Modules over PID

Let A be a ring, and M be an A-module. Recall that M is called *free* over A if it admits a (finite) basis over A. In other words M is free if and only if there exists a finite set $(v_1, \dots, v_r)$ of elements of M such that the natural A-linear map $\mathrm{A}^r \to \mathrm{M}$ defined by $(a_1, \dots, a_r) \mapsto \sum_{k=1}^r a_k v_k$ is a bijection (i.e., an isomorphism of A-modules). Such a set $(v_1, \dots, v_r)$ is called a basis for M over A. One can define freeness without references to elements: a module M is free if and only if there exist an integer $r \geqslant 1$ and an isomorphism of A-modules $\mathrm{A}^r \to \mathrm{M}$.

Be aware that some (most) modules are not free. The most simple is example to bear in mind is the following: let $\mathrm{A} = \mathbb{Z}$ be the ring of integers, and $n \geqslant 1$ be an integer. Then the A-module $\mathrm{M} := \mathbb{Z}/n\mathbb{Z}$ is finitely generated over A (by 1_M) but cannot be free over A (otherwise, M would be isomorphic to A^m for some $m \geqslant 1$... but, as sets, M is finite while A is not).

Here is a more elaborate example. Let k be a field, consider $\mathrm{A} := k[\mathrm{X}, \mathrm{Y}]$ be the ring of polynomials in two variables with coefficients in k. One can view $\mathrm{M} = \mathrm{A}$ as a module over itself. It is clear that M is a free A-module of rank 1 (a basis of M over A is given by (1_A)). Let N denote the sub-A-module of M generated by X and Y. This means that N consists of all polynomials of the form $\mathrm{P}(\mathrm{X}, \mathrm{Y}) \cdot \mathrm{X} + \mathrm{Q}(\mathrm{X}, \mathrm{Y}) \cdot \mathrm{Y}$ with $\mathrm{P}, \mathrm{Q} \in \mathrm{A}$. Then N is finitely generated: the A-linear map $\phi : \mathrm{A}^2 \to \mathrm{N}$ given by $(\mathrm{P}, \mathrm{Q}) \mapsto \mathrm{PX} + \mathrm{QY}$ is surjective. But N is not free! Indeed, one can check that $\operatorname{Ker} \phi$ is non zero (it contains $(-\mathrm{Y}, \mathrm{X})$ for instance), so that ϕ is not an isomorphism.

That being said, modules over PIDs are a bit more well-behaved:

Theorem 1. *Let* A *be a principal ideal domain, and let* M *be a free* A*-module of rank* m*. Let* N *be a sub-*A*-module of* M*. Then* N *is free, and the rank* n *of* N *satisfies* $0 \leqslant n \leqslant m$*.*

Proof. The proof requires a lemma.

Lemma 2. *Let* M *be a module over a principal ideal domain* A*. Let* $u : \mathrm{M} \to \mathrm{A}$ *be an* A*-linear map. Then there is an* A*-linear isomorphism*

$$\mathrm{M} \simeq \operatorname{Ker} u \times \operatorname{Im} u.$$

Proof. If u is the zero map $m \mapsto 0$, there is nothing to prove (since then $\operatorname{Im} u = \{0\}$ and $\operatorname{Ker} u = \mathrm{M}$). So we can assume that $u \neq 0$. The image $\operatorname{Im} u$ is then a non-zero submodule of A i.e., a non-zero ideal of A. Since A is principal, there exists $a \in \mathrm{A} \setminus \{0\}$ such that $\operatorname{Im} u = \mathrm{A} \cdot a$. Note that, as an A-module, $\mathrm{A} \cdot a \simeq \mathrm{A}$ (because A is integral). Hence, any element $b \in \mathrm{A} \cdot a$ can be written in a unique way as $b = r \cdot a$ with $r \in \mathrm{A}$. Since $a \in \operatorname{Im} u$, there exists $m_0 \in \mathrm{M}$ such that $u(m_0) = a$.

