Invariant Theory with Applications

Jan Draisma and Dion Gijswijt

September 17 2009

Contents

Chapter 1

Lecture1: introducing invariant theory

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

1.1 Polynomial functions

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

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

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

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

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

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

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

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

1.2 Representations

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

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

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

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

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

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

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

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

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

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

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

1.3 Invariant functions

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

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

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

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

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

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

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

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

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

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

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

the dual representation is given by the same matrices:

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

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

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

1.4 Conjugacy classes of matrices

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

Let

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

□

1.5 Exercises

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

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

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

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

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

Chapter 2

Lecture2: Symmetric polynomials

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

2.1 Symmetric polynomials

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

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

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

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

Some obvious examples of symmetric polynomials are

$$s_1 := x_1 + x_2 + \cdots + x_n \text{ and} \tag{2.2}$$

$$s_2 := x_1x_2 + x_1x_3 + \cdots + x_1x_n + \cdots + x_{n-1}x_n \tag{2.3}$$

More generally, for every $k = 1, \ldots, n$, the *k-th elementary symmetric polyno-*

mial

$$s_k := \sum_{i_1<\ldots<i_k} x_{i_1}\cdots x_{i_k} \tag{2.4}$$

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

$$\prod_{i=1}^{n}(t-x_i) = x^n + \sum_{k=1}^{n}(-1)^k s_k t^{n-k}. \tag{2.5}$$

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

$$x_1^2+\cdots+x_n^2 = s_1^2 - 2s_2. \tag{2.6}$$

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

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

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

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

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

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

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

$$s_n^{\alpha_n} s_{n-1}^{\alpha_{n-1}-\alpha_n}\cdots s_1^{\alpha_1-\alpha_2}. \tag{2.7}$$

The leading monomial of this polynomial equals

$$(x_1 \cdots x_n)^{\alpha_n}(x_1 \cdots x_{n-1})^{\alpha_{n-1}-\alpha_n} \cdots x_1^{\alpha_1-\alpha_2}, \tag{2.8}$$

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

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

$$\mathrm{lm}(s_1^{\alpha_1} \cdots s_n^{\alpha_n}) = x_1^{\alpha_1+\cdots+\alpha_n} x_2^{\alpha_2+\cdots+\alpha_n} \cdots x_n^{\alpha_n}. \tag{2.9}$$

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

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

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

$$\pi(x_1, \ldots, x_n) = (s_1(x_1, \ldots, x_n), \ldots, s_n(x_1, \ldots, x_n)). \tag{2.10}$$

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

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

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

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

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

2.2 Counting real roots

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

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

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

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

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

$$A := \begin{pmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{pmatrix}. \tag{2.11}$$

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

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

$$p = \prod_{1\le i<j\le n} (x_j - x_i), \tag{2.12}$$

since the right-hand side divides p, and the two polynomials have the same degree and the same nonzero coefficient for $x_1^0 x_2^1 \cdots x_n^{n-1}$. □

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

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

$$\mathrm{Bez}(f) = \left(p_{i+j-2}(\alpha_1, \ldots, \alpha_n)\right)_{i,j=1}^{n}, \tag{2.13}$$

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

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

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

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

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

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

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

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

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

$$\mathrm{Bez}(f) = A^{\mathsf{T}}A = \tilde{A}^{\mathsf{T}} \operatorname{diag}(m_1, \ldots, m_{2k+r})\tilde{A}, \tag{2.14}$$

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

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

$$\mathrm{Bez}(f) = B^{\mathsf{T}}B - C^{\mathsf{T}}C + i(C^{\mathsf{T}}B + B^{\mathsf{T}}C) = B^{\mathsf{T}}B - C^{\mathsf{T}}C. \tag{2.15}$$

We have

$$\operatorname{rank}(B) \le r + k, \qquad \operatorname{rank}(C) \le k. \tag{2.16}$$

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

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

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

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

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

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

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

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

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

2.3 Exercises

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

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

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

Exercise 2.3.2. Prove the following formulas due to Newton:

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

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

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

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

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

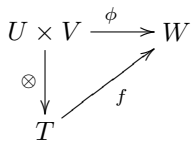
Chapter 3

Multilinear algebra

We review some constructions from linear algebra, in particular the tensor product of vector spaces. Unless explicitly stated otherwise, all our vector spaces are over the field $\mathbb{C}$ of complex numbers.

Definition 3.0.3. Let $V_1, \ldots, V_k, W$ be vector spaces. A map $\phi : V_1 \times \cdots \times V_k \to W$ is called *multilinear* (or *k-linear* or bilinear if $k = 2$ or trilinear if $k = 3$) if for each i and all $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k$ the map $V_i \to W,\ v_i \mapsto \phi(v_1, \ldots, v_k)$ is linear.

