Multilinear Algebra and Applications (2MMD20, Fall 2015)

Jan Draisma

CHAPTER 1

Basics

We will assume familiarity with the terms *field*, *vector space*, *subspace*, *basis*, *dimension*, and *direct sums*. If you are not sure what these terms mean, please look them up in lecture notes of earlier (linear) algebra courses.

1.1. Products and sums

Given a collection $(V_i)_{i\in I}$ of vector spaces labelled by a (possibly infinite) set I, their *direct product* $\prod_{i\in I} V_i$ (or sometimes $\times_{i\in I} V_i$) is defined as the set of tuples $(v_i)_{i\in I}$ with $v_i \in V_i$. This product is a vector space with component-wise addition and scalar multiplication. In the special case where all V_i are equal to a fixed vector space V we also write V^I (or sometimes $V^{\times I}$).

On the other hand, the *direct sum* $\bigoplus_{i\in I} V_i$ is defined as the subspace of $\prod_{i\in I} V_i$ consisting of all tuples in which only finitely many of the v_i are non-zero. In the case where all V_i are equal to V we write $V^{\oplus I}$.

Clearly, if I is finite, then the two notions coincide, while if I is infinite, then the direct sum is a proper subspace of the direct product. An amusing and disturbing exercise exploiting this is Exercise 1.5.10 below.

1.2. Linear maps

Given K-vector spaces V and W, a map $\phi : V \to W$ is called *linear* (or *K-linear* if we want to stress K) if $\phi(v + v') = \phi(v) + \phi(v')$ for all $v, v' \in V$ and $\phi(cv) = c\phi(v)$ for all $c \in K$ and $v \in V$.

We write $L(V, W)$ (or $L_K(V, W)$) for the set of K-linear maps $V \to W$. We also write $L(V)$ (or $L_K(V)$) for $L(V, V)$. The set $L(V, W)$ is itself a vector space over K with addition defined by $(\phi + \psi)(v) := \phi(v) + \psi(v)$ and scalar multiplication defined by $(c\phi)(v) := c(\phi(v))$.

Any linear map $\phi : V \to W$ has an *image* $\operatorname{im}\phi := \{\phi v \mid v \in V\} \subseteq W$ and a *kernel* $\ker\phi := \{v \in V \mid \phi v = 0\} \subseteq V$. These sets are subspaces of W and V, respectively, and they satisfy

$$\dim \operatorname{im}\phi + \dim \ker\phi = \dim V;$$

the so-called *Dimension Theorem*; see also Section 1.5. The map ϕ is called *surjective* if $\operatorname{im}\phi = W$ and *injective* if $\phi v = \phi v'$ implies $v = v'$. Since ϕ is linear, this latter condition is equivalent to the condition that $\ker\phi = \{0\}$.

A linear map $\phi : V \to W$ is called a (linear) *isomorphism* if there is a linear map $\psi : W \to V$ such that $\phi \circ \psi$ ("ψ followed by ϕ" or "ϕ after ψ") is the identity map $W \to W$ and $\psi \circ \phi$ is the identity map $V \to V$. This map ψ is necessarily unique,

and denoted ϕ^{-1}. A linear map ϕ is an isomorphism if and only if it is surjective and injective. An isomorphism maps any basis of V to a basis of W, showing that $\dim V = \dim W$.

Suppose we are given a basis $(v_j)_{j\in J}$ of V, with J some potentially infinite index set. Then we have a linear map $L(V,W) \to W^J,\ \phi \mapsto (\phi v_j)_{j\in J}$. This map is itself an isomorphism—indeed, if ϕ is mapped to the all-zero vector in W^J, then ϕ is zero on a basis of V and hence identically zero. Hence the map $L(V,W) \to W^J$ just described is injective. On the other hand, given any vector $(w_j)_{j\in J} \in W^J$ there is a linear map $\phi : V \to W$ with $\phi v_j = w_j$. Indeed, one defines ϕ as follows: given v, write it as $\sum_{j\in J} c_j v_j$ with only finitely many of the c_j non-zero (this can be done, and in a unique way, since the v_j form a basis). Then set $\phi v := \sum_i c_j w_j$. Hence the map $L(V,W) \to W^J$ is surjective. We conclude that it is, indeed, an isomorphism. In particular, this means that the dimension of $L(V,W)$ is that of W^J.

1.3. Matrices and vectors

Given a basis $(v_j)_{j\in J}$ of a K-vector space V, we have a linear isomorphism $\beta : K^{\oplus J} \to V$ sending a J-tuple $(c_j)_{j\in J}$ with only finitely many non-zero entries to the linear combination $\sum_j c_j v_j$. Thus we may represent an element v of the abstract vector space V by means of a J-tuple $\beta^{-1}v$ of numbers in K.

Given a basis $(w_i)_{i\in I}$ of a second K-vector space W with corresponding isomorphism $\gamma : K^{\oplus I} \to W$ and given a linear map $\phi : V \to W$, we may represent ϕ by a *matrix* A with rows labelled by the elements of I and the columns labelled by the elements of J, and with (i,j)-entry equal to the j-th coordinate of $\gamma^{-1}(\phi v_i)$. Note that every column has only finitely many non-zero entries. Conversely, by the results of the previous section, every $I \times J$-matrix whose columns are elements of $K^{\oplus I}$ is the matrix of a unique linear map $V \to W$. The fundamental relation between ϕ and A is that applying ϕ on an abstract vector v boils down to performing matrix-vector multiplication of A with the vector $\beta^{-1}v$ representing v. More precisely, the diagram

$$\begin{array}{ccc} V & \xrightarrow{\ \phi\ } & W \\ {\scriptstyle\beta}\uparrow & & \uparrow{\scriptstyle\gamma} \\ K^{\oplus J} & \xrightarrow[x\mapsto Ax]{} & K^{\oplus I} \end{array}$$

commutes, i.e., the composition of linear maps along both paths from $K^{\oplus J}$ to W give the same linear map. Here Ax stands for the product of the $I \times J$-matrix A with the vector $x \in K^{\oplus J}$. Notice that since x has only finitely non-zero entries, this product is well-defined, and since the columns of A are elements of $K^{\oplus I}$, the product is again an element of $K^{\oplus I}$. When $I = \{1,\dots,n\}$ and $J = \{1,\dots,m\}$ for some natural numbers m, n, this boils down to ordinary matrix-vector multiplication.

Matrices and column or row vectors will be used whenever we implement linear-algebraic algorithms on a computer. For instance, an m-dimensional subspace of an n-dimensional vector space W can be described as the image of an injective linear map $\phi : K^m \to W$ (here m is necessarily at most n). After a choice of basis $w_1,\dots,w_n$ of W, ϕ can be represented as an $n \times m$-matrix A as above (where for $v_j,\ j \in J = \{1,\dots,m\}$ we take the standard basis of $V = K^m$). Hence the

columns of this matrix represent a basis of $\operatorname{im}\phi$ relative to the chosen basis of W. The matrix A will have rank m.

EXERCISE 1.3.1. In this setting, two $n \times m$-matrices A and B, both of rank m, represent the same linear subspace of W if and only if there exists an invertible $m \times m$-matrix g such that $Ag = B$; prove this.

EXERCISE 1.3.2. Write a function `Intersect` in `Mathematica` which takes as input two full-rank matrices A and B with n rows and at most n columns, and which outputs an $n \times k$-matrix C of rank k that represents the *intersection* of the subspaces represented by A and by B. Also write a function `Add` that computes a full-rank matrix representing the *sum* of the spaces represented by A and by B.

Alternatively, a codimension-n subspace of an m-dimensional vector space V can be represented as the kernel of a surjective map $\phi : V \to K^n$ (now m is necessarily at least n); here we use the Dimension Theorem. Choosing a basis $v_1, \ldots, v_m$ of V and the standard basis in K^m, we may represent ϕ by an $n \times m$-matrix.

EXERCISE 1.3.3. In this setting, two $n \times m$-matrices A and B, both of rank n, represent the same linear subspace of V if and only if there exists an invertible $n \times n$-matrix h such that $hA = B$; prove this.

1.4. The dual

Given a K-vector space V, the *dual* V^* of V is $L_K(V, K)$, i.e., the set of all K-linear functions $V \to K$.

By the discussion of $L(V, W)$ in Section 1.2, V^* is itself a K-vector space. Moreover, V^* has the same dimension as K^J, if J is (the index set of) some basis of V, while V itself has the same dimension as $K^{\oplus J}$. If, in particular, J is finite, i.e., if V is finite-dimensional, then V and V^* have the *same* dimension. In general, since $K^{\oplus J}$ is a subspace of K^J, we still have the inequality $\dim V \leq \dim V^*$.

EXERCISE 1.4.4. Assume that J is infinite but countable. (A typical example is $V = K[t] = \langle 1, t, t^2, \ldots \rangle_K$, the K-space of polynomials in a single variable t, with $J = \mathbb{N}$ and basis $(t^j)_{j \in \mathbb{N}}$.) Prove that V^* does not have a countable basis.

The following generalisation is trickier.

EXERCISE 1.4.5. Assume that J is infinite but not necessarily countable. Prove that V^* has no basis of the same cardinality as J.

Whether V is finite-dimensional or not, there is *no* natural linear map $V \to V^*$ that does not involve further choices (e.g. of a basis). However, there *is* always a natural map $V \to (V^*)^*$ that sends $v \in V$ to the linear function $V^* \to K$ that sends $x \in V^*$ to $x(v)$. This natural map is linear, and also injective: if v is mapped to zero, then this means that for all $x \in V^*$ we have $x(v) = 0$, and this means that v itself is zero (indeed, otherwise v would be part of some basis $(v_j)_{j \in J}$, say as v_{j_0}, and one could define a linear function x to be 1 on v_{j_0} and arbitrary on the remaining basis elements; then $x(v) \neq 0$).

The fact that $V \to (V^*)^*$ is injective implies that its image has dimension $\dim V$. If V is *finite-dimensional*, then by the above we have $\dim V = \dim V^* = \dim(V^*)^*$, so

that the map $V \to (V^*)^*$ is actually a linear isomorphism. Informally, we express this by saying that *for finite-dimensional vector spaces V we have $V = (V^*)^*$*. In such a statement we mean that we have a natural (or *canonical*) isomorphism in mind from one space to the other, i.e., an isomorphism that does not involve further arbitrary choices. (We are being deliberately vague here about the exact mathematical meaning of the terms *natural* or *canonical*.)

EXERCISE 1.4.6. Replacing V by V^* in the above construction, we obtain an injective linear map $\psi : V^* \to ((V^*)^*)^*$. Find a canonical *left* inverse to this map, that is, a linear map $\pi : ((V^*)^*)^* \to V^*$ such that $\pi \circ \psi$ is the identity map on V^*.

A linear map $\phi : V \to W$ gives rise to a natural linear map $\phi^* : W^* \to V^*$, called the *dual* of ϕ and mapping $y \in W^*$ to the linear function on V defined by $v \mapsto y(\phi v)$. More succinctly, we have $\phi^* y := y \circ \phi$. It is convenient to see this in a diagram: if y is a linear function on W, then $\phi^* y$ fits into the following diagram:

$$\begin{array}{ccc} V & \xrightarrow{\phi} & W \\ & \searrow^{\phi^* y} & \downarrow{\scriptstyle y} \\ & & K. \end{array}$$

Note that duality *reverses* the arrow: from a map $V \to W$ one obtains a map $W^* \to V^*$. The kernel of ϕ^* is the set of all linear functions on W that vanish identically on $\operatorname{im} \phi$. For the image of ϕ^* see Exercise 1.5.9.

If $(v_j)_{j \in J}$ is a basis of V and $(w_i)_{i \in I}$ is a basis of W, then we have seen in Section 1.3 how to associate an $I \times J$-matrix A with the linear map ϕ, each *column* of which has only finitely many non-zero elements. We claim that the *transpose* A^T, which is a $J \times I$-matrix each of whose *rows* has a finite number of non-zero elements, can be used to describe the dual map ϕ^*, as follows. There is a linear map $\beta^* : V^* \to K^J$ defined as $\beta^* x = (x(v_j))_{j \in J}$ (and dual to the map β from Section 1.3; check that K^I is the dual of $K^{\oplus I}$), and similarly a linear map $\gamma^* : W^* \to K^I$. Then the diagram

$$\begin{array}{ccc} V^* & \xleftarrow{\phi^*} & W^* \\ {\scriptstyle \beta^*}\downarrow & & \downarrow{\scriptstyle \gamma^*} \\ K^J & \xleftarrow[x \mapsto A^T x]{} & K^I \end{array}$$

commutes. Note that the product $A^T x$ is well-defined for every $x \in K^I$, as each row of A^T has only finitely many non-zero entries. In the special case where I and J are finite, we recover the familiar fact that "the matrix of the dual map with respect to the dual basis is the transpose of the original matrix". In the infinite case, however, note that A^T is *not* the matrix associated to ϕ^* with respect to *bases* of W^* and V^*: e.g., the linear forms $x_i \in W^*, i \in I$ determined by $x_i(w_{i'}) = \delta_{i,i'}$ do *not* span W^* as soon as W is infinite-dimensional.

1.5. Quotients

Given a K-vector space V and a subspace U, the *quotient* V/U of V by U is defined as the set of cosets ("affine translates") $v + U$ with v running over V. Note that

$v+U = v'+U$ if and only if $v - v' \in U$. The quotient comes with a surjective map $\pi : V \to V/U$, $\pi v := v+U$, and often it is less confusing to write πv instead of $v+U$. The quotient V/U is a K-vector space with operations $\pi(v) + \pi(v') := \pi(v + v')$ and $c\pi(v) := \pi(cv)$ (since the left-hand sides in these definitions do not depend on the choices of v and v' representing $v + V$ and $v' + V$, one needs to check that the right-hand sides do not depend on these choices, either). The map π is linear with respect to this vector space structure.

If U' is any vector space complement of U in V, i.e., if every element $v \in V$ can be written in a unique way as $u + u'$ with $u \in U$ and $u' \in U'$ (notation: $V = U \oplus U'$, the direct sum), then the restriction $\pi|_{U'}$ of π to U' is an isomorphism $U' \to V/U$. Hence $\dim(V/U)$ satisfies

$$\dim U + \dim(V/U) = \dim U + \dim U' = \dim V.$$

Here we have implicitly used that U *has* a vector space complement U'; this holds if one assumes the Axiom of Choice, as we do throughout.

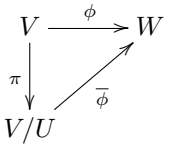
One application of this is, once again, the Dimension Theorem: If $\phi : V \to W$ is a linear map, then there is natural isomorphism $V/\ker\phi \to \operatorname{im}\phi$ sending $\pi(v)$ to $\phi(v)$. By the above with U replaced by $\ker\phi$ we find

$$\dim\ker\phi + \dim\operatorname{im}\phi = \dim\ker\phi + \dim(V/\ker\phi) = \dim V.$$

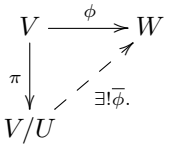
The following exercise gives a construction that we will use a number of times in this lecture.

Exercise 1.5.7. Let $\phi : V \to W$ be a linear map, and let U be a subspace of $\ker\phi$. Prove that there exists a unique linear map $\overline{\phi} : V/U \to W$ satisfying $\overline{\phi} \circ \pi = \phi$.

We say that ϕ *factorises* into π and $\overline{\phi}$, or that the diagram



commutes, i.e., that both paths from V to W yield the same linear map. The entire statement that ϕ factorises into π and a unique $\overline{\phi}$ is sometimes depicted as



Exercise 1.5.8. Let W be a subspace of V. Find a natural isomorphism between $(V/W)^*$ and the subspace W^0 of V^* consisting of linear functions that restrict to 0 on W (the *annihilator of* W), and prove that it is, indeed, an isomorphism.

The previous exercise shows that *the dual of a quotient is a subspace of the dual*, i.e., duality swaps the notions of quotient and subspace.

Exercise 1.5.9. Let $\phi : V \to W$ be a linear map. Find a natural isomorphism between the image $\operatorname{im}\phi^*$ of the dual of ϕ and $(V/\ker\phi)^*$, and prove that it is, indeed, an isomorphism.

Here is an amusing exercise, where the hint for the second part is that it appears in the section on quotients; which quotient is relevant here?

EXERCISE 1.5.10. A countable number of prisoners, labelled by the natural numbers $\mathbb{N}$, will have rational numbers tattooed on their foreheads tomorrow. Each can then see all other prisoners' tattoos, but not his own. Then, without any communication, they go back to their cells, where they individually guess the numbers on their own foreheads. Those who guess correctly are set free, the others have to remain in prison. Today the prisoners can agree on a strategy, formalised as a sequence of functions $g_i : \mathbb{Q}^{\mathbb{N}\setminus\{i\}} \to \mathbb{Q},\ i \in \mathbb{N}$ (one for each prisoner) describing their guess as a function of the tatooed numbers that they can see.

(1) Prove that there exists a strategy that guarantees that all but finitely many prisoners will be set free.
(2) Does there exist a strategy as in the previous part, where moreover all g_i are $\mathbb{Q}$-linear functions?
(3) We have implicitly assumed that prisoner i indeed "sees" an element of $\mathbb{Q}^{\mathbb{N}\setminus\{i\}}$, which requires that he can distinguish his fellow prisoners from one another. Does there exist a strategy when this is *not* the case?

Here is another amusing variant of this problem. The first part is standard (if you get stuck, ask your colleagues or look it up!), the second part is due to Maarten Derickx, a former Master's student from Leiden University.

EXERCISE 1.5.11. Tomorrow, the countably many prisoners will all be put in a queue, with prisoner 0 seeing all prisoners $1, 2, 3, 4, \ldots$ in front of him, prisoner 1 seeing all prisoners $2, 3, 4, \ldots$ in front of him, etc. Then each gets a black or white hat. Each sees the colours of all hats in front of him, but not his own colour or the colours behind. Then, each has to guess the colour of his hat: first 0 shouts his guess (black or white) so that all others can hear it, then 1, etc. Each prisoner hears all guesses behind him before it is his turn, but not whether the guesses were correct or not. Afterwards, those who guessed correctly are set free, and those who guessed incorrectly remain in prison. Today, the prisoners can decide on a strategy.

(1) For each finite natural number n find a strategy such that all prisoners will guess correctly except possibly the prisoners $0, n, 2n, \ldots$.
(2) Prove the existence of a strategy where all prisoners guess correctly except possibly prisoner 0.
(3) Generalise the previous exercise to other finite numbers of colours (fixed before the strategy meeting).

Given a linear map $\phi : V \to W$ and linear subspaces $V' \subseteq V$ and $W' \subseteq W$ such that ϕ maps V' into W', we have a unique induced linear map $\overline{\phi} : V/V' \to W/W'$ making the diagram

$$\begin{array}{ccc} V & \xrightarrow{\phi} & W \\ {\scriptstyle \pi_{V/V'}}\downarrow & & \downarrow{\scriptstyle \pi_{W/W'}} \\ V/V' & \xrightarrow[\overline{\phi}]{} & W/W' \end{array}$$

commutative (this is just Exercise 1.5.7 with W replaced by W/W' and ϕ replaced by $\pi_{W/W'} \circ \phi$). Suppose we have a basis $(v_j)_{j\in J}$ and a subset $J' \subseteq J$ such that

$(v_j)_{j\in J'}$ span (and hence form a basis of) V', and similarly a basis $(w_i)_{i\in I}$ of W with $I' \subseteq I$ indexing a basis of W'. Then the $\pi_{V/V'}v_j$ with j *not* in J form a basis of V/V', and the $\pi_{W/W'}w_i$ with $i \in I$ form a basis of W/W'. The matrix of $\overline{\phi}$ with respect to these latter bases is the sub-matrix of the matrix of ϕ with respect to the original bases, obtained by taking only the rows whose index is in $I \setminus I'$ and the columns whose index is in $J \setminus J'$. Schematically, the matrix of ϕ has the following block structure:

$$\begin{bmatrix} A_{I',J'} & A_{I',J\setminus J'} \\ 0 & A_{I\setminus I',J\setminus J'} \end{bmatrix},$$

where the 0 block reflects that V' is mapped into W' (and the matrix $A_{I',J'}$ is the matrix of the restriction $\phi|_{V'} : V' \to W'$), and where $A_{I\setminus I',J\setminus J'}$ is the matrix of $\overline{\phi}$.

CHAPTER 2

Group actions and linear maps

This chapter deals with group actions on $L(V, W)$ and on $L(V)$, where V and W are (finite-dimensional) vector spaces. The group action on $L(V)$ is by conjugation, and it will be further analysed in later chapters.

2.1. Actions and orbits

Let G be a group and let X be a set. An *action* of G on X is a map $G \times X \to X$, denoted $(g, x) \mapsto gx$, satisfying the axioms $1x = x$ and $g(hx) = (gh)x$ for all $x \in X$ and $g, h \in G$. Here 1 is the identity element of the group and gh is the product of g and h in the group. The *G-orbit* of $x \in X$ (or just orbit if G is fixed in the context) is $Gx := \{gx \mid g \in G\} \subseteq X$.

REMARK 2.1.12. The term *orbit* may be related to the special case where G is the group of rotations of three-space around the z-axis, and X is the unit sphere (the "earth") centered at the origin. Then orbits are trajectories of points on the earth's surface; they are all circles parallel to the (x, y)-plane, except the north pole and the south pole, which are *fixed points* of the action.

The *stabiliser* of $x \in X$ is $G_x := \{g \in G \mid gx = x\}$. An action gives rise (and is, in fact, equivalent) to a homomorphism $G \to \mathrm{Sym}(X)$, where $\mathrm{Sym}(X)$ is the group of all permutations of X, by means of $g \mapsto (x \mapsto gx)$; this homomorphism is called a *permutation representation* of G on X. If these notions are new to you, please look them up in textbooks or lecture notes for previous algebra courses!

The relation $x \sim y :\Leftrightarrow x \in Gy$ is an equivalence relation (if $x = gy$ then $y = g^{-1}x$ and if in addition $y = hz$ then $x = (gh)z$), hence the orbits $Gx,\ x \in X$ *partition* the set X. Often, in mathematics, *classifying* objects means describing the orbits of some group action in detail, while *normal forms* are representatives of the orbits.

2.2. Left and right multiplication

The most important example in this chapter is the case where $X = L(V, W)$ and $G = \mathrm{GL}(V) \times \mathrm{GL}(W)$. Here $\mathrm{GL}(V)$, the *General Linear group*, is the subset of $L(V)$ consisting of *invertible* linear maps, with multiplication equal to composition of linear maps. The action is defined as follows:

$$(g, h)\phi = h \circ \phi \circ g^{-1} (= h\phi g^{-1}).$$

To verify that this is an action, note first that $(1, 1)\phi = \phi$ (where the 1s stand for the identity maps on V and on W, respectively), and for the second axiom write

$$[(g, h)(g', h')]\phi = (gg', hh')\phi = hh'\phi(gg')^{-1} = h(h'\phi(g')^{-1})g^{-1} = (g, h)((g', h')\phi).$$

Check that things go wrong if we leave out the inverse in the definition.

Two maps $\phi, \psi \in L(V, W)$ are in the same orbit if and only if there exist linear maps $g \in \mathrm{GL}(V)$ and $h \in \mathrm{GL}(W)$ such that the following diagram commutes:

$$\begin{array}{ccc} V & \xrightarrow{\phi} & W \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle h} \\ V & \xrightarrow[\psi]{} & W; \end{array}$$

indeed, this is equivalent to $\psi = h\phi g^{-1}$. From the diagram it is clear that g must map $\ker\phi$ isomorphically onto $\ker\psi$, and also induce a linear isomorphism $V/\ker\phi \to V/\ker\psi$. Similarly, h must map $\operatorname{im}\phi$ isomorphically onto $\operatorname{im}\psi$ and induce a linear isomorphism $W/\operatorname{im}\phi \to W/\operatorname{im}\psi$.

Conversely, assume that $\dim\ker\phi = \dim\ker\psi$ and $\dim(V/\ker\phi) = \dim(V/\ker\psi)$ and $\dim(W/\operatorname{im}\phi) = \dim(W/\operatorname{im}\psi)$. Then we claim that ϕ and ψ are in the same orbit. Indeed, choose vector space complements V_1, V_2 of $\ker\phi$ and $\ker\psi$, respectively, and vector space complements W_1, W_2 of $\operatorname{im}\phi, \operatorname{im}\psi$, respectively. By the first two dimension assumptions, there exist linear isomorphisms $g' : \ker\phi \to \ker\psi$ and $g'' : V_1 \to V_2$, which together yield a linear isomorphism $g : V \to V$ mapping $\ker\phi$ onto $\ker\psi$. Then define $h' : \operatorname{im}\phi \to \operatorname{im}\psi$ by $h'(\phi(v_1)) := \psi(g''v_1)$ for all v_1. This is well-defined because ϕ maps V_1 isomorphically onto $\operatorname{im}\phi$. By the last dimension assumption there exists a linear isomorphism $h'' : W_1 \to W_2$, which together with h' gives a linear isomorphism $W \to W$ satisfying $h \circ \phi = \psi \circ g$, as required.

