Modules over principal ideal domains: *the Smith normal form approach*

1 Preliminaries from linear algebra

1.1 Generalities about equivalence of matrices

Definition 1.1. Two matrices $P, Q \in A^{n\times m}$ over a commutative ring A are *equivalent*, denoted $P \approx Q$, if and only if there are invertible matrices $S \in A^{n\times n}$ and $R \in A^{m\times m}$ such that $Q = SPR$.

Remark 1.2. By definition, a matrix $S \in A^{n\times n}$ is *invertible* if and only if there is a matrix $T \in A^{n\times n}$ such that ST and TS are identity matrices. Using the *adjoint matrix construction* one can show that a matrix S is invertible if and only if $\det S$ is invertible in A.

Example 1.3. The following row operations yield equivalent matrices (and so do the analogous column operations):

1. Swapping two rows.
2. Adding a multiple of a row to another.
3. Scaling a row by an invertible element of the ring.

The operation "substitute row i by a times row i plus b times row j", where $i \neq j$, is only guaranteed to yield an equivalent matrix if a is invertible. This is a difference to linear algebra over fields, where it would be sufficient that a is not zero.

Example 1.4. Let $P \approx P'$ and $Q \approx Q'$. Then we have the equivalence

$$\begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix} \approx \begin{pmatrix} P' & 0 \\ 0 & Q' \end{pmatrix}$$

of block diagonal matrices.

Definition 1.5. Let $P \in A^{n\times m}$ be a matrix over a commutative ring A. The *cokernel* of P is the A-module

$$\operatorname{cok} P := A^n / \operatorname{im} P.$$

Proposition 1.6. *Equivalent matrices have isomorphic cokernels: Let $P, Q \in A^{n\times m}$ be matrices over a commutative ring A. If $P \approx Q$, then $\operatorname{cok} P \cong \operatorname{cok} Q$.*

Proof. If $P \approx Q$, there are invertible matrices $S \in A^{n\times n}$, $R \in A^{m\times m}$ such that $Q = SPR$. One can check that mapping

$$\operatorname{cok} P \longrightarrow \operatorname{cok} Q, [x] \longmapsto [Sx]$$

results in a well-defined module isomorphism with inverse given by $[y] \mapsto [S^{-1}y]$. (Well-definedness is seen as follows: If $x \in \operatorname{im} P$, so $x = Pv$ for some $v \in A^m$, then $Sx = SPv = SPRR^{-1}v = QR^{-1}v \in \operatorname{im} Q$.) □

Proposition 1.7. *Let $D \in A^{n \times m}$ be a diagonal matrix over a commutative ring A. Then the cokernel of D is canonically isomorphic to the direct sum*

$$X := A/(d_1) \oplus \ldots \oplus A/(d_n),$$

where $d_1, \ldots, d_r$ with $r = \min\{n, m\}$ are the diagonal entries of D and $d_{r+1}, \ldots, d_n$ are zero.

Proof. An isomorphism is given by

$$\operatorname{cok} D \longrightarrow X, [(a_1, \ldots, a_n)] \longmapsto ([a_1], \ldots, [a_n]). \qquad \square$$

1.2 Equivalence operations related to greatest common divisors

In this subsection, let the base ring A be a Bézout domain according to the following definition:

Definition 1.8. A *Bézout domain* is an integral domain in which every finitely generated ideal is principal.

Example 1.9. Let $x, y \in A$. Let d be a generator of the ideal (x, y), i. e. a gcd of x and y. Let $d = sx + ty$, $x = dx'$, $y = dy'$. Then we have the equivalence

$$\begin{pmatrix} x & y \end{pmatrix} \approx \begin{pmatrix} d & 0 \end{pmatrix}$$

of (1×2)-matrices (and similarly with (2×1)-matrices). If $d = 0$, this equivalence is trivial in view of $x = y = 0$. If d is regular, this equivalence is witnessed by the identity

$$\begin{pmatrix} d & 0 \end{pmatrix} = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} s & -y' \\ t & x' \end{pmatrix},$$

where the (2×2)-matrix is indeed invertible as its determinant is $sx' + ty' = 1$. (Over Euclidean domains, this equivalence can also be witnessed by a sequence of the elementary column transformations of Example 1.3.)

