CHAPTER V

MODULES

1. Modules

Let A be a ring. A *left module* M over A consists of an abelian group (also denoted M) and a law of composition $A \times M \to M$ (denoted $(a, x) \mapsto ax$) such that

$$\begin{aligned} a(bx) = (ab)x \quad & \text{for } a, b \in A \text{ and } x \in M, \\ 1x = x \quad & \text{for } x \in M, \\ (a+b)x = ax + bx \quad & \text{for } a, b \in A \text{ and } x \in M, \\ a(x+y) = ax + ay \quad & \text{for } a \in A \text{ and } x, y \in M. \end{aligned}$$

The last statement asserts that $\rho(a) : M \to M$ defined by $\rho(a)(x) = ax$ is an endomorphism of the underlying abelian group of the module, while the first three statements assert that $\rho : A \to \mathrm{End}(M)$ is a ring homomorphism. Conversely, given such a homomorphism, we may define a module structure on M by setting $ax = \rho(a)(x)$.

Similarly, a *right* module M over A consists of an abelian group (also denoted M) and a binary operation $(x, a) \mapsto xa$ such that

$$\begin{aligned} (xb)a = x(ba) \quad & \text{for } a, b \in A \text{ and } x \in M, \\ x1 = x \quad & \text{for } x \in M, \\ x(a+b) = xa + xb \quad & \text{for } a, b \in A \text{ and } x \in M, \\ (x+y)a = xa + ya \quad & \text{for } a \in A \text{ and } x, y \in M. \end{aligned}$$

In this case $\rho : A \to \mathrm{End}(M)$ defined by $\rho(a)(x) = xa$ is not a homomorphism but an *anti-homomorphism*.

Note the similarities between these notions and the notion of a group acting on a set. However, in this case there is more structure to be taken account of.

If the ring A is commutative, we need not really distinguish between right and left modules since $ab = ba$, and every anti-homomorphism is also a homomorphism and vice versa. However, in the non-commutative case the distinction is often important.

Various elementary facts (like $0x = 0$ for all $x \in M$) follow easily from the definitions, and we shall assume these facts without further discussion.

Modules are best initially thought of as abelian groups with additional structure. In particular, we would expect most of the basic facts we derived earlier for groups (hence for abelian groups) to hold true.

Examples:

1) Let $A = \mathbf{Z}$. Then if M is any abelian group, we may define a $\mathbf{Z}$-module structure on M by $(n, x) \mapsto nx =$ the sum of n copies of x if n is positive, the negative of $-n$ copies of x if n is negative, and 0 if $n = 0$.
2) Let $A = k$ be a field. Then a module over k is called a vector space. We hope that you have studied vector spaces in an earlier course. Many facts about vector spaces remain true for modules, but in many ways the theory of modules is considerably richer. For example, if $a \neq 0$ in a field k, then

$ax = 0 \Rightarrow x = 0$ since we can multiply by a^{-1}. The corresponding fact in an arbitrary module is of course not generally true.

3) Let A be a ring. Then using multiplication in A to define the operation, we may view A either as a left or a right module over itself.

4 Let k be a field, and let $A = M_n(k)$ be the ring of $n \times n$ matrices with entires in k. Let $V = k^n$ visualized as $n \times 1$ matrices or column vectors. Let A act on V by matrix multiplication ax where $a \in A, x \in V$. Then V becomes a left A-module. If we visualize V as $1 \times n$ matrices or row vectors, then matrix multiplication xa makes V a right A-module.

We shall now proceed to develop the basic theory of modules as we did earlier for groups and rings. We shall generally restrict attention to left modules, but the theory for right modules is exactly the same with appropriate change of notation.

Let A be a ring and M a left A-module. An additive subgroup N of M (as abelian group) is called a *submodule* if $a \in A, x \in N \Rightarrow ax \in N$.

Examples:

1 If $A = \mathbf{Z}$ then a submodule is just a subgroup.

2 If A is a field, then a submodule is called a subspace.

3 If A is viewed as a left module over itself, then a submodule is just a left ideal. Similarly, if it is viewed as a right module over itself, then a submodule is a right ideal.

Let A be a ring and M a left A-module. If $\mathfrak{a}$ is an additive subgroup of A and N is an additive subgroup of M, then we define $\mathfrak{a}N$ to be the additive subgroups of M generated by all products ax with $a \in \mathfrak{a}$ and $x \in N$. The basic formulas developed earlier for products of additive subgroups in a ring are also true in this more general situation:

$$(\mathfrak{a}\mathfrak{b})N = \mathfrak{a}(\mathfrak{b}N), (\mathfrak{a} + \mathfrak{b})N = \mathfrak{a}N + \mathfrak{b}N, \text{ and } \mathfrak{a}(N + L) = \mathfrak{a}N + \mathfrak{a}L.$$

If S is a subset of a left module M over A, then as above, we may denote by AS the additive subgroup of M generated by all products ax with $a \in A$ and $x \in S$. It is easy to see that it is a submodule and it is called the submodule generated by S.

The sum $N + L$ and intersection $N \cap L$ of two submodules is a submodule. (It is even true that an intersection of infinitely many submodules is a submodule.)

Let M and M' be left A-modules. A function $f : M \to M'$ is called a module homomorphism if it is a homomorphism of additive groups and consistent with the actions:

$$\begin{aligned} f(x + y) &= f(x) + f(y) \quad \text{for } x, y \in M \\ f(ax) &= af(x) \quad \text{for } a \in A, x \in M. \end{aligned}$$

This may be simplified to

$$f(ax + y) = af(x) + f(y) \text{ for } a \in A, x, y \in M.$$

(If A is a field, recall that a module homomorphism is called a linear function or linear transformation.)

Let A be a ring, M a left A-module, and N a submodule. The factor group M/N (as additive abelian group) may be made into an A-module by defining $a(x + N) = ax + N$ for a coset $x + N \in M/N$. The canonical epimorphism is then a module homomorphism.

Let $f : M \to M'$ be a homomorphism of left A-modules. Then it is easy to check that $\operatorname{Ker} f$ is a submodule of M and $\operatorname{Im} f$ is a submodule of M'. Moreover, the group theoretic isomorphism $M/\operatorname{Ker} f \to \operatorname{Im} f$ of the first isomorphism theorem is easily seen to be a module isomorphism. Similarly, the various homomorphisms and isomorphisms in the second and third isomorphism theorems are also easily seen to be module homomorphisms.

There is a collection of simple notions concerning modules which have entered algebra from algebraic topology. These concepts are important in developing the theory of homology and cohomology groups, but they have also turned out to be useful ways to discuss certain aspects of module theory.

A sequence of module homomorphisms

$$M' \xrightarrow{f} M \xrightarrow{g} M''$$

is called *exact* (at M) if $\operatorname{Ker} g = \operatorname{Im} f$. A longer sequence is called exact if it is exact at each stage. An exact sequence of the form

$$0 \longrightarrow M' \xrightarrow{f} M \xrightarrow{g} M'' \longrightarrow 0$$

is called a *short* exact sequence. (Note that the homomorphisms on the left and right need not be further specified since there is only one homomorphism from the 0 module to any given module and there is only one homomorphism from any given module to the 0 module.) The assertion that the sequence is exact at M' is just that $\operatorname{Ker} f = \operatorname{Im} 0 = \{0\}$, i.e., that f is a monomorphism. The assertion that the sequence is exact at M is that $\operatorname{Ker} g = \operatorname{Im} f$ which is isomorphic to M'. The assertion that the sequence is exact at M'' is that $M' = \operatorname{Ker} 0 = \operatorname{Im} g$, i.e., that g is an epimorphism. Hence $M'' \cong M/\operatorname{Ker} g = M/\operatorname{Im} f$ and $\operatorname{Im} f \cong M'$. Hence such a short exact sequence generalizes the situation in which M' is a submodule of M and $M' = M/M'$. A useful short exact sequence in the case $A = \mathbf{Z}$ is

$$0 \longrightarrow \mathbf{Z} \xrightarrow{k} \mathbf{Z} \longrightarrow \mathbf{Z}/k\mathbf{Z} \longrightarrow 0$$

where the homomorphism on the left is shorthand for multiplication by the integer k.

