1

INTRODUCTION TO SOME BASIC IDEAS

1.1 General remarks concerning rings

It is assumed that the reader is familiar with the notions of *group* and *ring*. Nevertheless a few remarks concerning these concepts may help to establish certain conventions about notation and to clarify the attitude adopted in regard to certain extreme situations. Suppose then that R is a ring. The elements of R may be *added* and *multiplied* to give other elements of the same system. With respect to addition, R is a commutative group. The zero element of this group will be denoted by 0 although we may sometimes write 0_R if more than one ring is under consideration and we wish to be quite explicit. So far as multiplication is concerned, if α, β, γ are arbitrary elements of R, then we have the associative law $(\alpha\beta)\gamma = \alpha(\beta\gamma)$ and the two distributive laws $\alpha(\beta+\gamma) = \alpha\beta+\alpha\gamma$ and $(\beta+\gamma)\alpha = \beta\alpha+\gamma\alpha$. The ring R is said to be *commutative* if multiplication is commutative, that is to say if the relation $\alpha\beta = \beta\alpha$ always holds. In the early part of this book a good deal of the theory developed will apply only to modules over these more restricted rings, but in this chapter we shall not require that the commutative law be satisfied.

An element e belonging to R is called an *identity element* if

$$e\alpha = \alpha = \alpha e$$

for every α in R. It is immediately clear that a ring has at most one identity element. We shall confine our attention entirely to those rings in which such an element is present. The reader should therefore note that, *from this point onwards, when we speak of a ring it is to be understood that we always mean a ring with an identity element*. The identity element will be denoted by 1, though we may sometimes embellish this by writing 1_R if other rings are also being considered and we wish to avoid confusion.

Let R be a ring with zero element 0 and identity element 1. If $0 = 1$ and $\alpha \in R$, then $\alpha = \alpha 1 = \alpha 0 = 0$ and therefore 0 is the only element in the ring. In these circumstances, we say that R is a *null ring*.

Although null rings have only a trivial theory, it is convenient not to exclude them from consideration. For in this way certain results become not only more general, but also (and this is more important) easier to formulate.

1.2 Modules

We are now ready to introduce the central concept of our subject. Let R be a ring and M an additive group†. Suppose that, given an element r of R and an element x of M, we have some rule whereby we can form a 'product' rx which is again an element of M. (We can express this more formally in the following way. Let $R \times M$ consist of all ordered pairs (r, x), where $r \in R$ and $x \in M$, so that $R \times M$ is the so-called *Cartesian product* of R and M. Our supposition now amounts to assuming that we have a mapping of $R \times M$ into M together with the convention that the image of the pair (r, x) is to be denoted by rx.) What is now required is that the product rx shall have some natural connection with the ring structure of R and the group structure of M. The precise requirements are set out in the following

Definition. *If the situation is as described above, then M is said to be a 'left R-module' or a 'left module with respect to R' provided that the following four conditions are satisfied*:

(i) $(r_1+r_2)x = r_1x+r_2x$ *whenever* r_1, r_2 *belong to* R *and* x *belongs to* M;
(ii) $r(x_1+x_2) = rx_1+rx_2$ *whenever* $r \in R$ *and* x_1, x_2 *belong to* M;
(iii) $(r_1r_2)x = r_1(r_2x)$ *whenever* r_1, r_2 *belong to* R *and* x *belongs to* M;
(iv) $1x = x$ *for all* x *in* M.

In (iv), the symbol 1 denotes the identity element of R.

Naturally, there is an analogous concept called a *right R-module*. In this, the product of $r \in R$ and $x \in M$ is written as xr and (i), (ii), (iii) and (iv) are replaced by

(i)′ $x(r_1+r_2) = xr_1+xr_2$ *when* $x \in M$ *and* $r_1, r_2 \in R$;
(ii)′ $(x_1+x_2)r = x_1r+x_2r$ *when* $x_1, x_2 \in M$ *and* $r \in R$;
(iii)′ $x(r_1r_2) = (xr_1)r_2$ *when* $r_1, r_2 \in R$ *and* $x \in M$;
(iv)′ $x1 = x$ *for all* $x \in M$.

