

# Linear Algebra\*

William McCallum

## 1 Equivalence and Canonical Forms in Algebra

### 1.1 High School Algebra

In the transition from arithmetic to algebra we learn to let symbols stand for unknown numbers. This gives rise to the important new ideas of variables, algebraic expressions and algebraic equations.

Algebraic expressions are built up from numbers, variables, and the basic operations of multiplication, division, addition and subtraction. We have rules for transforming expressions into equivalent forms (expanding, factoring, collecting like terms), which give rise to identities such as  $(x + y)^2 = x^2 + 2xy + y^2$ . These rules are rooted in the basic laws satisfied by the basic operations. We tend to think of identities as giving different *forms* of some underlying thing, but what that thing is exactly is never made explicit in high school algebra. In fact, from a more advanced point of view, there are at least two sensible answers to the question: the thing could be a function or it could be an element of some abstract algebraic structure, such as a polynomial ring.

Different equivalent forms of algebraic expressions have different uses. For example, a quadratic expression can be put into completed square form

$$ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2 + \left(c - \frac{b^2}{4a}\right)$$

in order to see its maximum or minimum, or in factored form

$$ax^2 + bx + c = a(x - \alpha)(x - \beta)$$

in order to see its roots. A useful idea is the idea of a canonical form: a standardized form, preferably simple, such that there is exactly one such form in each equivalence class. For example, the canonical form of a quadratic expression is  $ax^2 + bx + c$ , and we might test whether two quadratic expressions are equivalent by reducing both to this canonical form. For canonical forms to be useful in detecting equivalence, we need to have some systematic way of reducing any object to its canonical form, preferably by an explicit algorithm.

### 1.2 Abstract Algebra

In the transition from algebra to abstract algebra, the operations themselves become unknown, no longer standing for multiplication and addition of numbers, but for abstract operations of various sorts, subject to certain rules. Correspondingly, the symbols themselves no longer stand necessarily for numbers, but for elements of the domain on which the operations are defined. We need letters for the domains themselves in addition to their elements:  $K$  for a field,  $G$  for a group,  $R$  for a ring,  $V$  for a vector space.

What's left after all this abstraction? The basic idea of an algebraic expression remains—indeed, the different sorts of algebraic objects are distinguished by what sorts of expressions

---

\*A rough guide to some bits of it

we are allowed to write: monomials in groups, polynomials in rings, linear combinations in modules and vector spaces.

At the same time, we introduce a higher level of algebraic operation: operations on algebraic structures themselves. We have quotients, direct products and sums, tensor products. We also have various identities between these: fundamental isomorphism theorems like

$$\frac{V + W}{W} \simeq \frac{V}{V \cap W}.$$

Again at this level, we seek for canonical forms of objects that make it possible to detect whether they are isomorphic. We also have the new notion of a homomorphism between two objects, a map that preserves some structure but loses or misses some.

There is a possibility of confusion between isomorphism and equality. For example, for vector spaces, it is true that  $V \oplus W_1 \simeq V \oplus W_2$  implies that  $W_1 \simeq W_2$ . But, supposing that  $V$ ,  $W_1$ , and  $W_2$  all lie in some larger vector space, so that the question makes sense, is it true that  $V \oplus W_1 = V \oplus W_2$  implies that  $W_1 = W_2$ ?

## 2 Vector Spaces and Linear Transformations

We consider a vector space  $V$  over a field  $K$ , with the two operations of vector addition and scalar multiplication. The laws governing these operations are such that every expression in a vector space can be reduced to one of the form

$$\lambda_1 v_1 + \cdots + \lambda_n v_n, \quad \lambda_i \in K, v_i \in V.$$

We call this basic form a linear combination of the vectors  $v_i$ , and the  $\lambda_i$  are the coefficients. In fact, using the distributive law, we can arrange that the  $v_i$  are distinct, so we could also write this more compactly as

$$\sum_{v \in S} \lambda_v v, \quad \lambda_v \in K, S \subset V.$$

We want to know how to tell if two such expressions represent the same element of  $V$ , so we distinguish those sets  $S$  for which this question is easy to answer: we say  $S$  is linearly independent if the only circumstances under which two such expressions are equivalent are when the coefficients are equal. The set of all linear combinations of the elements of  $S$  is the set generated by  $S$ , and it is in fact a subspace of  $V$ . A linearly independent set which generates all of  $V$  is a *basis* for  $V$ . Another way of saying that  $B$  is a basis is to say that the map

$$(\lambda_v)_{v \in B} \mapsto \sum_{v \in B} \lambda_v v$$

is an isomorphism between  $V$  and the subspace of  $K^B = \prod_{v \in B} K$  consisting of vectors that have zero component at all but finitely many positions. It turns out this provides a canonical form for  $K$ -vector spaces. That is, every vector space has a basis, any two bases have the same cardinality, and two vector spaces are isomorphic if and only if they have bases of the same cardinality. Properly speaking we should choose a set of each cardinality to represent our canonical form: for denumerable cardinals we choose  $\{1, \dots, n\} \subset \mathbb{N}$ , allowing  $n = \infty$  as a possibility. This amounts to choosing an ordering on the basis  $B$ . The cardinal  $n$  is called the *dimension* of the vector space. If  $n$  is finite, then  $V \simeq K^n$ .