Consider the map $\lambda : \operatorname{Ker} u \times \operatorname{Im} u \to \mathrm{M}$ defined by $\lambda(m, r \cdot a) \mapsto m + r \cdot m_0$. It is clear that λ is a morphism of A-modules. Let us now check that λ is bijective. This will provide the desired isomorphism.

Let $x = (m, r \cdot a) \in \operatorname{Ker} u \times \operatorname{Im} u$ be such that $\lambda(m, r \cdot a) = 0$. Then $m + r \cdot m_0 = 0$ in M. Suppose for a moment that $r \neq 0$, then we deduce that $u(m) = -r \cdot u(m_0)$. Since $m \in \operatorname{Ker} u$ and $u(m_0) = a \neq 0$, this contradicts the fact that A is integral. Hence $r = 0$ and $m = -rm_0 = 0$. Thus $x = 0$ and λ is injective.

Now let $m \in \mathrm{M}$ be an arbitrary element. Since $a = u(m_0)$ generates $\operatorname{Im} u$ as an A-module, we have $u(m) = r \cdot u(m_0)$ for some $r \in \mathrm{A}$. We write $m = (m - r \cdot m_0) + r \cdot m_0$, and let $m_1 := m - r \cdot m_0$. We have $u(m_1) = u(m) - u(r \cdot m_0) = u(m) - ru(m_0) = 0$ so that $m_1 \in \operatorname{Ker} u$. Hence $m = \lambda(m_1, r)$. Therefore λ is surjective. $\square$

We can now prove the Theorem by induction on the rank m of M. If $m = 0$, there is nothing to prove so we assume that $m \geqslant 1$.

Suppose that the Theorem holds for all free A-modules M' of rank m. Let us prove that the Theorem then holds for all free A-modules of rank $m+1$. Let M be an arbitrary free A-module of rank $m+1$, and let N be a submodule of M. By definition, we can find an A-linear isomorphism $\phi : \mathrm{M} \to \mathrm{A}^{m+1}$. Through ϕ, the submodule N of M is isomorphic to the submodule $\phi(\mathrm{N})$ of A^{m+1}. Hence, there is no loss of generality in assuming that $\mathrm{M} = \mathrm{A}^{m+1}$ and that N is a sub-A-module of A^{m+1}. Write $\pi' : \mathrm{A}^{m+1} \to \mathrm{A}$ for the projection on the last coordinate (defined by $(a_1, \dots, a_{m+1}) \mapsto a_{m+1}$). The map π' is clearly A-linear. We restrict π' to N and write π for the resulting map. We apply the Lemma to the A-linear map $\pi : \mathrm{N} \to \mathrm{A}$. We have an isomorphism $\mathrm{N} \simeq \mathrm{Ker}\,\pi \times \mathrm{Im}\,\pi$. It is clear that $\mathrm{Ker}\,\pi = \mathrm{N} \cap (\mathrm{A}^m \times \{0\})$ and that $\mathrm{A}^m \times \{0\} \simeq \mathrm{A}^m$. Hence $\mathrm{Ker}\,\pi$ is isomorphic to a sub-A-module of A^m. Since A^m is free of rank m, we may use our induction hypothesis: this yields that $\mathrm{Ker}\,\pi$, being a submodule of a free module of rank m, is a free A-module of rank n with $n \leqslant m$. Thus, there exists an isomorphism of A-modules $\mathrm{Ker}\,\pi \simeq \mathrm{A}^n$.

On the other hand, $\mathrm{Im}\,\pi$ is a submodule of A (which means that $\mathrm{Im}\,\pi$ is an ideal in A). Since A is a PID, $\mathrm{Im}\,\pi$ is principal: we can find $a \in \mathrm{A}$ such that $\mathrm{Im}\,\pi = \mathrm{A} \cdot a$. If $a = 0$, it is clear that $\mathrm{Im}\,\pi = \{0\}$ is free of rank 0. If $a \neq 0$, we have an isomorphism $\mathrm{Im}\,\pi \simeq \mathrm{A}$ which shows that $\mathrm{Im}\,\pi$ is free of rank 1.

