14

Modules and vector spaces

In this chapter, we introduce the basic definitions and results concerning modules over a ring R and vector spaces over a field F. The reader may have seen some of these notions before, but perhaps only in the context of vector spaces over a specific field, such as the real or complex numbers, and not in the context of, say, finite fields like $\mathbb{Z}_p$.

14.1 Definitions, basic properties, and examples

Throughout this section, R denotes a ring.

Definition 14.1. *An R-**module** is an abelian group M, which we shall write using additive notation, together with a **scalar multiplication** operation that maps $a \in R$ and $\alpha \in M$ to an element $a\alpha \in M$, such that the following properties are satisfied for all $a, b \in R$ and $\alpha, \beta \in M$:*

(i) $a(b\alpha) = (ab)\alpha$,

(ii) $(a + b)\alpha = a\alpha + b\alpha$,

(iii) $a(\alpha + \beta) = a\alpha + a\beta$,

(iv) $1_R\alpha = \alpha$.

One may also call an R-module M a **module over** R. Elements of R are often referred to as **scalars**, and elements of M may be called **vectors**.

Note that for an R-module M, for fixed $a \in R$, the map that sends $\alpha \in M$ to $a\alpha \in M$ is a group homomorphism with respect to the additive group operation of M; likewise, for fixed $\alpha \in M$, the map that sends $a \in R$ to $a\alpha \in M$ is a group homomorphism from the additive group of R into the additive group of M.

The following theorem summarizes a few basic facts which follow directly from the observations in the previous paragraph, and basic facts about group homomorphisms (see Theorem 8.20):

Theorem 14.2. *If M is a module over R, then for all $a \in R$, $\alpha \in M$, and $m \in \mathbb{Z}$, we have:*

(i) $0_R\alpha = 0_M$,

(ii) $a0_M = 0_M$,

(iii) $(-a)\alpha = -(a\alpha) = a(-\alpha)$,

(iv) $(ma)\alpha = m(a\alpha) = a(m\alpha)$.

Proof. Exercise. □

The definition of a module includes the **trivial** module, consisting of just the zero element 0_M. If R is the trivial ring, then any R-module is trivial, since for all $\alpha \in M$, we have $\alpha = 1_R\alpha = 0_R\alpha = 0_M$.

***Example* 14.1.** A simple but extremely important example of an R-module is the set $R^{\times n}$ of n-tuples of elements of R, where addition and scalar multiplication are defined component-wise—that is, for $\alpha = (a_1, \ldots, a_n) \in R^{\times n}$, $\beta = (b_1, \ldots, a_n) \in R^{\times n}$, and $a \in R$, we have

$$\alpha + \beta = (a_1 + b_1, \ldots, a_n + b_n) \text{ and } a\alpha = (aa_1, \ldots, aa_n). \quad \Box$$

***Example* 14.2.** The ring of polynomials $R[\mathtt{X}]$ over R forms an R-module in the natural way, with addition and scalar multiplication defined in terms of the addition and multiplication operations of the polynomial ring. □

***Example* 14.3.** As in Example 9.34, let f be a monic polynomial over R of degree $\ell \geq 0$, and consider the quotient ring $E := R[\mathtt{X}]/(f)$. Then E is a module over R, with addition defined in terms of the addition operation of R, and scalar multiplication defined by $a[g]_f := [ag]_f$, for $a \in R$ and $g \in R[\mathtt{X}]$. If $f = 1$, then E is trivial. □

***Example* 14.4.** If E is any ring containing R as a subring (i.e., E is an extension ring of R), then E is a module over R, with addition and scalar multiplication defined in terms of the addition and multiplication operations of E. □

***Example* 14.5.** If $M_1, \ldots, M_n$ are R-modules, then so is the direct product $M_1 \times \cdots \times M_n$, where addition and scalar product are defined component-wise. □

***Example* 14.6.** Any abelian group G, written additively, can be viewed as a $\mathbb{Z}$-module, with scalar multiplication defined in terms of the usual integer multiplication map (see parts (vi)–(viii) of Theorem 8.3). □

***Example* 14.7.** Let G be any group, written additively, whose exponent divides n. Then we may define a scalar multiplication that maps $[m]_n \in \mathbb{Z}_n$ and $\alpha \in G$ to $m\alpha$. That this map is unambiguously defined follows from the fact that G has exponent dividing n, so that if $m \equiv m' \pmod{n}$, we have $m\alpha - m'\alpha = (m - m')\alpha = 0_G$, since $n \mid (m - m')$. It is easy to check that this scalar multiplication operation indeed makes G into a $\mathbb{Z}_n$-module. □

***Example* 14.8.** Of course, viewing a group as a module does not depend on whether or not we happen to use additive notation for the group operation. If we specialize the previous example to the group $G = \mathbb{Z}_p^*$, where p is prime, then we may view G as a $\mathbb{Z}_{p-1}$-module. However, since the group operation itself is written multiplicatively, the "scalar product" of $[m]_{p-1} \in \mathbb{Z}_{p-1}$ and $\alpha \in \mathbb{Z}_p^*$ is the power α^m. □

14.2 Submodules and quotient modules

Again, throughout this section, R denotes a ring. The notions of subgroups and quotient groups extend in the obvious way to R-modules.

Definition 14.3. *Let M be an R-module. A subset N is a* ***submodule*** *of M if*

(i) N is a subgroup of the additive group M, and

(ii) N is closed under scalar multiplication; that is, for all $a \in R$ and $\alpha \in N$, we have $a\alpha \in N$.

It is easy to see that a submodule N of an R-module M is also an R-module in its own right, with addition and scalar multiplication operations inherited from M.