From now on we generally assume  $n$  is finite. The operations in the vector space  $K^n$  are defined in terms of the ordinary field operations in  $K$ . Thus the question posed above, of how to tell when two linear combinations give the same vector, is reduced by means of a basis to a question in  $K$ . The utility of this depends, of course, on being able to find a basis.

## 2.1 How do We Find a Basis?

Although every vector space has a basis, there is no one basis in general that is *the* basis—there are many choices of basis. On the other hand, although we know these bases exist, finding even one basis can be a problem in general.

In practice, many vector spaces come equipped with at least one basis:  $\mathbb{R}^n$  has the basis  $\{e_i\}$ , where  $e_i$  is the  $n$ -tuple with a 1 in the  $i$ -th place and 0 elsewhere. The vector space of polynomials in  $x$  over a field (or polynomials of degree less than or equal to a fixed number) has the basis  $1, x, x^2, \dots$  (although there are other natural choices). However, the vector space of solutions to an ordinary linear differential equation does not come equipped with any natural basis.

Suppose we have a finite dimensional vector space  $V$  equipped with an ordered basis  $B$ , and let  $W$  be the subspace generated by vectors  $v_j, j = 1, \dots, m$ , explicitly given in terms of the basis  $B$ . We want to find a basis for  $W$ . Using the isomorphism  $V \simeq K^n$  (where  $n$  is now a non-negative integer) we can reduce this question to case  $V = K^n$ , and

$$v_j = (\lambda_{ij}), \quad 1 \leq i \leq n, 1 \leq j \leq m.$$

The columns of the matrix  $M = (\lambda_{ij})$  are the vectors  $v_j$  represented with respect to the basis  $B$  of  $V$ . Suppose first that  $M$  is in row echelon form. Then the subset of vectors corresponding to columns where the echelon form steps down forms a basis for  $W$ . If  $M$  is not in row echelon form, we can make it so by elementary row operations. You might wonder, don't these change the vectors  $v_j$ ? That's one way of looking at it. Another way is that the vectors  $v_j$  remain unchanged, but the basis  $B$  is changed. For example, swapping rows corresponds to swapping the corresponding elements of the basis. Multiplying the  $i$ -th row by  $\lambda$  corresponds to multiplying the  $i$ -th basis element by  $\lambda^{-1}$ , since

$$\alpha_1 v_1 + \dots + (\lambda \alpha_i)(\lambda^{-1} v_i) + \dots + \alpha_n v_n.$$

Adding the  $i$ -th row to the  $j$ -th row corresponds to subtracting the  $j$ -th basis vector from the  $i$ -th, since

$$\begin{aligned} \alpha_1 v_1 + \dots + \alpha_i v_i + \dots + \alpha_n v_n &= \\ \alpha_1 v_1 + \dots + \alpha_i (v_i - v_j) + \dots + (\alpha_j + \alpha_i) v_j + \dots + \alpha_n v_n. \end{aligned}$$

## 2.2 Linear Transformations

A linear transformation  $T : W \rightarrow V$  is a function that preserves the vector space operations, i.e.,  $T(v + w) = T(v) + T(w)$  and  $T(\lambda v) = \lambda T(v)$ . The kernel and the image of  $T$  are defined in the usual way:

$$\begin{aligned} \ker T &= \{v \in W : T(v) = 0\} \\ \text{im } T &= \{w \in V : \exists v \in W, T(v) = w\}. \end{aligned}$$

Examples: on column vectors in  $\mathbb{R}^n$ , we can define linear maps by matrix multiplication; in fact, every linear map is defined this way (why?). On polynomials, the translation  $p(x) \mapsto p(ax + b)$  is a linear map. On  $C^\infty$  functions on  $\mathbb{R}$ , differentiation is a linear map (also  $\int_a^b$ ).

If  $W = K^m$  and  $V = K^n$  then  $T$  can be defined by an  $n \times m$  matrix. By means of the canonical form, we can represent any linear transformation by a matrix once we have chosen

bases for  $V$  and  $W$ . Explicitly, if  $\{w_1, \dots, w_m\}$  is an ordered basis for  $W$  and  $\{v_1, \dots, v_n\}$  is an ordered basis for  $V$ , then the matrix  $M = (\lambda_{ij})$  is given by

$$T(w_j) = \sum_{i=1}^n \lambda_{ij} v_i.$$

Row and column reduction on the matrix  $M = (\lambda_{ij})$  puts it in the form

$$J = \left( \begin{array}{c|c} I_k & 0 \\ \hline 0 & 0 \end{array} \right) \quad (1)$$

for some  $k \times k$  identity matrix  $I_k$ . We can record the row and column operations by starting out with identity matrices  $I_n$  and  $I_m$  positioned next to  $M$  like so:

$$\frac{0 \mid I_m}{I_n \mid M}$$

We record every row operation on  $M$  by also performing it on the matrix to the left, and every column operation by also performing it on the matrix above, until we have

$$\frac{0 \mid B}{A \mid J}$$

Then

$$ABM = J.$$

This follows from the following basic fact:

**LEMMA 2.1** *If we perform a series of elementary row operations (multiplying rows by scalars, permuting rows, and adding one row to another) on an identity matrix to get a matrix  $E$ , the left multiplication by  $E$  performs the same sequence of operations on any matrix  $M$ . A similar statement holds true for column operations and right multiplication.*

It is important to remember that the matrix  $J$  represents exactly the same linear map as the original matrix  $M$ , with respect to different bases  $(v'_i)$  and  $(w'_j)$ . We have already seen how the row operations can be interpreted as corresponding to operations on the basis, and in fact the  $A$  above is the inverse of the change-of-basis matrix from  $(v_i)$  to  $(v'_i)$ , that is, the matrix  $A'$  such that

$$(v_i)A' = (v'_i),$$

regarding  $(v_i)$  as a row of vectors and multiplying it into  $A'$  in the obvious way. The situation is simpler for  $B$ , since the column operations on  $M$  correspond to the same operations on the basis  $(w_i)$ , so  $B$  is the change of basis matrix from  $(w_i)$  to  $(w'_i)$ .

Thus the canonical form of a linear transformation between vector spaces is a matrix of the form (1). From (1) it is easy to read off the dimensions of the kernel and the image of  $T$ . In particular, we can deduce

$$\dim V = \dim \ker T + \dim \operatorname{im} T. \quad (2)$$

The linear transformations from  $V$  to  $W$  themselves form a vector space. What are the operations and what is its dimension?

### 2.3 Products, Quotients and Sums

The product of two vector spaces  $W_1$  and  $W_2$  is

$$W_1 \times W_2 = \{(w_1, w_2) : w_i \in W_i\}.$$

It comes equipped with projection maps

$$\begin{aligned} p_i : W_1 \times W_2 &\rightarrow W_i \\ (w_1, w_2) &\mapsto w_i \end{aligned}$$

In fact, strictly speaking, the direct product is the triple  $(W_1 \times W_2, p_1, p_2)$ .

We can also view  $W_1$  and  $W_2$  as being contained in  $W_1 \times W_2$ , via the maps

$$\begin{array}{ccc} i_1 : W_1 &\rightarrow & W_1 \times W_2 & i_2 : W_2 &\rightarrow & W_1 \times W_2 \\ w_1 &\mapsto & (w_1, 0) & w_2 &\mapsto & (0, w_2) \end{array}$$

Furthermore, under these identifications, every element of  $W_1 \times W_2$  can be expressed uniquely as a sum  $w_1 + w_2$  with  $w_i \in W_i$ . When viewed this way,  $W_1 \times W_2$  is denoted  $W_1 \oplus W_2$  and called the direct sum (again, strictly speaking the direct sum is the triple  $(W_1 \times W_2, i_1, i_2)$ ).

If  $W \subset V$ , the quotient is the same as the quotient as abelian groups, with scalar multiplication on the cosets  $W + v$  defined by  $\lambda(W + v) = W + \lambda v$ . For example, if  $V = \mathbb{R}^2$  and  $W$  is a line in  $V$  (through the origin, so it's a subspace), then  $V/W$  may be visualized as the family of lines parallel to  $V$ .

**Short Exact Sequences** We frequently break algebraic objects up into simpler pieces by considering a sub-object and the corresponding quotient. A useful way of describing this procedure is by means of sequences:

$$0 \rightarrow W \rightarrow V \rightarrow U \rightarrow 0.$$

Each arrow represents a homomorphism, and “exact” means that at each location the image of the arrow on the left is the kernel of the arrow on the right. Thus, the 0 on the left makes the next map injective, and the 0 on the right makes the previous arrow surjective. So  $W$  can be identified with a subspace of  $V$ , and then the fundamental isomorphism theorem gives an isomorphism  $U \simeq V/W$ . Thus the essential content of this representation is no more than the idea of taking a subspace  $W$  and considering the quotient  $V/W$ . But representing things this way is useful.

We say that a short exact sequence splits if there is a map  $U \rightarrow V$  which is the identity on  $U$  when composed with the projection  $V \rightarrow U$ . This gives a direct sum decomposition  $V = W \oplus U$  (how?).

### 2.4 Endomorphisms and Invariant Subspaces.

An endomorphism of  $V$  is a linear transformation from  $V$  to itself. These have a richer structure than maps from  $V$  to  $W$  because we can compose them with each other, giving a multiplication law. Thus the endomorphisms of  $V$  form a ring  $\text{End}(V)$ , which is also a vector space over  $K$  (such a structure is called an algebra).

Our canonical form for linear transformations depended on being able to choose different bases for  $V$  and  $W$ . When  $V = W$  we cannot necessarily assume that there is a basis for  $V$  with respect to which  $T$  is even diagonal, let alone diagonal with 1s and 0s. Thus the theory

of canonical forms for endomorphisms is much more interesting. In terms of matrices  $A$  and  $B$  above, we must have  $A = B^{-1}$ , we are looking for the simplest matrix of the form  $B^{-1}MB$ , rather than of the form  $AMB$ . The restriction that the change of basis matrix be the same on both sides leads to a richer theory of canonical forms.

