Modules

February 20, 2024

Table of Contents

1. Definitions and basic properties

Throughout this section, R is a ring.

1.1. Left and right modules, vector spaces, and algebras

1.1.1. *A left R-module*, or *a left module over R*, is an abelian group M, written additively, with an operation of "left multiplication by elements of R": a mapping $R \times M \longrightarrow M$, $(a, u) \mapsto au$, satisfying the properties $a(u+v) = au + av$, $(a+b)u = au + bu$, and $(ab)u = a(bu)$ for all $a, b \in R$ and $u, v \in M$. If R is unital (that is, contains the multiplicative identity 1), it is usually additionally required that $1u = u$ for all $u \in M$. The elements of R are often called *scalars*.

Notice that an R-module structure defines a left action of the multiplicative semigroup of R on the group M by homomorphisms of M. (That is, for any $a \in G$ the mapping $u \mapsto au$ is a homomorphism $M \longrightarrow M$, and the homomorphism corresponding to the product ab of elements $a, b \in R$ is the composition of the homomorphisms corresponding to these elements.)

1.1.2. *A right R-module* is defined similarly, with a right action of (the multiplicative semigroup of) R on M. In the case R is a commutative ring, the notions of a left and a right R-modules coincide, and a left=right R-module is simply called *an R-module*. Also, under an R-module we often understand a left (or a right) R-module when it is either clear from the context or does not matter which, left or right, action of R on M is meant.

1.1.3. *An R-bimodule*, or *a two-sided R-module*, is an abelian group M that has both structures, – of a left R-module and of a right R-module, – with the property $a(ub) = (au)b$ for all $a, b \in R$ and $u \in M$.

1.1.4. If F is a field, an F-module is called *an F-vector space*.

1.1.5. *An R-algebra A* is a ring whose additive group has a structure of an R-module satisfying the property $a(\alpha\beta) = (a\alpha)\beta = \alpha(a\beta)$ for all $a \in R$ and $\alpha, \beta \in A$.

1.2. Examples and constructions of modules

1.2.1. *The zero R-module* is the module $M = \{0\}$.

1.2.2. R itself is a left R-module (and a right R-module; and, actually, an R-bimodule).

1.2.3. Any left ideal in R is a left R-module (a right ideal is a right R-module, and a two-sided ideal is a bimodule).

1.2.4. Any abelian group G (written additively) is a $\mathbb{Z}$-module, by putting $nu = \underbrace{u + \cdots + u}_{n}$ for $n > 0$, $nu = -(-n)u$ for $n < 0$, and $0u = 0$, $u \in G$.

1.2.5. For any $n \in \mathbb{N}$, the direct product $R^n = \underbrace{R \times \cdots \times R}_{n}$, with the multiplication by elements of R defined by $a(b_1, \ldots, b_n) = (ab_1, \ldots, ab_n)$, $a \in R$, $(b_1, \ldots, b_n) \in R^n$, is called *a free R-module of rank n*. R^n also has a natural structure of a right R-module, and is, actually, an R-bimodule.

1.2.6. Let X be a set. The set of functions (mappings) $X \longrightarrow R$ has a structure of a left R-module (and of an R-bimodule), by putting $(af)(x) = af(x)$, $a \in R$, $x \in X$. This module is sometimes denoted by R^X. (The free R-module of rank n is a special case of this, corresponding to $X = \{1, \ldots, n\}$.) In the case R is a commutative ring, R^X is an R-algebra.

More generally, if X is a set and M is an R-module, then the set M^X of functions $X \longrightarrow M$ is an R-module.

1.2.7. The ring $R[x]$ of polynomials with coefficients from R is an R-module, and is an R-algebra if R is commutative. Same applies to the ring $R[x_1, \ldots, x_n]$ of polynomials in n variables over R.

1.2.8. Let G be a group and R be a commutative ring. *The group algebra* of G with coefficients from R is the R-module $RG = \left\{a_1 g_1 + \cdots + a_k g_k,\ a_i \in R,\ g_i \in G\right\}$ of formal linear combinations of elements of G with coefficients from R; the multiplication in RG is defined by $(ag)(bh) = (ab)(gh)$.

1.2.9. The group $\mathrm{Mat}_{m,n}(R)$ of $m \times n$ matrices with entries from R is an R-module (which is, actually, a free R-module of rank mn). If R is commutative, the ring $\mathrm{Mat}_{n,n}(R)$ of *square* $n \times n$ matrices with entries from R is an R-algebra.

1.2.10. Let F be a field, V be an n-dimensional F-vector space, and R be the ring of $n \times n$ matrices over F. Then V is a left R-module, with the standard multiplication Au of matrices and vectors.

1.2.11. The following example will be especially important to us in this course. Let F be a field, V be an F-vector space, and T be a linear transformation of V. Consider the action of (the multiplicative semigroup of) the polynomial ring $F[x]$ on V defined by

$$(a_n x^n + \cdots + a_1 x + a_0)u = a_n T^n(u) + \cdots + a_1 T(u) + a_0 u,$$

where $a_n x^n + \cdots + a_1 x + a_0 \in F[x]$ and $u \in V$. This action converts V into an $F[x]$-module.

1.2.12. The preceding example is easily generalizable: Let G be a group, R be a commutative ring, and RG be the group algebra of G with coefficients from R. Then any left (or right) action of G on an abelian group M by homomorphisms defines on M a structure of a left (respectively, right) RG-module.

1.2.13. If A is a unital ring and R is a unital subring of A that lies in the center of A and with $1_R = 1_A$, then A has a structure of an R-algebra. More generally, if $\varphi\colon R \longrightarrow A$ is a homomorphism of unital rings with $\varphi(R) \subseteq Z(A)$ and $\varphi(1_R) = 1_A$, then A has a structure of an R-algebra defined by $au = \varphi(a)u$, $a \in R$, $u \in A$. (It is easy to see that any unital R-algebra A can be constructed this way, by defining $\varphi(a) = a1_A$, $a \in R$.)

1.2.14. Every ring is a $\mathbb{Z}$-algebra.

1.2.15. If M is an R-module and S is a subring of R, then M has a structure of an S-module as well. The operation of converting an R-module into an S-module is called *reduction of scalars*.

1.3. Elementary properties of modules

The following properties of modules are easily verifiable:

Proposition. *let R be a ring and M be a left R-module. Then*

(i) *for any $u \in M$, $0u = 0$;*

(ii) *for any $a \in R$, $a0 = 0$;*

(iii) *for any $a \in R$ and $u \in M$, $(-a)u = a(-u) = -au$.*

1.4. Submodules

1.4.1. Let M be a left R-module. *A submodule of M* is a subgroup N of M which is a left R-module under the multiplication by scalars (elements of R) defined in M. For a subset N of M to be a submodule of M it is necessary and sufficient that N is a subgroup of M and is closed with respect to multiplication by scalars: $N - N \subseteq N$ and $RN \subseteq N$.

1.4.2. Examples. (i) M itself and the zero submodule $0 = \{0\}$ are submodules of M.

(ii) For R viewed as a left R-module, left submodules of R are the left ideals of R.

(iii) Let I be a left ideal in R; then the set $IM = \left\{\sum_{i=1}^{k} a_i u_i,\ k \in \mathbb{N},\ a_i \in I,\ u_i \in M\right\}$ is a submodule of M.

(iv) Let V be a vector space over a field F and let T be a linear transformation of V. Then V is a $F[x]$-module, with $xu = T(u)$, $u \in V$. A subset W of V is a submodule of this module iff W is a subgroup of V, is invariant under multiplication by scalars from F: $aW \subseteq W$ for all $a \in F$, and is invariant under multiplication by x: $xW = T(W) \subseteq W$; that is, iff W is a vector subspace of V invariant under T: $T(W) \subseteq W$.

1.4.3. Proposition. *The intersection of any collection of submodules of a module M is also a submodule of M.*

1.4.4. *The sum* of two submodules N_1 and N_2 of a module M is the set $N_1 + N_2 = \left\{u_1 + u_2 : u_1 \in N_1,\ u_2 \in N_2\right\}$. More generally, the sum of a collection $\{N_\alpha\}_{\alpha\in\Lambda}$ of submodules of a module M is the set of elements of M representable as a sum of elements of the members of this collection:

$$\sum_{\alpha\in\Lambda} N_\alpha = \left\{\sum_{i=1}^{k} u_{\alpha_i} : k \in \mathbb{N},\ \alpha_i \in \Lambda,\ u_i \in N_{\alpha_i},\ i = 1, \dots, k\right\}$$

1.4.5. Proposition. *The sum of any collection of submodules of a module M is also a submodule of M.*

1.5. Generating sets of modules

1.5.1. Let M be a left module and S be a subset of M. The minimal submodule of M containing S (namely, the intersection of all submodules of M containing S) is called *the submodule generated by S*. If R be unital, the submodule generated by S is the set of all finite sums of the form $\sum_{i=1}^{k} a_i s_i$ with $a_1, \dots, a_k \in R$ and $s_1, \dots, s_k \in S$; this set is denoted by RS. (If R is not unital, it is $RS + \mathbb{Z}S$.)

Let's introduce the notation $\sum_{\alpha\in\Lambda}^{\text{fin}} w_\alpha$ for $\sum_{\alpha\in\Lambda} w_\alpha$ in which $w_\alpha = 0$ for all but finitely many αs. Then (in the case $1 \in R$) we can write $RS = \left\{\sum_{s\in S}^{\text{fin}} a_s s : a_s \in R,\ s \in S\right\}$.

1.5.2. If $M = RS$ for some $S \subseteq M$, we say that M *is generated by* S, or that S *generates* M, or that S is *a generating set of* M. (In the case R is a field and so, M is an R-vector space, we say that S *spans* M.) If S is finite, we say that M is *finitely generated*.

1.5.3. The submodule of M generated by a collection $\{N_\alpha\}_{\alpha\in\Lambda}$ of submodules of M is, clearly, the sum $\sum_{\alpha\in\Lambda} N_\alpha$ of these submodules.

1.5.4. A (left) R-module M is said to be *cyclic* if it is generated by a single element, $M = Ru$ for some $u \in M$.

1.6. Quotient modules

1.6.1. Let M be a left R-module and let N be a submodule of M. The factor-group M/N is the group $\{\bar{u},\ u \in M\}$, where for $u \in M$, $\bar{u} = u + N$, the class of elements equivalent to u modulo N. M/N has a structure of a left R-module defined by $a\bar{u} = \overline{au}$; this module is denoted by M/N and is called *the quotient*, or *the factor* module of M by N.

1.6.2. Example. Let R be a non-commutative ring and I be a left ideal in R. If I is not a two-sided ideal, then R/I is not a ring, but is a left R-module.

1.7. Torsion elements of a module and the torsion submodule

1.7.1. An element u of a (left) R-module M is said to be *a torsion element* if $au = 0$ for some nonzero $a \in R$. If all elements of M are torsion, then M is called *a torsion module*; if no elements of M, except 0, are torsion, then M is said to be *torsion-free*.

1.7.2. If R is an integral domain and M is an R-module, then the torsion elements of M form a submodule of M; this submodule is called *the torsion submodule* of M and is denoted by $\operatorname{Tor}(M)$.

Proposition. *If R is an integral domain and M is an R-module, the quotient module $M/\operatorname{Tor}(M)$ is torsion-free.*

1.8. Annihilators

1.8.1. Let M be a left R-module and let P be a subset of M. *The annihilator of* P is the set $\mathrm{Ann}(P) = \{a \in R : aP = 0\}$; this is a left ideal in R. If N is a submodule of M, then $\mathrm{Ann}(N)$ is a two-sided ideal in R, and N has a structure of an $(R/\mathrm{Ann}(N))$-module, defined by $(a + \mathrm{Ann}(N))u = au$, $a \in R$, $u \in N$.

1.8.2. Proposition. *Let M be a module and N_1, N_2 be submodules of M. Then* $\mathrm{Ann}(N_1+N_2) = \mathrm{Ann}(N_1)\cap \mathrm{Ann}(N_2)$*, and* $\mathrm{Ann}(N_1 \cap N_2) \supseteq \mathrm{Ann}(N_1) + \mathrm{Ann}(N_2)$.

1.8.3. Let M be a left R-module and S be a subset of R; *the annihilator of S in M* is the set $\mathrm{Ann}(S) = \{u \in M : Su = 0\}$. If I is the left ideal of R generated by S, $I = (S) = RS$, then $\mathrm{Ann}(I) = \mathrm{Ann}(S)$. If I is a right ideal in R, then $\mathrm{Ann}(I)$ is a submodule of M.

1.8.4. Proposition. *Let M be an R-module and let I_1, I_2 be right ideals in R. Then* $\mathrm{Ann}(I_1 + I_2) = \mathrm{Ann}(I_1) \cap \mathrm{Ann}(I_2)$*, and* $\mathrm{Ann}(I_1 \cap I_2) \supseteq \mathrm{Ann}(I_1) + \mathrm{Ann}(I_2)$.

1.9. Homomorphisms of modules

1.9.1. Let M and N be left R-modules. A mapping $\varphi\colon M \longrightarrow N$ is called *an R-module homomorphism*, or just *a homomorphism*, if it satisfies $\varphi(u+v) = \varphi(u)+\varphi(v)$ for all $u, v \in M$ (that is, is a group homomorphism from M to N) and $\varphi(au) = a\varphi(u)$ for all $a \in R$ and $u \in M$.

1.9.2. Examples. (0) The zero mapping $\varphi(u) = 0$ for all $u \in M$ is *the zero homomorphism* from module M (to any other module).

(i) $\mathbb{Z}$-module homomorphisms of abelian groups, viewed as $\mathbb{Z}$-modules, is the same as group homomorphisms, since for any group homomorphism we automatically have $\varphi(na) = n\varphi(a)$ for any element a and $n \in \mathbb{Z}$.

(ii) If M is a module and N is a submodule of M, then we have *the embedding homomorphism* $\pi\colon N \longrightarrow M$ defined by $u \mapsto u$, $u \in N$.

(iii) If M is a module and N is a submodule of M, then we have *the factorization*, or *the projection* homomorphism $\pi\colon M \longrightarrow M/N$ defined by $u \mapsto \bar{u}$, $u \in M$.

(iv) If R is a commutative ring and M is an R-module, then for any $a \in R$ multiplication by a (the mapping $u \mapsto au$) is a homomorphism $M \longrightarrow M$. (This is not so, generally speaking, if R is not commutative.)

(v) If R is a unital ring, then for any R-module M and any element $u \in M$ there exists a unique homomorphism $\varphi\colon R \longrightarrow M$ that maps 1 to u, namely, φ is defined by $\varphi(a) = au$ for all $a \in R$.

(vi) Let R be a ring, M be an R-module, X be a set, and M^X be the set of functions $X \longrightarrow M$. Let $x_0 \in X$; then the mapping $M^X \longrightarrow M$ defined by $f \mapsto f(x_0)$ is a module homomorphism, called *the evaluation homomorphism*.

(vii) Homomorphisms of vector spaces are called *linear mappings*, or *linear transformations*.

(viii) Let V and W be two vector spaces over a field F, let T be a linear transformation of V and S be a linear transformation of W. Then V and W have a structure of $F[x]$-modules, by putting $xu = T(u)$ for $u \in V$ and $xv = S(v)$ for $v \in W$. An $F[x]$-module homomorphism between these two $F[x]$-modules is an F-linear mapping $\varphi\colon V \longrightarrow W$ satisfying $\varphi\circ T = S\circ\varphi$.

1.9.3. Any ring R is simultaneously an R-module, but module homomorphisms $R \longrightarrow R$ are not the same as ring homomorphisms. For example the mapping $\mathbb{Z} \longrightarrow \mathbb{Z}$ defined by $n \mapsto 2n$ is a $\mathbb{Z}$-module homomorphism but not a ring homomorphism; the mapping $\mathbb{Z}[x] \longrightarrow \mathbb{Z}[x]$ defined by $p(x) \mapsto p(x^2)$ is a ring homomorphism but not a $\mathbb{Z}[x]$-module homomorphism.

1.9.4. *An algebra homomorphism* from an R-algebra A to an R-algebra B is a mapping $\varphi\colon A \longrightarrow B$ which is an R-module homomorphism and a ring homomorphism: $\varphi(u + v) = \varphi(u) + \varphi(v)$, $\varphi(uv) = \varphi(u)\varphi(v)$, and $\varphi(au) = a\varphi(u)$ for any $u, v \in A$ and $a \in R$.

1.9.5. Proposition. *If S is a generating set of a module M, then any homomorphism φ from M is uniquely defined by its restriction $\varphi|_S$ on S.*

1.9.6. Proposition. *The composition $\psi\circ\varphi\colon M \longrightarrow K$ of two homomorphisms $\varphi\colon M \longrightarrow N$ and $\psi\colon N \longrightarrow K$ of modules is a homomorphism. If a homomorphism $\varphi\colon M \longrightarrow N$ of modules is an invertible mapping, then its inverse $\varphi^{-1}\colon N \longrightarrow M$ is also a homomorphism.*

1.9.7. Proposition. *Let $\varphi\colon M \longrightarrow N$ be a homomorphism of R-modules. Then for any submodule L of M, its image $\varphi(L)$ is a submodule of N, and for any submodule K of N, its preimage $\varphi^{-1}(K)$ is a submodule of M. In particular, the image $\varphi(M)$ of φ is a submodule of N, and the preimage $\varphi^{-1}(0)$ of 0 is a submodule of M.*

1.9.8. *The kernel* $\ker(\varphi)$ of a homomorphism $\varphi\colon M \longrightarrow N$ of R-modules is the submodule $\varphi^{-1}(0)$ of M. *The cokernel* $\operatorname{coker}(\varphi)$ of φ is the factor module $N/\varphi(M)$ of N.

1.9.9. A surjective homomorphism of modules is called *an epimorphism*, an injective homomorphism of modules is called *a monomorphism*, a bijective (i.e., invertible) homomorphism of modules is called *an isomorphism*. A self-homomorphism of a module (that is, a homomorphism of a module to itself) is called *an endomorphism*, and a self-isomorphism of a module is called *an automorphism*.

1.9.10. Proposition. *A homomorphism $\varphi\colon M \longrightarrow N$ of modules is a monomorphism iff $\ker(\varphi) = 0$, is an epimorphism iff $\operatorname{coker}(\varphi) = 0$, and is an isomorphism iff both $\ker(\varphi) = 0$ and $\operatorname{coker}(\varphi) = 0$.*

1.9.11. Two R-modules M and N are said to be *isomorphic* if there is an isomorphism $M \longrightarrow N$; this is denoted by $M \cong N$. (Isomorphic modules are often identified, and considered as "the same" module. For example, the modules $R[x]$ and $R[y]$ can both be called "the module of polynomials in one variable", whereas these are, of course, two distinct submodules of the module $R[x, y]$.)

1.10. Isomorphism theorems for modules

The isomorphism theorems for modules are the same as for the (abelian) groups (since modules are abelian groups!). To prove them, it suffices to check that the group isomorphisms appearing in these theorems are, actually, module homomorphisms as well. In fact, it is enough to check this for the first isomorphism theorem only, since the other isomorphism theorems are its corollaries.

1.10.1. The 1st Isomorphism Theorem. *Let $\varphi\colon M \longrightarrow N$ be a homomorphism of R-modules. Then $\operatorname{range}(\varphi) = \varphi(M) \cong M/\ker(\varphi)$. In more details, the mapping $M/\ker(\varphi) \longrightarrow N$ defined by $u + N \mapsto \varphi(u)$, $u \in M$, is an isomorphism between $M/\ker(\varphi)$ and $\varphi(M)$.*

1.10.2. The 2nd Isomorphism Theorem. *Let M be an R-module and N, K be submodules of M. Then $(N+K)/K \cong N/(N\cap K)$; namely, the mapping $u+(N\cap K) \mapsto u+K$ is an isomorphism between $N/(N\cap K)$ and $(N + K)/K$.*

1.10.3. The 3rd Isomorphism Theorem. *Let M be an R-module, N be a submodule of M, and K be a submodule of N. Then $M/N \cong (M/K)/(N/K)$; namely, the mapping $u + N \mapsto (u + K) + N/K$ is an isomorphism between M/N and $(M/K)/(N/K)$.*

1.11. Finitely generated modules as factors of R^n

Let R be a unital ring.

1.11.1. Let M be a cyclic module, $M = Ru$ for some $u \in M$. The mapping $\varphi\colon R \longrightarrow M$ defined by $\varphi(a) = au$ is then an epimorphisms of R-modules, so M is isomorphic to the quotient module R/I where $I = \ker(\varphi) = \operatorname{Ann}(u)$ is a left ideal of R.

1.11.2. If R is unital and a R-module M is generated by a finite set $\{u_1, \dots, u_n\}$, then the homomorphism $\varphi\colon R^n \longrightarrow M$ defined by $\varphi(a_1, \dots, a_n) = \sum_{i=1}^{n} a_i u_i$ is an epimorphisms, so M is isomorphic to a quotient module of R^n.

1.12. The module $\operatorname{Hom}(M, N)$ and the algebra $\operatorname{End}(M)$

Let R be a commutative ring (this is important here!).

1.12.1. For two R-modules M and N, the set of homomorphisms $M \longrightarrow N$ is denoted by $\operatorname{Hom}_R(M, N)$, or just $\operatorname{Hom}(M, N)$. For $\varphi, \psi \in \operatorname{Hom}(M, N)$ and $a \in R$, the homomorphisms $\varphi + \psi, a\varphi \in \operatorname{Hom}(M, N)$ are defined by $(\varphi + \psi)(u) = \varphi(u) + \psi(u)$, $(a\varphi)(u) = a\varphi(u)$, $u \in M$; these two operations induce on $\operatorname{Hom}(M, N)$ a structure of an R-module.

1.12.2. If R is a unital ring, then for any R-module M, $\operatorname{Hom}(R, M) \cong M$, were the isomorphism is defined by $\varphi \mapsto \varphi(1)$.

1.12.3. For an R-module M, the set $\mathrm{Hom}(M,M)$ of endomorphism of M is denoted by $\mathrm{End}_R(M)$, or just $\mathrm{End}(M)$. With the operation of composition playing the role of multiplication, $\psi\varphi = \psi\circ\varphi$, $\mathrm{End}(M)$ is a ring and an R-algebra (usually, noncommutative). The set $\mathrm{Aut}(M)$ of automorphisms of M is the set of units (invertible elements) of the ring $\mathrm{End}(M)$, and is a group under multiplication. M has a structure of a left $\mathrm{End}(M)$-module, by defining $\varphi u = \varphi(u)$, $u \in M$, $\varphi \in \mathrm{End}(M)$.

1.13. Schur's lemma

1.13.1. A module with no nontrivial (that is, not counting itself and 0) submodules is said to be *simple* or *irreducible*. A module is simple iff it is generated by every its nonzero element. A left module is simple iff it is isomorphic to R/I, where I is a maximal proper left ideal in R. (In the case R is commutative, this means that the only simple modules are fields.)

1.13.2. Shur's lemma. *If M and N are simple modules, then any homomorphism $M \longrightarrow N$ is either zero or an isomorphism.*

Proof. Let $\varphi\colon M \longrightarrow N$ be a homomorphism. Since M is simple, $\ker(\varphi) = 0$ or $\ker(\varphi) = M$. Since N is simple, $\varphi(M) = N$ or $\varphi(M) = 0$. If $\ker(\varphi) = M$ or $\varphi(M) = 0$, then $\varphi = 0$; if $\ker(\varphi) = 0$ and $\varphi(M) = N$, then φ is an isomorphism. ■

It follows that for a simple module M, $\mathrm{End}(M)$ is a division ring.

1.14. Commutative diagrams and exact sequences of modules

The following terminology turns out to be pretty handy.

1.14.1. A diagram of module homomorphisms (and actually, of any mappings) is said to be *commutative* if for any two modules in this diagram, the composition of the homomorphisms along any path connecting these two modules is independent of the path.

Example. The diagram of module homomorphisms

$$\begin{array}{ccccc} A & \xrightarrow{\varphi} & B & \xrightarrow{\psi} & C \\ {\scriptstyle\alpha}\downarrow & & {\scriptstyle\beta}\downarrow & & {\scriptstyle\gamma}\downarrow \\ A' & \xrightarrow{\varphi'} & B' & \xrightarrow{\psi'} & C' \end{array}$$

is commutative iff $\varphi'\circ\alpha = \beta\circ\varphi$, $\psi'\circ\beta = \gamma\circ\psi$, and, as a corollary, $\gamma\circ\psi\circ\varphi = \psi'\circ\varphi'\circ\alpha$.

1.14.2. A sequence $\ldots \longrightarrow A \xrightarrow{\varphi} B \xrightarrow{\psi} C \longrightarrow \ldots$ of module homomorphisms is said to be *exact at term* B if $\ker(\psi) = \varphi(A)$, and just *exact* if it is exact at all its terms. In particular, the sequence $0 \longrightarrow A \xrightarrow{\varphi} B$ is exact iff φ is a monomorphism, the sequence $A \xrightarrow{\varphi} B \longrightarrow 0$ is exact iff ψ is an epimorphism, and the sequence $0 \longrightarrow A \xrightarrow{\varphi} B \longrightarrow 0$ is exact iff φ is an isomorphism.

1.14.3. An exact sequence of the form $0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0$ is called *a short exact sequence*. This sequence expresses the idea that A is (isomorphic to) a submodule of B, and $B/A \cong C$.

1.14.4. For any module homomorphism $\varphi\colon A \longrightarrow B$, the sequence $0 \longrightarrow \ker(\varphi) \longrightarrow A \xrightarrow{\varphi} B \longrightarrow \mathrm{coker}(\varphi) \longrightarrow 0$ (where $\ker(\varphi) \longrightarrow A$ is the natural embedding and $B \longrightarrow \mathrm{coker}(\varphi)$ is the natural projection) is exact.

1.14.5. Any exact sequence can be "decomposed" into a sequence of short exact sequences: a sequence

$$\cdots \longrightarrow A_{i-1} \xrightarrow{\varphi_{i-1}} A_i \xrightarrow{\varphi_i} A_{i+1} \longrightarrow \cdots$$

of module homomorphisms is exact iff there are exact short sequences $0 \longrightarrow B_i \longrightarrow A_i \xrightarrow{\varphi_i} B_{i+1} \longrightarrow 0$, for some submodules B_i of A_i (in which case $B_i = \ker(\varphi_i) = \varphi_{i-1}(A_{i-1})$ for all i).

1.14.6. As an example of an application of commutative diagrams and exact sequences, here is the so-called

The short five lemma. *Suppose that the diagram of module homomorphisms*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \xrightarrow{\varphi} & B & \xrightarrow{\psi} & C & \longrightarrow & 0 \\ & & {\scriptstyle\alpha}\downarrow & & {\scriptstyle\beta}\downarrow & & {\scriptstyle\gamma}\downarrow & & \\ 0 & \longrightarrow & A' & \xrightarrow{\varphi'} & B' & \xrightarrow{\psi'} & C' & \longrightarrow & 0 \end{array}$$

is commutative with exact rows. Then

(i) *if α and γ are monomorphisms, then β is also a monomorphism;*
(ii) *if α and γ are epimorphisms, then β is also an epimorphism;*
(iii) *if α and γ are isomorphisms, then β is also an isomorphism.*

Proof. This sort of proving is called "diagram wandering":
(i) Let α and γ be injective. Let $b \in B$ be such that $\beta(b) = 0$; we need to show that $b = 0$. We have $\gamma(\psi(b)) = \psi'(\beta(b)) = 0$. Since γ is injective, $\psi(b) = 0$. Since the first raw is exact, $b = \varphi(a)$ for some $a \in A$. Now, $\varphi'(\alpha(a)) = \beta(\varphi(a)) = \beta(b) = 0$; since φ' and α are injective, $a = 0$, so $b = \varphi(a) = 0$.
(ii) Let α and γ be surjective. Let $b' \in B'$; we need to show that $b' = \beta(b)$ for some $b \in B$. Let $c' = \psi'(b')$. Since γ is surjective, $c' = \gamma(c)$ for some $c \in C$. Since ψ is surjective, there is $d \in B$ such that $\psi(d) = c$. Now $\psi'(\beta(d)) = \gamma(\psi(d)) = c' = \psi'(b')$, so $\psi'(b' - \beta(d)) = 0$. Since the second row is exact, there is $a' \in A'$ such that $\varphi'(a') = b' - \beta(d)$. Since α is surjective, $a' = \alpha(a)$ for some $a \in A$. Let $b = \varphi(a) + d$. Then

$$\beta(b) = \beta(\varphi(a)) + \beta(d) = \varphi'(\alpha(a)) + \beta(d) = \varphi'(a') + \beta(d) = b' - \beta(d) + \beta(d) = b'.$$

■

1.14.7. The following lemma (from which The short five lemma 1.14.6 follows) is important in *homological algebra*:

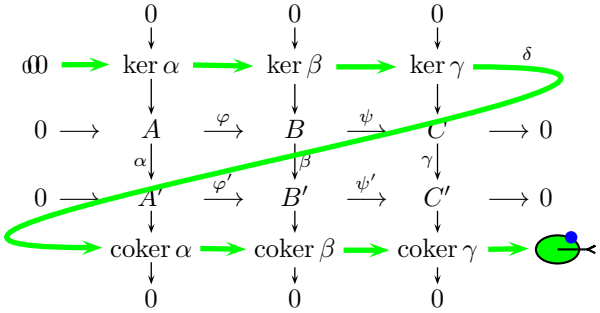
The snake lemma. *Suppose that the diagram of module homomorphisms*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \xrightarrow{\varphi} & B & \xrightarrow{\psi} & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle\alpha} & & \downarrow{\scriptstyle\beta} & & \downarrow{\scriptstyle\gamma} & & \\ 0 & \longrightarrow & A' & \xrightarrow{\varphi'} & B' & \xrightarrow{\psi'} & C' & \longrightarrow & 0 \end{array}$$

is commutative with exact rows. Then there is a homomorphism $\delta\colon \ker(\gamma) \longrightarrow \operatorname{coker}(\alpha)$ such that the sequence

$$0 \longrightarrow \ker(\alpha) \xrightarrow{\varphi} \ker(\beta) \xrightarrow{\psi} \ker(\gamma) \xrightarrow{\delta} \operatorname{coker}(\alpha) \xrightarrow{\bar\varphi'} \operatorname{coker}(\beta) \xrightarrow{\bar\psi'} \operatorname{coker}(\gamma) \longrightarrow 0$$

(where $\bar\varphi'$ and $\bar\psi'$ are the natural quotients of the homomorphisms φ' and ψ' to the quotient modules $\operatorname{coker}(\alpha)$ and $\operatorname{coker}(\beta)$ respectively) is exact, and the diagram



(with exact rows, exact columns, and the exact "snake") is commutative.