In module theory, we also commonly use the additional terminology (which could in fact have been defined earlier for groups). For $f : M \to M'$ a module homomorphism:

$$\begin{aligned} \operatorname{Coker} f &= M'/\operatorname{Im} f \\ \operatorname{Coim} f &= M/\operatorname{Ker} f \cong \operatorname{Im} f. \end{aligned}$$

Associated with a module homomorphism $f : M \to M'$ are two short exact sequences

$$\begin{aligned} 0 \to \operatorname{Ker} f \longrightarrow M \longrightarrow \operatorname{Coim} f \to 0 \\ 0 \to \operatorname{Im} f \longrightarrow M' \longrightarrow \operatorname{Coker} f \to 0 \end{aligned}$$

and the longer exact sequence

$$0 \to \operatorname{Ker} f \longrightarrow M \xrightarrow{f} M' \longrightarrow \operatorname{Coker} f \to 0$$

which can be thought of as obtained by piecing together the two previous short exact sequences.

One often has to deal with a complicated system of modules, submodules, factor modules, and homomorphisms of such. Given such a situation, we often want to prove that some crucial homomorphism is an isomorphism. For example, suppose $f : M \to L$ is a module homomorphism, M' is a submodule of M, and L' is a submodule of L such that $f(M') \subseteq L'$. Since $f(M') \subseteq L'$, it is not hard to see that f induces a module homomorphism $f'' : M/M' \to L/L'$. Let $f' : M' \to L'$ denote the restriction of f to M'. It is not hard to prove that if f' and f'' are monomorphisms, then f is also a monomorphism, and similarly for epimorphisms. A general method for dealing with such issues is the so-called *5-lemma* which we dissect below into two *4-lemmas.*

PROPOSITION. *Suppose in the diagram*

$$\begin{array}{ccccccc} M_1 & \xrightarrow{\alpha_1} & M_2 & \xrightarrow{\alpha_2} & M_3 & \xrightarrow{\alpha_3} & M_4 \\ \downarrow{\scriptstyle f_1} & & \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle f_3} & & \downarrow{\scriptstyle f_4} \\ L_1 & \xrightarrow[\beta_1]{} & L_2 & \xrightarrow[\beta_2]{} & L_3 & \xrightarrow[\beta_3]{} & L_4 \end{array}$$

the rows are exact and each square commutes. If f_2 and f_4 are monomorphisms, and f_1 is an epimorphism, then f_3 is a monomorphism.

PROOF. The argument proceeds by "diagram chasing." Let $x \in M_3$ go to 0 under f_3. Then since the right square commutes, it goes to 0 under $f_4\alpha_3$ i.e., $\alpha_3(x)$ goes to zero under f_4. Since the latter is a monomorphism, $\alpha_3(x) = 0$. Hence by the exactness of the top row, $x = \alpha_2(y)$ for some $y \in M_2$. By commutativity, $f_2(y)$ goes to 0 under β_2 so by the exactness of the bottom row, we have $\beta_2(y) = \beta_1(z)$ for some $z \in L_1$. Since f_1 is an epimorphism, $z = f_1(u)$ for some $u \in M_1$. Since the left hand square commutes, f_2 takes $\alpha_1(u)$ to $f_2(y)$ so since f_2 is a monomorphism, we have $y = \alpha_1(u)$. Hence, $x = \alpha_2(y) = \alpha_2(\alpha_1(u)) = 0$. It follows that f_3 is a monomorphism as claimed. □

Note: The proof is much easier to follow by pointing at the diagram on a blackboard or on paper.

PROPOSITION. *Suppose in the diagram*

$$\begin{array}{ccccccc} M_2 & \xrightarrow{\alpha_2} & M_3 & \xrightarrow{\alpha_3} & M_4 & \xrightarrow{\alpha_4} & M_5 \\ \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle f_3} & & \downarrow{\scriptstyle f_4} & & \downarrow{\scriptstyle f_5} \\ M_2 & \xrightarrow[\beta_2]{} & M_3 & \xrightarrow[\beta_3]{} & M_4 & \xrightarrow[\beta_4]{} & M_5 \end{array}$$

the rows are exact and each square commutes. If f_2 and f_4 are epimorphisms, and f_5 is an monomorphism, then f_3 is an epimorphism.

PROOF. Exercise.

PROPOSITION. *Suppose in the diagram*

$$\begin{array}{ccccccccc} M_1 & \xrightarrow{\alpha_1} & M_2 & \xrightarrow{\alpha_2} & M_3 & \xrightarrow{\alpha_3} & M_4 & \xrightarrow{\alpha_4} & M_5 \\ \downarrow{\scriptstyle f_1} & & \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle f_3} & & \downarrow{\scriptstyle f_4} & & \downarrow{\scriptstyle f_5} \\ M_1 & \xrightarrow[\beta_1]{} & M_2 & \xrightarrow[\beta_2]{} & M_3 & \xrightarrow[\beta_3]{} & M_4 & \xrightarrow[\beta_4]{} & M_5 \end{array}$$

the rows are exact and each square commutes. If f_2 and f_4 are isomorphisms, f_1 is an epimorphism, and f_5 is a monomorphism then f_3 is an isomorphism.

Direct products and direct sums.

Let $\{M_i \mid i \in I\}$ be an indexed family of left A-modules (where the indexing set I can be finite or infinite.) By definition, the set theoretic product $\prod M_i$ of the family consists of all *sequences*

$$(x_i)_{i\in I} \text{ with } x_i \in M_i.$$

(Technically a "sequence" is a function from I to the union of the sets M_i such that x_i, the value of the function at i, belongs to M_i.) We may make the set theoretic product into a left A-module by defining

$$(x_i) + (y_i) = (x_i + y_i) \text{ and } a(x_i) = (ax_i).$$

It is easy to check that this provides a module structure, and we call this module the *direct product* module.

Consider the subset of the product module consisting of all sequences (x_i) such that $x_i = 0$ for all but a finite number of $i \in I$. It is not hard to see that this forms an A-submodule of $\prod M_i$. This A-module is called the *direct sum* module and it is denoted $\bigoplus_{i\in I} M_i$.

The two concepts introduced above could just as well have been defined for abelian groups or for arbitrary groups. For rings, however, the "sum" as defined above would not have been a subring because it would not have contained the identity. It is important to have a more abstract way to describe these objects so that

we may see analogies when going from one such "category" to another. We do this by exhibiting so called universal mapping properties.

First we consider the product. Let $P = \prod M_i$ and let $p_i : P \to M_i$ denote the projection of an element onto its ith component. Assume L is any other module for which we are given module homomorphisms $f_i : L \to M_i$ for each $i \in I$. Then there is a unique module homomorphism $f : L \to P$ such that

$$\begin{array}{ccc} & L & \\ & \searrow^{f_i} & \\ \mathrm{f}\downarrow & & M_i \\ & \nearrow_{p_i} & \\ & P & \end{array} \qquad \text{commutes for each } i \in I.$$

For, if there is such a homomorphism, we must have $p_i(f(z)) = f_i(z)$ for each $i \in I$. Thus, $f(z) = (f_i(z))_{i \in I}$ so f is unique. On the other hand, this formula defines a function f which is easily seen to be a module homomorphism.

