Math 361 lecture for Wednesday, Week 8

Smith normal form

Motivation: Let $\mathfrak{a}$ be a nonzero ideal in a number $\mathfrak{O}_K$. It turns out that the structure of the quotient ring $\mathfrak{O}_K/\mathfrak{a}$ is interesting and useful. Our goal is to reduced the problem of determining its structure to finding a canonical form for an integer matrix using elementary row and column operations.

We have seen that $\mathfrak{O}_K$ is a free $\mathbb{Z}$-module of rank n, that is, there exist $\alpha_1, \ldots, \alpha_n \in \mathfrak{O}_K$ such that we get a $\mathbb{Z}$-module isomorphism $f\colon \mathfrak{O}_K \to \mathbb{Z}^n$ by sending $\alpha_i \to e_i$ and extending linearly, i.e., $\sum_{i=1}^n a_i\alpha_i \mapsto (a_1, \ldots, a_n)$. Since $\mathfrak{O}_K$ is Noetherian, $\mathfrak{a}$ is also finitely generated. By the structure theorem for finitely generated abelian groups (i.e., finitely generated $\mathbb{Z}$-modules), which we will prove next time, it will follow that there is an isomorphism $g\colon \mathfrak{a} \simeq \mathbb{Z}^n$. We then get a commutative diagram[1] with exact rows:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathfrak{a} & \longrightarrow & \mathfrak{O}_K & \longrightarrow & \mathfrak{O}_K/\mathfrak{a} & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle g}\wr & & \downarrow{\scriptstyle f}\wr & & \vdots & & \\ 0 & \longrightarrow & \mathbb{Z}^n & \xrightarrow{M} & \mathbb{Z}^n & \longrightarrow & \mathbb{Z}^n/\operatorname{im}(M) & \longrightarrow & 0, \end{array}$$

where M is an $n\times n$ matrix with integer entries. The vertical mapping on the right-hand side is given by $\overline{\alpha} \mapsto \overline{f(\alpha)}$. Since the diagrams commute and the rows are exact, it turns out the this mapping is a well-defined isomorphism. Using integer row an column operations, we will see that we can replace M with a diagonal matrix from which the structure of $\mathbb{Z}^n/\operatorname{im}(M)$ will be apparent.

Diagonalization of integer matrices. Let M be an $m\times n$ matrix with integer coefficients. The matrix M determines a $\mathbb{Z}$-linear mapping $\mathbb{Z}^n \to \mathbb{Z}^m$ via $v \mapsto Mv$. The *cokernel* of M (or its associated mapping) is the $\mathbb{Z}$-module

$$\operatorname{cok}(M) := \mathbb{Z}^m/\operatorname{im} M.$$

Recall that $\operatorname{im}(M)$ is the same as the column space of M. So the cokernel of M is the set of integer vectors $(a_1, \ldots, a_m)$ for which we add vectors as usual, but such that any vector that is a column of M is thought of as the zero vector.

Example 1.

- Let $M = [5]$, a 1×1 matrix. Then $\operatorname{cok}(M) = \mathbb{Z}/5\mathbb{Z}$.

[1] The diagram being commutative means that if there are two ways to get from one space in the diagram to another by composing functions that appear in the diagram, then those two compositions of functions are equal.

- Let $M = \mathrm{diag}(2,3)$, a 2×2 diagonal matrix. Then

$$\mathrm{cok}(M) = \mathbb{Z}^2 / \operatorname{Span}\left\{ \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \end{pmatrix} \right\} \xrightarrow{\sim} \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}$$
$$(a,b) \mapsto (a \bmod 2, b \bmod 3).$$

  Setting $(2,0)$ equal to $(0,0)$ in $\mathbb{Z}^2$, just means we can work modulo 2 in the first coordinate. Similarly, we can work modulo 3 in the second coordinate.

- Let $M = \mathrm{diag}(0,0,1,2,3)$. Then

$$\mathrm{cok}(M) \simeq \mathbb{Z}/0\mathbb{Z} \oplus \mathbb{Z}0\mathbb{Z} \oplus \mathbb{Z}/1\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z} \xrightarrow{\sim} \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}$$
$$(a,b,c,d,e) \mapsto (a,b,d,e).$$

  Here, we use the fact that $\mathbb{Z}/0\mathbb{Z} = \mathbb{Z}$ and $\mathbb{Z}/1\mathbb{Z} = \{0\}$. For instance, in $\mathrm{cok}(M)$, the third coordinate is always equivalent to 0 modulo 1, hence, we can drop that coordinate in our isomorphism.

Definition 2. The *integer row (resp., column) operations* on an integer matrix consist of the following:

1. swapping two rows (resp., columns);
2. negating a row (resp., column);
3. adding one row (resp., column) to a different row (resp., column).