Let U, V and T be vector spaces and let $\otimes : U \times V \to T$ be a bilinear map. The map $\otimes$ is said to have the *universal property* if for every bilinear map $\phi : U \times V \to W$ there exists a *unique* linear map $f : T \to W$ such that $\phi = f \circ \otimes$.



We will usually write $u \otimes v := \otimes(u, v)$ for $(u, v) \in U \times V$. Although $\otimes$ will in general not be surjective, the image linearly spans T.

Exercise 3.0.4. Show that if $\otimes : U \times V \to T$ has the universal property, the vectors $u \otimes v, u \in U, v \in V$ span T.

Given U and V, the pair $(T, \otimes)$ is *unique up to a unique isomorphism*. That is, given two bilinear maps $\otimes : U \times V \to T$ and $\otimes' : U \times V \to T'$ that both have the universal property, there is a unique linear isomorphism $f : T \to T'$ such that $f(u \otimes v) = u \otimes' v$ for all $u \in U, v \in V$. This can be seen as follows. Since $\otimes'$ is bilinear, there exists by the universal property of $\otimes$, a unique linear map $f : T \to T'$ such that $\otimes' = f \circ \otimes$. It suffices to show that f is a bijection. By the universal property of $\otimes'$ there is a linear map $f' : T' \to T$ such that $\otimes' = f' \circ \otimes$. Now $\otimes \circ f' \circ f = \otimes$, which implies that $f' \circ f : T \to T$ is the identity since the

image of $\otimes$ spans T (or alternatively, by using the universal property of $\otimes$, and the bilinear map $\otimes$ itself). Hence f is injective. Similarly, $f \circ f'$ is the identity on T' and hence f is surjective.

Definition 3.0.5. Let U, V be vector spaces. The *tensor product* of U and V is a vector space T together with a bilinear map $\otimes : U \times V \to T$ having the universal property. The space T, which is uniquely determined by U and V up to an isomorphism, is denoted by $U \otimes V$.

Often we will refer to $U \otimes V$ as the tensor product of U and V, implicitly assuming the map $\otimes : U \times V \to U \otimes V$.

So far, we have not shown that the tensor product $U \otimes V$ exists at all, nor did we gain insight into the dimension of this space in terms of the dimensions of U and V. One possible construction of $U \otimes V$ is as follows.

Start with the vector space F (for *free* or *formal*) formally spanned by pairs (u, v) as u, v run through U, V, respectively. Now take the subspace R (for *relations*) of F spanned by all elements of the form

$$(c_1 u + u', c_2 v + v') - c_1 c_2 (u, v) - c_1 (u, v') - c_2 (u', v) - (u', v') \tag{3.1}$$

with $c_1, c_2 \in \mathbb{C},\ v, v' \in V, u, u' \in U$. Now any map $\phi : U \times V \to W$ factors through the injection $i : U \times V \to F$ and a unique linear map $g : F \to W$. The kernel of g contains R if and only if ϕ is bilinear, and in that case the map g factors through the quotient map $\pi : F \to F/R$ and a unique linear map $f : F/R \to W$. Taking for $\otimes$ the bilinear map $\pi \circ i : (u, v) \mapsto u \otimes v$, the space F/R together with the map $\otimes$ is the tensor product of U and V.

As for the dimension of $U \otimes V$, let $(u_i)_{i \in I}$ be a basis of U. Then by using bilinearity of the tensor product, every element $T \in U \otimes V$ can be written as a $T = \sum_{i \in I} u_i \otimes w_i$ with w_i non-zero for only finitely many i. We claim that the w_i in such an expression are unique. Indeed, for $k \in I$ let ξ_k be the linear function on U determined by $u_i \mapsto \delta_{ik},\ i \in I$. The bilinear map $U \times V \to V,\ (u, v) \to \xi_k(u) v$ factors, by the universal property, through a unique linear map $f : U \otimes V \to V$. This map sends all terms in the expression $\sum_{i \in I} u_i \otimes w_i$ for T to zero except the term with $i = k$, which is mapped to w_k. Hence $w_k = f_k(t)$ and this shows the uniqueness of the w_k.

Exercise 3.0.6. Use a similar argument to show that if $(v_j)_{j \in J}$ is a basis for V, then the set of all elements of the form $u_i \otimes v_j,\ i \in I,\ j \in J$ form a basis of $U \otimes V$.

This exercise may remind you of matrices. Indeed, there is a natural map ϕ from $U \otimes V^*$, where V^* is the dual of V, into the space $\mathrm{Hom}(V, U)$ of linear maps $V \to U$, defined as follows. Given a pair $u \in U$ and $f \in V^*$, $\phi(u \otimes f)$ is the linear map sending v to $f(v)u$. Here we are implicitly using the universal property: the linear map $v \mapsto f(v)u$ depends bilinearly on f and u, hence there is a unique linear map $U \otimes V^* \to \mathrm{Hom}(V, U)$ that sends $u \otimes f$ to $v \mapsto f(v)u$. Note that if f and u are both non-zero, then the image of $u \otimes f$ is a linear map of rank one.