Note that any other module P' with homomorphisms $p'_i : P' \to M_i$ having the same universal mapping property just described would have to be isomorphic to the product P. For, using the universal property, we would have unique homomorphisms between P and P' in both directions making the appropriate diagrams commute. The composition $P \to P' \to P$ would make all the appropriate triangles commute, but so would the identity $P \to P$ make all those diagrams commute. Hence, by uniqueness, that composition would be the identity. Similarly, the composition in the other direction would be the identity. In other words, the universal mapping property defines the product up to isomorphism. The only reason to define it in the specific way we did is to show that at least one such object exists.

Consider next the sum. Let $S = \bigoplus M_i$. For each i, define $h_i : M_i \to S$ by $h_i(v) = (x_i)$ where $x_j = v$ for $j = i$ and $x_j = 0$ for $j \neq i$. Let K be any other module for which there are module homomorphisms $g_i : M_i \to K$ for each $i \in I$. Then, there is a unique module homomorphism $g : S \to K$ such that

$$\begin{array}{ccc} & K & \\ & \nwarrow_{g_i} & \\ \mathrm{g}\uparrow & & M_i \\ & \swarrow_{s_i} & \\ & S & \end{array} \qquad \text{commutes for each } i \in I.$$

We leave the proof as an exercise for the student. (Use the fact that S is generated as a module by the submodules M_i.)

As with the product, the sum is in fact defined up to isomorphism by the universal mapping property.

In the case of a finite indexing set, the two notions are of course the same. In particular, we are interested in having a criterion for determining when a given module M is isomorphic to the direct sum of two modules $M \cong M' \oplus M''$. In that case there are homomorphisms

$$p' : M \to M', \quad p'' : M \to M''$$

and

$$h' : M' \to M, \quad h'' : M'' \to M$$

defined by composing the homomorphisms defined above with the given isomorphism, and a simple calculation shows that $p' \circ h' = id_{M'}$ and $p'' \circ h'' = id_{M'}$. For example, if M is actually equal to the sum (product), then h' sends $x' \in M'$ into $(x', 0)$ and p' projects this back onto x'.

Generally, given $p : M \to M''$, we say that $h : M'' \to M$ *splits* p if $p \circ h = id_{M''}$.

PROPOSITION. *Assume $p : M \to M''$ is split by $h : M'' \to M$. Then $M \cong \operatorname{Ker} p \oplus M''$.*

PROOF. Let $M' = \operatorname{Ker} p$. Define $h' : M' \to M$ to be the inclusion of M' in M. Let $p' : M \to M'$ be defined by $p'(x) = x - h(p(x))$. Since $p(p'(x)) = p(x) - p(h(p(x))) = p(x) - p(x) = 0$ (—use $ph = id_{M'}$—),

it follows that $p'(x) \in \operatorname{Ker} p = M'$ as required. We will show that the homomorphisms p and p' have the universal mapping property for a product. Note that for any $x \in M$, we have

$$x = p'(x) + h(p(x)).$$

Suppose we are given $q : M \to M''$ and $q' : K \to M'$. If there is a homomorphism $f : K \to M$ such that $pf = q$ and $p'f = q'$ then

$$f(x) = p'(f(x)) + h(p(f(x))) = q'(x) + h(q(x)).$$

Hence, if there is such an f, it is unique. On the other hand, it is straightforward to check that if f is defined by

$$f(x) = q'(x) + h(q(x))$$

then it is a homomorphism with the desired properties. □

Notes: It would have been just as easy to show that the homomorphisms h and h' satisfy the universal mapping property for a sum. Also, in the representation

$$x = p'(x) + h(p(x)),$$

we have $p'(x) \in \operatorname{Ker} p = \operatorname{Im} h'$ and $h(p(x)) \in \operatorname{Im} h = \operatorname{Ker} p'$. Moreover, $\operatorname{Ker} p \cap \operatorname{Im} h = \{0\}$ so that M is the direct product of the two submodules $\operatorname{Ker} p$ and $\operatorname{Im} h$ is the sense described in group theory.

Exercises.

1. Suppose in the diagram

$$\begin{array}{ccccccc} M_2 & \xrightarrow{\alpha_2} & M_3 & \xrightarrow{\alpha_3} & M_4 & \xrightarrow{\alpha_4} & M_5 \\ \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle f_3} & & \downarrow{\scriptstyle f_4} & & \downarrow{\scriptstyle f_5} \\ L_2 & \xrightarrow[\beta_2]{} & L_3 & \xrightarrow[\beta_3]{} & L_4 & \xrightarrow[\beta_4]{} & L_5 \end{array}$$

the rows are exact and each square commutes. If f_2 and f_4 are epimorphims, and f_5 is an monomorphism, show that f_3 is an epimoprhism.

2. Free Modules

Let A be a ring. As usual, we shall discuss left A-modules. The corresponding theory for right modules is exactly the same except for notation. Recall in the case that A is a field that every module (vector space) has a basis. You were probably only shown the proof of this fact in the case of finitely generated modules (finite dimensional vector spaces), but it holds quite generally. (See the Exercises.) We want to develop a corresponding theory for modules over an arbitrary ring. Unfortunately, it is not generally true that every module over a general ring has a basis. Those modules for which it is true are called *free*.

We proceed as in the case of vector spaces. Let M be a left A-module. A subset X of M is said to generate (or span) M if and only if $M = AX$—i.e., if and only if every element $z \in M$ may be written $\sum a_x x$ where the sum ranges over X and all but a finite number of coefficients $a_x = 0$. As in the case of vector spaces such a sum is called a *linear combination* of the elements of X. A set X is said to be *linearly independent* if and only if $\sum a_x x = 0 \Rightarrow a_x = 0$ for all $x \in X$. If X is a linearly independent subset of M, then $\sum a_x x = \sum b_x x$ (where the coefficients on both sides are in A and all but a finite number are 0) implies $a_x = b_x$ for all $x \in X$. A subset X of M is called a *basis* if it is linearly independent and it generates M. Then every element in M may be expressed uniquely in the form $\sum a_x x$ where $a_x \in A$ and $a_x = 0$ for all but a finite number of $x \in X$. As mentioned above, M is said to be *free* if it has a basis.

Suppose M is free with basis X. Consider the family of modules $\{A_x \mid x \in X\}$ where for each $x \in X$, A_x is just A viewed as a left module over itself. Let S be the direct sum of this family as defined in the previous section. Define $\phi : M \to S$ by $\phi(z) = (a_x)_{x \in X}$ where $z = \sum a_x x$ is the unique representation of $z \in M$ in

terms of the basis. Since $a_x = 0$ for all but a finite number of $x \in X$, it follows that $\phi(x) \in S$. Also, it is easy to see that ϕ is a module homomorphism. In the other direction, define $\mu : S \to M$ by $\mu((a_x)) = \sum a_x x$. The result is well defined because the argument is in S and all but a finite number of coefficients are 0. Also, it is a module homomorphism, and it is clear that ϕ and μ are inverses so they are isomorphisms. Hence, if M is free, then it is isomorphic to a direct sum of copies of A. Conversely, it is not hard to see that any such direct sum is free. For if $S = \bigoplus_{i \in I} A_i$ where each $A_i = A$, consider the element d_i which has all components 0 except the ith component which is 1. It is easy to check that the set $\{d_i \mid i \in I\}$ is a basis for S. Note that by this construction, we may form a free module with a basis $\{d_i \mid i \in I\}$ in one-to-one correspondence with an arbitrary set I. Often we oversimplify and identify d_i with i so that we may think of I as imbedded in S as a basis. With that convention, we call S the free module on the set I. (Question for thought: do you think a direct product of copies of A is always free?) Let F be a free module with basis X. Since each element of F is uniquely expressible as a linear combination of the basis elements, a homomorphism from F into another module is determined completely by its values on the basis and these values may be specified arbitrarily.