Given a vector space  $V$  with an endomorphism  $T$ , an invariant subspace with respect to  $T$  is a subspace  $W \subset V$  such that  $T(W) \subset W$ . If we choose a basis adapted to  $W$  (that is, a basis for  $V$  the first  $k$  members of which form a basis for  $W$ ), then the corresponding matrix for  $T$  has an upper diagonal block structure which reflects the invariance (what exactly is this structure?).

We can go further and try to express  $V$  as a direct sum of two invariant subspaces  $W_1$  and  $W_2$ . This means that  $W_1 + W_2 = V$  and  $W_1 \cap W_2 = \{0\}$ , so that every vector  $v \in V$  has a unique expression  $w_1 + w_2$ ,  $w_i \in W_i$ . Then we can choose a basis for  $V$  consisting of the concatenation of bases for  $W_1$  and  $W_2$  (why?), and the corresponding matrix has a diagonal block structure. The most extreme example of this is if we can express  $V$  as a direct sum of one-dimensional invariant subspaces: in that case the corresponding matrix is diagonal.

So as a first step towards finding a canonical form for the matrix of  $T$  we seek invariant subspaces, and try to decompose  $V$  as a direct sum of such subspaces that are as small as possible. One way to find such a space is to start with a vector  $v$ , and keep adding vectors  $Tv, T^2v, \dots$ , until we find a vector  $T^k v$  which is a linear combination of the ones already added:

$$T^k v = \lambda_0 v + \lambda_1 Tv + \dots + \lambda_{k-1} T^{k-1} v. \quad (3)$$

Then the vectors  $T^i v$ ,  $0 \leq i \leq k-1$ , clearly generate an invariant subspace. How do we decide if there isn't a smaller such subspace contained in  $W$ ? And how do we know if there is a complementary subspace to  $W$ , or find one if there is? The equation (3) tells us that  $W$  can be regarded as a module over the polynomial ring  $K[t]$ , by letting  $t$  act as  $T$ . Studying the structure of this module for different invariant subspaces, and particular for  $V$  itself, is the key to answering our questions. So next we look at the structure of modules over rings.

### 3 Rings and Modules

A ring has the same operations as a field, addition and multiplication, but does not necessarily have multiplicative inverses and is not necessarily commutative.

Examples of rings:  $\mathbb{Z}$ ,  $K[t]$ ,  $K[t_1, \dots, t_n]$ ,  $M_n(K)$ . Having mentioned the last example, let us assume from now on that all rings are commutative.

A module is just a vector space over a ring.

Examples:  $\mathbb{Z}_N$ , any abelian group, ideals (modules that are contained in the ring), a vector space  $V$  is a module over the ring  $\text{End}(V)$ . Of course, just as with vector spaces,  $R^n$  can be regarded as a module over  $R$ .

We say a module is free (of finite rank) if it is isomorphic to  $R^n$  or, equivalently, it has a basis: a set  $x_1, \dots, x_n$  such that every element of the module can be uniquely expressed in the form  $r_1 x_1 + \dots + r_n x_n$ .

Is it true that modules behave just like vector spaces, i.e., that every module is free? Consider  $\mathbb{Z}/N\mathbb{Z}$ . If we take  $\{x_i\} = \{\bar{1}\}$ , then certainly every element can be written as  $n \cdot 1$ . But this not unique:  $(n+N) \cdot 1$  is the same element. The problem here is the phenomenon of *torsion*: the existence of elements  $m \in M$ ,  $r \in R$ , both not zero, such that  $r \cdot m = 0$ . Why does this phenomenon not arise in the theory of vector spaces?

The problem of torsion is unavoidable. If  $R$  is a ring that contains a non-zero non-unit  $x$  (i.e., not a field) then  $R/xR$  is a torsion module over  $R$ . But maybe if we get rid of torsion, things would be alright. Does every torsion-free module have a basis? No. Consider the

ideal  $(x, y) \subset K[x, y]$ . It has generators  $x$  and  $y$ , which are not linearly independent (since  $y \cdot x - x \cdot y = 0$ ), but we cannot eliminate either of them. (This does not prove that there isn't some other basis, but it gives you an idea of the problem. Exercise: prove that  $(x, y)$  is not a free  $K[x, y]$ -module.)

So maybe the problem is that we need to restrict to rings  $R$  which don't have ideals like  $(x, y)$  in them (again, the problem is unavoidable otherwise). For example, we could restrict to principal ideal domains. Is it true that every torsion-free module over a principal ideal domain is free? No. Think about  $\mathbb{Q}$  over  $\mathbb{Z}$ .

However, in the case of finitely generated modules over principal ideal domains, we have something approaching the theory of vector spaces. For example, it is true that a submodule  $N$  of a finitely generated free module  $M$  is itself free. To prove this we need the following lemma.