Sketch of the proof. δ is defined in the following way: Let $c \in \ker(\gamma)$. Let $b \in B$ be such that $\psi(b) = c$. Let $b' = \beta(b)$. Then $\psi'(b') = \gamma(\psi(b)) = \gamma(c) = 0$, so $b' = a'$ for some $a' \in A'$. Put $\delta(c) = a' \bmod \alpha(A) \in \operatorname{coker}(\alpha)$. It is now to show that δ is well defined (doesn't depend on the choise of b), that δ is a homomorphism, that "the snake" is exact, and that the obtained diagram is commutative; it's a lot of work.

---

2. The direct product and the direct sum of modules as universal objects

2.1. The direct sum and product of two modules

2.1.1. *The direct sum*, or *the direct product* of two R-modules M_1 and M_2 is the R-module $M_1 \oplus M_2 = M_1 \times M_2 = \{(u_1, u_2) : u_1 \in M_1,\ u_2 \in M_2\}$ with the componentwise addition and multiplication by scalars: $(u_1, u_2) + (v_1, v_2) = (u_1 + v_1, u_2 + v_2)$ and $a(u_1, u_2) = (au_1, au_2)$, for $u_1, v_1 \in M_1$, $u_2, v_2 \in N$, and $a \in R$.

2.1.2. Clearly, for any two modules M_1 and M_2, $M_1 \oplus M_2 \cong M_2 \oplus M_1$, under the isomorphism $(u_1, u_2) \mapsto (u_2, u_1)$. Also, for any three modules M_1, M_2 and M_3, $(M_1 \oplus M_2) \oplus M_3 \cong M_1 \oplus (M_2 \oplus M_3)$.

2.1.3. The submodule $M_1 \times \{0\}$ of the module $M_1 \oplus M_2$ is isomorphic to M_1, and is usually identified with M_1, so that M_1 can be assumed to be a submodule of $M_1 \oplus M_2$. Similarly, M_2 can be assumed to be a submodule of $M_1 \oplus M_2$ after identifying it with the submodule $\{0\} \times M_2$. Let $\eta_1 \colon M_1 \longrightarrow M_1 \oplus M_2$ and $\eta_2 \colon M_2 \longrightarrow M_1 \oplus M_2$ be the corresponding embeddings. Notice also that $M_1 \oplus M_2$ is generated by M_1 and M_2 (so that $M_1 \oplus M_2 = M_1 + M_2$), and $M_1 \cap M_2 = 0$.

2.1.4. Also, there are *natural projections* $\pi_1 \colon M_1 \oplus M_2 \longrightarrow M_1$ and $\pi_2 \colon M_1 \oplus M_2 \longrightarrow M_2$ defined by $\pi_1(u_1, u_2) = u_1$ and $\pi_2(u_1, u_2) = u_2$. For these projections one has $\ker(\pi_1) = M_2$ and $\ker(\pi_2) = M_1$, so that the sequences

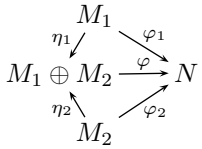
$$0 \longrightarrow M_1 \xrightarrow{\eta_1} M_1 \oplus M_2 \xrightarrow{\pi_2} M_2 \longrightarrow 0 \quad \text{and} \quad 0 \longrightarrow M_2 \xrightarrow{\eta_2} M_1 \oplus M_2 \xrightarrow{\pi_1} M_1 \longrightarrow 0$$

are exact, and we have the isomorphisms $(M_1 \oplus M_2)/M_1 \cong M_2$ and $(M_1 \oplus M_2)/M_2 \cong M_1$.

2.2. The universal properties of the direct sum and product

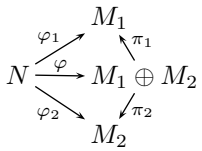
2.2.1. In the notation of the preceding section, the direct sum $M_1 \oplus M_2$ has the following "universal properties":

Theorem. (i) *Given a module N and homomorphisms $\varphi_1 \colon M_1 \longrightarrow N$ and $\varphi_2 \colon M_2 \longrightarrow N$, there exists a unique homomorphism $\varphi \colon M_1 \oplus M_2 \longrightarrow N$ that makes the diagram*



commutative (that is, such that $\varphi_1 = \varphi \circ \eta_1$ and $\varphi_2 = \varphi \circ \eta_2$). φ is defined by $\varphi(u_1, u_2) = \varphi_1(u_1) + \varphi_2(u_2)$, $u_1 \in M_1$, $u_2 \in M_2$.

(ii) *Given a module N and homomorphisms $\varphi_1 \colon N \longrightarrow M_1$ and $\varphi_2 \colon N \longrightarrow M_2$, there exists a unique homomorphism $\varphi \colon N \longrightarrow M_1 \oplus M_2$ that makes the diagram*



commutative (that is, such that $\varphi_1 = \pi_1 \circ \varphi$ and $\varphi_2 = \pi_2 \circ \varphi$). φ is defined by $\varphi(v) = (\varphi_1(v), \varphi_2(v))$, $v \in N$.

2.2.2. It follows that for any module N the set $\mathrm{Hom}(M_1 \oplus M_2, N)$ is in a one-to-one correspondence with the product $\mathrm{Hom}(M_1, N) \times \mathrm{Hom}(M_2, N)$ (each homomorphism $M_1 \oplus M_2 \longrightarrow N$ corresponds to a pair of homomorphisms $M_1 \longrightarrow N$ and $M_2 \longrightarrow N$), and the set $\mathrm{Hom}(N, M_1 \oplus M_2)$ is in a one-to-one correspondence with the product $\mathrm{Hom}(N, M_1) \times \mathrm{Hom}(N, M_2)$.

Proposition. *If R is a commutative ring, then for any R-modules M_1, M_2 and N the bijections defined above are module isomorphisms* $\mathrm{Hom}(M_1 \oplus M_2, N) \cong \mathrm{Hom}(M_1, N) \times \mathrm{Hom}(M_2, N)$ *and* $\mathrm{Hom}(N, M_1 \oplus M_2) \cong \mathrm{Hom}(N, M_1) \times \mathrm{Hom}(N, M_2)$.

2.3. Categories and universal objects

The notion of "universality" comes from the *category theory*.

2.3.1. *A category* consists of *objects* and *morphisms* $\varphi \colon A \longrightarrow B$ between objects. (Usually, objects are sets, and morphisms are mappings between these sets, but this is not required.) A category can be seen as a directed graph whose vertices are called "objects" and (directed) edges are called "morphisms". For any two morphisms $\varphi \colon A \longrightarrow B$ and $\psi \colon B \longrightarrow C$ their *composition* morphism $\psi \circ \varphi \colon A \longrightarrow C$ must be defined, and the operation of composition must be associative: $\tau \circ (\psi \circ \varphi) = (\tau \circ \psi) \circ \varphi$. Also, for every object A *the identity morphism* $\mathrm{Id}_A \colon A \longrightarrow A$ must exist so that $\varphi \circ \mathrm{Id}_A = \varphi$ for any morphism $\varphi \colon A \longrightarrow B$ and $\mathrm{Id}_A \circ \varphi = \varphi$ for any morphism $\varphi \colon B \longrightarrow A$.

2.3.2. Here are some examples of categories:

(i) In the *category of sets* the objects are sets and the morphisms are mappings between these sets.

(ii) In the *category of groups* the objects are groups and the morphisms are group homomorphisms.

(iii) In the *category of topological spaces* the objects are topological spaces and the morphisms are continuous mappings.

(iv) Let R be a ring; then *the category of (left) R-modules* consists of (left) R-modules as objects and their homomorphisms as morphisms.

(v) In the *category of sets with a marked element* the objects are pairs (A, a) where A is a set and $a \in A$ and the morphisms between two objects (A, a) and (B, b) are mappings $\varphi\colon A \longrightarrow B$ with $\varphi(a) = b$.

2.3.3. A morphism $\varphi\colon A \longrightarrow B$ is said to be *an isomorphism* if there is an *inverse morphism* $\psi\colon B \longrightarrow A$ such that $\varphi\circ\psi = \mathrm{Id}_B$ and $\psi\circ\varphi = \mathrm{Id}_A$; in this case, the objects A and B are said to be *isomorphic*.

2.3.4. An object A of a category is said to be *universal repelling*, or *initial*, if for any object B of this category there is a unique morphism $\varphi\colon A \longrightarrow B$. An object A of a category is said to be *universal attracting*, or *terminal*, if for any object B of this category there is a unique morphism $\varphi\colon B \longrightarrow A$.

If a universal repelling, or attracting, object exists, then it is unique up to an isomorphism (that is, any two such objects are isomorphic). Indeed, assume that A_1 and A_2 are two repelling objects in some category. Then there is a unique morphism $\varphi_1\colon A_1 \longrightarrow A_2$ and a unique morphism $\varphi_2\colon A_2 \longrightarrow A_1$. Then the composition $\varphi_1\circ\varphi_2$ is the unique morphism $A_1 \longrightarrow A_1$, which must be Id_{A_1}, and likewise, $\varphi_2\circ\varphi_1 = \mathrm{Id}_{A_2}$; hence, φ_1 is an isomorphism.

2.3.5. Example. Let S be a set, consider the category of groups G with mappings $S \longrightarrow G$: The objects in this category are pairs (G, η) where G is a group and η is a mapping $S \longrightarrow G$, and morphisms $(G, \eta) \longrightarrow (H, \tau)$ are homomorphisms $\varphi\colon G \longrightarrow H$ for which the diagram

$$\begin{array}{ccc} & S & \\ {\scriptstyle\eta}\swarrow & & \searrow{\scriptstyle\tau} \\ G & \xrightarrow{\varphi} & H \end{array}$$

is commutative. In this category the uiversal repelling object is the free group generated by S.

2.3.6. The direct sum $M_1 \oplus M_2$ of two R-modules M_1, M_2 is the universal repelling object in the category whose objects are the triplets $(N, \varphi_1, \varphi_2)$ where N is an R-module and $\varphi_1\colon M_1 \longrightarrow N$, $\varphi_2\colon M_2 \longrightarrow N$ are homomorphisms, with morphisms between $(N, \varphi_1, \varphi_2)$ and (K, ψ_1, ψ_2) being homomorphisms $\eta\colon N \longrightarrow K$ that make the diagram

$$\begin{array}{ccccc} & & M_1 & & \\ & {\scriptstyle\varphi_1}\swarrow & & \searrow{\scriptstyle\psi_1} & \\ N & & \xrightarrow{\sigma} & & K \\ & {\scriptstyle\varphi_2}\nwarrow & & \nearrow{\scriptstyle\psi_2} & \\ & & M_2 & & \end{array}$$

commutative.

2.3.7. Likewise, $M_1 \oplus M_2$ is the universal attracting object in the category of triplets $(N, \varphi_1, \varphi_2)$ where N is an R-module and $\varphi_1\colon N \longrightarrow M_1$, $\varphi_2\colon N \longrightarrow M_2$ are homomorphisms, with morphisms between $(N, \varphi_1, \varphi_2)$ and (K, ψ_1, ψ_2) being homomorphisms $\eta\colon N \longrightarrow K$ that make commutative the diagram

$$\begin{array}{ccccc} & & M_1 & & \\ & {\scriptstyle\varphi_1}\nearrow & & \nwarrow{\scriptstyle\psi_1} & \\ N & & \xrightarrow{\eta} & & K. \\ & {\scriptstyle\varphi_2}\searrow & & \swarrow{\scriptstyle\psi_2} & \\ & & M_2 & & \end{array}$$

2.4. The direct product and the direct sum of families of modules

The direct product and the direct sum of a finite collection of modules are the same, but when infinitely many modules are involved, the notions of a direct product and a direct sum differ!

2.4.1. Let $\{M_\alpha\}_{\alpha\in\Lambda}$ be a collection of R-modules. *The direct product* $\prod_{\alpha\in\Lambda} M_\alpha$ is the module

$$\prod_{\alpha\in\Lambda} M_\alpha = \Big\{(u_\alpha)_{\alpha\in\Lambda} : u_\alpha \in M_\alpha \text{ for all } \alpha\in\Lambda\Big\},$$

with the addition and the multiplication by scalars defined by

$$(u_\alpha)_{\alpha\in\Lambda} + (v_\alpha)_{\alpha\in\Lambda} = (u_\alpha + v_\alpha)_{\alpha\in\Lambda},\ a(u_\alpha)_{\alpha\in\Lambda} = (au_\alpha)_{\alpha\in\Lambda}.$$

For each $\alpha\in\Lambda$, M_α is still naturally identified with a submodule and with a quotient module of $\prod_{\alpha\in\Lambda} M_\alpha$ (and so, is a direct summand of this product); however, $\prod_{\alpha\in\Lambda} M_\alpha$ is no longer generated by the submodules M_α. The direct product $\prod_{\alpha\in\Lambda} M_\alpha$ is the universal attracting object in the category whose objects are the pairs $\big(N, (\varphi_\alpha)_{\alpha\in\Lambda}\big)$, where N is an R-module and $\varphi_\alpha\colon N\longrightarrow M_\alpha$, $\alpha\in\Lambda$, are homomorphisms, and morphisms between $(N, (\varphi_\alpha)_{\alpha\in\Lambda})$ and $(K, (\psi_\alpha)_{\alpha\in\Lambda})$ are homomorphisms $\eta\colon N\longrightarrow K$ satisfying $\varphi_\alpha = \psi_\alpha\circ\eta$ for all $\alpha\in\Lambda$. (In this category, the only morphism $\eta\colon (N, (\varphi_\alpha)_{\alpha\in\Lambda}) \longrightarrow \prod_{\alpha\in\Lambda} M_\alpha$ is that defined by $\eta(v) = (\varphi_\alpha(v))_{\alpha\in\Lambda}$, $v\in N$.)

2.4.2. *The direct sum* $\bigoplus_{\alpha\in\Lambda} M_\alpha$ is the submodule $\sum_{\alpha\in\Lambda} M_\alpha$ of the direct product $\prod_{\alpha\in\Lambda} M_\alpha$ that consists of elements $(u_\alpha)_{\alpha\in\Lambda}$ with $u_\alpha = 0$ for all but finitely many $\alpha\in\Lambda$:

$$\bigoplus_{\alpha\in\Lambda} M_\alpha = \Big\{(u_\alpha)_{\alpha\in\Lambda} : u_\alpha \in M_\alpha \text{ for all } \alpha\in\Lambda,\ u_\alpha = 0 \text{ for all but finitely many } \alpha\Big\}.$$

Every element of $\bigoplus_{\alpha\in\Lambda} M_\alpha$ is uniquely representable as a sum $\sum_{i=1}^{k} u_{\alpha_i}$ with distinct $\alpha_1,\dots,\alpha_k\in\Lambda$ and $u_{\alpha_i}\in M_{\alpha_i}$, $i=1,\dots,k$; or in the form $\sum_{\alpha\in\Lambda}^{\text{fin}} u_\alpha$ (that is, a sum $\sum_{\alpha\in\Lambda} u_\alpha$ where $u_\alpha = 0$ for all but finitely many αs) with $u_\alpha\in M_\alpha$ for all $\alpha\in\Lambda$. The direct sum $\bigoplus_{\alpha\in\Lambda} M_\alpha$ is the universal attracting object in the category whose objects are the pairs $\big(N, (\varphi_\alpha)_{\alpha\in\Lambda}\big)$, where N is an R-module and $\varphi_\alpha\colon M_\alpha\longrightarrow N$, $\alpha\in\Lambda$, are homomorphisms, and morphisms between $(N, (\varphi_\alpha)_{\alpha\in\Lambda})$ and $(K, (\psi_\alpha)_{\alpha\in\Lambda})$ are homomorphisms $\sigma\colon N\longrightarrow K$ satisfying $\psi_\alpha = \sigma\circ\varphi_\alpha$ for all $\alpha\in\Lambda$. (In this category, the only morphism $\sigma\colon \bigoplus_{\alpha\in\Lambda} M_\alpha \longrightarrow (N, (\varphi_\alpha)_{\alpha\in\Lambda})$ is that defined by $\sigma\big(\sum_{\alpha\in\Lambda} u_\alpha\big) = \sum_{\alpha\in\Lambda} \varphi_\alpha(u_\alpha)$; notice that this sum is finite.)

2.4.3. Let $\{M_\alpha\}_{\alpha\in\Lambda}$ be a family of R-modules, and let N be an R-module. Any family of homomorphisms $\varphi_\alpha\colon N\longrightarrow M_\alpha$, $\alpha\in\Lambda$, defines the homomorphism $\varphi\colon N\longrightarrow\prod_{\alpha\in\Lambda} M_\alpha$ by $\varphi(v) = (\varphi_\alpha(v))_{v\in N}$, and vice versa, any such homomorphism φ defines a family of homomorphisms $\{\varphi_\alpha\}_{\alpha\in\Lambda}$. We therefore have a one-to-one correspondence between $\prod_{\alpha\in\Lambda}\operatorname{Hom}(N, M_\alpha)$ and $\operatorname{Hom}\big(N, \prod_{\alpha\in\Lambda} M_\alpha\big)$,

$$\operatorname{Hom}\Big(N, \prod_{\alpha\in\Lambda} M_\alpha\Big) \leftrightarrow \prod_{\alpha\in\Lambda}\operatorname{Hom}(N, M_\alpha).$$

In the case R is commutative, this bijection is an R-module isomorphism.

2.4.4. Likewise, any family of homomorphisms $\varphi_\alpha\colon M_\alpha\longrightarrow N$, $\alpha\in\Lambda$, defines the homomorphism $\varphi\colon\prod_{\alpha\in\Lambda} M_\alpha\longrightarrow N$ by $\varphi\big(\sum_{\alpha\in\Lambda}^{\text{fin}} u_\alpha\big) = \sum_{\alpha\in\Lambda}^{\text{fin}}\varphi_\alpha(u_\alpha)$, $u_\alpha\subset M_\alpha$, $\alpha\in\Lambda$, and vice versa, any such homomorphism φ defines a family of homomorphisms $\{\varphi_\alpha\}_{\alpha\in\Lambda}$. We therefore have a one-to-one correspondence between $\prod_{\alpha\in\Lambda}\operatorname{Hom}(M_\alpha, N)$ and $\operatorname{Hom}\big(\bigoplus_{\alpha\in\Lambda} M_\alpha, N\big)$:

$$\operatorname{Hom}\Big(\bigoplus_{\alpha\in\Lambda} M_\alpha, N\Big) \leftrightarrow \prod_{\alpha\in\Lambda}\operatorname{Hom}(M_\alpha, N).$$

In the case R is commutative, this bijection is an R-module isomorphism.

2.5. The internal direct sum of two submodules and splitting short exact sequences

The direct sum of two modules defined in 2.1 is the so-called *external* direct sum; we will now deal with the *internal* one, appearing when a module already exists and only has to be recognized as a direct sum of its submodules.

2.5.1. Let M be a module and let M_1, M_2 be two submodules of M. We say that M *is a direct sum of* M_1 *and* M_2 and write $M = M_1 \oplus M_2$ if the triplet (M, ξ_1, ξ_2), where ξ_i are the embeddings $M_i \longrightarrow M$, is the universal repelling object in the category of triplets described in 2.3.6.

If $M = M_1 \oplus M_2$, we say that M_1 and M_2 are *direct summands* of M.

2.5.2. There are several criteria for a module to be a direct sum of two its submodules:

Theorem. *Let M be a module and M_1, M_2 be its submodules. Then the following are equivalent:*
(i) $M = M_1 \oplus M_2$*;*
(ii) $M = M_1 + M_2$ *and* $M_1 \cap M_2 = 0$*;*
(iii) *every element u of M is uniquely representable in the form $u = u_1 + u_2$ with $u_1 \in M_1$ and $u_2 \in M_2$;*
(iv) *for the projection homomorphism $\pi\colon M \longrightarrow M/M_1$, the restriction $\pi|_{M_2}$ is an isomorphism between M_2 and M/M_1.*

Proof. If $M = M_1 \oplus M_2$, then M is isomorphic to "the abstract", external direct sum $M_1 \oplus M_2$, under an isomorphism that is identical on M_1 and M_2; since statements (ii)-(iv) hold for the outer direct sum, they hold for M. So, (i) implies (ii)-(iv).

(ii) and (iii) are clearly equivalent: $M = M_1 + M_2$ means that every $u \in M$ is representable as $u_1 + u_2$ with $u_1 \in M_1$ and $u_2 \in M_2$. If $M_1 \cap M_2 = 0$, such a representaqtion is unique: if $u_1 + u_2 = v_1 + v_2$ with $u_1, v_1 \in M_1$ and $u_2, v_2 \in M_2$, then $u_1 - v_1 = v_2 - u_2 \in M_1 \cap M_2 = 0$, so $u_1 = v_1$ and $u_2 = v_2$. Conversely, if every $u \in M$ has a unique representation in the form $u_1 + u_2$ with $u_i \in M_i$, then since for $u \in M_1 \cap M_2$ we have $u = u + 0 = 0 + u$ with $u \in M_1$, $0 \in M_2$ and $0 \in M_1$, $u \in M_2$, we get that $u = 0$.

Next, (iii) implies (i): Given (iii), consider the homomorphism $\varphi\colon M_1 \oplus M_2 \longrightarrow M$, $\varphi(u_1, u_2) = u_1 + u_2$. Then φ is identical on M_1 and M_2: $\varphi(u_1) = u_1$ and $\varphi(u_2) = u_2$ for any $u_1 \in M_1$ and $u_2 \in M_2$. And since every $u \in M$ can be uniquely written as $u_1 + u_2$ with $u_1 \in M_1$ and $u_2 \in M_2$, φ is bijective.

Suppose (iv). Then $M_1 \cap M_2 = \ker\bigl(\pi|_{M_2}\bigr) = 0$ since $\pi|_{M_2}$ is injective. Let $u \in M$. Since $\pi|_{M_2}$ is surjective, there is $u_2 \in M_2$ such that $\pi(u) = \pi(u_2)$. Then $\pi(u - u_2) = 0$, so $u_1 = u - u_2 \in M_1$, and $u = u_1 + u_2$. Hence, (iv) implies (ii). ∎

2.5.3. If N is a submodule of a module M, we may ask whether N is a direct summand of M, that is, whether $M = N \oplus K$ for some submodule K of M. Similarly, if K is a quotient module of a module M, we may ask whether M is a direct product of (a copy of) K and some other submodule. Theorem 2.5.6 below helps recognize these situations.

2.5.4. For an epimorphism $\varphi\colon M \longrightarrow N$, if a homomorphism $\sigma\colon N \longrightarrow M$ is such that $\varphi\circ\sigma = \mathrm{Id}_N$, we say that σ is *a section* of φ:

$$M \underset{\sigma}{\overset{\psi}{\rightleftarrows}} N \longrightarrow 0.$$

If in a short exact sequence $0 \longrightarrow N \xrightarrow{\varphi} M \xrightarrow{\psi} K \longrightarrow 0$ the epimorphism ψ has a section, we say that this sequence *splits from the right*:

$$0 \longrightarrow N \longrightarrow M \underset{\sigma}{\overset{\psi}{\rightleftarrows}} K \longrightarrow 0.$$

2.5.5. For a monomorphism $\varphi\colon N \longrightarrow M$, if a homomorphism $\tau\colon M \longrightarrow N$ is such that $\tau\circ\varphi = \mathrm{Id}_N$, we say that τ is *a projection* for φ:

$$0 \longrightarrow N \underset{\tau}{\overset{\varphi}{\rightleftarrows}} M.$$

If in a short exact sequence $0 \longrightarrow N \xrightarrow{\varphi} M \xrightarrow{\psi} K \longrightarrow 0$ the monomorphism φ has a projection, we say that the sequence *splits from the left*:

$$0 \longrightarrow N \underset{\tau}{\overset{\varphi}{\rightleftarrows}} M \longrightarrow K \longrightarrow 0.$$

2.5.6. Theorem. (i) *If a short exact sequence* $0 \longrightarrow N \xrightarrow{\varphi} M \xrightarrow{\psi} K \longrightarrow 0$ *splits from the right, with* $\sigma\colon K \longrightarrow M$ *being a section of* ψ*, then* $M = N' \oplus K'$ *where* $N' = \varphi(N)$ *is isomorphic to* N *(under* φ*) and* $K' = \sigma(K)$ *is isomorphic to* K *(under* σ*).*

(ii) *If a short exact sequence* $0 \longrightarrow N \xrightarrow{\varphi} M \xrightarrow{\psi} K \longrightarrow 0$ *splits from the left, with* $\tau\colon M \longrightarrow N$ *being a projection for* φ*, then* $M = N' \oplus K'$ *where* $N' = \varphi(N)$ *is isomorphic to* N *(under* φ*) and* $K' = \ker(\tau)$ *is isomorphic to* K *(under* $\psi|_{K'}$*).*

Proof. (i) Consider the diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & N & \xrightarrow{\eta_1} & N \oplus K & \underset{\eta_2}{\overset{\pi_2}{\rightleftarrows}} & K & \longrightarrow & 0 \\ & & \updownarrow & & \downarrow\rho & & \updownarrow & & \\ 0 & \longrightarrow & N & \xrightarrow{\varphi} & M & \underset{\sigma}{\overset{\psi}{\rightleftarrows}} & K & \longrightarrow & 0 \end{array}$$

(where η_i are the embeddings and π_i are the projections). By the universal property of the direct sum, the homomorphisms $\varphi\colon N \longrightarrow M$ and $\sigma\colon K \longrightarrow M$ define a homomorphism $\rho\colon N \oplus K \longrightarrow M$ such that $\rho\circ\eta_1 = \varphi$ and $\rho\circ\eta_2 = \sigma$. This makes the diagram commutative. (The right square of the diagram is commutative since σ is a section of ψ: for any $u \in N$ and $v \in K$ we have $\rho(u,0) = \varphi(u)$ and $\rho(0,v) = \sigma(v)$, so $\rho(u,v) = \varphi(u) + \sigma(v)$, so $\psi(\rho(u,v)) = \psi(\varphi(u)) + \psi(\sigma(v)) = v = \pi_2(u,v)$.) By the short five lemma, ρ is an isomorphism, so $M = \varphi(N) \oplus \sigma(K)$. ■

As a corollary we get that a short exact sequence splits from the left iff it splits from the right; so, we can simply say that it *splits*.

2.6. The internal direct sum of a family of submodules

2.6.1. For finitely many modules $M_1, \ldots, M_k$, the direct sum and the direct product of R-modules $M_1, \ldots, M_k$ is

$$\bigoplus_{i=1}^{k} M_i = \prod_{i=1}^{k} M_i = \big\{(u_1, \ldots, u_k) : u_i \in M_i,\ i = 1, \ldots, k\big\}$$

with the componentwise addition and multiplication by scalars:

$$(u_1, \ldots, u_k) + (v_1, \ldots, v_k) = (u_1 + v_1, \ldots, u_k + v_k), \quad a(u_1, \ldots, u_k) = (au_1, \ldots, au_k),$$
$$u_i, v_i \in M_i,\ i = 1, \ldots, k,\ a \in R.$$

For each i, M_i is identified with a submodule and with a quotient module of $\bigoplus_{i=1}^{k} M_i$. Under the first of these identifications we have $M = M_1 + \ldots + M_k$ and $M_j \cap \sum_{\substack{i=1 \\ i \neq j}}^{k} M_i = 0$ for every $j = 1, \ldots, k$.

