10.3 GENERATION OF MODULES, DIRECT SUMS, AND FREE MODULES

Let R be a ring with 1. As in the preceding sections the term "module" will mean "left module." We first extend the notion of the sum of two submodules to sums of any finite number of submodules and define the submodule generated by a subset.

Definition. Let M be an R-module and let $N_1, \ldots, N_n$ be submodules of M.

(1) The *sum* of $N_1, \ldots, N_n$ is the set of all finite sums of elements from the sets N_i: $\{a_1 + a_2 + \cdots + a_n \mid a_i \in N_i \text{ for all } i\}$. Denote this sum by $N_1 + \cdots + N_n$.

(2) For any subset A of M let

$$RA = \{r_1a_1 + r_2a_2 + \cdots + r_ma_m \mid r_1, \ldots, r_m \in R,\ a_1, \ldots, a_m \in A,\ m \in \mathbb{Z}^+\}$$

(where by convention $RA = \{0\}$ if $A = \emptyset$). If A is the finite set $\{a_1, a_2, \ldots, a_n\}$ we shall write $Ra_1 + Ra_2 + \cdots + Ra_n$ for RA. Call RA the *submodule of M generated by A*. If N is a submodule of M (possibly $N = M$) and $N = RA$, for some subset A of M, we call A a *set of generators* or *generating set* for N, and we say N *is generated by* A.

(3) A submodule N of M (possibly $N = M$) is *finitely generated* if there is some finite subset A of M such that $N = RA$, that is, if N is generated by some finite subset.

(4) A submodule N of M (possibly $N = M$) is *cyclic* if there exists an element $a \in M$ such that $N = Ra$, that is, if N is generated by one element:

$$N = Ra = \{ra \mid r \in R\}.$$

Note that these definitions do not require that the ring R contain a 1, however this condition ensures that A is contained in RA. It is easy to see using the Submodule Criterion that for any subset A of M, RA is indeed a submodule of M and is the smallest submodule of M which contains A (i.e., any submodule of M which contains A also contains RA). In particular, for submodules $N_1, \ldots, N_n$ of M, $N_1 + \cdots + N_n$ is just the submodule generated by the set $N_1 \cup \cdots \cup N_n$ and is the smallest submodule of M containing N_i, for all i. If $N_1, \ldots, N_n$ are generated by sets $A_1, \ldots, A_n$ respectively, then $N_1 + \cdots + N_n$ is generated by $A_1 \cup \cdots \cup A_n$. Note that cyclic modules are, a fortiori, finitely generated.

A submodule N of an R-module M may have many different generating sets (for instance the set N itself always generates N). If N is finitely generated, then there is a smallest nonnegative integer d such that N is generated by d elements (and no fewer). Any generating set consisting of d elements will be called a *minimal set of generators for N* (it is not unique in general). If N is not finitely generated, it need not have a minimal generating set.

The process of generating submodules of an R-module M by taking subsets A of M and forming all finite "R-linear combinations" of elements of A will be our primary way of producing submodules (this notion is perhaps familiar from vector space theory where it is referred to as taking the *span* of A). The obstruction which made the analogous process so difficult for groups in general was the noncommutativity of group

operations. For abelian groups, G, however, it was much simpler to control the subgroup $\langle A \rangle$ generated by A, for a subset A of G (see Section 2.4 for the complete discussion of this). The situation for R-modules is similar to that of abelian groups (even if R is a noncommutative ring) because we can always collect "like terms" in elements of A, i.e., terms such as $r_1a_1 + r_2a_2 + s_1a_1$ can always be simplified to $(r_1 + s_1)a_1 + r_2a_2$. This again reflects the underlying abelian group structure of modules.

Examples

(1) Let $R = \mathbb{Z}$ and let M be any R-module, that is, any abelian group. If $a \in M$, then $\mathbb{Z}a$ is just the cyclic subgroup of M generated by a: $\langle a \rangle$ (compare Definition 4 above with the definition of a cyclic group). More generally, M is generated as a $\mathbb{Z}$-module by a set A if and only if M is generated as a group by A (that is, the action of ring elements in this instance produces no elements that cannot already be obtained from A by addition and subtraction). The definition of finitely generated for $\mathbb{Z}$-modules is identical to that for abelian groups found in Chapter 5.

(2) Let R be a ring with 1 and let M be the (left) R-module R itself. Note that R is a finitely generated, in fact cyclic, R-module because $R = R1$ (i.e., we can take $A = \{1\}$). Recall that the submodules of R are precisely the left ideals of R, so saying I is a cyclic R-submodule of the left R-module R is the same as saying I is a principal ideal of R (usually the term "principal ideal" is used in the context of commutative rings). Also, saying I is a finitely generated R-submodule of R is the same as saying I is a finitely generated ideal. When R is a commutative ring we often write AR or aR for the submodule (ideal) generated by A or a respectively, as we have been doing for $\mathbb{Z}$ when we wrote $n\mathbb{Z}$. In this situation $AR = RA$ and $aR = Ra$ (elementwise as well). Thus a Principal Ideal Domain is a (commutative) integral domain R with identity in which every R-submodule of R is cyclic.