LEMMA 3.1 *If*

$$0 \rightarrow N \xrightarrow{i} M \xrightarrow{p} P \rightarrow 0$$

*is an exact sequence of (finitely generated)  $R$ -modules, and if  $P$  and  $N$  are free, then  $M$  is free.*

PROOF. Let  $(x_1, \dots, x_r)$  be a basis for  $N$  and  $(y_1, \dots, y_s)$  be a basis for  $P$ . Choose  $y'_k \in M$  such that  $p(y'_k) = y_k$ . Then  $(i(x_1), \dots, i(x_r), y'_1, \dots, y'_s)$  is a basis for  $M$  (check this!). ■

THEOREM 3.2 *Let  $R$  be a principal ideal domain, and let  $N$  be a submodule of a finitely generated free module  $M$ . Then  $N$  is free.*

PROOF. Suppose  $M \simeq R^n$ , so we may regard  $N$  as a submodule of  $R^n$ . Consider the diagram

$$\begin{array}{ccccccc} 0 & \longrightarrow & R & \longrightarrow & R^n & \xrightarrow{p} & R^{n-1} \longrightarrow 0 \\ & & \uparrow & & \uparrow & & \uparrow \\ 0 & \longrightarrow & R \cap N & \longrightarrow & N & \longrightarrow & p(N) \longrightarrow 0 \end{array}$$

where  $p$  is projection onto the last  $n - 1$  factors. Then  $R \cap N$  is an  $R$ -submodule of  $R$  itself, which is nothing more or less than an ideal. So  $R \cap N = xR$  for some  $x \in R$ , since  $R$  is a principal ideal domain. Hence  $N \cap R$  is free. And  $p(N)$  is free by induction on  $n$  (where the base case  $n = 1$  again follows from the fact that  $R$  is a principal ideal domain). By Lemma 3.1,  $N$  is free. ■

Now, let  $M$  be an arbitrary finitely generated module over a principal ideal domain  $R$ , with  $n$  generators  $x_1, \dots, x_n$ . Then there is a surjective map  $R^n \rightarrow M$ . The kernel is a submodule, hence free by Theorem 3.2. So we get an exact sequence

$$0 \rightarrow R^m \xrightarrow{A} R^n \rightarrow M \rightarrow 0. \tag{4}$$

Here  $A$  is an  $n \times m$  matrix with coefficients in  $R$ . Just as in the case of vector spaces, we can use row and column operations to put this matrix in a simple form. Exactly which operations do we use? We use operations that correspond to change of basis: thus, we do not include multiplying a row or column by an arbitrary element of  $R$ , since that corresponds to

either multiplying or dividing the corresponding basis element by that constant (depending on whether we are talking about columns or rows). First, we can't divide in general, and second, multiplying a basis element by a non-unit in  $R$  could render it a non-basis. Specifically, the operations we allow are interchanging rows and columns, and adding a multiple of one row or column to another (we could, in fact, also allow multiplying a row or column by a *unit*).

**THEOREM 3.3** *Let  $R$  be a principal ideal domain, and let  $A$  be an  $n \times m$  matrix with coefficients in  $R$ . Then, by elementary row and column operations,  $A$  can be put in the form*

$$\left( \begin{array}{c|c} E & 0 \\ \hline 0 & 0 \end{array} \right)$$

where  $E$  is a diagonal matrix whose entries  $e_1, \dots, e_k$  satisfy  $e_i | e_{i+1}$ ,  $1 \leq i \leq k-1$ .

We give the main idea of the proof in the case that  $R$  is  $K[t]$  for some field  $K$ . First we find a non-zero element in the matrix and move it to the top left. Then we divide it successively into every entry of the matrix. Any time we get a nonzero remainder, we move it to the top left. Since the degrees of the remainders are strictly decreasing positive integers, we must eventually have a situation where the top left element divides every other entry of the matrix. Then subtract appropriate multiples of the top row and the left column to get 0s in the top row and left column at all positions other than the top left. Then we apply the same procedure inductively to the remaining bottom right sub-matrix.

Applying this to (4), we get the fundamental structure theorem for modules over principal ideal domains.

**THEOREM 3.4** *Let  $R$  be a principal ideal domain, and let  $M$  be a finitely generated  $R$ -module. Then there exists an integer  $N$  and elements  $e_1, \dots, e_k \in R$ , satisfying  $e_i | e_{i+1}$ ,  $1 \leq i \leq k-1$ , such that*

$$M \simeq R^N \oplus R/(e_1) \oplus \cdots \oplus R/(e_k).$$

The elements  $e_i$  are called the invariant factors of the matrix, or of the module, and the integer  $N$  is called the rank. They constitute a basic set of invariants for finitely generated modules over PIDs, and the theorem gives us a canonical form for such modules. To qualify as a canonical form, it must be unique in some sense. It turns out that  $N$  is unique (that is, two isomorphic modules have the same  $N$ ), and that the  $e_i$  are unique up to multiplication by a unit in  $R$ . (Exercise: prove this.)