Note that, if V is finite-dimensional, then the assumption that $\dim\ker\phi = \dim\ker\psi$ implies the other two dimension assumptions. If, moreover, W is also finite-dimensional, then the construction just given shows that the rank of ϕ completely determines its orbit: it consists of all linear maps of the same rank. After choosing bases, this is equivalent to the fact that *for $n \times m$-matrices A, B there exist invertible square matrices g and h with $B = hAg^{-1}$ if and only if A and B have the same rank.*

Reformulating things entirely in the setting of (finite) matrices, we write $\mathrm{GL}_n = \mathrm{GL}_n(K)$ for the group of invertible $n \times n$-matrices with entries in K. Then $\mathrm{GL}_m \times \mathrm{GL}_n$ acts on the space $M_{n,m} = M_{n,m}(K)$ of $n \times m$-matrices by means of $(g, h)A = hAg^{-1}$, and the orbit of the $n \times m$-matrix

$$I_{n,m,k} := \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix},$$

where the $k \times k$-block in the upper left corner is an identity matrix, is the set of all rank-k matrices in $M_{n,m}$.

Exercise 2.2.13. Let A be the matrix

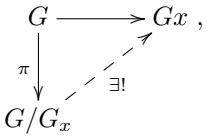
$$\begin{bmatrix} -21 & 3 & -57 & -84 & -117 \\ -3 & 21 & 39 & 63 & 93 \\ 27 & 51 & 199 & 308 & 443 \\ 69 & 93 & 423 & 651 & 933 \end{bmatrix} \in M_{4,5}(\mathbb{Q}).$$

(1) Determine the rank k of A.

(2) Determine invertible matrices g, h such that $hAg^{-1} = I_{4,5,2}$.

2.3. Orbits and stabilisers

A fundamental observation about general group actions is that for fixed $x \in X$ the map $G \mapsto Gx \subseteq X,\ g \mapsto gx$ factorises as follows:



where $\pi : G \to G/G_x$ is the projection $g \mapsto gG_x$ mapping g to the left coset gG_x and where the dashed map sends gG_x to gx—this is well-defined, since if $h = gg'$ with $g' \in G_x$, then $hx = (gg')x = g(g'x) = gx$ by the axioms. The dashed map is a bijection: it is surjective by definition of Gx and it is injective because $gx = hx$ implies that $x = g^{-1}(hx) = (g^{-1}h)x$ so that $g' := g^{-1}h$ lies in G_x and hence $h = gg' \in gG_x$.

In particular, if G is finite, then G/G_x and Gx have the same cardinality. Moreover, $|G/G_x|$ is the number of left cosets of G_x. As these partition G and all have the same cardinality $|G_x|$, we have $|G/G_x| = |G|/|G_x|$. Hence we find that

$$|G| = |G_x| \cdot |Gx|;$$

this fundamental equality that can be used to compute $|G|$ if you known $|G_x|$ and $|Gx|$, or $|Gx|$ if you know $|G|$ and $|G_x|$, etc.

2.4. Counting over finite fields

In this section we assume that $K = \mathbb{F}_q$, a field with q elements.

Exercise 2.4.14. The group $\mathrm{GL}_n(\mathbb{F}_q)$ is finite; prove that its order is $(q^n-1)(q^n-q)(q^n-q^2)\cdots(q^n-q^{n-1})$.

Exercise 2.4.15. Consider the action of $G = \mathrm{GL}_2(\mathbb{F}_q)$ on the set X of 1-dimensional linear subspaces of $\mathbb{F}_q^2$ defined by $gU := \{gu \mid u \in U\}$ for $g \in G$ and U a 1-dimensional linear subspace of $\mathbb{F}_q^2$.

(1) Show that X consists of a single orbit, and compute its cardinality.
(2) Show that the kernel of the corresponding permutation representation equals the centre Z consisting of all (non-zero) scalar matrices.
(3) Deduce that $\mathrm{GL}_2(\mathbb{F}_2)$ is isomorphic to the symmetric group on 3 letters.
(4) Deduce that the quotient of $\mathrm{GL}_2(\mathbb{F}_4)$ by its centre Z (of order 3) is isomorphic to the alternating group on 5 letters.

Note that the order can also be written as

$$q^{\binom{n}{2}}(q^n-1)(q^{n-1}-1)\cdots(q-1)$$

or as

$$q^{\binom{n}{2}}(q-1)^n[n]_q!,$$

where the *q-factorial* is defined as

$$[n]_q! := [n]_q[n-1]_q\cdots[1]_q$$

and the *q-bracket* $[a]_q$ is defined as $\frac{q^a-1}{q-1} = q^{a-1} + \cdots + q^1 + 1$.

Next we compute the order of the stabiliser in $\mathrm{GL}_n \times \mathrm{GL}_m$ of the matrix $I_{n,m,k}$; this is the group of tuples (g, h) such that $hI_{n,m,k} = I_{n,m,k}g$. Splitting into blocks:

$$h = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \text{ and } g = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$$

we find that

$$hI_{n,m,k} = \begin{bmatrix} h_{11} & 0 \\ h_{21} & 0 \end{bmatrix} \text{ and } I_{n,m,k}g = \begin{bmatrix} g_{11} & g_{12} \\ 0 & 0 \end{bmatrix}.$$

Hence it is necessary and sufficient that g_{12} and h_{21} both be zero, and that $g_{11} = h_{11}$. So take g_{21} an arbitrary element of $M_{m-k,k}$ and g_{11} an arbitrary element of GL_k and g_{22} an arbitrary element of GL_{m-k}; for this there are

$$q^{(m-k)k+\binom{k}{2}+\binom{m-k}{2}}(q-1)^{k+(m-k)}[k]_q![m-k]_q! = q^{\binom{m}{2}}(q-1)^m[k]_q![m-k]_q!$$

possibilities. Then $h_{11} = g_{11}$ is fixed, but for h_{12} and h_{22} there are still

$$q^{k(n-k)+\binom{n-k}{2}}(q-1)^{n-k}[n-k]_q! = q^{\binom{n}{2}-\binom{k}{2}}(q-1)^{n-k}[n-k]_q!$$

possibilities. Hence the number of matrices of rank equal to k equals

$$\frac{q^{\binom{m}{2}+\binom{n}{2}}(q-1)^{m+n}[m]_q![n]_q!}{q^{\binom{m}{2}+\binom{n}{2}-\binom{k}{2}}(q-1)^{m+n-k}[k]_q![m-k]_q![n-k]_q!} = q^{\binom{k}{2}}(q-1)^k\frac{[m]_q![n]_q!}{[m-k]_q![n-k]_q![k]_q!}.$$

EXERCISE 2.4.16. Compute the number of k-dimensional subspaces of $\mathbb{F}_q^n$. (Hint: these form a single orbit under $\mathrm{GL}_n(\mathbb{F}_q)$. Compute the order of the stabiliser in $\mathrm{GL}_n(\mathbb{F}_q)$ of the k-dimensional space spanned by the first k standard basis vectors.)

2.5. Invariants

Given an action of a group G on a set X, a function (or map) $f : X \to Y$ is called *invariant* if f is constant on orbits, or, equivalently, if $f(gx) = f(x)$ for all $x \in X$ and all $g \in G$. Here the co-domain Y can be anything: some finite set, a field, a vector space, an algebra, etc.

For example, we have seen that the function $M_{n,m}(K) \to \mathbb{N}$ that maps a matrix to its rank is an invariant under the action of $\mathrm{GL}_m \times \mathrm{GL}_n$ studied above. And in fact it is a *complete* invariant in the sense that it completely classifies the orbits.

2.6. Conjugation

In the coming weeks we will intensively study another group action, namely, the *conjugation* action of $\mathrm{GL}(V)$ on $L(V)$ defined by $gA := g \circ A \circ g^{-1}$. We actually never use the notation gA, because of potential confusion with our preferred short-hand gAg^{-1} for the right-hand side.

EXERCISE 2.6.17. Let $K = \mathbb{F}_q$ with q odd (so that $1 \neq -1$). Compute the cardinalities of of the orbits of the matrices

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \text{ and } \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

under the conjugation action of $\mathrm{GL}_2(\mathbb{F}_q)$.

Throughout the discussion of conjugation we will assume that V is a finite-dimensional vector space of dimension n. To warm up, here are a number of invariants for this action:

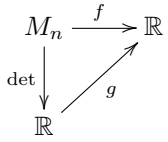
(1) rank: of course the rank of A equals that of gAg^{-1};
(2) determinant: we have $\det(gAg^{-1}) = (\det g)(\det A)(\det g)^{-1} = \det A$;
(3) trace: using $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for linear maps (or matrices) A, B, one finds that trace is invariant under conjugation;
(4) spectrum: this is the set of eigenvalues of A in some fixed algebraic closure $\overline{K}$ of K, and it is invariant under conjugation;
(5) characteristic polynomial: this is the degree-n polynomial $\det(tI - A) \in K[t]$, which remains the same when we conjugate A by g.

EXERCISE 2.6.18. Assume that $K = \mathbb{C}$. Using your knowledge of the Jordan normal form, argue that these invariants, even taken together, are not complete in the sense that they do not completely characterise orbits. Do so by giving two square matrices A, B of the same size with the property that the above invariants all coincide, but such that B is not conjugate to A.

EXERCISE 2.6.19. Let $P \in K[t]$ be any polynomial in one variable. Prove that the function $L(V) \to \{\text{yes,no}\}$ sending ϕ to the answer to the question "Is $P(\phi)$ the zero map?" is an invariant under conjugation. (Here $P(\phi)$ is defined by replacing the variable t in P by ϕ and interpreting powers of ϕ as repeated compositions of ϕ with itself.)

EXERCISE 2.6.20. Let SL_n (for *Special Linear group*) denote the subgroup of GL_n consisting of matrices of determinant 1. Let $\mathrm{SL}_n \times \mathrm{SL}_n$ act on M_n by left-and-right multiplication, i.e., $(g, h)A$ equals hAg^{-1}.

(1) Prove that $\det : M_n \to K$ is an invariant.
(2) Prove that det and rank together form a complete set of invariants.
(3) Assume that $K = \mathbb{R}$. Prove that for every *continuous* function $f : M_n \to \mathbb{R}$ invariant under the action there exists a continuous function $g : \mathbb{R} \to \mathbb{R}$ such that the diagram



commutes.

In the next two lectures, we will derive a complete set of invariants, and corresponding normal forms, for linear maps under conjugation. What makes these normal forms different from the Jordan normal form is that they are completely defined over the original field K, rather than over some extension field.

2.7. Symmetric polynomials

No discussion of invariants is complete without a discussion of *symmetric polynomials*. Let n be a natural number and let $G = \mathrm{Sym}(n)$ act on the polynomial ring $K[x_1, \ldots, x_n]$ in n variables by $\pi f := f(x_{\pi(1)}, \ldots, x_{\pi(n)})$, so that, for instance

$(1,2,3)(x_1^2x_2 - x_2x_3) = x_2^2x_3 - x_1x_3)$. A polynomial f is called *symmetric* if it is invariant under $\mathrm{Sym}(n)$, i.e., if $\pi f = f$ for all $\pi \in \mathrm{Sym}(n)$. In particular, the *elementary symmetric polynomials*

$$\begin{aligned} s_1 &:= x_1 + x_2 + \ldots + x_n \\ s_2 &:= x_1x_2 + x_1x_3 + \ldots + x_{n-1}x_n \\ &\vdots \\ s_k &:= \sum_{i_1<\ldots<i_k} x_{i_1} \cdots x_{i_k} \\ &\vdots \\ s_n &:= x_1x_2 \cdots x_n \end{aligned}$$

are symmetric polynomials. They are in fact a *complete set* of symmetric polynomials, in the following sense.

THEOREM 2.7.21. *If f is a symmetric polynomial in $x_1, \ldots, x_n$, then there exists a polynomial g in n variables $y_1, \ldots, y_n$ such that f is obtained from g by replacing y_i by s_i.*

In fact, the polynomial g is unique, but we will not need that. The relation of this theorem with the notion of invariants above is the following: $\mathrm{Sym}(n)$ acts by means of linear maps on K^n permuting the standard basis. Symmetric polynomials give rise to invariant polynomial functions $K^n \to K$, and the theorem states that all invariant polynomial functions can be expressed in the elementary symmetric polynomial functions. Here, strictly speaking, we have to assume that K is infinite, since if it is finite of order q then x_1^q, while clearly a different polynomial from x_1, defines the same polynomial function $K^n \to K$ as x_1 does.

PROOF. Define a linear order on monomials in the x_i as follows: $x_1^{a_1} \cdots x_n^{a_n}$ is larger than $x_1^{b_1} \cdots x_n^{b_n}$ if the first non-zero entry of $(a_1 - b_1, \ldots, a_n - b_n)$ is positive. Hence the largest monomial of s_k equals $x_1 \cdots x_k$. It is easy to see that this linear order is a *well-order*: there do not exist infinite, strictly decreasing sequences $m_1 > m_2 > \ldots$ of monomials.

The $\mathrm{Sym}(n)$-orbit of a monomial $x_1^{a_1} \cdots x_n^{a_n}$ consists of all monomials of the form $x_1^{a_{\pi(1)}} \cdots x_n^{a_{\pi(n)}}$. Among these, the monomial corresponding to permutations such that the sequence $a_{\pi(1)}, \ldots, a_{\pi(n)}$ decreases (weakly) is the largest monomial.

Now let $m = x_1^{a_1} \cdots x_n^{a_n}$ be the largest monomial with a non-zero coefficient in f. Since f is symmetric, all monomials in the $\mathrm{Sym}(n)$-orbit of m also have non-zero coefficients in f. Since m is the largest among these, we have $a_1 \geq \ldots \geq a_n$. Now consider the monomial $g_1 := y_n^{a_n} y_{n-1}^{a_{n-1}-a_n} \cdots y_1^{a_1-a_2}$. It is easy to see that the polynomial $g_1(s_1, \ldots, s_n)$ obtained by replacing y_k by s_k has the same largest monomial m as f. Hence subtracting a suitable scalar multiple we arrive at a polynomial

$$f_1 := f - cg_1(s_1, \ldots, s_n)$$

where the coeffient of m has become zero, so that the largest monomial is strictly smaller than m. By induction, this proves the theorem. □

The following exercise shows the link with invariants of matrices under conjugation.

EXERCISE 2.7.22. Take $K = \mathbb{C}$. Let $c_k : M_n(\mathbb{C}) \to \mathbb{C}$ be the i-th coefficient of the characteristic polynomial, i.e., we have

$$\det(tI - A) = t^n + c_1(A)t^{n-1} + \ldots + c_{n-1}(A)t + c_n(A),$$

so that, in particular, $c_n(A) = (-1)^n \det A$ and $c_1(A) = -\operatorname{tr}(A)$. Then the c_k are polynomial functions $M_n(\mathbb{C}) \to \mathbb{C}$ that are invariant under conjugation.

(1) Check that $c_k(A) = (-1)^k s_k(\lambda_1, \ldots, \lambda_n)$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues (listed with multiplicities) of A.

In the following parts, let $f : M_n(\mathbb{C}) \to \mathbb{C}$ be any polynomial function of the matrix entries that is invariant under conjugation.

(2) Prove that the restriction of f to the space of diagonal matrices, identified with $\mathbb{C}^n$, is a polynomial function of the elementary symmetric polynomials in those diagonal entries.
(3) Prove that f itself (on all of $M_n(\mathbb{C})$) is a polynomial in the functions $c_1, \ldots, c_n$. (Hint: you may use that polynomials are continuous functions, and that diagonalisable matrices are dense in $M_n(\mathbb{C})$.

CHAPTER 3

The minimal polynomial and nilpotent maps

In this chapter, we study two important invariants of linear maps under conjugation by invertible linear maps. These are the minimal polynomial on the one hand, which is defined for any linear map, and a combinatorial object called associated partition, which is defined only for so-called nilpotent linear maps.

3.1. Minimal polynomial

Throughout this chapter, V is a finite-dimensional vector space of dimension n over K, and ϕ is a linear map from V to itself. We write ϕ^0 for the identity map $I : V \to V$, ϕ^2 for $\phi \circ \phi$, ϕ^3 for $\phi \circ \phi \circ \phi$, etc. Let $p = p(t) \in K[t]$ be a polynomial in one variable t, namely $p = c_0 + c_1 t + \ldots + c_d t^d$. Then we define $p(\phi) := c_0 I + c_1 \phi + \ldots + c_d \phi^d$. We say that ϕ is a root of p if $p(\phi)$ is the zero map $V \to V$. A straightforward calculation shows that the map $K[t] \to L(V)$, $p \mapsto p(\phi)$ is an algebra homomorphism, i.e., it satisfies $(p+q)(\phi) = p(\phi) + q(\phi)$ and $(pq)(\phi) = p(\phi)q(\phi) = q(\phi)p(\phi)$, where the multiplication in $K[t]$ is multiplication of polynomials and the multiplication in $L(V)$ is composition of linear maps. Since $L(V)$ has finite dimension over K, the linear maps $I, \phi, \phi^2, \ldots$ cannot be all linearly independent. Hence the linear map $p \mapsto p(\phi)$ must have a non-zero kernel I. This kernel is an *ideal* in $K[t]$: if $q, r \in I$, which means that $q(\phi) = r(\phi) = 0$, then also $q + r \in I$ and $pq \in I$ for all $p \in K[t]$.

Let $p_{\min}$ be a non-zero polynomial of minimal degree in I that has leading coefficient equal to 1, i.e., which is *monic*. Every polynomial $p \in I$ is a multiple of $p_{\min}$. Indeed, by division with remainder we can write $p = qp_{\min} + r$ with $\deg r < \deg p_{\min}$. But then $r = p - qp_{\min}$ lies in I, and hence must be zero since $p_{\min}$ had the least degree among non-zero elements of I. Hence $p = qp_{\min}$. In particular, it follows that $p_{\min} \in I$ with the required properties (minimal degree and monic) is unique, and this is called the *minimal polynomial* of ϕ. The minimal polynomial of a square matrix is defined in exactly the same manner.

Exercise 3.1.23. Write a `Mathematica` function `MinPol` that takes as input a square integer matrix A and a number p that is either 0 or a prime, and that outputs the minimal polynomial of A over $\mathbb{Q}$ if $p = 0$ and the minimal polynomial of A over $K = \mathbb{Z}/p\mathbb{Z}$ if p is a prime.

Exercise 3.1.24. Show that $p_{\min}$ is an invariant of ϕ under the conjugation action of $\mathrm{GL}(V)$ on $L(V)$.

3.2. Chopping up space using $p_{\min}$

The minimal polynomial will guide us in finding the so-called *rational Jordan normal form* of ϕ. The first step is the following lemma.

LEMMA 3.2.25. *Suppose that ϕ is a root of $p \cdot q$, where p and q are coprime polynomials. Then V splits as a direct sum* $\ker p(\phi) \oplus \ker q(\phi)$.

Perhaps a warning is in order here: $(pq)(\phi) = 0$ does not imply that one of $p(\phi), q(\phi)$ must be zero.

PROOF. Write $1 = ap + bq$ for suitable $a, b \in K[t]$; these can be found using the extended Euclidean algorithm for polynomials. Then for $v \in V$ we have $v = a(\phi)p(\phi)v + b(\phi)q(\phi)v$. The first term lies in $\ker q(\phi)$, because

$$q(\phi)a(\phi)p(\phi)v = a(\phi)(p(\phi)q(\phi))v = 0$$

(while linear maps in general do not commute, polynomials evaluated in ϕ do—check this if you haven't already done so). Similarly, the second term $b(\phi)q(\phi)v$ lies in $\ker p(\phi)$. This shows that $\ker p(\phi) + \ker q(\phi) = V$. Finally, if v is in the intersection $\ker p(\phi) \cap \ker q(\phi)$, then the above shows that $v = 0 + 0 = 0$. Hence the sum in $\ker p(\phi) + \ker q(\phi)$ is direct, as claimed. □

EXERCISE 3.2.26. Prove that, in the setting of the lemma, $\operatorname{im} p(\phi) = \ker q(\phi)$ (and vice versa).

In the lemma, both subspaces $V_1 := \ker p(\phi)$ and $V_2 := \ker q(\phi)$ are *stable* under ϕ, which means that ϕ maps V_i into V_i for $i = 1, 2$. Indeed, if $v \in V_1$, then $p(\phi)\phi v = \phi p(\phi)v = 0$ so that $\phi v \in V_1$, as well.

Now factor $p_{\min}$ as $p_1^{m_1} \cdots p_r^{m_r}$ where the p_i are distinct, monic, irreducible polynomials in $K[t]$, and the exponents m_i are strictly positive. Write $q_i := p_i^{m_i}$. The polynomial q_1 is coprime with $q_2 \cdots q_r$. Hence we may apply the lemma and find

$$V = \ker q_1(\phi) \oplus \ker(q_2 \cdots q_r)(\phi).$$

Now let W be the second space on the right-hand side. This space is stable under ϕ, and the restriction $\phi|_W$ on it is a root of the polynomial $q_2 \cdots q_r$. Moreover, q_2 is coprime with $q_3 \cdots q_r$. Hence we may apply the lemma with V replaced by W, ϕ replaced by $\phi|_W$, p replaced by q_2 and q replaced by $q_3 \cdots q_r$ to find

$$W = \ker q_2(\phi|_W) \oplus \ker(q_3 \cdots q_r)(\phi|_W).$$

Now $\ker q_2(\phi) \subseteq V$ and $\ker(q_3 \cdots q_r)(\phi) \subseteq V$ are both contained in W, so that we may as well write

$$W = \ker q_2(\phi) \oplus \ker(q_3 \cdots q_r)(\phi),$$

and hence

$$V = \ker q_1(\phi) \oplus \ker q_2(\phi) \oplus \ker(q_3 \cdots q_r)(\phi).$$

Continuing in this fashion we find that

$$V = \bigoplus \ker q_i(\phi);$$

in what follows we write $V_i := \ker q_i(\phi)$ and $\phi_i : V_i \to V_i$ for the restriction of ϕ to V_i.

Now recall that we are trying to find invariants and normal forms for linear maps under conjugation. Suppose that we have a second element $\psi \in L(V)$ and want to find out whether it is in the same orbit as ϕ. First, we have already seen in Exercise 3.1.24 that both minimal polynomials must be the same. Moreover, if $\phi = g\psi g^{-1}$, then for each $i = 1, \ldots, r$, the linear map g maps $\ker q_i(\psi)$ isomorphically onto V_i. This proves that the dimensions of these spaces are invariants. Conversely, if these dimensions are the same for ϕ and for ψ, then one can choose arbitrary linear isomorphisms $h_i : \ker q_i(\psi) \to V_i$. These together give a linear isomorphism $h : V \to V$ with the property that for $\psi' := h\psi h^{-1}$ we have $\ker q_i(\psi') = V_i$. Write ψ_i' for the restriction of ψ' to V_i. Of course, ψ and ϕ are $\mathrm{GL}(V)$-conjugate if and only if ψ' and ϕ are, and this is true if and only if each ψ_i' is $\mathrm{GL}(V_i)$-conjugate to each ϕ_i. Indeed, if $\phi_i = g_i\psi_i' g_i^{-1}$, then the g_i together define an element g of $\mathrm{GL}(V)$ conjugating ψ' into ϕ. Conversely, a g that conjugates ψ' into ϕ necessarily leaves each V_i stable, and the restrictions g_i to the V_i conjugate ψ_i' into ϕ_i.

Now replace ϕ by a single ϕ_i, and V by V_i. The point of all this is that we have thus reduced the search for invariants and normal forms to the case where $\phi \in L(V)$ is a root of a polynomial p^m with p irreducible and monic. Then it follows that the minimal polynomial of ϕ itself must also be a power of p; assume that we have taken m minimal, so that $p_{\min} = p^m$. In the next chapter, we will split such a linear map ϕ into two commuting parts: a *semi-simple part* ϕ_s whose mimimal polynomial is p itself, and a *nilpotent part* ϕ_n. The remainder of the present chapter deals with the latter types of matrices.

3.3. The partition associated to a nilpotent map

Starting afresh, let V be a finite-dimensional vector space of dimension n, and let $\phi \in L(V)$ be a linear map such that $\phi^m = 0$ for some sufficiently large nonnegative integer m; such linear maps are called *nilpotent*. Note that any linear map conjugate to ϕ is then also nilpotent. Let m be the smallest natural number with $\phi^m = 0$, so that $p_{\min} = t^m$; m is also called the *nilpotency index* of ϕ.

Exercise 3.3.27. In the setting preceding this exercise, fix a non-zero vector $v \in V$ and let $p \geq 0$ be the largest exponent for which $\phi^p v$ is non-zero. Prove that $v, \phi v, \ldots, \phi^p v$ are linearly independent. Derive from this that m, the degree of $p_{\min}$, is at most n. If $p = n - 1$, give the matrix of ϕ with respect to the basis $v, \phi v, \ldots, \phi^{n-1} v$.

Write V_i for $\operatorname{im} \phi^i$; so for $i \geq m$ we have $V_i = \{0\}$. Then we have a chain of inclusions

$$V = V_0 \supseteq V_1 \supseteq \cdots \supseteq V_m = \{0\}.$$

Now note that $\phi(V_i) = V_{i+1}$, so that ϕ induces a surjective linear map $V_{i-1}/V_i \to V_i/V_{i+1}$. For $i \geq 1$ write $\lambda_i := \dim V_{i-1} - V_i$. By surjectivity and by the fact that $\phi^{m-1} \neq 0$ we have $\lambda_1 \geq \cdots \geq \lambda_m > 0 = \lambda_{m+1} = \ldots$. Moreover, by construction, we have $\sum_{i=1}^m \lambda_i = n = \dim V$. A non-increasing sequence $(\lambda_1, \ldots, \lambda_m)$ of positive numbers adding up to n is called a *partition* of n with *parts* $\lambda_1, \ldots, \lambda_m$. We call λ the partition *associated* to ϕ—though this is not completely standard terminology: often, the partition associated to ϕ is defined as the transpose of λ in the following sense.