Exercise 3.0.7. 1. Show that ϕ is injective. Hint: after choosing a basis $(u_i)_i$ use that a general element of $U \otimes V^*$ can be written in a unique way as $\sum_i u_i \otimes f_i$.

2. Show that ϕ is surjective onto the subspace of $\mathrm{Hom}(V, U)$ of linear maps of finite rank, that is, having finite-dimensional image.

Making things more concrete, if $U = \mathbb{C}^m$ and $V = \mathbb{C}^n$ and $u_1, \ldots, u_m$ is the standard basis of U and $v_1, \ldots, v_n$ is the standard basis of V with dual basis $x_1, \ldots, x_n$, then the tensor $u_i \otimes x_j$ corresponds to the linear map with matrix E_{ij}, the matrix having zeroes everywhere except for a 1 in position (i, j).

Remark 3.0.8. A common mistake is to assume that all elements of $U \otimes V$ are of the form $u \otimes v$. The above shows that in the finite-dimensional case the latter elements correspond to rank-one linear maps from V^* to U, or to rank-one matrices, while $U \otimes V$ consists of all finite-rank linear maps from V^* to U—a much larger set.

Next we discuss tensor products of linear maps. If $A : U \to U'$ and $B : V \to V'$ are linear maps, then the map $U \times V \to U' \otimes V',\ (u, v) \mapsto Au \otimes Bv$ is bilinear. Hence, by the universal property of $U \otimes V$ there exists a unique linear map $U \otimes V \to U' \otimes V'$ that sends $u \otimes v$ to $Au \otimes Bv$. This map is denoted $A \otimes B$.

Example 3.0.9. If $\dim U = m, \dim U' = m', \dim V = n, \dim V' = n'$ and if A, B are represented by an $m' \times m$-matrix $(a_{ij})_{ij}$ and an $n' \times n$-matrix $(b_{kl})_{kl}$, respectively, then $A \otimes B$ can be represented by an $m'n' \times mn$-matrix, with rows labelled by pairs (i, k) with $i \in [m'], k \in [n']$ and columns labelled by pairs (j, l) with $j \in [m], l \in [n]$, whose entry at position $((i, k), (j, l))$ is $a_{ij}b_{kl}$. This matrix is called the *Kronecker product* of the matrices $(a_{ij})_{ij}$ and $(b_{kl})_{kl}$.

Exercise 3.0.10. Assume, in the setting above, that $U = U', m' = m$ and $V = V', n' = n$ and A, B are diagonalisable with eigenvalues $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_n$, respectively. Determine the eigenvalues of $A \otimes B$.

Most of what we said about the tensor product of two vector spaces carries over *verbatim* to the tensor product $V_1 \otimes \cdots \otimes V_k$ of k. This tensor product can again be defined by a universal property involving k-linear maps, and its dimension is $\prod_i \dim V_i$. Its elements are called k-tensors. We skip the boring details, but do point out that for larger k there is no longer a close relationship with of k-tensors with linear maps—in particular, the *rank* of a k-tensor T, usually defined as the minimal number of terms in any expression of T as a sum of *pure tensors* $v_1 \otimes \cdots \otimes v_k$, is only poorly understood. For instance, *computing* the rank, which for $k = 2$ can be done using Gaussian elimination, is very hard in general. If all V_i are the same, say V, then we also write $V^{\otimes k}$ for $V \otimes \cdots \otimes V$ (k factors).

Given three vector spaces U, V, W, we now have several ways to take their tensor product: $(U \otimes V) \otimes W$, $U \otimes (V \otimes W)$ and $U \otimes V \otimes W$. Fortunately, these tensor products can be identified. For example, there is a unique linear

isomorphism $f : U \otimes V \otimes W \to (U \otimes V) \otimes W$ such that $f(u \otimes v \otimes w) = (u \otimes v) \otimes w$ for all $u \in U, v \in V, w \in W$.

Indeed, consider the trilinear map $U \times V \times W \to (U \otimes V) \otimes W$ defined by $(u, v, w) \mapsto (u \otimes v) \otimes w$. By the universal property, there is a unique linear map $f : U \otimes V \otimes W \to (U \otimes V) \otimes W$ such that $f(u \otimes v \otimes w) = (u \otimes v) \otimes w$ for all u, v, w.

Now for fixed $w \in W$, the bilinear map $\phi_w : U \times V \to U \otimes V \otimes W$ defined by $\phi_w(u, v) := u \otimes v \otimes w$ induces a linear map $g_w : U \otimes V \to U \otimes V \otimes W$ such that $u \otimes v$ is mapped to $u \otimes v \otimes w$. Hence the bilinear map $\phi : (U \otimes V) \times W \to U \otimes V \otimes W$ given by $\phi(x, w) := g_w(x)$ induces a linear map $g : (U \otimes V) \otimes W \to U \otimes V \otimes W$ sending $(u \otimes v) \otimes w$ to $u \otimes v \otimes w$. It follows that $f \circ g$ and $g \circ f$ are the identity on $(U \otimes V) \otimes W$ and $U \otimes V \otimes W$ respectively. This shows that f is an isomorphism.