As an application of this theorem we have the basic structure theorem for finitely generated abelian groups. An abelian group is just a  $\mathbb{Z}$ -module, and every finitely generated abelian group is isomorphic to

$$\mathbb{Z}^n \oplus \mathbb{Z}/e_1\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/e_k\mathbb{Z}. \quad (5)$$

There is another standard way of writing an abelian group. If  $e \in \mathbb{Z}$  factors into primes as  $e = p_1^{f_1} \cdots p_r^{f_r}$ , then

$$\mathbb{Z}/e\mathbb{Z} \simeq \mathbb{Z}/p_1^{f_1} \times \cdots \times \mathbb{Z}/p_r^{f_r}. \quad (6)$$

To prove this, we consider the obvious map from the left side to the right (the natural projection onto each factor is a quotient map by a subgroup) and show it is injective and surjective, using the Chinese Remainder Theorem.

Applying the decomposition in (6) to each torsion factor in (5), we get an alternative structure theorem.

**THEOREM 3.5** *Let  $R$  be a principal ideal domain, and let  $M$  be a finitely generated  $R$ -module. Then there exists an integer  $N$  and prime powers  $p_i^{f_i}$ ,  $i = 1, \dots, r$ , where the  $p_i$  are (not necessarily distinct) irreducible elements in  $R$  and the  $f_i$  are positive integers, such that*

$$M \simeq R^N \oplus \bigoplus_{i=1}^r R/(p_i^{f_i}).$$

## 4 Canonical Forms of Matrices

Let  $K$  be a field,  $V$  a finite dimensional vector space over  $K$ , and  $T : V \rightarrow V$  a linear transformation. We give  $V$  the structure of a module over  $K[t]$  by setting  $tv = Tv$ .

Let  $W$  be another vector space with a linear transformation  $S : W \rightarrow W$ , and regard  $W$  as a  $K[t]$ -module in the same way as above. Then  $W$  is isomorphic to  $V$  as a  $K[t]$ -module if and only if there is a vector space isomorphism  $A : V \rightarrow W$  such that  $AT = SA$ . (Prove this!)

Now let  $V = W = K^n$  and let  $T$  and  $S$  be the linear transformations given by matrices  $M, N \in M_n(K)$ . Then the two  $K[t]$ -module structures on  $V$  arising from  $T$  and  $S$  are isomorphic to each other if and only if the matrices  $M$  and  $N$  are similar.

Thus, canonical forms of matrices correspond to canonical forms of  $K[t]$ -modules. So now let's look at the latter and see what natural bases they have, and what the corresponding matrices are.

First, consider the case where  $V = K[t]/(f)$  for some polynomial

$$f = a_0 + a_1t + \dots + t^n.$$

(Notice that we have assumed  $f$  is monic. We can do this without loss of generality, since multiplying  $f$  by the inverse of its leading coefficient does not change the ideal it generates in  $K[t]$ .) A basis for  $V$  as a vector space is  $(1, t, \dots, t^{n-1})$ . Indeed, if an element of  $V$  is represented by a polynomial  $g$ , then we can divide  $f$  into  $g$  and find a remainder  $r$  of degree less than  $n$ . Since  $g - r$  is a multiple of  $f$ ,  $r$  represents the same element of  $V$  as  $g$  does, and since  $r$  has degree less than  $n$  it is a linear combination of the basis vectors. Thus the basis spans  $V$ . To see that it is linearly independent, consider a linear dependence relation

$$\alpha_0 + \alpha_1t + \dots + \alpha_{n-1}t^{n-1} = 0 \quad \text{in } V.$$

A polynomial represents the 0 element of  $V$  if and only if it is divisible by  $f$ , but a polynomial of degree less than  $n$  cannot be divisible by  $f$  unless it is zero. Thus  $\alpha_i = 0$  for  $1 \leq i \leq n-1$ . So our basis is indeed a basis.

Now, multiplication by  $t$  gives a linear transformation  $T$  of  $V$ . What is its matrix?

$$\begin{aligned} T(1) = t &= 0 \cdot 1 + 1 \cdot t + 0 \cdot t^2 + \dots + 0 \cdot t^{n-1} \\ T(t) = t^2 &= 0 \cdot 1 + 0 \cdot t + 1 \cdot t^2 + \dots + 0 \cdot t^{n-1} \\ &\vdots \\ T(t^{n-2}) = t^{n-1} &= 0 \cdot 1 + 0 \cdot t + 0 \cdot t^2 + \dots + 1 \cdot t^{n-1} \\ T(t^{n-1}) = t^n &= -a_0 \cdot 1 - a_1 \cdot t - a_2 \cdot t^2 - \dots - a_{n-1}t^{n-1}. \end{aligned}$$

Hence the matrix for  $t$  in this basis is

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix}. \tag{7}$$

In general, if  $V$  is isomorphic to a direct sum of cyclic modules as in Theorem 3.4 (there is no free part because  $K[t]$  is infinite dimensional as a vector space over  $K$ ), we get a block structure for the matrix, with blocks as in (7). This is known as the rational canonical form of the matrix. A pleasant exercise verifies that the characteristic polynomial of (7) is  $f$ , so the characteristic polynomial of a general matrix is the product of its invariant factors. Furthermore, since  $e_i | e_{i+1}$ , it is clear that the minimal polynomial is the last invariant factor  $e_k$ .

