MODULES OVER A PID

A module over a PID is an abelian group that also carries multiplication by a particularly convenient ring of scalars. Indeed, when the scalar ring is the integers, the module is precisely an abelian group. This writeup presents the structure theorem for finitely generated modules over a PID. Although the result is essentially similar to the theorem for finitely generated abelian groups from earlier in this course, the result for modules over a PID is not merely generality for its own sake. Its first and best known application bears on linear algebra, and in this context the PID is the polynomial ring $k[X]$ (k a field) rather than the integer ring $\mathbb{Z}$. The theorem cogently gives the rational canonical form and the Jordan canonical form of a linear transformation, whereas establishing these results by more elementary methods feels unexplanatory to me. The next writeup of this course will present this application.

This writeup's proof of the structure theorem for finitely generated modules over a PID is not the argument found in many texts. That argument, which confused me for a long time, is an algorithm that

- proceeds from an assumption that often is left tacit, that the module has a presentation, meaning a characterizing description in terms of generators and relations that fully determines it (we so assumed in our proof of the theorem for finitely generated abelian groups);
- in fact is only as algorithmic as the arithmetic of the PID;
- blurs the distinction between two different uniqueness questions in a way that is easily lost.

Hence my choice to present the proof here instead, based on an exposition in Pierre Samuel's algebraic number theory text.

1. The Sun-Ze Theorem Again

Theorem 1.1. *Let A be a commutative ring with 1. Let $\mathfrak{a}$ and $\mathfrak{b}$ be ideals of A such that $\mathfrak{a}+\mathfrak{b}=A$. Then the natural map*

$$A \longrightarrow A/\mathfrak{a} \times A/\mathfrak{b}, \quad x \longmapsto (x+\mathfrak{a}, x+\mathfrak{b})$$

induces an isomorphism

$$A/\mathfrak{ab} \xrightarrow{\sim} A/\mathfrak{a} \times A/\mathfrak{b}.$$

Proof. First we show that $\mathfrak{ab}=\mathfrak{a}\cap\mathfrak{b}$. The reasoning

$$(\mathfrak{ab}\subset\mathfrak{a} \text{ and } \mathfrak{ab}\subset\mathfrak{b}) \implies \mathfrak{ab}\subset\mathfrak{a}\cap\mathfrak{b}$$

holds with no reference to the given condition $\mathfrak{a}+\mathfrak{b}=A$. On the other hand, the condition implies that

$$a+b=1 \quad \text{for some } a\in\mathfrak{a},\ b\in\mathfrak{b},$$

so that every $c\in A$ takes the form $c=(a+b)c=cb+ac$, and consequently

$$c\in\mathfrak{a}\cap\mathfrak{b} \implies c=cb+ac\in\mathfrak{ab}+\mathfrak{ab}=\mathfrak{ab}.$$

Now, the map $A \longrightarrow A/\mathfrak{a} \times A/\mathfrak{b}$ taking each x to $(x+\mathfrak{a}, x+\mathfrak{b})$ certainly has kernel $\mathfrak{a}\cap\mathfrak{b} = \mathfrak{a}\mathfrak{b}$. We need to show that it surjects. Given $(y+\mathfrak{a}, z+\mathfrak{b}) \in A/\mathfrak{a}\times A/\mathfrak{b}$, the value (constructed using a and b from the previous paragraph)

$$x = by + az = \begin{cases} (1-a)y + az = y + a(z-y) \\ by + (1-b)z = z + b(y-z) \end{cases}$$

satisfies

$$\begin{aligned} x + \mathfrak{a} &= y + \mathfrak{a}, \\ x + \mathfrak{b} &= z + \mathfrak{b}. \end{aligned}$$

This completes the proof. □

A straightforward induction argument generalizes the result to any number of pairwise coprime ideals, as follows.