Expanding the above definition, we see that a subset N of M is a submodule if and only if for all $a \in R$ and all $\alpha, \beta \in N$, we have

$$\alpha + \beta \in N, \quad -\alpha \in N, \quad \text{and} \quad a\alpha \in N.$$

Observe that the condition $-\alpha \in N$ is redundant, as it is implied by the condition $a\alpha \in N$ with $a = -1_R$.

For $m \in \mathbb{Z}$, it is easy to see (verify) that not only are mM and $M\{m\}$ subgroups of M (see Theorems 8.6 and 8.7), they are also submodules of M. Moreover, for $a \in R$, $aM := \{a\alpha : \alpha \in M\}$ and $M\{a\} := \{\alpha \in M : a\alpha = 0_M\}$ are also submodules of M (verify).

Let $\alpha_1, \ldots, \alpha_n$ be elements of M. In general, the subgroup $\langle \alpha_1, \ldots, \alpha_n \rangle$ will not be a submodule of M. Instead, let us consider the set $\langle \alpha_1, \ldots, \alpha_n \rangle_R$, consisting of all R-**linear combinations** of $\alpha_1, \ldots, \alpha_n$, with coefficients taken from R:

$$\langle \alpha_1, \ldots, \alpha_n \rangle_R := \{a_1\alpha_1 + \cdots + a_n\alpha_n : a_1, \ldots, a_n \in R\}.$$

It is not hard to see (verify) that $\langle \alpha_1, \ldots, \alpha_n \rangle_R$ is a submodule of M containing $\alpha_1, \ldots, \alpha_n$; it is called the submodule **spanned** or **generated** by $\alpha_1, \ldots, \alpha_n$. Moreover, it is easy to see (verify) that any submodule containing $\alpha_1, \ldots, \alpha_n$ must contain $\langle \alpha_1, \ldots, \alpha_n \rangle_R$. As a matter of definition, we allow $n = 0$, in which case, the spanned submodule is $\{0_M\}$.

If N_1 and N_2 are submodules of M, then $N_1 + N_2$ and $N_1 \cap N_2$ are not only subgroups of M, they are also submodules of M (verify).

***Example* 14.9.** For integer $\ell \geq 0$, define $R[\mathtt{X}]_{<\ell}$ to be the set of polynomials of degree less than ℓ. The reader may verify that $R[\mathtt{X}]_{<\ell}$ is a submodule of the R-module $R[\mathtt{X}]$. If $\ell = 0$, then this submodule is the trivial submodule $\{0_R\}$. □

***Example* 14.10.** Let G be an abelian group. As in Example 14.6, we can view G as a $\mathbb{Z}$-module in a natural way. Subgroups of G are just the same thing as submodules of G, and for $a_1, \ldots, a_n \in G$, the subgroup $\langle a_1, \ldots, a_n \rangle$ is the same as the submodule $\langle a_1, \ldots, a_n \rangle_{\mathbb{Z}}$. □

***Example* 14.11.** Any ring R can be viewed as an R-module in the obvious way, with addition and scalar multiplication defined in terms of the addition and multiplication operations of R. With respect to this module structure, ideals of R are just the same thing as submodules of R, and for $a_1, \ldots, a_n \in R$, the ideal $(a_1, \ldots, a_n)$ is the same as the submodule $\langle a_1, \ldots, a_n \rangle_R$. □

***Example* 14.12.** Let $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_m$ be elements of an R-module. Assume that each α_i can be expressed as an R-linear combination of $\beta_1, \ldots, \beta_m$. Then the submodule spanned by $\alpha_1, \ldots, \alpha_n$ is contained in the submodule spanned by $\beta_1, \ldots, \beta_m$.

One can see this in a couple of different ways. First, the assumption that each α_i can be expressed as an R-linear combination of $\beta_1, \ldots, \beta_m$ means that the submodule $\langle \beta_1, \ldots, \beta_m \rangle_R$ contains the elements $\alpha_1, \ldots, \alpha_n$, and so by the general properties sketched above, this submodule must contain $\langle \alpha_1, \ldots, \alpha_n \rangle_R$.

One can also see this via an explicit calculation. Suppose that

$$\alpha_i = \sum_{j=1}^{m} c_{ij}\beta_j \quad (i = 1, \ldots, n),$$

where the c_{ij} are elements of R. Then for any element γ in the submodule spanned by $\alpha_1, \ldots, \alpha_n$, there exist $a_1, \ldots, a_n \in R$ with

$$\gamma = \sum_{i=1}^{n} a_i\alpha_i = \sum_{i=1}^{n} a_i \sum_{j=1}^{m} c_{ij}\beta_j = \sum_{j=1}^{m} \Big(\sum_{i=1}^{n} a_i c_{ij} \Big) \beta_j,$$

and hence γ is contained in the submodule spanned by $\beta_1, \ldots, \beta_m$. □

If N is a submodule of M, then in particular, it is also a subgroup of M, and we can form the quotient group M/N in the usual way (see §8.3). Moreover, because N is closed under scalar multiplication, we can also define a scalar multiplication on M/N in a natural way. Namely, for $a \in R$ and $\alpha \in M$, we define

$$a \cdot (\alpha + N) := (a\alpha) + N.$$

As usual, one must check that this definition is unambiguous, that is, if $\alpha \equiv \alpha' \pmod{N}$, then $a\alpha \equiv a\alpha' \pmod{N}$. But this follows from the fact that N is closed under scalar multiplication (verify). One can also easily check (verify) that with scalar multiplication defined in this way, M/N is an R-module; it is called the **quotient module of M modulo N**.