PROPOSITION. *Suppose F is a free module and $g : M \to M''$ is an epimorphism of modules. For each module homomorphism $f'' : F \to M''$ there is a module homomorphism $f : F \to M$ such that $f'' = g \circ f$.*

$$\begin{array}{ccc} & & F \\ & \exists f \swarrow & \downarrow f'' \\ M & \xrightarrow{g} & M'' \longrightarrow 0 \end{array}$$

PROOF. Choose a basis X for F. As mentioned above, we may define a homomorphism by specifying it on a basis. For each $x \in X$, choose $z \in M$ such that $g(z) = f''(x)$ and set $f(x) = z$. Then $g(f(x)) = f''(x)$ for each x in a basis, and hence $g \circ f = f''$ as required. □

Let M be any module. We may represent it as an epimorphic image of a free module as follows. Let X be any generating set for M. (There is always at least one such set, namely M itself, but usually we can find a generating set considerably smaller.) Let F be the free module on a set Y in one-to-one correspondence with X. Since Y may be viewed as imbedded in F, we may define a homomorphism $f : F \to M$ by sending an element $y \in Y$ to the corresponding element $x \in X$. Since any element of M can be written as a linear combinations of elements of X, it is clear that f is an epimorphism. Let R be the kernel of f. It follows that $M \cong F/R$. All this is reminiscent of the theory of free groups. By analogy with that case R is called the module of *relations*, and we say that we have a *presentation* of M as a factor module of a free module.

PROPOSITION. *Let F be a module. If F is free then every epimorphism $f : M \to F$ splits.*

PROOF. Suppose F is free and consider the diagram

$$\begin{array}{ccc} & & F \\ & \exists g \swarrow & \downarrow \mathrm{id} \\ M & \xrightarrow{f} & F \longrightarrow 0 \end{array}$$

The homomorphism which we know exists by the previous proposition splits f. □

Exercises.

1. Let V be a vector space over the field k. Show that V has basis. Hint: Use Zorn's Lemma to construct a maximal linearly independent set. Show that such a set must be a basis.

2. Let A be a ring and let $\mathfrak{a}$ be a 2-sided ideal in A. Let M be an A-module and let $X \subseteq M$. Set $\bar{X} = \{x + \mathfrak{a}M \mid x \in X\}$ be the corresponding subset of the $A/\mathfrak{a}$-module $M/\mathfrak{a}M$.
(a) Show that $\bar{X}$ generates $M/\mathfrak{a}M$ over $A/\mathfrak{a}$ if X generates M over A.

(b) Suppose X is a basis for M. Show that the mapping

$$x \mapsto x + \mathfrak{a}M$$

is one-to-one. Show also that $\bar{X}$ is linearly independent over $A/\mathfrak{a}$.

Conclude that $\bar{X}$ is a basis for $M/\mathfrak{a}M$.

3. Vector spaces

We assume that you are familiar with the basic concepts of linear algebra. In particular, you should know that every finitely generated vector space has a basis and that the number of elements in such a basis is an invariant of the vector space called the dimension. You should know the relationship between linear transformations and matrices. You should know something about the theory of determinants for square matrices (and the corresponding linear transformations.) In particular, you should know that a square matrix over a field is invertible if and only if its determinant is non-zero. We shall investigate some of these questions and related questions for other rings than fields, but we will assume familiarity with the field case and not dwell on these topics as much as we might were they totally unfamiliar.

4. Rank

Let A be a *commutative* ring and let F be a finitely generated free module. We shall show that the number of elements in a basis for F depends only on F, not on the basis. (The corresponding result is true for non-finitely generated free modules but is harder to prove because of the need to deal with transfinite cardinals.) The proof is based on a trick which reduces the problem to one for vector spaces where we already know what is true.

Let A be any ring and suppose $\mathfrak{a}$ is a 2-sided ideal in A. Suppose M is a left A-module; consider the submodule $\mathfrak{a}M$ and the factor module $M/\mathfrak{a}M$. Since each element of $\mathfrak{a}$ clearly acts as the 0-endomorphism of $M/\mathfrak{a}M$, it is not hard to see that $M/\mathfrak{a}M$ becomes a module over the factor ring $A/\mathfrak{a}$ in a natural way. Also, one may prove that if $X \subseteq M$ is linearly independent or generates M as an A-module, the same is true for the set $\bar{X} = \{x + \mathfrak{a}M \mid x \in X\}$. (See the Exercises.) It follows that if M is free over A with basis X, then $M/\mathfrak{a}M$ is free over $A/\mathfrak{a}$ with basis $\bar{X}$.

Suppose now that F is free and finitely generated over a commutative ring A. We know that there is at least one maximal ideal $\mathfrak{m}$ in A. Apply the above analysis to $F/\mathfrak{m}F$ as a module—i.e., vector space—over $A/\mathfrak{m}$—which is a field. By the invariance of dimension for vector spaces, we know that the number of elements in a basis for $F/\mathfrak{m}F$ does not depend on the basis. Since any basis for F over A yields a basis of the same size for $F/\mathfrak{m}F$ over $A/\mathfrak{m}$, we are done.

In the case of a free module over any arbitrary commutative ring, we call the number of elements in a basis the *rank* of the module. In the case of vector spaces rank and dimension mean the same thing.

If A is not commutative, we cannot necessarily define the concept of rank. The problem is that if $\mathfrak{m}$ is a maximal 2-sided ideal, all we can say is that $A/\mathfrak{m}$ is simple. It is possible to have a simple ring A which is isomorphic to the direct sum of two copies of itself as a left module. Hence the number of elements in a basis would not be independent of the basis for such a ring. However, for many important classes of non-commutative rings, it is possible to define the concept of rank for free modules. We shall study one such class in the last third of this course.

Suppose F is free over A of rank n. Then as we saw earlier, F is isomorphic to the direct sum of n copies of A. We denote that direct sum by A^n. As is common in the case of vector spaces, we may exhibit the elements of A^n either as n by 1 matrices (or column vectors) or as 1 by n matrices (or row vectors) with entries in A. Denote by e_i the element of A^n with i^{th} component 1 and all others 0. Then $\{e_i \mid i = 1, 2, \ldots, n\}$ is a basis for A^n called the canonical basis.

Exercises.

1. Show that $\mathbf{Q}$ is not a free abelian group.

5. Torsion

Let A be a domain, and let M be an A-module. An element $x \in M$ is called a *torsion* element provided $\exists a \in A$ such that $a \neq 0$ and $ax = 0$. Denote by M_t the set of all torsion elements of M. M_t is submodule of M. For, if $ax = 0$ and $by = 0$ then $ab(x + y) = 0$. Also, if $ax = 0$ then $a(bx) = 0$ for any $b \in A$.

We say that M is *torsion free* if $M_t = \{0\}$—i.e., if 0 is the only torsion element. We say M is a *torsion* module if $M = M_t$. The module M/M_t is always torsion free. For, $a(x + M_t) = ax + M_t = M_t$ (the 0 coset) if and only if $ax \in M_t$. However, if ax is a torsion element, there is $b \in A (b \neq 0)$ such that $b(ax) = 0$ so $ba \neq 0$ kills x and $x \in M_t$. Thus $x + M_t$ is the 0 coset as required.