Corollary 1.2. *Let A be a commutative ring with 1. Let $r \geq 2$ be an integer, and let $\mathfrak{a}_1, \cdots, \mathfrak{a}_r$ be ideals of A such that $\mathfrak{a}_i + \mathfrak{a}_j = A$ for all distinct pairs i, j. Then the natural map*

$$A \longrightarrow \prod_{i=1}^{r} A/\mathfrak{a}_i, \quad x \longmapsto (x+\mathfrak{a}_1, \cdots, x+\mathfrak{a}_r)$$

induces an isomorphism

$$A/\prod_{i=1}^{r} \mathfrak{a}_i \xrightarrow{\sim} \prod_{i=1}^{r} A/\mathfrak{a}_i.$$

For the induction, take pairs $(a_i, a_{r,i}) \in \mathfrak{a}_i \times \mathfrak{a}_r$ for $i = 1, \cdots, r-1$ such that

$$a_1 + a_{r,1} = 1, \quad a_2 + a_{r,2} = 1, \quad \cdots, \quad a_{r-1} + a_{r,r-1} = 1,$$

and multiply the equalities together to get

$$a_1 \cdots a_{r-1} + a_r = 1 \quad \text{where } a_r \in \mathfrak{a}_r.$$

Thus $(\mathfrak{a}_1 \cdots \mathfrak{a}_{r-1}) + \mathfrak{a}_r = A$ and so by the result for two ideals and then by induction on r,

$$A/\prod_{i=1}^{r} \mathfrak{a}_i \xrightarrow{\sim} A/\prod_{i=1}^{r-1} \mathfrak{a}_i \times A/\mathfrak{a}_r \xrightarrow{\sim} \prod_{i=1}^{r} A/\mathfrak{a}_i.$$

2. Finite Generation

Proposition 2.1. *Let A be a commutative ring with 1. Let M be an A-module. Then the following conditions are equivalent:*

(a) *Every nonempty family of A-submodules of M contains a maximal element under inclusion.*
(b) *Every A-submodule of M is finitely generated.*
(c) (*Noetherian property*) *Every increasing sequence $M_1 \subset M_2 \subset M_3 \subset \cdots$ of A-submodules of M is eventually stationary.*

Proof. (Sketch.) Assume (a). Let M' be an A-submodule of M. The family

$$\{\text{finitely generated } A\text{-submodules of } M'\}$$

contains a maximal element, which must be all of M', and so M' is finitely generated. This shows that (a) implies (b).

Assume (b). Consider an increasing sequence $M_1 \subset M_2 \subset M_3 \subset \cdots$, and let $M' = \bigcup_i M_i$. Because M' is finitely generated, and each generator lies in some M_i, the entire set of generators lies in some M_i, after which the sequence is stationary. This shows that (b) implies (c).

Assume (c). Consider any nonempty family of A-submodules of M. Let M_1 an element of the family. If M_1 is maximal in the family under inclusion then (a) holds. Otherwise $M_1 \subset M_2$ for some M_2 in the family, with the containment proper. If M_2 is maximal in the family under inclusion then (a) holds. Otherwise $M_2 \subset M_3$ for some M_3 in the family, with the containment proper, and so on. Because (c) holds, this process must terminate, producing a maximal element under inclusion in the family. □

Especially, if A is a PID, viewed as a module over itself, then certainly every A-submodule of A is finitely generated and thus every nonempty family of ideals of A contains a maximal element.

3. Basis, Rank, and the Main Theorem

Definition 3.1. *Let A be a PID, and let M be an A-module. A subset $\{e_i\}$ (possibly infinite) of M is a* **basis** *of M if every $x \in M$ has a unique expression as a finite A-linear combination of $\{e_i\}$. An A-module F that has a basis is called* **free**.

The definition here is not the mapping-theoretic definition of free module from an earlier handout, but of course the two are compatible.

Not all modules-over-PIDs are free, i.e., not all such modules have bases. For example, let $A = \mathbb{Z}$ and let $M = \mathbb{Z}/2\mathbb{Z}$. More interestingly, let $A = \mathbb{Z}$ and let $M = \mathbb{Q}$. We will return to this example later in the writeup. However, for finitely-generated modules, any submodule of a free module is again free, as follows.