Exercise 3.0.11. Let V be a vector space. Show that for all p, q there is a unique linear isomorphism $V^{\otimes p} \otimes V^{\otimes q} \to V^{\otimes (p+q)}$ sending $(v_1 \otimes \cdots \otimes v_p) \otimes (v_{p+1} \otimes \cdots \otimes v_{p+q})$ to $v_1 \otimes \cdots \otimes v_{p+q}$.

The direct sum $TV := \bigoplus_{k=0}^{\infty} V^{\otimes k}$ is called the *tensor algebra* of V, where the natural linear map $V^{\otimes k} \times V^{\otimes l} \to V^{\otimes k} \otimes V^{\otimes l} = V^{\otimes (k+l)}$ plays the role of (non-commutative but associative) multiplication. We move on to other types of tensors.

Definition 3.0.12. Let V be a vector space. A k-linear map $\omega : V^k \to W$ is called *symmetric* if $\omega(v_1, \ldots, v_k) = \omega(v_{\pi(1)}, \ldots, v_{\pi(k)})$ for all permutations $\pi \in \mathrm{Sym}(k)$.

The k-th *symmetric power* of V is a vector space $S^k V$ together with a symmetric k-linear map $V^k \to S^k V$, $(v_1, \ldots, v_k) \to v_1 \cdots v_k$ such that for all vector spaces W and symmetric k-linear maps $\psi : V^k \to W$ there is a unique linear map $\phi : S^k V \to W$ such that $\psi(u_1, \ldots, u_k) = \phi(u_1 \cdots u_k)$.

Uniqueness of the k-th symmetric power of V can be proved in exactly the same manner as uniqueness of the tensor product. For existence, let R be the subspace of $V^{\otimes k} := V \otimes \cdots \otimes V$ spanned by all elements of the form

$$v_1 \otimes \cdots \otimes v_k - v_{\pi(1)} \otimes \cdots \otimes v_{\pi(k)}, \ \pi \in \mathrm{Sym}(k).$$

Then the composition of the maps $V^k \to V^{\otimes k} \to V^{\otimes k}/R$ is a symmetric k-linear map and if $\psi : V^k \to W$ is any such map, then ψ factors through a linear map $V^{\otimes k} \to W$ since it is k-linear, which in turn factors through a unique linear map $V^{\otimes k}/R \to W$ since ψ is symmetric. This shows existence of symmetric powers, and, perhaps more importantly, the fact that they are quotients of tensor powers of V. This observation will be very important in proving the first fundamental theorem for $\mathrm{GL}(V)$.

There is also an analogue of the tensor product of maps: if A is a linear map $U \to V$, then the map $U^k \to S^k V$, $(u_1, \ldots, u_k) \mapsto Au_1 \cdots Au_k$ is multilinear and symmetric. Hence, by the universal property of symmetric powers, it factors through the map $U^k \to S^k U$ and a linear map $S^k U \to S^k V$. This map, which sends $u_1 \cdots u_k$ to $Au_1 \cdots Au_k$, is the k-th symmetric power $S^k A$ of A.

If $(v_i)_{i\in I}$ is a basis of V, then using multilinearity and symmetry every element t of S^kV can be written as a linear combination $\sum_\alpha c_\alpha v^\alpha$ of the elements $v^\alpha := \prod_{i\in I} v_i^{\alpha_i}$—the product order is immaterial—where $\alpha \in \mathbb{N}^I$ satisfies $|\alpha| := \sum_{i\in I}\alpha_i = k$ and only finitely many coefficients c_α are non-zero. We claim that the c_α are unique, so that the v^α, $|\alpha| = k$ a basis of V. Indeed, let $\alpha \in \mathbb{N}^I$ with $|\alpha| = k$. Then there is a unique k-linear map $\psi_\alpha : V^k \to \mathbb{C}$ which on a tuple $(v_{i_1}, \dots, v_{i_k})$ takes the value 1 if $|\{j \mid i_j = i\}| = \alpha_i$ for all $i \in I$ and zero otherwise. Moreover, ψ_α is symmetric and therefore induces a linear map $\phi_\alpha : S^kV \to \mathbb{C}$. We find that $c_\alpha = \phi_\alpha(t)$, which proves the claim.