14.3 Module homomorphisms and isomorphisms

Again, throughout this section, R is a ring. The notion of a group homomorphism extends in the obvious way to R-modules.

Definition 14.4. *Let M and M' be modules over R. An **R-module homomorphism** from M to M' is a map $\rho : M \to M'$, such that*

(i) ρ is a group homomorphism from M to M', and

(ii) for all $a \in R$ and $\alpha \in M$, we have $\rho(a\alpha) = a\rho(\alpha)$.

An R-module homomorphism is also called an **R-linear map**. We shall use this terminology from now on. Expanding the definition, we see that a map $\rho : M \to M'$ is an R-linear map if and only if $\rho(\alpha + \beta) = \rho(\alpha) + \rho(\beta)$ and $\rho(a\alpha) = a\rho(\alpha)$ for all $\alpha, \beta \in M$ and all $a \in R$.

Since an R-module homomorphism is also a group homomorphism on the underlying additive groups, all of the statements in Theorem 8.20 apply. In particular, an R-linear map is injective if and only if the kernel is trivial (i.e., contains only the zero element). However, in the case of R-module homomorphisms, we can extend Theorem 8.20, as follows:

Theorem 14.5. *Let $\rho : M \to M'$ be an R-linear map.*

(i) For any submodule N of M, $\rho(N)$ is a submodule of M'.

(ii) $\ker(\rho)$ is a submodule of M.

(iii) For any submodule N' of M', $\rho^{-1}(N')$ is a submodule of M.

Proof. Exercise. □

Theorems 8.21, 8.22, and 8.23 have natural R-module analogs:

Theorem 14.6. *If $\rho : M \to M'$ and $\rho' : M' \to M''$ are R-linear maps, then so is their composition $\rho' \circ \rho : M \to M''$.*

Proof. Exercise. □

Theorem 14.7. *Let $\rho_i : M \to M_i$, for $i = 1, \ldots, n$, be R-linear maps. Then the map $\rho : M \to M_1 \times \cdots \times M_n$ that sends $\alpha \in M$ to $(\rho_1(\alpha), \ldots, \rho_n(\alpha))$ is an R-linear map.*

Proof. Exercise. □

Theorem 14.8. *Let $\rho_i : M_i \to M$, for $i = 1, \ldots, n$, be R-linear maps. Then the map $\rho : M_1 \times \cdots \times M_n \to M$ that sends $(\alpha_1, \ldots, \alpha_n)$ to $\rho_1(\alpha_1) + \cdots + \rho_n(\alpha_n)$ is an R-linear map.*

Proof. Exercise. □

If an R-linear map $\rho : M \to M'$ is bijective, then it is called an **R-module isomorphism** of M with M'. If such an R-module isomorphism ρ exists, we say that M **is isomorphic to** M', and write $M \cong M'$. Moreover, if $M = M'$, then ρ is called an **R-module automorphism** on M.

Analogous to Theorem 8.24, we have:

Theorem 14.9. *If ρ is a R-module isomorphism of M with M', then the inverse function ρ^{-1} is an R-module isomorphism of M' with M.*

Proof. Exercise. □

Theorems 8.25, 8.26, 8.27, and 8.28 generalize immediately to R-modules:

Theorem 14.10. *If N is a submodule of an R-module M, then the natural map $\rho : M \to M/N$ given by $\rho(\alpha) = \alpha + N$ is a surjective R-linear map whose kernel is N.*

Proof. Exercise. □

Theorem 14.11. *Let ρ be an R-linear map from M into M'. Then the map $\bar{\rho} : M/\ker(\rho) \to \mathrm{img}(\rho)$ that sends the coset $\alpha + \ker(\rho)$ for $\alpha \in M$ to $\rho(\alpha)$ is unambiguously defined and is an R-module isomorphism of $M/\ker(\rho)$ with* $\mathrm{img}(\rho)$*.*

Proof. Exercise. □

Theorem 14.12. *Let ρ be an R-linear map from M into M'. Then for any submodule N contained in* $\ker(\rho)$*, the map $\bar{\rho} : M/N \to \mathrm{img}(\rho)$ that sends the coset $\alpha + N$ for $\alpha \in M$ to $\rho(\alpha)$ is unambiguously defined and is an R-linear map from M/N onto* $\mathrm{img}(\rho)$ *with kernel* $\ker(\rho)/N$*.*

Proof. Exercise. □

Theorem 14.13. *Let M be an R-module with submodules N_1, N_2. Then the map $\rho : N_1 \times N_2 \to N_1 + N_2$ that sends (α_1, α_2) to $\alpha_1 + \alpha_2$ is a surjective R-linear map. Moreover, if $N_1 \cap N_2 = \{0_M\}$, then ρ is an R-module isomorphism of $N_1 \times N_2$ with $N_1 + N_2$.*

Proof. Exercise. □

***Example* 14.13.** Let M be an R-module, and let m be an integer. Then the m-multiplication on M is not only a group homomorphism, but it is an R-linear map. □

***Example* 14.14.** Let M be an R-module, and let a be an element of R. The **a-multiplication map on** M is the map that sends $\alpha \in M$ to $a\alpha \in M$. This is an R-linear map whose image is aM, and whose kernel is $M\{a\}$. The set of all $a \in R$ for which $aM = \{0_M\}$ is called the **R-exponent of** M, and is easily seen to be an ideal of R (verify). □

***Example* 14.15.** Let M be an R-module, and let α be an element of M. Then the map $\rho : R \to M$ given by $\rho(a) = a\alpha$ is an R-linear map. The image of this map is $\langle \alpha \rangle_R$. The kernel of this map is called the **R-order of** α, and is easily seen to be an ideal of R (verify). □

