21 Smith Normal Form

21.1 Review

Last time, we looked at presentation matrices for a module. We saw that if we perform elementary row and column operations on the presentation matrix, then this leads to an isomorphic module. We then stated the theorem that we can use such operations to reduce any matrix to Smith normal form. We will prove this today; but first, let's look at a few examples.

21.2 Some Examples in $\mathbb{Z}$

We'll consider the case of 2×2 matrices over $\mathbb{Z}$ — consider the presentation matrix

$$B = \begin{bmatrix} a & c \\ b & d \end{bmatrix},$$

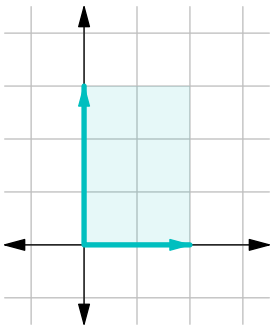
whose corresponding module is

$$M_B = \mathbb{Z}^2 / \operatorname{Span}((a,b)^t, (c,d)^t).$$

The simplest case is when B is diagonal:

Example 21.1
Consider the presentation matrix

$$B = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}.$$



In this case, we have

$$M_B \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z},$$

since given any vector (m, n), its coset mod $\operatorname{Span}((2,0)^t, (0,3)^t)$ is just given by taking the first component mod 2 and the second mod 3. (To be pedantic, the isomorphism is given by $(m, n) \mapsto (m \mod 2, n \mod 3)$ — given any vector, we can subtract multiples of our two vectors to bring it into the rectangle.)

Example 21.2
Consider the presentation matrix

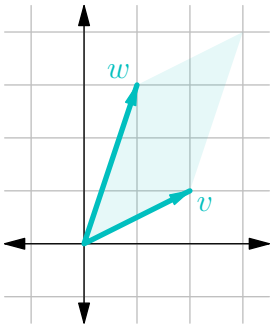
$$B = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}.$$

Now the corresponding module is $M_B = \mathbb{Z}/5\mathbb{Z}$ — the second coordinate is "useless" since we're allowed to subtract multiples of $(0,1)^t$, so we can always eliminate it. So to keep track of the coset of a vector, we only need to keep track of the first component mod 5.

Now we'll look at a more complicated example, where the original matrix is *not* diagonal.

Example 21.3
Consider the presentation matrix

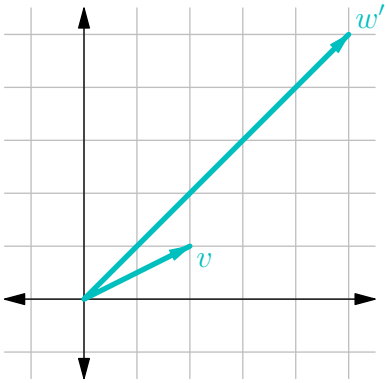
$$B = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}.$$



We can see $\det(B) = 5$, so we can use elementary column operations to make one column a multiple of 5 — if we add twice the first column to the second, then we get

$$B' = B\begin{bmatrix}1 & 2\\ 0 & 1\end{bmatrix} = \begin{bmatrix}2 & 5\\ 1 & 5\end{bmatrix}.$$

So we've replaced w with $w' = 5(1,1)^t$, without changing the lattice spanned by our vectors:



But $(2,1)^t$ and $(1,1)^t$ form a basis for $\mathbb{Z}^2$! So this means to get our lattice, we started with a basis for $\mathbb{Z}^2$, then fixed one of the basis vectors and scaled the other by 5. So this is isomorphic to the previous example — by changing the basis we use for $\mathbb{Z}^2$, we can rewrite v and w' as $(1,0)^t$ and $(0,5)^t$. So we have $M_B \cong \mathbb{Z}/5\mathbb{Z}$.

21.3 Smith Normal Form

Now we'll return to the general case, and prove the theorem.

Theorem 21.4
For a Euclidean domain R, any $n \times m$ matrix B can be reduced using elementary row and column operations to a matrix D, where $d_{ij} = 0$ for all $i \neq j$, and $d_{11} \mid d_{22} \mid \cdots$.

One example of a matrix written in this form (which is called Smith normal form) is

$$D = \begin{bmatrix}2 & 0 & 0 & 0\\ 0 & 6 & 0 & 0\\ 0 & 0 & 12 & 0\end{bmatrix}.$$

Note that if D can be obtained from B by elementary row and column operations, then we can write $D = ABC$, where A is an invertible $n \times n$ matrix and C is an invertible $m \times m$ matrix. For any such D, we then have $M_D \cong M_B$.

Note 21.5
In a PID, it's still possible to obtain D in Smith normal form with $D = ABC$ (for A and C invertible). However, it may not be possible to obtain D by using the elementary operations.

To motivate the proof, notice that the greatest common divisor of all the matrix entries does *not* change under elementary operations — it's clear that scaling by a unit doesn't change the gcd; meanwhile if we perform the operation $a_{ij} \longmapsto a_{ij} + ca_{kj}$, we have

$$\gcd(a_{ij}, a_{kj}) = \gcd(a_{ij} + ca_{kj}, a_{kj}).$$

(This is the same idea as in the Euclidean algorithm.) So if we are able to obtain a matrix D in Smith normal form, then we *must* have

$$d_{11} = \gcd(b_{ij})$$

(since d_{11} divides all other entries of D). So this suggests the main idea of the proof — we want to run some sort of Euclidean algorithm in order to isolate the gcd of all matrix entries in the top-left corner.