Putting these ingredients together, we conclude that N is isomorphic, as an A-module, to either $\mathrm{A}^n \times \{0\}$ or $\mathrm{A}^n \times \mathrm{A} \simeq \mathrm{A}^{n+1}$. In any case, N is free and its rank r satisfies $r \leqslant n+1 \leqslant m+1$. This completes the induction step, and concludes the proof of the Theorem. □

Theorem 3. *Let* A *be a* principal ideal domain *and* M *be a* free A*-module of rank m. Let* N *be a non-zero submodule of* N*. By the previous theorem,* N *is free, and the rank n of* N *satisfies $1 \leqslant n \leqslant m$. There exist*

- *a basis $(e_1, \dots, e_m)$ of* M *over* A*,*
- *and non-zero elements $a_1, \dots, a_n$ in* A*,*

such that

- *$(a_1 e_1, \dots, a_n e_n)$ is a basis of* N *over* A*,*
- *and, for all $i \in \{1, \dots, n-1\}$, a_i divides a_{i+1}.*

This theorem proves that there exists a basis of M which is "adapted to N". The ideals $\mathrm{A} \cdot a_1, \dots, \mathrm{A} \cdot a_n$ are called the invariant factors of M in N. One can show that they are uniquely determined by M and N (warning: the elements $a_1, \dots, a_n$ are only determined up to multiplication by units in A).

Proof. We prove the Theorem by induction on the rank of M. The induction step requires the following construction.

Let M be a free A-module of rank $m \geqslant 1$, and $\mathrm{N} \neq 0$ be a submodule of M. The set $\mathrm{Hom}_{\mathrm{A}}(\mathrm{M}, \mathrm{A})$ of A-linear maps $u : \mathrm{M} \to \mathrm{A}$ can be equipped with an A-module structure. We denote this A-module by $\mathrm{M}^\vee$ (we could call it the "dual of M").

For any $u \in \mathrm{M}^\vee$, the image $u(\mathrm{N})$ of u restricted to N is a sub-A-module of A i.e., $u(\mathrm{N})$ is an ideal of A. Since A is a principal ideal domain, we may find a generator $a_u \in \mathrm{A}$ of $u(\mathrm{N})$. Consider the family $\mathcal{F} := \{u(\mathrm{N}),\ u \in \mathrm{M}^\vee\}$ of ideals of A. The family $\mathcal{F}$ is non-empty since it contains the ideal 0 (the 0-map $v \mapsto 0$ is an element of $\mathrm{M}^\vee$). Now, by a corollary of Theorem 1.3.1, any non-empty family of ideals in a principal ideal domain admits a maximal element.

This means that there exist $u_{\mathrm{N}} \in \mathrm{M}^\vee$ and an element $a_{\mathrm{N}} \in \mathrm{A}$ with $u_{\mathrm{N}}(\mathrm{N}) = \mathrm{A} \cdot a_n$ such that, for all $v \in \mathrm{M}^\vee$, $v(\mathrm{N}) \subset u_{\mathrm{N}}(\mathrm{N})$. In other words, for all $v \in \mathrm{M}^\vee$, a_{N} divides a_v in A (or $\mathrm{A} \cdot a_v \subseteq \mathrm{A} \cdot a_{\mathrm{N}}$). By construction, we may find $e \in \mathrm{N}$ such that $u_{\mathrm{N}}(e) = a_{\mathrm{N}}$.

- Fact 1 : the element a_{N} is non-zero.