***Example* 14.16.** Consider again the R-module $R[\mathtt{X}]/(f)$ discussed in Example 14.3, where f is monic of degree ℓ. As an R-module, $R[\mathtt{X}]/(f)$ is isomorphic to $R[\mathtt{X}]_{<\ell}$ (see Example 14.9). Indeed, based on the observations in Example 9.34, the map $\rho : R[\mathtt{X}]_{<\ell} \to R[\mathtt{X}]/(f)$ that sends a polynomial $g \in R[\mathtt{X}]$ of degree less than ℓ to $[g]_f \in R[\mathtt{X}]/(f)$ is an isomorphism of $R[\mathtt{X}]_{<\ell}$ with $R[\mathtt{X}]/(f)$. Furthermore, $R[\mathtt{X}]_{<\ell}$ is isomorphic as an R-module to $R^{\times \ell}$.

Indeed, the map $\rho' : R[\mathtt{X}]_{<\ell} \to R^{\times \ell}$ that sends $g = \sum_{i=0}^{\ell-1} g_i \mathtt{X}^i \in R[\mathtt{X}]_{<\ell}$ to $(g_0, \ldots, g_{\ell-1}) \in R^{\times \ell}$ is an isomorphism of $R[\mathtt{X}]_{<\ell}$ with $R^{\times \ell}$. □

***Example* 14.17.** Let E and E' be ring extensions of the ring R. As we saw in Example 14.4, E and E' may be viewed as R-modules in a natural way. Suppose that $\rho : E \to E'$ is a ring homomorphism whose restriction to R is the identity map (i.e., $\rho(a) = a$ for all $a \in R$). Then ρ is an R-linear map. Indeed, for any $a \in R$ and $\alpha, \beta \in E$, we have $\rho(\alpha + \beta) = \rho(\alpha) + \rho(\beta)$ and $\rho(a\alpha) = \rho(a)\rho(\alpha) = a\rho(\alpha)$. □

14.4 Linear independence and bases

Throughout this section, R denotes a ring.

Definition 14.14. *We say that an R-module M is **finitely generated (over R)** if it is spanned by a finite number of elements, which is to say that $M = \langle \alpha_1, \ldots, \alpha_n \rangle_R$ for some $\alpha_1, \ldots, \alpha_n \in M$.*

*We say that a collection of elements $\alpha_1, \ldots, \alpha_n$ in M is **linearly dependent (over R)** if there exist $a_1, \ldots, a_n \in R$, not all zero, such that $a_1\alpha_1 + \cdots a_n\alpha_n = 0_M$; otherwise, we say that $\alpha_1, \ldots, \alpha_n$ are **linearly independent (over R)**.*

*We say that a collection $\alpha_1, \ldots, \alpha_n$ of elements in M is a **basis for** M **(over R)** if it is linearly independent and spans M.*

Note that in the above definition, the collection of elements $\alpha_1, \ldots, \alpha_n$ may contain duplicates; the collection may also be empty (i.e., $n = 0$), in which case, by definition, it is a basis for the trivial submodule $\{0_M\}$. Note that the ordering of the elements $\alpha_1, \ldots, \alpha_n$ makes no difference in any aspect of the definition.

***Example* 14.18.** Consider the R-module $R^{\times n}$. Define $\alpha_1, \ldots, \alpha_n \in R^{\times n}$ as follows:

$$\alpha_1 := (1, 0, \ldots, 0),\ \alpha_2 := (0, 1, 0, \ldots, 0), \ldots,\ \alpha_n := (0, \ldots, 0, 1);$$

that is, α_i has a 1 in position i and is zero everywhere else. It is easy to see that $\alpha_1, \ldots, \alpha_n$ form a basis for $R^{\times n}$. Indeed, for any $a_1, \ldots, a_n \in R$, we have $a_1\alpha_1 + \cdots + a_n\alpha_n = (a_1, \ldots, a_n)$, from which it is clear that the α_i span $R^{\times n}$ and are linearly independent. The vectors $\alpha_1, \ldots, \alpha_n$ form what is called the **standard basis** for $R^{\times n}$. □

***Example* 14.19.** Consider the $\mathbb{Z}$-module $\mathbb{Z}^{\times 3}$. In addition to the standard basis

$$(1,0,0),(0,1,0),(0,0,1),$$

the vectors

$$\alpha_1 := (1,1,1),\ \alpha_2 :=\ (0,1,0),\ \alpha_3 := (2,0,1)$$

also form a basis. To see this, first observe that for $a_1, a_2, a_3, b_1, b_2, b_3 \in \mathbb{Z}$, we have

$$(b_1, b_2, b_3) = a_1\alpha_1 + a_2\alpha_2 + a_3\alpha_3$$

if and only if

$$b_1 = a_1 + 2a_3,\ b_2 = a_1 + a_2,\ \text{and}\ b_3 = a_1 + a_3. \tag{14.1}$$

If (14.1) holds with $b_1 = b_2 = b_3 = 0$, then subtracting the equation $a_1+a_3 = 0$ from $a_1 + 2a_3 = 0$, we see that $a_3 = 0$, from which it easily follows that $a_1 = a_2 = 0$. This shows that the vectors are linearly independent. To show that they span $\mathbb{Z}^{\times 3}$, the reader may verify that for any given $b_1, b_2, b_3 \in \mathbb{Z}$, the values

$$a_1 := -b_1 + 2b_3,\ a_2 := b_1 + b_2 - 2b_3,\ a_3 := b_1 - b_3$$

satisfy (14.1).

