^{m-1} + \sigma_2(\xi)\lambda^{m-2} + \dots + (-1)^m \sigma_m(\xi). \tag{3}$$

Since $\det(\xi) = \det(g\xi g^{-1})$ the determinant (3) is an invariant function and then, also all its coefficients $\sigma_i(\xi) \in S[\xi]$ are invariant under conjugation.

Here $\sigma_1(\xi) = tr(\xi)$ is the *trace* and $\sigma_m(\xi) = \det(\xi)$ is the *determinant.* A first basic theorem is

Theorem 1.3. *The ring of polynomial invariants of a matrix ξ under conjugation is freely generated by the m functions $\sigma_i(\xi)$.*

$$S[\xi]^{GL(m)} = F[\sigma_1(\xi), \dots, \sigma_m(\xi)]. \tag{4}$$

In fact this is usually proved by a classical method which has far reaching generalizations (due to Chevalley), the *restriction Theorem to diagonal matrices.*

Inside $M_m(F)$ we have the diagonal matrices which we identify to F^m and inside $GL(m,F)$ we have the symmetric group S_m, of permutation matrices, which under conjugation induces the permutation action on F^m.

It follows that, if we restrict a G invariant function f on $M_m(F)$ to F^m, we obtain a symmetric function. If $x = diag(x_1, \dots, x_m)$ is a diagonal matrix with entries x_i its characteristic polynomial is given by formula (1), therefore the coefficients of the characteristic polynomial restrict (up to the sign) to the elementary symmetric functions $\sigma_i(x_1, \dots, x_m)$. When F is an infinite field we have

THEOREM 1.4. *If F is an infinite field, the restriction of the ring of G invariant polynomials on $M_m(F)$ to F^m, $S[\xi]^{GL(m)} \xrightarrow{\cong} F[x_1, \dots, x_m]^{S_m}$ is an isomorphism with the ring of symmetric polynomials $F[x_1, \dots, x_m]^{S_m} = F[\sigma_1(x), \dots, \sigma_m(x)]$.*

PROOF. The restriction map is surjective since the elementary symmetric functions generate the ring of symmetric functions. So it is enough to show that it is injective. Since the field is infinite we may assume that it is algebraically closed. Then consider the subset U of $M_m(F)$ of matrices with distinct eigenvalues, this is the affine open set where the discriminant is nonzero. The discriminant is a polynomial with integer coefficients in the coefficients of the characteristic polynomial.

If f is an invariant polynomial which restricted to the diagonal matrices is zero then it is also zero when restricted to U since every element of U is conjugate to a diagonal matrix. But U is Zariski open non empty, hence $f = 0$. $\square$

If $F = \mathbb{Q}$ the ring of symmetric functions is also generated by the *Newton functions*

$$\psi_k(x_1, \dots, x_m) := \sum_{i=1}^{m} x_i^k, \qquad \psi_k(\xi) = tr(\xi^k),\ k = 1, \dots, m.$$

There are *universal formulas* (cf. (21)) expressing the elements $\sigma_i(\xi)$ as polynomials *with rational coefficients* in the elements $tr(\xi^k)$, and conversely, the elements $tr(\xi^k)$ as polynomials *with integer coefficients* in the elements $\sigma_i(\xi)$.

REMARK 1.5. From the theory of symmetric functions it also follows that there are universal formulas expressing the elements $\sigma_i(\xi^j)$ as polynomials with integer coefficients in the elements $\sigma_h(\xi)$.

1.0.1. *The theorems over $\mathbb{Z}$ or over any field F for any number of $m \times m$ matrices.* We state now the Theorems which will be proved in this lecture note. Their proof is quite different in characteristic 0, where one can follow a very classical approach and in positive characteristic or over the integers where things are extremely more complicated.

Take a finite set $X = \{x_1, \dots, x_n\}$ which we call an *alphabet* (later on we are going to allow ourselves to even take X infinite countable). Denote by $F\langle X\rangle = F\langle x_1, \dots, x_n\rangle$ the free algebra with basis over F all the *words* in the alphabet X, also called *monomials.*

DEFINITION 1.6. We say that a monomial of positive degree (or a non empty word) w is *primitive* if it is not of the form $w = w_0^k,\ k > 1$.

Let W_p denote the set of primitive monomials.