Claim. By performing integer row and column operations, the matrix M can be transformed into a diagonal matrix D, i.e., $D_{ij} = 0$ for $i \neq j$. To make the final form unique, one may insist that the diagonal elements satisfy $D_{i,i} | D_{i+1,i+1}$ for all i. Start with the the identity matrix I_m and perform all of the same row operations on I_m as used in the reduction of M to D to create a matrix P. Similarly, start with I_n and perform the same column operations on it as used in the reduction of M to D to create a matrix Q. Then both P and Q have inverses that are integer matrices (equivalently, $\det(P) = \pm 1$ and $\det(Q) = \pm 1$), and

$$PMQ = D.$$

Roughly, the algorithm for reducing M to a diagonal matrix goes like this: First, permute rows, if necessary, to ensure that some nonzero entry is in the first row. Next, use column operations to put the gcd of the elements in the first row into the $1,1$-position of the matrix. Then use the first column to make the other entries in the first row equal to 0. Next, use row operations to put the gcd of the first column into the $1,1$-position, Then use row operations to make the other entries in the first column equal to 0. By this time, you may have put nonzero entries in the first row again. Repeat. Eventually, every entry in the first row and column besides the $1,1$-entry will be 0. The first row and column are now completely processed. Proceed inductively, using rows and columns operations not involving rows and columns not previously processed.

In terms of mappings, the above process yields the commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathbb{Z}^n & \xrightarrow{M} & \mathbb{Z}^m & \longrightarrow & \operatorname{cok}(M) & \longrightarrow & 0 \\ & & {\scriptstyle Q^{-1}}\downarrow\wr & & {\scriptstyle P}\downarrow\wr & & \vdots & & \\ 0 & \longrightarrow & \mathbb{Z}^n & \xrightarrow{D} & \mathbb{Z}^m & \longrightarrow & \operatorname{cok}(D) & \longrightarrow & 0, \end{array}$$

We have an isomorphism $\operatorname{cok}(M) \mapsto \operatorname{cok}(D)$ induced by P as follows: $\overline{v} \mapsto \overline{Pv}$. Since D is diagonal, we may easily write $\operatorname{cok}(D)$ as a product of cyclic groups of the form $\mathbb{Z}/n_i\mathbb{Z}$ for various integers n_i. (Going back to our motivating goal, note how the above procedure will allow us to identify $\mathfrak{O}_K/\mathfrak{a}$ as a product of cyclic groups.)

Example 3. We illustrate the procedure using the following matrix.

$$M = \begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 4 & -1 & -2 \\ -1 & -1 & 3 & -1 \\ 0 & -2 & -1 & 3 \end{pmatrix}.$$

Perform integer row and column operations to diagonalize M:

$$\begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 4 & -1 & -2 \\ -1 & -1 & 3 & -1 \\ 0 & -2 & -1 & 3 \end{pmatrix} \xrightarrow{c_1 \to c_1 + c_2} \begin{pmatrix} 1 & -1 & -1 & 0 \\ 3 & 4 & -1 & -2 \\ -2 & -1 & 3 & -1 \\ -2 & -2 & -1 & 3 \end{pmatrix}$$

$$\xrightarrow[c_3 \to c_3 + c_1]{c_2 \to c_2 + c_1} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 3 & 7 & 2 & -2 \\ -2 & -3 & 1 & -1 \\ -2 & -4 & -3 & 3 \end{pmatrix}$$

$$\xrightarrow[r_3 \to r_3 + 2r_1, r_4 \to r_4 + 2r_1]{r_2 \to r_2 - 3r_1} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 7 & 2 & -2 \\ 0 & -3 & 1 & -1 \\ 0 & -4 & -3 & 3 \end{pmatrix}$$

$$\xrightarrow{c_2 \to c_2 - 3c_3} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & -2 \\ 0 & -6 & 1 & -1 \\ 0 & 5 & -3 & 3 \end{pmatrix}$$

$$\xrightarrow[c_4 \to c_4 + 2c_2]{c_3 \to c_3 - 2c_2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -6 & 13 & -13 \\ 0 & 5 & -13 & 13 \end{pmatrix}$$

$$\xrightarrow[r_4\to r_4-5r_2]{r_3\to r_3+6r_2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 13 & -13 \\ 0 & 0 & -13 & 13 \end{pmatrix}$$

$$\xrightarrow{c_4\to c_4+c_3} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 13 & 0 \\ 0 & 0 & -13 & 0 \end{pmatrix}$$

$$\xrightarrow{r_4\to r_4+r_3} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 13 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Apply the row operations above to I_4 to get P and apply the column operations to I_4 to get Q:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow[r_3\to r_3+2r_1, r_4\to r_4+2r_1]{r_2\to r_2-3r_1} \begin{pmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow[r_4\to r_4-5r_2]{r_3\to r_3+6r_2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ -16 & 6 & 1 & 0 \\ 17 & -5 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow{r_4\to r_4+r_3} \begin{pmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ -16 & 6 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} = P.$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{c_1\to c_1+c_2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow[c_3\to c_3+c_1]{c_2\to c_2+c_1} \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow{c_2 \to c_2 - 3c_3} \begin{pmatrix} 1 & -2 & 1 & 0 \\ 1 & -1 & 1 & 0 \\ 0 & -3 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow[c_4 \to c_4 + 2c_2]{c_3 \to c_3 - 2c_2} \begin{pmatrix} 1 & -2 & 5 & -4 \\ 1 & -1 & 3 & -2 \\ 0 & -3 & 7 & -6 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow{c_4 \to c_4 + c_3} \begin{pmatrix} 1 & -2 & 5 & 1 \\ 1 & -1 & 3 & 1 \\ 0 & -3 & 7 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} = Q.$$