Examples:

1) $A = \mathbf{Z}$. Any abelian group is a $\mathbf{Z}$-module. Any finite abelian group is a torsion module. The additive group of the rationals $\mathbf{Q}$ is torsion free, but it is not free. (See the Exercises.) The group $\mathbf{Q}/\mathbf{Z}$ is a torsion module, but it is not finite. (It is not every finitely generated. Proof?)
2) Let V be a finite dimensional vector space over a field k. Let $B = \mathrm{End}_k(V)$ denote the ring of all linear transformations (module homomorphisms) of V into itself. (As usual, B may be identified with the ring of n by n matrices with entries in k.) Then V may be viewed as a left module over B where $fx = f(x)$ for $f \in B$ and $x \in V$. B is of course not a domain; it is not even commutative if $n > 1$. Let $f : V \to V$ be a fixed linear transformation of V, and consider the subring $k[f]$ of B generated by f. $k[f]$ is commutative, but it is still not a domain. We rectify this situation by considering $k[f]$ as the epimorphic image of $A = k[X]$ (where X is an indeterminate) under the ring homomorphism defined by $g(X) \mapsto g(f)$ for each $g(X) \in k[X]$. Then V becomes a left A-module under the operation defined by

$$g(X)v = g(f)(v) \quad \text{for } g(X) \in k[X] \text{ and } v \in V.$$

V is a torsion module over A. For, given any $v \in V$, the elements

$$v, f(v), f(f(v) = f^2(v), \ldots, f^i(v), \ldots$$

cannot form a linearly independent set, since V is finite dimensional. Hence, some linear combination

$$a_0v + a_1f(v) + \cdots + a_kf^k(v) = 0$$

i.e.,

$$(a_0\mathrm{id} + a_1f + \cdots + a_kf^k)v = 0$$

i.e.,

$$(a_0 + a_1X + \cdots + a_kX^k)v = 0.$$

Note that our primary interest in this example is the structure of V in reference to the linear transformation f. Our resort to the polynomial ring $k[X]$ is just to be able to work with a domain and its attendant advantages.

Exercises.

1. Prove that a free module over a domain is torsion free.

6. Modules over a PID

The structure of finitely generated modules over a PID is particularly easy to describe. Since $\mathbf{Z}$ is a PID, that means that we have a good structure theorem for finitely generated abelian groups. Similarly, since $k[X]$ is a PID if k is a field, using the analysis at the end of the previous section, we obtain a good structure theorem for the behavior of a vector space with respect to a fixed linear transformation.

We shall use two facts about PIDs which were developed while proving that every PID is a UFD. Let A be a PID.

(a) Let $d = \gcd(a, b)$ in A. Then the equation $ax + by = d$ has a solution x, y in A.
(b) Any strictly increasing chain of ideals in A must terminate. (Equivalently, any sequence $d_1, d_2, \ldots$, where at each stage, d_{i+1} is a non-associate divisor of d_i, must stop.)

THEOREM. *Let A be a PID. Any submodule M of a finitely generated free A-module F is also free. Moreover,*

$$\mathit{rank}(M) \leq \mathit{rank}(F).$$

Note: The theorem is in fact true for arbitrary free modules over a PID but the non-finitely generated case is harder to prove.

PROOF. We proceed by induction on $n = \operatorname{rank}(F)$. The theorem is clearly true for free modules of rank 0. (What does it mean to say that the empty set is a basis for a module?)

Let X be a basis for F with n elements and pick out a subset X' of X with $n-1$ elements. The submodule F' of F spanned by X' clearly has X' as a basis so it is free of rank $n-1$. Also, we have

$$F \cong F' \oplus A.$$

Denote by $p'' : F \to A$ the corresponding projection onto the second summand A. Note that $F' = \operatorname{Ker} p''$. Also, $p''(M)$ is an A-submodule of A—that is $p''(M)$ is an ideal in A. Since A is a PID, $p''(M) = Ax''$ for some $x'' \in A$. Consider the short exact sequence

$$0 \longrightarrow \operatorname{Ker} p'' \cap M \longrightarrow M \longrightarrow Ax'' \longrightarrow 0.$$

Since Ax'' is free, the sequence must split, that is

$$M \cong Ax'' \oplus (\operatorname{Ker} p'' \cap M).$$

However, $\operatorname{Ker} p'' \cap M = F' \cap M$ is a submodule of F' so by induction, we may assume it is free and has rank $\leq n-1$. Since a direct sum of free modules is certainly free and since the ranks add, we are done. □

Suppose now that M is an arbitrary finitely generated module over the PID A. We know that we can find a finitely generated free module F and an epimorphism $\phi : F \to M$. Let $R = \ker \phi$. To determine the structure of M, we shall investigate the isomorphic module F/R. Since both F and R are free, we shall attempt to capitalize on this fact by picking convenient bases for them. In fact, we shall show that it is always possible to pick a basis for F such that appropriate multiples of its elements form a basis for R.

First, choose a basis $\{x_1, \ldots, x_n\}$ for F. (Ordinarily, you would choose the basis so that it is carried onto some specified set of generators of M under the presentation $F \to M$.) Similarly, choose a spanning set $\{y_1, \ldots, y_m\}$ for R. Eventually, we will find a basis for R, but allowing a spanning set at this stage allows a bit more flexibilty. We may write

$$y_i = \sum_{j=1}^{n} s_{ji} x_j \qquad i = 1, 2, \ldots, m.$$

The coefficients s_{ji} form an $n \times m$ matrix S with entries in the ring A. Since A is commutative, and we need not distinguish between left and right modules, we may also write this as

$$y_i = \sum_{j=1}^{n} x_j s_{ji} \qquad i = 1, 2, \ldots, m,$$

which in turn may be summarized in a pseudo-matrix equation

$$[\, y_1 \quad y_2 \quad \ldots \quad y_m \,] = [\, x_1 \quad x_2 \quad \ldots \quad x_n \,] S.$$

The entries in the 'x' and y' matrices are elements of F, but the products are computed by the usual rule for multiplication of matrices.

Suppose we change to another basis $\{x'_1, x'_2, \ldots, x'_n\}$ for F. Again using pseudo-matrix notation, since the old basis elements can be expressed uniquely in terms of the new basis elements, we can write

$$[\, x_1 \quad x_2 \quad \ldots \quad x_n \,] = [\, x'_1 \quad x'_2 \quad \ldots \quad x'_n \,] P$$

where P is an $n \times n$ matrix with entries in A. Note that P must necessarily be an invertible matrix in $M_n(A)$, since we can reverse the roles of the two bases to express the new basis elements in terms of the old basis by means of a matrix P'. That P and P' are inverses follows easily by noting that the products PP' and $P'P$ express each basis *uniquely* in terms of itself.

Consider next spanning sets for R with m elements related to the first spanning set by

$$[\, y'_1 \quad y'_2 \quad \dots \quad y'_m \,] = [\, y_1 \quad y_2 \quad \dots \quad y_m \,]\, Q$$

where Q is an $m \times m$ *invertible* matrix in the ring $M_m(A)$. Every spanning set for R may be obtained by such an equation if we drop the assumption that Q is invertible. Indeed, it need not even be a square matrix, i.e., the number of y' need not be m. But we shall restrict attention to spanning sets for R obtained by invertible Q.

Putting these pseudo-matrix equations together, we obtain

$$[\, y'_1 \quad y'_2 \quad \dots \quad y'_m \,] = [\, x'_1 \quad x'_2 \quad \dots \quad x'_n \,]\, PSQ$$

We shall show below that the $n \times m$ matrix $S' = PSQ$ may be made *diagonal* by suitable choice of the invertible matrices P and Q. Just what we mean by 'diagonal' will be clear from what follows.

Suppose first that $n \geq m$. In this case, S' will have the form

$$S' = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & d_m \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

where there are $n - m$ rows of zero at the bottom (or no such rows if $n = m$). Then the pseudo-matrix equation

$$[\, y'_1 \quad y'_2 \quad \dots \quad y'_m \,] = [\, x'_1 \quad x'_2 \quad \dots \quad x'_n \,]\, S'$$

tells us that

$$y'_1 = d_1 x'_1, y'_2 = d_2 x'_2, \dots, y'_m = d_m x'_m.$$

If some of the d_i are zero, we may clearly assume that they occur at the end, and that $d_1, d_2, \dots, d_s$ are all nonzero for some s. Then it is clear that the set of $y'_i = d_i x_i, i = 1, \dots, s$ span R (since the remaining $y'_i = 0$). It is also not hard to see that they form a linearly independent set (since A is a domain). Hence, they form a basis for R.