The vectors

$$(1,1,1),(0,1,0),(1,0,1)$$

do not form a basis, as they are linearly dependent: the third vector is equal to the first minus the second.

The vectors $(1,0,12),(0,1,30),(0,0,18)$ are linearly independent, but do not span $\mathbb{Z}^{\times 3}$ — the last component of any $\mathbb{Z}$-linear combination of these vectors must be divisible by $\gcd(12,30,18) = 6$. These vectors do, however, form a basis for the $\mathbb{Q}$-module $\mathbb{Q}^{\times 3}$. □

***Example* 14.20.** If R is non-trivial, the ring of polynomials $R[\mathtt{X}]$ is not finitely generated as an R-module, since any finite set of polynomials spans only polynomials of some bounded degree. □

***Example* 14.21.** Consider the submodule $R[\mathtt{X}]_{<\ell}$ of $R[\mathtt{X}]$, where $\ell \geq 0$. If $\ell = 0$, then $R[\mathtt{X}]_{<\ell}$ is trivial; otherwise, $1, \mathtt{X}, \ldots, \mathtt{X}^{\ell-1}$ form a basis. □

***Example* 14.22.** Consider again the ring $E = R[\mathtt{X}]/(f)$, where $f \in R[\mathtt{X}]$ is monic of degree $\ell \geq 0$. If $f = 1$, then E is trivial; otherwise, $1, \eta, \eta^2, \ldots, \eta^{\ell-1}$, where $\eta := [\mathtt{X}]_f \in E$, form a basis for E over R. □

The next theorem highlights a critical property of bases:

Theorem 14.15. *If $\alpha_1, \ldots, \alpha_n$ form a basis for M, then the map $\rho : R^{\times n} \to M$ that sends $(a_1, \ldots, a_n) \in R^{\times n}$ to $a_1\alpha_1 + \cdots + a_n\alpha_n \in M$ is an R-module isomorphism of $R^{\times n}$ with M. In particular, every element of M can be expressed in a unique way as $a_1\alpha_1 + \cdots + a_n\alpha_n$, for $a_1, \ldots, a_n \in R$.*

Proof. To show this, one has to show (1) that ρ is an R-linear map, which follows immediately from the definitions, (2) that ρ is injective, which follows immediately from the linear independence of $\alpha_1, \ldots, \alpha_n$, and (3) that ρ is surjective, which follows immediately from the fact that $\alpha_1, \ldots, \alpha_n$ span M. □

The following theorems develop important connections among the notions of spanning, linear independence, and linear maps.

Theorem 14.16. *Suppose that $\alpha_1, \ldots, \alpha_n$ span an R-module M and that $\rho : M \to M'$ is an R-linear map.*

(i) ρ is surjective if and only if $\rho(\alpha_1), \ldots, \rho(\alpha_n)$ span M'.

(ii) If $\rho(\alpha_1), \ldots, \rho(\alpha_n)$ are linearly independent, then ρ is injective.

Proof. Since the α_i span M, every element of M can be expressed as $\sum_i a_i\alpha_i$, where the a_i are in R. It follows that the image of ρ consists of all elements of M' of the form $\rho(\sum_i a_i\alpha_i) = \sum_i a_i\rho(\alpha_i)$. That is, the image of ρ is the submodule of M' spanned by $\rho(\alpha_1), \ldots, \rho(\alpha_n)$, which implies (i).

For (ii), suppose that ρ is not injective. Then $\rho(\alpha) = 0_{M'}$ for some $\alpha \neq 0_M$, and since the α_i span M, we can write $\alpha = \sum_i a_i\alpha_i$, where the a_i are in R. Since α is non-zero, some of the a_i must be non-zero. So we have $0_{M'} = \rho(\sum_i a_i\alpha_i) = \sum_i a_i\rho(\alpha_i)$, and hence $\rho(\alpha_1), \ldots, \rho(\alpha_n)$ are linearly dependent. □

Theorem 14.17. *Suppose $\rho : M \to M'$ is an injective R-linear map and that $\alpha_1, \ldots, \alpha_n \in M$ are linearly independent. Then $\rho(\alpha_1), \ldots, \rho(\alpha_n)$ are linearly independent.*

Proof. Suppose that $0_{M'} = \sum_i a_i\rho(\alpha_i) = \rho(\sum_i a_i\alpha_i)$. Then, as $\ker(\rho) = \{0_M\}$, we must have $\sum_i a_i\alpha_i = 0_M$, and as the α_i are linearly independent, all the a_i must be zero. □

Theorem 14.18. *Let $\alpha_1, \ldots, \alpha_n$ be a basis for an R-module M, and let $\rho : M \to M'$ be an R-linear map.*

(i) ρ is surjective if and only if $\rho(\alpha_1), \ldots, \rho(\alpha_n)$ span M'.

(ii) ρ is injective if and only if $\rho(\alpha_1), \ldots, \rho(\alpha_n)$ are linearly independent.

(iii) ρ is an isomorphism if and only if $\rho(\alpha_1), \ldots, \rho(\alpha_n)$ form a basis for M'.

Proof. (i) follows immediately from part (i) of Theorem 14.16. (ii) follows from part (ii) of Theorem 14.16 and Theorem 14.17. (iii) follows from (i) and (ii). □

Exercise 14.1. Show that if a finite collection of elements of an R-module is linearly independent, then any sub-collection is also linearly independent.

Exercise 14.2. Assume R is non-trivial. Show that if a finite collection of elements of an R-module contains the zero element, or contains two identical elements, then it is not linearly independent.