Therefore,

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ -16 & 6 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \quad \text{and} \quad Q = \begin{pmatrix} 1 & -2 & 5 & 1 \\ 1 & -1 & 3 & 1 \\ 0 & -3 & 7 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We then have

$$PMQ = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ -16 & 6 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 4 & -1 & -2 \\ -1 & -1 & 3 & -1 \\ 0 & -2 & -1 & 3 \end{pmatrix} \begin{pmatrix} 1 & -2 & 5 & 1 \\ 1 & -1 & 3 & 1 \\ 0 & -3 & 7 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 13 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} =: D.$$

Therefore,

$$\operatorname{cok}(M) \simeq \operatorname{cok}(D) \simeq \mathbb{Z}/1\mathbb{Z} \oplus \mathbb{Z}/1\mathbb{Z} \oplus \mathbb{Z}/13\mathbb{Z} \oplus \mathbb{Z} \simeq \mathbb{Z} \oplus \mathbb{Z}/13\mathbb{Z}.$$

The explicit isomorphism $\operatorname{cok}(M) \to \operatorname{cok}(D)$ given by the matrix P

$$\operatorname{cok}(M) = \mathbb{Z}^4/\operatorname{im}(M) \to \mathbb{Z}/1\mathbb{Z} \oplus \mathbb{Z}/1\mathbb{Z} \oplus \mathbb{Z}/13\mathbb{Z} \oplus \mathbb{Z} \qquad \to \qquad \mathbb{Z}/13\mathbb{Z} \oplus \mathbb{Z}$$

$$(a,b,c,d) \mapsto P \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} a \\ -3a+b \\ -16a+6b+c \\ a+b+c+d \end{pmatrix} \mapsto (-16a+6b+c, a+b+c+d).$$

Check that each column of M is sent to $(0,0)$ under this mapping, and thus the mapping is well-defined.

Smith normal form.

Definition 4. An $m \times n$ integer matrix M is in *Smith normal form* if

$$M = \mathrm{diag}(s_1, \ldots, s_k, 0, \ldots, 0),$$

a diagonal matrix, where $s_1, \ldots, s_k$ are positive integers such that $s_i | s_{i+1}$ for all i. The s_i are called the *invariant factors* of M.

Example 5. The matrix

$$M := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

is in Smith normal form with invariant factors $s_1 = 1$, $s_2 = 2$, and $s_3 = 12$.

We have

$$\mathrm{cok}(M) := \mathbb{Z}^5 / \mathrm{im}(M) \simeq \mathbb{Z}/1\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z} \times \mathbb{Z}^2 \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z} \times \mathbb{Z}^2.$$

So $\mathrm{cok}(M)$ has rank $r = 2$ and its invariant factors are 2 and 12.

Note that 1 is an invariant factor of M but not of $\mathrm{cok}(M)$. By definition, the invariant factors of a finitely generated abelian group are greater than 1; the invariant factors of M equal to 1 do not affect the isomorphism class of $\mathrm{cok}(M)$ since $\mathbb{Z}_1$ is the trivial group.

An algorithm for computing the Smith normal form is presented below. For our purposes, however, the diagonalization procedure given above is sufficient.

Computing the Smith normal form. The Smith normal form of an integer matrix exists and it unique. (Uniqueness can be shown by relating the invariant factors to the greatest common divisors of the $i \times i$ minors of the matrix for each i.) We show existence here in the form of an algorithm. Let M be an $m \times n$ integer matrix.

Step 1. By permuting rows and columns we may assume that m_{11} is the smallest nonzero entry in absolute value. By adding integer multiples of the first row to other rows or the first column to other columns, attempt to make all entries in the first row and first column except the $(1,1)$-entry equal to 0. If during the process any nonzero entry in the matrix appears with absolute value less than m_{11}, permute rows and columns to bring that entry into the $(1,1)$-position. In this way, m_{11} remains the smallest nonzero entry. Since the succession of values for m_{11} are nonzero and decreasing in magnitude, the process eventually terminates with a matrix of the form

$$\left(\begin{array}{c|ccc} m_{11} & 0 & \cdots & 0 \\ \hline 0 & & & \\ 0 & & & \\ \vdots & & M' & \\ 0 & & & \end{array}\right)$$

where M' is an $(m-1) \times (n-1)$ integer matrix. Negating the first row, if necessary, we take $m_{11} > 0$.

Step 2. If there is an entry of M' that is not divisible by m_{11}, say m_{ij}, then add column j to column 1 and go back to Step 1. Again, since the $(1,1)$-entry is nonzero and decreases in magnitude, this new process terminates. Therefore, we may assume that m_{11} divides every entry of M'.

Step 3. Apply Steps 1 and 2 to M', and thus, by recursion, we produce an equivalent matrix in Smith normal form.