Suppose instead that $n < m$. Then S' has the form

$$S' = \begin{bmatrix} d_1 & 0 & \dots & 0 & 0 & \dots 0 \\ 0 & d_2 & \dots & 0 & 0 & \dots 0 \\ \vdots & \vdots & \dots & \vdots & 0 & \dots 0 \\ 0 & 0 & \dots & d_n & 0 & \dots 0 \end{bmatrix}$$

where there are $m - n$ columns of zero at the end. In this case, we have

$$y'_1 = d_1 x'_1, y'_2 = d_2 x'_2, \dots, y'_n = d_n x'_n, y'_{n+1} = \dots = y'_m = 0.$$

Again we can assume some of the $d_i = 0$ and place these at the end. As above, we find that a subset of the y'_i consisting of multiples of the corresponding x'_i forms a basis for R.

Assuming that a matrix with entries in a PID may be diagonalized as described above, we have the following result.

THEOREM. *Let A be a PID and suppose M is a finitely generated module over A. Then M is isomorphic to a direct sum of modules*

$$M \cong A/d_1A \oplus A/d_2A \oplus \cdots \oplus A/d_rA$$

where $d_1, d_2, \ldots, d_r$ are non-units in A. (Some of the d_i may be zero.)

Note. A module isomorphic to A/dA for some $d \in A$ is called a *cyclic* module. It is clearly generated by a single element.

PROOF. Assume as above that $d_1, d_2, \ldots, d_s$ are non-zero and $d_s, d_{s+1}, \ldots, d_n = 0$. Then

$$F = Ax_1' \oplus Ax_2' \oplus \ldots Ax_s' \oplus \ldots Ax_n'$$
$$R = d_aAx_1' \oplus d_2Ax_2' \oplus \ldots d_sAx_s' \oplus \ldots 0$$

It follows easily that

$$F/R \cong A/d_1A \oplus A/d_2A \oplus \cdots \oplus A/d_nA.$$

(The last $n - s$ terms yield the direct sum of that many copies of A. i.e., a free A-module of rank $n - s$.) If any d_i is a unit, then $A/d_iA \cong 0$, and we may omit that term.

We must still prove the required diagonalization theorem. In so doing, we shall add an additional requirement about the d_i; that after suitable rearrangment each divides the next. Then the d_i turn out to be unique and are called the *invariant factors* of the matrix S.

THEOREM. *(Invariant Factors Theorem) Let A be a PID and let S be an $n \times m$ matrix with entries in A. There exist an invertible $n \times n$ matrix P and an invertible $m \times m$ matrix Q such that $S' = PTQ$ is diagonal and moreover the diagonal entries satisfy $d_1 \mid d_2 \mid \cdots \mid d_l$. Moreover, the diagonal entries so derived are unique up to associates.*

PROOF. We first address the issue of the existence of a diagonalization. The uniqueness question will be deferred until later.

We proceed by applying *row and column operations* to the matrix S as in ordinary linear algebra over a field. Such operations may be performed by multiplying S on either the left or right by appropriate invertible *basic matrices*. There are four such types of basic matrices.

(i) The matrix $E_{i,j}(c)$ obtained by adding c times the ith row of the identiy matrix to the jth row, where $c \in A$. If we multiply a matrix by this matrix on the left, the result is the matrix obtained by adding c times the ith row to the jth row. If we multiply instead on the right, the result is the matrix obtained by adding c times the jth column to the ith column. (Note the switch between i and j.)

(ii) The matrix $E_i(c)$ obtained by multiplying the ith row of the identity matrix by the invertible element $c \in A$. If we multiply a matrix by this matrix on the left, the result is the matrix obtained by mulitplying the ith row by c. If we multiply instead on the right, the result is the matrix obtained by multiplying the ith column by c.

(iii) The matrix E_{ij} obtained by interchanging the ith and jth rows of the identity matrix. If a matrix is multiplied on the left by such a matrix, its ith and jth rows are interchanged. If a matrix is multiplied on the right instead, the corresponding columns are interchanged. Such a matrix is called a *permutation matrix*.

These matrices and the corresponding operations suffice for linear algebra over a field, indeed, as we shall see, over a Euclidean domain. They are called *elementary matrices* and the correspond row or column operations are called *elementary operations*. Each such matrix is invertible, and the corresponding row or column operation is reversible. But for a PID, we need one additional type of basic matrix.

(iv) Assume $a, b \in A$ and $d = \gcd(a, b)$. Then we may write $d = ax + by$ for appropriate $x, y \in A$. Consider

the matrix

$$\begin{array}{c} \\ i \\ \\ j \\ \\ \end{array} \begin{array}{c} \begin{array}{cccccccc} & & & i & & j & & \end{array} \\ \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & 1 & & & & & & \vdots \\ 0 & 0 & \dots & x & \dots & y & \dots & 0 \\ \vdots & & & & & & & \vdots \\ 0 & 0 & & -b/d & \dots & a/d & \dots & 0 \\ \vdots & & & & & & \vdots & \\ 0 & & & 0 & & 0 & & 1 \end{bmatrix} \end{array}$$

We leave it to the student to show this matrix is invertible.

If a matrix S has a in row i and some column and b in row j and the same column, then multiplying S by the above matrix on the left results in a matrix which has $ax + by = d$ in row i and that column and $(-b/d)a + b(a/d) = 0$ in row j and the same column. Simiarly, analogous results in a given row may be produced by multiplying by an appropriate matrix on the right.

To construct the desired invertible matrices P and Q, we perform an appropriate sequence of row and column operations. P then is then obtained by taking the product of the corresponding elementary matrices for the row operations and Q is obtained by the product of the corresponding column operations, and then *inverting*. (You should think about why the inversion is necessary.)

If $S = 0$, there is nothing to do, so assume it has at least one non-zero column.

Start with the first column of S. If it consists only of 0's, we may exchange it with a nonzero column. If the first entry of this column is zero, we may exchange the first row with a row in which the first entry is not zero. Consider the other non-zero entries in the first column. By multiplying by a sequence of type (iv) matrices on the left, we put the gcd of the set of non-zero entries in the first entry and make all other entries zero.

Apply the same reasoning to the first row of the new S. That will yield a matrix with all entries in the first row zero except for the entry in the $1, 1$ postion, which will divide the previous such entry. Unfortunately the needed row or column operations *may produce non-zero entries in the first column of the new S*. But we can iterate this procedure, alternating between clearing the first column and clearning the first row. At each stage the entry in the $1, 1$ postion will divide the corresponding entry at the previous stage. In a PID we can't have an infinite sequence of non-zero elements, each of which divides the previous element, unless at some point one of those elements is a unit. Hence, either the process stops because both the first row and first column have been cleared, or the first entry is a unit. In the latter case, we may just apply type (i) row or column operations to clear the first row or first column as needed without affecting the already cleared column or row. Call the entry in the $1, 1$ position d_1. Then S has been reduced to a matrix of the form

$$\begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & s_{2,2} & \dots & s_{2,m} \\ \vdots & \vdots & \dots & \vdots \\ 0 & s_{n,2} & \dots & s_{n,m} \end{bmatrix}$$

We may now proceed by induction on the size of the matrix to diagnalize S.

Finally we must show that the diagonal entries can be chosen so they divide each other in order. Consider d_1 and d_2. Choose s, $t \in A$ such that

$$d_1 s + d_2 t = d = \gcd(d_1, d_2).$$

Ignoring everything but the first two rows and columns, we have

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} = \begin{bmatrix} d_1 & d_2 \\ 0 & d_2 \end{bmatrix}$$

and

$$\begin{bmatrix} d_1 & d_2 \\ 0 & d_2 \end{bmatrix} \begin{bmatrix} s & -d_2/d \\ t & d_1/d \end{bmatrix} = \begin{bmatrix} d & 0 \\ d_2 t & d_1 d_2/d \end{bmatrix}$$

Note that $d \mid d' = d_1 d_2/d$, and $d \mid d_2 t$. By subtracting an appropriate multiple of the first row from the second, we may put the matrix in diagonal form with the first diagonal entry dividing the second. Arguing the same way for d_1 and each of the other diagonal entries, we see that we can arrange that d_1 divides each of the diagonal entries. To complete the argument, argue inductively on the submatrix starting in the $2, 2$ position.