Example 1.10. By iterating Example 1.9, every matrix is equivalent to a matrix of the form

$$\begin{pmatrix} d & 0 & \cdots & 0 \\ \star & \star & \cdots & \star \\ \vdots & \vdots & & \vdots \\ \star & \star & \cdots & \star \end{pmatrix},$$

where d is a gcd of the original entries of the first row. Similarly, we can arrange for the first column to consist entirely of zeros with the exception of the first entry.

Example 1.11. Let $x, y \in A$. Let d be a generator of the ideal (x, y), i. e. a gcd of x and y. Let $d = sx + ty$, $x = dx'$, $y = dy'$. Let $p = xy' = x'y$ be the least common multiple of x and y. Note that $d \mid p$. Then we have the equivalence

$$\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix} \approx \begin{pmatrix} d & 0 \\ 0 & p \end{pmatrix},$$

so we can always arrange the first diagonal element to divide the second. If $d = 0$, this equivalence is trivial in view of $x = y = d = p = 0$. If d is regular, this equivalence is witnessed by the identity

$$\begin{pmatrix} s & t \\ -y' & x' \end{pmatrix} \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix} \begin{pmatrix} 1 & -ty' \\ 1 & sx' \end{pmatrix} = \begin{pmatrix} d & 0 \\ 0 & p \end{pmatrix},$$

where the two (2×2)-matrices are invertible as their determinant is one.

Example 1.12. By iterating Example 1.11, every diagonal matrix is equivalent to a diagonal matrix where successive diagonal entries divide each other.

2 The Smith normal form

Definition 2.1. A matrix $P = (a_{ij})_{ij} \in A^{n \times m}$ is *in Smith normal form* if and only if all off-diagonal entries are zero and successive diagonal entries divide each other: $a_{11} \mid a_{22} \mid \ldots \mid a_{rr}$ where $r = \min\{n, m\}$.

Theorem 2.2. *Let A be a principal ideal domain. Let $P \in A^{n \times m}$ be a matrix. Then P is equivalent to a matrix in Smith normal form.*

Proof. By Exercise 1.12, it suffices to show that P is equivalent to a diagonal matrix. We proceed by induction. The cases $n = 0$ and $m = 0$ are trivial. Let $n \geq 1$ and $m \geq 1$. By Example 1.10, the matrix P is equivalent to a matrix whose first row consists entirely of zeros, except at the top left. Applying Example 1.10 again, but transposed, we can then arrange for the first column to contain only zeros, except at the top left.

However, this second transformation might destroy the zeros obtained in the first step. We can fix this issue by applying Example 1.10 yet again for the first row—at the cost of destroying the zeros obtained in the first column. Continuing in this fashion, we would obtain a chain of equivalences of the form

$$P \approx \underbrace{\begin{pmatrix} d_1 & 0 & \cdots & 0 \\ \star & \star & \cdots & \star \\ \vdots & \vdots & & \vdots \\ \star & \star & \cdots & \star \end{pmatrix}}_{=: P_1} \approx \underbrace{\begin{pmatrix} d_2 & \star & \cdots & \star \\ 0 & \star & \cdots & \star \\ \vdots & \vdots & & \vdots \\ 0 & \star & \cdots & \star \end{pmatrix}}_{=: P_2} \approx \underbrace{\begin{pmatrix} d_3 & 0 & \cdots & 0 \\ \star & \star & \cdots & \star \\ \vdots & \vdots & & \vdots \\ \star & \star & \cdots & \star \end{pmatrix}}_{=: P_3} \approx \cdots,$$

where d_1 is a gcd of the entries of the first row of P, d_2 is a gcd of the entries of the first column of P_1, d_3 is a gcd of the entries of the first row of P_2 and so on. Are we stuck?