It will be clear that to every result concerning left R-modules there is a corresponding result for right R-modules and vice versa.

† The term 'additive group' is used here to mean a *commutative* group in which the law of composition is written as addition.

Of course, whenever one is concerned *simultaneously* with left modules and right modules, then it is necessary to make careful distinctions. When, however, all our modules have the ring operating on the same side, we shall normally develop the theory in terms of left modules. This will be the case for the remainder of the present chapter. The reader should therefore note that, *for the remainder of Chapter* 1, *the term 'R-module' will always signify a left R-module*, as defined above, unless there is a definite statement to the contrary.

Let M be an R-module. The zero element of M will usually be denoted by 0. However, it may be necessary, on occasion, to distinguish between the zero element of M and that of R. We then employ the symbols 0_M and 0_R.

Proposition 1. *Let M be an R-module. Then for all $r \in R$ and $x \in M$, we have*

(*a*) $0_R x = 0_M$;
(*b*) $r0_M = 0_M$;
(*c*) $(-r)x = -(rx) = r(-x)$;
(*d*) $(-r)(-x) = rx$.

Proof. Since $0_R + 0_R = 0_R$, the definition of an R-module shows that

$$0_R x = (0_R + 0_R)x = 0_R x + 0_R x,$$

whence $0_R x = 0_M$ because M is a group. Next, from $0_M + 0_M = 0_M$ follows

$$r0_M = r(0_M + 0_M) = r0_M + r0_M,$$

whence $r0_M = 0_M$. Thus (*a*) and (*b*) are proved. Again $0_R = r + (-r)$ and therefore, by (*a*),

$$0_M = 0_R x = (r + (-r))x = rx + (-r)x$$

which yields $(-r)x = -(rx)$. A similar argument shows that

$$r(-x) = -(rx).$$

Finally, using (*c*), we obtain

$$(-r)(-x) = -((-r)x) = -(-(rx)) = rx.$$

This completes the proof.

1.3 Homomorphisms and isomorphisms

Let M and N be modules with respect to a ring R and let $f: M \to N$ be a mapping of M *into* N so that with each element x of M there is associated a definite element $f(x)$ of N.

Definition. *The mapping $f: M \to N$ is said to be an 'R-homomorphism' or a 'homomorphism of R-modules' if it satisfies the following two conditions*:

(i) $f(x_1+x_2) = f(x_1)+f(x_2)$ *whenever* $x_1,\ x_2$ *belong to* M;

(ii) $f(rx) = rf(x)$ *whenever* $r \in R$ *and* $x \in M$.

An R-homomorphism $f: M \to N$ is also known as an *R-linear* mapping. Sometimes we speak simply of a homomorphism of M into N when it is quite clear which is the ring of operators.

Let $f: M \to N$ be a homomorphism of the R-module M into the R-module N. Since $f(x_1+x_2) = f(x_1)+f(x_2)$ whenever x_1, x_2 are in M, we see, in particular, that f is a homomorphism of the additive group of M into the additive group of N. Certain consequences follow from this fact alone. To begin with, from $0_M + 0_M = 0_M$, we obtain

$$f(0_M)+f(0_M) = f(0_M)$$

and therefore

$$f(0_M) = 0_N. \tag{1.3.1}$$

Again, if $x \in M$, then $0_M = x+(-x)$. Applying the mapping f and making use of (1.3.1), we see that $0_N = f(x)+f(-x)$ whence

$$f(-x) = -f(x). \tag{1.3.2}$$

Finally, if $x, y \in M$, then $y-x = y+(-x)$ so that $f(y-x) = f(y)+f(-x)$ which, by virtue of (1.3.2), yields

$$f(y-x) = f(y)-f(x). \tag{1.3.3}$$

In certain situations, homomorphisms can be combined. For example, if $f: M \to N$ and $g: N \to P$ are homomorphisms of R-modules, then we obtain a mapping $h: M \to P$ by first applying f and afterwards applying g. Thus, by definition, $h(x) = g(f(x))$ from which it is easily verified that $h(x_1+x_2) = h(x_1)+h(x_2)$ and $h(rx) = rh(x)$ whenever x_1, x_2, x belong to M and r belongs to R. The new mapping is therefore a homomorphism. It is called the *product* of the original mappings and is denoted by† gf. Observe that

$$(gf)(x) = g(f(x)) \tag{1.3.4}$$

for all $x \in M$.