This may remind you of polynomials. Indeed, if $V = \mathbb{C}^n$ and $x_1, \dots, x_n$ is the basis of V^* dual to the standard basis of V, then S^kV^* is just the space of *homogeneous polynomials* in the x_i of degree k. The algebra of all polynomial functions on V therefore is canonically isomorphic to $SV^* := \bigoplus_{k=0}^\infty S^kV^*$. The product of a homogeneous polynomials of degree k and homogeneous polynomials of degree l corresponds to the unique bilinear map $S^kV^* \times S^lV^* \to S^{k+l}V^*$ making the diagram

$$\begin{array}{ccc} (V^*)^{\otimes k} \times (V^*)^{\otimes l} & \longrightarrow & (V^*)^{\otimes k+l} \\ \downarrow & & \downarrow \\ S^kV^* \times S^lV^* & \dashrightarrow & S^{k+l}V^* \end{array}$$

commute, and this corresponds to multiplying polynomials of degrees k and l. Thus SV^* is a quotient of the tensor algebra TV (in fact, the maximal commutative quotient).

Above we have introduced S^kV as a *quotient* of $V^{\otimes k}$. This should not be confused with the *subspace* of $V^{\otimes k}$ spanned by all *symmetric tensors*, defined as follows. For every permutation $\pi \in S_k$ there is a natural map $V^k \to V^k$ sending $(v_1, \dots, v_k)$ to $(v_{\pi^{-1}(1)}, \dots, v_{\pi^{-1}(k)})$. Composing this map with the natural k-linear map $V^k \to V^{\otimes k}$ yields another k-linear map $V^k \to V^{\otimes k}$, and hence a linear map $V^{\otimes k} \to V^{\otimes k}$, also denoted π. A tensor ω in $V^{\otimes k}$ is called symmetric if $\pi\omega = \omega$ for all $\pi \in S_k$. The restriction of the map $V^{\otimes k} \to S^kV$ to the subspace of symmetric tensors is an isomorphism with inverse determined by $v_1 \cdots v_k \mapsto \frac{1}{k!}\sum_{\pi\in S_k} \pi(v_1 \otimes \cdots v_k)$. (Note that this inverse would not be defined in characteristic less than k.)

Exercise 3.0.13. Show that the subspace of symmetric tensors in $V^{\otimes k}$ is spanned by the tensors $v \otimes v \cdots \otimes v$, where $v \in V$.

3.1 Exercises

Exercise 3.1.1. Let $U \otimes V$ be the tensor product of the vector spaces U and V. Let $u_1, \dots, u_s$ and $u'_1, \dots, u'_t$ be two systems of linearly independent vectors in U and let $v_1, \dots, v_s$ and $v'_1, \dots, v'_t$ be two systems of linearly independent vectors in V. Suppose that

$$u_1 \otimes v_1 + \cdots + u_s \otimes v_s = u'_1 \otimes v'_1 + \cdots + u'_t \otimes v'_t. \tag{3.2}$$

Show that $s = t$.

Exercise 3.1.2. a) Let $T \in V_1 \otimes V_2 \otimes V_3$ be an element of the tensor product of V_1, V_2 and V_3. Suppose that there exist $v_1 \in V_1$, $v_3 \in V_3$, $T_{23} \in V_2 \otimes V_3$ and $T_{12} \in V_1 \otimes V_2$ such that

$$T = v_1 \otimes T_{23} = T_{12} \otimes v_3. \tag{3.3}$$

Show that there exist a $v_2 \in V_2$ such that $T = v_1 \otimes v_2 \otimes v_3$.

b) Suppose that $T \in V_1 \otimes V_2 \otimes V_3$ can be written as a sum of at most d_1 tensors of the form $v_1 \otimes T_{23}$, where $v_1 \in V_1, T_{23} \in V_2 \otimes V_3$, and also as a sum of at most d_3 tensors of the form $T_{12} \otimes v_3$, where $v_3 \in V_3, T_{12} \in V_1 \otimes V_2$. Show that T can be written as the sum of at most $d_1 d_3$ tensors of the form $v_1 \otimes v_2 \otimes v_3$, where $v_i \in V_i$.

Exercise 3.1.3. Let U, V, W be vector spaces. Denote by $B(U \times V, W)$ the linear space of bilinear maps from $U \times V$ to W. Show that the map $f \mapsto f \circ \otimes$ is a linear isomorphism between $\mathrm{Hom}(U \otimes V, W)$ and $B(U \times V, W)$.

Exercise 3.1.4. Let U, V be finite dimensional vector spaces. Show that the linear map $\phi : U^* \otimes V^* \to (U \otimes V)^*$ given by $\phi(f \otimes g)(u \otimes v) := f(u)g(v)$ is an isomorhism.

Chapter 4

Representations

Central objects in this course are linear representations of groups. We will only consider representations on complex vector spaces. Recall the following definition.

Definition 4.0.5. Let G be a group and let V be a vector space. A *(linear) representation of G on V* is a group homomorphism $\rho : G \to \mathrm{GL}(V)$.

If ρ is a representation of G, then the map $(g, v) \mapsto \rho(g)v$ is an action of G on V. A vector space with an action of G by linear maps is also called a *G-module*. Instead of $\rho(g)v$ we will often write gv.