We have now proved everything asserted in the theorem except for the uniqueness. The uniqueness may be proved by methods similar to those used. However, we shall use an alternate argument which will be developed in below. □

Note that the above process may actually be embodied in an effective algorithm provided we have a way to solve equations of the form $ax + by = d = \gcd(a, b)$ in A. In a Euclidean domain, we may always use the Euclidean algorithm to solve such equations. (If you have never encountered the Euclidean algorithm, you should look it up immediately in a book on elementary number theory.) Hence for Euclidean domains, the above argument is actually a *constructive* proof of the existence of invariant factors. In fact, for a Euclidean domain, one may accomplish the reduction to diagonal form without use of the type (iv) imbedded 2 by 2 matrices; the elementary row and column operations suffice because the steps in the Euclidean algorithm can be accomplished by subtraction of multiples.

Let us return to the cyclic decompostion

$$M \cong A/d_1A \oplus A/d_2A \oplus \dots A/d_sA \oplus A^{n-s}$$

obtained previously. The last $n-s$ summands (if any) give a free module as indicated. The first s summands clearly form a torsion module since any element in that sum is killed by $d_1 \dots d_s$. Using the direct sum decomposition, it is not hard to see that every torsion element is contained in this sum so the torsion submodule of M

$$M_t \cong A/d_1A \oplus A/d_2A \oplus \dots \oplus A/d_sA \text{ and } M/M_t \cong A^{n-s}.$$

If M is torsion free, i.e., $M_{=}\{0\}$, then $M \cong A^n$ is free. Thus, we have

Corollary. *Every finitely generated torsion free module over a PID is free.*

If M is not finitely generated, the corresponding result is false. For, take $A = \mathbf{Z}$ and $M = \mathbf{Q}$.

Uniqueness questions. By the above theorems, we know that we may write a finitely generated module over a PID as

$$M \cong A/d_1A \oplus A/d_2A \oplus \dots \oplus A/d_sA \oplus A^{n-s}$$

where we may assume $d_1 \mid d_2 \mid \dots \mid d_s$ and no d_i is a unit. We shall show that under this assumption the invariant factors $d_1, d_2, \dots, d_s$ are unique up to associates. From this it will follow that the invariant factors in the diagonalization of a matrix over A are unique up to associates if they satisfy the divisibility condition. For associated with any diagonalization of the matrix is a decomposition of the appropriate F/R, so the module theorem implies the matrix theorem. (There are a few technical difficulties having to do with the number of invariants of the matrix which are units. That number can be recovered from the available data. The details are left to the student.)

We start with a few more general comments which are interesting in their own right. Let A be a PID, and suppose that M is an A-module. For each irreducible element p in A, define

$$M_p = \{x \in M \mid p^i x = 0 \text{ for some positive integer } i\}.$$

It is a routine task to check that M_p is a submodule of M.

THEOREM. *Let A be a PID and let M be a A-module. Then $M_t = \bigoplus_p M_p$ where the sum is taken over a set of irreducible elements of A containing one irreducible for each equivalence class of associates.*

PROOF. Let $M' = \bigoplus_p M_p$. Clearly, M' is a submodule of M_t. Let $x \in M_t$. Then the set $\{a \in A \mid ax = 0\}$ is clearly an ideal so it is of the form Ad for some $d \in A$. Ad is called the *order* ideal of x, and we say x has order d. d is of course only unique up to associates. Let p be an irreducible factor of d, and suppose $d = p^e d'$ where d' is relatively prime to p. Then we can solve the equation

$$p^e s + d't = 1$$

in A. We have

$$x = p^e sx + d'tx$$

and $p^e d'tx = 0$ so $d'tx \in M_p$. Also, $x' = p^e sx$ satisfies $d'x = 0$. Continuing in this way, we may write x as a sum of elements in various M_p for non-associate primes plus an additional element whose order divides the order of the additional element obtained at the previous stage. Since we can't have an infinite sequence of proper divisors in a PID, the process must stop with a decomposition of x as desired. This shows that M_t is the sum (possibly not direct) of the M_p.

To show that the sum is direct, it suffices to show that a representation of the form

$$x = \sum x_p \text{ where } x_p \in M_p,$$

and where p ranges over a set of irreducibles as above, is unique. Suppose

$$\sum x_p = \sum y_p$$

where x_p and $y_p \in M_p$ and as usual all but a finite number of terms in each sum are zero. Then

$$\sum (x_p - y_p) = 0.$$

For each term in the sum, pick a power m_p of p to kill that term. Fix an irreducible q and let $m' = \prod m_p$ where the product is over all p occurring in the sum *except* $p = q$. Then m' kills off every term in the sum except $x_q - y_q$. Hence, we must have $m'(x_q - x_q) = 0$ also. However, if $q^e(x_q - y_q) = 0$, since $\gcd(m', q^e) = 1$, it is easy to see that $x_q - y_q = 0$ as required. □

M_p is called the *p-primary* component of M, and the above decomposition is called the *primary decomposition*.

We approach the problem of uniqueness by noting that $n - s$ is uniquely determined by M since it is the rank of M/M_t which is free. Hence, we need only show how to recover the invariants $d_1, \ldots, d_s$ from M_t. Suppose

$$d_s = p_1^{e_{s1}} p_2^{e_{s2}} \ldots p_m^{e_{sm}}$$

is a decomposition of d_s as a product of powers of non-associate irreducible elements of A. Since A is a PID (hence UFD), this representation is unique up to order and factors which are units. Since we may clearly always change to an associate, we shall ignore unit factors. Then for $i = 1, 2, \ldots, s-1$, we may write

$$d_i = p_1^{e_{i1}} p_2^{e_{i2}} \ldots p_m^{e_{im}}$$

where because of the divisibility conditions we have

$$0 \le e_{1j} \le e_{2j} \le \cdots \le e_{sj} \quad \text{for } j = 1, 2, \ldots, m.$$

(Here we have again used the fact that d_i can always be changed by a unit.) Suppose now that

$$M_t = Ax_1 \oplus Ax_2 \oplus \cdots \oplus Ax_s$$

where x_i has order d_i. Consider the primary decomposition of Ax_i

$$Ax_i = M_{i1} \oplus M_{i2} \oplus \cdots \oplus M_{im}$$

where M_{ij} is the p_j-primary component of Ax_i. Then since as a direct summand, M_{ij} is an epimorphic image of $Ax_i \cong A/d_iA$, it is clear that M_{ij} is of the form A/hA for some $h \in A$. Since every element of M_{ij} is of order a power of p_j, it is not hard to see that h is an associate of

$$p_j^{e_{ij}}$$

which is the exact power of p_j dividing d_i. Also, by rearranging the terms in the direct sum decomposition of M_t, we have

$$M_t = N_1 \oplus N_2 \oplus \cdots \oplus N_m$$

where

$$N_j = M_{1j} \oplus M_{2j} \oplus \cdots \oplus M_{kj}.$$

Since N_j is clearly p_j-primary, it is not hard to see that N_j is in fact the p_j-primary component of M. Thus, N_j is determined just by M, not by any particular direct sum decomposition of M. Hence, to show that the d_i are unique up to associates, it will suffice to show that N_j completely determines the orders

$$p_j^{e_{ij}}$$

of the M_{ij} for $i = 1, 2, \ldots, s$. This follows from the following lemma.