2.6.2. $\bigoplus_{i=1}^{k} M_i$ is the universal repelling object in the category whose objects are the $k+1$-tuples $(N, \varphi_1, \ldots, \varphi_k)$, where N is an R-module and $\varphi_i\colon M_i \longrightarrow N$, $i = 1, \ldots, k$, are homomorphisms, and morphisms between $(N, \varphi_1, \ldots, \varphi_k)$ and $(K, \psi_1, \ldots, \psi_k)$ are homomorphisms $\sigma\colon N \longrightarrow K$ satisfying $\psi_i = \sigma\circ\varphi_i$ for all $i = 1, \ldots, k$:

$$\begin{array}{ccc} & & M_i \\ & {\scriptstyle\varphi_i}\swarrow \quad \searrow{\scriptstyle\psi_i} & \\ N & \xrightarrow{\sigma} & K. \end{array}$$

(In this category, the only morphism $\sigma\colon \bigoplus_{i=1}^{k} M_i \longrightarrow (N, \varphi_1, \ldots, \varphi_k)$ is that defined by $\sigma(u_1, \ldots, u_k) = \varphi_1(u_1) + \cdots + \varphi_k(u_k)$.)

2.6.3. $\prod_{i=1}^{k} M_i = \bigoplus_{i=1}^{k} M_i$ is also the universal attracting object in the category whose objects are the $k+1$-tuples $(N, \varphi_1, \ldots, \varphi_k)$, where N is an R-module and $\varphi_i\colon N \longrightarrow M_i$, $i = 1, \ldots, k$, are homomorphisms, and morphisms between $(N, \varphi_1, \ldots, \varphi_k)$ and $(K, \psi_1, \ldots, \psi_k)$ are homomorphisms $\eta\colon N \longrightarrow K$ satisfying $\varphi_i = \psi_i\circ\eta$ for all $i = 1, \ldots, k$:

$$\begin{array}{ccc} & & M_i \\ & {\scriptstyle\varphi_i}\nearrow \quad \nwarrow{\scriptstyle\psi_i} & \\ N & \xrightarrow{\eta} & K. \end{array}$$

(In this category, the only morphism $\eta\colon (N, \varphi_1, \ldots, \varphi_k) \longrightarrow \bigoplus_{i=1}^{k} M_i$ is that defined by $\eta(v) = \big(\varphi_1(v), \ldots, \varphi_k(v)\big)$, $v \in N$.)

2.6.4. Let M be a module and M_α, $\alpha \in \Lambda$, be a (finite or infinite) family of submodules of M. We say that M is *an (internal) direct sum* of this family and write $M = \bigoplus_{\alpha\in\Lambda} M_\alpha$ if M, with the embeddings $M_\alpha \longrightarrow M$, $\alpha \in \Lambda$, is a universal repelling object in the category described in 2.4.2.

2.6.5. Theorem. *Let M be a module and M_α, $\alpha \in \Lambda$, be a family of submodules of M. Then the following are equivalent:*

(i) $M = \bigoplus_{\alpha\in\Lambda} M_\alpha$.

(ii) $M = \sum_{\alpha\in\Lambda} M_\alpha$ *(that is, M is generated by the modules M_α) and for every $\alpha \in \Lambda$, $M_\alpha \cap \sum_{\substack{\beta\in\Lambda\\ \beta\neq\alpha}} M_\beta = 0$ (which is equivalent to saying that for any distinct $\alpha, \alpha_1, \ldots, \alpha_k$, $M_\alpha \cap \sum_{i=1}^k M_{\alpha_i} = 0$).*

(iii) *Every nonzero $u \in M$ is uniquely representable in the form $u = \sum^{\text{fin}}_{\alpha\in\Lambda} u_\alpha$ with $u_\alpha \in M_\alpha$ for every α. (Equivalently, every nonzero element u of M is uniquely representable in the form $u = u_1 + \ldots + u_k$ with nonzero $u_i \in M_{\alpha_i}$, $i = 1, \ldots, k$, and distinct $\alpha_1, \ldots, \alpha_k \in \Lambda$.)*

2.7. The Chinese remainder theorem and p-primary components of modules

Here are two situations where direct sums naturally appear.

2.7.1. If M is a left R-module and I is a left ideal of R, then the set

$$IM = \Big\{\sum_{i=1}^k a_i u_i : k \in \mathbb{N},\ a_i \in I,\ u_i \in M,\ i = 1, \ldots, k\Big\}$$

is a submodule of M.

2.7.2. The Chinese remainder theorem. *Let R be a commutative unital ring, let M be an R-module, and let $I_1, \ldots, I_n$ be ideals in R which are pairwise comaximal (that is, $I_i + I_j = R$ for all $i \neq j$). Then $I_1M \cap \cdots \cap I_kM = (I_1 \cdots I_n)M$ and $M/(I_1 \cdots I_n)M \cong \bigoplus_{i=1}^n (M/I_iM)$, under the homomorphism that maps $u + (I_1 \cdots I_n)M$ to $\big(u + I_1M, \ldots, u + I_nM\big)$, $u \in M$. In particular, if $(I_1 \cdots I_n)M = 0$, then $M \cong \bigoplus_{i=1}^n (M/I_iM)$.*

Proof. We know that I_1 and $I_2 \cdots I_n$ are comaximal, so it suffices to prove the theorem for the case $n = 2$ and then use induction.

Let I and J be comaximal, let $a \in I$ and $b \in J$ be such that $a+b = 1$. We clearly have $IJM \subseteq IM \cap JM$, let now $u \in IM \cap JM$; then $u = au + bu \in aJM + bIM \subseteq IJM$, so $IM \cap JM \subseteq IJM$.

Consider the natural homomorphism $\varphi\colon M \longrightarrow (M/IM) \oplus (M/JM)$. We have $\ker(\varphi) = IM \cap JM$, so it remains to show that φ is surjective. Given $v, w \in M$, put $u = av + bw$; then $u = v - bv + bw = v \bmod JM$ and $u = av + w - aw = w \bmod IM$, so $\varphi(u) = (w \bmod IM, w \bmod JM)$. ■

2.7.3. Let R be a PID and let M be a torsion R-module. For a prime $p \in R$, *the p-primary component* of M is the submodule $M_p = \big\{u \in M : p^k u = 0 \text{ for some } k \in \mathbb{N}\big\}$. Then M is a direct sum of its nonzero primary components: $M = \bigoplus_{p:M_p\neq 0} M_p$.

In the case M has a anonzero annihilator $a = p_1^{r_1} \cdots p_k^{r_k}$, where p_i are distinct primes in R and $r_i \geq 1$ for all i, we have $M_{p_i} = \operatorname{Ann}(p_i^{r_i})$ for all i, and $M = \bigoplus_{i=1}^k M_{p_i}$.

---

3. Free modules

In this section R is assumed to be a unital ring.

3.1. Free modules of finite rank

3.1.1. Let $n \in \mathbb{N}$. *The free R-module of rank n* is the direct product R^n (which is the same as the direct sum $\sum_{i=1}^n R$) of n copies of R; more generally, an R-module M is said to be *free of rank n* if $M \cong R^n$.

3.1.2. Let B be a subset of a module M. *A linear combination* of elements of B is a (finite) sum of the form $\sum^{\text{fin}}_{v\in B} a_v v$ with $a_v \in R$, $v \in B$. A subset B of a module M is called *a basis of M* if $M = \bigoplus_{v\in B} Rv$, that is, if every element $u \in M$ is uniquely representable as a linear combination of element of B, $u = \sum_{v\in B} a_v v$; the scalars a_v, $v \in B$, are called *the coordinates of u in the basis B*.

3.1.3. The elements $e_1 = (1, 0, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$, …, $e_n = (0, \dots, 0, 1)$ form *the standard basis* of R^n: any $u = (a_1, \dots, a_n) \in R^n$ is uniquely representable in the form $u = \sum_{i=1}^{n} a_i e_i$.

3.1.4. If M is free of rank n and $\varphi\colon R^n \longrightarrow M$ is an isomorphism, let $u_i = \varphi(e_i)$, $i = 1, \dots, n$; then $\{u_1, \dots, u_n\}$ is a basis of M. Conversely, if M has a basis $\{u_1, \dots, u_n\}$ of cardinality n, we can construct an isomorphism $M \longrightarrow R^n$ by mapping $u = \sum_{i=1}^{n} a_i u_i \in M$ to the n-tuple $(a_1, \dots, a_n) \in R^n$ of its coordinates in the basis $\{u_1, \dots, u_n\}$. Hence, M is free of rank n iff M has a basis of cardinality n.

3.2. Free modules generated by sets

Let S be a set, finite or infinite.

3.2.1. *The free R-module generated by S* is the direct sum $\bigoplus_{s\in S} R$; let us denote it by $\mathcal{F}_R(S)$. It consists of functions $S \longrightarrow R$, $s \mapsto a_s$, such that $a_s = 0$ for all but finitely many $s \in S$. For each $s \in S$ let e_s be the function that is equal to 1 at s and to 0 at all other elements of S; then $\{e_s\}_{s\in S}$ is a basis of $\mathcal{F}_R(S)$: every element of this module is uniquely representable as a linear combination $\sum_{s\in S}^{\text{fin}} a_s e_s$. Identifying each $s \in S$ with the corresponding e_s, we may assume that $S \subseteq \mathcal{F}_R(S)$ and is a basis of this module, so that the elements of $\mathcal{F}_R(S)$ take the form $\sum_{s\in S} a_s s$.

3.2.2. We call an R-module M *free* if it is isomorphic to a free module. This is so iff M has a basis B, in which case M is (isomorphic to) the module $\mathcal{F}_R(B)$.

3.2.3. *The rank* of a free module M is defined as the cardinality of any its basis. (Generally speaking, the rank is not defined uniquly: there is a ring R for which $R^2 \cong R$ as R-modules. However if the ring R is commutative and unital, the rank is well defined.)

3.2.4. Let M be an R-module and let $\eta\colon S \longrightarrow M$ be a mapping. Then η is uniquely extendible to a homomorphism $\mathcal{F}_R(S) \longrightarrow M$, by putting $\eta\big(\sum_{s\in S}^{\text{fin}} a_s s\big) = \sum_{s\in S}^{\text{fin}} a_s \eta(s)$. This means that $\mathcal{F}_R(S)$, with the natural embedding $S \longrightarrow \mathcal{F}_R(S)$, is the universal repelling object in the category of the pairs (M, η), where η is a mapping $S \longrightarrow M$, with morphisms between two objects (M, η) and (N, θ) being homomorphisms $\varphi\colon M \longrightarrow N$ satisfying $\theta = \varphi \circ \eta$:

$$\begin{array}{ccc} & S & \\ {\scriptstyle\eta}\swarrow & & \searrow{\scriptstyle\theta} \\ M & \xrightarrow{\varphi} & N. \end{array}$$

3.2.5. In particular, if S is a subset of M, we have a unique homomorphism $\varphi\colon \mathcal{F}_R(S) \longrightarrow M$ that is identical on S: $\varphi(s) = s$ for all $s \in S$. If S is a generating set of M, then φ is an epimorphism, and M is isomorphic to a quotient module of $\mathcal{F}_R(S)$. We therefore have the following theorem:

Theorem. *Every module is isomorphic to a quotient module of a free module; if a module M is generated by a set S, then M is isomorphic to a quotient module of $\mathcal{F}_R(S)$.*

3.3. Maximal free submodules and the rank of a module

3.3.1. A subset B of a module M is said to be *linearly independent* if a linear combination $\sum_{u\in B}^{\text{fin}} a_u u$ of elements of B is equal to 0 only if $a_u = 0$ for all $u \in B$.

3.3.2. Lemma. *A subset B of a module M is a basis of M iff B is linearly independent and generates M.*

3.3.3. If B is a linearly independent subset of M and N is the submodule of M generated by B, then B is a basis in N and so, N is free.

3.3.4. A standard application of Zorn's lemma gives:

Theorem. *Every module M has a maximal linearly independent subset. Moreover, every linearly independent subset of M is contained in a maximal one.*

3.3.5. I will call the module generated by a maximal linearly independent subset of M *a maximal free submodule of M*. (Though such a module may not be, in fact, a maximal element in the set of free submodules of M! Consider, for instance, $\mathbb{Q}$ as a $\mathbb{Z}$-module.)

3.3.6. Proposition. *Let N be a free submodule of a module M. Then N is a maximal free submodule of M iff M/N is a torsion module.*

Proof. Let B be a basis of N. For $u \in M$, the set $B \cup \{u\}$ is linearly dependent iff $au + \sum_{v\in B}^{\text{fin}} a_v v = 0$ for some $a \neq 0$, iff $au = 0 \bmod N$ for some $a \neq 0$, iff $\bar{u} = u \bmod N$ is a torsion element in M/N. So, B is a maximal linearly independent subset of M iff M/N is a torsion module. ∎

3.3.7. If R is an integral domain, then *the rank* of an R-module M, $\operatorname{rank}_R M$ or $\operatorname{rank} M$, is defined as the rank of its maximal free submodule, that is, the cardinality of a maximal linearly independent subset. We will see later (in 6.4.3) that $\operatorname{rank} M$ is well defined.

3.4. Vector spaces and dimension

3.4.1. Theorem. *Any vector space is a free module; any maximal linearly independent subset of a vector space is a basis of this space.*

Proof. Let F be a field and V be a vector space. Let N be a maximal free F-submodule of V. Then V/N is a torsion module; but since F has no nontrivial ideals, there are no nontrivial torsion F-modules, so $V/N = 0$. Hence, $V = N$ and V is free. ■

3.4.2. Theorem. *Any subspace W of a vector space V is a free summand of V: there exists a subspace W' of V such that $V = W \oplus W'$.*

3.4.3. Theorem. *Any two bases in a vector space have the same cardinality.*

The proof of this theorem is based on *the replacement lemma* (or *theorem*), and, in the case of infnite dimensional spaces, requires Zorn's lemma.

The cardinality of any basis of an F-vector space V (that is, the rank of V as a free F-module) is called *the dimension* of V and is denoted by $\dim V$; the theorem says that $\dim V$ is well defined. Two vector F-spaces are isomorphic iff they have the same dimension.

3.4.4. As a corollary, we get:

Corollary. *If R is a commutative unital ring, then the rank of a free R-module is well defined. (That is, any two maximal linearly independent subsets in M have the same cardinality.)*

Proof. Let I be a maximal ideal in R, let F be the field R/I. If M is an R-module, then M/IM has a structure of an R/I-module, that is, is an F-vector space. And if M is free, then $\operatorname{rank}_R M = \dim_F(M/IM)$, and so is well defined. (If $M \cong R^n$, then $M/IM \cong R^n/(IR^n) = R^n/I^n \cong (R/I)^n = F^n$. In the case M has infinite rank, given a basis B of M, $B \bmod IM$ is a basis in M/IM. Indeed, $B \bmod IM$ generates (spans) M/IM over R and so over $F = R/I$; and the fact that any finite subset of B is linearly indpendent modulo IM follows from the finite rank case.) ■

3.4.5. If V is a finite dimensional vector space and W is a subspace of V, then $\dim V = \dim W + \dim(V/W)$, and if $\dim W = \dim V$, then $W = V$.

3.4.6. Proposition. *If $\varphi\colon V \longrightarrow W$ is a homomorphism of finite dimensional vector spaces, then* $\dim \varphi(V) = \dim V - \dim \ker(\varphi)$.

3.4.7. As a corollary, we obtain:

Proposition. *If $\varphi\colon V \longrightarrow W$ is a homomorphism of vector spaces with $\dim V = \dim W < \infty$, then φ is an isomorphism iff φ is a monomorphism iff φ is an epimorphism.*

---

4. Tensor product of modules

To avoid unpleasant complications, I'll only consider tensor products of modules over commutative rings; for tensor products of modules over non-commutative rings see the book.

In this section, R will be a commutative unital ring.

4.1. Bilinear mappings of modules

4.1.1. Let M_1, M_2, N be R-modules. A mapping $\beta\colon M_1 \times M_2 \longrightarrow N$ is said to be *bilinear* if

$$\beta(u_1 + v_1, u_2) = \beta(u_1, u_2) + \beta(v_1, u_2); \quad \beta(au_1, u_2) = a\beta(u_1, u_2);$$
$$\beta(u_1, u_2 + v_2) = \beta(u_1, u_2) + \beta(u_1, v_2); \quad \beta(u_1, au_2) = a\beta(u_1, u_2)$$

for all $u_1, v_1 \in M_1$, $u_2, v_2 \in M_2$, and $a \in R$; in other words, if for any $u_2 \in M_2$ the mapping $u_1 \mapsto \beta(u_1, u_2)$ is a homomorphism $M_1 \longrightarrow N$, and for any $u_1 \in M_1$ the mapping $u_2 \mapsto \beta(u_1, u_2)$ is a homomorphism $M_2 \longrightarrow N$.

Example. If A is an R-algebra, then the multiplication in A is a bilinear mapping $A \times A \longrightarrow A$.

4.1.2. The composition of a bilinear mapping and a homomorphism is a bilinear mapping.

4.2. The tensor product of two modules

4.2.1. Let M_1 and M_2 be two R-modules. *The tensor product* $M_1 \otimes_R M_2$, or just $M_1 \otimes M_2$, is the R-module that is the universal repelling object in the category of bilinear mappings from $M_1 \times M_2$; in other words, $M_1 \otimes M_2$ is the R-module with a bilinear mapping $\beta\colon M_1 \times M_2 \longrightarrow M_1 \otimes M_2$ such that for any R-module N with a bilinear mapping $\gamma\colon M_1 \times M_2 \longrightarrow N$ there exists a unique homomorphism $\varphi\colon M_1 \otimes M_2 \longrightarrow N$ with the property that $\gamma = \varphi \circ \beta$:

$$\begin{array}{ccc} & M_1 \times M_2 & \\ & {\scriptstyle\beta}\swarrow \quad \searrow{\scriptstyle\gamma} & \\ M_1 \otimes M_2 & \xrightarrow{\ \varphi\ } & N. \end{array} \tag{4.1}$$

4.2.2. As a universal object, the tensor product is unique up to isomorphism; but its existence is not evident. Here is a direct construction of $M_1 \otimes_R M_2$. Consider the free module $\mathcal{F}_R(M_1 \times M_2)$, that is, the module of formal linear combinations of the pairs $(u_1, u_2) \in M_1 \times M_2$. Let $\mathcal{K}$ be the submodule of $\mathcal{F}_R(M_1 \times M_2)$ generated by the set of elements of the form

$$\begin{array}{ll} (u_1 + v_1, u_2) - (u_1, u_2) - (v_1, u_2), & (au_1, u_2) - a(u_1, u_2), \\ (u_1, u_2 + v_2) - (u_1, u_2) - (u_1, v_2), & (u_1, au_2) - a(u_1, u_2) \end{array}$$

for $u_1, v_1 \in M_1$, $u_2, v_2 \in M_2$, and $a \in R$. Put $M = \mathcal{F}_R(M_1 \times M_2)/\mathcal{K}$, and define $\beta\colon M_1 \times M_2 \longrightarrow M$ by $\beta(u_1, u_2) = \overline{(u_1, u_2)} = (u_1, u_2) + \mathcal{K}$. Then β is bilinear: modulo $\mathcal{K}$, we have $\overline{(u_1 + v_1, u_2)} = \overline{(u_1, u_2)} + \overline{(v_1, u_2)}$, etc. Now, given any bilinear mapping $\gamma\colon M_1 \times M_2 \longrightarrow N$, by the universal property of free modules, there is a unique homomorphism $\widetilde{\varphi}\colon \mathcal{F}_R(M_1 \times M_2) \longrightarrow N$ such that $\widetilde{\varphi}(u_1, u_2) = \gamma(u_1, u_2)$ for all $u_1 \in M_1$, $u_2 \in M_2$. Because of bilinearity of γ, the elements of $\mathcal{K}$ are mapped by $\widetilde{\varphi}$ to 0, that is, $\mathcal{K} \subseteq \ker(\varphi)$; this implies that $\widetilde{\varphi}$ factorizes to a unique homomorphism $\varphi\colon M_1 \otimes M_2 \longrightarrow N$ for which the diagram (4.1) is commutative: $\varphi\big(\overline{(u_1, u_2)}\big) = \gamma(u_1, u_2)$.

4.2.3. Elements of the module $M_1 \otimes M_2$ are called *tensors*. The image $\beta(M_1 \times M_2)$ of $M_1 \times M_2$ generates $M_1 \otimes M_2$. For an element $(u_1, u_2) \in M_1 \times M_2$, its image in $M_1 \otimes M_2$ is denoted by $u_1 \otimes u_2$; tensors of this sort are called *simple*. Therefore, $M_1 \otimes M_2$ is generated by the set of simple tensors: every tensor is a linear combination of simple tensors. Thus, any homomorphism from $M_1 \otimes M_2$ is defined by its values on the simple tensors.

Moreover, if M_1 is generated by a set S_1 and M_2 by a set S_2, then $M_1 \otimes M_2$ is generated by simple tensors of the form $u_1 \otimes u_2$ with $u_1 \in S_1$ and $u_2 \in S_2$. Indeed, for every simple tensor $w = v_1 \otimes v_2 \in M_1 \otimes M_2$ we have $v_1 = \sum_{i=1}^{k} a_i u_{1,i}$ with $u_{1,1}, \ldots, u_{1,k} \in S_1$ and $v_2 = \sum_{j=1}^{l} b_j u_{2,j}$ with $u_{2,1}, \ldots, u_{2.l} \in S_2$, so $w = \sum_{i,j} a_i b_j u_{1,i} \otimes u_{2,j}$, which is a linear combination of tensors of the needed form. Since every tensor is a linear combination of simple tensors, this holds true for every tensor either.

4.3. Elementary properties of tensor multiplication and examples of tensor products

In what follows, M, M_1, M_2, M_3 are R-modules.

4.3.1. Remark. How can we construct a homomorphism $M_1 \otimes M_2 \longrightarrow M$? We first construct a bilinear mapping $\gamma\colon M_1 \times M_2 \longrightarrow M$, and then the universal property of tensor product will guarantee that there is a homomorphism $\varphi\colon M_1 \otimes M_2 \longrightarrow M$ such that $\varphi(u_1 \otimes u_2) = \gamma(u_1, u_2)$ for all $u_1 \in M_1$ and $u_2 \in M_2$. How can we prove that this homomorphism is an isomorphism? We either construct its inverse, or prove that it is both an epi- and monomorphism.

4.3.2. By definition, in the tensor product $M_1 \otimes M_2$,

$$(u_1 + v_1) \otimes u_2 = u_1 \otimes u_2 + v_1 \otimes u_2, \quad u_1 \otimes (u_2 + v_2) = u_1 \otimes u_2 + u_1 \otimes v_2, \quad (au_1) \otimes u_2 = u_1 \otimes (au_2) = a(u_1 \otimes u_2)$$

for any $u_1, v_1 \in M_1$, $u_2, v_2 \in M_2$, $a \in R$.

4.3.3. For any $u_1 \in M_1$ and any $u_2 \in M_2$ we have $u_1 \otimes 0 = 0 \otimes u_2 = 0$; indeed, $u_1 \otimes 0 = u_1 \otimes (0 + 0) = u_1 \otimes 0 + u_1 \otimes 0$. It follows that for any module M, $M \otimes 0 = 0 \otimes M = 0$.

4.3.4. $R \otimes M \cong M$, where the isomorphism is given by $a \otimes u \mapsto au$.

Proof. Define $\beta\colon R \times M \longrightarrow M$ by $\beta(a,u) = au$. β is bilinear, thus induces a homomorphism $\varphi\colon R \otimes M \longrightarrow M$ with $\varphi(a \otimes u) = au$, $a \in R$, $u \in M$. The homomorphism $\psi\colon M \longrightarrow R \otimes M$ defined by $\psi(u) = 1 \otimes u$ is the inverse of φ on simple tensors and thus is the inverse of φ. So, φ is an isomorphism. ∎

4.3.5. $M_1 \otimes M_2 \cong M_2 \otimes M_1$, where the isomorphism is defined by $u_1 \otimes u_2 \mapsto u_2 \otimes u_1$.

Proof. Define $\beta\colon M_1 \times M_2 \longrightarrow M_2 \otimes M_1$ by $\beta(u_1,u_2) = u_2 \otimes u_1$. Then β is a bilinear mapping, thus it induces a homomorphism $\varphi\colon M_1 \otimes M_2 \longrightarrow M_2 \otimes M_1$ with $\varphi(u_1 \otimes u_2) = u_2 \otimes u_1$ for all $u_1 \in M_1$, $u_2 \in M_2$. Similarly, there is a homomorphism $\psi\colon M_2 \otimes M_1 \longrightarrow M_1 \otimes M_2$ with $\varphi(u_2 \otimes u_1) = u_1 \otimes u_2$ for all $u_1 \in M_1$, $u_2 \in M_2$. Since ψ is the inverse of φ on the generators (the simple tensors), ψ is the inverse of φ and φ is an isomorphism. ∎

4.3.6. $(M_1 \otimes M_2) \otimes M_3 \cong M_1 \otimes (M_2 \otimes M_3)$, where the isomorphism is defined by $(u_1 \otimes u_2) \otimes u_3 \mapsto u_1 \otimes (u_2) \otimes u_3)$.

Proof. For each $z \in M_3$, define $\beta_z\colon M_1 \times M_2 \longrightarrow M_1 \otimes (M_2 \otimes M_3)$ by $\beta_z(u_1,u_2) = u_1 \otimes (u_2 \otimes z)$, $u_1 \in M_1$, $u_2 \in M_2$. β_z is bilinear, and thus induces a homomorphism $\varphi_z\colon M_1 \otimes M_2 \longrightarrow M_1 \otimes (M_2 \otimes M_3)$ with $\varphi_z(u_1 \otimes u_2) = u_1 \otimes (u_2 \otimes z)$, $u_1 \in M_1$, $u_2 \in M_2$. Now define a mapping $\beta\colon (M_1 \otimes M_2) \times M_3 \longrightarrow M_1 \otimes (M_2 \otimes M_3)$ by $\beta(w,z) = \varphi_z(w)$, $z \in M_3$, $w \in M_1 \otimes M_2$; in particular, $\beta((u_1 \otimes u_2), z) = u_1 \otimes (u_2 \otimes z)$ for all $u_1 \in M_1$, $u_2 \in M_2$, $z \in M_3$. It is easy to check that β is bilinear, and thus induces a homomorphism $\varphi\colon (M_1 \otimes M_2) \otimes M_3 \longrightarrow M_1 \otimes (M_2 \otimes M_3)$ with $\varphi((u_1 \otimes u_2) \otimes z) = u_1 \otimes (u_2 \otimes z)$ for all $u_1 \in M_1$, $u_2 \in M_2$, $z \in M_3$. Similarly, there is a homomorphism $\psi\colon M_1 \otimes (M_2 \otimes M_3) \longrightarrow (M_1 \otimes M_2) \otimes M_3$ with $\psi(u_1 \otimes (u_2 \otimes z)) = (u_1 \otimes u_2) \otimes z$ for all $u_1 \in M_1$, $u_2 \in M_2$, $z \in M_3$. So, φ and ψ are inverses of each other on the generators of $(M_1 \otimes M_2) \otimes M_3$ and $M_1 \otimes (M_2 \otimes M_3)$, and so, are inverses of each other; hence, φ is an isomorphism. ∎

4.3.7. $(M_1 \oplus M_2) \otimes M_3 \cong (M_1 \otimes M_3) \oplus (M_2 \otimes M_3)$, where the isomorphism is defined by $(u_1,u_2) \otimes u_3 \mapsto (u_1 \otimes u_3, u_2 \otimes u_3)$.

Proof. Define $\beta\colon (M_1 \oplus M_2) \times M_3 \longrightarrow (M_1 \otimes M_3) \oplus (M_2 \otimes M_3)$ by $\beta((u_1,u_2),u_3) = \big(u_1 \otimes u_3, u_2 \otimes u_3\big)$. It is easy to check that β is bilinear, hence it induces a homomorphism $\varphi\colon (M_1 \oplus M_2) \otimes M_3 \longrightarrow (M_1 \otimes M_3) \oplus (M_2 \otimes M_3)$ with $\varphi((u_1,u_2) \otimes u_3) = \big(u_1 \otimes u_3, u_2 \otimes u_3\big)$. In particular, $\varphi((u_1,0) \otimes u_3) = (u_1 \otimes u_3, 0)$ and $\varphi((0,u_2) \otimes u_3) = (0, u_2 \otimes u_3)$ for all $u_i \in M_i$, $i = 1,2,3$.