  Proof. Let us choose a basis $(g_1, \dots, g_m)$ of M over A (such a basis exists by hypothesis), and denote by $p_i : \mathrm{M} \to \mathrm{A}$ the i-th coordinate function. This map is characterised by $p_i(g_j) = \delta_{ij}$ for all $1 \leqslant i, j \leqslant m$ and the fact that it is A-linear. We have $p_i \in \mathrm{M}^\vee \setminus \{0\}$. Since $\mathrm{N} \neq 0$, there is an index i such that $p_i(\mathrm{N}) \neq 0$. For this index i, we have $0 \subsetneq p_i(\mathrm{N}) \subset u_\mathrm{N}(\mathrm{N})$, by maximality of $u_\mathrm{N}(\mathrm{N})$. In particular, the element a_N cannot be zero since the ideal it generates contains a non-zero ideal $p_i(\mathrm{N})$. $\square$

- Fact 2 : for all $v \in \mathrm{M}^\vee$, a_N divides $v(e)$ in A.

  Proof. Let $v \in \mathrm{M}^\vee$. The ring A is principal, so we may introduce $d := \gcd(a_\mathrm{N}, v(e))$. It suffices to show that $d = a_\mathrm{N}$, up to a unit of A. By Bézout's theorem, there exist $\alpha, \beta \in \mathrm{A}$ such that $d = \alpha a_\mathrm{N} + \beta v(e)$. By construction $a_\mathrm{N} = u_\mathrm{N}(e)$, so that $d = \alpha u_\mathrm{N}(e) + \beta v(e) = (\alpha u_\mathrm{N} + \beta v)(e)$. Since $\mathrm{M}^\vee$ is an A-module, the map $w := \alpha u_\mathrm{N} + \beta v$ belongs to $\mathrm{M}^\vee$, and the identity we have just proved shows that $d \in w(\mathrm{N})$. Therefore $\mathrm{A} \cdot d \subset w(\mathrm{N})$ because $w(\mathrm{N})$ is an ideal in A. We have $\mathrm{A} \cdot a_n \subset \mathrm{A} \cdot d$ because d divides a_n. Moreover, the maximality of $u_\mathrm{N}(\mathrm{N})$ implies that $w(\mathrm{N}) \subset u_\mathrm{N}(\mathrm{N})$. We thus have a chain of inclusions: $\mathrm{A} \cdot a_\mathrm{N} \subset \mathrm{A} \cdot d \subset w(\mathrm{N}) \subset u_\mathrm{N}(\mathrm{N}) = \mathrm{A} \cdot a_\mathrm{N}$. Hence $\mathrm{A} \cdot a_n = \mathrm{A} \cdot d$, so that a_N and d differ by a unit in A. $\square$

Let us now choose a basis $(g_1, \ldots, g_m)$ for M over A and write, as above, $p_i : \mathrm{M} \to \mathrm{A}$ for the i-th coordinate function $(1 \leqslant i \leqslant m)$. Applying Fact 2 to $v = p_i$ yields that a_N divides $p_i(e)$ in A: hence there exists $b_i \in \mathrm{A}$ such that $p_i(e) = b_i \cdot a_\mathrm{N}$. We let $f := \sum_{i=1}^m b_i \cdot g_i \in \mathrm{M}$. We have $e = \sum p_i(e) \cdot g_i = a_\mathrm{N} \cdot f$. Moreover, $u_\mathrm{N}(f) = 1$ since $u_\mathrm{N}(e) = a_\mathrm{N}$ and A is integral.

- Fact 3 : we have $\mathrm{M} = \operatorname{Ker} u_n + \mathrm{A} \cdot f$, the sum being direct (i.e. $(\operatorname{Ker} u_n) \cap \mathrm{A} \cdot f = 0$).