Proposition 3.2. *Let A be a PID, and let F be a free A-module of finite rank. If S is an A-submodule of F then also S is free, and its rank is at most the rank of A.*

Proof. Let n denote the rank of F. When $n = 1$ the submodule S is free on at most one generator because A is a PID. For $n \geq 2$, identify F with $A^{\oplus n}$ and let $\pi : A^{\oplus n} \longrightarrow A^{\oplus(n-1)}$ be the projection to the last $n-1$ components. Since $\ker(\pi|_S) = \ker(\pi) \cap S$ is up to isomorphism an ideal of A, it is free on at most one generator. Also, πS is free on at most $n-1$ generators by induction, and so its mapping property guarantees a section $\iota : \pi S \longrightarrow S$, an A-module map such that $\pi \circ \iota = 1_S$. Because $\pi \circ \iota = 1_S$, the section has trivial kernel and so its image $\iota(\pi S)$ is isomorphic to πS, again free on at most $n-1$ generators. The decomposition $s = (s - (\iota \circ \pi)(s)) + (\iota \circ \pi)(s)$ for any $s \in S$ shows that $S = \ker(\pi|_S) \oplus \iota(\pi S)$ is free on at most n generators, as claimed; the sum is direct because $\pi \circ (\iota \circ \pi) = (\pi \circ \iota) \circ \pi = \pi$. (That is, if $\pi(\iota(\pi(s))) = 0$ then $\pi(s) = 0$ and so $\iota(\pi(s)) = 0$.) □

Now we proceed to our main theorem.

Theorem 3.3. *Let A be a PID. Let F be free A-module of finite rank n, and let S be an A-submodule of F, which therefore is free as well and has rank $m \leq n$. There exist a basis $(e_1, \cdots, e_n)$ of F and a chain of ideals of A,*

$$\mathfrak{a}_1 \supset \cdots \supset \mathfrak{a}_m,$$

such that

$$F = Ae_1 \oplus \cdots \oplus Ae_m \oplus \cdots \oplus Ae_n,$$
$$S = \mathfrak{a}_1 e_1 \oplus \cdots \oplus \mathfrak{a}_m e_m.$$

The ideals $\mathfrak{a}_1, \cdots, \mathfrak{a}_m$ *are uniquely determined by* F *and* S.

The idea of the proof is as follows. In the desired decompositions of F and S the ideal $\mathfrak{a}_1$ is large, so the argument starts by using its setting to concoct a large $\mathfrak{a}_1$ associated to F and S. Doing so entails a generator a_1 of $\mathfrak{a}_1$ in A and an element e_1' of S. Next the argument shows that e_1' takes the form $e_1' = a_1 e_1$ where $e_1 \in F$. Then Ae_1 will be the first summand of F and $\mathfrak{a}_1 e_1 = Aa_1 e_1 = Ae_1'$ the first summand of S, and the decompositions will follow easily.

Proof. We take $S \neq 0$, since otherwise the result is immediate. For every A-map $w : F \longrightarrow A$, the submodule-image $w(S)$ is an ideal of A. The key to the argument is that (as noted at the end of the previous section) among the ideals $\{w(S)\}$, some particular ideal $u(S)$ is maximal where $u : F \longrightarrow A$ is an A-map.[1] Certainly $u(S) \neq 0$ since for any basis $(x_1, \cdots, x_n)$ of F, some projection-image $\pi_i(S)$ is nonzero; such a basis exists because $S \neq 0$ and so $F \neq 0$. Let

$$\mathfrak{a}_1 = u(S) \quad \text{(maximal in the ideals } \{w(S)\}).$$

Since A is a PID, $\mathfrak{a}_1$ takes the form

$$\mathfrak{a}_1 = a_1 A \qquad \text{where } a_1 = u(e_1') \text{ for some } e_1' \in S.$$

As explained, we want to show that $e_1' = a_1 e_1$ for some $e_1 \in F$.

To do so, we establish that any A-map $v : F \longrightarrow A$ takes e_1' to a multiple of a_1. Indeed, since the sum of ideals is again an ideal,

$$\begin{aligned} \mathfrak{a}_1 = u(e_1')A \subset u(e_1')A + v(e_1')A &= (\alpha u(e_1') + \beta v(e_1'))A \quad \text{for some } \alpha, \beta \in A \\ &= w(e_1')A \qquad\qquad\qquad\quad \text{where } w = \alpha u + \beta v \\ &\subset w(S). \end{aligned}$$

But $w(S)$ cannot properly contain $\mathfrak{a}_1$ because the latter is maximal among A-map images of S in A. So in particular the first containment in the display must be equality, giving $v(e_1') \in u(e_1')A = a_1 A$ as desired.