On the other hand, we have homomorphisms $\psi_1\colon M_1 \otimes M_3 \longrightarrow (M_1 \oplus M_2) \otimes M_3$ and $\psi_2\colon M_2 \otimes M_3 \longrightarrow (M_1 \oplus M_2) \otimes M_3$ with $\psi_1(u_1 \otimes u_3) = (u_1,0) \otimes u_3$ and $\psi_2(u_2 \otimes u_3) = (0,u_2) \otimes u_3$, for all $u_i \in M_i$, $i = 1,2,3$. Hence (by the universal property of the direct sum) there is a homomorphism $\psi\colon (M_1 \otimes M_3) \oplus (M_2 \otimes M_3) \longrightarrow (M_1 \oplus M_2) \otimes M_3$ such that $\psi(u_1 \otimes u_3, 0) = (u_1,0) \otimes u_3$ and $\psi(0, u_2 \otimes u_3) = (0,u_2) \otimes u_3$ for all $u_i \in M_i$, $i = 1,2,3$. Thus, φ and ψ are inverses of each other on the generators of $(M_1 \oplus M_2) \otimes M_3$ and $(M_1 \otimes M_3) \oplus (M_2 \otimes M_3)$, and so, φ is an isomorphism. ∎

4.3.8. For any $n \in \mathbb{N}$, $M \otimes R^n \cong M^n$. (This follows by induction from 4.3.7 and 4.3.4: $M \otimes R^n = M \otimes (R \oplus \cdots \oplus R) \cong (M \otimes R) \oplus \cdots \oplus (M \otimes R) \cong M \oplus \cdots \oplus M = M^n$.)

For any $n_1, n_2 \in \mathbb{N}$, $M_1^{n_1} \otimes M_2^{n_2} \cong (M_1 \otimes M_2)^{n_1 n_2}$. (This follows by induction from 4.3.7.)

4.3.9. For an infinite collection of modules, tensor product also commutes with the direct sum (but not with the direct product!): Let $\{M_\alpha\}_{\alpha\in\Lambda}$ be a collection of modules and N be a module; then $\big(\bigoplus_{\alpha\in\Lambda} M_\alpha\big) \otimes N \cong \bigoplus_{\alpha\in\Lambda} M_\alpha \otimes N$.

Proof. We cannot use induction (at least, not the ordinary one), but the proof in 4.3.7 can just be copied. Define $\beta\colon \big(\bigoplus_{\alpha\in\Lambda} M_\alpha\big) \times N \longrightarrow \bigoplus_{\alpha\in\Lambda} M_\alpha \otimes N$ by $\beta((u_\alpha)_{\alpha\in\Lambda}, v) = (u_\alpha \otimes v)_{\alpha\in\Lambda}$. It is easy to check that β is bilinear, hence it induces a homomorphism $\varphi\colon \big(\bigoplus_{\alpha\in\Lambda} M_\alpha\big) \otimes N \longrightarrow \bigoplus_{\alpha\in\Lambda} M_\alpha \otimes N$ with $\varphi((u_\alpha)_{\alpha\in\Lambda} \otimes v) = (u_\alpha \otimes v)_{\alpha\in\Lambda}$.

On the other hand, for every $\alpha \in \Lambda$ we have the homomorphism $\psi_\alpha\colon M_\alpha \otimes N \longrightarrow \big(\bigoplus_{\alpha\in\Lambda} M_\alpha\big) \otimes N$ defined by $\psi_\alpha(u \otimes v) = (u_\delta)_{\delta\in\Lambda} \otimes v$ with $u_\delta = u$ if $\delta = \alpha$ and 0 otherwise. Hence (by the universal property of the direct sum!) there is a homomorphism $\psi\colon \bigoplus_{\alpha\in\Lambda}(M_\alpha \otimes N) \longrightarrow \big(\bigoplus_{\alpha\in\Lambda} M_\alpha\big) \otimes N$ satisfying, in particular, $\psi\big((u_\alpha \otimes v)_{\alpha\in\Lambda}\big) = \sum_{\alpha\in\Lambda}^{\text{fin}} \varphi_\alpha(u_\alpha \otimes v) = (u_\alpha)_{\alpha\in\Lambda} \otimes v$. Thus, φ and ψ are inverses of each other on the generators, so are inverses of each other, and so, φ is an isomorphism. ∎

4.3.10. If I is an ideal of R, then $(R/I)\otimes M \cong M/(IM)$, where the isomorphism is given by $(a \bmod I)\otimes u \mapsto (au \bmod IM)$. $M/(IM)$ can be considered as an (R/I)-module.

Proof. The mapping $\beta\colon (R/I)\times M \longrightarrow M/(IM)$, $\beta(a \bmod I, u) = au \bmod IM$, is well defined, since if $a = b \bmod I$ then $au = bu \bmod IM$. β is bilinear, so defines a homomorphism $\varphi\colon (R/I)\otimes M \longrightarrow M/IM$ with $\varphi((a \bmod I)\otimes u) = au \bmod IM$ for all $a \in R$, $u \in M$.

Let us define a homomorphism $\widetilde{\psi}\colon M \longrightarrow (R/I)\otimes M$ by $\widetilde{\psi}(u) = (1 \bmod I)\otimes u$. For any $a \in I$, $u \in M$ we have $\widetilde{\psi}(au) = (1 \bmod I)\otimes(au) = (a \bmod I)\otimes u = 0$, so $\widetilde{\psi}(IM) = 0$, and $\widetilde{\psi}$ factorizes to a homomorphism $\psi\colon M/IM \longrightarrow (R/I)\otimes M$. We can check that ψ is the inverse of φ on simple tensors: $\psi(\varphi((a \bmod I)\otimes u)) = \psi(au \bmod IM) = (1 \bmod I)\otimes au = (a \bmod I)\otimes u$, and $\varphi(\psi(u \bmod IM)) = \varphi((1 \bmod I)\otimes u) = u \bmod IM$ for all $a \in R$, $u \in M$. Hence, ψ is the inverse of φ, and φ is an isomorphism. ■

4.3.11. For any $n, n \in \mathbb{N}$, $\mathbb{Z}_n \otimes_{\mathbb{Z}} \mathbb{Z}_m \cong \mathbb{Z}_d$ where $d = \gcd(n, m)$.

Proof. The bilinear mapping $\mathbb{Z}_n \times \mathbb{Z}_m \longrightarrow \mathbb{Z}_d$, $(a, b) \mapsto ab$, is well defined, thus it induces a homommorphism $\varphi\colon \mathbb{Z}_n \otimes \mathbb{Z}_m \cong \mathbb{Z}_d$ with $\varphi(a\otimes b) = ab \bmod \mathbb{Z}_d$. Clearly, φ is surjective. To construct the inverse mapping, we define $\psi(c) = c\otimes 1$, $c \in \mathbb{Z}$, and check that it factorizes to a homomorphism from $\mathbb{Z}_d$: for this end, we write $d = kn + lm$ and see that $\psi(d) = (kn + lm)\otimes 1 = k(n\otimes 1) + l(1\otimes m) = 0$.

It is however easier to use 4.3.10: $\mathbb{Z}_n \otimes \mathbb{Z}_m = (\mathbb{Z}/(n))\otimes \mathbb{Z}_m \cong \mathbb{Z}_m/((n)\mathbb{Z}_m) = \mathbb{Z}_m/(d\mathbb{Z}_m) \cong \mathbb{Z}_d$. ■

4.3.12. If I and J are ideals in M, then $(R/I)\otimes(R/J) \cong R/(I+J)$. (This can be proved directly or deduced from 4.3.10.)

4.3.13. Let R be an integral domain, let F be the field of fractions of R, and let M be an R-module. Then the kernel of the natural homomorphism $M \longrightarrow F\otimes_R M$, $u \mapsto 1\otimes u$, $u \in M$, is the torsion submodule of M.

Proof. Let $\varphi\colon M \longrightarrow F\otimes M$, $\varphi(u) = 1\otimes u$. It is easy to see that $\mathrm{Tor}(M) \subseteq \ker(\varphi)$. Indeed, if $u \in \mathrm{Tor}(M)$, let $a \neq 0$ be such that $au = 0$, then $\varphi(u) = 1\otimes u = \left(a\frac{1}{a}\right)\otimes u = \frac{1}{a}\otimes(au) = 0$.

Now, assume that $u \in \ker(\varphi)$, that is, $1\otimes u = 0$ in $F\otimes M$. This means that $(1, u)$ is contained in the "relations submodule" $\mathcal{K}$ of $\mathcal{F}_R(F\times M)$, the kernel of the projection $\mathcal{F}_R(F\times M) \longrightarrow F\otimes M$, that is, is a linear combination of elements of $\mathcal{F}_R(F\times M)$ of the form $(q_1 + q_2, v) - (q_1, v) - (q_2, v)$, etc., with $q_i \in F$, $v \in M$. Let d be a common multiple of the denominators of all elements of F involved in this linear combination; then all elements of this linear combination are contained in $\mathcal{F}_R(R\frac{1}{d}\times M)$. This means that $1\otimes u = 0$ in the product $R\frac{1}{d}\otimes M$. We have $R\frac{1}{d} \cong R$, under the isomorphism $a \mapsto da$; so $R\frac{1}{d}\otimes M \cong R\otimes M \cong M$ under the isomorphism defined by $a\otimes v \mapsto da\otimes v \mapsto dav$, $a \in R$, $v \in M$. Hence, since $1\otimes u = 0$ in $R\frac{1}{d}\otimes M$, we have that $du = 0$ in M. So, $u \in \mathrm{Tor}(M)$. ■

4.3.14. For any three R-modules M_1, M_2, and N there is a natural isomorphism $\mathrm{Hom}(M_1\otimes M_2, N) \cong \mathrm{Hom}\big(M_1, \mathrm{Hom}(M_2, N)\big)$.

Indeed, given a homomorphism $\varphi\colon M_1\otimes M_2 \longrightarrow N$, for every $u \in M_1$ we have a homomorphism $\varphi_u\colon M_2 \longrightarrow N$ defined by $\varphi_u(v) = \varphi(u\otimes v)$. This gives a mapping $\psi\colon M_1 \longrightarrow \mathrm{Hom}(M_2, N)$, $u \mapsto \varphi_u$, and it is easy to see that ψ is a homomorphism: for any $u_1, u_2 \in M_1$, $\varphi_{u_1+u_2}(v) = \varphi((u_1+u_2)\otimes v) = \varphi(u_1\otimes v + u_2\otimes v) = \varphi_{u_1}(v) + \varphi_{u_2}(v)$ for all $v \in M_2$, so $\psi(u_1 + u_2) = \psi(u_1) + \psi(u_2)$, and similarly $\psi(au) = a\psi(u)$. So, we have a mapping $\mathrm{Hom}(M_1\otimes M_2, N) \longrightarrow \mathrm{Hom}\big(M_1, \mathrm{Hom}(M_2, N)\big)$, and we can see that this mapping is also a homomorphism: for $\varphi_1, \varphi_2 \in \mathrm{Hom}(M_1\otimes M_2, N)$, $\varphi_1 + \varphi_2$ is mapped to $\psi \in \mathrm{Hom}\big(M_1, \mathrm{Hom}(M_2, N)\big)$ such that, for all $u \in M_1$ and $v \in M_2$, $\psi(u)(v) = (\varphi_1 + \varphi_2)(u\otimes v) = \varphi_1(u\otimes v) + \varphi_2(u\otimes v) = \psi_1(u)(v) + \psi_2(u)(v)$, so $\psi(u) = \psi_1(u) + \psi_2(u)$, so $\psi = \psi_1 + \psi_2$, where ψ_i is the image of φ_i, $i = 1, 2$; similarly, for $\varphi \in \mathrm{Hom}(M_1\otimes M_2, N)$ and $a \in R$, the image of $a\varphi$ is $a\psi$ where ψ is the image of φ in $\mathrm{Hom}\big(M_1, \mathrm{Hom}(M_2, N)\big)$.

The inverse mapping $\mathrm{Hom}\big(M_1, \mathrm{Hom}(M_2, N)\big) \longrightarrow \mathrm{Hom}(M_1\otimes M_2, N)$ is defined in the following way: Given a homomorphism $\psi\colon M_1 \longrightarrow \mathrm{Hom}(M_2, N)$, define a mapping $\beta\colon M_1\times M_2 \longrightarrow N$ by $\beta(u, v) = \psi(u)(v)$. β is bilinear: for any $u, u_1, u_2 \in M_1$, $v, v_1, v_2 \in M_2$ and $a \in R$, $\beta(u_1 + u_2, v) = \psi(u_1 + u_2)(v) = (\psi(u_1) + \psi(u_2))(v) = \psi(u_1)(v) + \psi(u_2)(v) = \beta(u_1, v) + \beta(u_2, v)$ and $\beta(au, v) = \psi(au)(v) = (a\psi(u))(v) = a\psi(u)(v) = a\beta(u)(v)$ since β is a homomorphism, and $\beta(u, v_1 + v_2) = \psi(u)(v_1 + v_2) = \psi(u)(v_1) + \psi(v)(v_2) = \beta(u, v_1) + \beta(u, v_2)$ and $\beta(u, av) = \psi(u)(av) = a\psi(u)(v) = a\beta(u, v)$ since $\beta(u)$ is a homomorphism for every $u \in M_1$. Thus β induces a homomorphism $\varphi\colon M_1\otimes M_2 \longrightarrow N$ with $\varphi(u\otimes v) = \psi(u)(v)$ for all $u \in M_1$ and $v \in M_2$. And clearly, the constructed mapping $\psi \mapsto \varphi$ is the inverse of the homomorphism $\varphi \mapsto \psi$ above: for all $u \in M_1$ and $v \in M_2$ we have $\psi(u)(v) = \varphi(u\otimes v)$ and $\varphi(u\otimes v) = \psi(u)(v)$.

4.4. Extension of scalars

4.4.1. Let M be an R-module and A be an R-algebra. Then the tensor product $A \otimes_R M$ has a structure of an A-module, defined by $\alpha(\beta \otimes u) = (\alpha\beta) \otimes u$. This operation of passing from an R-module M to the A-module $A \otimes_R M$ is called *an extension of scalars*.

4.4.2. Examples. (i) If V is an $\mathbb{R}$-vector space, the $\mathbb{C}$-vector space $\mathbb{C} \otimes_{\mathbb{R}} V$ is called *the complexification* of V. $\mathbb{C} \otimes_R V$ is spanned by tensors of the form $1 \otimes u$ and $i \otimes u$, $u \in V$, and, after identification V with $1 \otimes V$, can be written as $V \oplus iV$.

(ii) If M is a free R-module and A is an R-algebra, then $A \otimes_R M$ is a free A-module, of the same rank.

(iii) If R is an integral domain and F is its field of fractions, then for any R-module M, $F \otimes_R M$ is an F-vector space.

(iv) 4.3.10 above also gives an example of an extension of scalars: R/I is an R-algebra and $(R/I) \otimes M$ is an (R/I)-module.

4.4.3. If A is a unital R-algebra, the A-module $A \otimes_R M$ with the homomorphism $M \longrightarrow A \otimes_R M$ of R-modules is also a universal repelling object, namely, in the category of A-modules N together with an R-module homomorphism $\varphi\colon M \longrightarrow N$. (Given such a module and a homomorphism, the corresponding homomorphism $A \otimes_R M \longrightarrow N$ is defined by $\alpha \otimes u \mapsto \alpha\varphi(u)$, $\alpha \in A$, $u \in M$.)

4.5. The tensor product of two algebras

4.5.1. If A_1 and A_2 are two R-algebras, then $A_1 \otimes_R A_2$ has a structure of an R-algebra as well, where the multiplication is defined by $(u_1 \otimes u_2)(v_1 \otimes v_2) = (u_1v_1) \otimes (u_2v_2)$.

4.5.2. Examples. (i) If A is an R-algebra, then $A \otimes_R R[x] \cong A[x]$, the algebra of polynomials over A.

(ii) The product $R[x] \otimes R[y]$ is isomorphic to the algebra $R[x, y]$ of polynomials in two variables.

4.6. The tensor product of two homomorphisms

4.6.1. Let $\varphi_1\colon M_1 \longrightarrow N_1$ and $\varphi_2\colon M_2 \longrightarrow N_2$ be two homomorphisms of R-modules. Then a homomorphism $\varphi\colon M_1 \otimes_R M_2 \longrightarrow N_1 \otimes_R N_2$ is defined by $\varphi(u_1 \otimes u_2) = \varphi_1(u_1) \otimes \varphi_2(u_2)$. φ is called *the tensor product* of φ_1 and φ_2 and is denoted by $\varphi_1 \otimes \varphi_2$.

4.6.2. We therefore have a mapping $\mathrm{Hom}(M_1, N_1) \times \mathrm{Hom}(M_2, N_2) \longrightarrow \mathrm{Hom}(M_1 \otimes M_2, N_1 \otimes N_2)$ defined by $(\varphi_1, \varphi_2) \mapsto \varphi_1 \otimes \varphi_2$; this mapping is bilinear, and, hence, defines a homomorphism $\mathrm{Hom}(M_1, N_1) \otimes \mathrm{Hom}(M_2, N_2) \longrightarrow \mathrm{Hom}(M_1 \otimes M_2, N_1 \otimes N_2)$ (which may be neither injective nor surjective).

4.7. The tensor product of several modules

4.7.1. Let $M_1, \ldots, M_k, N$ be R-modules. A mapping $\mu\colon M_1 \times \cdots \times M_k \longrightarrow N$ is said to be *multilinear*, or *polylinear*, if for every $i \in \{1, \ldots, k\}$, every $(u_1, \ldots, u_k) \in M_1 \times \cdots \times M_k$, every $v_i \in M_i$, and every $a \in R$,

$$\mu(u_1, \cdots, u_{i-1}, u_i + v_i, u_{i+1}, \ldots, u_k) = \mu(u_1, \cdots, u_{i-1}, u_i, u_{i+1}, \ldots, u_k) + \mu(u_1, \cdots, u_{i-1}, v_i, u_{i+1}, \ldots, u_k)$$

and

$$\mu(u_1, \cdots, u_{i-1}, au_i, u_{i+1}, \ldots, u_k) = a\mu(u_1, \cdots, u_{i-1}, u_i, u_{i+1}, \ldots, u_k).$$

4.7.2. Given k R-modules $M_1, \ldots, M_k$, the tensor product $M_1 \otimes_R M_2 \otimes_R \cdots \otimes_R M_k$ is defined the same way as in the case $k = 2$, namely, as the universal repelling object in the category of R-modules N together with multilinear mappings $M_1 \times \cdots \times M_k \longrightarrow N$.

4.7.3. It is easy to see that, actually,

$$M_1 \otimes M_2 \otimes \cdots \otimes M_{k-1} \otimes M_k \cong \big(\ldots((M_1 \otimes M_2) \otimes M_3) \cdots \otimes M_{k-1}\big) \otimes M_k,$$

and, in light of 4.3.6, is also isomorphic to the sequential tensor product of these modules performed in any other order.

4.8. The tensor algebra of a module

Tensor multiplication allows to construct an algebra from any module. (And moreover, the best possible, the universal one!)

4.8.1. Let M be an R-module. Put $\mathcal{T}^0(M) = R$, $\mathcal{T}^1(M) = M$, $\mathcal{T}^2(M) = M \otimes M$, and for each $k \in \mathbb{N}$, $\mathcal{T}^k(M) = M^{\otimes k} = \underbrace{M \otimes \cdots \otimes M}_{k}$. The elements of $\mathcal{T}^k(M)$ are called *k-tensors*.

4.8.2. Define

$$\mathcal{T}(M) = R \oplus M \oplus M^{\otimes 2} \oplus M^{\otimes 3} \oplus \cdots = \bigoplus_{k=0}^{\infty} \mathcal{T}^k(M).$$

As an R-module, $\mathcal{T}(M)$ is generated by 1 and the simple tensors $u_1 \otimes \cdots \otimes u_k$, $k \in \mathbb{N}$, $u_1, \ldots, u_k \in M$. The multiplication in $\mathcal{T}(M)$ is just the tensor multiplication $\otimes$:

$$(u_1 \otimes \cdots \otimes u_k) \otimes (v_1 \otimes \cdots \otimes v_l) = u_1 \otimes \cdots \otimes u_k \otimes v_1 \otimes \cdots \otimes v_l.$$

Under this multiplication, $\mathcal{T}(M)$ becomes an R-algebra, called *the tensor algebra* of M.

4.8.3. Examples. (i) The tensor algebra $\mathcal{T}(R)$ is isomorphic to the algebra $R[x]$, under the isomorphism that maps $a_1 \otimes \cdots \otimes a_k$ to $a_1 \cdots a_k x^k$.

(ii) Let $M = \mathcal{F}_R(\{x_1, \ldots, x_d\})$, the free R-module generated by the set $\{x_1, \ldots, x_d\}$. Then $\mathcal{T}(M)$ is the algebra of polynomials over R of *noncommuting* variables $x_1, \ldots, x_d$.

4.8.4. An algebra A which is a direct sum $A = \bigoplus_{k=0}^{\infty} A_k$ of its submodules with the property that for every k and l, $A_k A_k \subseteq A_{k+l}$ is called *graded*. $\mathcal{T}(M)$, with the decomposition $\mathcal{T}(M) = \bigoplus_{k=0}^{\infty} \mathcal{T}^k(M)$, is a graded algebra.

4.8.5. For a given R-module M, the tensor algebra $\mathcal{T}(M)$ is the universal repelling object in the category of unital R-algebras A with an R-module homomorphism $\eta\colon M \longrightarrow A$, where morphisms $(A, \eta)\colon (B, \tau)$ are R-algebra homomorphisms $\varphi\colon A \longrightarrow B$ satisfying $\varphi \circ \eta = \tau$ and $\varphi(1) = 1$. (If $\eta\colon M \longrightarrow A$ is a homomorphism, then the unique homomorphism $\varphi\colon \mathcal{T}(M) \longrightarrow A$ extending η is defined by $\varphi(u_1 \otimes \cdots \otimes u_k) = \eta(u_1) \cdots \eta(u_k)$.)

4.8.6. The tensor algebra $\mathcal{T}$ has a *functorial* property: any homomorphism $\varphi\colon M \longrightarrow N$ extends to an R-algebra homomorphism $\mathcal{T}(M) \longrightarrow \mathcal{T}(N)$, by $u_1 \otimes \cdots \otimes u_k \mapsto \varphi(u_1) \otimes \cdots \cdots \otimes \varphi(u_k)$.

4.9. The symmetric and the exterior algebras of a module

The tensor algebra $\mathcal{T}(M)$ of a module M is, generally speaking, noncommutative; it can be made commutative by proper factorization.

4.9.1. An ideal I in a graded algebra $A = \bigoplus_{k=0}^{\infty} A_k$ is said to be *graded* if $I = \bigoplus_{k=0}^{\infty} (I \cap A_k)$. If I is a two-sided graded ideal in A, then A/I is a graded algebra, with $A/I = \bigoplus_{k=0}^{\infty} (A_k/I_k)$.

4.9.2. Consider the two-sided ideal $\mathcal{C}(M)$ in $\mathcal{T}(M)$ generated by the tensors of the form $u \otimes v - v \otimes u$, $u, v \in M$. The quotient ring $\mathcal{T}(M)/\mathcal{C}(M)$ is called *the symmetric algebra* of M and is denoted by $\mathcal{S}(M)$. It is easy to see that $\mathcal{C}(M)$ is a graded ideal, so $\mathcal{S}(M)$ is a graded algebra, $\mathcal{S}(M) = \bigoplus_{k=0}^{\infty} \mathcal{S}^k(M)$, with $\mathcal{S}^k(M) = \mathcal{T}^k(M)/\big(\mathcal{T}^k(M) \cap \mathcal{C}(M)\big)$. For each k, $\mathcal{S}^k(M)$ is the universal repelling object in the category of modules N with a *symmetric* k-linear mapping $\mu\colon M \longrightarrow N$, that is, satisfying $\mu(u_{\sigma(1)}, \ldots, u_{\sigma(k)}) = \mu(u_1, \ldots, u_k)$, $\sigma \in S_k$.

As a ring, $\mathcal{S}(M)$ is generated by $\mathcal{S}^0(M) = R$ and $\mathcal{S}^1(M) = M$; as an R-module, $\mathcal{S}(M)$ is generated by 1 and simple tensors $u_1 \otimes \cdots \otimes u_k$, $u_i \in M$, for all k (more exactly, by the equivalence classes of these tensors). Since in $\mathcal{S}(M)$, $u \otimes v = v \otimes u$, it is commutative; and it is easy to see that $\mathcal{S}(M)$ is the universal repelling object in the category of commutative unital R-algebras A with R-module homomorphisms $M \longrightarrow A$.

4.9.3. Example. For $M = \mathcal{F}_R(\{x_1, \ldots, x_d\})$, $\mathcal{S}(M) = R[x_1, \ldots, x_d]$.

4.9.4. Let $\mathcal{A}(M)$ be the ideal of $\mathcal{T}(M)$ generated by the tensors of the form $u \otimes u$, $u \in M$. The quotient ring $\mathcal{T}(M)/\mathcal{A}(M)$ is called *the exterior algebra* of M and is denoted by $\Lambda(M)$. It is easy to see that $\mathcal{A}(M)$ is a graded ideal, so $\Lambda(M)$ is a graded algebra, $\Lambda(M) = \bigoplus_{k=0}^{\infty} \Lambda^k(M)$, with $\Lambda^k(M) = \mathcal{T}^k(M)/\big(\mathcal{T}^k(M) \cap \mathcal{A}(M)\big)$. The multiplication on Λ, induced by the multiplication $\otimes$ in $\mathcal{T}(M)$, is denoted by $\wedge$: *the wedge product* of $\omega_1 \in \Lambda^k(M)$ and $\omega_2 \in \Lambda^l(M)$ is $\omega_1 \wedge \omega_2 \in \Lambda^{kl}(M)$. For each k, $\Lambda^k(M)$ is the universal repelling object in the category of modules N with an *alternating* k-linear mapping $\mu\colon M \longrightarrow N$, that is, satisfying $\mu(u_{\sigma(1)}, \dots, u_{\sigma(k)}) = \operatorname{sign}(\sigma)\mu(u_1, \dots, u_k)$, $\sigma \in S_k$.

As an R-module, $\Lambda(M)$ is generated by 1 and the simple tensors $u_1 \wedge \dots \wedge u_k$, $u_i \in M$, for all k. The operation $\wedge$ is "anticommutative": it has the property that $u_2 \wedge u_1 = -u_1 \wedge u_2$ for any $u_1, u_2 \in M$, and for any $k, l \in \mathbb{N}$, $\omega_1 \in \Lambda^k(M)$ and $\omega_2 \in \Lambda^l(M)$, $\omega_2 \wedge \omega_1 = (-1)^{kl}\omega_1 \wedge \omega_2$. Graded algebras with such a multiplication are called *alternating*; $\Lambda(M)$ is, therefore, an alternating algebra, and is, in fact, the universal repelling object in the category of alternating R-algebras A with R-module homomorphisms $M \longrightarrow A$.

4.9.5. Example. A valuable example of an exterior algebra is the algebra of differential forms on, say, an open domain $U \subseteq \mathbb{R}^d$: this is the exterior algebra of the module of (continuous or differentiable) covector fields on U over the ring of (continuous or differentiable) functions on U.

4.9.6. The symmetric and the exterior algebras $\mathcal{S}$ and Λ have a *functorial* property: any homomorphism $\varphi\colon M \longrightarrow N$ extends to R-algebra homomorphisms $\mathcal{S}(M) \longrightarrow \mathcal{S}(N)$ and $\Lambda(M) \longrightarrow \Lambda(N)$.

4.10. Symmetric and alternating tensors

An alternative way of constructing symmetric and alternating tensors is by passing to subalgebras of $\mathcal{T}(M)$ instead of quotient algebras.

4.10.1. For each $k \in \mathbb{N}$, the symmetric group S_k acts on $\mathcal{T}^k(M)$ by permuting the entries of tensors: $\sigma\colon u_1 \otimes \dots \otimes u_k \mapsto u_{\sigma(1)} \otimes \dots u_{\sigma(k)}$. The tensors in $\mathcal{T}^k(M)$ invariant under this action, $\omega \in \mathcal{T}^k(M)$ such that $\sigma(\omega) = \omega$ for all $\sigma \in S_k$, are said to be *symmetric*; they form a submodule of $\mathcal{T}^k(M)$, which I will denote by $\mathcal{ST}^k(M)$. (The difference between, say, $\mathcal{S}^2(M)$ and $\mathcal{ST}^2(M)$ is that, in the first case, we deal with "symmetric" tensors of the form $u_1 \otimes u_2$ where $\otimes$ is assumed to be commutative, so that in this algebra $u_1 \otimes u_2 = u_2 \otimes u_1$; in the second case we deal with "usual" tensors of the the form $u_1 \otimes u_2 + u_2 \otimes u_1$.)

4.10.2. *The symmetrization* Sym_k of a k-tensor is the operation defined by

$$\operatorname{Sym}_k(u_1 \otimes \dots \otimes u_k) = \sum_{\sigma \in S_k} u_{\sigma(1)} \otimes \dots \otimes u_{\sigma(k)};$$

this is a homomorphism from $\mathcal{T}^k(M)$ to $\mathcal{ST}^k(M)$. Sym_k may not be surjective; however, for sure, its image contains the submodule $k!\mathcal{ST}^k(M)$ of $\mathcal{ST}^k(M)$, and so, in the case $k!$ is a unit in R, Sym_k is surjective. The kernel of Sym_k can be shown to be $\mathcal{C}^k(M)$, so, Sym_k induces an isomorphism between "the module of symmetric k-tensors" $\mathcal{S}^k(M)$ and the submodule $\operatorname{Sym}_k(\mathcal{T}^k(M))$ of $\mathcal{ST}^k(M)$.