Now suppose that  $K$  is algebraically closed. The second form of the fundamental structure theorem, Theorem 3.5, leads to the Jordan canonical form. Since  $K$  is algebraically closed, the prime powers are  $(t - \lambda)^k$ . We consider again just a single component of the form  $V = K[t]/((t - \lambda)^k)$  and choose the basis  $(1, (t - \lambda), \dots, (t - \lambda)^{k-1})$ . Then

$$\begin{aligned} T(1) = t &= \lambda \cdot 1 + 1 \cdot (t - \lambda) + 0 \cdot (t - \lambda)^2 + \cdots + 0 \cdot (t - \lambda)^{k-1} \\ T(t - \lambda) = t(t - \lambda) &= 0 \cdot 1 + \lambda \cdot (t - \lambda) + 1 \cdot (t - \lambda)^2 + \cdots + 0 \cdot (t - \lambda)^{k-1} \\ &\vdots \\ T((t - \lambda)^{k-1}) = t(t - \lambda)^{k-1} &= 0 \cdot 1 + 0 \cdot (t - \lambda) + 0 \cdot (t - \lambda)^2 + \cdots + \lambda(t - \lambda)^{k-1} \end{aligned}$$

where in the last line we have used the fact that  $(t - \lambda)^k$  represents 0 in  $V$ . The corresponding matrix is the (slightly un-)usual Jordan block

$$A = \begin{pmatrix} \lambda & 0 & \cdots & 0 & 0 \\ 1 & \lambda & \cdots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & 0 \\ \vdots & \vdots & \ddots & \lambda & \vdots \\ 0 & 0 & \cdots & 1 & \lambda \end{pmatrix}.$$

(If you want the completely usual form, you need to use this basis in reverse order, which puts the 1s above the diagonal.)

## 5 Duality Pairings, and Determinants

We saw how the leap from high-school algebra to abstract algebra was to make the operations themselves unknowns, which led to us introducing operations on algebraic structures. The next leap is to regard the structures themselves as variables. We have already done this to certain extent: we let  $V$  stand for an unknown vector space, but so far it has been in some sense fixed. Now we want to treat  $V$  as a true variable.

For example, we have seen that  $\text{Hom}(V, W)$  is a vector space. We think of this as a two-variable function from vector spaces to vector spaces, and we want to study the properties of this function. Let's fix one of the variables, say put  $W = K$ , then we get the dual vector

space  $V^*$ . The map  $V \mapsto V^*$  is what we call a functor on vector spaces. Given a linear map  $T : V \rightarrow W$ , we get a dual map going in the other direction

$$\begin{aligned} T^* : W^* &\rightarrow V^* \\ f &\mapsto f \circ T. \end{aligned}$$

Thus the functor not only maps vector spaces to vector spaces, but it maps vector space homomorphisms to vector space homomorphisms. Moreover, this map respects the algebraic structure of homomorphisms, in the sense that  $(S \circ T)^* = T^* \circ S^*$ . (The reversal of order is necessary because the direction of the maps is reversed. We call the functor *contravariant* in this case: we also have *covariant* functors, which preserve the direction of maps and the order of composition.)

More generally,  $\text{Hom}(V, W)$  is a functor in both  $V$  and  $W$ , contravariant in the first variable and covariant in the second variable.

What is a basis for  $V^*$ ? Given a basis  $(v_1, \dots, v_n)$  for  $v$ , every linear map  $V \rightarrow K$  (linear functional) is determined by its value on the  $v_i$ . If we define  $v_i^*$  to be the functional that takes the value 1 on  $v_i$  and 0 on the other basis elements, then the functional that takes the value  $\alpha_i$  on  $v_i$  can be written

$$\sum_{i=1}^n \alpha_i v_i^*.$$

Thus  $(v_i^*)$  is a basis for  $V^*$ . Now, given a basis  $(w_1, \dots, w_m)$  for  $W$ ,  $(v_i^* w_j)$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ , is a basis for  $\text{Hom}(V, W)$ . So the dimension of  $\text{Hom}(V, W)$  is  $nm$  (which makes sense, since it is isomorphic to the space of  $m \times n$  matrices).

Now,  $V$  is isomorphic to  $V^*$  since they have the same dimension. However, there is no *natural* isomorphism: one is as good as another. On the other hand, there *is* a natural isomorphism

$$V \simeq (V^*)^* \tag{8}$$

which takes  $v$  to the functional  $V^*$  that maps  $f$  to  $f(v)$ . Notice that the description of this isomorphism makes no mention of a basis.

Why do we care about natural isomorphisms? There are situations where you consider families of vector spaces, and even though each one has a basis, there is no sensible universal choice of a basis. For example, if we attach a series of one-dimensional vector spaces to a circle in such a way that they form a Möbius strip, there is no continuous choice of basis. In general, if we have two vector spaces  $V$  and  $W$ , and a linear map  $T : V \rightarrow W$ , and we have chosen isomorphisms  $V \simeq V^*$  and  $W \simeq W^*$ , so that  $T^*$  can be regarded as a map  $W \rightarrow V$  we have no guarantee that there is any sensible relationship between  $T$  and  $T^*$ . On the other hand, if we use the isomorphism (8) to identify  $(T^*)^*$  with a map  $V \rightarrow W$ , then we find  $(T^*)^* = T$ . Another way of saying this is that there is an isomorphism of *functors* between the double duality functor and the identity functor, whereas the duality functor is not isomorphic to the identity functor.