Now consider cyclic equivalence of monomials that is

DEFINITION 1.7. We say that two monomials w_1, w_2 are *cyclically equivalent,* and we write $w_1 \stackrel{c}{\sim} w_2$ if $w_1 = ab,\ w_2 = ba$ for some monomials a, b.

LEMMA 1.8. *If w is primitive we have exactly $\ell(w)$ distinct monomials equivalent to w^k for all $k \geq 1$.*

If w_1, w_2 are both primitive and $w_1^{k_1} \stackrel{c}{\sim} w_2^{k_2}$ are cyclically equivalent we have $k_1 = k_2$ and $w_1 \stackrel{c}{\sim} w_2$ are cyclically equivalent.

PROOF. This is an easy exercise left to the reader. □

REMARK 1.9. In a class of cyclically equivalent primitive words there is a unique minimum word in lexicographic order, this is called a *Lyndon word* (cf. [**40**]).

Let W_0 denote a set of representatives of primitive monomials up to cyclic equivalence (for instance we can choose the Lyndon words).

We order W_0 by degree and then lexicographically.

By the universal property the free algebra can be evaluated in any associative algebra and in particular evaluates on $m \times m$ matrices ξ_i to matrix valued polynomial functions. Then, given any $f(x_1, \ldots, x_n) \in F\langle x_1, \ldots, x_n\rangle$, the coefficients $\sigma_i(f(\xi_1, \ldots, \xi_n))$ are clearly invariants. We then have:

THEOREM 1.10 (FFT). *The ring of invariants $S_m[\xi_1, \ldots, \xi_n]^{GL(m)}$ of n-tuples of matrices is generated by the elements $\sigma_i(M),\ i = 1, \ldots, m$ with M any* primitive monomial *in $\xi_1, \ldots, \xi_n$.*

If F has characteristic 0, the ring of invariants of n-tuples of matrices is also generated by the elements $tr(M)$ with M any monomial *in $\xi_1, \ldots, \xi_n$.*

This Theorem, in characteristic 0 is classical and appears in the papers of C. Procesi [**45**] and K. S. Sibirskii [**59**]. In positive characteristic it is due to S. Donkin [**20**], [**21**] as a consequence of a general theory which will be presented in §7 and of some identities to be presented in §14. The proof of this theorem in all characteristics is given in section 15.2.

REMARK 1.11. For each i and matrices A, B we have $\sigma_i(AB) = \sigma_i(BA)$ that is $\sigma_i(M)$ depends on M only up to cyclic equivalence (cf. Definition 1.7).

PROOF. In fact since this is a formal identity it is enough to prove it for A invertible and then $AB = A(BA)A^{-1}$ and the statement follows from the fact that σ_i is conjugation invariant. □

For the second fundamental theorem one needs a *substitutional calculus* on the polynomial ring in the infinitely many variables:

$$\mathbb{S} = F[\sigma_i(p)],\ i = 1, \ldots, \infty,\ p \in W_0.$$

Let us denote by $F_+\langle X\rangle$ the free algebra *without 1.* One defines (cf. Proposition 4.24) degree i polynomial maps $\sigma_i : F_+\langle X\rangle \to \mathbb{S} = F[\sigma_i(p)]$ which on a given primitive monomial p coincide with $\sigma_i(p)$.

THEOREM 1.12 (SFT). *The ring of invariants of n-tuples of $m \times m$ matrices is the quotient of the formal ring $\mathbb{S} = F[\sigma_i(p)],\ i = 1, \ldots, \infty,\ p \in W_0$ modulo the ideal generated by all values $\sigma_i(f),\ \forall i > m,\ \forall f \in F_+\langle X\rangle$ (here F is an infinite field).*

This Theorem, in characteristic 0 is due independently to C. Procesi [**45**] and Y. Razmyslov [**50**]. In positive characteristic it is due to A. Zubkov [**75**].

Since σ_i is a polynomial map of degree i, it gives rise under polarization to maps $\sigma_{i_1,\dots,i_k}(M_1,\dots,M_k)$, defined for all choices of indices $(i_1,\dots i_k)$ and $M_1,\dots M_k$, belonging to the natural basis W of monomials of $F_+\langle X\rangle$.

The Theorem also holds for F any field or $F=\mathbb{Z}$ taking the ideal generated by all values $\sigma_{i_1,\dots,i_k}(M_1,\dots,M_k)$, with $\sum_j i_j > m$.

