The Smith normal form

and its use in computing simplicial homology

Sai Sivakumar

May 27, 2022

outline

1. the Smith normal form
   - proof of existence for Euclidean domains
   - proof of existence for PIDs
2. some computational remarks
3. application to computing simplicial homology

the Smith normal form

what the Smith normal form is

Theorem (Smith normal form)

Let R be a principal ideal domain and let A be an $m \times n$ matrix with entries from R.

There exist invertible $m \times m$ and $n \times n$ matrices U and V respectively such that $UAV = S$, where

$$S = \left(\begin{array}{ccc|c} a_1 & & & \\ & \ddots & & 0 \\ & & a_k & \\ \hline & 0 & & 0 \end{array}\right)$$

and $a_1 \mid a_2 \mid \cdots \mid a_k$. The a_i are unique up to associates.

proof of existence for Euclidean domains

Let R be a Euclidean domain and let A be an $m \times n$ matrix with entries from R.

The following elementary row/column operations on A may be achieved by multiplication on the left/right by invertible ("elementary") matrices:

1. Exchanging rows j and k
2. Multiplying row i by any unit of R.
3. Add $q \cdot$ row j to row k for $q \in R$ and $j \neq k$

Same operations for columns in place of rows.

Let A be a nonzero matrix (every zero matrix is in Smith normal form) with

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

We want to transform this matrix by way of elementary row/column operations into a matrix A' with $A'_{11} \mid A'_{ij}$ for $1 \leq i \leq m$ and $1 \leq j \leq n$.

That is, the first entry of A' divides every other entry of A'.

Let $\alpha(A) := \min\{N(a_{ij}) \colon a_{ij} \neq 0_R\}$, and call any entry a_{ij} a *minimal entry* if $N(a_{ij}) = \alpha(A)$. We show that $\alpha(A)$ may be decreased by elementary row/column operations if and only if a minimal entry does not divide every entry in A:

If a minimal entry a_{ij} does not divide every entry in A, we first handle the case where a_{ij} does not divide an entry a_{ik} for $j \neq k$ (another entry in its row).

By the division algorithm on R we have $a_{ik} = qa_{ij} + r$ for $q, r \in R$ with $0 < N(r) < N(a_{ij})$.

Then we can add $-q \cdot$ column j to column k to form a new matrix A' with $\alpha(A')$ being at most $N(r)$ since one of its entries is $r = a_{ik} - qa_{ij}$. But $\alpha(A') = N(r) < N(a_{ij})$, which means we have decreased $\alpha(A)$.

A similar procedure can be done if a_{ij} does not divide another entry in its column.

If a_{ij} divides entries in its row and column but does not divide an entry found outside of its row and column, say a_{sk} for $i \neq s, j \neq k$, we can reduce the situation to one from before.

Use elementary row/column operations to achieve the following picture:

$$\begin{matrix} a_{ij} & \cdots & a_{ik} \\ \vdots & & \vdots \\ a_{sj} & \cdots & a_{sk} \end{matrix} \to \begin{matrix} a_{ij} & \cdots & a_{ik} \\ \vdots & & \vdots \\ 0 & \cdots & a_{sk} - qa_{ik} \end{matrix} \to \begin{matrix} a_{ij} & \cdots & a_{sk} + (1-q)a_{ik} \\ \vdots & & \vdots \\ 0 & \cdots & a_{sk} - qa_{ik} \end{matrix}$$

Now a_{ij} does not divide an element in its row, which we already handled.

We show the converse by the contrapositive.

If a minimal entry a_{ij} does divide every entry in A, then a_{ij} divides all entries in any matrix A' obtained by applying elementary row/column operations to A. As a result, there is no way to reduce $N(a_{ij}) = \alpha(A)$.

Hence we can take a matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

and by elementary row/column operations form a matrix

$$A' = \begin{pmatrix} a'_{11} & \cdots & a'_{1n} \\ \vdots & \ddots & \vdots \\ a'_{m1} & \cdots & a'_{mn} \end{pmatrix}$$

with a'_{11} dividing all entries of A'.

By using more elementary row/column operations form another matrix

$$B = \left(\begin{array}{c|ccc} a'_{11} & 0 & \cdots & 0 \\ \hline 0 & a'_{22} & \cdots & a'_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a'_{m2} & \cdots & a'_{mn} \end{array}\right) = \left(\begin{array}{c|c} a'_{11} & 0 \\ \hline 0 & B' \end{array}\right).$$

By induction the smaller matrix B' can be made into Smith normal form.

Hence there exist invertible matrices U, V such that

$$UAV = S = \left(\begin{array}{ccc|c} a_1 & & & \\ & \ddots & & 0 \\ & & a_k & \\ \hline & 0 & & 0 \end{array}\right).$$

The divisibility relations $a_1 \mid a_2 \mid \cdots \mid a_k$ come from the fact that in B, a'_{11} divided every entry of B'. As a result, a'_{11} will divide every entry of a matrix obtained by applying row/column operations to B'.

Uniqueness of the Smith normal form up to associates comes from the fact that divisibility holds up to associates.

proof of existence for PIDs

We need a few lemmas. Let R be a principal ideal domain, and let x, y be nonzero elements of R.

1. There exists an invertible matrix W such that
$$W \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \gcd(x, y) \\ 0 \end{pmatrix}.$$
2. There exist invertible matrices U, V such that
$$U \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix} V = \begin{pmatrix} \gcd(x, y) & 0 \\ 0 & \operatorname{lcm}(x, y) \end{pmatrix},$$
with $\operatorname{lcm}(x, y) := xy / \gcd(x, y)$.