Definition 4.0.6. Let U and V be G-modules. A linear map $\phi : U \to V$ is called a *G-module morphism* or a *G-linear map* if $\phi(gv) = g\phi(v)$ for all $v \in V$ and $g \in G$. If ϕ is invertible, then it is called an isomorphism of G-modules. The linear space of all G-linear maps from U to V is denoted $\mathrm{Hom}(U, V)^G$.

The multilinear algebra constructions from Section 3 carry over to representations. For instance, if $\rho : G \to \mathrm{GL}(U)$ and $\sigma : G \to \mathrm{GL}(V)$ are representations, then the map $\rho \otimes \sigma : G \to \mathrm{GL}(U \otimes V)$, $g \mapsto \rho(g) \otimes \sigma(g)$ is also a representation. Similarly, for any natural number k the map $g \mapsto S^k\rho(g)$ is a representation of G on S^kV. Also, the dual space V^* of all linear functions on V carries a natural G-module structure: for $f \in V^*$ and $g \in G$ we let gf be the linear function defined by $gf(v) = f(g^{-1}v)$. The inverse ensures that the action is a left action rather than a right action: for $g, h \in G$ and $v \in V$ we have

$$(g(hf))(v) = (hf)(g^{-1}v) = f(h^{-1}g^{-1}v) = f((gh)^{-1}v) = ((gh)f)(v),$$

so that $g(hf) = (hg)f$.

Exercise 4.0.7. Show that the set of fixed points in $\mathrm{Hom}(U, V)$ under the action of G is precisely $\mathrm{Hom}(U, V)^G$.

Example 4.0.8. Let V, U be G-modules. Then the space $\mathrm{Hom}(V, U)$ of linear maps $V \to U$ is a G module with the action $(g\phi)(v) := g\phi(g^{-1}v)$. The space

$U \otimes V^*$ is also a G-module with action determined by $g(u \otimes f) = (gu) \otimes (gf)$. The natural map $\Psi : U \otimes V^* \to \mathrm{Hom}(V, U)$ determined by $\Psi(u \otimes f)v = f(v)u$ is a morphism of G-modules. To check this it suffices to observe that

$$\Psi(g(u \otimes f))v = \Psi((gu) \otimes (gf))v = (gf)(v) \cdot gu = f(g^{-1}v) \cdot gu$$

and

$$(g\Psi(u \otimes f))v = g\Psi(u \otimes f)(g^{-1}v) = g(f(g^{-1}v)u) = f(g^{-1}v) \cdot gu.$$

The map Ψ is an G-module isomorphism of $U \otimes V^*$ with the space of finite-rank linear maps from V to U. In particular, if U or V is finite-dimensional, then Ψ is an isomorphism.

Example 4.0.9. Let G be a group acting on a set X. Consider the vectorspace

$$\mathbb{C}X := \{ \sum_{x \in X} \lambda_x x \mid \lambda_x \in \mathbb{C} \text{ for all } x \in X \text{ and } \lambda_x = 0 \text{ for almost all } x \} \quad (4.1)$$

consisting of all formal finite linear combinations of elements from X. The natural action of G given by $g(\sum_x \lambda_x x) := \sum_x \lambda_x gx$ makes $\mathbb{C}X$ into a G module. In the special case where $X = G$ and G acts on itself by multiplication on the left, the module $\mathbb{C}G$ is called the *regular representation of* G.

Definition 4.0.10. A *G-submodule* of a G-module V is a G-stable subspace, that is, a subspace U such that $gU \subseteq U$ for all $g \in G$. The *quotient* V/U then carries a natural structure of G-module, as well, given by $g(v + U) := (gv) + U$.

Definition 4.0.11. A G-module V is called *irreducible* if it has precisely two G-submodules (namely, 0 and V).

Exercise 4.0.12. Show that for finite groups G, every irreducible G-module has finite dimension.

Note that, just like 1 is not a prime number and the empty graph is not connected, the zero module is not irreducible. In this course we will be concerned only with G-modules that are either finite-dimensional or *locally finite*.

Definition 4.0.13. A G-module V is called *locally finite* if every $v \in V$ is contained in a finite-dimensional G-submodule of V.

Proposition 4.0.14. *For a locally finite G-module V the following statements are equivalent.*

1. *for every G-submodule U of V there is a G-submodule W of V such that $U \oplus W = V$;*
2. *V is a (potentially infinite) direct sum of finite-dimensional irreducible G-submodules.*

In this case we call V *completely reducible*; note that we include that condition that V be locally finite in this notion.