$\lambda = (5, 5, 2, 1)$ $\qquad$ $\lambda' = (4, 3, 2, 2, 2)$

FIGURE 1. A Young diagram of a partition and its transpose.

DEFINITION 3.3.28. Given any partition $\lambda = (\lambda_1, \ldots, \lambda_m)$ of n, we define a second partition $\lambda' = (\lambda'_1, \ldots, \lambda'_p)$ of n by $p := \lambda_1$ and $\lambda'_k := |\{i \in \{1, \ldots, m\} \mid \lambda_i \geq k\}|$. The partition λ' is called the *transpose* of λ.

Why this is called the transpose is best illustrated in terms of *Young diagrams* depicting λ and λ'; see Figure 1.

EXERCISE 3.3.29. Prove that λ' is, indeed, a partition of n, and that $(\lambda')' = \lambda$.

A straightforward check shows that the associated partition is an invariant under conjugation. The following theorem states that it is a *complete* invariant for nilpotent linear maps.

THEOREM 3.3.30. *Two nilpotent linear maps $\phi, \psi \in L(V)$ are in the same orbit of the conjugation action of $\mathrm{GL}(V)$ if and only if the partitions of n as constructed above from ϕ and ψ are the same. Morever, every partition of n occurs in this manner.*

The proof of the first part of this theorem will be carried out in the next subsection. For the second part, it suffices to exhibit an $n \times n$-matrix giving rise to any partition $\lambda = (\lambda_1, \ldots, \lambda_m)$ of n.

Let $\lambda' = (\lambda'_1, \ldots, \lambda'_p)$ be the transpose of λ. Let A be the block-diagonal $n \times n$-block matrix

$$\begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_p \end{bmatrix} \text{ where } J_i = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix},$$

with λ'_i rows and columns (the block consists of a single zero if λ'_i equals 1). We claim that A gives rise the partition λ in the manner above. Indeed, J_i contributes 1 to $\dim \operatorname{im} A^{k-1} - \dim \operatorname{im} A^k$ if $\lambda'_i \geq k$, and zero otherwise. Hence $\dim \operatorname{im} A^{k-1} - \dim \operatorname{im} A^k$ is the number of indices i for which λ'_i is at least k. This gives the transpose of λ', which is λ itself again.

3.4. Completeness of the associated partition

Given a nilpotent linear map $\phi \in L(V)$ with associated partition $\lambda = (\lambda_1, \ldots, \lambda_m)$ whose transpose is $\lambda' = (\lambda'_1, \ldots, \lambda'_p)$, we will prove that there exists a basis of V with respect to which ϕ has the matrix A constructed above. This implies that any other nilpotent $\psi \in L(V)$ with the same associated partition λ is conjugate to ϕ.

We proceed as follows:

(1) First, choose a basis $z_1, \ldots, z_{\lambda_m}$ of V_{m-1}.
(2) Next, choose $y_1, \ldots, y_{\lambda_m} \in V_{m-2}$ such that $\phi y_i = z_i$. This is possible because $\phi : V_{m-2} \to V_{m-1}$ is surjective. Now $y_1 + V_{m-1}, \ldots, y_{\lambda_m} + V_{m-1}$ are linearly independent elements of V_{m-2}/V_{m-1}, because their ϕ-images in $V_{m-1}/V_m = V_{m-1}$ are linearly independent. Choose a basis $y_{\lambda_m+1} + V_{m-1}, \ldots, y_{\lambda_{m-1}} + V_{m-1} \in V_{m-2}/V_{m-1}$ of the kernel of the map $\overline{\phi} : V_{m-2}/V_{m-1} \to V_{m-1}/V_m = V_{m-1}$. This means that each $\phi(y_i) = 0$. The elements $y_1, \ldots, y_{\lambda_{m-1}}$ thus found represent a basis of V_{m-2}/V_{m-1}.
(3) The pattern is almost clear by now, but there is a catch in the next step: choose $x_1, \ldots, x_{\lambda_{m-1}} \in V_{m-3}$ such that $\phi x_i = y_i$, and further a basis $\tilde{x}_{\lambda_{m-1}+1} + V_{m-2}, \ldots, \tilde{x}_{\lambda_{m-2}} + V_{m-2} \in V_{m-3}/V_{m-2}$ of the kernel of induced map $\overline{\phi} : V_{m-3}/V_{m-2} \to V_{m-2}/V_{m-1}$. This means that each $\phi(\tilde{x}_i)$ is in V_{m-1}, i.e., is a linear combination of $z_1, \ldots, z_{\lambda_m}$. Hence, by subtracting the corresponding linear combinations of $y_1, \ldots, y_{\lambda_m}$ from $\tilde{x}_i$ we obtain a vector $x_i \in V_{m-3}$ that is in the kernel of ϕ. The vectors $x_1 + V_{m-2}, \ldots, x_{\lambda_{m-2}} + V_{m-2}$ are a basis of V_{m-3}/V_{m-2}.
(4) ... (in the next step, one chooses $\tilde{w}_{\lambda_{m-2}+1}, \ldots, \tilde{w}_{\lambda_{m-3}}$ that map to a basis of the kernel of $\overline{\phi} : V_{m-4}/V_{m-3} \to V_{m-3}/V_{m-2}$. This means that each $\phi(\tilde{w}_i)$ is a linear combination of $y_1, \ldots, y_{\lambda_{m-1}}, z_1, \ldots, z_{\lambda_m}$. Subtracting the corresponding linear combination of $x_1, \ldots, x_{\lambda_{m-1}}, y_1, \ldots, y_{\lambda_m}$, one finds vectors w_i properly in the kernel of ϕ, etc.) ... until:
(5) Choose $a_1, \ldots, a_{\lambda_2} \in V_0$ such that $\phi a_i = b_i$, and extend to a basis $a_1 + V_1, \ldots, a_{\lambda_1} + V_1$ of V_0/V_1 such that $\phi(a_i) = 0$ for $i > \lambda_2$.

By construction, $a_1, \ldots, a_{\lambda_1}, b_1, \ldots, b_{\lambda_2}, \ldots, z_1, \ldots, z_{\lambda_m}$ form a basis of V. Indeed, let $v \in V$. Then there is a unique linear combination a of $a_1, \ldots, a_{\lambda_1}$ such that $v - a \in V_1$, and a unique linear combination b of $b_1, \ldots, b_{\lambda_2}$ such that $v - a - b \in V_2$, etc. We conclude that v is a unique linear combination of the vectors above.

Moreover, each a_k generates a Jordan block as follows: ϕ maps $a_k \mapsto b_k \mapsto \ldots \mapsto 0$, where the zero appears when the corresponding λ_i is smaller than k, so that there are no basis vectors indexed k in that step. In other words, the number of Jordan blocks is equal to λ_1, and the k-th Jordan block has size equal to the number of i for which λ_i is $\geq k$. This is precisely the k-th part λ'_k of the transposed partition.

EXERCISE 3.4.31. Write a `Mathematica` program that takes as input a nilpotent matrix and computes the associated partition. Apply this program to the matrix

$$\begin{bmatrix} 1 & 1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 \\ 1 & 0 & -1 & -1 & 0 \\ 1 & 0 & -1 & -1 & 0 \end{bmatrix}.$$

EXERCISE 3.4.32. Let $n = 3$ and let $K = \mathbb{F}_q$ be a field with q elements.

(1) How many orbits does $\mathrm{GL}_3(K)$ have by conjugation on the set of nilpotent 3×3-matrices? For each orbit, give a representative matrix A as above.

(2) For each of these matrices A, compute the cardinality of the stabiliser $\{g \in \mathrm{GL}_3(K) \mid gAg^{-1} = A\}$ of A, and deduce the cardinality of these orbits.

As observed out by Jorn van der Pol in Fall 2011, the cardinalities of the orbits add up to q^6, which gives a check that you carried out the right computation in the second part of the exercise. This is, in fact, a general result.

THEOREM 3.4.33. *The number of nilpotent $n \times n$-matrices over $\mathbb{F}_q$ equals $q^{n(n-1)}$.*

Various proofs of this fact are known in the literature, but here is one found by Andries Brouwer and Aart Blokhuis after Jorn's observation, (also implicit in existing literature).

PROOF. Denote the number of nilpotent $n \times n$-matrices over $\mathbb{F}_q$ by a_n. We first derive a linear relation among $a_0, \ldots, a_n$, and then by a combinatorial argument show that the numbers $q^{k(k-1)}$ satisfy this relation. Since $a_0 = 1$ (the empty matrix is nilpotent) this proves the theorem.

To derive the relation among $a_0, \ldots, a_n$, let ϕ be a general $n \times n$-matrix, defining a linear map $K^n \to K^n$. Let d be the exponent of t in the minimal polynomial $p_{\min}$ of ϕ, and write $p_{\min} = t^d q$. By Lemma 3.2.25 we can write $V = V_0 \oplus V_1$ with $V_0 = \ker \phi^d$ and $V_1 = \ker q(\phi)$. The fact that q is not divisible by t means that it is of the form $c + tq_1$ with c a non-zero constant and q_1 a polynomial. This means that the restriction of ϕ to V_1 is invertible (with inverse $-\frac{1}{c}q_1(\phi)$), while of course the restriction of ϕ to V_0 is nilpotent. Thus to every linear map $\phi : K^n \to K^n$ we associate the four-tuple $(V_0, V_1, \phi|_{V_0}, \phi|_{V_1})$ with the third entry a nilpotent linear map and the fourth entry an invertible linear map. Conversely, splitting K^n into a direct sum $U_0 \oplus U_1$ and choosing any nilpotent linear map $\psi_0 : U_0 \to U_0$ and any invertible linear map $\psi_1 : U_1 \to U_1$ we obtain a "block-diagonal" linear map $\psi : K^n \to K^n$ which restricts to ψ_0 on U_0 and to ψ_1 on U_1. Thus we have found two ways to count linear maps $K^n \to K^n$: first, the straightforward way by counting matrix entries, which gives q^{n^2}, and second, the detour through splitting K^n into a direct sum $U_0 \oplus U_1$ where U_0 has dimension k, say, for which we denote the number of possibilities by b_k, then choosing a nilpotent linear map in $L(U_0)$ (a_k possibilities) and an invertible linear map in $L(U_1)$ (c_k possibilities). This gives the equality

$$q^{n^2} = \sum_{k=0}^{n} a_k b_k c_k.$$

In this expression we know c_k and b_k explicitly, namely,

$$c_k = |\mathrm{GL}_{n-k}|$$
$$b_k = \frac{|\mathrm{GL}_n|}{|\mathrm{GL}_k||\mathrm{GL}_{n-k}|},$$

where the expression for b_k is explained by choosing a basis of K^n and then letting U_0 be the span of the first k columns and U_1 be the span of the last k columns, and accordingly dividing by the number of bases of U_0 and U_1, respectively. Thus we obtain

$$b_k c_k = \frac{|\mathrm{GL}_n|}{|\mathrm{GL}_k|} = (q^n - 1) \cdots (q^{k+1} - 1) q^{\binom{n}{2} - \binom{k}{2}}.$$

Hence to prove that $a_k = q^{k(k-1)}$ it suffices to prove the equality

$$q^{n^2} = \sum_{k=0}^{n} (q^n - 1) \cdots (q^{k+1} - 1) q^{k(k-1)+\binom{n}{2}-\binom{k}{2}}.$$

For this, we start all over and enumerate general $n \times n$-matrices yet in a different manner, namely, as follows. If ϕ is a matrix, then let k be minimal such that the first $n-k$ columns of ϕ are linearly independent. So $k = 0$ if ϕ is invertible and $k = n$ if the first column of k is the zero column. Then the $(n-k+1)$st column (if existent) is a linear combination of the first $n-k$ columns, and the remaining columns are arbitrary. This yields the equality

$$q^{n^2} = \sum_{k=0}^{n} (q^n - 1)(q^n - q) \cdots (q^n - q^{n-k-1}) q^{n-k} q^{n(k-1)},$$

where the factor q^{n-k} counts the number of possibilities for the $(n-k+1)$st column. Note that the term with $k = 0$ is also valid because then the last two factors cancel. Now we claim that the k-th terms in the two found expressions for q^{n^2} coincide. Indeed, the k-th term in the latter is equal to

$$(q^n - 1)(q^{n-1} - 1) \cdots (q^{k+1} - 1) q^{\frac{1}{2}(n-k)(n-k-1)+(n-k)+n(k-1)}.$$

Finally, the exponents of q in both k-th terms are equal to $\binom{n}{2} + \binom{k}{2}$. This concludes the proof of the theorem. □

CHAPTER 4

Rational Jordan normal form

This chapter concludes our dicussion started in Chapter 3 of linear maps under the action of conjugation with invertible linear maps. We derive a complete set of invariants under a technical assumption on the field, namely, that it is *perfect*—this is needed from Section 4.3 on.

4.1. Semisimple linear maps

In Chapter 3 we have seen nilpotent linear maps and their orbits. In this section we will get acquainted with the other extreme: *semisimple* linear maps. In the sections following this one, we combine results on both types of maps to derive the *rational Jordan normal form*.

A polynomial $p \in K[t]$ is called *square-free* if it has no factor of the form q^2 with q a non-constant polynomial. Writing p as $p_1^{m_1} \cdots p_r^{m_r}$ where the p_i are distinct irreducible polynomials and the m_i are positive integers, p is square-free if and only if all m_i are equal to 1. The polynomial $f := p_1 \cdots p_r$ is then called the *square-free part* of f. Note that f divides p, while p in turn divides f^m where $m := \max_i m_i$.

Now let V be a finite-dimensional vector space. A linear map $\phi \in L(V)$ is called *semisimple* if there exists a square-free polynomial p such that $p(\phi) = 0$. As such a polynomial is automatically a scalar multiple of the minimal polynomial $p_{\min}$ of ϕ, semisimplicity of ϕ is equivalent to the requirement that $p_{\min}$ itself is square-free. Write $p_{\min} = p_1 \cdots p_r$ where the p_i are distinct and irreducible. In Chapter 3 we have seen that then $V = \bigoplus_{i=1}^r V_i$ where each $V_i := \ker p_i(\phi)$ is a ϕ-stable subspace of V, so if we choose bases in the subspaces V_i separately, then the matrix of ϕ becomes block-diagonal with r diagonal blocks, one for each V_i. Let us concentrate on these blocks now: the restriction ϕ_i of ϕ to V_i is a root of p_i, so if we now zoom in and write V for V_i and ϕ for ϕ_i and p for p_i, then the minimal polynomial of ϕ is the irreducible polynomial p. The following example gives fundamental examples of linear maps of this type.

Example 4.1.34. Let $W = K[t]/(p)$ be the quotient of $K[t]$ by the ideal consisting of all multiples of a monic polynomial p. Let $\psi = \psi_p : W \to W$ be the linear map defined by $\psi(f + (p)) := tf + (p)$. Then we claim that ψ has minimal polynomial p. Indeed, let $q = c_e t^e + c_{e-1} t^{e-1} + \ldots + c_0$ be any polynomial. Then we have

$$\begin{aligned}(q(\psi))(f + (p)) &= (c_e \psi^e + c_{e-1} \psi^{e-1} + \ldots + c_0 I)(f + (p)) \\ &= (c_e t^e + c_{e-1} t^{e-1} + \ldots + c_0) f + (p) \\ &= qf + (p).\end{aligned}$$

Hence ψ is a root of q if and only if qf is a multiple of p for every polynomial f, which happens if and only if q is a multiple of p (substitute $f = 1$ for the "only if" direction). Now write $p = t^d + b_{d-1}t^d + \cdots + b_0$. Then the matrix of ψ_p with respect to the monomial basis $1 + (p), t + (p), \ldots, t^{d-1} + (p)$ of W equals

$$C_p := \begin{bmatrix} 0 & & & & -b_0 \\ 1 & 0 & & & -b_1 \\ & \ddots & \ddots & & \vdots \\ & & 1 & 0 & -b_{d-2} \\ & & & 1 & -b_{d-1} \end{bmatrix},$$

called the *companion matrix* of p.

Returning to the discussion above, where ϕ has an irreducible polynomial p as minimal polynomial, we claim that we can choose a basis of V such that the matrix of ϕ becomes block-diagonal with diagonal blocks that are all copies of the companion matrix C_p. This means, in particular, that $\dim V$ must be a multiple ld of the degree d of p. Why would that be true? For an analogy, recall the argument why the cardinality of a finite field K is always a prime power a^l: it is a vector space of dimension l over its subfield consisting of $0, 1, 2, \ldots, a-1$ where a is the (necessarily prime) characteristic of K. Here the argument is similar: in addition to the vector space structure over K, we can give V a vector space structure over the field $F := K[t]/(p)$ containing K. Note that this is, indeed, a field because p is irreducible. We do not need to re-define addition in V, but we do need to define scalar multiplication with an element $f + (p)$ of F. This is easy: given $v \in V$ we set $(f + (p))v := f(\phi)v$. This is well-defined, since if we add a multiple bp to f, with $b \in K[t]$, then $(bp)(\phi)$ maps v to zero. One readily verifies all the axioms to show that this gives V the structure of an F-vector space. Let $v_1, \ldots, v_l$ be a basis of V as an F-vector space, and set $W_i := Fv_i$. This is a 1-dimensional F-vector space, and since F is a d-dimensional K-vector space, so is W_i. Since every element of V can be written as a unique F-linear combination of $v_1, \ldots, v_l$, we have the direct sum decomposition $V = W_1 \oplus \ldots \oplus W_l$ of V into K-subspaces. Each W_i is ϕ-stable, so if we choose bases of the W_i separately, then the matrix of ϕ becomes block-diagonal with l blocks of size $d \times d$ on the diagonal. Finally, when we take for W_i the basis $v_i, \phi v_i, \ldots, \phi^{d-1}v_i$, then these block matrices are exactly the companion matrices C_p, as claimed above.

Summarising all of this section, we have proved the following structure theorem for semisimple linear maps.

THEOREM 4.1.35. *Assume that $\phi \in L(V)$ is semisimple with minimal polynomial $p_1 \cdots p_r$ where the p_i are irreducible. Let d_i be the degree of p_i and set $l_i := \frac{1}{d_i} \dim \ker p_i(\phi)$. Then the l_i are natural numbers satisfying $l_1 d_1 + \ldots l_r d_r = \dim V$,*

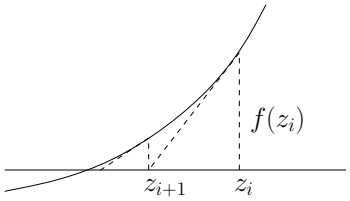


FIGURE 1. The Newton method for finding roots of a function.

and V has a basis with respect to which the matrix of ϕ is

$$\begin{bmatrix} C_{p_1} & & & & & \\ & \ddots & & & & \\ & & C_{p_1} & & & \\ & & & C_{p_2} & & \\ & & & & \ddots & \\ & & & & & C_{p_r} \end{bmatrix},$$

where the number of blocks containing the companion matrix C_{p_i} equals l_i.

4.2. Hensel lifting

Recall the Newton method for finding roots of a real-valued function (see Figure 1): if z_i is the current approximation of a root, then the next approximation is found by

$$z_{i+1} := z_i - \frac{f(z_i)}{f'(z_i)}$$

where f' denotes the derivative of f. Under suitable initial conditions, which in particular imply that the denominator $f'(z_i)$ never becomes zero, the sequence $z_1, z_2, z_3, \ldots$ converges to a nearby root of f. Remarkably, the same formula works beautifully in the following completely different set-up.

THEOREM 4.2.36 (Hensel lifting). *Let $f, q \in K[t]$ be polynomials for which the derivative f' is co-prime with q, and assume that $q|f$. Then for every $i = 1, 2, 3, \ldots$ there exists a polynomial $z_i \in K[t]$ such that $f(z_i) = 0 \mod q^i$ and $z_i = t \mod q$.*

Here $f(z)$ is the result of substituting the polynomial z for the variable t in f. So for instance, if $f = t^2$ and $z = t + 1$, then $f(z) = (t + 1)^2$. The property $f(z_i) = 0 \mod q^i$ expresses that $f(z_i)$ gets "closer and closer to zero" as $i \to \infty$, while $z_i = t \mod q$ expresses that the approximations of the root thus found are "nearby" the first approximation t.

PROOF. We will use the Taylor expansion formula:

$$f(t + s) = f(t) + sf'(t) + s^2 g(t, s)$$

for some two-variable polynomial $g \in K[t, s]$—check that you understand this! Define $z_1 := t$, so that $f(z_1) = f(t) = 0 \mod q$ since $q|f$. Assume that we have found z_i with the required properties. Since f' and q are co-prime, by the extended Euclidean algorithm one finds polynomials a and b such that $af' + bq = 1$. This

means that $af' = 1 \mod q$, which leads us to let a play the role of $1/f'$ in the formula above. We therefore define, for $i \geq 1$,

$$z_{i+1} := z_i - a(z_i)f(z_i),$$

and we note that $z_{i+1} = z_i \mod q^i$, so that certainly $z_{i+1} = z_1 \mod q = t \mod q$. Next we compute

$$\begin{aligned} f(z_{i+1}) &= f(z_i - a(z_i)f(z_i)) \\ &= f(z_i) - [a(z_i)f(z_i)]f'(z_i) + [a(z_i)f(z_i)]^2 g(z_i, a(z_i)f(z_i)) \\ &= f(z_i) - [1 - b(z_i)q(z_i)]f(z_i) \mod q^{i+1} \\ &= b(z_i)q(z_i)f(z_i), \end{aligned}$$

where the third equality follows from $f(z_i)^2 = 0 \mod q^{2i}$ and $2i \geq i+1$. Now use that $z_i = t \mod q$, which implies $q(z_i) = q(t) \mod q$, or in other words q divides $q(z_i)$. Since q^i divides $f(z_i)$, we find that q^{i+1} divides the last expression, and we are done. □

4.3. Jordan decomposition

We will use Hensel lifting to split a general linear map into a semisimple and nilpotent map. However, we will do this under an additional assumption on our ground field K, namely, that it is *perfect*. This means that either the characteristic of K is zero, in which case there is no further restriction, or the characteristic of K is a prime a, in which we impose the condition that every element of K is an a-th power. Every finite field K of characteristic a is perfect, because taking the a-th power is in fact an $\mathbb{F}_a$-linear map $K \to K$ with trivial kernel, so that it is a bijection and in particular surjective. A typical example of a non-perfect field is the field $\mathbb{F}_a(s)$ of rational functions in a variable s. Indeed, the element s in this field has no p-th root (but of course it does in some extension field).

We will use the following characterisation of square-free polynomials over perfect fields.

Exercise 4.3.37. Assuming that K is perfect, prove that $p \in K[t]$ is square-free if and only if it is coprime with its derivative p'. (Hint: use the decomposition of p into $p_1^{m_1} \cdots p_r^{m_r}$ with irreducible and coprime p_i, and apply the Leibniz rule. If the characteristic of K is zero, then the derivative of a non-constant polynomial is never zero, and you may use this fact. However, if the characteristic of K is a positive prime number a, then derivatives of non-constant polynomials *are* zero if all exponents of t appearing in it are multiples of a. For this case, use the fact that K is perfect.)

We can now state the main result of this section.

Theorem 4.3.38 (Jordan decomposition). *Let V be a finite-dimensional vector space over a perfect field K and let $\phi \in L(V)$. Then there exist unique $\phi_n, \phi_s \in L(V)$ such that*

(1) $\phi = \phi_s + \phi_n$,
(2) ϕ_n *is nilpotent,*
(3) ϕ_s *is semisimple, and*

(4) $\phi_s\phi_n = \phi_n\phi_s$ *(i.e., they commute).*

The proof will in fact show that ϕ_s can be taken equal to $z(\phi)$ for a suitable polynomial $z \in K[t]$. Being a polynomial in ϕ, ϕ_s then automatically commutes with ϕ_s and with $\phi - \phi_s = \phi_n$.

PROOF. Let $p_{\min}$ be the minimal polynomial of ϕ and let f be its square-free part. Then f and f' are coprime by Exercise 4.3.37; this is where we use perfectness of K. We may therefore apply Hensel lifting with q equal to f. Hence for each $i \geq 1$ there exists a polynomial $z_i \in K[t]$ with $f(z_i) = 0 \mod f^i$ and $z_i = t \mod f$. Take i large enough, so that f^i is divisible by $p_{\min}$. Then $f^i(\phi)$ is the zero map, hence so is $f(z_i(\phi))$. Set $\phi_s := z_i(\phi)$. Then we have $f(\phi_s) = 0$, and since f is square-free, ϕ_s is semisimple. Next, since $t - z_i$ is a multiple of f, we find that $(t - z_i)^i$ is a multiple of f^i, and therefore $(\phi - \phi_s)^i$, which is obtained from $(t - z_i(t))^i$ by substituting ϕ for t, is the zero map. Hence $\phi_n := \phi - \phi_s$ is nilpotent. As remarked earlier, ϕ_s is a polynomial in ϕ and hence commutes with ϕ and with ϕ_n. This proves the existence part of the theorem. The uniqueness is the content of the following exercise. □

EXERCISE 4.3.39. Assume that $\phi = \psi_s + \psi_n$ where ψ_s is semisimple and ψ_n is nilpotent, and where $\psi_n\psi_s = \psi_s\psi_n$.

(1) Prove that ψ_s and ψ_n both commute with both ϕ_s and ϕ_n constructed in the proof above.
(2) Prove that $\phi_s - \psi_s$, which equals $\psi_n - \phi_n$, is nilpotent.