Knowing that $v(e_1')$ is always a multiple of a_1 in A as v varies, we now show that e_1' itself is a multiple of a_1 in F. Consider any basis $(x_1, \cdots, x_n)$ of F. Specialize v to each projection π_i to get that the projections are all multiples of a_1,

$$\pi_i(e_1') = a_1 \alpha_i \text{ where } \alpha_i \in A, \quad i = 1, \cdots, n.$$

Thus, as desired,

$$e_1' = \sum_i a_1 \alpha_i x_i = a_1 \sum_i \alpha_i x_i = a_1 e_1 \quad \text{where } e_1 = \sum_i \alpha_i x_i.$$

Since $e_1' = a_1 e_1$ and $a_1 = u(e_1')$, it follows that $u(e_1) = 1$. Consequently the decomposition of each element of F as

$$x = u(x)e_1 + (x - u(x)e_1), \quad x \in F$$

[1] Although $u(S)$ is maximal among the images of S under A-maps from F to A, and although it is an ideal of A, it needn't be a *maximal ideal of* A in the algebra textbook sense. In particular, it could be all of A, e.g., if $S = F$.

gives the direct sums

$$F = Ae_1 \oplus \ker(u),$$
$$S = \mathfrak{a}_1 e_1 \oplus (\ker(u) \cap S).$$

Indeed, the decomposition gives the first sum because the intersection is trivial, and it gives the second sum as well because $u(S) = \mathfrak{a}_1$.

Induction on n gives a basis $(e_2, \cdots, e_n)$ of $\ker(u)$ and ideals $\mathfrak{a}_2 \supset \cdots \supset \mathfrak{a}_m$ of A such that $\ker(u) \cap S = \mathfrak{a}_2 e_2 \oplus \cdots \oplus \mathfrak{a}_m e_m$. If $m \geq 2$ then we need to show that $\mathfrak{a}_1 \supset \mathfrak{a}_2$. Define

$$w : F \longrightarrow A, \quad w(\textstyle\sum_{i=1}^m a_i e_i) = a_1 + a_2.$$

Then $w(S) = \mathfrak{a}_1 + \mathfrak{a}_2$ contains $\mathfrak{a}_1$ and hence is $\mathfrak{a}_1$ by maximality, and contains $\mathfrak{a}_2$.

As for uniqueness, in the context of multilinear algebra (which we will learn later in the semester) the ideals $\mathfrak{a}_i$ have a clear intrinsic description in terms of F and S, making their uniqueness clear. These ideas will be explained fully in a later writeup. For now, in the current writeup, we do two things: First, we show quickly that the *first* elementary divisor is unique, and second, we do give a uniqueness proof here for the sake of completeness. But the reader who intends to go on to the relevant multilinear algebra argument is encouraged to skip the one here. After seeing the multilinear algebra argument, the reader is invited to look at the one here and see that in fact it *is* the multilinear algebra but with no language yet at hand to express the ideas.

For the uniqueness of the first elementary divisor, *any* decomposition

$$F = \bigoplus_{i=1}^n Ae_i, \qquad S = \bigoplus_{i=1}^m \mathfrak{a}_i e_i, \qquad \mathfrak{a}_1 \supset \cdots \supset \mathfrak{a}_m$$

shows that $\mathfrak{a}_1$ is the image of S under an A-linear map from F to A and it contains all such images of S. That is, $\mathfrak{a}_1$ is *uniquely* maximal among all such images. This intrinsic characterization of $\mathfrak{a}_1$ in terms of F and S inherently connotes its uniqueness.