No, for $(d_1) \subseteq (d_2) \subseteq (d_3) \subseteq \ldots$ is an ascending sequence of ideals. As the base ring A is Noetherian, this sequence stabilizes, in particular there is an index i such that $(d_i) = (d_{i+1})$. So d_i is a gcd of the entries of both the first row of P_i and the first column of P_i. So the elementary row or the elementary column operations from Example 1.3 suffice to establish the equivalence

$$P \approx P_i \approx \begin{pmatrix} d_i & 0 & \cdots & 0 \\ 0 & \star & \cdots & \star \\ \vdots & \vdots & & \vdots \\ 0 & \star & \cdots & \star \end{pmatrix}.$$

The lower right block is equivalent to a diagonal matrix by the induction hypothesis, and Example 1.4 allows us to conclude. □

3 Modules over principal ideal domains

We will use the following facts: Principal ideal domains are Noetherian, and a ring is Noetherian iff every ideal is finitely generated.

Lemma 3.1. *Let A be a Noetherian ring. Let $n \in \mathbb{N}$. Then every submodule $U \subseteq A^n$ is finitely generated.*

Proof. By induction on n. In the base case $n = 0$, the submodule $U = \{0\}$ is generated by the empty family.

In the case $n \geq 1$, the submodule

$$N := \{(a_1, \ldots, a_{n-1}, 0) \,|\, a_1, \ldots, a_{n-1} \in A\} \subseteq A^n$$

is isomorphic (by the module isomorphism $i : (a_1, \ldots, a_{n-1}, 0) \mapsto (a_1, \ldots, a_{n-1})$) to A^{n-1}. So $U \cap N$ can be regarded as a submodule of A^{n-1} and is hence, by the induction hypothesis, finitely generated (more precisely, the submodule $i[U \cap N] \subseteq A^{n-1}$ is finitely generated and $U \cap N$ is isomorphic to $i[U \cap N]$).

The isomorphism theorem implies that the injective module homomorphism

$$U/(U \cap N) \longrightarrow A, [(a_1, \ldots, a_n)] \longmapsto a_n$$

induces an isomorphism of $U/(U \cap N)$ with its image. This image is an ideal of A and hence finitely generated.

As both $U \cap N$ and $U/(U \cap N)$ are finitely generated, Exercise 8.3(a) implies that U is finitely generated as well. □

Corollary 3.2. *Let M be a finitely generated module over a Noetherian ring A. Then M is isomorphic to the cokernel of a matrix.*

Proof. As M is finitely generated, there is a finite generating family $(x_1, \ldots, x_n)$ for M. By one of the isomorphism theorems, the induced surjective module homomorphism

$$\begin{array}{rrcl} f: & A^n & \longrightarrow & M \\ & (a_1, \ldots, a_n) & \longmapsto & \sum_{i=1}^n a_i x_i \end{array}$$

gives rise to an isomorphism

$$A^n/\ker(f) \to M.$$

By the lemma, there is a finite generating family $(y_1, \ldots, y_m)$ for the submodule $\ker(f) \subseteq A^n$. We conclude by observing that $\ker(f) = \operatorname{im}(P)$ with $P = (y_1 \mid \ldots \mid y_m) \in A^{n \times m}$ the matrix which has the y_i as its columns. □

Theorem 3.3. *Let M be a finitely generated module over a principal ideal domain A. Then M is isomorphic to a direct sum*

$$M \cong A/(d_1) \oplus \ldots \oplus A/(d_n)$$

for some number $n \in \mathbb{N}$ and ring elements $d_1, \ldots, d_n$ such that $d_1 \mid \ldots \mid d_n$.