Finally, Ziplies [**73**], [**74**] discovered a very compact and surprising formulation of the previous Theorems which he proved in characteristic 0 and to which Vaccarino [**69**] gave its final form. In these notes we show that the following Theorem is valid whenever F is any field or the integers:

THEOREM 1.13. (1) *The map* $\det\colon F\langle x_1,\dots,x_n\rangle \to S_m[\xi_1,\dots,\xi_n]^{GL(m)}$, $f(x_1,\dots,x_n)\mapsto \det(f(\xi_1,\dots,\xi_n))$ *is multiplicative.*
(2) *The induced map*

$$\det : (F\langle x_1,\dots,x_n\rangle^{\otimes m})^{S_m} \to S_m[\xi_1,\dots,\xi_n]^{GL(m)}$$

is surjective (FFT).
(3) *Its kernel is the ideal generated by all commutators (SFT).*

Finally one has a non commutative variation of the theorems. One considers the algebra $\mathcal{P}_{m,n}$ of $GL(m)$ *equivariant polynomial maps* $f : M_m(F)^n \to M_m(F)$:

$$f(g\xi_1 g^{-1},\dots,g\xi_n g^{-1}) = gf(\xi_1,\dots,\xi_n)g^{-1},\quad \forall g\in GL(m,F).$$

Then the coordinate maps $\xi_i : (\xi_1,\dots,\xi_n)\to\xi_i$ are among these maps as well as the invariants (since $F\cdot 1_m\subset M_m(F)$). One has

THEOREM 1.14. (1) *The non commutative algebra* $\mathcal{P}_{m,n}$ *of equivariant polynomial maps is generated by the coordinates* ξ_i *over the ring of invariants.*
(2) *Every relation in this algebra can be* deduced *from the Cayley–Hamilton identities*

$$CH_k(\xi) := \xi^k - \sigma_1(\xi)\xi^{k-1} + \sigma_2(\xi)\xi^{k-2} + \ldots + (-1)^k\sigma_k(\xi),\ \forall k\ge m.$$

(3) *In characteristic 0 it is enough to consider* $CH_m(\xi)$.
(4) *If* F *is finite, or* $\mathbb{Z}$, *one has to replace in 2) the polynomial* $CH_k(\xi)$ *with all its possible polarized forms* $CH_{h_1,\dots,h_r}(M_1,\dots,M_r)$, $\sum h_i\ge m$.

1.0.2. *The plan of these notes.* These notes are organized into 6 Parts. In the first we prove the Theorems in characteristic 0. In this case only basic methods of representation theory are used, in particular tensor symmetry and Schur–Weyl duality.

In Part 2 we introduce the notion of quasi hereditary algebras due to Cline, E. ; Parshall, B. ; Scott, L., [**9**], [**10**], [**11**], [**12**], [**13**] and explain some of the properties of their representations, in particular we introduce the notion of good filtration which plays a central role in our notes.

In Part 3 we study a particular but very important class of quasi hereditary algebras, that of Schur algebras, introduced by Green in [**27**]. This allows us to introduce the notion of representations of the group $GL(m,F)$ having a *good filtrations.* This is a special case of a more general theory which holds for arbitrary reductive algebraic groups, see the book of Donkin [**19**]. In particular the Theorem that the tensor product of two modules with good filtration, here proved in a combinatorial way 11.1, has a good filtration is due to Wang [**71**] in this case and the final result for general reductive groups to Olivier Mathieu [**42**] using the deep ideas of *Frobenius splittings* by Mehta and Ramanathan [**43**].

This is based on results of the cohomology of line bundles on the flag variety, which are beyond the purpose of these notes for which we refer to the book [**34**] by Jantzen.

As mentioned above we develop this theory, by combinatorial methods, only for the linear group, thus avoiding all the deep problems of algebraic geometry and homological algebra. This suffices for our purposes. The reader who is familiar with the work of Donkin and others on good filtrations may want to skip this part.

In Part 4 we develop the special tools needed to apply the theory of good filtrations to our situation and we prove the First Fundamental Theorem.

Part 5 is devoted to the work of Zubkov and the Second Fundamental Theorem.

Finally Part 6 develops the notion of the Schur algebras of a free algebra and then explains the theorem of Ziplies and Vaccarino in a characteristic free way.

Finally we should point out that, especially in dimension 3 and for the orthogonal group there is an extensive literature of matrix invariants in continuum physics and elasticity theory, mostly by Anthony James Merrill Spencer and Ronald Rivlin, see [**61**], [**62**], [**63**], [**64**], [**65**].