4.10.3. The submodule $\mathcal{AT}^k(M)$ of *alternating* k-tensors is defined similarly: we say that a tensor $\omega \in \mathcal{T}^k(M)$ is alternating if $\sigma(\omega) = \operatorname{sign}(\sigma)\omega$ for every $\sigma \in S_k$.

4.10.4. *The alternation*, or *the skew-symmetrization* Alt_k of an k-tensor is defined by

$$\operatorname{Alt}_k(u_1 \otimes \dots \otimes u_k) = \sum_{\sigma \in S_k} \operatorname{sign}(\sigma) u_{\sigma(1)} \otimes \dots \otimes u_{\sigma(k)};$$

this is a homomorphism from $\mathcal{T}^k(M)$ to $\mathcal{AT}^k(M)$. Alt_k may not be surjective; however, for sure, its image contains the submodule $k!\mathcal{AT}^k(M)$ of $\mathcal{AT}^k(M)$, and so, in the case $k!$ is a unit in R, Alt_k is surjective. The kernel of Alt_k can be shown to be $\mathcal{A}^k(M)$, so, Alt_k induces an isomorphism between $\Lambda^k(M)$ and the submodule $\operatorname{Alt}_k(\mathcal{T}^k(M))$ of $\mathcal{AT}^k(M)$.

4.11. A word about tensor multiplication of modules over a noncommutative ring

4.11.1. If R is noncommutative, the tensor multiplication of R-modules becomes more complicated. If we multiply two left or two right R-modules, we loose noncommutativity of R: $(abu)\otimes v = (bu)\otimes(av) = u\otimes(bav) = (bau)\otimes v$. If M is a right and N is a left R-module, using, instead of bilinear mappings, the so-called *balanced* mappings, with the property that $\beta(ua, v) = \beta(u, av)$, so that $(ua)\otimes v = u\otimes(av)$, we obtain the product $M\otimes_R N$ which is not an R-module but only an abelian group. But if M is a bimodule and N is a left R-module (or if M is a right R-module and N is a bimodule), then $M\otimes_R N$ gets a structure of a left (respectively, right) R-module, by $a(u\otimes v) = (au)\otimes v$ (respectively, $(u\otimes v)a = u\otimes(va)$).

4.11.2. We may also ignore the R-module structures and multiply R-modules over $\mathbb{Z}$, $M\otimes_{\mathbb{Z}} N$. Multiplying this way two left R-modules we produce a group with two different R-module structures: $a(u\otimes v)$ can be defined as $(au\otimes v)$, or as $u\otimes(av)$. By multiplying over $\mathbb{Z}$ a left and a right R-modules, we get an R-bimodule, with $a(u\otimes v)b = (au)\otimes(vb)$.

---

5. Elements of homological algebra: flat, projective, and injective modules

In this section all modules are over a commutative ring R.

This may look strange, but it happens that, given a submodule N of a module M and some other module K, $N\otimes K$ is no longer a submodule, and is not even isomorphic to a submodule, of the module $M\otimes K$. (For example, $\mathbb{Z}$ is a submodule of $\mathbb{Q}$, but $\mathbb{Q}\otimes_{\mathbb{Z}}\mathbb{Z}_2 = 0$ whereas $\mathbb{Z}\otimes_{\mathbb{Z}}\mathbb{Z}_2\cong\mathbb{Z}_2\neq 0$.) In the language of exact sequences, this means that the exactness of a sequence $0\longrightarrow N\xrightarrow{\varphi} M$ does not imply the exactness of the sequence $0\longrightarrow N\otimes K\xrightarrow{\varphi\otimes\mathrm{Id}_L} M\otimes K$. However, the surjectivity of a homomorphism is preserved by tensor multiplication: if a sequence $M\xrightarrow{\varphi} N\longrightarrow 0$ is exact, then the sequence $M\otimes K\xrightarrow{\varphi\otimes\mathrm{Id}_K} N\otimes K\longrightarrow 0$ is also exact. There is a categorial language to express such behavior.

5.1. Co- and contravariant functors, left and right exact

5.1.1. *A covariant functor* Φ from category $\mathcal{A}$ to category $\mathcal{B}$ assigns an object $\Phi(A)$ in $\mathcal{B}$ to every object A of $\mathcal{A}$ (we could say that Φ is a mapping from $\mathcal{A}$ to $\mathcal{B}$, but, unfortunately, categories are not necessarily sets!) and a morphism $\Phi(\varphi)\colon\Phi(A)\longrightarrow\Phi(B)$ to each morphism $\varphi\colon A\longrightarrow B$ in $\mathcal{A}$, preserving the compositions of morphisms: $\Phi(\psi\circ\varphi) = \Phi(\psi)\circ\Phi(\varphi)$, and sending identity morphisms to identity morphisms: $\Phi(\mathrm{Id}_A) = \mathrm{Id}_{\Phi(A)}$.

5.1.2. *A contravariant functor* Φ from category $\mathcal{A}$ to category $\mathcal{B}$ assigns an object $\Phi(A)$ in $\mathcal{B}$ to every object A of $\mathcal{A}$ and a morphism $\Phi(\varphi)\colon\Phi(B)\longrightarrow\Phi(A)$ to each morphism $\varphi\colon A\longrightarrow B$ in $\mathcal{A}$, reversing the composition: $\Phi(\psi\circ\varphi) = \Phi(\varphi)\circ\Phi(\psi)$, and sending identity morphisms to identity morphisms: $\Phi(\mathrm{Id}_A) = \mathrm{Id}_{\Phi(A)}$.

5.1.3. Here are some example of functors:

(i) The *forgetting* (covariant) functor from the category of groups (or the category of topological spaces; or any other category of sets with a structure) the is functor that assigns to a group G the set G.

(ii) The fundamental group $\pi_1(X)$ is a covariant functor from the category of path-connected topological spaces with a marked point to the category of groups.

(iii) $\mathcal{T}$ is a covariant functor from the category of R modules (where R is a commutative unital ring) to the category of graded unital R-algebras.

(iv) To give an example of a contravariant functor, fix a set Z, and to each set X assign the set $\Phi(X)$ of functions $X\longrightarrow Z$. For a mapping $\varphi\colon X\longrightarrow Y$, $\Phi(\varphi)\colon\Phi(Y)\longrightarrow\Phi(X)$ is defined by $\Phi(\varphi)(f) = f\circ\varphi$.

5.1.4. A covariant functor Φ from a category of modules to a category of modules (actually, from one *abelian* category – a category where kernels and cokernels make sense, – to another abelian category) is said to be *left exact* if for every exact sequence $0\longrightarrow A\xrightarrow{\varphi} B\xrightarrow{\psi} C$ the sequence $0\longrightarrow\Phi(A)\xrightarrow{\Phi(\varphi)}\Phi(B)\xrightarrow{\Phi(\psi)}\Phi(C)$ is also exact; and is said to be *right exact* if for every exact sequence $A\xrightarrow{\varphi} B\xrightarrow{\psi} C\longrightarrow 0$ the sequence $\Phi(A)\xrightarrow{\Phi(\varphi)}\Phi(B)\xrightarrow{\Phi(\psi)}\Phi(C)\longrightarrow 0$ is exact; and is said to be *exact* if it is both left and right exact, that is, transforms short exact sequences to short exact sequences.

5.1.5. A contravariant functor Φ from a category of modules to a category of modules is said to be *left exact* if for every exact sequence $A \xrightarrow{\varphi} B \xrightarrow{\psi} C \longrightarrow 0$ the sequence $0 \longrightarrow \Phi(C) \xrightarrow{\Phi(\psi)} \Phi(B) \xrightarrow{\Phi(\varphi)} \Phi(A)$ is exact, is said to be *right exact* if for every exact sequence $0 \longrightarrow A \xrightarrow{\varphi} B \xrightarrow{\psi} C$ the sequence $\Phi(C) \xrightarrow{\Phi(\psi)} \Phi(B) \xrightarrow{\Phi(\varphi)} \Phi(A) \longrightarrow 0$ is exact, and is said to be *exact* if it is both left and right exact, that is, transforms short exact sequences to short exact sequences.

5.1.6. It follows from 1.14.5 that if a functor is exact, then it converts all exact sequences into exact sequences.

5.2. The functor $\otimes K$ and flat modules

5.2.1. Let K be an R-module; the covariant functor $\otimes K$ from the category of R-modules to itself maps a module M to the module $M \otimes K$ and a homomorphism $\varphi\colon M \longrightarrow N$ to the homomorphism $(\varphi \otimes \mathrm{Id}_K)\colon M \otimes K \longrightarrow N \otimes K$.

5.2.2. Proposition. *For any module K the functor $\otimes K$ is right exact: for every exact sequence $A \xrightarrow{\varphi} B \xrightarrow{\psi} C \longrightarrow 0$ of modules the sequence $A \otimes K \xrightarrow{\varphi\otimes\mathrm{Id}_K} B \otimes K \xrightarrow{\psi\otimes\mathrm{Id}_K} C \otimes K \longrightarrow 0$ is exact as well.*

It is easy to see that $\psi \otimes \mathrm{Id}_K$ is surjective and that $(\psi \otimes \mathrm{Id}_K)\circ(\varphi \otimes \mathrm{Id}_K) = 0$. It is not however clear why $\ker(\psi \otimes \mathrm{Id}_K)$ coincides with the image of $\varphi \otimes \mathrm{Id}_K$; we will prove this in 5.4.8.

5.2.3. The functor $\otimes K$ may not be left exact: for instance, the sequence $0 \longrightarrow \mathbb{Z} \longrightarrow \mathbb{Q}$ is exact, but $0 \longrightarrow \mathbb{Z} \otimes \mathbb{Z}_2 \longrightarrow \mathbb{Q} \otimes \mathbb{Z}_2$ is not. If the functor $\otimes K$ is exact, the module K is said to be *flat*: that is, K is flat iff for every monomorphism $N \xrightarrow{\varphi} M$ the homomorphism $N \otimes K \xrightarrow{\varphi\otimes\mathrm{Id}_K} M \otimes K$ is also injective.

5.2.4. There exists the following flatness criterion:

Proposition. *An R-module K is flat iff for every ideal I of R the natural homomorphism $I \otimes K \longrightarrow K$, $a \otimes u \mapsto au$, is injective (and so, defines an isomorphism $I \otimes K \cong IK$), and iff this is true for every finitely generated ideal of R.*

(I won't prove this criterion.)

5.2.5. Here are some simple facts about flat modules:

(i) R is a flat R module.

(ii) Every free module is flat. (This follows from (i) and (iii).)

(iii) Any direct sum of flat modules is flat. Indeed, if $K = \bigoplus_{\alpha\in\Lambda} K_\alpha$ where each K_α is flat and $\varphi\colon A \longrightarrow B$ is a monomorphism, then $A\otimes K \cong \bigoplus_{\alpha\in\Lambda}(A\otimes K_\alpha)$, $B\otimes K \cong \bigoplus_{\alpha\in\Lambda}(B\otimes K_\alpha)$, and $\varphi\otimes\mathrm{Id}_K$ acts "componentwisely": for $\omega = \sum^{\mathrm{fin}}_{\alpha\in\Lambda} \omega_\alpha \in A \otimes K$ with $\omega_\alpha \in A \otimes K_\alpha$ for all α we have $(\varphi \otimes \mathrm{Id}_K)(\omega) = \sum^{\mathrm{fin}}_{\alpha\in\Lambda}(\varphi \otimes \mathrm{Id}_{K_\alpha})(\omega_\alpha)$ with $(\varphi \otimes \mathrm{Id}_{K_\alpha})(\omega_\alpha) \in B \otimes K_\alpha$ for all α. Thus $\varphi \otimes \mathrm{Id}_K(\omega) = 0$ iff $(\varphi \otimes \mathrm{Id}_{K_\alpha})(\omega_\alpha) = 0$ for all α, iff $\omega_\alpha = 0$ for all α, iff $\omega = 0$.

(iv) Any direct summand of a flat module is flat; in particular, any direct summand of any free module is flat. Indeed, assume that $K = K_1 \oplus K_2$ is flat, and let $\varphi\colon A \longrightarrow B$ be a monomorphism. Then $A \otimes K \cong (A \otimes K_1) \oplus (A \otimes K_2)$ and $B \otimes K \cong (B \otimes K_1) \oplus (B \otimes K_2)$, with $\varphi \otimes \mathrm{Id}_K$ acting componentwise: $\varphi \otimes \mathrm{Id}_K = (\varphi \otimes \mathrm{Id}_{K_1}) \oplus (\varphi \otimes \mathrm{Id}_{K_2})$. Since $\varphi \otimes \mathrm{Id}_K$ is injective, $\varphi \otimes \mathrm{Id}_{K_1}$ and $\varphi \otimes \mathrm{Id}_{K_2}$ are also injective.

(v) If R is an integral domain and F is its field of fractions, then F is a flat R-module. Indeed, let $\varphi\colon A \longrightarrow B$ be a monomorphism, and let $\omega \in A \otimes F$ be such that $(\varphi \otimes \mathrm{Id}_F)(\omega) = 0$. ω can be written in the form $u \otimes \frac{1}{d}$ for some $u \in A$ and nonzero $d \in R$. Then $(\varphi \otimes \mathrm{Id}_F)(\omega) = \varphi(u) \otimes \frac{1}{d}$. Then $d\varphi(u) \otimes \frac{1}{d} = \varphi(u) \otimes 1$. By 4.3.13, $\varphi(u) \otimes 1 = 0$ iff $\varphi(u) \in \mathrm{Tor}(B)$, that is, $\varphi(au) = a\varphi(u) = 0$ for some nonzero $a \in R$. But φ is injective, so $au = 0$, so $\omega = u \otimes \frac{1}{d} = au \otimes \frac{1}{ad} = 0$.

5.2.6. If R is an integral domain, then flat R-modules are torsion-free. Indeed, assume that $u \neq 0$ is a torsion element of an R-module K, let $au = 0$ for $a \neq 0$. Consider the monomorphism $\varphi\colon R \longrightarrow R$ defined by $\varphi(b) = ab$, $b \in R$. Then $(\varphi \otimes \mathrm{Id}_K)(1 \otimes u) = a \otimes u = 1 \otimes (au) = 0$, whereas $1 \otimes u \neq 0$ since it corresponds to u under the isomorphism $R \otimes K \cong K$.

5.2.7. Moreover, if R is a PID, we have:

Proposition. *If R is a PID, then an R-module is flat iff it is torsion-free.*

Proof. (This isn't a fair proof, it is based on the unproved Proposition 5.2.4.) Let K be a torsion free module and let $I = (a)$ be an ideal in R. Every tensor $\omega = I \otimes K$ can be written as $\omega = a \otimes u$ for some $u \in M$. The image of ω in R is then au and is $\neq 0$ unless $a = 0$ or $u = 0$. ■

If R is ID but not a PID, a torsion-free module may not be flat: such is the ideal (x, y) in the ring $F[x, y]$.

5.3. The functor $\mathrm{Hom}(K, \cdot)$ and projective modules

Given an R-module K, there are two more natural functors from the category of R-modules to itself: the covariant functor $\mathrm{Hom}(K, \cdot)$, and the contravariant functor $\mathrm{Hom}(\cdot, K)$.

5.3.1. The functor $\mathrm{Hom}(K, \cdot)$ maps a module M to the module $\mathrm{Hom}(K, M)$ and a homomorphism $\varphi\colon M \longrightarrow N$ to the homomorphism $\mathrm{Hom}(K, M) \longrightarrow \mathrm{Hom}(K, N)$ defined by $\eta \mapsto \varphi\circ\eta$:

$$\begin{array}{ccc} M & \xrightarrow{\varphi} & N \\ & \nwarrow^{\eta} & \uparrow{\scriptstyle \varphi\circ\eta} \\ & & K. \end{array}$$

5.3.2. Proposition. *For any module K the functor $\mathrm{Hom}(K, \cdot)$ is left exact: whenever $0 \longrightarrow A \xrightarrow{\varphi} B \xrightarrow{\psi} C$ is an exact sequence of modules, the sequence $0 \longrightarrow \mathrm{Hom}(K, A) \xrightarrow{\varphi\circ\cdot} \mathrm{Hom}(K, B) \xrightarrow{\psi\circ\cdot} \mathrm{Hom}(K, C)$ is also exact.*

Proof. Since φ is injective, for $\alpha \in \mathrm{Hom}(K, A)$, $\varphi\circ\alpha = 0$ only if $\alpha = 0$, so $\mathrm{Hom}(K, A) \xrightarrow{\varphi\circ\cdot} \mathrm{Hom}(K, B)$ is injective.

$$\begin{array}{ccccccc} 0 \longrightarrow & A & \xrightarrow{\varphi} & B & \xrightarrow{\psi} & C \\ & & \nwarrow^{\alpha} & \uparrow{\scriptstyle \beta} & \nearrow_{\gamma} & \\ & & & K & & \end{array}$$

For $\alpha \in \mathrm{Hom}(K, A)$, we have $(\psi\circ\cdot)(\varphi\circ\cdot)(\alpha) = (\psi\circ\cdot)(\varphi\circ\alpha) = \psi\circ\varphi\circ\alpha$, which is $= 0$ since $\psi\circ\varphi = 0$.

Finally, let $\beta \in \mathrm{Hom}(K, B)$ be such that $\psi\circ\beta = 0$. Then $\beta(K) \subseteq \ker(\psi) = \mathrm{Im}(\varphi)$. Define $\alpha\colon K \longrightarrow A$ by $\alpha(u) = \varphi^{-1}(\beta(u))$ (which is well defined since φ is injective), then $\beta = \varphi\circ\alpha$. ∎

5.3.3. The functor $\mathrm{Hom}(P, \cdot)$ may not be right exact; if it is exact, the module P is said to be *projective*. P is projective iff whenever $\varphi\colon B \longrightarrow C$ is an epimorphism, the corresponding homomorphism $\mathrm{Hom}(P, B) \xrightarrow{\varphi\circ\cdot} \mathrm{Hom}(P, C)$ is surjective too, that is, any homomorphism $\gamma\colon P \longrightarrow C$ "lifts" to a homomorphism $\beta\colon P \longrightarrow B$ such that $\gamma = \varphi\circ\beta$:

$$\begin{array}{ccccc} B & \xrightarrow{\varphi} & C & \longrightarrow & 0. \\ & \nwarrow^{\beta} & \uparrow{\scriptstyle \gamma} & & \\ & & P & & \end{array}$$

5.3.4. The criterion of projectivity is, actually, very simple:

Proposition. *A module P is projective iff P is a direct summand of a free module.*

Proof. Let P be a direct summand of a free module: let $F = P \oplus P'$ be free. To prove that P is projective, consider an epimorphism $\varphi\colon B \longrightarrow C$ and let $\gamma\colon P \longrightarrow C$ be a homomorphism. Extend γ to a homomorphism $F \longrightarrow C$ by putting $\gamma|_{P'} = 0$. Let B a basis in F; for each $u \in B$ choose an element $v_u \in \varphi^{-1}(\gamma(u))$, and define a homomorphism $\beta\colon F \longrightarrow B$ by $\beta(u) = v_u$. Then $\varphi\circ\beta = \gamma$, and in particular $\varphi\circ\beta|_P = \gamma|_P$.

$$\begin{array}{ccccc} B & \xrightarrow{\varphi} & C & \longrightarrow & 0. \\ & \nwarrow^{\beta} & \uparrow{\scriptstyle \gamma} & & \\ & & P \oplus P' = F & & \end{array}$$

Conversely, let P be a projective module. Find a free module F with an epimorphism $\pi\colon F \longrightarrow P$. Since P is projective, the identity mapping $P \longrightarrow P$ lifts to a homomorphism $\beta\colon P \longrightarrow F$, so that $\pi\circ\beta = \mathrm{Id}_P$. This means that β is a section of π, that is, the exact sequence $0 \longrightarrow \ker(\pi) \longrightarrow F \xrightarrow{\pi} P \longrightarrow 0$ splits. Hence, P is a direct summand of F.

$$\begin{array}{ccccc} F & \xrightarrow{\pi} & P & \longrightarrow & 0. \\ & \nwarrow^{\beta} & \uparrow{\scriptstyle \mathrm{Id}_P} & & \\ & & P & & \end{array}$$

∎

5.3.5. Examples. (i) Every free module is projective.

(ii) $\mathbb{Z}_2$ is a projective non-free $\mathbb{Z}_6$-module.

(iii) Here is how one can get an example of a projective non-free module over an integral domain. Let R be an ID and let I, J be proper comaximal ideals of R such that IJ is principal, but I is not (and so, $IJ \cong R$ as a module and I is not a free R-module). Define $\varphi\colon I \oplus J \longrightarrow R$ by $\varphi(u,v) = u+v$; we have $\ker(\varphi) = I \cap J = IJ$. Since R is free, φ splits, thus $I \oplus J \cong R \oplus IJ \cong R^2$. So, I and J are projective. (As a concrete example of the scheme just described, we can take $R = \mathbb{Z}[\sqrt{-5}]$, $I = \big(3, 1+\sqrt{-5}\big)$ and $J = \big(3, 1-\sqrt{-5}\big)$.)

5.4. The functor $\mathrm{Hom}(\cdot, K)$ and injective modules

5.4.1. The functor $\mathrm{Hom}(\cdot, K)$ maps a module M to the module $\mathrm{Hom}(M,K)$ and a homomorphism $\varphi\colon M \longrightarrow N$ to the homomorphism $\mathrm{Hom}(N,K) \longrightarrow \mathrm{Hom}(M,K)$ defined by $\eta \mapsto \eta\circ\varphi$:

$$\begin{array}{ccc} & & K \\ & {\scriptstyle \eta\circ\varphi}\nearrow & \uparrow{\scriptstyle \eta} \\ M & \xrightarrow{\varphi} & N. \end{array}$$

5.4.2. Proposition. *For any module K the functor $\mathrm{Hom}(\cdot, K)$ is left exact: whenever $A \xrightarrow{\varphi} B \xrightarrow{\psi} C \longrightarrow 0$ is an exact sequence of modules, the sequence $0 \longrightarrow \mathrm{Hom}(C,K) \xrightarrow{\cdot\circ\psi} \mathrm{Hom}(B,K) \xrightarrow{\cdot\circ\varphi} \mathrm{Hom}(A,K)$ is also exact.*

Proof. If $\gamma \in \mathrm{Hom}(C,K)$ is such that $\gamma\circ\psi = 0$, then $\gamma = 0$ since ψ is surjective.

$$\begin{array}{ccccccc} & & & & K & & \\ & & {\scriptstyle\alpha}\nearrow & & \uparrow{\scriptstyle\beta} & \nwarrow{\scriptstyle\gamma} & \\ A & \xrightarrow{\varphi} & B & \xrightarrow{\psi} & C & \longrightarrow & 0. \end{array}$$

For $\gamma \in \mathrm{Hom}(C,K)$ we have $(\cdot\circ\varphi)(\gamma\circ\psi) = \gamma\circ\psi\circ\varphi = 0$ since $\psi\circ\varphi = 0$.

Finally, let $\beta \in \mathrm{Hom}(B,K)$ be such that $\beta\circ\varphi = 0$. Then $\ker(\beta) \supseteq \varphi(A) = \ker(\psi)$, so β factorizes to a homomorphism $\gamma\colon C \longrightarrow K$ so that $\beta = \gamma\circ\psi$. ■

5.4.3. A sort of converse of Proposition 5.4.2 (which we will need) is also true:

Proposition. *If a sequence $A \xrightarrow{\varphi} B \xrightarrow{\psi} C \longrightarrow 0$ is such that the sequence $0 \longrightarrow \mathrm{Hom}(C,K) \xrightarrow{\cdot\circ\psi} \mathrm{Hom}(B,K) \xrightarrow{\cdot\circ\varphi} \mathrm{Hom}(A,K)$ is exact for all modules K, then $A \xrightarrow{\varphi} B \xrightarrow{\psi} C \longrightarrow 0$ is exact.*

Proof. The proof is by contraposition. If ψ is not surjective, for the (nonzero) projection homomorphism $\gamma\colon C \longrightarrow C/\psi(B)$ we have $\gamma\circ\psi = 0$, so $\mathrm{Hom}(C,K) \xrightarrow{\cdot\circ\psi} \mathrm{Hom}(B,K)$ is not injective.

If $\psi\circ\varphi \neq 0$, then $\big((\cdot\circ\varphi)\circ(\cdot\circ\psi)\big)(\mathrm{Id}_C) = \psi\circ\varphi \neq 0$, so $(\cdot\circ\varphi)\circ(\cdot\circ\psi) \neq 0$.

Finally, assume that $\varphi(A) \underset{\neq}{\subseteq} \ker(\psi)$. Let β be the factorization homomorphism $B \longrightarrow K = B/\varphi(A)$. Then $(\cdot\circ\varphi)(\beta) = \beta\circ\varphi = 0$, however since $\ker(\psi) \not\subseteq \varphi(A)$, $\beta \neq \gamma\circ\psi$ for any $\gamma \in \mathrm{Hom}(C,K)$. ■

5.4.4. The functor $\mathrm{Hom}(\cdot, K)$ may not be right exact; if it is exact, the module Q is said to be *injective*. Q is injective iff whenever $\varphi\colon A \longrightarrow B$ is a monomorphism, the corresponding homomorphism $\mathrm{Hom}(B,Q) \xrightarrow{\cdot\circ\varphi} \mathrm{Hom}(A,Q)$ is surjective, that is, any homomorphism $\alpha\colon A \longrightarrow Q$ can be extended to a homomorphism $\beta\colon B \longrightarrow Q$, so that $\alpha = \beta\circ\varphi$:

$$\begin{array}{ccccc} & & & & Q \\ & & {\scriptstyle\alpha}\nearrow & & \uparrow{\scriptstyle\beta} \\ 0 & \longrightarrow & A & \xrightarrow{\varphi} & B. \end{array}$$

In other words, Q is injective if for any module B and its submodule A any homomorphism $A \longrightarrow Q$ is extendible to a homomorphism $B \longrightarrow Q$.

5.4.5. A module M is said to be *divisible* if $aM = M$ for every $a \in R$ that is not a zero divisor in R.

Every injective module is divisible. Indeed, let $a \in R$ be not a zero divisor; then the multiplication by a, $\varphi(b) = ab$, is a monomorphism $\varphi\colon R \longrightarrow R$. If Q is an injective module and $u \in Q$, the homomorphism $\alpha\colon R \longrightarrow M$ defined by $\alpha(1) = u$ extends to a homomorphism $\beta\colon R \longrightarrow M$ such that $\beta\circ\varphi = \alpha$. Let $v = \beta(1)$; then $u = \alpha(1) = \beta(\varphi(1)) = \beta(a) = av$.

5.4.6. The converse is true when R is a PID:

Proposition. *If R is a PID then an R-module Q is injective iff it is divisible.*

I won't prove this criterion; the proof requires Zorn's lemma.

5.4.7. We can however prove the following fact:

Proposition. *If R is an integral domain and F is a field that contains R, then F is an injective R-module.*

Proof. Let A be a submodule of a module B and let $\alpha \in \mathrm{Hom}_R(A, F)$. First, assume that B/A is a torsion module. Then α can be extended to a homomorphism $\beta\colon B \longrightarrow F$ in the following way: for $u \in B$, if $a \neq 0$ is such that $au \in A$, put $\beta(u) = \frac{1}{a}\alpha(au)$. β is well defined since if also $bu \in A$, then $\frac{1}{b}\alpha(bu) = \frac{1}{ab}\alpha(abu) = \frac{1}{a}\alpha(au)$. And β is a homomorphism since for any $u, v \in M$ with $au = bv \in A$, $a, b \neq 0$, we have $\beta(u+v) = \frac{1}{ab}\alpha(abu + abv) = \frac{1}{ab}\alpha(abu) + \frac{1}{ab}\alpha(abv) = \beta(u) + \beta(v)$ and for any $c \in R$, $\beta(cu) = \frac{1}{a}\alpha(acu) = \frac{c}{a}\alpha(au) = c\beta(u)$.