Proof. First assume (1). Let $\mathcal{M}$ be the collection of all finite-dimensional irreducible G-submodules of V. The collection of subsets S of $\mathcal{M}$ for which the sum $\sum_{U\in S} U$ is direct satisfies the condition of Zorn's Lemma: the union of any chain of such subsets S is again a subset of $\mathcal{M}$ whose sum is direct. Hence by Zorn's Lemma there exists a maximal subset S of $\mathcal{M}$ whose sum is direct. Let U be its (direct) sum, which is a G-submodule of V. By (1) U has a direct complement W, which is also a G-submodule. If W is non-zero, then it contains a non-zero finite-dimensional submodule (since it is locally finite), and for dimension reasons the latter contains an irreducible G-submodule W'. But then $S \cup \{W'\}$ is a subset of $\mathcal{M}$ whose sum is direct, contradicting maximality of S. Hence $W = 0$ and $V = U = \bigoplus_{M\in S} M$, which proves (2).

For the converse, assume (2) and write V as the direct sum $\bigoplus_{M\in S} M$ of irreducible finite-dimensional G-modules. Let U be any submodule of V. Then the collections of subsets of S whose sum intersects U only in 0 satisfies the condition of Zorn's Lemma. Hence there is a maximal such subset S'. Let W be its sum. We claim that $U + W = V$ (and the sum is direct by construction). Indeed, if not, then some element M of S is not contained in $U + W$. But then $M \cap (U+V) = \{0\}$ by irreducibility of M and therefore the sum of $S' \cup \{M\}$ still intersects U trivially, contradicting the maximality of S'. This proves (1). □

Remark 4.0.15. It is not hard to prove that direct sums, submodules, and quotients of locally finite G-modules are again locally finite, and that they are also completely reducible if the original modules were.

Example 4.0.16. A typical example of a module which is not completely reducible is the following. Let G be the group of invertible upper triangular 2×2-matrices, and let $V = \mathbb{C}^2$. Then the subspace spanned by the first standard basis vector e_1 is a G-submodule, but it does not have a direct complement that is G-stable.

Note that the group in this example is infinite. This is not a coincidence, as the following fundamental results show.

Proposition 4.0.17. *Let G be a finite group and let V be a finite-dimensional G-module. Then there exists a Hermitian inner product $(.|.)$ on V such that $(gu|gv) = (u|v)$ for all $g \in G$ and $u, v \in V$.*

Proof. Let $(.|.)'$ be any Hermitian inner product on V and take

$$(u|v) := \sum_{g\in G} (gu|gv)'.$$

Straightforward computations shows that $(.|.)$ is G-invariant, linear in its first argument, and semilinear in its second argument. For positive definiteness, note that for $v \neq 0$ the inner product $(v|v) = \sum_{g\in G}(gv|gv)$ is positive since every entry is positive. □

Theorem 4.0.18. *For a finite group G any G-module is completely reducible.*

Proof. Let V be a G-module. Then every $v \in V$ lies in the finite-dimensional subspace spanned by its *orbit* $Gv = \{gv \mid g \in G\}$, which moreover is G-stable. Hence V is locally finite. By Zorn's lemma there exists a submodule U of V which is maximal among all direct sums of finite-dimensional irreducible submodules of V. If U is not all of V, then let W be a finite-dimensional submodule of V not contained in U, and let $(.|.)$ be a G-invariant Hermitian form on W. Then $U \cap W$ is a G-submodule of W, and therefore so is the orthogonal complement $(U \cap W)^{\perp}$ of $U \cap W$ in W—indeed, one has $(gw|U \cap W) = (w|g^{-1}(U \cap W)) \subseteq (w|U \cap W) = \{0\}$ for $g \in G$ and $w \in (U \cap W)^{\perp}$, so that $gw \in (U \cap W)^{\perp}$. Let W' be an irreducible submodule of $(U \cap W)^{\perp}$. Then $U \oplus W'$ is a larger submodule of V which is the direct sum of irreducible submodules of V, a contradiction. Hence $V = U$ is completely reducible. □

4.1 Schur's lemma and isotypic decomposition

The following easy observation due to the German mathematician Issai Schur (1875-1941) is fundamental to representation and invariant theory.

Lemma 4.1.1 (Schur's Lemma)**.** *Let V and U be irreducible finite-dimensional G modules for some group G. Then either V and U are isomorphic and* $\mathrm{Hom}(V,U)^G$ *is one-dimensional, or they are not isomorphic and* $\mathrm{Hom}(V,U)^G = \{0\}$.

Proof. Suppose that $\mathrm{Hom}(V,U)^G$ contains a non-zero element ϕ. Then $\ker \phi$ is a G-submodule of V unequal to all of V and hence equal to $\{0\}$. Also, $\mathrm{im}\, \phi$ is a G-submodule of U unequal to $\{0\}$, hence equal to U. It follows that ϕ is an isomorphism of G-modules. Now suppose that ϕ' is a second element of $\mathrm{Hom}(V,U)^G$. Then $\psi := \phi' \circ \phi^{-1}$ is a G-morphism from U to itself; let $\lambda \in \mathbb{C}$ be an eigenvalue of it. Then $\psi - \lambda I$ is a G-morphism from U to itself, as well, and its kernel is a non-zero submodule, hence all of U. This shows that $\psi = \lambda I$ and therefore $\phi' = \lambda\phi$. Hence $\mathrm{Hom}(V,U)^G$ is one-dimensional, as claimed. □

If G is a group and V is a completely reducible G-module, then the decomposition of V as a direct sum of irreducible G-modules need not be unique. For instance, if V is the direct sum $U_1 \oplus U_2 \oplus U_3$ where the first two are isomorphic irreducible modules and the third is an irreducible module not isomorphic to the other two, then V can also be written as $U_1 \oplus \Delta \oplus U_3$, where $\Delta = \{u_1 + \phi(u_1) \mid u_1 \in U_1\}$ is the diagonal subspace of $U_1 \oplus U_2$ corresponding to an isomorphism ϕ from U_1 to U_2.