It turns out that the sum of *commuting* semisimple linear maps is again semisimple; we will get back to that in a later chapter.

(3) Prove that $-\phi_s$ is semisimple, and conclude that so is $\phi_s - \psi_s$.
(4) From the fact that $\phi_s - \psi_s$ is both semisimple and nilpotent, deduce that it is zero.

This proves the uniqueness of the Jordan decomposition.

4.4. Rational Jordan normal form

We can now prove the main result of this chapter.

THEOREM 4.4.40. *Let V be a finite-dimensional vector space over a perfect field K and let $\phi \in L(V)$. Write $p_{min} = p_1^{m_1} \cdots p_r^{m_r}$ for the minimal polynomial of ϕ, where the p_i are distinct, monic, irreducible polynomials and where the m_i are positive integers. For each $i = 1, \ldots, r$ write $V_i := \ker(p_i(\phi)^{m_i})$ and write ϕ_i for the restriction of ϕ to V_i. Let $(\phi_i)_s + (\phi_i)_n$ be the Jordan decomposition of ϕ. Let $\lambda^{(i)}$ denote the partition of $\dim V_i$ associated to the nilponent map $(\phi_i)_n$. Then the map that sends ϕ to the unordered list of pairs $(p_i, \lambda^{(i)})$ is a complete invariant for the action of $\mathrm{GL}(V)$ on $L(V)$ by conjugation.*

We have already convinced ourselves that this map is, indeed, an invariant. To prove completeness, we have to show that we can find a basis of V with respect to which the matrix of ϕ "depends only on those pairs". We do this by zooming in on a single V_i, which we call V to avoid indices. Similarly, we write ϕ for ϕ_i, p, m for p_i, m_i, as well as ϕ_s, ϕ_n, λ for their i-indexed counterparts.

Recall from the proof of the Jordan decomposition that ϕ_s is a root of p, the square-free part of $p_{\min} = p^m$. Thus we may give V the structure of a vector space over $F := K[t]/(p)$, with multiplication given by $(f + (p))v := f(\phi_s)v$, as in the discussion in Section 4.1. The K-linear nilpotent map ϕ_n is even F-linear, since it commutes with ϕ_s. Let λ_F be the partition of $\dim_F V$ associated to ϕ_n, with $\lambda'_F = (a_1, \ldots, a_s)$ its dual. Then our structure theory for nilpotent maps shows that there exists an F-basis $v_1, \ldots, v_l$ of V with respect to which ϕ_n is block-diagonal with block sizes $a_1, \ldots, a_s$, where the j-th block equals

$$\begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix},$$

with a_j rows and columns. Now in each one-dimensional F-space Fv_i choose the K-basis $v_i, \phi_s v_i, \ldots, \phi_s^{d-1} v_i$, where d is the degree of p. With respect to the resulting

K-basis of V, the matrix of ϕ_n becomes block-diagonal with the same shape, except that the zeroes and ones in the matrix above now stand for $d \times d$-zero matrices and $d \times d$-identity matrices, respectively. Moreover, the matrix of ϕ_s relative to the resulting K-basis is block-diagonal with all diagonal blocks equal to the $d \times d$-companion matrix C_p. We conclude that the matrix of ϕ with respect to the K-basis thus found is block-diagonal with block sizes $a_1, \ldots, a_s$, where the j-th block equals

$$\begin{bmatrix} C_p & I & & & \\ & C_p & I & & \\ & & \ddots & \ddots & \\ & & & C_p & I \\ & & & & C_p \end{bmatrix},$$

with $a_j \cdot d$ rows and columns. We have thus found the required normal form: zooming out again to the setting in the theorem, we see that we can find a matrix for our original, arbitrary linear map ϕ, with diagonal blocks of the above type, where p runs from p_1 to p_r. This matrix is called the *rational Jordan normal form* of ϕ. There is one minor issue, though: we have used the partition λ_F rather than the original partition λ. The following exercise explains the relation, showing in particular that the latter determines the former.

Exercise 4.4.41. Let V be a finite-dimensional vector space over F of dimension l, where F is a field extension of K of finite dimension d over K. Let $\phi \in L_F(V)$ be a nilpotent linear map; then ϕ may also be regarded as a nilpotent K-linear map. Describe the relation between the associated partitions λ_F (when ϕ is regarded F-linear) and λ_K (when it is regarded merely K-linear).

Exercise 4.4.42. In this exercise you may use the full power of Mathematica. Download the large matrix A over $\mathbb{Q}$ from the course webpage.

(1) Find the minimal polynomial of A.
(2) Perform Hensel lifting to find the a polynomial $z(x)$ such that $A_s := z(A)$ is semisimple and $A - A_s$ is nilpotent. Also find A_s and A_n.
(3) For each irreducible factor of the minimal polynomial, compute the associated partition (over $\mathbb{Q}$).
(4) Find the rational Jordan normal form B of A over $\mathbb{Q}$.
(5) Find a linear map $g \in \mathrm{GL}(V)$ such that $gAg^{-1} = B$.

CHAPTER 5

Tensors

In many engineering applications, data are stored as multi-dimensional arrays of numbers, where two-dimensional arrays are just matrices. The mathematical notion capturing such data objects, without reference to an explicit basis and therefore amenable to linear algebra operations, is that of a *tensor*. Tensors are the subject of this and later chapters.

5.1. Multilinear functions

Let $V_1, \ldots, V_k, W$ be vector spaces over a field K. A map $f : V_1 \times \cdots \times V_k \to W$ is called *multilinear* if it satisfies

$$\begin{aligned} &f(v_1, \ldots, v_{j-1}, u_j + v_j, v_{j+1}, \ldots, v_k) \\ &= f(v_1, \ldots, v_{j-1}, u_j, v_{j+1}, \ldots, v_k) + f(v_1, \ldots, v_{j-1}, v_j, v_{j+1}, \ldots, v_k) \text{ and} \\ &f(v_1, \ldots, v_{j-1}, cv_j, v_{j+1}, \ldots, v_k) = cf(v_1, \ldots, v_{j-1}, v_j, v_{j+1}, \ldots, v_k) \end{aligned}$$

for all $j = 1, \ldots, k$ and all $u_j \in V_j$ and all $v_i \in V_i,\ i = 1, \ldots, k$ and all $c \in K$. Put differently, f is linear in each argument when the remaining arguments are fixed. Multilinear maps to the one-dimensional vector space $W = K$ are also called *multilinear forms*.

Multilinear maps $V_1 \times \cdots \times V_k \to W$ form a vector space with the evident addition and scalar multiplication. To get a feeling for how large this vector space is, let B_j be a basis of V_j for each $j = 1, \ldots, k$. Then a multilinear map f is uniquely determined by its values $f(u_1, \ldots, u_k)$ where u_j runs over B_j (express a general vector $v_j \in V_j$ as a linear combination of the u_j and make repeated use of the multilinearity of f). These values can be recorded in a $B_1 \times B_2 \times \cdots \times B_k$-indexed table of vectors in W. Conversely, given any such table t, there exists a multilinear f whose values at the tuples $(u_1, \ldots, u_k)$ with $u_j \in B_j,\ j = 1, \ldots, k$ are recorded in t, namely, the map that sends $(\sum_{u_1 \in B_1} c_{u_1} u_1, \ldots, \sum_{u_k \in B_k} c_{u_k} u_k)$ to the expression

$$\sum_{u_1 \in B_1, \ldots, u_k \in B_k} c_{u_1} \cdots c_{u_k} t(u_1, \ldots, u_k).$$

Hence we have a linear isomorphism between the space of multilinear maps and $W^{B_1 \times \cdots \times B_k}$. In particular, the dimension of the former space equals $|B_1||B_2| \cdots |B_k| \cdot \dim W$.

Example 5.1.43. Take all V_j equal to $V = \mathbb{R}^k$, where k is also the number of vector spaces under consideration. Then the function $f : V^k \to \mathbb{R}$ assigning to $(v_1, \ldots, v_k)$ the determinant of the matrix with columns $v_1, \ldots, v_k$ is multilinear. This function has the additional property that interchanging two arguments results

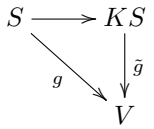
in multiplication by -1; such *alternating* forms will be studied in more detail in Chapter 8.

There are many more (non-alternating) multilinear forms: any function $g : V^k \to \mathbb{R}$ of the form $g(v_1, \dots, v_k) = (v_1)_{i_1} \cdots (v_k)_{i_k}$ with fixed $1 \le i_1, \dots, i_k \le k$ is multilinear. There are k^k of such functions, and they span the space of all multilinear functions $V^k \to \mathbb{R}$, confirming the computation above.

5.2. Free vector spaces

Given any set S, one may construct a K-vector space KS, the *free vector space spanned by* S of *formal* K*-linear combinations* of elements of S. Such linear combinations take the form $\sum_{s \in S} c_s s$ with all $c_s \in K$ and only finitely many among them non-zero, and the vector space operations are the natural ones. Note that we have already encountered this space in disguise: it is canonically isomorphic to $K^{\oplus S}$. The space KS comes with a natural map $S \to KS$ sending s to $1 \cdot s$, the formal linear combination in which s has coefficient 1 and all other elements have coefficient 0.

Here is a so-called *universal property* of KS: given any *ordinary* map g from S into a K-vector space V, there is a unique K*-linear* map $\tilde{g} : KS \to V$ that makes the diagram



commute. This formulation is somewhat posh, but analogues of it out to be very useful in more difficult situations, as we will see in the next section.

5.3. Tensor products

We will construct a space $V_1 \otimes \cdots \otimes V_k$ with the property that *multilinear* maps $V_1 \times \cdots \times V_k \to W$ correspond bijectively to *linear* maps $V_1 \otimes \cdots \otimes V_k \to W$. For this, we start by taking S to be the *set* $V_1 \times \cdots \times V_k$, where for a moment we forget the vector space structure of V_i. Then set $H := KS$ (for *Huge*), the free vector space formally spanned by all k-tuples $(v_1, \dots, v_k)$ with $v_j \in V_j,\ j = 1, \dots, k$. This is an infinite-dimensional vector space, unless the Cartesian product is a finite set (which happens only if either all V_i are zero or K is a finite field and moreover all V_i are finite-dimensional).

Next we define R (for *Relations*) to be the subspace of H spanned by all elements of the forms

$$1 \cdot (v_1, \dots, v_{j-1}, u_j + v_j, v_{j+1}, \dots, v_k) + (-1) \cdot (v_1, \dots, v_{j-1}, u_j, v_{j+1}, \dots, v_k)$$
$$+ (-1) \cdot (v_1, \dots, v_{j-1}, v_j, v_{j+1}, \dots, v_k) \text{ and}$$
$$1 \cdot (v_1, \dots, v_{j-1}, cv_j, v_{j+1}, \dots, v_k) + (-c) \cdot (v_1, \dots, v_{j-1}, v_j, v_{j+1}, \dots, v_k)$$

for all $j = 1, \dots, k$ and all $u_j \in V_j$ and all $v_i \in V_i,\ i = 1, \dots, k$ and all $c \in K$. Note the similarity with the definition of multilinear functions. When reading the expressions above, you should surpress the urge to think of the k-tuples as vectors;

the addition and scalar multiplication are taking place in the space H where the k-tuples are just symbols representing basis elements.

Finally, we set $V_1 \otimes \cdots \otimes V_k := H/R$, the quotient of the huge vector space H by the relations R. This vector space is called the *tensor product* of $V_1, \ldots, V_k$, and its elements are called *tensors*. The tensor product comes with a map $V_1 \times \cdots \times V_k \to V_1 \otimes \cdots \otimes V_k$ sending $(v_1, \ldots, v_k)$ to the equivalence class modulo R of $(v_1, \ldots, v_k) \in S \subseteq H$. This equivalence class is denoted $v_1 \otimes \cdots \otimes v_k$, called the *tensor product* of the vectors v_i, and a tensor of this form is called a *pure* tensor. By construction, every tensor is a linear combination of pure tensors.

EXERCISE 5.3.44. (1) Prove that $v_1 \otimes \cdots \otimes v_k$ is the zero tensor as soon as one of the v_i is the zero vector in V_i.
(2) Prove that, in fact, every tensor is a *sum* of pure tensors (so no coefficients are needed in the linear combination).

We claim that the map $(v_1, \ldots, v_k) \mapsto v_1 \otimes \cdots \otimes v_k$ is multilinear. Indeed, by the relations spanning R, we have

$$v_1 \otimes \cdots \otimes (u_j + v_j) \otimes \cdots \otimes v_k = v_1 \otimes \cdots \otimes u_j \otimes \cdots \otimes v_k + v_1 \otimes \cdots \otimes v_j \otimes \cdots \otimes v_k$$
$$\text{and } v_1 \otimes \cdots \otimes (cv_j) \otimes \cdots \otimes v_k = c(v_1 \otimes \cdots \otimes v_j \otimes \cdots \otimes v_k),$$

which is exactly what we need for the map to be multilinear. The most important property of the tensor product thus constructed is the following universal property (compare this with the easier universal property of KS).

THEOREM 5.3.45. *Given any vector space W and any* multilinear *map $f : V_1 \times \cdots \times V_k \to W$, there is a unique* linear *map $\overline{f} : V_1 \otimes \cdots \otimes V_k \to W$ such that the diagram*

$$\begin{array}{ccc} V_1 \times \cdots \times V_k & \longrightarrow & V_1 \otimes \cdots \otimes V_k \\ {\scriptstyle f}\downarrow & \swarrow_{\overline{f}} & \\ W & & \end{array}$$

commutes.

PROOF. Such a linear map $\overline{f}$ must satisfy

$$\overline{f}(v_1 \otimes \cdots \otimes v_k) = f(v_1, \ldots, v_k)$$

for all $v_i \in V_i,\ i = 1, \ldots, k$. Since pure tensors span the tensor product, this shows that only one such linear map $\overline{f}$ can exist, i.e., this shows uniqueness of $\overline{f}$. To prove existence, consider the following diagram:

$$\begin{array}{ccccc} S = V_1 \times \cdots \times V_k & \longrightarrow & H = KS & \longrightarrow & H/R \\ & {\scriptstyle f}\searrow & \downarrow{\scriptstyle \tilde{f}} & \swarrow_{\overline{f}} & \\ & & W. & & \end{array}$$

Here the existence of a linear map $\tilde{f}$ making the left-most triangle commute follows from the universal property of KS discussed earlier. We claim that $\ker \tilde{f}$ contains

R. Indeed, we have

$$\begin{aligned}&\tilde{f}((v_1,\dots,v_{j-1},u_j+v_j,v_{j+1},\dots,v_k)-(v_1,\dots,v_{j-1},u_j,v_{j+1},\dots,v_k)\\&-(v_1,\dots,v_{j-1},v_j,v_{j+1},\dots,v_k))\\&=f(v_1,\dots,v_{j-1},u_j+v_j,v_{j+1},\dots,v_k)-f(v_1,\dots,v_{j-1},u_j,v_{j+1},\dots,v_k)\\&-f(v_1,\dots,v_{j-1},v_j,v_{j+1},\dots,v_k)\\&=0,\end{aligned}$$

because f is multilinear, and similarly for the second type of elements spanning R. Thus $R\subseteq\ker\tilde{f}$ as claimed. But then, by Exercise 1.5.7 $\tilde{f}$ factorises through $H\to H/R=V_1\otimes\cdots\otimes V_k$ and the required linear map $\overline{f}:V_1\otimes\cdots\otimes V_k\to W$. □

The universal property at first seems somewhat abstract to grasp, but it is fundamental in dealing with tensor products. In down-to-earth terms it says that, to define a linear map $\phi:V_1\otimes\cdots\otimes V_k\to W$, it suffices to prescribe $\phi(v_1\otimes\cdots\otimes v_k)$ for all v_i, and the only restriction is that the resulting expression is multilinear in the v_i.

5.4. Two spaces

In this section we consider the tensor product $U\otimes V$. In this context, multilinear maps are called bilinear. Here is a first fundamental application of the universal property.

PROPOSITION 5.4.46. *Let $(u_i)_{i\in I}$ be a basis of U. Then for every tensor $\omega\in U\otimes V$ there exist unique $(v_i)_{i\in I}$, with only finitely many non-zero, such that $\omega=\sum_{i\in I}u_i\otimes v_i$.*

PROOF. This is the same as saying that the map $\psi:V^{\oplus I}\to U\otimes V$ sending $(v_i)_i$ to $\sum_i u_i\otimes v_i$ is bijective. Note that this map is linear. We will use the universal property to construct its inverse. For this let $f:U\times V\to V^{\oplus I}$ be the map defined by $f(\sum_i c_iu_i,v)=(c_iv)_{i\in I}$. This map is clearly bilinear, so that it factorises as

$$\begin{array}{ccc}U\times V & \longrightarrow & U\otimes V\\ {\scriptstyle f}\downarrow & \swarrow{\scriptstyle \overline{f}} & \\ V^{\oplus I}; & & \end{array}$$

and we claim that $\overline{f}$ is the inverse of ψ. Indeed, we have

$$\overline{f}(\psi((v_i)_i))=\overline{f}(\sum_i u_i\otimes v_i)=\sum_i f(u_i,v_i)=(v_i)_i,$$

which shows that $\overline{f}\circ\psi$ is the identity on $V^{\oplus I}$; and

$$\psi(\overline{f}((\sum_i c_iu_i)\otimes v))=\psi((c_iv)_{i\in I})=\sum_{i\in I}u_i\otimes c_iv=(\sum_{i\in I}c_iu_i)\otimes v,$$

which shows that $\psi\circ\overline{f}$ is the identity on a spanning set of $U\otimes V$, and hence by linearity the identity everywhere. □

This proposition has the following immediate corollary.

COROLLARY 5.4.47. *If $(u_i)_{i\in I}$ is a basis of U and $(w_j)_{j\in J}$ is a basis of V, then $(u_i \otimes w_j)_{i\in I, j\in J}$ is a basis of $U \otimes V$.*

PROOF. Indeed, by the proposition, every element $\omega \in U \otimes V$ can be written as $\sum_{i\in I} u_i \otimes v_i$ with $v_i \in V$. Writing $v_i = \sum_{j\in J} c_{ij} w_j$ gives

$$\begin{aligned}\omega &= \sum_{i\in I} u_i \otimes v_i \\ &= \sum_{i\in I} u_i \otimes (\sum_{j\in J} c_{ij} w_j) \\ &= \sum_{i\in I, j\in J} c_{ij}(u_i \otimes w_j).\end{aligned}$$

This shows that the $u_i \otimes w_j$ span $U \otimes V$. Moreover, if the latter expression is the zero vector, then by the proposition we have $\sum_{j\in J} c_{ij} w_j = 0$ for all i, and hence, since the w_j form a basis, all c_{ij} are zero. This proves linear independence. □

This shows that $\dim(U \otimes V) = \dim U \cdot \dim V$.

EXAMPLE 5.4.48. A common mistake when encountering tensors for the first time is the assumption that *every* tensor is pure. This is by no means true, as we illustrate now. Take $U = V = K^2$, both equipped with the standard basis e_1, e_2. A pure tensor is of the form

$$(ae_1 + be_2) \otimes (ce_1 + de_2) = ace_1 \otimes e_1 + ade_1 \otimes e_2 + bce_2 \otimes e_1 + bde_2 \otimes e_2.$$

The tensor's coefficients with respect to the basis $(e_i \otimes e_j)_{i,j=1,2}$ of $U \otimes V$ can be recorded in the matrix

$$\begin{bmatrix} ac & ad \\ bc & bd \end{bmatrix},$$

and we note that the determinant of this matrix is $(ac)(bd) - (ad)(bc) = 0$. Thus if $\omega = \sum_{i,j} c_{ij} e_i \otimes e_j \in U \otimes V$ is a tensor for which the coefficient matrix $(c_{ij})_{ij}$ has determinant unequal to zero, then ω is not a pure tensor.

EXERCISE 5.4.49. Extend the previous discussion to $U = K^m$ and $V = K^n$, as follows: prove that the coefficient matrix $(c_{ij})_{ij}$ of a tensor with respect to the basis $(e_i \otimes e_j)_{ij}$ is that of a pure tensor if and only if the matrix has rank 0 or 1.

As the preceding example shows, there is a tight connection between tensor products of two spaces and matrices. At the more abstract level of linear maps, this connection goes as follows. Let U and V be vector spaces, and consider the map $f : U^* \times V \to L(U, V)$ that sends a pair (x, v) to the linear map $u \mapsto x(u)v$. This linear map depends bilinearly on x and v, so f is bilinear, hence by the universal property there exists a unique linear map $\Psi = \overline{f} : U^* \otimes V \to L(U, V)$ that maps $x \otimes v$ to $u \mapsto x(u)v$. We claim that Ψ is linear isomorphism from $U^* \otimes V$ onto the set of linear maps $\phi : U \to V$ of *finite* rank.

EXERCISE 5.4.50. Prove that the linear maps $\phi : U \to V$ of finite rank, i.e., for which $\operatorname{im}\phi$ is finite-dimensional, do indeed form a linear subspace of $L(U, V)$.

First of all, $x \otimes v$ is mapped by Ψ to a linear map whose image is spanned by v, hence its rank is at most 1. An arbitrary element ω of $U^* \otimes V$ is a linear

combination of finitely many such tensors, and hence mapped by Ψ to a finite linear combination of rank-at-most-1 linear maps. Hence $\Psi\omega$ indeed has finite rank. Second, let $\phi : U \to V$ be a linear map of finite rank r. Let $v_1, \ldots, v_r$ be a basis of $\operatorname{im}\phi$ and let $u_1, \ldots, u_r \in U$ be elements such that $\phi(u_i) = v_i$. Then the u_i are linearly independent (because the v_i are), and we may extend $u_1, \ldots, u_r$ to a basis of U. Let $x_i,\ i = 1, \ldots, r$ be the element of U^* defined by $x_i(u_j) = \delta_{ij}$ for $j = 1, \ldots, r$ and $x_i(u) = 0$ for all further elements in the chosen basis of U. Then $\phi = \Psi(x_1 \otimes v_1 + \ldots + x_r \otimes v_r)$. This shows that Ψ is surjective onto the space of finite-rank linear maps. Finally, for injectivity of Ψ, suppose that $\Psi(\omega)$ is zero, and write $\omega = \sum_{i=1}^{l} x_i \otimes v_i)$, where by Proposition 5.4.46 we may assume that the x_i are linearly independent. Then there exist $u_j \in U,\ j = 1, \ldots, r$ such that $x_i(u_j) = \delta_{ij}$, and we have

$$0 = \Psi(\omega)u_j = \sum_{i=1}^{r} x_i(u_j)w_i = w_j$$

for all $j = 1, \ldots, r$. Hence $\omega = 0$. This concludes the proof that Ψ is a linear isomorphism from $U^* \otimes V$ to the space of finite-rank linear maps $U \to V$.

This relation between tensors and linear maps strongly suggest the following definition of rank of a tensor; we will generalise this notion in the next chapter to more tensor factors.

Definition 5.4.51. The *rank* of a tensor $\omega \in U \otimes V$ is the minimal number of terms in any expression of ω as a sum $u_1 \otimes v_1 + \ldots + u_r \otimes v_r$.

Exercise 5.4.52. Define a natural linear map $U^* \to V$ associated to ω, and prove that the rank of ω is equal to the rank of that linear map.

Specialising to $U = K^m$ and $V = K^n$, the exercise above shows that the rank of ω equals the rank of a certain linear map $(K^m)^* \to K^n$. The matrix of this linear map is the $n \times m$-matrix whose *transpose* is the coefficient matrix of ω relative to the basis $(e_i \otimes e_j)_{ij}$. Thus we find that the rank of ω equals the rank of this coefficient matrix.

The following exercise is a prelude to material on *communication complexity*.

Exercise 5.4.53. Fix a natural number n. Let Disj_n be the $2^n \times 2^n$-matrix whose rows and columns are labelled by all subsets of the numbers $\{1, \ldots, n\}$, with a 1 at position (S, T) if $S \cap T$ is empty and a 0 otherwise.

(1) Make a programme that on input n computes Disj_n in Mathematica.
(2) Experiment with this programme, and formulate a conjecture what the rank of Disj_n should be.
(3) Prove that conjecture.

CHAPTER 6

Tensor rank

Tensor rank is a ubiquitous notion in applications ranging from electrical engineering and computer science to phylogenetics. In this chapter we will encounter one application, having to do with complexity theory.

6.1. Higher-dimensional tensors

We start by determining a basis of arbitrary tensor products. Let $V_1, V_2, \ldots, V_k$ be vector spaces over K.