So if we want an isomorphism  $V \rightarrow V^*$ , then, we have to choose one: it is not naturally determined by what we already have in hand. Therefore, it makes sense to want to know the nature of the domain from which we are making a choice—to consider the entire vector space  $\text{Hom}(V, V^*)$ .

Given an element  $T \in \text{Hom}(V, V^*)$ , we can define a pairing  $\langle \cdot, \cdot \rangle : V \times V \rightarrow K$  by

$$\langle v, w \rangle = T(v)(w). \tag{9}$$

A pairing is a map  $V \times W \rightarrow K$  which is separately linear in each variable when you hold the other variable fixed. (Note that this implies that it is *not* linear as map  $V \times W \rightarrow K$ .) Example: inner product on  $\mathbb{R}^n$ . Pairings form a vector space in a natural way. Let's call it  $L^2(V, K)$ . Conversely, any pairing defines a linear map  $T$ , by (9). Thus we have an isomorphism

$$L^2(V) \simeq \text{Hom}(V, V^*).$$

This *is* a natural isomorphism: it is defined without any reference to a basis. Thus, choosing a pairing on  $V$  is the same thing as choosing a map  $V \rightarrow V^*$ . The pairings which correspond to isomorphisms under this identification are called non-degenerate pairings.

The moral of the story is that when dealing with abstract vector spaces, it's important to know when you are making a choice and when you are getting something for free. This is hard to do if you think about  $\mathbb{R}^n$  always, since it comes equipped with a basis, and with an isomorphism  $\mathbb{R}^n \rightarrow (\mathbb{R}^n)^*$ .

As an illustration of this functorial point of view, consider the determinant. The determinant of a matrix has a well-known formula. The question arises: is the formula something we are choosing, or does it arise naturally? Are there other determinants? We'll look now for a coordinate-free definition of the determinant. So, since a square matrix is a map  $K^n \rightarrow K^n$ , we are looking at some way of defining  $\det(T)$  for an arbitrary endomorphism  $T : V \rightarrow V$ . Does it depend on the choice of a basis for  $V$  or not?

A hint to get us started is to notice that the determinant is characterized by its properties under row and column operations, and by the fact that  $\det(I) = 1$ . We can summarize these properties by saying that the determinant is a multilinear, alternating map on its columns or rows. Consider, for a moment, the case  $n = 2$ , where we have a bilinear pairing.

There is yet another way of looking at bilinear pairings. We noticed before that bilinear pairings are not linear on  $V \times W$ : a bilinear pairing is a map on  $B : V \times W \rightarrow K$  which satisfies the "wrong rules" for linearity, for example:

$$B(v_1 + v_2, w) = B(v_1, w) + B(v_2, w).$$

If we define a vector space structure on pairs  $(v, w)$  that satisfies these rules, we get the tensor product  $V \otimes W$ , consisting of symbols  $v \otimes w$  such that

$$(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w, \text{ etc.}$$

Then

$$L^2(V, K) = \text{Hom}(V \otimes V, K).$$

More generally,  $L^n(V, K) = \text{Hom}(V \otimes \cdots \otimes V, K) = \text{Hom}(V^{\otimes n}, K)$ .

There is an analogous construction for alternating multilinear forms, the wedge product. It is generated (in the case  $n = 2$ ) by symbols  $v \wedge w$  subject to the same conditions as tensors, and the extra condition  $v \wedge w = -w \wedge v$ . The space of alternating  $n$ -multilinear forms on  $V$  is then isomorphic to  $\text{Hom}(V \wedge \cdots \wedge V, K)$ , where the wedge is taken  $n$  times. Now suppose that  $n$  is the dimension of the vector space  $V$ , so that  $V$  has a basis  $v_1, \dots, v_n$ . Then it is not hard to see that every  $n$ -fold wedge product is a linear multiple of  $v_1 \wedge \cdots \wedge v_n$ . Thus  $V \wedge \cdots \wedge V$  is at most one-dimensional. Furthermore, the existence of the determinant shows that there is a non-trivial  $n$ -linear alternating form on  $V$ , hence the space is in fact one-dimensional. Given this, we define  $\det(T)$ , for a linear transformation  $T : V \rightarrow V$ , to be the functorially induced map  $V \wedge \cdots \wedge V \rightarrow V \wedge \cdots \wedge V$ . Since this is an endomorphism of a one-dimensional vector space, it is multiplication by a scalar. We call that scalar  $\det(T)$ .

(The key here is that the matrix of an endomorphism of a one-dimensional vector space is independent of the choice of a basis!)

Note that, although this is a definition of the determinant that does not depend on the choice of a basis, we needed to choose a basis to prove, for example, that  $\det(T)$  is not always zero. This is often the case: we need coordinates to do actual computations, but it is useful to know when things can be defined in a coordinate-free manner.