Exercise 14.3. Assume R is trivial and that M is an R-module (which must also be trivial). Show that any finite collection of zero or more copies of 0_M is a basis for M.

Exercise 14.4. Let $\rho : M \to M'$ be an R-linear map. Show that if $\alpha_1, \ldots, \alpha_n \in M$ are linearly dependent, then $\rho(\alpha_1), \ldots, \rho(\alpha_n) \in M'$ are also linearly dependent.

14.5 Vector spaces and dimension

Throughout this section, F denotes a field.

A module over a field is also called a **vector space**. In particular, an F-module is called an **F-vector space**, or a **vector space over F**.

For vector spaces over F, one typically uses the terms **subspace** and **quotient space**, instead of (respectively) submodule and quotient module; likewise, one usually uses the terms **F-vector space homomorphism**, **isomorphism** and **automorphism**, as appropriate.

Throughout the rest of this section, V denotes a vector space over F.

We now develop the basic theory of dimension for *finitely generated* vector spaces. The following two theorems provide the keys to this theory.

Theorem 14.19. *If V is finitely generated, then any finite set of vectors that spans V contains a subset that is a basis.*

Proof. We give an "algorithmic" proof. Let $\alpha_1, \ldots, \alpha_n$ be a given set of vectors that spans V. Let S_0 be the empty set, and for $i = 1, \ldots, n$, do the following: if α_i does not belong to the subspace spanned by S_{i-1}, set $S_i := S_{i-1} \cup \{\alpha_i\}$, and otherwise, set $S_i := S_{i-1}$. We claim that S_n is a basis for V.

First, we show that S_n spans V. To do this, first note that for $i = 1, \ldots, n$, if α_i is not in S_n, then by definition, α_i is a linear combination of vectors in

$S_{i-1} \subseteq S_n$. In any case, each α_i is a linear combination of the vectors in S_n. Since any element β of V is a linear combination of $\alpha_1, \ldots, \alpha_n$, and each α_i is a linear combination of elements of S_n, it follows (see Example 14.12) that β is a linear combination of elements of S_n.

Second, we show that S_n is linearly independent. Suppose it were not. Then we could express 0_V as a non-trivial linear combination of elements in S_n. Let us write this as

$$0_V = a_1\alpha_1 + a_2\alpha_2 + \cdots + a_n\alpha_n,$$

where the only non-zero coefficients a_i are those with $\alpha_i \in S_n$. If j is the highest index with $a_j \neq 0_F$, then by definition $\alpha_j \in S_n$. However, we see that α_j is in fact in the span of S_{j-1}; indeed,

$$\alpha_j = (-a_j^{-1}a_1)\alpha_1 + \cdots + (-a_j^{-1}a_{j-1})\alpha_{j-1},$$

and by definition, the only terms with non-zero coefficients are those corresponding to the vectors in S_{j-1}. This means that we would not have added α_j to S_j at step j, which means α_j is not in S_n, a contradiction. □

Theorem 14.20. *If V has a basis of size n, then any collection of $n+1$ elements of V is linearly dependent.*

Proof. Let $\alpha_1, \ldots, \alpha_n$ be a basis, and let $\beta_1, \ldots, \beta_{n+1}$ be any collection of $n+1$ vectors. We wish to show that $\beta_1, \ldots, \beta_{n+1}$ are linearly dependent.

Since the α_i span V, we know that β_1 is a linear combination of the α_i, say, $\beta_1 = a_1\alpha_1 + \cdots a_n\alpha_n$. If all the a_i were zero, then we would have $\beta_1 = 0_V$, and so trivially, $\beta_1, \ldots, \beta_{n+1}$ would be linearly dependent (see Exercise 14.2). So assume that not all a_i are zero, and for convenience, let us say that $a_1 \neq 0_F$. It follows that α_1 is a linear combination of $\beta_1, \alpha_2, \ldots, \alpha_n$; indeed,

$$\alpha_1 = a_1^{-1}\beta_1 + (-a_1^{-1}a_2)\alpha_2 + \cdots + (-a_1^{-1}a_n)\alpha_n.$$

It follows that $\beta_1, \alpha_2, \ldots, \alpha_n$ span V (see Example 14.12).

Next, consider β_2. This is a linear combination of $\beta_1, \alpha_2, \ldots, \alpha_n$, and we may assume that in this linear combination, the coefficient of one of $\alpha_2, \ldots, \alpha_n$ is non-zero (otherwise, we find a linear dependence among the β_j), and for convenience, let us say that the coefficient of α_2 is non-zero. As in the previous paragraph, it follows that $\beta_1, \beta_2, \alpha_3, \ldots, \alpha_n$ span V.

Continuing in this way, we find that $\beta_1, \ldots, \beta_n$ are either linearly dependent or they span V. In the latter case, we find that β_{n+1} is a linear combination of $\beta_1, \ldots, \beta_n$, and hence, the vectors $\beta_1, \ldots, \beta_n, \beta_{n+1}$ are linearly dependent. □

We stress that the proofs of Theorems 14.19 and 14.20 both made critical use of the assumption that F is a *field*. An important corollary of Theorem 14.20 is the following:

Theorem 14.21. *If V is finitely generated, then any two bases have the same size.*

Proof. If one basis had more elements than another, then Theorem 14.20 would imply that the first basis was linearly dependent, which contradicts the definition of a basis. □

Theorem 14.21 allows us to make the following definition:

Definition 14.22. *If V is finitely generated, the common size of any basis is called the* ***dimension*** *of V, and is denoted $\dim_F(V)$.*

Note that from the definitions, we have $\dim_F(V) = 0$ if and only if V is the trivial vector space (i.e., $V = \{0_V\}$). We also note that one often refers to a finitely generated vector space as a **finite dimensional** vector space. We shall give preference to this terminology from now on.