THEOREM 6.1.54. *If $B_1, \ldots, B_k$ are bases of $V_1, \ldots, V_k$, respectively, then $B := \{u_1 \otimes \cdots \otimes u_k \mid u_j \in B_j \text{ for all } j\}$ is a basis of $V_1 \otimes \cdots \otimes V_k$.*

PROOF. Given vectors $v_j \in V_j,\ j = 1, \ldots, k$, write $v_j = \sum_{u \in B_j} c_u u$. Then the pure tensor $v_1 \otimes \cdots \otimes v_k$ can be expanded as

$$\sum_{u_1 \in B_1, \ldots, u_k \in B_k} c_{u_1} \cdots c_{u_k} u_1 \otimes \cdots \otimes u_k,$$

a linear combination of B. Since pure tensors span the tensor product, so does B. On the other hand, to prove linear independence, assume that

$$0 = \sum_{u_1 \in B_1, \ldots, u_k \in B_k} c_{u_1, \ldots, u_k} u_1 \otimes \cdots \otimes u_k.$$

For each B_j and each $u \in B_j$ let $x_u \in V_j^*$ be the linear form that takes the value 1 on u and the value 0 on all other elements of B_j. Fix any k-tuple $(w_1, \ldots, w_k) \in B_1 \times \cdots \times B_k$. Then function f sending $(v_1, \ldots, v_k) \in V_1 \times \cdots V_k$ to $x_{w_1}(v_1) x_{w_2}(v_2) \cdots x_{w_k}(v_k)$ is multilinear, and hence factorises through a linear form $\overline{f} : V_1 \otimes \cdots \otimes V_k \to K$. Applying this linear form to both sides of the equality above yields

$$0 = \sum_{u_1 \in B_1, \ldots, u_k \in B_k} c_{u_1, \ldots, u_k} x_{w_1}(u_1) \cdots x_{w_k}(u_k) = c_{w_1, \ldots, w_k},$$

which proves linear independence of B. □

This proposition shows that one may think of a tensor in $V_1 \otimes \cdots \otimes V_k$ as a $B_1 \times \cdots \times B_k$-indexed table of numbers with only finitely many non-zero entries. In fact, in applications the V_j are often all equal to some K^{n_j} and the B_j are standard bases, and with these choices a tensor is nothing but at $n_1 \times \cdots \times n_k$-array of numbers from K.

6.2. Definition of rank

The central definition in this chapter is the following.

DEFINITION 6.2.55. The *rank* of a tensor $\omega \in V_1 \otimes \cdots \otimes V_k$, denoted $\operatorname{rk}\omega$, is the minimal number r in any decomposition of ω as a sum $\omega_1 + \ldots + \omega_r$ of pure tensors.

Hence the zero tensor has rank zero, a non-zero pure tensor $v_1 \otimes \cdots \otimes v_k$ has rank one, a non-pure tensor that can be written as the sum of two pure tensors has rank two, etc. We have seen in the previous chapter that when $k = 2$, the rank of a tensor $\omega \in V_1 \otimes V_2$ equals the rank of the corresponding linear map $V_1^* \to V_2$. Specialising even further, when $V_j = K^{n_j},\ j = 1, 2$ equipped with the standard basis, the rank equals the rank of the coefficient matrix of ω with respect to the basis $e_{i_1} \otimes e_{i_2},\ i_1 = 1, \ldots, n_1,\ i_2 = 1, \ldots, n_2$. So rank is well-understood for tensor products of two spaces. The situation is much more complicated for tensor products of more spaces, as the following example illustrates.

EXAMPLE 6.2.56. Let $\omega \in \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$ be a non-zero tensor. With respect to the basis $e_1 \otimes e_1 \otimes e_1, e_1 \otimes e_1 \otimes e_2, \ldots, e_2 \otimes e_2 \otimes e_2$ the tensor ω is represented by a $2 \times 2 \times 2$-block of complex numbers. Write ω as $e_1 \otimes A + e_2 \otimes B$ where $A, B \in \mathbb{C}^2 \otimes \mathbb{C}^2$ are represented by 2×2-matrices.

First, if ω is a pure tensor $u \otimes v \otimes w$ with $u, v, w \in \mathbb{C}^2$, then we have $A = u_1(v \otimes w)$ and $B = u_2(v \otimes w)$ and together they span the one-dimensional space $Kv \otimes w$ spanned by the rank-one tensor $v \otimes w$. Conversely, when the span of A and B is one-dimensional and spanned by a pure tensor, then ω is pure (and hence of rank one).

Second, when the span of A and B is one-dimensional but not spanned by a pure tensor, then it is spanned by a rank-two tensor (since 2×2-matrices have rank at most two) $C = v_1 \otimes w_1 + v_2 \otimes w_2$. Writing $A = aC$ and $B = bC$ we have

$$\omega = (ae_1 + be_2) \otimes v_1 \otimes w_1 + (ae_1 + be_2) \otimes v_2 \otimes w_2,$$

so that ω has rank at most two. By the above it is not pure, hence it has rank two.

Third, when A and B are linearly independendent, and their span $\langle A, B \rangle_{\mathbb{C}}$ also has a basis consisting of two pure tensors $C_1 = v_1 \otimes w_1, C_2 = v_2 \otimes w_2$, then writing $A = a_1C_1 + a_2C_2$ and $B = b_1C_1 + b_2C_2$ we have

$$\begin{aligned} \omega &= e_1 \otimes (a_1C_1 + a_2C_2) + e_2 \otimes (b_1C_1 + b_2C_2) \\ &= (a_1e_1 + b_1e_2) \otimes C_1 + (a_2e_1 + b_2e_2) \otimes C_2 \\ &= (a_1e_1 + b_1e_2) \otimes v_1 \otimes w_1 + (a_2e_1 + b_2e_2) \otimes v_2 \otimes w_2, \end{aligned}$$

showing that ω has rank at most two, hence equal to two since it is not pure.

EXERCISE 6.2.57. Conversely, show that if A and B are linearly independent and ω has rank two, then the span $\langle A, B \rangle_{\mathbb{C}}$ has a basis of two rank-one tensors.

Finally, we have to analyse the case where A and B are linearly independent, and their span does not contain two linearly independent rank-one tensors. In particular, at least one of A and B has rank at least two; assume that this is true for A. Now consider the polynomial $p_\omega(x) := \det(xA + B) = f(A, B)x^2 + g(A, B)x + h(A, B)$, where we identify A and B with their 2×2-matrices. This is a quadratic

polynomial in x, whose coefficients are homogeneous quadratic polynomials in the entries of A and B. The coefficient of x^2 is $\det(A)$, which we have assumed non-zero, so that p_ω is really a quadratic polynomial. The assumption that $\langle A, B\rangle_{\mathbb{C}}$ does not contain two linearly independent rank-one matrices is equivalent to the condition that p_ω does not have two distinct roots, which in turn is equivalent to the condition that the *discriminant* $\Delta := g(A,B)^2 - 4f(A,B)h(A,B)$ of p_ω is zero. This discriminant equals

$$\begin{aligned}\Delta = {} & a_{11}^2 b_{22}^2 + a_{22}^2 b_{11}^2 + a_{12}^2 b_{21}^2 + a_{21}^2 b_{12}^2 \\ & - 2a_{11}a_{12}b_{21}b_{22} - 2a_{11}a_{21}b_{12}b_{22} - 2a_{11}a_{22}b_{11}b_{22} \\ & - 2a_{12}a_{21}b_{12}b_{21} - 2a_{12}a_{22}b_{11}b_{21} - 2a_{21}a_{22}b_{11}b_{12} \\ & + 4a_{11}a_{22}b_{12}b_{21} + 4a_{12}a_{21}b_{11}b_{22},\end{aligned}$$

and this quartic polynomial in the entries of ω is called *Cayley's hyperdeterminant*. We conclude that if ω does not fall into any of the three previous three cases, then its hyperdeterminant is zero.

EXERCISE 6.2.58. Show that, in the last case, ω has rank three.

This example makes a few important points. First, that tensor rank for more than two factors is a more complicated notion than that for two factors (matrices); indeed, it is known that it is *NP-hard* already for three factors [**2**]. Second, over the complex numbers, that tensors of rank at most some fixed bound r do in general not form a closed subset (they do for matrices, since they are characterised by the vanishing of all $(r+1)\times(r+1)$-subdeterminants, a closed condition on the matrix entries). Indeed, a $2\times 2\times 2$-tensor ω of the last type above, on which Cayley's hyperdeterminant vanishes, will have arbitrarily close tensors with non-zero hyperdeterminant. Those all have rank at most 2, while ω has rank 3. Third, the rank of a tensor may depend on the field one is working over, as the following exercise shows.

EXERCISE 6.2.59. Prove that if a tensor $\omega \in \mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R}^2$ has negative hyperdeterminant (which means that p_ω has negative discriminant), then it has rank at least three over the real numbers, while if we allow the pure tensors in the decomposition of ω to be complex, it has rank at most two.

6.3. Rank under tensor operations

Although tensor rank is hard to grasp, it obeys a number of easy inequalities. These inequalities involve the following three basic operations on tensors.

Tensor product. Given vector spaces $V_1, \dots, V_l, V_{l+1}, \dots, V_k$, there is a natural bilinear map

$$(V_1 \otimes \cdots \otimes V_l) \times (V_{l+1} \otimes \cdots \otimes V_k) \to V_1 \otimes \cdots \otimes V_k,\ (\omega, \mu) \mapsto \omega \otimes \mu$$

determined uniquely by sending $(v_1 \otimes \cdots \otimes v_l, v_{l+1} \otimes \cdots \otimes v_k)$ to $v_1 \otimes \cdots \otimes v_l \otimes v_{l+1} \otimes \cdots v_k$. Since the latter expression is l-linear in $(v_1, \dots, v_l)$ and $(k-l)$-linear in $(v_{l+1}, \dots, v_k)$, the existence of this bilinear map follows from the universal properties of $V_1 \otimes \cdots \otimes V_l$ and $V_{l+1} \otimes \cdots \otimes V_k$; we omit the details. The image $\omega \otimes \mu$ of (ω, μ) is called the *tensor product* of ω and μ. The following exercise gives a fundamantal inequality for the rank of a tensor product of tensors.

EXERCISE 6.3.60. Prove that $\operatorname{rk}(\omega \otimes \mu) \leq \operatorname{rk}(\omega)\operatorname{rk}(\mu)$.

In other words: tensor rank is *sub-multiplicative*. To the best of my knowledge, it is *not known* whether the inequality can be strict.

Flattening. In a sense, flattening is the inverse operation of tensor product: the bilinear map $(V_1 \otimes \cdots \otimes V_l) \times (V_{l+1} \otimes \cdots \otimes V_k) \to V_1 \otimes \cdots \otimes V_k$ above factors through a linear map $(V_1 \otimes \cdots \otimes V_l) \otimes (V_{l+1} \otimes \cdots \otimes V_k) \to V_1 \otimes \cdots \otimes V_k$. This linear map is, in fact, a linear isomorphism. Let us denote the inverse by $\flat$. The pre-image $\flat\omega$ of $\omega \in V_1 \otimes \cdots \otimes V_k$ is called a *flattening* of ω. It is a tensor in a tensor product of two vector spaces, namely, $U := V_1 \otimes \cdots \otimes V_l$ and $W := V_{l+1} \otimes \cdots \otimes V_k$ and, as such, has a well-understood rank (namely, that of a matrix). It turns out that $\operatorname{rk}\flat\omega \leq \operatorname{rk}\omega$. Indeed, this follows immediately from the fact that $\flat$ maps a pure tensor $v_1 \otimes \cdots \otimes v_k$ into a pure tensor (or matrix) $(v_1 \otimes \cdots \otimes v_l) \otimes (v_{l+1} \otimes \cdots \otimes v_k)$.

EXERCISE 6.3.61. Give an example of a rank-two tensor in $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$ whose flattening to an element of $(\mathbb{C}^2) \otimes (\mathbb{C}^2 \otimes \mathbb{C}^2)$ has rank one.

Here is a partial converse to the inequality above.

PROPOSITION 6.3.62. *Let ω be a tensor in $V_1 \otimes \cdots \otimes V_k$ and for $i = 1, \ldots, k$ let $\flat_i$ denote the flattening $V_1 \otimes \cdots \otimes V_k \to (V_i) \otimes (V_1 \otimes \cdots \otimes \widehat{V_i} \otimes \cdots \otimes V_k)$, where the hat means omission of the factor V_i. Then $\operatorname{rk}\omega \leq \prod_{i=1}^{k} \operatorname{rk}(\flat_i\omega)$.*

PROOF. The flattening $\flat_i\omega$ defines a linear map $(V_1 \otimes \cdots \otimes \widehat{V_i} \otimes \cdots \otimes V_k)^* \to V_i$; let U_i be the image of this map. Then $\dim U_i = \operatorname{rk}\flat_i\omega$ by the interpretation of tensor rank for factors. We claim that ω lies in $U_1 \otimes \cdots \otimes U_k$, a tensor product that is naturally contained in $V_1 \otimes \cdots \otimes V_k$. Indeed, for each i choose a linear map $\pi_i : V_i \to U_i$ that restricts to the identity on U_i. Then, regarding $\flat_i\omega$ as a linear map $(V_1 \otimes \cdots \otimes \widehat{V_i} \otimes \cdots \otimes V_k)^* \to U_i \subseteq V_i$, we have $\pi_i \circ \flat_i\omega = \flat_i\omega$, since π_i is the identity on the image of $\flat_i\omega$. The maps $\pi_1, \ldots, \pi_k$ together define a linear map $\psi : V_1 \otimes \cdots \otimes V_k \to U_1 \otimes \cdots \otimes V_k$, determined by sending $v_1 \otimes \cdots \otimes v_k$ to $\pi_1(v_1) \otimes \cdots \otimes \pi_k(v_k)$ (note that we use the universal property). By the argument just given, $\psi\omega = \omega$, so that $\omega \in U_1 \otimes \cdots \otimes U_k$ as claimed. This claim implies the desired result: after choosing bases B_i in all U_i, ω is a linear combination of the corresponding $|B_1| \cdots |B_k|$ pure tensors forming a basis of $U_1 \otimes \cdots \otimes U_k$. □

Contraction. A linear function $x \in V_i^*$ gives rise to a linear map $\psi_{i,x} : V_1 \otimes \cdots \otimes V_k \to V_1 \otimes \cdots \otimes \widehat{V_i} \otimes \cdots \otimes V_k$ sending $v_1 \otimes \cdots \otimes v_k$ to $x(v_i) \cdot v_1 \otimes \cdots \otimes \widehat{v_i} \otimes \cdots \otimes v_k$. A tensor of the form $\psi_{i,x}\omega$ for some i and some $x \in V_i^*$ is called a *contraction* of ω.

EXERCISE 6.3.63. Show that $\operatorname{rk}\psi_{i,x}\omega \leq \operatorname{rk}\omega$.

The following converse to this exercise is open.

CONJECTURE 6.3.64. *Fix a natural number r. Then there exists a universal bound k_0, depending on r (and perhaps on the ground field), such that for any $k > k_0$ and any tensor ω in any k-fold tensor product $V_1 \otimes \cdots \otimes V_k$ we have $\operatorname{rk}\omega \leq r$ if and only if all contractions of ω have rank at most r.*

6.4. Communication complexity

The discussion that follows is based on the first chapter of [**3**]. Suppose that each of two players, Alice and Bob, holds one argument to a two-argument function $f : \{1, \dots, m\} \times \{1, \dots, n\} \to \{0, 1\}$, and that they want to evaluate f with as little communication as possible between them. Here the function f is known to both players, and we will think of it as an $m \times n$-matrix, of which Alice holds a row index i and Bob holds a column index j. One possible protocol would be that Alice sends her row index to Bob, Bob computes f and sends the result back to Alice. This would require $\lceil \log_2(m) \rceil + 1$ bits being exchanged. Here is an example, however, where one can do much better.

EXAMPLE 6.4.65. Suppose that G is a finite, undirected graph on $V = \{1, \dots, n\}$ without loops or multiple edges. Alice's argument consists of a subset C of V which is a clique in the graph (pairwise connected vertices), and Bob's argument consists of a subset I of V which is an independent set (pairwise non-connected vertices). The function f counts the cardinality $|C \cap I|$, which is at most 1. Sending Alice's entire coclique as a 0/1-vector indexed by V would yield n bits of information being transmitted. However, as Yannakakis pointed out [**4**], they can do better.

The key point is the following. Suppose Alice and Bob have narrowed down the search for an intersection to a subset $V' \subseteq V$, so that they know for sure that *if* C and I share a vertex, then it lies in the intersection of $C' := V' \cap C$ and $I' := V' \cap I$. Then if C' contains a vertex with at most $|V'|/2$ neighbours in V', Alice sends the name of such a vertex u to Bob, using $O(\log_2 n)$ bits. Of course, if u also lies in I', then Bob returns "yes", and the protocol is finished. If $u \notin I'$, then they may replace V' by the set of neighbours of u in V' and return to the beginning. If C' does not contain such a vertex, then Alice transmits this information to Bob (using some constant number of bits). He then checks whether I' contains a vertex with at most $|V'|/2$ non-neighbours in V'. If so, then he sends the name of such a vertex v to Alice, using $O(\log_2 n)$ bits. If v lies in C', then Alice reports "yes", and the protocol is finished; otherwise, they may both replace V' by the non-neighbours of v in V' and start from the beginning. Finally, if neither Alice nor Bob can choose a vertex to transmit, then all elements in C' have more than $|V'|/2$ neighbours, while all elements in I' have at less than $|V'|/2$ neighbours, so that $I' \cap C'$ is empty. Bob transmits this information to Alice, and we are done.

With each choice of u or v, the cardinality of V' is (at least) halved. This means that Alice and Bob come to a conclusion after at most $O(\log_2 n)$ such steps. The number of bits transmitted is therefore $O((\log_2 n)^2)$.

For the purpose of this section, a *protocol* is a finite, rooted, binary tree (i.e., all internal edges have valency two), where each internal vertex v is labelled either by A and a function a_v from $\{1, \dots, m\}$ to the two-element set of children of v or by B and a function b_v from $\{1, \dots, n\}$ to the two-element set of children of v; and where each leaf is labelled either 0 or 1. *Running* the protocol on a pair (i, j) means starting at the root $r =: v$, repeatedly moving to $a_v(i)$ or $b_v(j)$ according as the current vertex is labelled A or B, until one ends in a leaf labelled 0 or 1, which is then the output of the protocol. Thus the protocol defines a function $\{1, \dots, m\} \times \{1, \dots, n\} \to \{0, 1\}$, and if this function is identical to the function f that we started with, then we say that the protocol *computes* f. The *height*

of the tree, which is the longest distance in the tree from a root to a leaf, is the *communication complexity* of the protocol. The *communication complexity* of a function f is the minimal communication complexity of any protocol computing it. This is the framework from [**5**]. It is easy (but boring) to translate the above informal protocol for the clique vs. independent set problem into a protocol as just described.

THEOREM 6.4.66. *The communication complexity of any protocol computing f is at least $\log_2 \operatorname{rk} f$, when f is considered an $m \times n$-matrix with entries zero and one.*

Note that we have not specified the field here. This is deliberate; the rank over any field gives a lower bound on the complexity.

PROOF. For a fixed leaf l of the protocol, consider the set R_l of all input pairs (i, j) on which the protocol's run ends in l. Clearly the sets R_l form a partition of $\{1, \ldots, m\} \times \{1, \ldots, n\}$. We claim that each R_l is a *rectangle*, which means that it is of the shape $I_l \times J_l$ with I_l a subset of the rows and J_l a subset of the columns. To prove this, it is sufficient to prove that if (i_0, j_0) and (i_1, j_1) are both in R_l, then also (i_0, j_1) lies in R_l. But this is clear: at internal vertices labelled A, the chosen child depends only on the first argument. Since the first argument of (i_0, j_1) agrees with that of (i_0, j_0), the same child is chosen. Similarly, at internal vertices labelled B, the same child is chosen because the second argument of (i_0, j_1) agrees with that of (i_1, j_1).

Let A_l be the 0/1 matrix with ones in the positions labelled $I_l \times J_l$ and zeroes elsewhere. Then A_l has rank one (or zero, if the leaf l is never reached). Moreover, let c_l be the output of the protocol at leaf l. Then we have the equality of matrices

$$f = \sum_l c_l A_l,$$

where the sum runs over all leaves l. The number of terms is at most 2 raised to the height of the tree, which yields the desired inequality. □

EXERCISE 6.4.67. Let $m = n = 2^l$, and f is the function that takes two l-bit strings to the parity (0 or 1) of their inner product. Using the rank lower bound, prove that the communication complexity of any protocol computing f is at least (roughly) l. Hint: compute the square of the matrix representing f.

EXERCISE 6.4.68. Using Exercise 5.4.53 give a lower bound on the communication complexity on the problem of deciding whether a subset $S \subseteq \{1, \ldots, n\}$ held by Alice is disjoint from a subset $T \subseteq \{1, \ldots, n\}$ held by Bob.

EXERCISE 6.4.69. Let $m = n$, and let $f(i, j)$ be the function that is 1 if $i \leq j$ and 0 if $i > j$. Prove that the communication complexity of any protocol for f is at least (roughly) $\log_2 n$.

All definitions of protocol, communication complexity, etc. generalise in a straightforward manner to the case of $k > 2$ players: the internal vertices of the tree can now have k different labels, corresponding to whose "move" it is. The leaves are still labelled by the output, which we still assume equal to 0 or 1 (this requirement can be weakened, as well). In the case of two players, the edge leading to a leaf furthest away from the root contributed one bit to the communication complexity, which

made sense since that bit still needed to be communicated to the other player. In the k-player setting, one either has to adjust this by adding $(k-1)$ to the height of the tree (in case of one-to-one communication channels), or else one assumes a *broadcast channel* (sometimes called a *blackboard channel* in communication complexity literature) where all broadcast bits are received by all players. The log-rank theorem relating the communication complexity to the matrix of f generalises to this setting, except that now f is represented by a $n_1 \times \cdots \times n_k$-block of zeroes and ones, and that the rank of this block is the *tensor rank*.

EXERCISE 6.4.70. This exercise concerns a small communication problem with three players, each holding one bit of input. The function that they want to compute is 1 on the triple $(0,0,0),(0,1,1),(1,0,1)$ and zero on all other triples. Find the exact communication complexity of this function, as well the rank of this tensor.

When trying to bound the communication complexity of k-argument function, often so-called *partition arguments* are used: the players are partitioned into two groups $1,\ldots,l$ and $l+1,\ldots,k$, and each group is regarded one single player (communication among the members of the group is regarded "free"). The communication complexity of the resulting two-argument function is clearly a *lower bound* to the communication complexity of the original function.

EXERCISE 6.4.71. Discuss the relation of this partition argument with the behaviour of tensor rank under flattening.

EXERCISE 6.4.72. Suppose three players A,B,C hold non-zero elements a,b,c of $\mathbb{Z}/p\mathbb{Z}$, where p is a large prime, and want to decide whether $abc = 1 \mod p$. Give a protocol (in words, no complete tree needed) with communication complexity some constant multiple of $\log_2 p$, and show that any protocol has communication complexity at least some (potentially other) constant multiple of $\log_2 p$.

The relations between tensor rank, communication complexity, flattenings, and partition arguments have been thoroughly dealt with in [**1**].

Bibliography

[1] Jan Draisma, Eyal Kushilevitz, and Enav Weinreb. Partition arguments in multiparty communication complexity. *Theor. Comput. Sci.*, 412(24):2611–2622, 2011.

[2] Johan Håstad. Tensor rank is NP-complete. *J. Algorithms*, 11(4):644–654, 1990.

[3] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge Univ. Press., Cambridge, 1997.

[4] Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *J. Comput. System Sci.*, 43(3):441–466, 1991.

[5] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *Proceedings of the eleventh annual ACM symposium on Theory of computing, Atlanta, Georgia, United States*, pages 209–213, New York, 1979. ACM.

CHAPTER 7

Matrix multiplication and tensors of linear maps

7.1. Tensors, direct sums, and duals

One aspect of tensor products that we have not yet stressed is their behaviour with respect to direct sums and duals. In this section, for simplicity of notation, we consider tensor products of two spaces, but everything generalises in a straightforward manner.

Let U, V be vector spaces over K and suppose that we have a subspace $U_1 \subseteq U$. Then the map $U_1 \times V \to U \otimes V,\ (u_1, v) \mapsto u_1 \otimes v$ is bilinear and hence factorises through a linear map $U_1 \otimes V \to U \otimes V$. This linear map is in fact injective (as one can see, for instance, from the fact that it maps a basis of pure tensors $u \otimes v$ with u running over a basis of U_1 and v running over a basis of V into a subset of a basis of $U \otimes V$), and it allows us to see $U_1 \otimes V$ as a subspace of $U \otimes V$.

Now a second subspace $U_2 \subseteq U$ such that $U = U_1 \oplus U_2$ gives rise to a second subspace $U_2 \otimes V$ of $U \otimes V$, and we claim that

$$((U_1 \oplus U_2) \otimes V =)\ U \otimes V = (U_1 \otimes V) \oplus (U_2 \otimes V).$$

This follows, for instance, by taking bases B_1, B_2 of U_1, U_2, respectively, and a basis C of V, observing that $U_1 \otimes V$ is spanned by $\{u \otimes v \mid u \in B_1, v \in C\}$ and $U_2 \otimes V$ is spanned by $\{u \otimes v \mid u \in B_2, v \in V\}$, and recalling that the union of these two sets is a basis of $U \otimes V$. Thus, *tensor product distributes over direct sum.*

Next we discuss duals. Given a pair $(x, y) \in U^* \times V^*$, there is a unique linear function on $U \otimes V$ that sends a pure tensor $u \otimes v$ to $x(u)y(v)$ (by the universal property of $U \otimes V$). This linear function itself depends bilinearly on (x, y), so that we obtain a linear map $\Psi : U^* \otimes V^* \to (U \otimes V)^*$ sending (x, y) to the linear function mapping $u \otimes v$ to $x(u)y(v)$. We claim that Ψ is injective. Indeed, assume that $\Psi(\omega) = 0$ and write $\omega = \sum_{i=1}^{r} x_i \otimes y_i$, where we may assume that both the x_i and the y_i are linearly independent. If $r > 0$ then find a vector $u_1 \in U$ such that $x_i(u_1) = \delta_{i1}$ and a vector $v_1 \in V$ such that $y_i(v_1) = \delta_{i1}$ (why do these exist?). Then $\Psi(\omega)(u_1 \otimes v_1) = 1 \cdot 1 \neq 0$, a contradiction. Thus we find that ω is zero.