There are certain kinds of homomorphisms which play particularly important roles.

Definition. *A homomorphism $f: M \to N$ of R-modules is called a 'monomorphism', an 'injection' or an 'embedding' if distinct elements of M always have distinct images in N.*

† Note the order of the terms.

Thus the characteristic property of a monomorphism $f: M \to N$ is that $x_1 \neq x_2$ (where $x_1, x_2 \in M$) implies that $f(x_1) \neq f(x_2)$. We shall shortly give a different characterization but first we need another

Definition. *Let $f: M \to N$ be a homomorphism of R-modules. The set of elements of M which are mapped on to the zero element of N is called the 'kernel' of f and is denoted by* $\mathrm{Ker} f$.

It is clear, from (1.3.1), that $\mathrm{Ker} f$ always contains 0_M.

Lemma 1. *Let $f: M \to N$ be a homomorphism of R-modules. Then in order that f should be a monomorphism it is necessary and sufficient that* $\mathrm{Ker} f$ *contain only the element* 0_M.

Proof. If f is a monomorphism, then 0_M maps into 0_N. Also if $x \in M$ and $x \neq 0_M$, then $f(x) \neq 0_N$. Thus 0_M is the only element of $\mathrm{Ker} f$. Now assume that $\mathrm{Ker} f = \{0_M\}$, i.e. that $\mathrm{Ker} f$ is the set whose only member is 0_M. If, in this situation, $x_1, x_2 \in M$ and $f(x_1) = f(x_2)$, then, by (1.3.3),

$$f(x_2 - x_1) = f(x_2) - f(x_1) = 0_N.$$

and therefore $x_2 - x_1$ belongs to $\mathrm{Ker} f$. Accordingly $x_2 - x_1 = 0_M$ or $x_2 = x_1$. This completes the proof.

Definition. *A homomorphism $f: M \to N$ of R-modules is called an 'epimorphism' or a 'surjection' if each element of N is the image of at least one element of M.*

Thus $f: M \to N$ is an epimorphism if f maps M *on to* N. The reader will see, as the subject develops, that the notion of an epimorphism is, in a certain sense, dual to that of a monomorphism.

Suppose now that we have a homomorphism $f: M \to N$ which is both a monomorphism and an epimorphism. Then distinct elements of M have distinct images in N and each element of N is the image of at least one (and therefore of exactly one) element of M. In other terms, the homormorphism $f: M \to N$ gives a *one-one* mapping of the set M on to the set N. It follows that there exists a well defined inverse mapping $f^{-1}: N \to M$. *We contend that f^{-1} is also a homomorphism of R-modules.*

For suppose that y_1, y_2 belong to N and that r belongs to R. We can then choose $x_1, x_2 \in M$ so that $f(x_1) = y_1$ and $f(x_2) = y_2$. Then

$$f(x_1 + x_2) = f(x_1) + f(x_2) = y_1 + y_2$$

and therefore

$$f^{-1}(y_1 + y_2) = x_1 + x_2 = f^{-1}(y_1) + f^{-1}(y_2).$$

Again, from $f(rx_1) = rf(x_1) = ry_1$ follows

$$f^{-1}(ry_1) = rx_1 = rf^{-1}(y_1).$$

This establishes the contention.

It is convenient to embody some of these observations in the

Definition. *A homomorphism $f: M \to N$, of R-modules, which is both a monomorphism and an epimorphism is called an 'isomorphism' of M on to N. In such a situation, f will set up a one-one correspondence between the set M and the set N. The uniquely determined inverse mapping $f^{-1}: N \to M$ is then an isomorphism of N on to M.*

If $f: M \to N$ is an isomorphism and $f^{-1}: N \to M$ is the inverse isomorphism, then it is clear that f is the inverse of f^{-1}.