To summarize the main results in this section up to this point: if V is finite dimensional, it has a basis, and any two bases have the same size, which is called the dimension of V. The next theorem is simple consequences of these results.

Theorem 14.23. *Suppose that V is of finite dimension n, and let $\alpha_1, \ldots, \alpha_n \in V$. The following are equivalent:*

(i) *$\alpha_1, \ldots, \alpha_n$ are linearly independent.*

(ii) *$\alpha_1, \ldots, \alpha_n$ span V.*

(iii) *$\alpha_1, \ldots, \alpha_n$ form a basis for V.*

Proof. Let W be the subspace spanned by $\alpha_1, \ldots, \alpha_n$.

First, let us show that (i) implies (ii). Suppose $\alpha_1, \ldots, \alpha_n$ are linearly independent. Also, by way of contradiction, suppose that $W \subsetneq V$. Choose $\beta \in V \setminus W$. Then it follows that $\alpha_1, \ldots, \alpha_n, \beta$ are linearly independent; indeed, if we had a relation $0_V = a_1\alpha_1 + \cdots + a_n\alpha_n + b\beta$, then we must have $b = 0_F$ (otherwise, $\beta \in W$), and by the linear independence of $\alpha_1, \ldots, \alpha_n$, all the a_i must be zero as well. But then we have a set of $n + 1$ linearly independent vectors in V, which is impossible by Theorem 14.20.

Second, let us prove that (ii) implies (i). Let us assume that $\alpha_1, \ldots, \alpha_n$ are linearly dependent, and prove that $W \subsetneq V$. By Theorem 14.19, we can find a basis for W among the α_i, and since the α_i are linearly dependent, this basis must contain strictly fewer than n elements. Hence, $\dim_F(W) < \dim_F(V)$, and therefore, $W \subsetneq V$.

The theorem now follows from the above arguments, and the fact that, by definition, (iii) holds if and only if both (i) and (ii) hold. □

We next examine the dimension of subspaces of finite dimensional vector spaces.

Theorem 14.24. *If V is finite dimensional, and W is a subspace of V, then W is also finite dimensional, and $\dim_F(W) \leq \dim_F(V)$. Moreover, $\dim_F(W) = \dim_F(V)$ if and only if $W = V$.*

Proof. To see this, suppose $\dim_F(V) = n$, and assume that W is non-trivial. We shall construct a basis $\alpha_1, \ldots, \alpha_m$ for W, where $m \leq n$. We can take α_1 to be any non-zero vector in W, α_2 to be any vector in W not in the subspace spanned by α_1, and so on. More generally, at stage $i = 1, 2, \ldots$, we take α_i to be any element of W not in the subspace spanned by $\alpha_1, \ldots, \alpha_{i-1}$. It is easy to see that at each stage i, the vectors $\alpha_1, \ldots, \alpha_i$ are linearly independent: if we had a relation $a_1\alpha_1 + \cdots a_j\alpha_j = 0_V$, where $j \leq i$ and $a_j \neq 0_F$, this would imply that α_j lies in the subspace generated by $\alpha_1, \ldots, \alpha_{j-1}$, which contradicts the definition of how α_j was selected. Because of Theorem 14.20, this process must halt at some stage $m \leq n$, and since the process does halt, it must be the case that $\alpha_1, \ldots, \alpha_m$ span W.

That proves that W is finite dimensional with $\dim_F(W) \leq \dim_F(V)$. It remains to show that these dimensions are equal if and only if $W = V$. Now, if $W = V$, then clearly $\dim_F(W) = \dim_F(V)$. Conversely, if $\dim_F(W) = \dim_F(V)$, then by Theorem 14.23, any basis for W must already span V. □

Theorem 14.25. *If V is finite dimensional, and W is a subspace of V, then the quotient space V/W is also finite dimensional, and*

$$\dim_F(V/W) = \dim_F(V) - \dim_F(W).$$

Proof. Suppose that S is a finite set of vectors that spans V. Then $\{\alpha + W : \alpha \in S\}$ is a finite set of vectors that spans V/W. It follows from Theorem 14.19 that V/W has a basis, say, $\alpha_1 + W, \ldots, \alpha_\ell + W$. Suppose that $\beta_1, \ldots, \beta_m$ is a basis for W. The theorem will follow immediately from the following:

Claim. The vectors

$$\alpha_1, \ldots, \alpha_\ell,\ \beta_1, \ldots, \beta_m \tag{14.2}$$

form a basis for V.

To see that these vectors span V, consider any element γ of V. Then since $\alpha_1 + W, \ldots, \alpha_\ell + W$ span V/W, we have $\gamma \equiv \sum_i a_i\alpha_i \pmod{W}$ for some $a_1, \ldots, a_\ell \in F$. If we set $\beta := \gamma - \sum_i a_i\alpha_i \in W$, then since $\beta_1, \ldots, \beta_m$ span W, we have $\beta = \sum_j b_j\beta_j$ for some $b_1, \ldots, b_m \in F$, and hence $\gamma = \sum_i a_i\alpha_i + \sum_j b_j\beta_j$. That proves that the vectors (14.2) span V. To prove they are linearly independent, suppose we have a relation of the form $\sum_i a_i\alpha_i + \sum_j b_j\beta_j = 0_V$, where $a_1, \ldots, a_\ell \in F$ and $b_1, \ldots, b_m \in F$. If any of the a_i were non-zero, this would contradict the assumption that $\alpha_1+W, \ldots, \alpha_\ell+W$ are linearly independent. So assume that all the a_i are zero. If any of the b_j were non-zero, this would contradict the assumption that $\beta_1, \ldots, \beta_m$ are linearly independent. Thus, all the a_i and all the b_j must be zero, which proves that the vectors (14.2) are linearly independent. That proves the claim. □