Proof of Theorem 21.4. Recall that in a Euclidean domain, we have a size function $\sigma : R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ such that we can divide with remainder — given any a and b, with $b \neq 0$, we can write $a = bq + r$ where $r = 0$ or $\sigma(r) < \sigma(b)$. For convenience, we write $|a|$ instead of $\sigma(a)$.

If $B = 0$, we are done, so assume $B \neq 0$. Then we'll use induction on the size of the matrix; the key step is the following.

Lemma 21.6
By row and column operations, we can arrive at a matrix B' such that $b'_{11} = \gcd(b_{ij}) = \gcd(b'_{ij})$.

Proof. By permuting the rows and columns, we can ensure that $b_{11} \neq 0$, and b_{11} is the nonzero element of minimal size. Now if $b_{11} \mid b_{ij}$ for all i and j, then we're done, so assume not.

Now the main idea is to modify the matrix to make a *smaller* element appear (which we can again move to the top-left corner by rearranging rows and columns). First, if there is an entry with $b_{11} \nmid b_{ij}$ in the first row or column, then we can perform a row or column operation to reduce it — if $b_{ij} = qb_{11} + r$, then we can subtract q times the first row or column from the row or column of b_{ij}, which replaces b_{ij} with r.

If not, then we can use b_{11} to eliminate all other entries in the first row and column (by subtracting multiples of the first column and row), to get a matrix of the form

$$\begin{bmatrix} b_{11} & 0 & \cdots & 0 \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \cdots & * \end{bmatrix}.$$

Now there must be some b_{ij} with $b_{11} \nmid b_{ij}$. We can add its row to the first row; this adds 0 to b_{11}, so we now have a matrix

$$\begin{bmatrix} b_{11} & * & b_{ij} & \cdots & * \\ 0 & * & * & \cdots & * \\ 0 & * & * & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & * & * & \cdots & * \end{bmatrix}.$$

Then we can again add a multiple of the first column to the jth column in order to produce an entry with smaller size.

So either way, we've now produced a matrix with an element smaller in size than b_{11}. Now permute the rows and columns to move this element to the position of b_{11}. Then we repeat the process. At every step, we replace b_{11} with a nonzero entry of smaller size. Since the size is always a nonnegative integer, at some point this process must terminate; this means that b_{11} now divides all entries. □

To complete the proof of Theorem 21.4, we induct on the size of B. By Lemma 21.6, we can replace B with a matrix B' where b'_{11} divides b'_{ij} for all i and j.

Now using row and column operations, we can eliminate the first row and column (meaning that we make b'_{i1} and b'_{1j} all zero). So we get

$$\begin{bmatrix} b'_{11} & * & \cdots & * \\ * & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{bmatrix} \rightsquigarrow \begin{bmatrix} b'_{11} & 0 & \cdots & 0 \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \cdots & * \end{bmatrix}.$$

Now we ignore the first row and column, and use elementary operations on rows and columns $2, \ldots, n$. (These do not affect the first row or column.) By the induction assumption, we can reduce the submatrix of $*$s to Smith normal form (since b'_{11} already divides all other entries, this will remain true when we perform the operations). □

The theorem has more theoretical value than computational value, but we will compute an example nonetheless.

Example 21.7
We have

$$\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 3 \\ 0 & -5 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 \\ 0 & -5 \end{bmatrix}.$$

Example 21.8
We have

$$\begin{bmatrix} 4 & 2 & 6 \\ 1 & 2 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 3 \\ 4 & 2 & 6 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 \\ 4 & -6 & -6 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & -6 & -6 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & -6 & 0 \end{bmatrix}.$$

21.4 Applications

As a corollary, by taking R to be $\mathbb{Z}$, we can classify all finitely presented abelian groups.

Corollary 21.9
Every finitely presented abelian group is isomorphic to

$$\mathbb{Z}/d_1\mathbb{Z} \times \mathbb{Z}/d_2\mathbb{Z} \times \cdots \times \mathbb{Z}/d_n\mathbb{Z} \times \mathbb{Z}^a,$$

for some positive integers d_i with $d_1 \mid d_2 \mid \cdots \mid d_n$.

Sometimes, it's more useful to write this classification in a different form. Recall that the Chinese Remainder Theorem states that if $n = ab$ with $\gcd(a, b) = 1$, then

$$\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/a\mathbb{Z} \times \mathbb{Z}/b\mathbb{Z}.$$

So then if we factor $d_i = p_1^{a_{i1}} \cdots p_n^{a_{in}}$, we can decompose $\mathbb{Z}/d_i\mathbb{Z}$ as a product of groups of the form $\mathbb{Z}/p^m\mathbb{Z}$ (cyclic groups of prime power order).

Another application of Theorem 21.4 is in the case $R = F[x]$, where F is a field. A finitely generated module over R must then be of the form

$$R^a \oplus R/(P_1) \oplus \cdots \oplus R/(P_n),$$

where $P_1 \mid \cdots \mid P_n$. Alternatively, again using the Chinese Remainder Theorem, we can instead assume that each P_i is a power of an irreducible polynomial.

In particular, consider $F = \mathbb{C}$; then the only irreducible polynomials are linear, so we must have $P_i = (x - \lambda_i)^{n_i}$. If we only consider finite-dimensional modules, then as we said earlier, a module over $F[x]$ is equivalent to a (finite-dimensional) vector space along with one linear operator (corresponding to the action of x). So it turns out that this classification of modules actually implies the Jordan decomposition theorem — we will discuss this in more detail next lecture.

Resource: Algebra II Student Notes

Spring 2022

Instructor: Roman Bezrukavnikov
Notes taken by Sanjana Das and Jakin Ng