Now consider the general situation: $A \subseteq B$ and $\alpha\colon A \longrightarrow F$. Let $B/A = C$, let C' be a maximal free submodule of C, let B' be the preimage of C' in B. We have the exact sequence $0 \longrightarrow A \longrightarrow B' \longrightarrow C' \longrightarrow 0$; since C' is free, this sequence splits, and $B' = A \oplus C''$ for some submodule C'' isomorphic to C'. We may now extend α to a homomorphism $\beta'\colon B' \longrightarrow F$ by putting $\beta'|_{C''} = 0$. Finally, $B/B' \cong C/C'$ is a torsion module, so we may extend β' to $\beta\colon B \longrightarrow F$. ∎

5.4.8. We can now prove that for any module K, the functor $\otimes K$ (or equivalently $K\otimes$) is right exact:

Proof. Let $A \xrightarrow{\varphi} B \xrightarrow{\psi} C \longrightarrow 0$ be an exact sequence. By Proposition 5.4.2, for any module N the sequence

$$0 \longrightarrow \mathrm{Hom}(C, N) \xrightarrow{\cdot\circ\psi} \mathrm{Hom}(B, N) \xrightarrow{\cdot\circ\varphi} \mathrm{Hom}(A, N)$$

is exact. Now, by Proposition 5.3.2, the sequence

$$0 \longrightarrow \mathrm{Hom}(K, \mathrm{Hom}(C, N)) \xrightarrow{\cdot\circ\psi\circ\cdot} \mathrm{Hom}(K, \mathrm{Hom}(B, N)) \xrightarrow{\cdot\circ\varphi\circ\cdot} \mathrm{Hom}(K, \mathrm{Hom}(A, N))$$

is also exact. But $\mathrm{Hom}(K, \mathrm{Hom}(M, N))$ is *functorially* isomorphic to $\mathrm{Hom}(M \otimes K, N)$ (see 5.4.9 below), so that the sequence

$$0 \longrightarrow \mathrm{Hom}(K \otimes C, N) \xrightarrow{\cdot\circ(\mathrm{Id}_K \otimes\psi)} \mathrm{Hom}(K \otimes B, N) \xrightarrow{\cdot\circ(\mathrm{Id}_K \otimes\psi)} \mathrm{Hom}(K \otimes A, N)$$

is exact. Since this is true for any module N, by Proposition 5.4.3, the sequence

$$K \otimes A \xrightarrow{\mathrm{Id}_K \otimes\varphi} K \otimes B \xrightarrow{\mathrm{Id}_K \otimes\psi} K \otimes C \longrightarrow 0$$

is exact. ∎

5.4.9. Let us show that $\mathrm{Hom}(K, \mathrm{Hom}(M, N))$ is, indeed, functorially isomorphic to $\mathrm{Hom}(M \otimes K, N)$ with respect to M, meaning that for any homomorphism $\varphi\colon M \longrightarrow M'$ the induced diagram

$$\begin{array}{ccc} \mathrm{Hom}(K \otimes M', N) & \xrightarrow{\sim} & \mathrm{Hom}\big(K, \mathrm{Hom}(M', N)\big) \\ \downarrow & & \downarrow \\ \mathrm{Hom}(K \otimes M, N) & \xrightarrow{\sim} & \mathrm{Hom}\big(K, \mathrm{Hom}(M, N)\big) \end{array}$$

is commutative. (It is functorial with respect to K and N as well, but let's confine ourselves to M.) Let $\eta \in \mathrm{Hom}(K \otimes M', N)$. The image of η in $\mathrm{Hom}(K \otimes M, N)$ is $\eta\circ(\mathrm{Id}_K \otimes\varphi)$, and then in $\mathrm{Hom}(K, \mathrm{Hom}(M, N))$ is τ defined by $\tau(u)(v) = \eta\circ(\mathrm{Id}_K \otimes\varphi)(u \otimes v) = \eta(u \otimes \varphi(v))$ for all $u \in K$ and $v \in M$.

On the other hand, the image of η in $\mathrm{Hom}(K, \mathrm{Hom}(M', N))$ is τ' defined by $\tau'(u)(v') = \eta(u \otimes v')$ for all $u \in K$ and $v' \in M'$. φ induces a homomorphism $\mathrm{Hom}(M', N) \longrightarrow \mathrm{Hom}(M, N)$ by $\tau \mapsto \tau\circ\varphi$, thus the image of η in $\mathrm{Hom}(K, \mathrm{Hom}(M, N))$ is τ'' defined by $\tau''(u)(v) = (\tau'(u)\circ\varphi)(v) = \tau'(u)(\varphi(v)) = \eta(u \otimes \varphi(v))$ for all $u \in K$ and $v \in M$. Hence, $\tau'' = \tau$.

5.5. Dual modules and homomorphisms; contra- and covariant tensors

5.5.1. For an R-module M, the module $\mathrm{Hom}_R(M,R)$ is called *the dual module* of M and is denoted by M^*. Its elements, homomorphisms $M \longrightarrow R$, are called *covectors*, *linear forms*, or *linear functionals* on M.

If R is noncommutative, for a left module M its dual $M^* = \mathrm{Hom}_R(M,R)$ is also defined, but is a right R-module by $(fa)(u) = f(u)a$, $a \in R$, $f \in M^*$, $u \in M$.

5.5.2. We have

(i) $R^* \cong R$;

(ii) for any R-modules $M_1, \dots, M_n$, $\left(\bigoplus_{i=1}^n M_i\right)^* \cong \bigoplus_{i=1}^n M_i^*$; in particular, for any R-module M and $n \in \mathbb{N}$, $(M^n)^* \cong (M^*)^n$;

(iii) for any family $\{M_\alpha\}_{\alpha\in\Lambda}$ of R-modules, $\left(\bigoplus_{\alpha\in\Lambda} M_\alpha\right)^* \cong \prod_{\alpha\in\Lambda} M_\alpha^*$;

(iv) for any $n \in \mathbb{N}$, $(\mathbb{Z}_n)^* = 0$ as a $\mathbb{Z}$-module (and $\cong \mathbb{Z}_n$ as a $\mathbb{Z}_n$-module);

(v) $\mathbb{Q}^* = 0$ as a $\mathbb{Z}$-module (and $\cong \mathbb{Q}$ as a $\mathbb{Q}$-module).

5.5.3. Given a vector $u \in M$ and a covector $f \in M^*$, they produce the scalar $f(u) \in R$. This operation can be considered as *a pairing* of M^* and M: $M^* \times M \longrightarrow R$, $(f,u) \mapsto f(u)$, and this pairing is linear with respect to M^*: $(f_1+f_2,u) = (f_1,u)+(f_2,u)$ and $(af,u) = a(f,u)$. Thus, elements of M can also be viewed as linear forms on M^*: for $u \in M$ we have a homomorphism $\hat{u}\colon M^* \longrightarrow R$ by defining $\hat{u}(f) = f(u)$; the mapping $u \mapsto \hat{u}$ is a natural homomorphism $M \longrightarrow (M^*)^* = M^{**}$, called *the double duality homomorphism*. The double duality homomorphism is, generally speaking, neither injective nor surjective.

5.5.4. For a set $S \subseteq M$, *the annihilator* of S in M^* is the submodule $\mathrm{Ann}(S) = S^\perp = \{f \in M^* : f(u) = 0 \text{ for all } u \in S\}$ of M^*. If N is the module generated by S, then $\mathrm{Ann}(N) = \mathrm{Ann}(S)$. For any two submodules N_1 and N_2 of M, $\mathrm{Ann}(N_1+N_2) = \mathrm{Ann}(N_1)\cap\mathrm{Ann}(N_2)$ and $\mathrm{Ann}(N_1\cap N_2) \supseteq \mathrm{Ann}(N_1)+\mathrm{Ann}(N_2)$.

5.5.5. The pairing $M^* \times M \longrightarrow R$ is bilinear, thus defines a homomorphism $M^* \otimes M \longrightarrow R$ called *the contraction* of tensors: the contraction of a tensor $\omega = \sum_{i=1}^k a_i f_i \otimes u_i \in M^* \otimes M$ is the scalar $\sum_{i=1}^k a_i f_i(u_i) \in R$.

5.5.6. For two modules M and N we have a natural homomorphism $N \otimes M^* \longrightarrow \mathrm{Hom}(M,N)$, defined in the following way: a tensor $\omega = \sum_{i=1}^k v_i \otimes f_i \in N \otimes M^*$ sends $u \in M$ to $\sum_{i=1}^k f_i(u)v_i \in N$. (This mapping is, actually, nothing else but the contraction of $u \in M$ with the M^*-components of ω.)

5.5.7. From 4.3.14, for any modules M and N we have a natural isomorphism $(M \otimes N)^* \cong \mathrm{Hom}(M, N^*)$.

5.5.8. Any homomorphism $\varphi\colon M \longrightarrow N$ of R-modules induces *the dual homomorphism* $\varphi^*\colon N^* \longrightarrow M^*$ by putting $\varphi^*(f) = f\circ\varphi$, $f \in N^*$:

$$\begin{array}{ccc} & & R \\ & {\scriptstyle\varphi^*(f)}\nearrow & \uparrow{\scriptstyle f} \\ M & \xrightarrow{\varphi} & N. \end{array}$$

In the language of "pairing", this reads as $(\varphi^*(f),u) = (f,\varphi(u))$ for all $u \in M$ and $f \in N^*$.

5.5.9. The dual of the composition of two homomorphisms, $(\psi\circ\varphi)^*$, is equal to the composition of the duals, $\varphi^*\circ\psi^*$. (The operation $\cdot^*$ of taking the dual is a contravariant functor from the category of R-modules to itself.)

5.5.10. If $\varphi\colon M \longrightarrow N$ is an epimorphism of R-modules, the dual homomorphism φ^* is injective. (The duality functor $\cdot^*$ is left exact.)

If N is a submodule of a module M, the dual of the embedding $\varphi\colon N \longrightarrow M$ is the homomorphism $\varphi^*\colon M^* \longrightarrow N^*$, $\varphi^*(f) = f|_N$. We have $\ker(\varphi^*) = \mathrm{Ann}(N)$. φ^* does not have to be surjective; it is surjective if R is an injective R-module.

5.5.11. Tensor products $M^{\otimes k} \otimes (M^*)^{\otimes l}$, of k copies of a module M and l copies of its dual M^* with $k,l \geq 0$, often appear in applications. Elements of such a tensor product are called *k-times contravariant l-times covariant tensors*, or just *(k,l)-tensors* over M.

---

6. Linear algebra: homomorphisms of free modules of finite rank

In this section, R will be assumed to be a commutative unital ring. We will develop a sort of "linear algebra" for free modules of finite rank; it generalizes the conventional linear algebra of finite dimensional vector spaces. (Notice that any vector space is a free module over the corresponding field.)

I will call elements of R-modules *vectors* (and elements of R *scalars*).

6.1. Homomorphisms of free modules of finite rank and matrices

6.1.1. *The standard free module of rank* n is R^n, its elements are n-tuples $(a_1, \dots, a_n)$ of scalars. It is often convenient to write a vector $u = (a_1, \dots, a_n) \in R^n$ as a column, $u = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$.

6.1.2. The vectors $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, $e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, …, $e_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$ form *the standard basis* of R^n. For $u = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix} \in R^n$ we have $u = \sum_{i=1}^n a_i e_i$, that is, $a_1, \dots, a_n$ are *the coordinates* of u in the standard basis.

6.1.3. We have $\mathrm{Hom}(R, R) \cong R$, under the isomorphism that maps Id_R to 1. For any $n, m \in \mathbb{N}$ we therefore have

$$\mathrm{Hom}(R^n, R^m) \cong \mathrm{Hom}(R^n, R)^m \cong \mathrm{Hom}(R, R)^{nm} \cong R^{nm},$$

that is, is a free module. The standard basis in this module consists of homomorphisms $\mathrm{Id}_R \in \mathrm{Hom}(R, R)$ corresponding to distinct components of this sum; these are the homomorphisms $\varphi_{i,j}\colon R^n \longrightarrow R^m$, $i = 1, \dots, m$, $j = 1, \dots, n$, defined by $\varphi_{i,j}(e_k) = \begin{cases} e'_i \text{ if } k = j \\ 0 \text{ otherwise,} \end{cases}$ where $\{e_1, \dots, e_n\}$ is the standard basis in R^n and $\{e'_1, \dots, e'_m\}$ is the standard basis in R^m. Every homomorphism $\varphi\colon R^n \longrightarrow R^m$ is uniquely representable as a linear combination of these homomorphisms, $\varphi = \sum_{\substack{i=1,\dots,m \\ j=1,\dots,n}} a_{i,j}\varphi_{i,j}$, $a_{i,j} \in R$.

6.1.4. The $m \times n$ table $A_\varphi = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix}$ of coordinates of φ in the basis $\{\varphi_{i,j},\ i = 1, \dots, m,\ j = 1 \dots, n\}$ is called *the matrix* of φ; for each j, the j-th column of A_φ is the vector $\varphi(e_j) \in R^m$. Given $u \in R^n$, we write $A_\varphi u$ for $\varphi(u)$ and call this *the product* of the matrix A_φ and the vector u.

Conversely, any $m \times n$ matrix A defines a homomorphism $\varphi\colon R^n \longrightarrow R^m$ by putting $\varphi(e_j)$ to be the j-th column of A, $j = 1, \dots, n$. The one-to-one correspondence $\varphi \mapsto A_\varphi$ between the module $\mathrm{Hom}_R(R^n, R^m)$ and the module $\mathrm{Mat}_{m,n}(R)$ of $m \times n$ matrices over R is a module isomorphism.

6.1.5. The result of the application of a homomorphism $\varphi = \sum_{\substack{i=1,\dots,m \\ j=1,\dots,n}} a_{i,j}\varphi_{i,j}$ to a vector $u = \sum_{j=1}^n b_j e_j \in R^n$ is the vector

$$\varphi(u) = \sum_{\substack{1\le i\le m \\ 1\le j\le n}} a_{i,j}\varphi_{i,j}\big(\textstyle\sum_{k=1}^n b_k e_k\big) = \sum_{\substack{1\le i\le m \\ 1\le j\le n}} a_{i,j}\big(\textstyle\sum_{k=1}^n b_k \varphi_{i,j}(e_k)\big) = \sum_{\substack{1\le i\le m \\ 1\le j,k\le n}} a_{i,j}b_k\varphi_{i,j}(e_k)$$
$$= \sum_{\substack{1\le i\le m \\ 1\le j\le n}} a_{i,j}b_j e'_i = \sum_{i=1}^m \Big(\textstyle\sum_{j=1}^n a_{i,j}b_j\Big)e'_i;$$

that is, the product of the matrix A_φ and the vector u is

$$A_\varphi u = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n a_{1,j}b_j \\ \vdots \\ \sum_{j=1}^n a_{m,j}b_j \end{pmatrix}.$$

6.1.6. Let $\varphi\colon R^n \longrightarrow R^m$ and $\psi\colon R^m \longrightarrow R^k$ be two homomorphisms and let $A_\varphi = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix}$ and $A_\psi = \begin{pmatrix} b_{1,1} & \dots & b_{1,m} \\ \vdots & & \vdots \\ b_{k,1} & \dots & b_{k,m} \end{pmatrix}$ be the corresponding matrices. Then the matrix of the composition $\psi\circ\varphi\colon R^n \longrightarrow R^k$ is the $k \times n$ matrix

$$A_{\psi\circ\varphi} = \begin{pmatrix} \Sigma_{i=1}^m b_{1,i}a_{i,1} & \dots & \Sigma_{i=1}^m b_{1,i}a_{i,n} \\ \vdots & & \vdots \\ \Sigma_{i=1}^m b_{k,i}a_{i,1} & \dots & \Sigma_{i=1}^m b_{k,i}a_{i,n} \end{pmatrix},$$

which is called *the product* of the matrices A_ψ and A_φ and is denoted by $A_\psi A_\varphi$.

6.1.7. In the case $n = m = k$, the matrix multiplication just introduced defines an R-algebra structure on the module $\mathrm{Mat}_{n,n}(R)$, which makes it isomorphic to the algebra $\mathrm{End}_R(R^n)$.

6.1.8. The matrix of the identity mapping $\mathrm{Id}_{\mathbb{R}^n}\colon R^n \longrightarrow \mathbb{R}^n$ is $I_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$, called *the unit* $n \times n$-matrix.

If A and B are $n \times n$-matrices such that $AB = BA = I_n$, then B is said to be *the inverse* of A, and is denoted by A^{-1}. A homomorphism $\varphi\colon R^n \longrightarrow R^n$ is invertible iff A_φ is an invertible matrix, in which case $A_{\varphi^{-1}} = A_\varphi^{-1}$.

6.2. Change of basis and the transition matrix

6.2.1. A module M is free of rank n, that is, $M \cong R^n$, iff M has a basis of cardinality M; every such (an ordered) basis $B = \{u_1, \dots, u_n\}$ defines an isomorphism $M \eta R^n$ which maps every vector $u = \sum_{i=1}^n a_i u_1 \in M$ to the n-tuple $(a_1, \dots, a_n) \in R^n$ of its coordinates with respect to B.

6.2.2. Now let M be a free R-module of rank n and N be a free R-module of rank m, and let $\varphi\colon M \longrightarrow N$ be a homomorphism. Let B be a basis in M, C be a basis in N, and $\eta\colon M \longrightarrow R^n$, $\tau\colon N \longrightarrow R^m$ be the corresponding isomorphisms. Then we have the commutative diagram

$$\begin{array}{ccc} M & \xrightarrow{\varphi} & N \\ {\scriptstyle\eta}\downarrow & & \downarrow{\scriptstyle\tau} \\ R^n & \longrightarrow & R^m, \end{array}$$

which defines a homomorphism $\tau \circ \varphi \circ \eta^{-1}\colon R^n \longrightarrow R^m$. The matrix A of this homomorphism is called *the matrix of φ with respect to the bases B and C*; for each j, the j-th column of A is the vector of coordinates of $\varphi(u_j)$, where u_j is the j-th element of B, with respect to the basis C.

6.2.3. Now let M be a free R-module of rank n, let $B = \{u_1, \dots, u_n\}$ and $B' = \{u'_1, \dots, u'_n\}$ be two bases of M, let η and η' be the corresponding homomorphisms $M \longrightarrow R^n$. We then get the *change of basis* isomorphism $\eta' \circ \eta^{-1}\colon R^n \longrightarrow R^n$:

$$\begin{array}{ccc} & M & \\ {\scriptstyle\eta}\swarrow & & \searrow{\scriptstyle\eta'} \\ R^n & \xrightarrow{\sim} & R^n. \end{array}$$

The matrix P of this isomorphism is called *the transition matrix*; multiplied by the "old" coordinates $(a_1, \dots, a_n)$ (with respect to basis B) of a vector $u \in M$ it gives "the new" coordinates $(a'_1, \dots, a'_n)$ of u (with respect to B')

$$\begin{pmatrix} a'_1 \\ \vdots \\ a'_n \end{pmatrix} = P \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

6.2.4. A transition matrix is always invertible: its inverse P^{-1} is the transition matrix from the basis B' to the basis B. (Often, P^{-1} and P are switched, and it is P^{-1} that is called the transition matrix from B to B'.) On the other hand, any invertible matrix $P \in \mathrm{Mat}_{n,n}(R)$ is a transition matrix for some change of basis; indeed, P defines an automorphism φ of R^n; given an isomorphism $\eta\colon M \longrightarrow R^n$, define $\eta' = \varphi \circ \eta$; then P is the transition matrix from the basis B corresponding to η to the basis B' corresponding to η'. (The columns of P are the coordinates of the elements of B with respect to B'.)

6.2.5. Now let M be a free R-module of rank n and N be a free R-module of rank m, and let $\varphi\colon M \longrightarrow N$ be a homomorphism. Let B' and C' be two other bases in M and N respectively, and let P and Q be the transition matrices from B to B' and from C to C' respectively. Let A be the matrix of φ with respect to the bases B and C and A' be the matrix of φ with respect to the bases B' and C'. Then $A' = QAP^{-1}$.

6.2.6. In particular, if $N = M$, $C = B$, and $C' = B'$, then $Q = P$ and $A' = PAP^{-1}$.

6.2.7. Two $n \times n$ matrices A and A' are said to be *similar* (or *conjugate*) if $A' = PAP^{-1}$ for some invertible $P \in \mathrm{Mat}_{n,n}(R)$. We see that two matrices are similar iff they represent, in (potentially) distinct bases, the same endomorphism of a free R-module of rank n.

6.2.8. Let $\varphi\colon V \longrightarrow W$ be a homomorphism of vector spaces. Choose a basis $B = \{u_1, \dots, u_n\}$ in V such that $\{u_{k+1}, \dots, u_n\}$ is a basis of $\ker(\varphi)$. The vectors $\{\varphi(u_1), \dots, \varphi(u_k)\}$ are linearly independent in W; choose a basis $C = \{v_1, \dots, v_m\}$ in W such that $v_i = \varphi(u_i)$, $i = 1, \dots, k$. Then matrix of φ with respect to the bases B and C is the $m \times n$ matrix $\left(\begin{smallmatrix} I_k & O \\ O & O \end{smallmatrix}\right)$, where I_k is the $k \times k$ unit matrix and O denotes zero matrices of different sizes.

As a corollary, we obtain that for any matrix $A \in \mathrm{Mat}_{m,n}(F)$ over a field F there exist invertible matrices $P \in \mathrm{Mat}_{n,n}(F)$ and $Q \in \mathrm{Mat}_{m,m}(F)$, such that QAP has the form $\left(\begin{smallmatrix} I_k & O \\ O & O \end{smallmatrix}\right)$.

6.3. The dual module of a free module of finite rank

6.3.1. The dual module $(R^n)^*$ of R^n, $n \in \mathbb{N}$, is also isomorphic to R^n. The standard basis in $(R^n)^*$ is $\{f_1, \dots, f_n\}$, where for each i the linear form $f_i\colon R^n \longrightarrow R$ is defined by $f_i(e_j) = \begin{cases} 1 \text{ if } j = i \\ 0 \text{ if } j \neq i. \end{cases}$ The row of coordinates of $f \in M^*$ with respect to this basis is $\big(f(e_1), \dots, f(e_n)\big)$. For every i, f_i is the "reading the ith coordinate" linear form: for $u = (a_1, \dots, a_n)$, $f_i(u) = a_i$.

It is customary to write the coordinates of vectors (elements of R^n) as columns, $\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$, and the coordinates of covectors (elements of $(R^n)^*$) in the dual basis as rows, $(b_1 \dots b_n)$. (And indeed, for a covector f having coordinates $b_1, \dots, b_n$, $(b_1 \dots b_n)$ is the matrix of f as an element of $\operatorname{Hom}(R^n, R)$.) Then the result $f(u)$ of pairing of a covector $f = (b_1 \dots b_n)$ and a vector $u = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ is the matrix product $(b_1 \dots b_n)\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \sum_{i=1}^n a_i b_i$.

6.3.2. The dual module of the free module $\mathcal{F}_R(S)$ generated by a set S is the direct product $\prod_{s \in S} R$; if S is infinite, this module has larger rank than $\mathcal{F}_R(S)$, and does not have to be a free module.

6.3.3. Now let M be a free R-module of rank n and let $B = \{u_1, \dots, u_n\}$ be a basis in M. Then the dual module M^* is also free of rank n, and has a basis $B^* = \{f_1, \dots, f_n\}$ that corresponds to B: for each i, $f_i(u_j) = \begin{cases} 1 \text{ if } j = i \\ 0 \text{ if } j \neq i \end{cases}$. B^* is called *the dual basis* for B. For every i, f_i is the "reading the ith coordinate" linear form: if $u \in M$ has coordinates $(a_1, \dots, a_n)$, then $f_i(u) = a_i$.

For $u \in M$, $u = \sum_{i=1}^n a_i u_i$, and $f \in M^*$, $f = \sum_{i=1}^n b_i f_i$, we have $f(u) = \sum_{i=1}^n a_i b_i = (b_1 \dots b_n)\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$.

6.3.4. For a free module M of finite rank, the double duality homomorphism $M \longrightarrow M^{**}$ is an isomorphism, and for any basis B in M, its dual-of-the-dual basis B^{**} coincides with B under this isomorphism.

6.3.5. If M is a free module of finite rank, then, identifying M^{**} with M, for any subset $S \subseteq M^*$ we have its annihilator submodule $\operatorname{Ann}(S) = S^\perp \subseteq M$.

For any submodule N of M we have $\operatorname{Ann}(\operatorname{Ann}(N)) \supseteq N$.

6.3.6. If V is a vector space over a field F and W is a subspace of V, then the natural homomorphism $V^* \longrightarrow W^*$ is an epimorphism (since F, as an F-module, is injective). Hence, $W^* \cong V^*/\operatorname{Ann}(W)$.

For finite dimensional vector spaces this implies that $\dim \operatorname{Ann}(W) = \dim V - \dim W$.

6.3.7. If V is a finite dimensional vector space and W is a subspace of V, then $\operatorname{Ann}(\operatorname{Ann}(W)) = W$. In this case, for any two subspaces W_1 and W_2 of V we have $\operatorname{Ann}(W_1 + W_2) = \operatorname{Ann}(W_1) \cap \operatorname{Ann}(W_2)$ and $\operatorname{Ann}(W_1 \cap W_2) = \operatorname{Ann}(W_1) + \operatorname{Ann}(W_2)$.

6.3.8. For an $m \times n$ matrix $A = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix}$, *the transpose* of A is the $n \times m$ matrix $A^T = \begin{pmatrix} a_{1,1} & \dots & a_{m,1} \\ \vdots & & \vdots \\ a_{1,n} & \dots & a_{m,n} \end{pmatrix}$.

6.3.9. Let M and N be free modules of finite rank, let B be a basis in M, C be a basis in N, and let B^* and C^* be the dual bases in M^* and N^* for B and C respectively. Let $\varphi\colon M \longrightarrow N$ be a homomorphism and let A_φ be the matrix of φ with respect to the bases B and C. Then the matrix A_{φ^*} of the dual homomorphism $\varphi^*\colon N^* \longrightarrow M^*$ with respect to C^* and B^* is the transpose A_φ^T of A_φ. Indeed, for every i and j, the (i,j)-th entry of A_φ is $(f_i, \varphi(u_j)) = (\varphi^*(f_i), u_j)$, which is the (j,i)-th entry of the matrix A_{φ^*}.

It follows that for any $k \times m$ matrix A and $m \times n$ matrix D, $(AD)^T = D^T A^T$.

6.3.10. If B and B' are two bases in a free module M of finite rank and P is the transition matrix from B to B', then the transition matrix from the dual basis B^* to the dual basis $(B')^*$ in M^* is $(P^T)^{-1}$.

6.4. The rank of modules, submodules, dual modules and homomorphisms

In this section let R be an integral domain and F be the field of fractions of R.

6.4.1. Let V be a finite dimensional vector space; then $\dim V^* = \dim V$ and $V^{**} \cong V$ naturally. If W is a subspace of V, then $\dim V = \dim W + \dim(V/W)$; $\dim W = \dim V$ iff $W = V$; the natural homomorphism $V^* \longrightarrow W^*$ is surjective.

If $\varphi\colon V \longrightarrow W$ is a linear mapping (a homomorphism), the rank of φ, $\operatorname{rank} \varphi$, is defined as $\dim \varphi(V)$, and is equal to $\dim V - \dim(\operatorname{Ker} \varphi)$. Such a homomorphism φ is the composition of an epimorphism $V \longrightarrow \varphi(V)$ and the monomorphism $\varphi(V) \longrightarrow W$; so the dual mapping $\varphi^*\colon W^* \longrightarrow V^*$ is the composition of the

epimorphism $W^* \longrightarrow \varphi(V)^*$ and a monomorphism $\varphi(V)^* \longrightarrow V^*$:

Thus, $\operatorname{rank}\varphi^* = \dim\varphi^*(W^*) = \dim(\varphi(V)^*) = \dim\varphi(V) = \operatorname{rank}\varphi$.

If A is a matrix of φ, then the column rank of A (the dimension of the space spanned by the columns of A) equals the rank of φ; since the matrix of φ^* (with respect to the dual bases) is the transpose A^T of A, we obtain that the row rank of A (the dimension of the space spanned by the rows of A) equals its column rank.

6.4.2. Now, let M be an R-module and $V = F \otimes_R M$, an F-vector space. By 4.3.13, the kernel of the natural homomorphism from M to V is the torsion submodule $\operatorname{Tor}(M)$ of M; thus, $M/\operatorname{Tor}(M)$ is contained in V (more exactly, is naturally isomorphic to an R-submodule of V).

6.4.3. Let M be an R-module and $V = F \otimes_R M$. Let N be a maximal free submodule of M, so that M/N is a torsion module. Since F is a flat R-module, the exact sequence $0 \longrightarrow N \longrightarrow M \longrightarrow M/N \longrightarrow 0$ produces the exact sequence $0 \longrightarrow F \otimes N \longrightarrow F \otimes M \longrightarrow F \otimes (M/N) \longrightarrow 0$; we have $F \otimes N \cong F^n$ where $n = \operatorname{rank} N = \operatorname{rank} M$ and $F \otimes (M/N) = 0$, so $F \otimes M \cong F^n$. (Here n may be infnite.) This proves that $\operatorname{rank}_R M = \dim_F V$, and so, is well defined. This also shows that if B is a maximal linearly independent subset of M, then $1 \otimes B = \{1 \otimes u,\ u \in B\}$ is a basis in V.

6.4.4. Let N be a submodule of a module M. Then we have the exact sequence $0 \longrightarrow N \longrightarrow M \longrightarrow M/N \longrightarrow 0$, and so, the exact sequence $0 \longrightarrow F \otimes N \longrightarrow F \otimes M \longrightarrow F \otimes (M/N) \longrightarrow 0$. Thus the F-vector space $W = F \otimes N$ is a subspace of $V = F \otimes M$, $F \otimes (M/N) \cong V/W$, and $\operatorname{rank} M = \dim V = \dim W + \dim(V/W) = \operatorname{rank} N + \operatorname{rank}(M/N)$.