  Proof. It is clear that $\operatorname{Ker} u_n + \mathrm{A} \cdot f \subset \mathrm{M}$. For any $x \in \mathrm{M}$, we can write $x = u_\mathrm{N}(x) \cdot f + (x - u_\mathrm{N}(x) \cdot f)$. One readily checks that $u_\mathrm{N}(x) \cdot f \in \mathrm{A} \cdot f$ and that $x - u_\mathrm{N}(x) \cdot f \in \operatorname{Ker} u_\mathrm{N}$. Hence $\mathrm{M} = \operatorname{Ker} u_\mathrm{N} + \mathrm{A} \cdot f$. It is also clear that $(\operatorname{Ker} u_\mathrm{N}) \cap \mathrm{A} \cdot f = 0$, since $u_\mathrm{N}(f) = 1 \neq 0$ in A. $\square$

- Fact 4 : we have $\mathrm{N} = (\operatorname{Ker} u_n \cap \mathrm{N}) + \mathrm{A} \cdot a_\mathrm{N} f$, the sum being direct.

  Proof. The proof is very similar to that of the previous fact. Here, one decomposes any $y \in \mathrm{N}$ as $y = b \cdot a_\mathrm{N} f + (y - u_\mathrm{N}(y) \cdot f)$, where $b \in \mathrm{A}$ is such that $u_\mathrm{N}(y) = b a_\mathrm{N}$. $\square$

We can now prove the Theorem. Assume that, for some $m \geqslant 1$, the statement holds for free A-modules of rank m. Let us prove that the theorem holds for free A-modules of rank $m+1$.

Let M be a free A-module of rank $m+1$, and let $\mathrm{N} \subset \mathrm{M}$ be a non-zero submodule of M. Consider the submodule $\mathrm{M}' = \operatorname{Ker} u_\mathrm{N}$ of M which we constructed above. By the previous Theorem, M' is a free A-module. And Fact 3 shows that M' has rank m (a slight reformulation of Fact 3 indeed shows that $\mathrm{M} \simeq \mathrm{M}' \times \mathrm{A}$). The construction also provides $a_1 := a_\mathrm{N} \in \mathrm{A} \smallsetminus \{0\}$ and a $e_1 := f \in \mathrm{M}$.

We are now in a position to apply the induction hypothesis: M' is a free A-module of rank m, and $\mathrm{N}' := \mathrm{N} \cap \mathrm{M}'$ is a submodule of M'. The induction hypothesis then yields that there exist a basis $(e_2, \ldots, e_m)$ of M' over A and non-zero elements $a_2, \ldots, a_n$ of A such that

- $(a_2 e_2, \ldots, a_n e_n)$ is a basis for N over A,
- and a_i divides a_{i+1} for all $2 \leqslant i \leqslant n-1$.

Fact 3 above shows that $(e_1, e_2, \ldots, e_m)$ is a basis for M over A. Fact 4 proves that $(a_1 e_1, a_2 e_2, \ldots, a_n e_n)$ is a basis for N over A. It remains to prove that a_1 divides a_2 in A.

Consider the A-linear map $v : \mathrm{M} \to \mathrm{A}$ defined by $v(e_1) = v(e_2) = 1$ and $v(e_i) = 0$ for $i \geqslant 3$. Then we have $a_1 = a_\mathrm{N} = a_\mathrm{N} v(e_1) = v(a_\mathrm{N} e_1) = v(f)$ so $a_1 \in v(\mathrm{N})$. Therefore, by maximality of $u_\mathrm{N}(\mathrm{N})$, we have $\mathrm{A} \cdot a_\mathrm{N} = \mathrm{A} \cdot a_1 \subset v(\mathrm{N}) \subset u_\mathrm{N}(\mathrm{N}) \subset \mathrm{A} \cdot a_\mathrm{N}$. Hence $v(\mathrm{N}) = \mathrm{A} \cdot a_1$. We also have $a_2 = v(a_2 e_2) \in v(\mathrm{N})$. Thus $\mathrm{A} \cdot a_2 \subset \mathrm{A} \cdot a_1$, which exactly means that a_1 divides a_2 in A.

This concludes the proof of the second theorem. $\square$