For the full-but-underpowered uniqueness proof, suppose we have $(e_1, \cdots, e_n)$ and $\mathfrak{a}_1 \supset \cdots \supset \mathfrak{a}_m$ as above. For each $k \in \{1, \cdots, m\}$, for every alternating A-multilinear map $w_k : F^{\oplus k} \longrightarrow A$, the submodule-image $w_k(S)$ is an ideal of A. For each k, let $u_k : F^{\oplus k} \longrightarrow A$ be such a map such that

$$u_k(S^{\oplus k}) \text{ is maximal in the ideals } \{w_k(S^{\oplus k})\}, \quad k = 1, \cdots, m.$$

Note that each u_k is described intrinsically in terms of F and S, with no reference to the basis elements e_i or the ideals $\mathfrak{a}_i$. The claim is that

$$u_1(S) = \mathfrak{a}_1, \quad u_2(S \oplus S) = \mathfrak{a}_1\mathfrak{a}_2, \quad \cdots, \quad u_m(S^{\oplus m}) = \mathfrak{a}_1\mathfrak{a}_2\cdots\mathfrak{a}_m.$$

To establish the claim, write any element of S as

$$x = \sum_{j=1}^m a_j e_j, \quad \text{each } a_j \in \mathfrak{a}_j.$$

Thus, since all the u_k are multilinear, any $u_k(x_1, \cdots, x_k)$ is a sum of terms of the form

$$a_{j_1} \cdots a_{j_k} u_k(e_{j_1}, \cdots, e_{j_k}), \quad j_1, \cdots, j_k \text{ distinct},$$

and because $\mathfrak{a}_1 \supset \cdots \supset \mathfrak{a}_m$, all such terms lie in $\mathfrak{a}_1 \cdots \mathfrak{a}_k$. This gives the containment $u_k(S^{\oplus k}) \subset \mathfrak{a}_1 \cdots \mathfrak{a}_k$. To show that the containment is equality, fix $k \in \{1, \cdots, m\}$

and consider a particular alternating A-multilinear map $w_k : F^{\oplus k} \longrightarrow A$, the determinant function on the left k-by-k square subblock of a k-by-m matrix of coefficients,

$$w_k\left(\sum_{i=1}^m c_{1i}e_i, \cdots, \sum_{i=1}^m c_{ki}e_i\right) = \det([c_{ij}]_{i,j=1,\cdots,k}).$$

Let a_i generate $\mathfrak{a}_i$ for $i = 1, \cdots, k$. Then $w_k(a_1e_1, \cdots, a_ke_k) = a_1 \cdots a_k$ generates $\mathfrak{a}_1 \cdots \mathfrak{a}_k$. Thus $\mathfrak{a}_1 \cdots \mathfrak{a}_k$ is not a proper superideal of $u_k(S^{\oplus k})$.

Since $\mathfrak{a}_1$ and $\mathfrak{a}_1\mathfrak{a}_2$ are intrinsic to F and S, so is $\mathfrak{a}_2$ since the PID A is a UFD. Continuing in this fashion shows that all of $\mathfrak{a}_1, \cdots \mathfrak{a}_m$ are intrinsic to F and S. □

After presenting the standard consequences of Theorem 3.3 in the next section, we will comment at some length on what has happened in this writeup—and what hasn't happened, which can easily get lost—in the last section. For now we say briefly that the proof as given is not an argument that anybody would think up from scratch while investigating the phenomena. Rather, it is the sort of argument that can be found *post hoc* once one already knows that the theorem is true and therefore inevitably has a light, graceful verification. Here *light, graceful* applies only to the proof of the first part of the theorem; as discussed above, the good proof of uniqueness is set in a multilinear algebra context that we don't yet know.

4. Consequences

Corollary 4.1. *Let A be a PID. Let M be a finitely generated A-module. Then*

(a) *There exist nonnegative integers m and r, and ideals*

$$\mathfrak{a}_1 \supset \cdots \supset \mathfrak{a}_m$$

such that

$$M \approx A/\mathfrak{a}_1 \oplus \cdots \oplus A/\mathfrak{a}_m \oplus A^{\oplus r}.$$

(b) *If M is torsion-free then M is free.*

(c) *Furthermore, M decomposes as*

$$M \approx \bigoplus M_i, \quad \text{each } M_i = \begin{cases} A/\mathfrak{p}^e A & \mathfrak{p}_i \subset A \text{ maximal, } e \geq 1, \\ A. \end{cases}$$

The decomposition of M into such factors is uniquely determined by the integers m and r, and by the ideals $\mathfrak{a}_i$.

