8. Modules over PID, part II. Smith Normal Form.

8.1. **Proof of the Smith Normal Form theorem.**

Theorem (Smith Normal Form (SNF)). *Let R be a PID, $k, n \in \mathbb{N}$ and $A \in Mat_{k\times n}(R)$. Then A can be written as $A = CDB$ where $B \in GL_n(R)$, $C \in GL_k(R)$ and $D \in Mat_{k\times n}(R)$ is equal to* $\begin{pmatrix} \begin{matrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & a_m \end{matrix} & 0 \\ 0 & 0 \end{pmatrix}$ *for some $m \le \min\{n, k\}$ and nonzero $a_1, \dots, a_m$ with $a_1 \mid a_2 \mid \dots \mid a_m$. The matrix D is called the Smith Normal Form of A. Its entries $a_1, \dots, a_m$ are uniquely determined up to multiplication by units.*

Today we will prove the existence part of this theorem. For simplcity, we will present the proof under the extra assumption that R is a Euclidean domain (the argument is the general case is similar).

Let us introduce the following operations on the set $Mat_{k\times n}(R)$:

1. $\mathcal{E}_{ij}(r)$, $i \neq j$: add j^{th} row multiplied by r to i^{th} row
2. $\mathcal{F}_{ij}(r)$, $i \neq j$: flip i^{th} and j^{th} rows
3. $\mathcal{E}'_{ij}(r)$, $i \neq j$: add i^{th} column multiplied by r to j^{th} column
4. $\mathcal{F}'_{ij}(r)$, $i \neq j$: flip i^{th} and j^{th} columns

Operations (1) and (2) will be called row reductions and operations (3) and (4) column reductions.

It is easy to see that

- $\mathcal{E}_{ij}(r)$ = multiplication on the left by $E_{ij}(r)$ = the matrix with 1's on the diagonal, r at the (i, j)-entry and 0's everywhere else
- $\mathcal{F}_{ij}$ = multiplication on the left by F_{ij} = the matrix obtained by flipping i^{th} and j^{th} rows of the identity matrix
- $\mathcal{E}'_{ij}(r)$ = multiplication on the right by $E_{ij}(r)$
- $\mathcal{F}_{ij}$ = multiplication on the right by F_{ij}

Claim. *Using operations (1)-(4) one can reduce any $k \times n$ matrix A to the form $diag_{k,n}(a_1, \dots, a_m)$=* $\begin{pmatrix} \begin{matrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & a_m \end{matrix} & 0 \\ 0 & 0 \end{pmatrix}$ *with $a_1 \mid a_2 \mid \dots \mid a_m$.*

Suppose we proved the claim and A is reduced to $D = diag_{k,n}(a_1, \dots, a_m)$ using p row reductions and q column reductions for some p and q. Then there exist matrices $C_1, \dots, C_p, B_1, \dots, B_q$ each of which is equal to $E_{ij}(r)$ or F_{ij} for some i, j & r s.t.

$$C_p \dots C_1 A B_1 \dots B_q = D.$$

All B_k's and C_k's are clearly invertible, so $A = CDB$ where $C = (C_p \dots C_1)^{-1}$ and $B = (B_1 \dots B_q)^{-1}$, as desired in the SNF Theorem.

Proof of the Claim. Recall that we consider the case R=Euclidean domain, and let N be a Euclidean norm on R.
Initial step: Find nonzero entry of A with smallest possible norm and move it to position (1,1) using flips, call it a_1.

Case 1: All entries of A are divisible by a_1.

Then using operations $\mathcal{E}_{1j}(r)$ and $\mathcal{E}'_{j1}(r)$, that is, subtracting suitable multiples of the first row (resp. column) from other rows (resp. columns), we can put zeroes everywhere in the first row and first columnn except for $(1,1)$-entry, so our matrix is of the form $\begin{pmatrix} a_1 & 0 \\ 0 & \widetilde{A} \end{pmatrix}$. By induction the matrix $\widetilde{A}$ can be put into SNF using reductions, so A can be reduced to the form $\begin{pmatrix} \begin{matrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & a_m \end{matrix} & 0 \\ 0 & 0 \end{pmatrix}$ with $a_2 \mid a_3 \mid \dots \mid a_m$. It remains to show that $a_1 \mid a_2$.
By assumption a_1 divides all entries of A. When we apply a row or column reduction, the entries of the new matrix are R-linear combinations of the entries of the old matrix. Thus a_1 divides all entries of the matrix $diag_{k \times n}(a_1, \dots, a_m)$, and in particular $a_1 \mid a_2$.

Case 2: One of the entries of A is not divisible by a_1, call it *bad entry.*

Subcase 1: Bad entry exists in row_1: $a_1 \nmid a_{1j}$ for some j. Then write $a_{1j} = qa_1 + r$ with $0 < N(r) < N(a_1)$. After subtracting the first column multiplied by q from the j^{th} column, we get r in the position $(1, j)$. Then we go back to the initial step. The process cannot go forever since $N(r) < N(a_1)$ and N has values in $\mathbb{Z}_{\geq 0}$.

Subcase 2: Bad entry exists in $column_1$. This is analogous to Subcase 1.