Two R-modules M and N are said to be *isomorphic* if there exists an isomorphism of M on to N. As already observed, this relation is symmetrical for there will then exist an isomorphism of N on to M. The symbol $M \approx N$ is frequently used to indicate that M and N are isomorphic.

We take this opportunity to introduce some simple but useful terminology. If A is a subset of a set B, then we obtain a mapping $j: A \to B$ by putting $j(a) = a$ for every $a \in A$. This mapping is called the *inclusion mapping* of A into B. In the special case where $A = B$, this yields the *identity mapping* of B. It will be convenient to denote the identity mapping of B by i_B. Thus $i_B(b) = b$ for every b in B.

If M is an R-module, then i_M is an isomorphism of M on to itself, hence $M \approx M$. We have already observed that if $M \approx N$ then $N \approx M$. Again, if $f: M \to N$ is an isomorphism of M on to N and $g: N \to P$ is an isomorphism of N on to P, then $gf: M \to P$ is an isomorphism of M on to P. Thus $M \approx N$ and $N \approx P$ together imply that $M \approx P$. This shows that $\approx$ has the properties of an equivalence relation. Indeed, from our point of view, isomorphic R-modules are simply copies of one another and have identical properties.

Lemma 2. *Let $f: M \to N$ and $g: N \to M$ be homomorphisms of R-modules. In order that f should be an isomorphism whose inverse is g, it is necessary and sufficient that $gf = i_M$ and $fg = i_N$.*

Proof. It is clear that if f is an isomorphism and $g = f^{-1}$, then both gf and fg are identity maps. Now suppose that $gf = i_M$, $fg = i_N$. If $y \in N$, then $y = i_N(y) = f(g(y))$ so that y is the image of $g(y)$. Accord-

ingly f is an epimorphism. Now assume that $x, x' \in M$ and $f(x) = f(x')$. Then

$$x = i_M(x) = g(f(x)) = g(f(x')) = i_M(x') = x'.$$

This shows that f is also a monomorphism and hence an isomorphism. It is clear that $g = f^{-1}$.

1.4 Submodules

Let M and N be R-modules and suppose that M is a subset of N. It does not follow (from what has so far been said) that the module structure of M is in any way related to that of N but clearly we shall have an interesting situation if the two happen to be compatible. This leads to the

Definition. *Let the situation be as described above. We say that 'M is a submodule of N' if the inclusion mapping $M \to N$ is a homomorphism of R-modules. If, in addition, $M \neq N$ (so that M is strictly contained in N) then M is called a 'proper submodule' of N.*

For example, since the identity map of N is a homomorphism of R-modules, N is a submodule of itself. Again, if M is a submodule of N and N is a submodule of P, then M must be a submodule of P.

Suppose that M is a submodule of N. Then $M \subseteq N$ and the inclusion mapping is an R-homomorphism. Let x_1, x_2 belong to M and let r be an element of R. Since the image of the sum of x_1 and x_2 is the sum of their separate images, we see that $x_1 + x_2$ is the same whether we regard x_1, x_2 as elements of M or as elements of N. Also, the image of rx_1 is r times the image of x_1. The interpretation of this is that rx_1 is the same element whether we consider x_1 as belonging to M or as belonging to N. A similar observation applies to $x_1 - x_2$. Again, by (1.3.1), $0_M = 0_N$ which means that M and N have their zero element in common.

A submodule of N is, in particular, a non-empty subset of N. However not every non-empty subset of N can be endowed with the structure of a submodule. We have, in fact, the following result.

Proposition 2. *Let N be an R-module and A a non-empty subset of N. Then A can be given the structure of a submodule of N if and only if the following two conditions are both satisfied:*

(i) *whenever a_1 and a_2 belong to A, then $a_1 + a_2$ also belongs to A;*
(ii) *whenever a belongs to A and r belongs to R, then ra belongs to A.*

Proof. If A is a submodule of N, then the remarks made earlier show that the conditions are satisfied. Now suppose that (i) and (ii) hold. By (i), the sum of any two elements of A is again an element of A and,

so far as A is concerned, addition is both commutative and associative because these laws hold in the larger system N. Since A is not empty, we can choose $\alpha \in A$ and then $0_R \alpha = 0_N$ belongs to A by virtue of (ii). Thus A contains an element which is neutral for addition. Furthermore, if $a \in A$, then $(-1)\,a$ belongs to A, by (ii), and we have $a + (-1)\,a = 0_N$. These remarks amount to a verification that A is an additive abelian group.