LEMMA. *Let A be a PID and suppose p is an irreducible element of A. Suppose also that M is an A-module with a direct sum decomposition*

$$M = Ax_1 \oplus Ax_2 \oplus \cdots \oplus Ax_s$$

where x_i has order

$$p^{e_i} \quad i = 1, 2, \ldots, s$$

where $e_1 \leq e_2 \leq \cdots \leq e_s$. Then $e_1, \ldots, e_s$ are determined by M and not by the particular direct sum decomposition.

PROOF. Consider the chain of submodules

$$\ldots p^{h+1}M \subseteq p^hM \subseteq \cdots \subseteq p^2M \subseteq pM \subseteq M.$$

Since p kills each factor module

$$p^jM/p^{j+1}M$$

it follows that the factor module is in fact a module over A/pA. Since p is irreducible, and since A is a PID, it follows that pA is maximal and A/pA is a field. Hence, $p^jM/p^{j+1}M$ is a vector space over A/pA. Let $l_j = \dim p^jM/p^{j+1}M$ for $j = 0, 1, 2, \ldots$ Now use the direct sum decomposition assumed in the statement of the lemma, and the fact that

$$\dim p^jAx_i/p^{j+1}Ax_i = 0 \text{ or } 1$$

depending on whether or not $p^jx_i = 0$, i. e. on whether or not $j \geq e_i$. It is clear that by determining the differences $l_j - l_{j+1}$ for $j = 0, 1, \ldots$ we may determine the numbers e_i. □

THEOREM. *Let M be a finitely generated module over a PID A. The invariants $d_1 \mid d_2 \mid \cdots \mid d_s$ (none a unit or zero) associated with the direct sum decomposition*

$$M \cong A/d_1A \oplus A/d_2A \oplus \cdots \oplus A/d_sA \oplus A^{n-s}$$

are uniquely (up to associates) determined by M. Also $n - s = \operatorname{rank}(M/M_t)$.

Application to Linear Algebra. One may use the above theory to derive the usual results about canonical forms in linear algebra. We shall outline briefly how to do this, but we shall not attempt to recover all those theorems, which you should have seen in a Linear Algebra course.

Let k be a field and let V be a finite dimensional vector space over k. Let $f : V \to V$ be a k-linear function on V, and let $\{v_1, v_2 \ldots, v_n\}$ be a basis for V over k. Then the matrix C of f is defined by

$$f(v_i) = \sum_{j=1}^{n} c_{ji} v_j \qquad i = 1, 2, \ldots, n.$$

As before, let $k[X]$ act on V by $Xv = vX = f(v)$. Consider the free $k[X]$-module $F = k[X]^n$ with the standard basis elements e_i chosen with ith component 1 and the other components zero. (Then e_i may be visualized as the ith column of the identity matrix in $M_n(k[X])$.) Map F onto V by sending $e_i \mapsto v_i$, and let R be the kernel. Then

$$\sum_{j=1}^{n} c_{ji} e_j - X e_i \mapsto \sum_{j=1}^{n} c_{ji} v_j - f(v_i) = 0$$

for each $i = 1, 2, , \ldots, n$. Consider the *characteristic matrix* with entries in $k[X]$:

$$C(X) = \begin{bmatrix} c_{11} - X & a_{12} & \ldots & c_{1,n} \\ c_{21} & c_{22} - X & \ldots & c_{2,n} \\ \vdots & \vdots & \ldots & \vdots \\ c_{n1} & c_{n2} & \ldots & c_{nn} - X \end{bmatrix}$$

By the above calculation, its columns are in R. In fact, they span R. For, let R' denote the submodule of R they do span, and let $V' = F/R'$. Note that $V = F/R$ is an epimorphic image of V'. From the Invariant Factors Theorem, there are invertible matrices P and Q in $M_n(k[X])$ such that $PC(X)Q$ is a diagonal $n \times n$ matrix with $d_1(X)|d_2(X)| \ldots |d_n(X)$ on the diagonal. (There must be n invariant factors by degree considerations. Some might be units; indeed some must be units as we shall see below.) Take determinants; we get

$$d_1(X) d_2(X) \ldots d_n(X) = \text{unit} \det(C(X)).$$

But $\det(C(X))$ is the characteristic polynomial of the original linear function f, so it is not the zero polynomial. Moreover, the sum of the degrees of the $d_i(X)$ must add up to the degree of the characteristic polynomial, which is n. It follows that none of the $d_i(X)$ is zero, and

$$V' \cong k[X]/d_1(X) \oplus k[X]/d_2(X) \oplus \cdots \oplus k[X]/d_n(X).$$

It is not hard to see that $\dim_k k[X]/d_i(X) = \deg d_i(X)$, so it follows that $\dim_k V' = n$. Hence, $V' \cong V$, and $R' = R$.

From this analysis, we can derive a canonical form for f. Namely, find the invariant factors of the characteristic matrix $C(X)$. Each term $k[X]/d_i(X)$ in the cyclic decomposition of F/R corresponds to an subspace of V invariant under f. Moreover with respect to an appropriate basis, f restricted to this subspace has the same matrix as multiplication by X does on $k[X]/d_i(X)$. However, the latter matrix is quite easy to exhibit. It is what is called the *companion matrix* to the polynomial $d_i(X)$. In fact, by keeping track of the row and column operations in $M_n(k[X])$, one determine just what the appropriate bases are in V.

Exercises.

1. Let M be the direct product of denumerably many copies of $\mathbf{Z}$. Show that M is not free by following the steps outlined below.

(a) Take as true the fact, stated but not proved in the section, that any subgroup N of a free abelian group M is free, whether M is finitely generated or not. Under this assumption, assume M is free, fix a prime p, and consider the subgroup N of M consisting of all sequences

$$(a_1, a_2, \ldots, a_n, \ldots)$$

such that the power of p dividing a_n goes to infinity as n goes to infinity or such that $a_n = 0$ for all sufficiently large n. Then under the stated assumptions, N is also free. Show that if $\{x_i\}_{i\in I}$ is a basis for N then the set of cosets $\{\bar{x}_i\}_{i\in I}$ in N/pN is a basis for the latter vector space over the field $\mathbf{Z}/p\mathbf{Z}$.

(b) With the notation as in (a), let e_i be the element of N which is one in the ith position and zero elsewhere. Show that N/pN has a denumerable basis over $\mathbf{Z}/pZ$. Under the assumption that all bases for a vector space have the same cardinality, show that every basis of N is denumerable.

(c) From (b), conclude that N is denumerable. Show on the other hand that N contains a subgroup with the same cardinality as M which by the Cantor diagonlization argument is non-denumerable. Hint: Consider the monomorphism $M \to M$ which sends

$$(a_1, a_2, \dots, a_n, \dots) \mapsto (pa_1, p^2a_2, \dots, p^n a_n, \dots).$$

2. Let p be a prime number and let J_p be the subgroup of $\mathbf{Q}/\mathbf{Z}$ consisting of all elements of order a power of p. Verify from the primary decomposition theorem that $\mathbf{Q}/\mathbf{Z}$ is isomorphic to the direct sum of the subgroups J_p with p ranging over the set of primes. Show that each J_p is not a finitely generated abelian group.

3. Suppose M is the abelian group generated by $\{x_1, x_2, x_3\}$ subject to the relations

$$2x_1 + 4x_2 - 2x_3 = 0$$
$$5x_1 - 3x_2 + 2x_3 = 0.$$

Find a decomposition of M as a direct sum of cyclic groups.

4. Find the invariant factors over $\mathbf{Z}$ of the matrix

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

5. Use the theory described in the section to show that if $V = \mathbf{R}^2$ and $f : V \to V$ is given with respect to the standard basis by the matrix

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$$

then V is a cyclic $\mathbf{R}[X]$-module. What is its order? Find a matrix representation for f as a companion matrix.