In particular, if N is a submodule of M with $\operatorname{rank}(N) = \operatorname{rank}(M)$, then M/N is a torsion module.

6.4.5. If a module M is generated by a set S, then $\operatorname{rank} M \le |S|$; in particular, a finitely generated module has finite rank. (The converse is not true, as the example $R = \mathbb{Z}$, $M = \mathbb{Q}$ shows.)

6.4.6. The dual module M^* of a module M is always torsion-free: for $f \in M^*$, if $af = 0$ for some nonzero a, that is, $af(u) = 0$ for all $u \in M$, then, since R has no zero divisors, $f(u) = 0$ for all $u \in M$. So, M^* is contained in $F \otimes_R M^*$.

Also, for any $f \in M^*$ and any $u \in \operatorname{Tor}(M)$ we have $f(u) = 0$, so $M^* = (M/\operatorname{Tor}(M))^*$.

6.4.7. Let M be an R-module, $V = F \otimes_R M$, $M^* = \operatorname{Hom}_R(M, R)$ and $V^* = \operatorname{Hom}_F(V, F)$. There is a natural R-module monomorphism $M^* \longrightarrow V^*$, $f \mapsto \tilde{f}$ where $\tilde{f}(\alpha \otimes u) = \alpha f(u)$, $\alpha \in F$, $u \in M$; it extends to an F-vector space monomorphism $\eta\colon F \otimes_R M^* \longrightarrow V^*$. If M has finite rank, this implies that $\operatorname{rank}_R M^* = \dim_F(F \otimes_R M^*) \le \dim_F V^* = \dim_F V = \operatorname{rank}_R M$.

If M is finitely generated, the image of M^* spans V^*, and η is an isomorphism, $F \otimes_R M^* \cong V^*$. (This may not be so if M is not finitely generated: consider $R = \mathbb{Z}$ and $M = F = V = \mathbb{Q}$.) So, in this case, $\operatorname{rank}_R M^* = \dim_F V^* = \dim_F V = \operatorname{rank}_R M$.

6.4.8. Every nonzero element of V defines a nonzero element of V^{**}, thus, if M is finitely generated, by 6.4.7, every element of $M \setminus \operatorname{Tor}(M)$ defines a nonzero element of M^{**}. Hence, the kernel of the double duality homomorphism $M \longrightarrow M^{**}$ is $\operatorname{Tor}(M)$, and $\operatorname{rank} M^{**} = \operatorname{rank} M$.

6.4.9. Let N be a submodule of M and let $W = F \otimes_R N$. The natural homomorphism $\pi\colon M^* \longrightarrow N^*$ (the dual of the embedding $N \longrightarrow M$) induces the homomorphism $\mathrm{Id}_F \otimes \pi\colon F \otimes_R M^* \longrightarrow F \otimes_R N^*$. Since $W \subseteq V$, we also have an epimorphism $\tau\colon V^* \longrightarrow W^*$ of F-spaces, and get the commutative diagram

$$\begin{array}{ccc} F \otimes_R M^* & \xrightarrow{\mathrm{Id}_F \otimes \pi} & F \otimes_R N^* \\ {\scriptstyle \eta_M}\searrow & & \swarrow{\scriptstyle \eta_N} \\ & V^* \xrightarrow{\ \tau\ } W^*, & \end{array}$$

where η_M and η_N are the monomorphisms described in 6.4.7.

In the case M is finitely generated, η_M is an isomorphism, so $\tau \circ \eta_M$ is surjective, so η_N is an isomorphism too, and $\mathrm{Id}_F \otimes \pi$ is surjective; hence, $\operatorname{coker}(\pi) = N^*/\pi(M^*)$ is a torsion module. Since $\ker\pi = \operatorname{Ann}(N)$, we get that $M^*/\operatorname{Ann}(N)$ is isomorphic to a submodule $N^\star$ of N^* such that $N^*/N^\star$ is a torsion module.

It follows that if N is a submodule of a finitely generated module M, then $\operatorname{rank} N^* = \dim W^* = \dim W = \operatorname{rank} N$, and $\operatorname{rank}\operatorname{Ann}(N) = \operatorname{rank} M^* - \operatorname{rank} N^* = \operatorname{rank} M - \operatorname{rank} N$.

6.4.10. Given an R-module homomorphism $\varphi\colon M \longrightarrow N$, $\operatorname{rank}\varphi(M)$ is called *the rank of* φ and is denoted by $\operatorname{rank}\varphi$. It follows from 6.4.4 that if $\operatorname{rank} M < \infty$, then $\operatorname{rank}\varphi = \operatorname{rank} M - \operatorname{rank}\ker(\varphi)$.

6.4.11. Let $\varphi\colon M \longrightarrow N$ be a homomorphism of R-modules. The dual homomorphism $\varphi^*\colon N^* \longrightarrow M^*$ is the composition of two homomorphisms: $N^* \longrightarrow \varphi(M)^*$ and $\varphi(M)^* \longrightarrow M^*$.

Since $M \longrightarrow \varphi(M)$ is surjective, $\varphi(M)^* \longrightarrow M^*$ is injective. $\varphi(M)$ is a submodule of N; if N is finitely generated, $\operatorname{rank}(\varphi^*(N^*)) = \operatorname{rank}(\varphi(M)^*)$ by 6.4.9, so $\operatorname{rank}\varphi^* = \operatorname{rank}\varphi(M)^*$; in the case M or N is finitely generated, this equals $\operatorname{rank}\varphi(M) = \operatorname{rank}\varphi$. We obtain:

Theorem. *If $\varphi\colon M \longrightarrow N$ is a homomorphism of R-modules and N is finitely generated, then* $\operatorname{rank}\varphi^* = \operatorname{rank}\varphi$.

6.4.12. For a matrix $A \in \operatorname{Mat}_{m,n}(R)$ over an integral domain R, *the column space* of A is the submodule of R^m generated by the columns of A; this is the image of the homomorphism, defined by A. *The column rank* of A is the rank of the column space of A; it is equal to the rank of the homomorphism defined by A.

The row space of A is the submodule R^n generated by the rows of A, or, equivalently, it is the column space of the transpose A^T of A; *the row rank* of A is the rank of its row space; it is equal to the rank of the dual of the homomorphism defined by A.

If $\varphi\colon M \longrightarrow N$ is a homomorphism of free modules of finite rank and A_φ is the matrix of φ with respect to some bases in M and N, then, in the corresponding coordinates, $\varphi(M)$ is the column subspace of A_φ, and so, $\operatorname{rank}\varphi$ equals the column rank of A_φ.

6.4.13. By 6.3.9, the row space of a matrix A is the image of the dual φ^* of the homomorphism φ defined by A. By 6.4.11, we get:

Theorem. *For any matrix over an integral domain, its row rank equals its column rank.*

6.5. The tensor product of free modules of finite rank

6.5.1. Let $m, n \in \mathbb{N}$. The tensor product $R^m \otimes R^n$ of free modules R^m and R^n is isomorphic to $(R \otimes R)^{mn} \cong R^{mn}$. The standard basis in $R^m \otimes R^n$ is $\{e_i \otimes e'_j,\ i = 1\ldots, m,\ j = 1, \ldots, n\}$, where $\{e_1, \ldots, e_m\}$ is the standard basis of R^m and $\{e'_1, \ldots, e'_n\}$ is the standard basis of R^n. Every tensors from $R^m \otimes R^n$ has form $\omega = \sum_{i=1,\ldots,n} a_{i,j} e_i \otimes e'_j$, $a_{i,j} \in R$, and its coordinates form the $m \times n$ matrix $\begin{pmatrix} a_{1,1} & \ldots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \ldots & a_{m,n} \end{pmatrix}$.

6.5.2. Now let M and N be free R-modules of ranks m and n respectively, let $B = \{u_1, \ldots, u_m\}$ be a basis in M and $C = \{v_1, \ldots, v_n\}$ be a basis in N. Then the module $M \otimes N$ is free of rank mn, with the basis $B \otimes C = \{u_i \otimes v_j,\ i = 1\ldots, m,\ j = 1, \ldots, n\}$, so that every tensor in $M \otimes N$ has the form $\omega = \sum_{i=1,\ldots,n} a_{i,j} u_i \otimes v_j$, $a_{i,j} \in R$, and its coordinates form the $m \times n$ matrix $\begin{pmatrix} a_{1,1} & \ldots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \ldots & a_{m,n} \end{pmatrix}$.

6.5.3. Similarly, if $M_1, \ldots, M_k$ are free R-modules of ranks $m_1, \ldots, m_k$ respectively, then $M_1 \otimes \cdots \otimes M_k$ is a free module of rank $m_1 \cdots m_k$, and any choice $B_1, \ldots, B_k$ of bases in $M_1, \ldots, M_k$ induces the basis $B_1 \otimes \cdots \otimes B_k$ in $M_1 \otimes \cdots \otimes M_k$; with respect to this basis, the coordinates of a tensor $\omega \in M_1 \otimes \cdots \otimes M_k$ form a k-dimensional $m_1 \times \cdots \times m_k$ matrix $(a_{i_1,\ldots,i_k})$.

6.5.4. In particular, given a free module M of rank n with a basis fixed, the (k,l)-tensors, that is, elements of the tensor product $M^{\otimes k} \otimes (M^*)^{\otimes l}$, are representable by $(k+l)$-dimensional $n \times \cdots \times n$ matrices $\left(a^{i_1,\ldots,i_k}_{j_1,\ldots,j_l}\right)^n_{i_1,\ldots,i_k,j_1,\ldots,j_l=1}$ (it's traditional to use superscripts for the contravariant indices and subscripts for covariant indices), and tensors are often identified with the matrices of their coordinates. (This is just this sort of tensors that appears in differential geometry: the Riemann metric $(g_{i,j})$, the Riemann curvature $(R^i_{j,k,l})$, or the Christoffel symbols $(\Gamma^i_{j,k})$.)

6.5.5. In old fasioned books tensors are defined as multidimensional tables $(a_{i_1,\dots,i_k})$ that "change in certain way when coordinates are changed". How do they change? If $P=(b_{i,j})$ is the transition matrix from old to new coordinates in the r-th component M_r of $M_1\otimes\dots\otimes M_k$, then a tensor $\omega=(a_{i_1,\dots,i_k})$ of this product transforms to $(a'_{i_1,\dots,i_k})=\bigl(\sum_{j_r} b_{i_r,j}a_{i_1,\dots,i_{r-1},j,i_{r+1},\dots,i_k}\bigr)$ (the result of contraction of the corresponding components of $P\otimes\omega=(b_{i,j}a_{i_1,\dots,i_k})\in M_r\otimes M_r^*\otimes M_1\otimes\dots\otimes M_k$).

In the case tensor ω is from the product $M^{\otimes k}\otimes(M^*)^{\otimes l}$, where change of coordinates happens simultaneously in all factors, one has to apply P to contravariant (vector) components of ω and $(P^T)^{-1}$ to its covariant (covector) components. For example, for $\omega\in\operatorname{End}(M)=M\otimes M^*$ represented by a matrix $A=(a_{i_1,i_2})$, if $P=(b_{i,j})$ and $(P^T)^{-1}=(c_{i,j})$, the transformation of A is given by $\bigl(\sum_{j_1,j_2} b_{i_1,j_i}c_{i_2,j_2}a_{j_1,j_2}\bigr)$ (which is just the matrix product $PA((P^T)^{-1})^T=PAP^{-1}$).

6.5.6. For a free module M of rank n with a basis $\{u_1,\dots,u_n\}$ and dual basis $\{f_1,\dots,f_n\}$ in M^*, the elements of the tensor product $M^*\otimes M$, which have form $\omega=\sum_{i,j=1}^n a_{i,j}f_i\otimes u_j$, are represented by $n\times n$-matrices $(a_{i,j})_{i,j=1}^n$ (or rather $(a_i^j)_{i,j=1}^n$).

The result of contraction (see 5.5.5 above) of a tensor $\sum_{i,j=1}^n a_{i,j}f_i\otimes u_j\in M^*\otimes M$ is $\sum_{i,j=1}^n a_{i,j}f_i(u_j)=\sum_{i=1}^n a_{i,i}$, that is, is the trace of the corresponding matrix.

6.5.7. For any free modules M_1 and M_2 of finite rank, we have the natural isomorphism $(M_1\otimes M_2)^*\cong M_1^*\otimes M_2^*$; indeed, the natural homomorphism $\operatorname{Hom}(M_1,R)\otimes\operatorname{Hom}(M_2,R)\longrightarrow\operatorname{Hom}(M_1\otimes M_2,R\otimes R)$ is an isomorphism.

6.6. Homomorphisms and multilinear forms as tensors

6.6.1. Let M and N be free R-modules of ranks n and m respectively, then the natural homomorphism $N\otimes M^*\longrightarrow\operatorname{Hom}(M,N)$ (see 5.5.6 above) is an isomorphism; moreover, if bases $B=\{u_1,\dots,u_n\}$ in M and $C=\{v_1,\dots,v_m\}$ in N are chosen and $B^*=\{f_1,\dots,f_n\}$ is the dual of B basis in M^*, then under this isomorphism, for any i and j, the basis tensor $v_i\otimes f_j$ corresponds to the basis homomorphism $\varphi_{i,j}$ and the matrix of coordinates of a tensor $\omega\in N\otimes M^*$ with respect to the basis $C\otimes B^*$ is just the matrix of the corresponding homomorphism with respect to the bases B and C. So, in this case, homomorphisms $M\longrightarrow N$ can interpreted as tensors from $N\otimes M^*$.

6.6.2. Let $\varphi\colon M\longrightarrow N$ and $\psi\colon N\longrightarrow K$ be homomorphisms of free modules of finite rank, considered as tensors from $N\otimes M^*$ and $K\otimes N^*$ respectively. Then the composition $\psi\circ\varphi\colon M\longrightarrow K$ is the contraction of the $N^*\otimes N$-components of the tensor $\psi\otimes\varphi\in K\otimes N^*\otimes N\otimes M^*$ (which produces a tensor in $K\otimes M^*$).

6.6.3. Let M and N be free modules of finite rank, let $\varphi\colon M\longrightarrow N$ be a homomorphism, and let $\varphi^*\colon N^*\longrightarrow M^*$ be the dual homomorphism. Then, as tensors, $\varphi\in N\otimes M^*$ and $\varphi^*\in M^*\otimes N^{**}\cong M^*\otimes N$ are obtained from each other simply by transposing the factors, $v\otimes f\mapsto f\otimes v$. (This explains why the matrix of φ^* is the transpose of the matrix of φ.)

6.6.4. Let φ be an endomorphism of a free module M of finite rank. As noticed in 6.5.6, the trace of the matrix of φ, in any basis, is the contraction of the corresponding tensor from $M\otimes M^*$. (This proves the fact that similar matrices have the same trace.) We define *the trace* of φ as the result of this contraction (that is, as the trace of the matrix of φ in any basis).

6.6.5. For any modules M_1, M_2, and N, bilinear mappings $M_1\times M_2\longrightarrow N$ are in a one-to-one correspondence with homomorphisms $M_1\otimes M_2\longrightarrow N$. If all these modules are free of finite rank, then $\operatorname{Hom}\bigl(M_1\otimes M_2,N\bigr)\cong N\otimes M_1^*\otimes M_2^*$. and we have an isomorphism between this module and the module $\operatorname{Bil}_R(M_1\times M_2,N)$ of bilinear mappings $M_1\times M_2\longrightarrow N$. (The result of the application of a tensor $v\otimes f_1\otimes f_2\in N\otimes M_1^*\otimes M_2^*$ to a pair $(u_1,u_2)\in M_1\times M_2$ is the vector $f_1(u_1)f_2(u_2)v\in N$.)

6.6.6. In particular, for any modules M_1 and M_2. *the bilinear forms* on $M_1\times M_2$, that is, bilinear mappings $M_1\times M_2\longrightarrow R$, are naturally identified with homomorphisms $M_1\otimes M_2\longrightarrow R$, that is, elements of $(M_1\otimes M_2)^*$. Such a form β defines *a pairing* of M_1 and M_2: for vectors $u\in M_1$ and $v\in M_2$ we get the scalar $\beta(u,v)\in R$, homomorphisms $\varphi\colon M_1\longrightarrow M_2^*$ and $\psi\colon M_1\longrightarrow M_2^*$ by $\varphi(u)(v)=\psi(v)(u)=\beta(u,v)$, $u\in M_1$, $v\in M_2$. The pairing is said to be *perfect* if these homomorphisms are isomorphisms. Thus, a perfect pairing of M_1 and M_2 identifies M_1 with M_2^* and M_2 with M_1^*.

In the case M_1 and M_2 are free of finite rank, the bilinear forms $M_1\times M_2$ can be seen as tensors from $M_1^*\otimes M_2^*$.

6.6.7. A bilinear form on a moudle M is a bilinear mapping $M \times M \longrightarrow R$; in the case M is free of finite rank, it is an elements of $M^* \otimes M^* = (M^*)^{\otimes 2}$. A bilinear form β is *symmetric*, $\beta(u_1, u_2) = \beta(u_2, u_1)$ for all $u_1, u_2 \in M$, iff the corresponding tensor is symmetric, and is *alternating*, $\beta(u_1, u_2) = -\beta(u_2, u_1)$ for all $u_1, u_2 \in M$, iff the corresponding tensor is alternating.

If a bilinear form on M is such that the corresponding pairing $M \times M \longrightarrow R$ is perfect, it defines an isomorphism $M \longrightarrow M^*$, that is, allows us to identify vectors (elements of M) and covectors (elements of M^*).

6.7. The tensor algebras of free modules of finite rank

Let M be a free R-module of rank n, with a basis $B = \{u_1, \dots, u_n\}$.

6.7.1. For any $k \in \mathbb{N}$, the tensor power $\mathcal{T}^k(M) = M^{\otimes k}$ is a free module of rank n^k, with basis (induced by B) $\{u_{i_1} \otimes \dots \otimes u_{i_k},\ 1 \le i_1, \dots, i_k \le n\}$. The tensor algebra $\mathcal{T}(M)$ is therefore a free R-module (of infinite rank).

6.7.2. The symmetric algebra $\mathcal{S}(M)$ of M is also a free R-module of infinite rank; for each $k \in \mathbb{N}$, a basis of the symmetric k-power $\mathcal{S}^k(M)$ is $\{u_{i_1} \otimes \dots \otimes u_{i_k},\ 1 \le i_1 \le \dots \le i_k \le n\}$; the rank of $\mathcal{S}^k$ is $\binom{k+n-1}{k}$.

6.7.3. The exterior algebra $\Lambda(M)$ of M is also a free R-module, but of finite rank. For each $k \in \mathbb{N}$, a basis of the exterior k-power $\Lambda^k(M)$ is $\{u_{i_1} \wedge \dots \wedge u_{i_k},\ 1 \le i_1 < \dots < i_k \le n\}$; the rank of Λ^k is therefore equal to $\binom{n}{k}$ for $k \le n$, and is equal to 0 for $k > n$.

6.7.4. The senior wedge power $\Lambda^n(M)$ of M is a free R-module of rank 1, that is, is isomorphic to R. It is generated by the single element $u_1 \wedge \dots \wedge u_n$.

6.7.5. For any modules M_1, M_2 and any k, a homomorphism $\varphi\colon M_1 \longrightarrow M_2$ induces a homomorphism $\wedge^k\varphi\colon \Lambda^k M_1 \longrightarrow \Lambda^k M_2$ defined by $\wedge^k\varphi\big(u_1 \wedge \dots \wedge u_k\big) = \varphi(u_1) \wedge \dots \wedge \varphi(u_k)$. For a composition $\psi\circ\varphi$ of two homomorphisms, $\wedge^k(\psi\circ\varphi) = \wedge^k\psi \circ \wedge^k\varphi$.

If φ is surjective, then $\wedge^k\varphi$ is surjective too. If φ is injective, $\wedge^k\varphi$ may not be injective; but it is if R is an integral domain and M_1, M_2 are free. (This fact is easy to check for vector spaces, and then M_1 and M_2 can be seen as submodules of $F \otimes_R M_1$ and $F \otimes_R M_2$ respectively, where F is the field of fractions of R.)

6.8. The determinant of endomorphisms of free modules of finite rank

Let R be a commutative unital ring and let M be a free R-module of rank n, with a basis $B = \{u_1, \dots, u_n\}$.

6.8.1. From 6.7.5 we deduce:

Proposition. *If R is an integral domain, then for any k, a set $\{v_1, \dots, v_k\}$ is linearly independent in M iff $v_1 \wedge \dots \wedge v_k \neq 0$.*

6.8.2. Let φ be an endomorphism of M. Then $\wedge^n\varphi$ is an endomorphism of $\Lambda^n(M) \cong R$, so it is defined by a multiplication by a scalar $d \in R$: for any $\omega \in \Lambda^m(M)$, $\wedge^n\varphi(\omega) = d\omega$; in particular, if $\{u_1, \dots, u_n\}$ is a basis of M, then

$$\wedge^n\varphi\big(u_1 \wedge \dots \wedge u_n\big) = \varphi(u_1) \wedge \dots \wedge \varphi(u_k) = d(u_1 \wedge \dots \wedge u_n).$$

This scalar d is called *the determinant* of φ and is denoted by $\det\varphi$.

If $A = (a_{i,j})_{i,j=1}^n$ is the matrix of φ, then $\det\varphi = \sum_{\sigma\in S_n} \operatorname{sign}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)}$. This sum is called *the determinant of* A, $\det A$; that is, the determinant of a matrix A is the determinant of the homomorphism defined by this matrix.

6.8.3. From the definition of det we have:

(i) $\det\varphi$ only depends on φ, not on the basis in M. It follows that similar matrices have the same determinant.

(ii) $\det \mathrm{Id}_M = 1$.

(iii) $\det(c\varphi) = c^n \det\varphi$ for any $c \in \mathbb{R}$.

(iv) For any $\varphi, \psi \in \mathrm{End}(M)$, $\det(\varphi\circ\psi) = \det\varphi\cdot\det\psi$; for two $n\times n$ matrices A and B, $\det(AB) = \det A \det B$.

(v) If R is an integral domain, $\det\varphi \neq 0$ iff $\operatorname{rank}\varphi = n$; for a square matrix A, $\det A \neq 0$ iff the columns of A are linearly independent.

(vi) An endomorphism φ of a free module of finite rank is invertible iff $\det\varphi$ is a unit in R; a square matrix A is invertible iff $\det A$ is a unit in R.

(vii) $\det\varphi^* = \det\varphi$; for a square matrix A, $\det A^T = \det A$.

6.8.4. Not all statements in 6.8.3 are evident. (v) follows from Proposition 6.8.1.

(vii) can be established in the following way: Let $\{u_1, \dots, u_n\}$ be a basis in a (free) module M and $\{f_1, \dots, f_n\}$ be the dual basis in M^*. There is a natural pairing of $\Lambda^n M$ and $\Lambda^n(M^*)$, which identifies $\Lambda^n(M^*)$ with $(\Lambda^n(M))^*$ and satisfies $f_1 \wedge \dots \wedge f_n(u_1 \wedge \dots \wedge u_n) = 1$. Thus for any $\varphi \in \mathrm{End}(M)$ we can write

$$\det \varphi = (f_1 \wedge \dots \wedge f_n)\big(\wedge^n \varphi(u_1 \wedge \dots \wedge u_n)\big) = (u_1 \wedge \dots \wedge u_n)\big((\wedge^n \varphi)^*(f_1 \wedge \dots \wedge f_n)\big).$$

It can be shown that the operation $\wedge^n$ commutes with the operation of "dualization", $(\wedge^n \varphi)^* = \wedge^n(\varphi^*)$, thus,

$$(u_1 \wedge \dots \wedge u_n)\big((\wedge^n \varphi)^*(f_1 \wedge \dots \wedge f_n)\big) = (u_1 \wedge \dots \wedge u_n)\big((\wedge^n \varphi^*)(f_1 \wedge \dots \wedge f_n)\big) = \det \varphi^*.$$

So, for any endomorphism φ of a free module of finite rank, $\det \varphi^* = \det \varphi$.

6.8.5. The "if" part of 6.8.3(vi) is also not obvious. We have a natural *pairing* of the modules M and $\Lambda^{n-1}(M)$, – a bilinear mapping $M \times \Lambda^{n-1}(M) \longrightarrow \Lambda^n(M) \cong R$, defined by $(u, \omega) \mapsto u \wedge \omega$. This pairing defines an homomorphism $M \longrightarrow \mathrm{Hom}\big(\Lambda^{n-1}(M), \Lambda^n(M)\big)$, by $u(\omega) = u \wedge \omega$, which is an isomorphism (the pairing is *perfect*).

Let $\varphi \in \mathrm{End}(M)$, and let $\psi \in \mathrm{End}(M)$ be "the dual" of $\wedge^{n-1}\varphi \in \mathrm{End}(\Lambda^{n-1}(M))$ in the above sense: $\psi(u) \wedge \omega = u \wedge (\wedge^{n-1}\varphi(\omega))$ for all $u \in M$ and $\omega \in \Lambda^{n-1}(M)$. ($\psi$ is *the adjoint* homomorphism of $\wedge^{n-1}\varphi$ with respect to the pairing above.) Then for any $u \in M$ and $\omega \in \Lambda^{n-1}(M)$ we have

$$\psi(\varphi(u)) \wedge \omega = \varphi(u) \wedge (\wedge^{n-1}\varphi(\omega)) = \wedge^n \varphi(u \wedge \omega) = (\det \varphi) u \wedge \omega.$$

Since the pairing is perfect, this implies that $\psi(\varphi(u)) = (\det \varphi)u$ for all $u \in M$, that is, $\psi \circ \varphi = \det \varphi \cdot \mathrm{Id}_M$.

It follows that if $d = \det \varphi$ is a unit in R, then $d^{-1}\psi$ is the left inverse of φ. Hence, in the monoid of endomorphisms of M with invertible determinant, each element has a left inverse; this implies that this monoid is, actually, a group, and $d^{-1}\psi = \varphi^{-1}$. (We can also obtain a formula for the matrix of φ^{-1} in terms of the matrix of φ.)

6.8.6. The following operations on a matrix are called *elementary column operations*:

(i) switching two columns;

(ii) multiplying a column by a scalar;

(iii) adding a multiple of one column to another column.

Each of these operations on a matrix A can be performed by multiplying A from the left by an invertible matrix.

Elementary row operations are defined accordingly; they can be performed by multiplying A by a invertible matrices from the right.

6.8.7. Let A be a square matrix over R. The column operations affect the determinant of A the following way:

(i) If matrix A' is obtained from A by switching two columns, then $\det A' = -\det A$;

(ii) if matrix A' is obtained from A by multiplying a column by a scalar c, then $\det A' = c \det A$;

(iii) if matrix A' is obtained from A by adding a multiple of one of its columns to another, then $\det A' = \det A$.

Since $\det A = \det A^T$, in (i)-(iii), "columns" can be replaced by "rows".

---

7. The theory of finitely generated modules over PIDs and normal forms of matrices

A commutative unital ring is said to be *a Principal Ideal Domain*, or *a PID*, if R is an integral domain such that every ideal in R is principal (generated by a single element). Every ED (an Euclidean domain) is a PID. Examples of PIDs (and EDs) are $\mathbb{Z}$ and $F[x]$ – the ring of polynomials over a field F.

7.1. Submodules of a free module of finite rank over a PID

7.1.1. The following theorem is the basic result of this section:

Theorem. *Let R be a PID, let M be a free R-module of rank n, and N be a nonzero submodule of M. Then N is also free, of rank $k \le n$. Moreover, there is a basis $\{u_1, \dots, u_n\}$ of M and scalars $a_1, \dots, a_k \in R$ with $a_1 \mid a_2 \mid \dots \mid a_k$ such that $\{a_1 u_1, \dots, a_k u_k\}$ is a basis in N.*

We will also see below that, though the basis $\{u_1, \dots, u_n\}$ is not defined uniquely, the scalars $a_1, \dots, a_k$ are defined uniquely (up to association – multiplication by units in R, of course).

Proof. We will use the double duality isomorphism and identify M^{**} with M.

Recall that, being a PID, R is a Noetherian ring, which implies that every family of ideals in R has a maximal element, – an ideal not contained in any other ideal of that family.

We will use induction on n. For $n = 1$, we have $M \cong R$, and we may assume that $M = R$. Then N is an ideal in R; since R is a PID, there is an element $a_1 \in R$, such that $N = (a_1)$. Thus, the basis $\{1\}$ of $M = R$ and the scalar a_1 satisfy the assertion of the theorem.

Now let $n \geq 2$. For every $f \in M^*$, $f(N)$ is an ideal of R. Since R is Noetherian, there exists a linear form $h \in M^*$ such that the ideal $h(N)$ is maximal in the family $\{f(N),\ f \in M^*\}$ of ideals of R; let a_1 be such that $(a_1) = h(N)$. There is f such that $f(N) \neq 0$; so, $a_1 \neq 0$. Since $a_1 \in h(N)$, there exists $v_1 \in N$ such that $a_1 = h(v_1)$.