Finally, if $a \in A$ and $r \in R$, then, by (ii), the product ra also belongs to A. It is now clear that the module axioms are satisfied by A and that the inclusion mapping $A \to N$ is a homomorphism. Accordingly, A is a submodule of N.

If we take for A the subset $\{0\}$ consisting simply of the zero element of N, then it is immediately clear that (i) and (ii) hold and therefore $\{0\}$ is a submodule. Indeed, modules which have no non-zero elements play an important role in the general theory. Such a module is called a *null module* or a *zero module*.

In the next proposition, we consider an indexed family $\{L_i\}_{i \in I}$ of submodules of a given R-module N. This means that each L_i is a submodule of N and the individual submodules are labelled by means of the elements of a set I, called the *index set*. The index set may be completely arbitrary. In particular it is not required to contain only a finite number of members. If $i \neq i'$ belong to I, it is not assumed that L_i and $L_{i'}$ are necessarily distinct.

Proposition 3. *Let N be an R-module and $\{L_i\}_{i \in I}$ an indexed family of submodules of N. Then their intersection $\bigcap_{i \in I} L_i$ is also a submodule of N.*

Proof. Let $L = \bigcap_i L_i$ so that L consists of those elements which belong to every L_i. Since every submodule of N contains the zero element, we have $0 \in L$ and therefore L is not empty. To show that it is a submodule, we need only verify that conditions (i) and (ii) of Proposition 2 are satisfied. To this end, let x, y belong to L and let r belong to R. Then for each $i \in I$, we have $x \in L_i$ and $y \in L_i$. But L_i is a submodule, hence $x + y$ and rx also belong to L_i. This holds for every $i \in I$. Consequently $x + y \in L$ and $rx \in L$. Thus we have checked that the conditions of Proposition 2 are satisfied and the proof is complete.

Once again, suppose that $\{L_i\}_{i \in I}$ is a family of submodules of an R-module N. For each $i \in I$, choose $x_i \in L_i$ subject to the condition that, *for only finitely many i, shall x_i be different from zero.* We sometimes describe this situation by saying that $x_i = 0$ *for almost all i.*

In these circumstances we are able to form the sum $\sum_i x_i = x$ say, in spite of the fact that the index set I may be infinite. Let L consist of all elements x which can be obtained as sums in this way. *We contend that L is a submodule of N and that $L_i \subseteq L$ for every i.* For let $i_0 \in I$ and take x_{i_0} to be an arbitrary element of L_{i_0}. By putting $x_i = 0$ whenever $i \neq i_0$, we obtain a family $\{x_i\}_{i \in I}$ such that $x_i \in L_i$ and $\sum_i x_i = x_{i_0}$. Thus $x_{i_0} \in L$ and therefore $L_{i_0} \subseteq L$. In particular, L is not empty.

Now suppose that $y, z \in L$ and $r \in R$. To complete the proof we need only show that $y+z$ and ry belong to L. However, we can write $y = \sum_i y_i, z = \sum_i z_i$, where y_i, z_i belong to L_i and only a finite number of the y_i and z_i are non-zero. Then $y+z = \sum_i (y_i + z_i)$ whence, since $y_i + z_i \in L_i$ (L_i is a submodule), we have $y+z \in L$. Likewise $ry = \sum_i ry_i$ and $ry_i \in L_i$. Consequently $ry \in L$ and the proof is complete.

The submodule L, which has just been constructed, is called the *sum* of the L_i and is denoted by $\sum_{i \in I} L_i$. Not only does the sum contain each of the summands L_i, but it is clearly the *smallest submodule* of N which has this property.