When the dimensions of U and of V are finite, we find that the map $U^* \otimes V^* \to (U \otimes V)^*$ is an isomorphism, by means of which we may identify the two spaces. Thus, for finite-dimensional vector spaces, *tensor product commutes with taking duals.*

7.2. Tensor products of linear maps

Let U, V, W, X be vector spaces over K, and let $\phi \in L(U, V)$ and $\psi \in L(W, X)$ be linear maps. Then we may define a bilinear map $f : U \times W \to V \otimes X$ by

$$f(u, w) = \phi(u) \otimes \psi(w).$$

By the universal property of the tensor product $U \otimes W$ this map factorises through a linear map $\overline{f} : U \otimes W \to V \otimes X$. This linear map is denoted $\phi \otimes \psi$ and called the tensor product of ϕ and ψ. In this chapter we derive some of its properties and applications.

At first glance, confusion may potentially arise from the fact that we write $\phi \otimes \psi$ both for the linear map just constructed, which is an element of $L(U \otimes W, V \otimes X)$, and for the tensor product of ϕ and ψ in the tensor product $L(U, V) \otimes L(W, X)$. But on closer inspection the linear map $\phi \otimes \psi$ depends bilinearly on (ϕ, ψ), so by the universal property of $L(U, V) \otimes L(W, X)$ there is a linear map Ψ from this tensor product to $L(U \otimes W, V \otimes X)$ that maps a pure tensor $\phi \otimes \psi$ to the linear map $\phi \otimes \psi$ defined above.

Exercise 7.2.73. Prove that the linear map $\Psi : L(U, V) \otimes L(W, X) \to L(U \otimes W, V \otimes X)$ sending a pure tensor $\phi \otimes \psi$ to the linear map sending a pure tensor $u \otimes w$ to $\phi(u) \otimes \psi(w)$ is injective. When all spaces involved are finite-dimensional, conclude that Ψ is a linear isomorphism.

Here are two fundamental properties of $\phi \otimes \psi$.

Proposition 7.2.74. *The kernel of $\phi \otimes \psi$ equals $(\ker \phi) \otimes W + U \otimes (\ker \psi) \subseteq U \otimes W$, and the image of $\phi \otimes \psi$ equals $\operatorname{im} \phi \otimes \operatorname{im} \psi \subseteq V \otimes X$.*

Proof. Since $(\phi \otimes \psi)(u \otimes w) = \phi(u) \otimes \psi(w)$ lies in the subspace $\operatorname{im} \phi \otimes \operatorname{im} \psi$ of $V \otimes X$, and since pure tensors $u \otimes w$ span the space $U \otimes W$, the map $\phi \otimes \psi$ maps that entire space into $\operatorname{im} \phi \otimes \operatorname{im} \psi$. Moreover, since pure tensors $\phi(u) \otimes \psi(w)$ span $\operatorname{im} \phi \otimes \operatorname{im} \psi$, the map $\phi \otimes \psi$ is surjective onto $\operatorname{im} \phi \otimes \operatorname{im} \psi$. This proves that $\operatorname{im}(\phi \otimes \psi) = \operatorname{im} \phi \otimes \operatorname{im} \psi$.

For the kernel, first note that $\ker \phi \otimes W$ and $U \otimes \ker \psi$ are clearly contained in $\ker(\phi \otimes \psi)$. To prove that they span it, let U_1 be a vector space complement of $\ker \phi$ in U and let W_1 be a vector space complement of $\ker \psi$ in W. Then we have

$$\begin{aligned} U \otimes W &= (\ker \phi \oplus U_1) \otimes (\ker \psi \oplus W_1) \\ &= (\ker \phi \otimes \ker \psi) \oplus (\ker \phi \otimes W_1) \oplus (U_1 \otimes \ker \psi) \oplus (U_1 \otimes W_1) \\ &= ((\ker \phi \otimes W) + (U \otimes \ker \psi)) \oplus (U_1 \otimes W_1), \end{aligned}$$

where the $+$ indicates a not necessarily direct sum. So we need only show that $\ker(\phi \otimes \psi)$ does not intersect $U_1 \otimes W_1$. But this is immediate from the fact that $\phi|_{U_1}$ and $\psi|_{W_1}$ are injective: as u_1 runs through a basis of U_1 and w_1 runs through a basis of W_1, the element $(\phi \otimes \psi)(u_1 \otimes w_1) = \phi(u_1) \otimes \psi(w_1)$ runs through part of a suitable basis of $V \otimes X$. □

From this proposition we conclude that, in particular, the rank of $\phi \otimes \psi$ is the product of the ranks of ϕ and of ψ. This has the following application to tensor rank.

COROLLARY 7.2.75. *Let U, V, W, X be finite-dimensional vector spaces and let $\omega \in U \otimes V$ and $\mu \in W \otimes X$ be tensors. Then the rank of the tensor product $\omega \otimes \mu \in U \otimes V \otimes W \otimes X$ equals $\operatorname{rk}\omega \cdot \operatorname{rk}\mu$.*

Compare this with the remark after Exercise 6.3.60: if tensor rank in general is not multiplicative, then a counter-example would involve at least *five* vector spaces. Finite-dimensionality is not really needed for the corollary, but makes the proof easier to formulate.

PROOF. We have already seen that tensor rank is submultiplicative, so we need only show that the rank of $\omega \otimes \mu$ is not *less* than $\operatorname{rk}\omega \cdot \operatorname{rk}\mu$. The product $\omega \otimes \mu$ lives in

$$U \otimes V \otimes W \otimes X = U \otimes W \otimes V \otimes X,$$

where we have re-arranged the tensor factors. Now consider the *flattening* $\flat(\omega \otimes \mu)$ of $\omega \otimes \mu$ in $(U \otimes W) \otimes (V \otimes X)$. We claim that this flattening has rank $\operatorname{rk}\omega \cdot \operatorname{rk}\mu$; since tensor rank cannot go up under flattening, this gives the desired inequality. To prove the claim, we re-interpret ω as an element of $(U^*)^* \otimes V = L(U^*, V)$ and μ as an element fo $(W^*)^* \otimes X = L(W^*, X)$, where the equality signs are justified by finite dimensionality. The ranks of ω and μ are equal to the ranks of these linear maps. On the other hand, the flattening $\flat(\omega \otimes \mu)$ can be re-interpreted as an element of $L((U \otimes W)^*, (V \otimes X)) = L(U^* \otimes W^*, V \otimes X)$, namely the map sending a pure tensor $x \otimes y \in U^* \otimes W^*$ to the tensor $\omega(x) \otimes \mu(y)$. But this linear map is just the tensor product of the linear maps ω and μ, and by the above its rank equals the product of the ranks of those linear maps. □

7.3. Extending scalars

Given a K-linear map $\psi \in L(W, X)$, and an extension field F of K, we may apply the tensor product construction with $U = V = F$ and $\phi = 1$, the identity $F \to F$. The resulting K-linear map $1 \otimes \psi : F \otimes U \to F \otimes W$ is in fact F-linear if we define the F-scalar multiplication on $F \otimes W$ by $c(d \otimes W) := (cd) \otimes W$ for $c, d \in F$ (note that for $c \in K$ this agrees with the K-scalar multiplication by properties of the tensor product).

The resulting F-linear map $1 \otimes \psi$ is denoted ψ_F, and the procedure leading from ψ to ψ_F is called *extension of scalars*. At the level of matrices this is a trivial operation: given K-bases $(w_j)_j$ and $(x_i)_i$ of W and X, respectively, the elements $1 \otimes w_j$ and $1 \otimes x_i$ form F-bases of $F \otimes W$ and $F \otimes V$, respectively, and the matrix of ϕ_F with respect to those bases is just the matrix of ϕ with respect to the original bases, only now with its entries interpreted as elements of the larger field $F \supseteq K$. Nevertheless, extension of scalars is a useful operation, partly due to the following proposition, where we take $X = W$.

PROPOSITION 7.3.76. *Let W be finite-dimensional. A linear map $\psi \in L(W)$ is semisimple if and only if ϕ_F is diagonalisable for F equal to the algebraic closure of K.*

PROOF. If ψ is semisimple, then by the results of chapter 4 there is a basis of W with respect to which the matrix of ψ has a block diagonal form, with companion matrices C_p of irreducible polynomials along the diagonal. Thus to prove that ϕ_F

is diagonalisable over the algebraic closure F of K it suffices to diagonalise C_p over F. Now C_p is the matrix of the linear map $\psi : K[t]/(p) \to K[t]/(p)$ sending $f + (p)$ to $tf + (p)$. Over F the polynomial p factors as $(t - t_1) \cdots (t - t_d)$ with distinct $t_1, \ldots, t_d \in F$. Then setting $f_j := \prod_{i \neq j}(t - t_i) + (p)$ we find $(t - t_j)f_j = 0 \mod p$ so that $\psi f_j = t_j f_j$. Since they are eigenvectors with distinct eigenvalues, the f_j are linearly independent, and since there are d of them, they form a basis of $F[t]/(p)$ diagonalising ψ.

Conversely, assume that ϕ_F is diagonalisable over F with distinct eigenvalues $\lambda_1, \ldots, \lambda_r$. Then its minimal polynomial over F equals $(t - \lambda_1) \cdots (t - \lambda_r)$, which is square-free. The minimal polynomial of ϕ is a divisor of this, and therefore also square-free. Hence ϕ is semisimple. □

EXERCISE 7.3.77.
(1) Prove by induction that two commutating diagonalisable linear maps A, B on a finite-dimensional F-vector space W can be diagonalised by one and the same basis.
(2) Conclude that $A + B$ is then also diagonalisable.
(3) Prove that the sum of two commuting semisimple linear maps on a finite-dimensionsional K-vector space W is semisimple.
(4) Use this to prove that the Jordan decomposition of a linear map into commuting semisimple and nilpotent parts is unique (see also Exercise 4.3.39).

7.4. Kronecker products and eigenvalues

Returning to the discussion of tensors of linear maps, assume that U, V, W, X are finite-dimensional and that bases $(u_j)_j, (v_i)_i, (w_l)_l, (x_k)_k$ have been chosen. Let $A = (a_{ij})$ be the matrix of ϕ and let $B = (b_{kl})$ be the matrix of ψ. Then

$$(\phi \otimes \psi)(u_j \otimes w_l) = \sum_{i,k} a_{ij} b_{kl} v_i \otimes x_k,$$

so that the matrix of $\phi \otimes \psi$ equals $(a_{ij}b_{kl})_{(i,k),(j,l)}$, where the pair (i, k) plays the role of row index and the pair (j, l) plays the role of column index. This matrix is called the *Kronecker product* of A and B. By the above, its rank equals the product of the ranks of A and of B. The following proposition gives further spectral information about $A \otimes B$.

PROPOSITION 7.4.78. *Assume that $U = V$ and $W = X$, so that $\phi \otimes \psi \in L(U \otimes W)$, and that both spaces are finite-dimensional. Then the eigenvalues of $\phi \otimes \psi$ in the algebraic closure F of K, counted with multiplicities, are exactly the numbers $c_i d_j$ with c_i running through the eigenvalues of ϕ (with multiplicities) and d_j running through the eigenvalues of ψ (with multiplicities).*

PROOF. If ϕ and ψ are both diagonalisable, then this is easy: if $u_1, \ldots, u_m$ is a basis of U consisting of ϕ eigenvectors with eigenvalues $c_1, \ldots, c_m$ and $w_1, \ldots, w_n$ is an eigenbasis of ψ with eigenvalues $d_1, \ldots, d_n$, then $u_i \otimes w_j,\ i = 1, \ldots, m,\ j = 1, \ldots, n$ is an eigenbasis of $\phi \otimes \psi$.

In the general case, choose F-bases for $F \otimes U$ and $F \otimes W$ with respect to which the matrices A and B of ϕ, ψ are upper triangular. Their eigenvalues can then be read off from the diagonal. Then the Kronecker product $A \otimes B$ is also upper triangular

with respect to a suitable ordering, and its diagonal entries are exactly the products $c_i d_j$, with the right multiplicities. □

EXERCISE 7.4.79. Let $\phi \in L(U)$ and $\psi \in L(W)$ with U, W finite-dimensional of dimensions m and n, respectively. Find and prove a formula for $\det(\phi \otimes \psi)$ in terms of $\det \phi, \det \psi, m, n$.

Here is a very nice application of Kronecker products.

EXERCISE 7.4.80. A *Hadamard matrix* is an $n \times n$-matrix A with entries -1 and 1 satisfying $AA^T = nI$. An example is

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Prove that if A is an $n \times n$-Hadamard matrix and B is an $m \times m$-Hadamard matrix, then their Kronecker product $A \otimes B$, with a suitable ordering of rows and columns, is an $mn \times mn$-Hadamard matrix.

Hadamard matrices are used in design of experiments. They are conjectured to exist whenever n is a multiple of 4; unfortunately the preceding exercise gives Hadamard matrices only for a rather sparse set of natural numbers.

7.5. Complexity of matrix multiplication

Computing the product of two $n \times n$-matrices in the ordinary manner needs n^3 multiplications of scalars, as well as $n^2(n-1)$ scalar additions. In this section we disregard the additions, and sketch *Strassen's algorithm* for computing the product with $O(n^{\log_2 7})$ multiplications [**2**]. The basic observating is that the product

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix} = \begin{bmatrix} m_1 - m_2 + m_5 + m_7 & m_2 + m_4 \\ m_3 + m_5 & m_1 - m_3 + m_4 + m_6 \end{bmatrix}$$

where

$$\begin{aligned} m_1 &= (a+d)(e+h) = ae + ah + de + dh \\ m_2 &= (a+b)h = ah + bh \\ m_3 &= (c+d)e = ce + de \\ m_4 &= a(f-h) = af - ah \\ m_5 &= d(g-e) = dg - de \\ m_6 &= (c-a)(e+f) = ce + cf - ae - af \\ m_7 &= (b-d)(g+h) = bg + bh - dg - dh. \end{aligned}$$

This is true for *scalars* $a, \ldots, h$, but also for *square matrices* $a, \ldots, h$ of the same size. Thus if we let $T(n)$ denote the number of multiplications of scalars needed for multiplying two matrices, then the above shows that $T(2k) \leq 7T(k)$. As a consequence, $T(2^l) \leq 7^l$. Hence to compute the product of two $n \times n$-matrices, let l be minimal with $2^l > n$, pad the matrices with zeroes to a $2^l \times 2^l$-matrix, perform the above trick recursively, and finally drop the padded zeroes. This gives an algorithm with $O(n^{\log_2 7})$ scalar multiplications. Here $\log_2 7$ is approximately 2.807; the best known exponent (with a much less practical algorithm) is about 2.376, due to Coppersmith and Winograd [**1**].

EXERCISE 7.5.81. Let U, V, W be copies of the vector space K^4, with bases a, b, c, d, e, f, g, h, and i, j, k, l, respectively. Show that the tensor

$$a\otimes e\otimes i+b\otimes g\otimes i+a\otimes f\otimes j+b\otimes h\otimes j+c\otimes e\otimes k+d\otimes g\otimes k+c\otimes f\otimes l+d\otimes h\otimes l \in U\otimes V\otimes W$$

has rank at most 7.

Bibliography

[1] Don Coppersmith and Shmuel Winograd. Matrix multiplication via arithmetic progressions. *J. Symb. Comput.*, 9(3):251–280, 1990.

[2] Volker Strassen. Gaussian elimination is not optimal. *Numer. Math.*, 13:354–356, 1969.

CHAPTER 8

Alternating tensors

8.1. Introduction

In this chapter we discuss a certain quotient of a tensor power $V^{\otimes k} := V \otimes \cdots \otimes V$ of a vector space V. This quotient, denoted $\bigwedge^k V$, is called the k-th *alternating* or *exterior* power of V, and its elements are called *alternating tensors* (though there is some ambiguity in this latter terminology; see below). The image of a pure tensor $v_1 \otimes \cdots \otimes v_k$ in $\bigwedge^k$ is denoted $v_1 \wedge \cdots \wedge v_k$, called the *wedge product* or *alternating product* of the v_i, and sometimes still called a pure tensor. Some important things to remember are:

- The element $\omega = v_1 \wedge \cdots \wedge v_k$ is zero if and only if $v_1, \ldots, v_k$ are linearly dependent. If ω is non-zero, then a second pure tensor $w_1 \wedge \cdots \wedge w_k$ is a non-zero scalar multiple of ω if and only if the span of $v_1, \ldots, v_k$ equals that of $w_1, \ldots, w_k$. In this sense, non-zero pure tensors in $\bigwedge^k V$ (up to non-zero scalars) are in bijection with k-dimensional subspaces of V.
- There is a unique bilinear map $\bigwedge^k V \times \bigwedge^l V \to \bigwedge^{k+l} V$ mapping $(v_1 \wedge \cdots \wedge v_k, w_1 \wedge \cdots \wedge w_l) \mapsto v_1 \wedge \cdots \wedge v_k \wedge w_1 \wedge \cdots \wedge w_l$. Taking these maps together makes $\bigoplus_{k \geq 0} \bigwedge^k V$ into an associative algebra called the *Grassmann algebra* on V.
- If $\mu \in \bigwedge^k V$ and $\omega \in \bigwedge^q V$ are non-zero pure tensors representing a k-dimensional subspace $U \subseteq V$ and an l-dimensional subspace $W \subseteq V$, respectively, then their product $\mu \wedge \omega$ in the Grassman algebra is zero if and only if $U \cap W \neq \{0\}$.
- If V is finite-dimensional with basis $v_1, \ldots, v_n$, and for every subset $I = \{i_1 < \ldots < i_k\} \subseteq \{1, \ldots, n\}$ of cardinality k we set $u_I := u_{i_1} \wedge \cdots \wedge u_{i_k}$, then the u_I form a basis of $\bigwedge^k V$. In particular, the dimension of the latter space is $\binom{n}{k}$, and the dimension of the Grassmann algebra is $2^n = \sum_{k=0}^n \binom{n}{k}$.
- A linear map $\phi : V \to W$ induces linear maps $\bigwedge^k \phi : \bigwedge^k V \to \bigwedge^k W$ sending $v_1 \wedge \cdots \wedge v_k$ to $(\phi v_1) \wedge \cdots \wedge (\phi v_k)$ for every k. This holds, in particular, when $W = V$ and when $\dim V$ is finite and equal to k. In this case $\dim \bigwedge^k V = 1$ so that the linear map $\bigwedge^k \phi$ is just multiplication by an element of K. You already know this number as the *determinant* of ϕ. This is a beautiful, coordinate-free definition of the determinant of a linear map.

Proofs of these statements, and others, are below. Alternating powers turn out to be tremendously useful in applications to combinatorics.

8.2. Definition and first properties of alternating powers

Fix a vector space V and let S (for "relationS") be the subspace of $V^{\otimes k}$ spanned by all pure tensors $u_1 \otimes \cdots \otimes u_k$ in which two of the factors are equal, i.e., for which there exist $i < j$ with $u_i = u_j$. We set $\bigwedge^k V := V^{\otimes k}/S$, and we write $v_1 \wedge \cdots \wedge v_k$ for the image of a pure tensor $v_1 \otimes \cdots \otimes v_k$. By construction, the space $\bigwedge^k V$ is spanned by these k-fold *wedge products* $v_1 \wedge \cdots \wedge v_k$, and these are also called *pure alternating tensors*. General elements of $\bigwedge^k V$ are called *alternating k-tensors*.

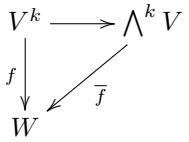
A multilinear map $f : V^k \to W$ is called *alternating* if $f(v_1, \ldots, v_k)$ is zero as soon as two of the v_i are identical. By the above, the map $V^k \to \bigwedge^k V,\ (v_1, \ldots, v_k) \mapsto v_1 \wedge \cdots \wedge v_k$ is (multilinear and) alternating.

EXERCISE 8.2.82. (1) Prove that if a multilinear map $f : V^k \to W$ is alternating, then $f(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) = -f(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k)$ for all $i < j$ and all $v_1, \ldots, v_k \in V$ (here only the arguments v_i and v_j have been swapped). Conversely, prove that if a multilinear map f has the latter property for all $i < j$, then it is alternating, provided that the characteristic of K is not 2.

(2) Prove that (regardless of the characteristic) a multilinear map $f : V^k \to W$ is alternating if and only if $f(v_1, \ldots, v_k) = 0$ whenever some *consecutive* arguments v_i, v_{i+1} coincide.

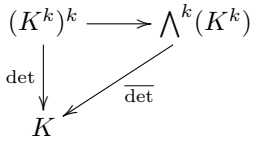
We have the following *universal property* characterising $\bigwedge^k V$. This follows from the universal property of $V^{\otimes k}$ and some reasoning analoguous to the reasoning that we used there; we omit the proof.

PROPOSITION 8.2.83. *Given any alternating k-linear map $f : V^k \to W$, there is a unique linear map $\bigwedge^k V \to W$ that makes the diagram*



commute.

EXAMPLE 8.2.84. Take $V = K^k$. Then the map $\det : V^k = (K^k)^k \to K$ sending an k-tuple $(v_1, \ldots, v_k)$ of column vectors to the determinant of the matrix with columns $v_1, \ldots, v_k$ is multilinear and alternating: if a matrix has two identical columns, then its determinant is zero. By the universal property, this means that there is a unique linear map $\overline{\det} : \bigwedge^k (K^k) \to K$ such that the diagram



commutes. Since det is not the zero map, neither is $\overline{\det}$. In particular, $\bigwedge^k K^k$ is non-zero. We will soon see that it is one-dimensional, and how fundamental this example is.

PROPOSITION 8.2.85. *Let $\phi : V \to W$ be a linear map and let k be a natural number. Then there exists a unique linear map $\bigwedge^k \phi : \bigwedge^k V \to \bigwedge^k W$ mapping $v_1 \wedge \cdots \wedge v_k$ to $(\phi v_1) \wedge \cdots \wedge (\phi v_k)$.*

This follows immediately from the universal property applied to the alternating multilinear map sending $(v_1, \ldots, v_k)$ to $(\phi v_1) \wedge \cdots \wedge (\phi v_k)$. A useful spacial case is that where V is a subspace of W and ϕ is the inclusion map. By the following exercise, this means that we can view $\bigwedge^k V$ as a subspace of $\bigwedge^k W$.

EXERCISE 8.2.86. Prove that

(1) if $\psi : W \to U$ is a second linear map, then $\bigwedge^k(\psi \circ \phi) = (\bigwedge^k \psi) \circ (\bigwedge^k \phi)$;
(2) if ϕ is injective, then so is $\bigwedge^k \phi$;
(3) if ϕ is surjective, then so is $\bigwedge^k \phi$; and
(4) if ϕ is invertible, then so is $\bigwedge^k \phi$, with inverse $\bigwedge^k(\phi^{-1})$.

Let B be a basis of V and assume that we have fixed a linear order $<$ on B. By Theorem 6.1.54 the tensors $v_1 \otimes \cdots \otimes v_k$ with all v_i running through B form a basis of $V^{\otimes k}$. As a consequence, their images $v_1 \wedge \cdots \wedge v_k$ span $\bigwedge^k V$. But they are not a basis (unless $k \leq 1$), because permuting two factors only multiplies the alternating product by -1, and when two v_i are equal, the image of $v_1 \otimes \cdots \otimes v_k$ in $\bigwedge^k V$ is zero. This leads to the following theorem.

THEOREM 8.2.87. *The pure tensors $v_1 \wedge \cdots \wedge v_k$ with $v_1, \ldots, v_k \in B$ and $v_1 < \ldots < v_k$ form a basis of $\bigwedge^k V$.*

We will refer to this basis as the "standard basis" of $\bigwedge^k V$ corresponding to B. If $|B| = n$ is finite, we will often label its elements by the indices $1, \ldots, n$ and use double indices ($v_{i_1} \wedge \cdots \wedge v_{i_k}$ with $i_1 < \ldots < i_k$).

PROOF. By the above reasoning, these alternating products span $\bigwedge^k V$. To prove linear independence, let $v_1 < \ldots < v_k$ be elements of B. We will construct a linear map with domain $\bigwedge^k V$ that is non-zero on $v_1 \wedge \ldots \wedge v_k$ and zero on all other proposed basis elements of $\bigwedge^k V$. For this let ϕ be the linear map from V to K^k that maps v_i to e_i and all elements of $B \setminus \{v_1, \ldots, v_k\}$ to 0. Then $\bigwedge^k \phi$ maps $v_1 \wedge \cdots \wedge v_k$ to $e_1 \wedge \cdots \wedge e_k$, and all other proposed basis elements to zero. Now $e_1 \wedge \cdots \wedge e_k$ spans $\bigwedge^k K^k$, and we have seen above that the latter space is non-zero. Hence $e_1 \wedge \cdots \wedge e_k$ is non-zero, as desired. □

This theorem immediately implies that if $\dim V =: n$ is finite, then $\dim \bigwedge^k V$ equals $\binom{n}{k}$. In particular, this dimension is 1 for $k = 0$, then grows with k until k reaches $\lfloor \frac{n}{2} \rfloor$, stays the same at $\lceil \frac{n}{2} \rceil$, and decreases all the way down to 1 for $k = n$ and 0 for $k > n$.