Submodules of a finitely generated module need not be finitely generated: take M to be the cyclic R-module R itself where R is the polynomial ring in infinitely many variables $x_1, x_2, x_3, \ldots$ with coefficients in some field F. The submodule (i.e., 2-sided ideal) generated by $\{x_1, x_2, \ldots\}$ cannot be generated by any finite set (note that one must show that *no* finite subset of this ideal will generate it).

(3) Let R be a ring with 1 and let M be the free module of rank n over R, as described in the first section. For each $i \in \{1, 2, \ldots, n\}$ let $e_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$, where the 1 appears in position i. Since

$$(s_1, s_2, \ldots, s_n) = \sum_{i=1}^{n} s_i e_i$$

it is clear that M is generated by $\{e_1, \ldots, e_n\}$. If R is commutative then this is a *minimal* generating set (cf. Exercises 2 and 27).

(4) Let F be a field, let x be an indeterminate, let V be a vector space over F and let T be a linear transformation from V to V. Make V into an $F[x]$-module via T. Then V is a *cyclic* $F[x]$-module (with generator v) if and only if $V = \{p(x)v \mid p(x) \in F[x]\}$, that is, if and only if every element of V can be written as an F-linear combination of elements of the set $\{T^n(v) \mid n \geq 0\}$. This in turn is equivalent to saying $\{v, T(v), T^2(v), \ldots\}$ span V as a vector space over F.

For instance if T is the identity linear transformation from V to V or the zero linear transformation, then for every $v \in V$ and every $p(x) \in F[x]$ we have $p(x)v = \alpha v$ for some $\alpha \in F$. Thus if V has dimension > 1, V cannot be a cyclic $F[x]$-module.

For another example suppose V is affine n-space and T is the "shift operator" described in Section 1. Let e_i be the i^{th} basis vector (as usual) numbered so that T is defined by $T^k(e_n) = e_{n-k}$ for $1 \leq k < n$. Thus V is spanned by the elements $e_n, T(e_n), \ldots, T^{n-1}(e_n)$, that is, V is a cyclic $F[x]$-module with generator e_n. For $n > 1$, V is not, however, a cyclic F-module (i.e., is not a 1-dimensional vector space over F).

Definition. Let $M_1, \ldots, M_k$ be a collection of R-modules. The collection of k-tuples $(m_1, m_2, \ldots, m_k)$ where $m_i \in M_i$ with addition and action of R defined componentwise is called the *direct product* of $M_1, \ldots, M_k$, denoted $M_1 \times \cdots \times M_k$.

It is evident that the direct product of a collection of R-modules is again an R-module. The direct product of $M_1, \ldots, M_k$ is also referred to as the *(external) direct sum* of $M_1, \ldots, M_k$ and denoted $M_1 \oplus \cdots \oplus M_k$. The direct product and direct sum of an infinite number of modules (which are different in general) are defined in Exercise 20.

The next proposition indicates when a module is isomorphic to the direct product of some of its submodules and is the analogue for modules of Theorem 9 in Section 5.4 (which determines when a group is the direct product of two of its subgroups).

Proposition 5. Let $N_1, N_2, \ldots, N_k$ be submodules of the R-module M. Then the following are equivalent:

(1) The map $\pi : N_1 \times N_2 \times \cdots \times N_k \to N_1 + N_2 + \cdots + N_k$ defined by

$$\pi(a_1, a_2, \ldots, a_k) = a_1 + a_2 + \cdots + a_k$$

is an isomorphism (of R-modules): $N_1 + N_2 + \cdots + N_k \cong N_1 \times N_2 \times \cdots \times N_k$.

(2) $N_j \cap (N_1 + N_2 + \cdots + N_{j-1} + N_{j+1} + \cdots + N_k) = 0$ for all $j \in \{1, 2, \ldots, k\}$.

(3) Every $x \in N_1 + \cdots + N_k$ can be written *uniquely* in the form $a_1 + a_2 + \cdots + a_k$ with $a_i \in N_i$.

Proof: To prove (1) implies (2), suppose for some j that (2) fails to hold and let $a_j \in (N_1 + \cdots + N_{j-1} + N_{j+1} + \cdots + N_k) \cap N_j$, with $a_j \neq 0$. Then

$$a_j = a_1 + \cdots + a_{j-1} + a_{j+1} + \cdots + a_k$$

for some $a_i \in N_i$, and $(a_1, \ldots, a_{j-1}, -a_j, a_{j+1}, \ldots, a_k)$ would be a nonzero element of $\ker \pi$, a contradiction.

Assume now that (2) holds. If for some module elements $a_i, b_i \in N_i$ we have

$$a_1 + a_2 + \cdots + a_k = b_1 + b_2 + \cdots + b_k$$

then for each j we have

$$a_j - b_j = (b_1 - a_1) + \cdots + (b_{j-1} - a_{j-1}) + (b_{j+1} - a_{j+1}) + \cdots + (b_k - a_k).$$

The left hand side is in N_j and the right side belongs to $N_1 + \cdots + N_{j-1} + N_{j+1} + \cdots + N_k$. Thus

$$a_j - b_j \in N_j \cap (N_1 + \cdots + N_{j-1} + N_{j+1} + \cdots + N_k) = 0.$$

This shows $a_j = b_j$ for all j, and so (2) implies (3).