Sometimes we are concerned with only a *finite* number of submodules, say $L_1, L_2, \ldots, L_s$. In such a situation it may be more convenient to use the alternative symbols $L_1 \cap L_2 \cap \ldots \cap L_s$ and $L_1 + L_2 + \ldots + L_s$ for their intersection and sum respectively.

Let U be a subset of an R-module M. By Proposition 3, the intersection L of all the submodules of M which contain U is also a submodule. *L, which is the smallest submodule containing U*, is called the *submodule generated by U*. If it happens that the submodule generated by U is M itself, then we say that *U is a system of generators for M*. Note that the submodule generated by the empty set is just $\{0_M\}$.

Proposition 4. *Let U be a subset of an R-module M and let $x \in M$. Then x belongs to the submodule generated by U if and only if we have a relation $x = r_1 u_1 + r_2 u_2 + \ldots + r_s u_s$, where the r_i are elements of R and the u_i belong to U.*

Remark. To cover the case where U is the empty set, we adopt the convention that an empty sum has the value zero.

Proof. Let L be the submodule generated by U, and L' the aggregate of all elements which can be written in the form $r_1 u_1 + r_2 u_2 + \ldots + r_s u_s$ with $r_i \in R$ and $u_i \in U$. It is clear that $L' \subseteq L$. Also, if $u \in U$, then

$u = 1u$ belongs to L' and therefore $U \subseteq L'$. Accordingly, if we can show that L' is a submodule of M, then (because L is the smallest submodule containing U) we shall have $L \subseteq L'$ and the proof will be complete.

Let $x, x^* \in L'$ and $r \in R$. By the definition of L, we can write

$$x = r_1 u_1 + \ldots + r_s u_s \quad \text{and} \quad x^* = r_1^* u_1^* + \ldots + r_t^* u_t^*,$$

where $r_i, r_j^* \in R$ and $u_i, u_j^* \in U$. Then

$$x + x^* = r_1 u_1 + \ldots + r_s u_s + r_1^* u_1^* + \ldots + r_t^* u_t^*$$

and

$$rx = (rr_1) u_1 + \ldots + (rr_s) u_s.$$

This shows that $x + x^*$ and rx both belong to L. Accordingly (Proposition 2) L' is a submodule of M and the required result follows.

Definition. *An R-module which can be generated by a finite number of elements is said to be 'finitely generated'. If an R-module can be generated by one element alone, then it is said to be 'singly generated.'*

The elements $x_1, x_2, \ldots, x_s$ generate an R-module M (it is supposed that the x_i belong to M) if and only if every element of the module can be expressed in the form $r_1 x_1 + r_2 x_2 + \ldots + r_s x_s$. Some important problems are connected with questions of finite generation. We must, however, postpone discussion of these for the time being.

Let $f: M \to N$ be a homomorphism of R-modules and suppose that A is a subset of M while B is a subset of N. We recall that $f(A)$ is used to denote the set of all elements of N which are images of elements of A, while $f^{-1}(B)$ denotes the set consisting of those elements of M whose images are contained in B. As is customary, we refer to $f(A)$ as the *image* of A and to $f^{-1}(B)$ as the *inverse image* of B. This provides an opportunity to introduce a concept which is complementary to $\operatorname{Ker} f$.

Definition. *If $f: M \to N$ is a homomorphism, then $f(M)$ is called the 'image' of the mapping f and is denoted by* $\operatorname{Im} f$.

Thus f is an epimorphism if and only if $\operatorname{Im} f = N$. Should $\operatorname{Im} f = \{0_N\}$, then we say that f is a *null homomorphism*.

Certain basic facts, about images and inverse images of submodules, will now be established.

Proposition 5. *Let $f: M \to N$ be a homomorphism of R-modules, let A be a submodule of M and B a submodule of N. Then $f(A)$ and $f^{-1}(B)$ are submodules of N and M respectively. In particular,* $\operatorname{Im} f$ *is a submodule of N while* $\operatorname{Ker} f$ *is a submodule of M. Finally, $f^{-1}(f(A)) = A + \operatorname{Ker} f$ and $f(f^{-1}(B)) = B \cap \operatorname{Im} f$.*