Theorem 14.26. *If V is of finite dimension, then any linearly independent set of elements of V can be extended to form a basis for V.*

Proof. This is actually implicit in the proof of the previous theorem. Let $\beta_1, \ldots, \beta_m \in V$ be linearly independent. Let W be the subspace of V spanned by $\beta_1, \ldots, \beta_m$, so that $\beta_1, \ldots, \beta_m$ form a basis for W. As in the proof of the previous theorem, we can choose $\alpha_1, \ldots, \alpha_\ell \in V$ such that $\alpha_1 + W, \ldots, \alpha_\ell + W$ form a basis for the quotient space V/W, so that

$$\alpha_1, \ldots, \alpha_\ell,\ \beta_1, \ldots, \beta_m$$

form a basis for V. □

***Example* 14.23.** Suppose that F is finite, say $|F| = q$, and that V is finite dimensional, say $\dim_F(V) = n$. Then clearly $|V| = q^n$. If W is a subspace with $\dim_F(W) = m$, then $|W| = q^m$, and by Theorem 14.25, $\dim_F(V/W) = n-m$, and hence $|V/W| = q^{n-m}$. Just viewing V and W as additive groups, we know that the index of W in V is $[V : W] = |V/W| = |V|/|W| = q^{n-m}$, which agrees with the above calculations. □

We next consider the relation between the notion of dimension and linear maps.

Theorem 14.27. *If V is of finite dimension n, and V is isomorphic to V', then V' is also of finite dimension n.*

Proof. If $\alpha_1, \ldots, \alpha_n$ is a basis for V, then by Theorem 14.18, $\rho(\alpha_1), \ldots, \rho(\alpha_n)$ is a basis for V'. □

Theorem 14.28. *If $\rho : V \to V'$ is an F-linear map, and if V and V' are finite dimensional with* $\dim_F(V) = \dim_F(V')$*, then we have:*

ρ is injective if and only if ρ is surjective.

Proof. Let $\alpha_1, \ldots, \alpha_n$ be a basis for V. By Theorem 14.18, we know that ρ is injective if and only if $\rho(\alpha_1), \ldots, \rho(\alpha_n)$ are linearly independent, and that ρ is surjective if and only if $\rho(\alpha_1), \ldots, \rho(\alpha_n)$ span V'. Moreover, by Theorem 14.23, we know that the vectors $\rho(\alpha_1), \ldots, \rho(\alpha_n)$ are linearly independent if and only if they span V'. The theorem now follows immediately. □

This last theorem turns out to be extremely useful in a number of settings. Generally, of course, if we have a function $f : A \to B$, injectivity does not imply surjectivity, nor does surjectivity imply injectivity. If A and B are finite sets of equal size, then these implications do indeed hold. Theorem 14.28 gives us another important setting where these implications hold, with finite dimensionality playing the role corresponding to finiteness.

Theorem 14.28 may be generalized as follows:

Theorem 14.29. *If V is finite dimensional, and $\rho : V \to V'$ is an F-linear map, then* $\mathrm{img}(\rho)$ *is a finite dimensional vector space, and*

$$\dim_F(V) = \dim_F(\mathrm{img}(\rho)) + \dim_F(\ker(\rho)).$$

Proof. As the reader may verify, this follows immediately from Theorem 14.25, together with Theorems 14.27 and 14.11. □

Intuitively, one way to think of Theorem 14.29 is as a "law of conservation" for dimension: any "dimensionality" going into ρ that is not "lost" to the kernel of ρ must show up in the image of ρ.

Exercise 14.5. Show that if $V_1, \ldots, V_n$ are finite dimensional vector spaces, then $V_1 \times \cdots \times V_n$ has dimension $\sum_{i=1}^n \dim_F(V_i)$.

Exercise 14.6. Show that if V is a finite dimensional vector space with subspaces W_1 and W_2, such that $W_1 + W_2 = V$ and $W_1 \cap W_2 = \{0_V\}$, then $\dim_F(V) = \dim_F(W_1) + \dim_F(W_2)$.

Exercise 14.7. The theory of dimension for finitely generated vector spaces is quite elegant and powerful. There is a theory of dimension (of sorts) for modules over an arbitrary, non-trivial ring R, but it is much more awkward and limited. This exercise develops a proof of one aspect of this theory: if an R-module M has a basis at all, then any two bases have the same size.

To prove this, we need the fact that any non-trivial ring has a maximal ideal (this was proved in Exercise 9.30 for countable rings). Let n, m be positive integers, let $\alpha_1, \ldots, \alpha_m$ be elements of $R^{\times n}$, and let I be an ideal of R.

(a) Show that if $\alpha_1, \ldots, \alpha_m$ span $R^{\times n}$, then every element of $I^{\times n}$ can be expressed as $a_1\alpha_1 + \cdots a_m\alpha_m$, where $a_1, \ldots, a_m$ belong to I.

(b) Show that if $m > n$ and I is a maximal ideal, then there exist $a_1, \ldots, a_m \in R$, not all in I, such that $a_1\alpha_1 + \cdots a_m\alpha_m \in I^{\times n}$.

(c) From (a) and (b), deduce that if $m > n$, then $\alpha_1, \ldots, \alpha_m$ cannot be a basis for $R^{\times n}$.

(d) From (c), conclude that any two bases for a given R-module M must have the same size.