Subcase 3: All entries in row_1 and $column_1$ are divisible by a_1. Then as in Case 1 we reduce A to the form $\begin{pmatrix} a_1 & 0 \\ 0 & \widetilde{A} \end{pmatrix}$. If $\widetilde{A}$ has bad entry a_{ij}, add i^{th} row to the first row, which puts us back in Subcase 1. $\square$

8.2. **Using SNF Theorem for finding compatible bases in the submodule theorem.**

Problem. *Let R be a Euclidean domain, F a f.g. free R-module and N a submodule of F. Want: find (algorithmically) a basis $\{y_1, \ldots, y_n\}$ of F and elements $a_1, \ldots, a_m \in R$ with $a_1 \mid a_2 \mid \ldots \mid a_m$ and $m \leq n$ s.t. $\{a_1 y_1, \ldots, a_m y_m\}$ is a basis for N. The bases of F and N with this property will be called compatible.*

Of course, the existence of such bases is guaranteed by Theorem 7.1

Example 8.1: Let $R = \mathbb{Z}$, $F = \mathbb{Z}e_1 \oplus \mathbb{Z}e_2$ (the free $\mathbb{Z}$-module with basis $\{e_1, e_2\}$) and N the submodule of F generated by z_1, z_2, z_3 where $z_1 = 7e_1 + 3e_2$, $z_2 = 3e_1 + 7e_2$ and $z_3 = 4e_1 + 4e_2$.

Let us find compatible bases for F and N. The initial basis for F is $\{e_1, e_2\}$, and the initial generating set for N is $\{z_1, z_2, z_3\}$, and we can write

$$\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = A \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \text{ where } A = \begin{pmatrix} 7 & 3 \\ 3 & 7 \\ 4 & 4 \end{pmatrix}$$

Now let us put A into SNF using row and column reductions. As can be seen from the proof of Theorem 7.1, each row reduction represents a change of a generating set of N, and each column reduction represents a change of basis of F, and at each stage of our process we have equality

$$\begin{pmatrix} z_1' \\ z_2' \\ z_3' \end{pmatrix} = A' \begin{pmatrix} e_1' \\ e_2' \end{pmatrix} \qquad (***)$$

where $\{e_1', e_2'\}$ is the current basis of F, $\{z_1', z_2', z_3'\}$ is the current generating set of N and A' is the current matrix.

Note that we only need to keep track of how the basis of F changes since the current generating set of N is determined by the current basis of F and the current matrix via (***). Because of this, we shall try to use as few column reductions as possible (since row reductions do not change the basis of F).

Let us now implement this algorithm in our example:

$$\begin{pmatrix} 7 & 3 \\ 3 & 7 \\ 4 & 4 \end{pmatrix} \xrightarrow{\mathcal{E}_{1,2}(-2)} \begin{pmatrix} 1 & -11 \\ 3 & 7 \\ 4 & 4 \end{pmatrix} \xrightarrow{\mathcal{E}_{2,1}(-3) \ \& \ \mathcal{E}_{3,1}(-4)} \begin{pmatrix} 1 & -11 \\ 0 & 40 \\ 0 & 48 \end{pmatrix} \xrightarrow{\mathcal{E}'_{1,2}(11)}$$
$$\begin{pmatrix} 1 & 0 \\ 0 & 40 \\ 0 & 48 \end{pmatrix} \xrightarrow{\mathcal{E}_{3,2}(-1)} \begin{pmatrix} 1 & 0 \\ 0 & 40 \\ 0 & 8 \end{pmatrix} \xrightarrow{\mathcal{F}_{2,3}} \begin{pmatrix} 1 & 0 \\ 0 & 8 \\ 0 & 40 \end{pmatrix} \xrightarrow{\mathcal{E}_{3,2}(-5)} \begin{pmatrix} 1 & 0 \\ 0 & 8 \\ 0 & 0 \end{pmatrix}.$$

So, we found that $a_1 = 1$ and $a_2 = 8$.

Our reduction of A to SNF involved only one column reduction (third transition), so we only need to see how the basis changed at that step. The new basis $\{e_1', e_2'\}$ satisifes the matrix equation:

$$\begin{pmatrix} 1 & -11 \\ 0 & 40 \\ 0 & 48 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 40 \\ 0 & 48 \end{pmatrix} \begin{pmatrix} e_1' \\ e_2' \end{pmatrix},$$

and so $e_1' = e_1 - 11e_2$ and $e_2' = e_2$.

Thus, if we let $y_1 = e_1 - 11e_2$ and $y_2 = e_2$, then $\{y_1, y_2\} = \{e_1 - 11e_2, e_2\}$ is a basis of F and $\{y_1, 8y_2\} = \{e_1 - 11e_2, 8e_2\}$ is a basis of N.

Verification: Let us check the asnwer (in case we made a computational mistake). It is clear that $\{e_1 - 11e_2, e_2\}$ is a basis of F, so we only need to check that $\{e_1 - 11e_2, 8e_2\}$ is a basis of N. We need to check that

(i) $e_1 - 11e_2$ and $8e_2$ lie in N

(ii) Initial generators of N are linear combinations of $e_1 - 11e_2$ and $8e_2$

(iii) $e_1 - 11e_2$ and $8e_2$ are linearly independent over $\mathbb{Z}$

We have

(i) $e_1 - 11e_2 = z_1 - 2z_2 \in N$ and $8e_2 = z_2 + z_3 - z_1 \in N$

(ii) $z_1 = 7(e_1 - 11e_2) + 10 \cdot 8y_2$, $z_2 = 3(e_1 - 11e_2) + 5 \cdot 8y_2$, $z_3 = 4(e_1 - 11e_2) + 6 \cdot 8y_2$,

(iii) is clear