8.3. Duals and quotients

Taking duals and alternating powers almost commute. More precisely, there is a natural linear map $\bigwedge^k(V^*) \mapsto (\bigwedge^k V)^*$, determined by sending $x_1 \wedge \cdots \wedge x_k$ to the

linear function on $\bigwedge^k V$ determined by

$$v_1 \wedge \cdots \wedge v_k \mapsto \sum_{\pi \in S_k} \operatorname{sgn}(\pi) x_1(v_{\pi(1)}) \cdots x_k(v_{\pi(k)}).$$

Note that the right-hand side is alternating and k-linear in $(v_1, \ldots, v_k)$, so that such a linear function exists by the universal property of $\bigwedge^k V$. Moreover, the map sending $(x_1, \ldots, x_k)$ to that linear function is itself alternating and k-linear, which guarantees the existence of a unique linear map $\bigwedge^k (V^*) \to (\bigwedge^k V)^*$ as just described.

PROPOSITION 8.3.88. *The map just described is injective. If V is finite-dimensional, then it is also surjective.*

PROOF. We first prove this for finite-dimensional V, and then reduce the general case to this case. For finite-dimensional V, let $v_1, \ldots, v_n$ be a basis of V, and let $x_1, \ldots, x_n$ be the dual basis of V^*. Then the standard basis element $x_{i_1} \wedge \cdots \wedge x_{i_k} \in \bigwedge^k (V^*)$ $(i_1 < \ldots < i_k)$ gets mapped to the linear function that maps the standard basis element $v_{j_1} \wedge \ldots \wedge v_{j_k}$ $(j_1 < \ldots < j_k)$ to 1 if each j_l equals the corresponding i_l and to 0 otherwise. In other words, the standard basis of $\bigwedge^k (V^*)$ is mapped to the basis dual to the standard basis of $\bigwedge^k V$. In particular, the linear map is an isomorphism.

In the general case, let $\omega \in \bigwedge^k (V^*)$ be a tensor mapped to zero. Since ω is a finite linear combination of pure tensors, it already lives in $\bigwedge^k U$ for some finite-dimensional subspace U of V^*. Let W be the subspace of all $v \in V$ with $x(v) = 0$ for all $x \in U$. Then U is canonically isomorphic to $(V/W)^*$, and we have the commutative diagram

$$\begin{array}{ccc} \bigwedge^k U & \longrightarrow & (\bigwedge^k (V/W))^* \\ \downarrow & & \downarrow \\ \bigwedge^k V^* & \longrightarrow & (\bigwedge^k V)^* \end{array}$$

where the vertical maps are inclusions (or at least canonical and injective: the right-most one is the dual of the surjective linear map $\bigwedge^k V \to \bigwedge^k (V/W)$ coming from the surjection $V \to V/W$) and where upper map is injective by the previous, finite-dimensional case. By assumption, $\omega \in \bigwedge^k U$ is mapped to zero when first the left-most vertical map and then the bottom map is applied. But then it is also mapped to zero by the composition of the other two maps. Since both of these are injective, ω is zero. □

EXERCISE 8.3.89. Assume that V is infinite-dimensional.

(1) Prove that, in the above construction, every element of $\bigwedge^k V^*$ is mapped to a linear function on $\bigwedge^k V$ that factorises through the surjective map $\bigwedge^k V \to \bigwedge^k (V/U)$ and a linear function $\bigwedge^k (V/U) \to K$ for some subspace U of finite codimension in V. [Note that the kernel of $\bigwedge^k V \to \bigwedge^k (V/U)$ is spanned by pure tensors of the form $u \wedge v_2 \wedge \cdots \wedge v_k$ with $u \in U$ and $v_2, \ldots, v_k \in V$. This kernel contains, but is typically larger than $\bigwedge^k U$.]

(2) Prove that, conversely, every linear function on $\bigwedge^k V$ that has such a factorisation for some subspace U of finite codimension in V is the image of some element of $\bigwedge^k(V^*)$. [Apply the finite-dimensional part of the proposition to V/U.]
(3) Assume that V is countable-dimensional, with basis $v_1, v_2, \ldots$. Let f be the linear function on $\bigwedge^2 V$ that maps the standard basis elements $v_{2i-1} \wedge v_{2i}$, $i = 1, 2, \ldots$ to 1 and all other standard basis elements to zero. Prove that there does not exist a non-zero vector $u \in V$ such that $f(u \wedge v) = 0$ for all $v \in V$. Conclude that f does not lie in the image of the map $\bigwedge^2(V^*) \to (\bigwedge^2 V)^*$.

8.4. Grassmann algebra and the Plücker embedding

There is a well-defined bilinear map $\bigwedge^k V \times \bigwedge^l V \to \bigwedge^{k+l} V$ mapping $(v_1 \wedge \cdots \wedge v_k, w_1 \wedge \cdots \wedge w_l) \mapsto v_1 \wedge \cdots \wedge v_k \wedge w_1 \wedge \cdots \wedge w_l$—indeed, for fixed $w_1, \ldots, w_l$ the right-hand side is k-linear and alternating in $v_1, \ldots, v_k$ and vice versa, hence the existence of such a bilinear map follows from the universal properties of both alternating powers. By convenient abuse of notation, we write $\omega \wedge \mu$ for the image of (ω, μ) under this map.

Taking these maps together makes $\bigwedge V := \bigoplus_{k \geq 0} \bigwedge^k V$ into an associative algebra: the product of $\omega \in \bigwedge^k V$ and $\mu \in \bigwedge^l V$ is $\omega \wedge \mu$, and this product is extended linearly. If V is finite-dimensional of dimension n, then $\bigwedge^k V = 0$ for $k > n$, and we find that $\bigwedge V$ has dimension 2^n.

We will see another beautiful application of the Grassmann algebra in Chapter 9, but for now we will focus on how we can use it to compute efficiently with finite-dimensional subspaces of a vector space V. For this let U be a k-dimensional subspace of V. The inclusion $U \to V$ gives an inclusion $\bigwedge^k U \to \bigwedge^k V$. The first of these vector spaces has dimension one, and is spanned by any pure tensor $\omega = u_1 \wedge \cdots \wedge u_k$ with $u_1, \ldots, u_k$ a basis of U. (In particular, the alternating product of any other basis of U is scalar multiple of ω.) If W is a k-dimensional subspace of V different from U, then we claim that $\bigwedge^k W$ is a one-dimensional subspace of $\bigwedge^k V$ different from $\bigwedge^k U$. This is almost immediate: choose a basis $v_1, \ldots, v_n$ of V such that $v_1, \ldots, v_k$ form a basis of U and $v_m, \ldots, v_{m+k-1}$ for some $m \leq k+1$ form a basis of W (for this one typically first chooses a (possibly empty) basis $v_m, \ldots, v_k$ of their intersection). The dis-equality $W \neq U$ implies $m \neq 1$. But then $\bigwedge^k U, \bigwedge^k W$ are spanned by the distinct standard basis elements $v_1 \wedge \cdots \wedge v_k$ and $v_m \wedge \cdots v_{m+k-1}$ of $\bigwedge^k V$, respectively. Hence they are not equal. This proves the following proposition.

PROPOSITION 8.4.90. *The map $U \mapsto \bigwedge^k U$ from the set of k-dimensional subspaces of V to the set of one-dimensional subspaces of $\bigwedge^k V$ is an embedding. Its image consists of those one-dimensional subspaces that are spanned by pure tensors.*

This embedding is called the *Plücker embedding*. In the following exercise you get acquainted with the smallest interesting example.

EXERCISE 8.4.91. Let $V = K^4$ thought of as row vectors, and let $k = 2$. Let a_1, a_2 be elements of K^4 spanning a two-dimensional subspace U, and let $A = (a_{ij})$ be the 2×4-matrix with rows a_1, a_2, respectively.

(1) Verify that the coefficient in $a_1 \wedge a_2$ of the standard basis element $e_j \wedge e_l$ $(1 \leq j < l \leq 4)$ equals the determinant $d_{jl} := a_{1j}a_{2l} - a_{2j}a_{1l}$.
(2) What happens with these determinants when A is multiplied from the left with an invertible 2×2-matrix g? (This corresponds to a change of basis in U.)
(3) Verify by computation that the determinants d_{jl} satisfy the following fundamental relation:
$$d_{12}d_{34} - d_{13}d_{24} + d_{14}d_{23} = 0.$$
(This is called a *Grassmann-Plücker relation*.)
(4) Conversely, given six elements $d_{jl} \in K$ $(1 \leq j < l \leq 4)$ satisfying the previous equation (and not all zero), show that there exists a unique subspace U having some basis a_1, a_2 such that $d_{jl} = a_{1j}a_{2l} - a_{2j}a_{1l}$ for all $1 \leq j < l \leq 4$. (Hint: argue first that you may assume that A contains two columns that contain an identity matrix, and then reconstruct the remaining entries of A.)

This exercise shows that *the set of two-dimensional subspaces of a four-dimensional space "is" the set of those one-dimensional subspaces of a certain six-dimensional space that satisfy a certain quadratic equation*. This "quadratic hypersurface" (strictly speaking, in projective five-space) is called the *Klein quadric*.

We conclude this chapter with a proposition relating products in the Grassmann algebra with intersections of subspaces.

PROPOSITION 8.4.92. *Let $\omega \in \bigwedge^k V$ and $\mu \in \bigwedge^l V$ be non-zero pure tensors representing subspaces U and W of V, respectively. Then $U \cap W = \{0\}$ if and only if $\omega \wedge \mu \neq 0$.*

PROOF. If $U \cap W = \{0\}$, then there exists a basis $v_1, \ldots, v_n$ of V such that $v_1, \ldots, v_k$ is a basis of U and $v_{k+1}, \ldots, v_{k+l}$ is a basis of W. Then ω is a non-zero scalar multiple of $v_1 \wedge \cdots \wedge v_k$ and μ is a non-zero scalar multiple of $v_{k+1} \wedge \cdots \wedge v_{k+l}$, hence $\omega \wedge \mu$ is a non-zero scalar multiple of $v_1 \wedge \cdots \wedge v_{k+l}$. Since the latter vector is part of the standard basis of $\bigwedge^{k+l} V$, it is not zero, hence neither is $\omega \wedge \mu$.

Conversely, if $v \in U \cap W$ is a non-zero vector, then for suitable $u_2, \ldots, u_k$ we have $\omega = v \wedge u_2 \wedge \cdots \wedge u_k$ and for suitable $w_2, \ldots, w_l$ we have $\mu = v \wedge w_2 \wedge \cdots \wedge w_l$. Then $\omega \wedge \mu$ is the alternating product of $v, u_2, \ldots, u_k, v, w_2, \ldots, w_l$, hence zero since v appears twice. □

EXERCISE 8.4.93. Take $V = K^n$. Let ω be a non-zero element of $\bigoplus_{k=1}^n \bigwedge^k V$, where we have excluded the summand $\bigwedge^0 V = K$.

(1) Prove that there exists an $m > 1$ for which ω^m, where the power is taken in the Grassmann algebra, is zero, and ω^{m-1} is not yet zero.
(2) Find such an ω for which m of the preceding part is maximal. (Hint: pure non-zero tensors will have $m = 2$, which is certainly not maximal.)

CHAPTER 9

Applications of alternating tensors

In this short chapter we discuss two beautiful applications of alternating tensors.

9.1. Bollobás's theorem

The first application is Lovász's proof of a 1965 theorem due to Bollobás; this part of the chapter is based on [**1**, Chapters 5 and 6]. In this chapter, a *simple graph* is understood to be undirected and without loops or multiple edges. First consider the following elementary statement.

EXERCISE 9.1.94. If in a simple graph every collection of at most three edges has a common vertex, then the graph has a vertex that lies in all edges.

There are various ways to generalise this statement; we will use the following terminology. An *r-uniform set system* on a set X is a collection $\mathcal{F}$ of cardinality-r subsets of X. Thus a simple graph on X is a 2-uniform set system. A subset S of X is said to *cover* $\mathcal{F}$ if every element $\mathcal{F}$ has non-empty intersection with S. Thus a single vertex contained in all edges of a graph covers all edges of that graph. Now the elementary observation above can be generalised by considering covering sets with $s > 1$ and/or r-uniform set systems with $r > 2$. Here is the first generalisation.

THEOREM 9.1.95 (Erdös-Hajnal-Moon, 1964). *Let s be a natural number, and let G be a simple graph on a set X. If every family of at most $\binom{s+2}{2}$ edges of G is covered by some cardinality-s subset of X, then there is a cardinality-s subset of X that covers* all *edges.*

Note that the special case with $s = 1$ is the previous exercise.

EXERCISE 9.1.96. Give an example of a graph, depending on s, that shows that the statement would be incorrect if $\binom{s+2}{2}$ were replaced by $\binom{s+2}{2} - 1$.

The following theorem generalises the previous one to larger r.

THEOREM 9.1.97 (Bollobás, 1965). *Let $\mathcal{F}$ be an r-uniform set system on a set X and let s be a natural number. If every subset of $\mathcal{F}$ of cardinality at most $\binom{s+r}{r}$ is covered by some cardinality-s subset of X, then all of $\mathcal{F}$ is covered by some cardinality-s subset of X.*

EXERCISE 9.1.98. Give an example of an r-uniform set, depending on r and s, that shows that the statement would be incorrect if $\binom{s+r}{r}$ were replaced by $\binom{s+r}{r} - 1$.

PROOF OF BOLLOBÁS'S THEOREM DUE TO LÓVÁSZ. We will prove this for finite X; below is an exercise that reduces the case of infinite X to this case. Assume

that $\mathcal{F}$ is *not* covered by any cardinality-s subset of X. Then $\mathcal{F}$ has some *minimal* sub-family $\mathcal{F}'$ (with respect to inclusion) with the property that $\mathcal{F}'$ is not covered by any cardinality-s subset of X. To arrive at a contradiction, it suffices to prove that $|\mathcal{F}'| \leq \binom{r+s}{r}$. Let $\mathcal{F}' = \{A_1, \ldots, A_m\}$, where the A_i are distinct cardinality-r subsets of X. By minimality of $\mathcal{F}'$, for every $i = 1, \ldots, m$ the sub-family $\mathcal{F}' \setminus \{A_i\}$ *is* covered by some cardinality-s set B_i. We thus have

$$A_i \cap B_j \begin{cases} \neq \emptyset \text{ if } i \neq j, \text{ and} \\ = \emptyset \text{ if } i = j. \end{cases}$$

Now comes Lovász's beautiful trick. Let K be a sufficiently large field (we will see below what cardinality suffices) and let $V = KX$ be the vector space formally spanned by X. For $i, j = 1, \ldots, m$ let $U_i \subseteq V$ be the r-dimensional subspace spanned by A_i and let W_j be the s-dimensional subspace spanned by B_j. Then, by the above, we have

$$U_i \cap W_j \begin{cases} \neq \{0\} \text{ if } i \neq j, \text{ and} \\ = \{0\} \text{ if } i = j. \end{cases}$$

Now let $\phi : V \to K^{r+s}$ be a linear map that maps each U_i to an r-dimensional subspace of K^{r+s} and each W_j to an s-dimensional subspace of K^{r+s}, and that satisfies $\phi(U_i) \cap \phi(W_i) = \{0\}$ for all i. (Below you will prove that such a linear map exists when the cardinality of K is sufficiently large.)

Let μ_i span the one-dimensional subspace $\bigwedge^r \phi(U_i) \subseteq \bigwedge^r K^{r+s}$ and let λ_i span the one-dimensional subspace $\bigwedge^s \phi(W_i) \subseteq \bigwedge^s K^{r+s}$. Then, by the above and by properties of the alternating product proved in the previous chapter, we have

$$\mu_i \wedge \lambda_j \begin{cases} = 0 \text{ if } i \neq j, \text{ and} \\ \neq 0 \text{ if } i = j; \end{cases}$$

here first case follows from the fact that ϕ maps the non-zero space $U_i \cap W_j,\ i \neq j$ into a non-zero common subspace $\phi(U_i \cap W_j)$ of $\phi(U_i)$ and $\phi(W_j)$, and the second case follows from the requirement that ϕ "keep U_i and W_i apart".

We claim that $\mu_1, \ldots, \mu_m$ are linearly independent (and so are $\lambda_1, \ldots, \lambda_m$). Indeed, if $\sum_i c_i \mu_i = 0$, then by the above we find that, for each j,

$$0 = (\sum_i c_i \mu_i) \wedge \lambda_j = c_j \mu_j \wedge \lambda_j,$$

and $c_j = 0$ since $\mu_j \wedge \lambda_j$ is non-zero. Since the μ_i are linearly independent, m is at most the dimension of $\bigwedge^r K^{r+s}$, which equals $\binom{r+s}{r}$. □

Exercise 9.1.99.

(1) Prove that, for ϕ to satisfy the conditions in the proof, it suffices that $\ker \phi$ does not intersect any of the $(r+s)$-dimensional subspaces $U_i \oplus W_i,\ i = 1, \ldots, m$ of V.
(2) Let M be the $(r+s) \times |X|$-matrix of ϕ with respect to the standard basis of K^{r+s} and the basis X of V. Show that the condition in the previous exercise is equivalent to the condition that the $(r+s) \times (r+s)$ matrix obtained by taking the columns of M corresponding to $A_i \cup B_i$ has non-zero determinant.
(3) Prove the existence of a matrix M with the property in the previous part, provided that $|K| \geq |X|$. (Hint: Vandermonde matrices.)

EXERCISE 9.1.100. Reduce the case of infinite X to the case of finite X.

9.2. Unique coclique extension property

This section contains an application of exterior powers to certain graphs whose study is popular among Eindhoven combinatorialists. Recall that a *coclique* in a (simple) graph is a set of vertices that are pairwise *not* connected.

Fix natural numbers n and $k \leq n/2$. Then the *Kneser graph* $\mathcal{K}(n,k)$ has as vertex set the collection of all cardinality-k subsets of $[n] := \{1,\ldots,n\}$, and two are connected by an edge if they are disjoint. Thus cocliques in the Kneser graph are k-uniform set systems on $[n]$ that pairwise intersect. By a famous theorem of Erdös-Ko-Rado (with a beautiful proof due to Katona) the the maximal size of a coclique is $\binom{n-1}{k-1}$, which is attained by the set of all cardinality-k subsets containing 1. If $k = n/2$, then this maximal size of a coclique $\mathcal{C}$ is $\frac{1}{2}\binom{n}{k}$, and there are many cocliques attaining this bound, all of the following form: for every $n/2$-subset A of $[n]$, $\mathcal{C}$ contains exactly one of the sets A and $[n] \setminus A$.

Next we turn our attention to q-analogues. Fix a field F. Then the *F-Kneser graph* $\mathcal{K}_F(n,k)$ has as vertex set the collection of all k-dimensional subspaces of F^n, and two subspaces U, V are connected if $U \cap V = \{0\}$.

A coclique $\mathcal{C}$ in $\mathcal{K}(n,k)$ can be "thickened" to a subset $\mathcal{C}_F$ of $\mathcal{K}_F(n,k)$ defined as the set of all k-dimensional subspaces U of F^n such that $U \cap \langle \{e_i \mid i \in A\}\rangle \neq \{0\}$ for all $A \in \mathcal{C}$. In words: to every set $A \in \mathcal{C}$ we associate the vector space U_A spanned by all standard basis vectors with index in A, and then $\mathcal{C}_F$ is defined as the set of all k-dimensional subspaces that are *not* connected, in $\mathcal{K}_F(n,k)$, to *any* of the spaces U_A, $A \in \mathcal{C}$. Note that, since $\mathcal{C}$ is a coclique, the U_A themselves are elements of $\mathcal{C}$. The following theorem gives a sufficient condition for $\mathcal{C}_F$ to be a coclique in $\mathcal{K}_F(n,k)$.

THEOREM 9.2.101. *Suppose that $\mathcal{C}$ is* maximal *(with respect to inclusion) among all cocliques in the ordinary Kneser graph $\mathcal{K}(k,n)$. Then $\mathcal{C}_F$ is a (neccessarily maximal) coclique in the F-Kneser graph $\mathcal{K}(k,n,F)$.*

EXAMPLE 9.2.102. Let $n = 5$ and $k = 2$ and let $\mathcal{C} = \{\{1,2\},\{1,3\},\{2,3\}\}$. This is a maximal coclique in the ordinary Kneser graph (though not of maximal cardinality). Then the set $\mathcal{C}_F$ contains the space $\langle e_1 + e_2, e_2 - e_3\rangle$ (which has non-trivial intersection with each of $\langle e_1, e_2\rangle, \langle e_1, e_3\rangle, \langle e_2, e_3\rangle$) but not, for example, the space $\langle e_1, e_2 + e_4\rangle$ (which only has non-trivial intersection with the first two).

PROOF. We will prove only the case where $n = 2k$; another special case of the general case will be among the exercises. We have to prove that for any $U, W \in \mathcal{C}_F$ we have $U \cap W \neq \{0\}$. Let $\mu \in \bigwedge^k F^n$ be the pure tensor (unique up to scalars) representing U and let $\lambda \in \bigwedge^k F^n$ be the pure tensor representing W. Then we have to prove that $\mu \wedge \lambda = 0$. To see this, expand μ on the standard basis $\{e_{i_1} \wedge \cdots \wedge e_{i_k} \mid 1 \leq i_1 < \ldots < i_k \leq n\}$ of $\bigwedge^k K^n$. Let $A = \{j_1 < \ldots < j_k\}$ be a cardinality-k set not in $\mathcal{C}$. Then by maximality of $\mathcal{C}$ the complement $\{i_1 < \ldots < i_k\}$ of A does lie in $\mathcal{C}$. By definition of $\mathcal{C}_F$ we find that U intersects U_A, which means that

$$\mu \wedge (e_{i_1} \wedge \cdots \wedge e_{i_k}) = 0.$$

Now a standard basis vector of $\bigwedge^k K^n$ has alternating product zero with $e_{i_1}\wedge\cdots\wedge e_{i_k}$ unless it equals $e_{j_1}\wedge\cdots\wedge e_{j_k}$, in which case it is $\pm e_1\wedge\cdots\wedge e_n\neq 0$. Hence the coefficient in μ of the $e_{j_1}\wedge\cdots\wedge e_{j_k}$ is zero, as claimed. This proves that μ lies in the span of the standard basis vectors corresponding to elements of $\mathcal{C}$. The same holds for λ. As $\mathcal{C}$ is a coclique in the ordinary Kneser graph, the alternating product of any two standard basis vectors corresponding to elements of $\mathcal{C}$ is zero. Hence $\mu\wedge\lambda=0$, as well. This means that $U\cap W\neq\{0\}$, as required. We have thus proved that $\mathcal{C}_F$ is a coclique. That it is maximal follows from the fact that we have added to the $U_A,\ A\in\mathcal{C}$ *all* subspaces that are not connected to any of these. □

Exercise 9.2.103. Let n be arbitrary and set $k:=2$.

(1) Describe all (inclusion-)maximal cocliques in the Kneser graph $\mathcal{K}(n,k)$ (so *not* only those of maximal cardinality!).
(2) Prove the theorem in this case.

Exercise 9.2.104. In this exercise we prove the Erdös-Ko-Rado theorem, as well as its $\mathbb{F}_q$-analogue, under the additional assumption that k divides n. First let $\mathcal{C}$ be a coclique in the ordinary Kneser graph $\mathcal{K}(k,n)$.

(1) Argue that in any ordered partition $(A_1,\ldots,A_{n/k})$ of $[n]$ into n/k parts of cardinality k at most one of the parts A_i is in $\mathcal{C}$.
(2) Prove $\mathcal{C}$ contains at most $1/(\frac{n}{k})=\frac{k}{n}$ times the total number of k-sets (which is $\binom{n}{k}$). (Hint: consider pairs $((A_1,\ldots,A_{n/k}),A)$ where $(A_1,\ldots,A_{n/k})$ is a partition of $[n]$ into parts of size k and where A is an element of $\mathcal{C}$ that appears among the A_i; count such pairs in two different ways.)

Next let $\mathcal{C}$ be a coclique in the $\mathbb{F}_q$-Kneser graph $\mathcal{K}(n,k,\mathbb{F}_q)$. Recall that $\mathbb{F}_q$ has a field extension $\mathbb{F}_{q^k}$, that the $\frac{n}{k}$-dimensional vector space $(\mathbb{F}_{q^k})^{n/k}$ is n-dimensional when regarded as a vector space over $\mathbb{F}_q$, and that a one-dimension subspace of $(\mathbb{F}_{q^k})^{n/k}$ is k-dimensional over $\mathbb{F}_q$.

(3) Fix any $\mathbb{F}_q$-linear isomorphism $\phi:(\mathbb{F}_q)^n\to(\mathbb{F}_{q^k})^{n/k}$ (this is the $\mathbb{F}_q$-analogue of the ordered partition above). Prove that $\mathcal{C}$ contains at most one element U (a k-dimensional subspace of $(\mathbb{F}_q)^n$) such that $\phi(U)$ is a one-dimensional subspace over $\mathbb{F}_{q^k}$.
(4) Prove that $\mathcal{C}$ contains at most $1/(\frac{q^n-1}{q^k-1})$ times the total number of k-dimensional subspaces of $\mathbb{F}_q^n$, and that this gives the upper bound $\frac{(q^{n-1}-1)\cdots(q^{n-k+1}-1)}{(q^{k-1}-1)\cdots(q-1)}$ on the cardinality of $\mathcal{C}$.

Bibliography

[1] László Babai and Péter Frankl. *Linear Algebra Methods in Combinatorics with Applications to Geometry and Computer Science*. Department of Computer Science, The University of Chicago, Chicago, 1992. available at `Library Genesis`.