I claim now that a_1 divides $f(v_1)$ for all $f \in M^*$. Indeed, put $I = v_1(M^*) = \{f(v_1),\ f \in M^*\}$. I is an ideal in R; let $I = (b)$. Then $a_1 = h(v_1) \in I$, so $b \mid a_1$. Let $f \in M^*$ be such that $f(v_1) = b$, then $f(N) \ni b$, so $f(N) \supseteq (a_1)$; but the ideal (a_1) is maximal in the family of ideals of the form $f(N)$, so $f(N) = (a_1)$, so $a_1 \mid b$, so $I = (a_1)$.

Since a_1 divides $v_1(f) = f(v_1)$ for all $f \in M^*$, v_1, as an element of M^{**}, is divisible by a_1. But $M^{**} = M$, so v_1 is divisible by a_1 in M: there exists $u_1 \in M$ such that $v_1 = a_1 u_1$. We then have $h(u_1) = 1$.

Let $M' = \ker(h)$. Then $Ru_1 \cap M' = 0$, and any vector $u \in M$ can be written as $u = h(u)u_1 + (u - h(u)u_1)$, where $h(u)u_1 \in Ru_1$ and $u - h(u)u_1 \in M'$, so $M = Ru_1 \oplus M'$.

Let $N' = M' \cap N$. If $u \in N$, then $a_1 \mid h(u)$, so, in the decomposition $u = h(u)u_1 + (u - h(u)u_1)$, $h(u)u_1 \in a_1 R u_1 = Rv_1 \subseteq N$ and $u - h(u)u_1 \in N'$; so, $N = Rv_1 \oplus N'$, and $\operatorname{rank}(N') = \operatorname{rank}(N) - 1 = k - 1$. We will now use induction on k to prove that N is free: If $k = 1$, then $\operatorname{rank}(N') = 0$; but since M has no torsion, $N' = 0$, so $N = Rv_1$ and is free. If $k \geq 2$, by induction on k, N' is a free submodule of M, so N is also a free submodule of M.

We have proved that every submodule of M is free. So, M' is free; it has rank $n - 1$, and N' is a submodule of M' of rank $k - 1$. By induction on n, there is a basis $\{u_2, \dots, u_n\}$ in M' and scalars $a_2, \dots, a_k \in R$ such that $a_2 \mid \dots \mid a_n$ and $\{a_2 u_2, \dots, a_k u_k\}$ is a basis in N'. Then $\{u_1, u_2, \dots, u_n\}$ is a basis in $M = Ru_1 \oplus M'$, and $\{a_1 u_1 = v_1, a_2 u_2, \dots, a_k u_k\}$ is a basis in $N = Rv_1 \oplus N'$.

It remains to show that $a_1 \mid a_2$. Define $f \in M^*$ by $f(x_1 u_1 + \dots + x_n u_n) = x_1 + x_2$. Then $f(a_1 u_1) = a_1$, so $(a_1) \subseteq f(N)$, so $(a_1) = f(N)$. But $a_2 = f(a_2 u_2) \in f(N)$, so $a_2 \in (a_1)$. ■

7.1.2. If N is a submodule of a free module M of finite rank over a PID, then N may not be a direct summand of M. It however follows from Theorem 7.1.1 that there is a submodule $\widetilde{N}$ of M which is a direct summand of M, contains N, and is such that $\widetilde{N}/N$ is a torsion module. (Namely, $\widetilde{N}$ is the submodule generated by $\{u_1, \dots, u_k\}$.)

If M/N is torsion-free, then $\widetilde{N} = N$, N is a direct summand of M, and there is a basis $\{u_1, \dots, u_n\}$ in M such that $\{u_1, \dots, u_k\}$ is a basis of N.

7.1.3. Let R be a PID and let $\varphi\colon M \longrightarrow N$ be a homomorphism of free R-modules M and N of ranks n and m respectively. Let $K = \ker(\varphi)$. The module M/K is isomorphic to a submodule of N, so has no torsion; hence, K is a direct summand in M, $M = M' \oplus K$. Let $L = \varphi(M)$, then $\varphi|_{M'}$ is an isomorphism between M' and L. Find a basis $\{v_1, \dots, v_m\}$ in N and scalars $a_1, \dots, a_k \in R$ with $a_1 \mid \dots \mid a_k$ such that $\{a_1 v_1, \dots, a_k v_k\}$ is a basis in L. For every i, let $u_i = (\varphi|_{M'})^{-1}(a_i v_i) \in M'$, then $\{u_1, \dots, u_k\}$ is a basis in M'. Choose a basis $\{u_{k+1}, \dots, u_n\}$ in K, then $\{u_1, \dots, u_k, \dots, u_n\}$ is a basis in M. Then the $m \times n$ matrix of φ with respect to the bases $\{u_1, \dots, u_n\}$ in M and $\{v_1, \dots, v_m\}$ in N has form

$$\begin{pmatrix} a_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & & & \\ 0 & 0 & \dots & a_k & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix}. \tag{7.1}$$

7.1.4. We obtain as a corollary that for any matrix $A \in \operatorname{Mat}_{m,n}(R)$ there exist invertible matrices $Q \in \operatorname{Mat}_{m,m}(R)$ and $P \in \operatorname{Mat}_{n,n}(R)$ such that the matrix $A' = QAP^{-1}$ has form (7.1) with $a_1 \mid \dots \mid a_k$. This

matrix A' is called *the Smith normal form* of A. (We will see below that the Smith normal form of a matrix is uniquely defined.)

7.1.5. In the case R is an ED (a Euclidean domain), there is an effective algorithm allowing to find the Smith normal form of any $m \times n$ matrix A over R, which uses the division algorithm in EDs along with the following elementary row/column operations: switching two columns, switching two rows, adding a multiple of a column to another column, and adding a multiple of a row to another row. (Multiplying rows/columns by units is also allowed.) Each of these operations can be performed by a multiplication of A by an invertible $n \times n$ matrix from the right (column operations) or by an invertible $m \times m$ matrix from the left (row operations), and corresponds to a certain elementary change of basis in R^n (column operations) or in R^m (row operations). Thus, the algorithm allows to find the bases in R^n and R^m with respect to which the matrix takes the form (7.1).

Let N be the Euclidean norm on R; but for convenience, let's assume that $N(0) = \infty$. We start with a nonzero $m \times n$ matrix A; the entries of A and of the matrices obtained after each operations will be denoted by $a_{i,j}$.

(i) If there is (i,j) with $N(a_{i,j}) < N(a_{1,1})$, find (i,j) for which $N(a_{i,j})$ is minimal, and switch rows 1 and i and columns 1 and j; else

(ii) If there is j such that $a_{1,1} \nmid a_{1,j}$, write $a_{1,j} = ca_{1,1} + r$ with $N(r) < N(a_{1,1})$, subtract c·(column 1) from column j, and switch columns 1 and j; else

(iii) If there is i such that $a_{1,1} \nmid a_{i,1}$, write $a_{i,1} = ca_{1,1} + r$ with $N(r) < N(a_{1,1})$, subtract c·(row 1) from row i, and switch rows 1 and i; else

(iv) (We are here if all entries in column 1 and in row 1 are divisible by $a_{1,1}$.) If there is (i,j) such that $a_{1,1} \nmid a_{i,j}$, write $a_{1,j} = ba_{1,1}$ and subtract $(b-1)$·(column 1) from column j, write $a_{i,j} = ca_{1,1} + r$ with $N(r) < N(a_{1,1})$, subtract c·(row 1) from row i, and switch rows 1 and i and columns 1 and j; else

(v) (We are here if all entries of the matrix are divisible by $a_{1,1}$.) Subtract a multiple of column 1 from all other columns to get all entries in the first row, except $a_{1,1}$, equal to 0, and subtract a multiple of row 1 from all other rows to get all entries in the first column, except $a_{1,1}$, equal to 0. (If needed, the first row can now be multiplied by a unit, to make $a_{1,1}$ "look better".) If m or $n = 1$, or if the submatrix $(a_{i,j})_{\substack{2 \le i \le m \\ 2 \le j \le n}}$ is zero, stop; otherwise pass to this submatrix.

(vi) Start over.

During this process, each step makes $N(a_{1,1})$ (or the size of the matrix) smaller, so the process terminates after finitely many steps.

7.1.6. If we want to reduce a matrix to the form (7.1) without requiring that $a_1 \mid \ldots \mid a_k$, the algorithm in 7.1.5 can be essentially shortened by removing item (iv) from it.

7.1.7. Let N be a submodule of rank k of a free module M of rank m over a PID R. Find a finite set of generators of N and construct an epimorphism $\varphi\colon R^n \longrightarrow N$. Then the matrix of φ, with respect to some bases in R^n and M, is an $m \times n$ matrix whose columns generate N. Finding bases in R^n and M in which the matrix of φ has form (7.1) is equivalent to finding bases in N and M satisfying the assertion of Theorem 7.1.1.

7.2. The fundamental theorem of finitely generated modules over PIDs; invariant factors and elementary divisors of a module

7.2.1. Theorem I – existence. *Any finitely generated module M over a PID R is a direct sum of cyclic submodules, $M \cong R^l \oplus R/(a_1) \oplus \cdots \oplus R/(a_m)$, where $l = \operatorname{rank} M$ and the nonzero nonunit scalars $a_1, \ldots, a_m \in R$ satisfy $a_1 \mid \cdots \mid a_m$.*

We will see that the scalars $a_1, \ldots, a_m$ are defined uniquely up to association (multiplication by units); they are called *the invariant factors* of M.

Proof. Assume that M is generated by n elements, then $M \cong R^n/N$ for some submodule N of R^n. Using Theorem 7.1.1, find a basis $\{u_1, \ldots, u_n\}$ in R^n and (nonzero) scalars $a_1, \ldots, a_k$ such that $a_1 \mid \cdots \mid a_k$ and $\{a_1u_1, \ldots, a_ku_k\}$ is a basis in N. Then $M \cong R^n/N \cong R/(a_1) \oplus \cdots \oplus R/(a_k) \oplus R^{n-k}$. If, for some i, a_i is a unit, then $R/(a_i) = 0$ and can be removed from this sum, and we get $M \cong R/(a_{i_1}) \oplus \cdots \oplus R/(a_{i_m}) \oplus R^{n-k}$ where all a_{i_j} are already non-units. ■

7.2.2. It follows that any finitely generated module over a PID is a direct sum of a free submodule and the torsion submodule. It then follows that if such a module is torsion-free, then it is free.

7.2.3. Let R be a PID. Let a be a nonzero nonunit element of R, and let $a = p_1^{r_1} \cdots p_k^{r_k}$, be the factorization of a where $p_1, \dots, p_k$ are distinct irreducible elements of R and $r_1, \dots, r_k \in \mathbb{N}$. The ideals $I_i = (p_i^{r_i})$, $i = 1, \dots, k$, are comaximal in R, and $(I_1 \cdots I_k)M = 0$, so by The Chinese remainder theorem 2.7.2, $R/(a) \cong R/(p_1^{r_1}) \oplus \cdots \oplus R/(p_k^{r_k})$.

Let now M be a finitely generated module over R. After constructing such an isomorphism for all a_i in the assertion of Theorem 7.2.1, we may write $M \cong R^l \oplus \bigoplus_{i=1}^{m} \bigoplus_{j=1}^{k_i} R/(p_{i,j}^{r_{i,j}})$ for some (not necessarily distinct) irreducible elements $p_{i,j} \in R$ and integers $r_{i,j} \in \mathbb{N}$. The scalars $p_{i,j}^{r_{i,j}}$ are called *the elementary divisors* of M.

7.2.4. Lemma. *Let R be a PID, p be a prime element of R, and $M = R/(p^r)$ for some $r \in \mathbb{N}$. Then $p^s M/p^{s+1}M \cong R/(p)$ for any $0 \le s \le r-1$ and $= 0$ for any $s \ge r$. If q is another prime element of R (so that q and p are not associates), then $q^s M/q^{s+1}M = 0$ for all s.*

Proof. If $s \ge r$, we have $p^s M = 0$, so $p^s M/p^{s+1}M = 0$. If $s < r$, we have $(p^r) \subseteq (p^s)$, so $p^s M \cong (p^s)/(p^r)$, and $p^s M/p^{s+1}M \cong (p^s)/(p^{s+1})$. The isomorphism $R \longrightarrow (p^s)$, $a \mapsto p^s a$, maps (p) onto (p^{s+1}), thus $(p^s)/(p^{s+1}) \cong R/(p)$.

For any q coprime with p and any s, $q^s M \cong ((q^s) + (p^r))/(p^r) = R/(p^r) = M$, so $q^s M/q^{s+1}M \cong M/M = 0$. ∎

7.2.5. Based on Lemma 7.2.4, we obtain the uniqueness of the decomposition $M \cong \bigoplus_{i=1}^{k}(R/p_i^{r_i})$ of a torsion module M:

Theorem II – uniqueness 1. *The elementary divisors of a finitely generated module over a PID are defined uniquely (up to permutation and association).*

Proof. Let M be a finitely generated module over a PID R, let $M' = \mathrm{Tor}(M)$, let $M' \cong R/(p_1^{r_1}) \oplus \cdots \oplus R/(p_k^{r_k})$ where $p_1, \dots, p_k$ are (not necessarily distinct) irreducible elements of R and $r_1, \dots, r_k \in \mathbb{N}$. Then by Lemma 7.2.4, for every irreducible $p \in R$ and any $s \in \mathbb{N}$, the number of i for which $p_i^{r_i} = p^s$ (up to association) is $\dim_{R/(p)}(p^{s-1}M'/p^s M') - \dim_{R/(p)}(p^s M'/p^{s+1}M')$. ∎

7.2.6. By 7.2.3, knowing the invariant factors of a finitely generated module M over a PID, we easily find the elementary divisors of M by decomposing the invariant factors to products of powers of irreducibles. Conversely, having elementary divisors of M, we easily reconstruct the invariant factors of M: if we list the elementary divisors this way:

$$p_1^{r_{1,m}},\ \dots,\ p_1^{r_{1,1}},\quad p_2^{r_{2,m}},\ \dots,\ p_2^{r_{2,1}},\quad \dots,\ p_k^{r_{k,m}},\ \dots,\ p_k^{r_{k,1}},$$

where p_i are distinct irreducibles in R, and $r_{k,l}$ are (possibly, zero) integers satisfying $r_{i,m} \ge r_{i,m-1} \ge \cdots \ge r_{i,1} \ge 0$ for every i and $r_{i,m} > 0$ for all i, then we have no other choice but to put $a_m = p_1^{r_{1,m}} \cdots p_k^{r_{k,m}}$, …, $a_1 = p_1^{r_{1,1}} \cdots p_k^{r_{k,1}}$.

7.2.7. We obtain:

Theorem II – uniqueness 2. *The invariant factors of a finitely generated module over a PID are defined uniquely (up to association).*

7.2.8. It follows that for a submodule N of a fintely generated free module M over a PID, the scalars $a_1, \dots, a_k$ in Theorem 7.1.1 are uniquely defined (up to association): if $c_1, \dots, c_m$ are the invariant factors of the module M/N, then $(a_1, \dots, a_k) = (1, \dots, 1, c_1, \dots, c_m)$.

7.2.9. How to find the invariant factors of a module? Let M be a finitely generated module over a PID R defined via generators and relations: M is defined as a set of formal R-linear combinations of elements $v_1, \dots, v_n$ satisfying a finite family of linear relations,

$$\sum_{i=1}^{n} a_{i,1} v_i = \cdots = \sum_{i=1}^{n} a_{i,m} v_i = 0.$$

This is equivalent to defining M as the quotient of the free module R^n by the submodule N, generated by the vectors $\begin{pmatrix} a_{1,1} \\ \vdots \\ a_{n,1} \end{pmatrix}$, ..., $\begin{pmatrix} a_{1,m} \\ \vdots \\ a_{n,m} \end{pmatrix}$. The matrix $A = \begin{pmatrix} a_{1,1} \dots a_{1,m} \\ \vdots \quad \vdots \\ a_{n,1} \dots a_{n,m} \end{pmatrix}$ is called *the relations matrix* of M. Now, as described in 7.1.3, A can be reduced, by left and right multiplication by invertible matrices, to its Smith normal form (7.1), with $a_1 \mid \dots \mid a_k$; the nonunit scalars out of $a_1, \dots, a_k$ are just the invariant factor of M.

In the case R is an ED, the reduction of A to its Smith normal form can be made using row/column operations from 7.1.5; by tracking these operations we can find an explicit presentation of M as a direct sum of its cyclic submodules.

7.3. The rational normal form of the matrix of a linear transformation of a finite dimensional vector space

We will now apply the results of subsection 7.2 in linear algebra – to establish the so-called *normal forms* of square matrices over a field (the forms, to which any square matrix can be reduced by a change of basis).

Let F be a field, let V be an n-dimensional F-vector space, and let T be a linear transformation of V (an F-module homomorphism $V \longrightarrow V$).

7.3.1. The transformation T defines an $F[x]$-module structure on V by putting $p(x)u = p(T)(u)$, $p \in F[x]$. The F-basis of V generates V as an $F[x]$-module, so V is generated by $\leq n$ elements.

7.3.2. A submodule of the $F[x]$-module V is a subspace W of V invariant under the action of $F[x]$, $F[x]W \subseteq W$; for this, it is necessary and sufficient that $T(W) \subseteq W$. If a basis $\{u_1, \dots, u_n\}$ of V is such that $\{u_1, \dots, u_k\}$ is a basis of W, then with respect to this basis the matrix of T has form $\begin{pmatrix} A_1 & C \\ O & A_2 \end{pmatrix}$, where A_1 is the matrix of $T|_W$ with respect to the basis $\{u_1, \dots, u_k\}$.

7.3.3. If the $F[x]$-module V is a direct sum of two its submodules (that is, T-invariant subspaces) $V = W_1 \oplus W_2$, and a basis $\{u_1, \dots, u_n\}$ of V is such that $\{u_1, \dots, u_k\}$ is a basis of W_1 and $\{u_{k+1}, \dots, u_n\}$ is a basis of W_2, then the matrix of T with respect to this basis has form $\begin{pmatrix} A_1 & O \\ O & A_2 \end{pmatrix}$, where A_1 is the matrix of $T|_{W_1}$ with respect to the basis $\{u_1, \dots, u_k\}$ and A_2 is the matrix of $T|_{W_2}$ with respect to the basis $\{u_{k+1}, \dots, u_n\}$.

7.3.4. Since $F[x]$ is an infinite dimensional F-vector space, V contains no copy of $F[x]$, so, is a torsion $F[x]$-module. The annihilator $\mathrm{Ann}(V)$ is an ideal in $F[x]$, generated by a monic polynomial $m_T \in F[x]$. m_T is called *the minimal polynomial* of T: we have $m_T(T)u = 0$ for all $u \in V$, so $m_T(T) = 0$, and any other polynomial $p \in F[x]$ satisfying $p(T) = 0$ is divisible by m_T.

7.3.5. Since the ring $F[x]$ is a PID (and even an ED), the theory of finite generated modules over PIDs applies to V, and we get the following theorem:

Theorem. *V is representable as a direct sum $V = W_1 \oplus \dots \oplus W_m$ of T-invariant subspaces which are cyclic $F[x]$-modules: for each i, $W_i \cong R/(p_i)$, where p_i are nonconstant polynomials from $F[x]$ satisfying $p_1 \mid p_2 \mid \dots \mid p_m$.*

The polynomials $p_1, \dots, p_m$ are called *the invariant factors* of T; they are uniquely defined up to multiplication by a constant, so, are uniquely defined if assumed to be monic. For each i, p_i is the minimal polynomial of $T|_{W_i}$; the senior invariant factor p_m, being divisible by all other, is just the minimal polynomial m_T of T. V is a cyclic $F[x]$-module iff it has a single invariant factor, iff $\deg m_T = n$.

7.3.6. Let W be an $F[x]$-submodule of V, that is, a subspace of V invariant under the action of T: $T(W) \subseteq W$. W is a cyclic $F[x]$-module iff there is $u \in W$ such that $F[x]u = W$; such vector u is called *a cyclic vector* of $T|_W$. We have $W \cong F[x]/(p)$ for some polynomial $p = x^k + a_{k-1}x^{k-1} + \dots + a_1 x + a_0 \in F[x]$, of degree k. As an F-vector space, $F[x]/(p)$ has dimension k and basis $\{\bar{1}, \bar{x}, \dots, \bar{x}^{k-1}\}$ (where $\bar{q}$ denotes the class $q + (p)$ of q modulo (p)). So, $\dim_F W = k$, and $\{u, T(u), \dots, T^{k-1}(u)\}$ is an F-basis of W, and $p(T)(u) = T^k(u) + a_{k-1}T^{k-1}(u) + \dots + a_1 T(u) + a_0 u = 0$.

Under the action of T, we have

$$u \mapsto T(u) \mapsto T^2(u) \mapsto \dots \mapsto T^k(u) = -a_0 - a_1 t(u) - \dots - a_{k-1}T^{k-1}(u),$$

so in the basis $\{u, T(u), \dots, T^{k-1}(u)\}$, the $k \times k$ matrix of $T|_W$ is

$$C_p = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & 0 & \dots & 0 & -a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & -a_{k-2} \\ 0 & 0 & 0 & \dots & 1 & -a_{k-1} \end{pmatrix}. \tag{7.2}$$

Matrix C_p is called *the companion matrix* of the monic polynomial $p = x^k + a_{k-1}x^{k-1} + \dots + a_1x + a_0$. We have $\det C_p = \pm a_0$, and, since $p(T|_W) = 0$, $p(C_p) = 0$.

7.3.7. In the decomposition $V = W_1 \oplus \dots \oplus W_m$ of V into a sum of cyclic $F[x]$-modules, for each i, choose a basis in W_i as described in 7.3.6, and let B be the union of these bases. Then with respect to basis B the matrix of T has the block-diagonal form

$$\begin{pmatrix} C_{p_1} & 0 & \dots & 0 \\ 0 & C_{p_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & C_{p_m} \end{pmatrix}, \tag{7.3}$$

where for each i, C_{p_i} is the companion matrix of the i-th invariant factor p_i of T. This form of a matrix is called *the Frobenius normal form* or *the rational normal form*; for any transformation T, the rational normal form of its matrix is unique.

7.3.8. As a corollary, we obtain that any square matrix is similar to a unique matrix of the form (7.3) with $p_1 \,\big|\, \dots \,\big|\, p_m$, and that two matrices are similar iff they have the same rational normal form.

7.3.9. The invariant factors, along with the rational normal form of a matrix are field independent: if A is a matrix over a field F_1 and F_2 is a field containing F_1, then the invariant factors and the rational normal form of A over F_2 are the same as over F_1.

7.3.10. Instead of invariant factors of T (that is, of the $F[x]$-module V) we as well can use the elementary divisors thereof. As a result, we will also obtain a matrix of the form (7.3), but now with polynomials p_i, instead of dividing each the next, being powers of irreducible polynomials, $p_i = q_i^{r_i}$. Such a form of the matrix is also unique for T (up to permutations of blocks), but is field dependent.

7.4. The Smith normal form of $x - T$ and the characteristic polynomial

How can the rational normal form of the matrix of a linear transformation be found? One possible method is to utilize the theory from 7.2.9.

We preserve notation from the previous section.

7.4.1. Let $A = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & & \vdots \\ a_{n,1} & \dots & a_{n,n} \end{pmatrix}$ be a matrix of T in an arbitrary basis $B = \{u_1, \dots, u_n\}$ of V. Then for each i, $xu_i = T(u_i) = a_{1,i}u_1 + \dots + a_{n,i}u_n$; these relations form the matrix $\begin{pmatrix} x-a_{1,1} & -a_{1,2} & \dots & -a_{1,n} \\ -a_{1,1} & x-a_{2,2} & \dots & -a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n,1} & -a_{n,2} & \dots & x-a_{n,n} \end{pmatrix} = xI - A$, where I is the unit $n \times n$ matrix.

I claim that this is the complete relations matrix of the $F[x]$-module V, that is, that all relations of V follow from the relations above. Indeed, let N be the submodule of the free module $F[x]^n$ generated by these relations, and let $M = F[x]^n/N$. Then in M, $\begin{pmatrix} x \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \end{pmatrix}$, $\begin{pmatrix} 0 \\ x \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{n,2} \end{pmatrix}$, etc. So, x (and so any its power) in this module, in any coordinate position, can be replaced by a vector from F^n, and so M is an F-vector space of dimension $\le n$. But since V has all the relations from N, there is an epimorphism $M \longrightarrow V$; and since $\dim V = n$, we have that $V \cong M$.

Thus, by 7.2.9, using the row/column operations from 7.1.5, we can reduce the matrix $xI - A$ to its Smith normal form. Since $\operatorname{rank}_{F[x]} V = 0$, the obtained matrix has no zero columns, and we will actually get a diagonal matrix

$$\begin{pmatrix} c_1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & c_2 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & c_l & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & p_1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & p_2 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & p_m \end{pmatrix} \tag{7.4}$$

where $c_1, \dots, c_l \in F$ and $p_1, \dots, p_m$ are nonzero nonconstant polynomials with $p_1 \mid \cdots \mid p_m$; these are the invariant factors of T. After multiplying $p_1, \dots, p_m$ by suitable constants (elements of F, which are units in $F[x]$), we may and will assume that they all are monic.

7.4.2. Let A be the matrix of T and let C be the Smith normal form of $xI - A$. Since the row/column operations from 7.1.5 preserve, up to multiplication by an element of F (a unit of $F[x]$), the determinant of the matrix (or since the matrices P and Q from 7.1.4 are invertible in $\mathrm{Mat}_{n,n}(F[x])$ and so, their determinants are elements of F), and since the invariant factors are assumed to be monic, we have $\det C = \det(xI - A)$. The polynomial $c_T(x) = \det(xI - A)$, of degree n, is called *the characteristic polynomial* of T (and of A); we see that $c_T = \det C = p_1 p_2 \cdots p_m$, that is, up to multiplication by an element of F, is the product of the elementary divisors of T.

7.4.3. In particular, the minimal polynomial $m_T = p_m$ of T divides its characteristic polynomial c_T. Since $m_T(T) = 0$, this implies

The Cayley-Hamilton theorem. *If T is a linear transformation of a finite dimensional vector space and c_T is the characteristic polynomial of T, then $c_T(T) = 0$.*

7.5. The Jordan normal form of a matrix

In the case the characteristic polynomial of T splits into a product of linear factors, T has another standard normal form.

7.5.1. Assume that V is a cyclic $F[x]$-module and that the minimal polynomial of T has form $c_T(x) = (x-\lambda)^n$ for some $\lambda \in F$. Consider the transformation $S = T - \lambda I$; the minimal polynomial of S is x^n. (So, $S^n = 0$; S is said to be *nilpotent*.) Since the degree of the minimal polynomial of S is $n = \dim V$, V is cyclic under the action of S as well; let u be a cyclic vector of V. Then S acts on u the following way: $u \mapsto Su \mapsto S^2(u) \mapsto \cdots \mapsto S^{n-1}(u) \mapsto S^n(u) = 0$, and $\{u, S(u), \dots, S^{n-1}(u)\}$ is a basis of V. If we reverse the order of elements of this basis, that is, in the basis $\{S^{n-1}(u), \dots, S(u), u\}$, the matrix of S is $\begin{pmatrix} 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}$, the $n \times n$ matrix with 1s above the main diagonal. Accordingly, the matrix of $T = S + \lambda I$ in this basis is $\begin{pmatrix} \lambda & 1 & \dots & 0 & 0 \\ 0 & \lambda & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & \dots & 0 & \lambda \end{pmatrix}$. A matrix of this form is called *a Jordan cell*.

7.5.2. Now assume that T is a linear transformation of a vector space V and that the characteristic polynomial of T splits into a product of linear factors, $c_T(x) = \prod_{i=1}^{n}(x - \lambda_i)$, $\lambda_1, \dots, \lambda_n \in F$. Then every elementary divisor of T has form $(x - \lambda_i)^{r_i}$ for some i and some $r_i \in \mathbb{N}$, and so, in a suitable basis, the matrix of T is block-diagonal $\begin{pmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_k \end{pmatrix}$ with each J_i being a Jordan cell. This matrix is called *the Jordan normal form* of the matrix of T.

7.5.3. A field F is said to be *algebraically closed* if every polynomial over F has a root in F; equivalently, if every polynomial over F splits into a product of linear factors. (The fundamental theorem of algebra says that $\mathbb{C}$, the field of complex numbers, is algebraically closed.)

In the case F is *an algebraically closed field*, every transformation of an F-vector space has a Jordan normal form (and so, every square matrix over F does).

7.5.4. The roots of the characteristic polynomial of a transformation T are called *eigenvalues* of T. So, if T has a Jordan normal form, the diagonal elements of its Jordan cells are the eigenvalues of T.

If λ is an eigenvalue of T, then $\det(\lambda I - T) = 0$, so $\lambda I - T$ is not invertible, so it has a nontrivial kernel, so there is a nonzero vector $u \in V$ such that $T(u) = \lambda u$. Such a vector u is called *the eigenvector* of T corresponding to eigenvalue λ.