However, there is always a coarser decomposition of V into G-modules which *is* unique. For this, start with any decomposition of V as a direct sum of irreducible finite-dimensional submodules. Let $\{U_i\}_{i \in I}$ be a set of representatives of the isomorphism classes of G-modules that occur in this decomposition. Taking together isomorphic summands, we obtain a decomposition

$$V = \bigoplus_{i \in I} V_i, V_i = \bigoplus_{j \in J_i} V_{ij}, \tag{4.2}$$

where $V_{ij} \cong U_i$ for every $i \in I$ and $j \in J_i$. Now let W be any finite-dimensional irreducible submodule U of V and consider for any $i \in I, j \in I_i$ the projection π_{ij} on V_{ij} along the other components in the direct sum. That is, any $v \in V$ can be uniquely written as $v = \sum_{k \in I, l \in I_k} v_{kl}$ (where only finitely many summands are nonzero), and $\pi_{ij} v := v_{ij}$. It is easy to see that π_{ij} is a G-linear map and hence by Schur's lemma, we know that $\pi_{ij} U = 0$ when $U \ncong U_i$. It follows that $U = \sum_{i \in I, j \in I_i} \pi_{ij} U$ is contained in V_i, where $U \cong U_i$. In particular, U is isomorphic to some U_i.

We can conclude that V_i is the (non-direct) sum of all submodules isomorphic to U_i and hence does not depend on the initial decomposition of V into irreducible submodules. The space V_i is called the *isotypic component of V of type U_i*, and it has the following pretty description. The map $\mathrm{Hom}(U_i, V)^G \times U_i \to V,\ (\phi, u) \mapsto \phi(u)$ is bilinear, and therefore gives rise to a linear map $\Psi : \mathrm{Hom}(U_i, V)^G \otimes U_i \to V$. This linear map is a linear isomorphism onto V_i.

Exercise 4.1.2. Let U, V, W be G-modules. Show that $\mathrm{Hom}(U \oplus V, W)^G \cong \mathrm{Hom}(U, W) \oplus \mathrm{Hom}(V, W)$ and $\mathrm{Hom}(W, U \oplus V) \cong \mathrm{Hom}(W, U) \oplus \mathrm{Hom}(W, V)$.

Exercise 4.1.3. Show that the map $\Psi : \mathrm{Hom}(U_i, V)^G \otimes U_i \to V_i$ is a linear isomorphism.

4.2 Exercises

Exercise 4.2.1.

- Let V be a G-module and $\langle , \rangle$ a G-invariant inner product on V. Show that for any two non-isomorphic, irreducible submodules $V_1, V_2 \subset V$ we have $V_1 \perp V_2$, that is, $\langle v_1, v_2 \rangle = 0$ for all $v_1 \in V_1$, $v_2 \in V_2$.
- Give an example where $V_1 \not\perp V_2$ for (isomorphic) irreducible G-modules V_1 and V_2.

Exercise 4.2.2. Let the symmetric group on 3 letters S_3 act on $\mathbb{C}[x_1, x_2, x_3]_2$ by permuting the variables. This action makes $\mathbb{C}[x_1, x_2, x_3]_2$ into a S_3-module. Give a decomposition of this module into irreducible submodules.

Exercise 4.2.3. Let G be an abelian group. Show that every irreducible G-module has dimension 1. Show that G has a *faithful* irreducible representation if and only if G is cyclic. A representation ρ is called faithful if it is injective.

Exercise 4.2.4. Let G be a finite group and V an irreducible G-module. Show that there is a *unique* G-invariant inner product on V, unique up to multiplication by scalars.

Exercise 4.2.5. Let G be a finite group, and let $\mathbb{C}G$ be the regular representation of G and let $\mathbb{C}G = W_1^{m_1} \oplus \cdots \oplus W_k^{m_k}$ be the isotypic decomposition of $\mathbb{C}G$. Show that for every irreducible G-module W, there is an i such that W is isomorphic to W_i and show that $m_i = \dim W_i$. Hint: for all $w \in W$ the linear map $\mathbb{C}G \to W$ given by $\sum_g \lambda_g g \mapsto \sum_g \lambda_g g w$ is a G-linear map.