CHAPTER 10

Symmetric bilinear forms

We already know what a *bilinear form* β on a K-vector space V is: it is a function $\beta : V \times V \to K$ that satisfies $\beta(u+v,w) = \beta(u,w) + \beta(v,w)$ and $\beta(u,v+w) = \beta(u,v) + \beta(u,w)$ and $\beta(cv,w) = \beta(v,cw) = c\beta(v,w)$ for all $u,v,w \in V$ and $c \in K$. In this chapter we study such forms that are, moreover, *symmetric* in the sense that $\beta(v,w) = \beta(w,v)$ holds for all $v,w \in V$.

10.1. Gram matrix

Linear combinations of symmetric bilinear forms on V are again symmetric bilinear forms. Hence these forms form a vector sub-space of the space $(V \otimes V)^*$ of all bilinear forms (remind yourself why the space of bilinear forms is canonically the same as the dual of $V \otimes V$!). Together with a basis $(v_i)_{i\in I}$ of V, the bilinear form determines the *Gram matrix* $(\beta(v_i,v_j))_{i,j\in I}$ of β with respect to the v_i. This is a symmetric matrix. Conversely, given any symmetric matrix $A = (a_{ij})_{i,j\in I}$ with rows and columns labelled by I, there is a unique symmetric bilinear form with Gram matrix A with respect to the v_i. This form is defined by

$$\beta(\sum_{i\in I} c_i v_i, \sum_{j\in I} d_j v_j) = \sum_{i\in I, j\in I} c_i a_{ij} d_j = c^T A d,$$

where in the last equality we think of c, d as a column vectors (with only finitely many non-zero entries). In particular, if $|I|$ is finite and equal to n, then a choice of basis gives rise to a linear isomorphism between the space of symmetric bilinear forms and the space of symmetric $n \times n$-matrices, which has dimension $\binom{n+1}{2} = n + \binom{n}{2}$.

10.2. Radicals, orthogonal complements, and non-degenerate forms

Let β be a symmetric bilinear form on V. Two vectors v, w are called *perpendicular* or *orthogonal* (with respect to β) if $\beta(v,w) = 0$. For a subspace (or even a subset) $U \subseteq V$ we write $U^\perp$ for the set of all $v \in V$ such that $\beta(v,U) = \{0\}$ (which, by symmetry of the form, is equivalent to $\beta(U,v) = \{0\}$). Then $U^\perp$ is called the *orthogonal complement* or *orthoplement* of U, although it need not be a vector space complement (for this reason, some people prefer the term *perp*).

EXAMPLE 10.2.105. (1) Let $K = \mathbb{F}_2$, $V = K^n$ and $\beta(v,w) := \sum_{i=1}^n v_i w_i$. Assume that n is even. Then the vector $(1,\dots,1)$ is perpendicular to itself (a non-zero vector with this property is called *isotropic*), hence it is contained in its own orthoplement.

(2) This is not just a positive-characteristic phenomenon; the same happens with $V = \mathbb{C}^2$ and $\beta(v, w) = v_1w_1+v_2w_2$. Then the vector $(1, i)$ is isotropic. (Note that $v_1\overline{w_1}+v_2\overline{w_2}$ is *not* linear in w! We'll get back to such *Hermitian forms* later.)
(3) Even on the real vector space $V = \mathbb{R}^2$ with $\beta(v, w) = v_1w_2+v_2w_1$ isotropic vectors, such as $(1, 0)$, exist.
(4) But of course in $V = \mathbb{R}^n$ with $\beta(v, w) = \sum_i v_iw_i$ no isotropic vectors exist.

The *radical* $\operatorname{rad}\beta$ of β is defined as the orthoplement $V^\perp$ of V itself. The form is called *non-degenerate* if $\operatorname{rad}\beta = \{0\}$.

Exercise 10.2.106. Verify that, given a basis $(v_i)_{i\in I}$ of V, the radical of β consists of all linear combinations $\sum_i c_iv_i$ such that the column vector c lies in the kernel of the Gram matrix of β with respect to the v_i.

The following lemma shows how to "get rid of a radical".

Lemma 10.2.107. *The form β induces a unique, well-defined bilinear form $\overline{\beta}$ on $V/\operatorname{rad}\beta$ that makes the diagram*

$$\begin{array}{ccc} V \times V & \xrightarrow{\beta} & K \\ {\scriptstyle \pi\times\pi}\downarrow & \nearrow_{\overline{\beta}} & \\ (V/\operatorname{rad}\beta) \times (V/\operatorname{rad}\beta), & & \end{array}$$

commute (here $\pi : V \to V/\operatorname{rad}\beta$ is the natural projection).

Proof. Uniqueness follows from surjectivity of π. Existence (or "well-definedness") follows from the fact that $\beta(u, v)$ equals $\beta(u', v')$ for all $u' \in u + \operatorname{rad}\beta$ and $v' \in v + \operatorname{rad}\beta$. □

The dimension of $V/\operatorname{rad}\beta$ is also called the *rank* of β.

Exercise 10.2.108. Show that, if V is finite-dimensional, then the rank of β is the rank of the Gram matrix of β with respect to any basis of V.

We now collect a number of facts about non-degenerate symmetric bilinear forms.

Lemma 10.2.109. *If β is a non-degenerate symmetric bilinear form on a finite-dimensional vector space V, then the map $V \to V^*$, $v \mapsto \beta(v, .)$ is a linear isomorphism.*

Proof. An element of the kernel of this map is perpendicular to all of V, and hence 0 since β is non-degenerate. This shows that the map is injective. It is also surjective since $\dim V = \dim V^*$ for finite-dimensional V. □

Although an orthoplement is not always a complement to a vector space, it does have the right dimension, provided that the form is non-degenerate.

Lemma 10.2.110. *Let β be a non-degenerate symmetric bilinear form on a finite-dimensional vector space V, and let U be a subspace of V. Then* $\dim U + \dim U^\perp = \dim V$.

PROOF. Let $\phi : V \to V^*$ be the isomorphism $v \mapsto \beta(v, .)$, and let $\pi : V^* \to U^*$ be the (surjective) restriction map of linear functions (dual to the inclusion $U \to V$). Now $U^\perp$ is the kernel of $\pi \circ \phi$. By the dimension theorem, we have $\dim U^\perp + \dim \operatorname{im}(\pi \circ \phi) = \dim V$. As π and ϕ are surjective, so is $\pi \circ \phi$, so the dimension of the image equals $\dim U^* = \dim U$. □

EXERCISE 10.2.111. Use the previous lemma to prove that for any symmetric bilinear form β on a finite-dimensional vector space V and for any subspace U of V we have $(U^\perp)^\perp = U$.

EXERCISE 10.2.112. The n inhabitants of Odd-Town enjoy forming clubs. To avoid the foundation of all 2^n possible clubs (all subsidised by the government, of course), the government of the country to which Odd-Town belongs has imposed the following two restrictions on clubs:

(1) the number of members of every club should be *odd*; and
(2) the number of common members of any two distinct clubs should be *even*.

In particular, these rules implies that two distinct clubs cannot have exactly the same members.

(1) What is the maximal number of clubs that the inhabitants of Odd-Town can form? (Hint: represent the clubs by vectors over $\mathbb{F}_2$.)
(2) Let N be that maximal number. Can any collection of less than N clubs satisfying the rules be extended to a collection of N clubs satisfying the rules?

EXERCISE 10.2.113. In Reverse-Odd-Town, which belongs to the same country as Odd-Town, a local administrator has miscopied the nation-wide rules mentioned before. Indeed, he has interchanged the words *even* and *odd*! Answer the same questions for Reverse-Odd-Town. (This time, you may have to make a case distinction between even and odd n.)

10.3. Group action, equivalence

Our old friend $\mathrm{GL}(V)$ acts on the space of symmetric bilinear forms by

$$(g\beta)(v, w) := \beta(g^{-1}v, g^{-1}w).$$

Two symmetric bilinear forms β, γ on V are called *equivalent* if they are in the same $\mathrm{GL}(V)$-orbit. The rest of this chapter will be concerned with describing the equivalence classes over easy fields. Note, first of all, that the dimension (and the co-dimension) of the radical are invariants of symmetric bilinear forms under this group action.

LEMMA 10.3.114. *Let $(v_i)_{i\in I}$ be a basis of V and let β be a symmetric bilinear form on V. A second symmetric bilinear form γ on V is equivalent to β if and only if there exists a basis $(w_i)_{i\in I}$ of V such that the Gram matrix of β with respect to the v_i equals the Gram matrix of γ with respect to the w_i.*

PROOF. First, assume that $g\beta = \gamma$. This means that $\gamma(gv_i, gv_j) = \beta(v_i, v_j)$ for all i, j, hence if we set $w_i := gv_i$, then the Gram matrix of γ with respect to the w_i equals the Gram matrix of β with respect to the v_i. For the converse, assume

that $(w_i)_{i\in I}$ is a basis with the required property. There is a (unique) $g \in \mathrm{GL}(V)$ satisfying $gv_i = w_i$. Then $g\beta = \gamma$. □

The proof of this lemma is worth working out in detail for $V = K^n$. Let β be a symmetric bilinear form on V and let A be the Gram matrix of β with respect to the the standard basis. Also let $g \in \mathrm{GL}_n(K) = \mathrm{GL}(V)$. Then the Gram matrix B of $g\beta$ with respect to the standard basis has entries

$$b_{ij} = \beta(g^{-1}e_i, g^{-1}e_j) = (g^{-1}e_i)^T A(g^{-1}e_j) = e_i^T(g^{-T}Ag^{-1})e_j,$$

so $B = g^{-T}Ag^{-1}$ where g^{-T} is the inverse of the transpose of A (which is also the transpose of the inverse). In particular, that Gram matrix is *not* an invariant of the action. However, note that the determinant $\det(B)$ equals $\det(g^{-1})^2 \det(A)$. Hence if $\det(A)$ is a square in K, then so is $\det(B)$, and if $\det(A)$ is not a square in K, then neither is $\det(B)$.

Moreover, consider a second basis $he_1, \ldots, he_n$, where $h \in \mathrm{GL}_n$. With respect to this basis β has Gram matrix equal to h^TAh, and $g\beta$ has Gram matrix $g^{-T}h^TAhg^{-1}$. Again, the determinants of these matrices are squares if and only if $\det A$ is a square. Thus *the "squareness" of the determinant of a Gram matrix of a symmetric bilinear form β does not depend on the basis, and does not change when β is replaced by an element $g\beta$ in its orbit.* This squareness is therefore a well-defined invariant of symmetric bilinear forms (and can be transfered from K^n to abstract n-dimensional vector spaces over K by means of a linear isomorphism).

Example 10.3.115. Consider the two symmetric bilinear forms on $\mathbb{F}_5^2$ for with Gram matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix},$$

with respect to the standard basis. These are not in the same orbit under $\mathrm{GL}_2(\mathbb{F}_5)$ because the first Gram matrix has determinant 1, which is a square, while the second Gram matrix has determinant 2, which is not a square.

Exercise 10.3.116. Prove that two symmetric bilinear forms β, γ on a finite-dimensional vector space V are equivalent if and only if

(1) $\dim \operatorname{rad}\beta = \dim \operatorname{rad}\gamma$; and
(2) there exists a vector space isomorphism $\phi : V/\operatorname{rad}\beta \to V/\operatorname{rad}\gamma$ such that $\overline{\gamma}(\phi(v + \operatorname{rad}\beta), \phi(w + \operatorname{rad}\beta)) = \overline{\beta}(v + \operatorname{rad}\beta, w + \operatorname{rad}\beta)$.

This exercise reduces the study of (equivalence classes of) arbitrary symmetric bilinear forms to (equivalence classes of) non-degenerate ones.

10.4. Quadratic forms

A symmetric bilinear form β on V gives rise to a so-called *quadratic form* $q : V \to K$ defined by $q(v) := \beta(v, v)$. Conversely, the *polarisation identity*

$$\beta(v, w) = \frac{1}{4}(q(v + w) - q(v - w))$$

shows that q determines β, *provided that the characteristic of K is not* 2. In the remainder of this chapter we will therefore assume that $\operatorname{char} K \neq 2$.

10.5. Non-degenerate forms over algebraically closed fields

In this section we assume that K is algebraically closed with $\operatorname{char} K \neq 2$ and we let V be an n-dimensional vector space over K. Then we have the following classification result.

THEOREM 10.5.117. *The rank is a complete invariant of symmetric bilinear forms on V under the action of $\mathrm{GL}(V)$. In other words, for every symmetric bilinear form there exists a basis $v_1, \ldots, v_n$ of V with respect to which the form has Gram matrix*

$$\begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix},$$

where I_k is the $k \times k$-identity matrix and k is the rank of β. In particular, since k can take the values $0, \ldots, n$, there are exactly $n+1$ orbits.

PROOF. In view of the last paragraph of Section 10.3, it suffices to prove this theorem under the condition that β has full rank n. By Lemma 10.3.114 it suffices, in this case, to prove the existence of a basis $w_1, \ldots, w_n$ with respect to which β has Gram matrix I_n. This goes by induction as follows. If $n = 0$, then nothing needs to be done. Assume $n > 0$ and assume that the theorem holds for non-degenerate symmetric bilinear forms on $(n-1)$-dimensional spaces.

Since β is non-degenerate, the associated quadratic form q is not identically zero. Hence there exists a $v \in V$ with $q(v) = \beta(v, v) = c \neq 0$. Let $d \in K$ be a square root of c (here we use that K is algebraically closed) and set $v_n := \frac{1}{d}v$. Then $\beta(v_n, v_n) = 1$ as required. Now set $U := \langle v_n \rangle^{\perp}$. By Lemma 10.2.110 the dimension of U equals $n-1$. Moreover, $v_n \notin U$. Hence V can be decomposed as $U \oplus \langle v_n \rangle_K$. We claim that the restriction of β to U is non-degenerate. Indeed, since the orthogonal complement $U^{\perp}$ of U in V equals $\langle v_n \rangle$ (see Exercise 10.2.111) and v_n does not lie in U, no non-zero vector in U is perpendicular to all of U. Hence we may apply the induction hypothesis to the restriction of β to U, and find a basis $v_1, \ldots, v_{n-1}$ of U with respect to which that restriction has Gram matrix I_{n-1}. Then β has Gram matrix I_n with respect to $v_1, \ldots, v_n$, as desired. □

10.6. Non-degenerate forms over finite fields

In this section $K = \mathbb{F}_q$ is a finite field with q an odd prime power (so that the characteristic is not 2).

THEOREM 10.6.118. *Let V be an n-dimensional vector space over $\mathbb{F}_q$, and let β be a symmetric bilinear form on V. Then the pair consisting of the rank of β and the* squareness *of the (determinant of the Gram matrix of the) induced bilinear form $\overline{\beta}$ on $V/\operatorname{rad}\beta$ is a complete invariant under the action of $\mathrm{GL}(V)$. In other words, fix any non-square $c \in \mathbb{F}_q^*$. Then for any non-zero symmetric bilinear form β there is a basis of V with respect to which the Gram matrix of β equals either*

$$\begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix} \text{ or } \begin{bmatrix} c & 0 & 0 \\ 0 & I_{k-1} & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

where k is the rank of the form. In particular, there are precisely $2n+1$ orbits.

PROOF. Again, it suffices to prove this when $k = n$, that is, for non-degenerate forms. We proceed by induction. For $n = 0$ nothing needs to be done, because there are no non-zero bilinear forms. For $n = 1$, fix a basis v_1 of V. If $a := \beta(v_1, v_1)$ is a square, then take a square root d of a and observe that the quadratic form of β evaluates to 1 on the vector $\frac{1}{d}v_1$. If a is not a square, then $a = cd^2$ for some $d \in K^*$ (check that you know and understand these elementary facts about finite fields). Then the quadratic form of β evaluates to c on the vector $\frac{1}{d}v_1$. This settles the cases where $n \leq 1$; these are the base cases of our induction. Now assume that the statement is true for $n - 1$, and we set out to prove it for $n \geq 2$.

First, using the exact same proof as for the case where $\mathbb{F}_q$ is algebraically closed, we find that there exists a basis $w_1, \ldots, w_n$ for which β has a *diagonal* Gram matrix. We will only use the space W spanned by w_1, w_2. The restriction of β to W has Gram matrix

$$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$$

with $a, b \neq 0$ by non-degeneracy. If a is a square, then we may take a square root d, set $v_n := \frac{1}{d}w_1$ and proceed as in the proof in the case of algebraically closed fields. Similarly if b is a square. Hence assume that not both are squares. Then consider

$$\beta(c_1w_1 + c_2w_2, c_1w_1 + c_2w_2) = ac_1^2 + bc_2^2 =: r.$$

As c_1 runs through all elements of $\mathbb{F}_q^*$, the first term $s := ac_1^2$ runs trough all non-squares. Similarly, also the second term $t := bc_2^2$ runs through all non-squares as c_2 runs through $\mathbb{F}_q^*$. We claim that, for suitable choices of c_1, c_2 the element r is a non-zero square. This is equivalent to the claim that there exist non-squares s, t such that $s + t$ is a non-zero square. Let s be any non-square (for instance, a), let p be the prime dividing q, and let m be the smallest positive integer such that m is not a square modulo p (half of the integers coprime with p have this property). Then $t := (m - 1)s$ is not a square but $s + t = ms$ is. We conclude that some $w \in W$ has the property that $\beta(w, w)$ is a square d^2, and we may set $v_n := \frac{1}{d}w$, and we may proceed by induction as before. □

EXERCISE 10.6.119. Let A be the symmetric matrix

$$\begin{bmatrix} 12 & 49 & 67 \\ 49 & 52 & 43 \\ 67 & 43 & 13 \end{bmatrix}$$

over the field $\mathbb{F}_{71}$.

(1) Determine the rank of the symmetric bilinear form $\beta(v, w) = v^T Aw$ on $V = \mathbb{F}_{71}^3$.
(2) Determine the squareness of the induced form on $V/\operatorname{rad}\beta$.
(3) Determine $g \in \mathrm{GL}_3(\mathbb{F}_{71})$ such that gAg^T has the form in Theorem 10.6.118.

EXERCISE 10.6.120. Determine the order of the group of invertible 3×3-matrices g over $\mathbb{F}_q$ that satisfy

$$g^T g = I.$$

CHAPTER 11

More on bilinear forms

11.1. Some counting

In this section we count non-degenerate symmetric bilinear forms on $V = \mathbb{F}_q^n$, where q is a power of an odd prime. Let a_n denote that number. Then we have $a_0 = 1$ (the zero form on a zero-dimensional vector space is non-degenerate) and $a_1 = q-1$ (because the form is determined by its (non-zero) value on (v_1, v_1) with v_1 a basis of the space).

PROPOSITION 11.1.121. *The numbers a_n satisfy the following recurrence relation:*

$$a_n = (q^n - q^{n-1})a_{n-1} + (q^{n-1} - 1)q^{n-1}a_{n-2}.$$

PROOF. For a non-degenerate form β there are two possibilities:

(1) $c := \beta(e_1, e_1) \neq 0$: then $W := e_1^\perp$ is a vector space complement of Ke_1 in V, and the restriction $\beta|_W$ is non-degenerate. Moreover, β is uniquely determined by the triple $(c, W, \beta|_W)$ where c is a non-zero number, W is a codimension-one subspace not containing e_1, and $\beta|_W$ is a non-degenerate form on W; and any such triple gives rise to some β. Thus the number of β's with $\beta(e_1, e_1) \neq 0$ equals the number of triples, which is $(q-1) \cdot \left(\frac{q^n-1}{q-1} - \frac{q^{n-1}-1}{q-1}\right) \cdot a_{n-1}$. This is the first term in the recurrence.
(2) $\beta(e_1, e_1) = 0$: then $W := e_1^\perp$ is a codimension-one subspace containing e_1, and the restriction of β to W has Ke_1 as radical. Now β is uniquely determined by W, the induced non-degenerate form $\overline{\beta|_W}$ on the $(n-2)$-dimensional space W/Ke_1, and the restriction of β to $\{v_n\} \times V$, where Kv_n is any vector space complement of W in V. These three pieces of data are counted by $\frac{q^{n-1}-1}{q-1}$ and a_{n-2} and $(q-1)q^{n-1}$, respectively. To see the last factor, note that in the Gram matrix with respect to the basis (v_1 followed by a basis of W followed by v_n) the last bit counts the choices for the n entries in the last column (or row). The first of these, which is $\beta(v_1, v_n)$ has to be non-zero (or else β would be degenerate). This yields the second term in the recurrence.

□

EXERCISE 11.1.122. Prove that the numbers a_n also satisfy the following recurrence relation:

$$a_n = q^{\binom{n+1}{2}} - \sum_{k=1}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q a_{n-k},$$

where, as before,

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{(q^n-1)\cdots(q^{n-k+1}-1)}{(q^k-1)\cdots(q-1)}.$$

is the number of k-dimensional subspaces of V.

EXERCISE 11.1.123. Prove that the numbers a_n are as follows:

$$a_{2m+1} = (q^{2m+1}-1)(q^{2m-1}-1)\cdots(q-1)q^{m(m+1)}$$

for $n = 2m+1$ odd, and

$$a_{2m} = (q^{2m-1}-1)(q^{2m-3}-1)\cdots(q-1)q^{m(m+1)}$$

for $n = 2m$ even.

With the result of this exercise we can compute the orders of so-called *finite orthogonal groups*, defined as follows: given a non-degenerate symmetric bilinear form β on V, we define

$$\mathrm{O}(\beta) := \{g \in \mathrm{GL}(V) \mid g\beta = \beta\}.$$

Now for $h \in \mathrm{GL}(V)$ we have $\mathrm{O}(h\beta) = h\mathrm{O}(\beta)h^{-1}$, i.e., the orthogonal groups corresponding to forms in one $\mathrm{GL}(V)$-orbit are all $\mathrm{GL}(V)$-conjugate (and hence isomorphic) to each other. Since there are only two orbits of such forms (with square and non-square Gram matrix determinant, respectively), there are really only two different orthogonal groups.

In fact, if $n = \dim V$ is odd, then multiplying β with a fixed non-square c has the effect of multiplying the determinant of the Gram matrix with the non-square c^n. Hence then the map $\beta \mapsto c\beta$ gives a bijection between the two orbits of non-degenerate forms. Moreover, $\mathrm{O}(\beta) = \mathrm{O}(c\beta)$, so when n is odd there is up to isomorphism only one orthogonal group, somewhat ambiguously denoted $\mathrm{O}_n(\mathbb{F}_q)$. By Lagrange and Exercise 11.1.123, its order equals

$$|\mathrm{O}_{2m+1}(\mathbb{F}_q)| = \frac{|\mathrm{GL}(V)|}{\frac{1}{2}(q^{2m+1}-1)(q^{2m-1}-1)\cdots(q-1)q^{m(m+1)}} = 2(q^{2m}-1)\cdots(q^2-1)q^{m^2}.$$

On the other hand, if $n = 2m$ is even, then there are really two types of finite orthogonal groups: the stabiliser of a form with square Gram matrix determinant is denoted $\mathrm{O}^+_{2m}(\mathbb{F}_q)$, and the stabiliser of a form with non-square Gram matrix determinant is denoted $\mathrm{O}^-_{2m}(\mathbb{F}_q)$. We will not discuss the details.

EXERCISE 11.1.124. It makes sense to define orthogonal groups for degenerate forms as well.

(1) Determine the structure of matrices in $\mathrm{GL}_5(\mathbb{F}_q)$ stabilising the bilinear form whose Gram matrix with respect to the standard basis of $\mathbb{F}_q^5$ is

$$\begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & 0 \end{bmatrix}.$$

(2) Determine the order of the group in the previous part.

11.2. Real symmetric bilinear forms

We briefly discuss symmetric bilinear forms on the real vector space $\mathbb{R}^n$. Let p and q be natural numbers with $p+q \leq n$, and denote by $\beta_{p,q}$ the symmetric bilinear form defined by

$$\beta_{p,q}(x,y) := \left(\sum_{i=1}^{p} x_i y_i\right) - \left(\sum_{i=p+1}^{p+q} x_i y_i\right).$$

The radical of $\beta_{p,q}$ is spanned by the last $n-(p+q)$ standard basis vectors. The classification result is as follows.

THEOREM 11.2.125. *Any symmetric bilinear form on $\mathbb{R}^n$ is equivalent to exactly one $\beta_{p,q}$.*

PROOF. Last chapter's argument for diagonalising the Gram matrix carries through, except that we cannot take square roots of negative numbers. But it does show that any symmetric bilinear form on V is equivalent to *some* $\beta_{p,q}$. So all we need to do is show that $\beta_{p,q}$ cannot be equal to $\beta_{p',q'}$ for $(p,q) \neq (p',q')$. This is the content of the following exercise. □

A symmetric bilinear form β on a real vector space V is called *positive definite* if $\beta(v,v) > 0$ for all non-zero $v \in V$. It is called *negative definite* if $\beta(v,v) < 0$ for all non-zero $v \in V$.

EXERCISE 11.2.126.
(1) Show that the restriction of $\beta_{p,q}$ to the span of the first p standard basis vectors is positive definite.
(2) Show that any subspace $W \subseteq \mathbb{R}^n$ of dimension strictly larger than p contains a non-zero vector w for which $\beta_{p,q}(w,w) \leq 0$.
(3) Conclude that p is the maximal dimension of a subspace of $\mathbb{R}^n$ restricted to which $\beta_{p,q}$ is positive definite. Do something similar for q.
(4) Conclude from the previous part that $\beta_{p,q}$ cannot be equivalent to $\beta_{p',q'}$.