Proof. By Corollary 3.2, the module M is isomorphic to the cokernel of a matrix P. By Theorem 2.2, this matrix is equivalent to a matrix D in Smith normal form. By Proposition 1.6, the cokernel of P is isomorphic to the cokernel of D. By Proposition 1.7, the cokernel of D is isomorphic to a direct sum of the desired form:

$$M \cong \operatorname{cok} P \cong \operatorname{cok} D \cong A/(d_1) \oplus \ldots \oplus A/(d_n). \qquad \square$$

3.1 Uniqueness

Except for superfluous summands of the form $A/(d) = 0$ where d is a unit of A in the decomposition provided by Theorem 3.3, their number n and the ideals (d_i) are uniquely determined by M. This fact is a consequence of the following three observations over general commutative rings.

Proposition 3.4. *Let A be a commutative ring. Let $n \in \mathbb{N}$. Then the A-module A^n can be generated by a family of length less than n only in the case $A = 0$.*

Proof. This statement is a ring analogue of the vector space result that a vector space of dimension n cannot be generated by a family of length less than n and will be proven later in the course. □

Lemma 3.5. *Let A be a commutative ring. Let $I_1 \subseteq I_2 \subseteq \ldots \subseteq I_n \subsetneq (1)$ be proper ideals of A. Then the number n is uniquely determined as the minimal length of a generating family for the module*

$$M = A/I_1 \oplus \ldots \oplus A/I_n.$$

Proof. A generating family of length n is $(b_1, \ldots, b_n)$, where $b_i = (0, \ldots, 0, [1], 0, \ldots, 0)$ with the entry $[1]$ at position i. So the minimal length of a generating family is at most n.

For the converse inequality, assume that M can be generated by a family of length less than n. Then so can be the module

$$A/I_n \oplus \ldots \oplus A/I_n = (A/I_n)^n.$$

Indeed, if $([x_1], \ldots, [x_k])$ is a generating family of M with $x_1, \ldots, x_k \in A$, then $([x_1], \ldots, [x_k])$ is a generating family of $(A/I_n)^n$, where the equivalence class brackets now all denote equivalence classes in A/I_n. The same family also generates $(A/I_n)^n$ as an A/I_n-module. But this is only possible if $A/I_n = 0$ which amounts to $I_n = (1)$, a contradiction. □

For the proof of the following proposition it is useful to introduce, for an ideal $J \subseteq A$ and an element $x \in A$, the notation $(J : x) := \{a \in A \,|\, ax \in J\}$. The set $(J : x)$ is a superset of J and again an ideal of A. It is improper if and only if $x \in J$.

Proposition 3.6. *Let M be a module over a commutative ring A. Assume that M is isomorphic to a module of the form*

$$M \cong A/I_1 \oplus \ldots \oplus A/I_n$$

for some ideals $I_1 \subseteq I_2 \subseteq \ldots \subseteq I_n$. Then I_k is uniquely determined by the description

$$I_k = \{x \in A \,|\, xM \text{ has a generating family of length less than } k\}.$$

Proof. Let $x \in A$. For an arbitrary ideal $J \subseteq A$, the canonical surjective module homomorphism $A \to x(A/J)$ given by $a \mapsto x[a]$ has kernel $(J : x)$ and hence induces an isomorphism $A/(J : x) \cong x(A/J)$. Applying this observation summandwise, we obtain an isomorphism

$$xM \cong A/(I_1 : x) \oplus \ldots \oplus A/(I_n : x).$$

By Lemma 3.5, we obtain the following chain of equivalences: The module xM has a generating family of length less than k iff $(I_k : x) = \ldots = (I_n : x) = (1)$, iff $x \in I_k, \ldots, I_n$, iff $x \in I_k$. □

References

[1] Anonymous. Proving uniqueness in the structure theorem for finitely generated modules over a principal ideal domain. math.SE, 2019.

[2] Ray Mines, Fred Richman, Witenburg. A course in constructive algebra. Springer, 1988.

[3] Marc Nieper-Wißkirchen. Abstrakte Galois-Theorie, Springer. 2021.

[4] Marc Nieper-Wißkirchen. Lineare Algebra I, Vorlesungsskript. 2009.

[5] Richard Stanley. Smith normal form and combinatorics. 2016.