Proof. (a) M is the image of a free A-module F of finite rank under a map whose kernel S is a submodule of F. Thus

$$\begin{aligned} M &\approx \frac{Ae_1 \oplus \cdots \oplus \ Ae_m \oplus Ae_{m+1} \oplus \cdots \oplus Ae_n}{\mathfrak{a}_1e_1 \oplus \cdots \oplus \mathfrak{a}_me_m} \\ &\approx A/\mathfrak{a}_1 \oplus \cdots \oplus A/\mathfrak{a}_m \oplus A^{\oplus(n-m)}. \end{aligned}$$

(b) By (a), the torsion-free case is $M \approx A^{\oplus t}$.

(c) Factor each $\mathfrak{a}_i$ as a product $\prod_j \mathfrak{p}_{ij}^{e_{ij}}$, and apply the Sun-Ze theorem to each $A/\mathfrak{a}_i$. □

Regarding item (b) of the corollary, the standard example of a nonfree torsion-free module over a PID is the $\mathbb{Z}$-module $\mathbb{Q}$, mentioned earlier in the writeup.

5. Comments

Given a PID A and a finitely generated A-module M, a number of questions present themselves:

(1) We want to know whether M is the third joint of a *short exact sequence* whose middle joint is free,

$$0 \longrightarrow S \longrightarrow F \longrightarrow M \longrightarrow 0, \quad F \text{ free.}$$

The answer is *yes*. Every A-module M is the image of a free module F, regardless of whether the commutative ring-with-unit A is a PID and regardless of whether M is finitely generated. The first joint S of the short exact sequence is the kernel of the map from F onto M.

(2) Granting such a short exact sequence, we want a *presentation*

$$M = (\mathcal{E} \,|\, \mathcal{R})$$

where $\mathcal{E}$ is a minimal set of generators and $\mathcal{R}$ is a minimal set of linear combinations of $\mathcal{E}$, and the linear combinations of $\mathcal{E}$ that are 0 in M are precisely the linear combinations of $\mathcal{R}$. There should be no nontrivial linear relations among the elements of $\mathcal{R}$.

Theorem 3.3(a) says that the submodule S in the short exact sequence is also free, making the short exact sequence a *one-stage free resolution* of M. So a presentation arises by taking $\mathcal{E}$ as the basis of a minimal-rank free F that maps to M and taking $\mathcal{R}$ as the corresponding generators of S. In fact, this holds regardless of whether M is finitely generated—the induction argument in the proof readily becomes transfinite induction.

(3) Granting a presentation, we want to know whether there exists a presentation in some standard form.

Yes. This is the content of Corollary 4.1(a) and (c), whose proofs follow from Theorem 3.3(b) and then the Sun-Ze theorem.

(4) Granting that there exists a presentation in standard form, we want a useful algorithm to put any presentation into standard form.

The argument here does not speak directly to this question. In the proof of Theorem 3.3, the choice of an ideal that is maximal in a set of ideals is not constructive.

Although an elaboration of the matrix-based approach that we saw earlier for abelian groups does scale up to finitely generated modules over a PID, the scaled-up version is only as algorithmic as the arithmetic of the PID. If the PID is Euclidean then the abelian group method works essentially verbatim, and if the PID is not Euclidean but has some other feature that makes its arithmetic easy then the more general matrix method is at least applicable even though it is more complicated. This handout does not present the matrix method because it is not necessarily algorithmic, and because it addresses only this question (4) but not the other questions here. Its many algorithm-like details can distract a student from its actual non-algorithmic-ness, and from the rest of the issues in play.

(5) Granting a presentation in standard form, we want to know whether the integers m and r and the ideals $\mathfrak{a}_i$ are unique.

Yes, but we have **not** *shown this.* The point is that Corollary 4.1(a) does not say that the integers m, r and the $\mathfrak{a}_i$ are uniquely determined by M, only that they are uniquely determined by the resolution of M, i.e., by F and S. Hence Corollary 4.1(c) does not say that the factors $A/\mathfrak{p}^e A$ and A are uniquely determined by M.

Of course they are, but showing this requires further work. A standard elementary approach is illustrated in Gallian's treatment of the abelian group case. Another approach is to consider the possible variations among resolutions of M, in particular among minimal resolutions, and then to show that those variations do not affect the invariants.