1. Since R is a PID, $(x, y) = (d)$ with d associate to $\gcd(x, y)$. Without loss of generality, take $d = \gcd(x, y)$. Thus there exist $\alpha, \beta \in R$ with $\alpha x + \beta y = d$. Since $d \mid x$ and $d \mid y$ there exist $p, q \in R$ with $x = dp$ and $y = dq$. It also follows that $\alpha p + \beta q = 1_R$.

   Take
$$W = \begin{pmatrix} \alpha & \beta \\ -q & p \end{pmatrix}.$$

   It follows that
$$W \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \gcd(x, y) \\ 0 \end{pmatrix}$$

   with $\det(W) = 1_R$.

2. Let d, α, β, p, q be given as before. Then with

$$U = \begin{pmatrix} 1 & 1 \\ -bq & 1 - bq \end{pmatrix}, \quad V = \begin{pmatrix} a & -q \\ b & p \end{pmatrix},$$

it follows that

$$U \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix} V = \begin{pmatrix} \gcd(x, y) & 0 \\ 0 & \operatorname{lcm}(x, y) \end{pmatrix}$$

with $\det(U) = \det(V) = 1_R$.

Another lemma: Let R be a (commutative) Noetherian ring and let D be a nonempty subset of R. Show that there exists an element $d \in D$ which is "minimal with respect to division"; that is, if there exists $d' \in D$ with $d' \mid d$ then $d \mid d'$ also.

To prove this lemma we must recall what a Noetherian ring is.

A Noetherian ring is one which satisfies the ascending chain condition: For any increasing sequence of ideals $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ of R there exists an n such that $I_n = I_{n+1} = \cdots$; that is, the chain stabilizes.

This is equivalent to the "maximal principle": Every nonempty subset of ideals of R has a maximal element.

Observe that the maximal principle implies the ascending chain condition pretty quickly, but the converse requires Zorn's lemma.

By taking the set of ideals $\{(a) \colon a \in D\}$ and invoking the fact that R is Noetherian we obtain a maximal ideal (d) for some $d \in D$ which satisfies the property we want:

If there is some $d' \in D$ with $d' \mid d$, we have that $(d) \subseteq (d')$. But by maximality of (d) we must have $(d) = (d')$ which yields that $d \mid d'$ as desired.

the proof

With the lemmas proven we may continue with the proof.

Let R be a PID (so R is Noetherian), and let A be a matrix with entries from R.

Let D be given by the set of all entries appearing in *any* matrix B *similar* to A, and let d be an element of D which is minimal with respect to division in the sense of the previous lemma.

Let A' be a matrix similar to A such that d appears as the first entry; here $A' = U'AV'$ for invertible U', V'.

We show that d divides every entry in the matrix A'.

Suppose some entry a_{sk} does not divide d. We can modify one of the first two lemmas to obtain matrices S, T such that $SA'T$ contains $\gcd(d, a_{sk})$ as an entry. But $\gcd(d, a_{sk}) \mid d$, and since d was chosen minimally with respect to division, it follows that $d \mid \gcd(d, a_{sk})$, from which it follows that d divides a_{sk}. Contradiction.

Since d divides every entry in A', we may apply elementary row/column operations to reduce A' into a matrix

$$B = \left(\begin{array}{c|c} d & 0 \\ \hline 0 & B' \end{array}\right)$$

where d divides every entry in B'. By induction reduce B' to Smith normal form to obtain the result.

The divisibility relations hold since d divided every entry of B' and will divide every entry of a matrix obtained by applying elementary row/column operations to B'.

The Smith normal form is unique up to associates since divisibility holds up to associates.

some computational remarks

the row reduction algorithm

For integer matrices the basic algorithm is just Gaussian elimination (or just the elementary row/column operations we saw earlier).

Sample pseudocode for reducing $\mathbb{Z}/2\mathbb{Z}$-valued $n_{p-1} \times n_p$ matrices:

```
void REDUCE(x)
  if there exist k ≥ x, l ≥ x with N_p[k, l] = 1 then
    exchange rows x and k;  exchange columns x and l;
    for i = x + 1 to n_{p−1} do
      if N_p[i, x] = 1 then add row x to row i endif
    endfor;
    for j = x + 1 to n_p do
      if N_p[x, j] = 1 then add column x to column j endif
    endfor;
    REDUCE(x + 1)
  endif.
```

In each recursive call there are at most $(n_{p-1} + n_p)$ elementary row/column operations so the number of elementary operations is at most $(n_{p-1} + n_p) \min \{n_{p-1}, n_p\}$.

By accounting for the length of the columns/rows in the matrix the computational complexity is bounded above by a cubic polynomial in n_{p-1}, n_p.

recent-ish results

Dumas, Heckenbach, Saunders, and Welker various efficient Smith normal form algorithms `link`

Storjohann and Labahn fast Las Vegas algorithm for polynomial matrices `link`

application to computing simplicial homology

chain complexes

Recall that a *chain complex* $\mathcal{C}$ is a sequence

$$\cdots \xrightarrow{\partial_{n+2}} C_{n+1} \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} \cdots$$

of abelian groups C_i with homomorphisms ∂_i such that $\partial_i \circ \partial_{i+1} = 0$.

We define the *homology groups* by $H_i(\mathcal{C}) = \ker \partial_i / \operatorname{im} \partial_{i+1}$.

standard bases for free chain complexes

When the groups C_i of a chain complex are free of finite rank, there exist subgroups U_i, V_i, W_i of C_i such that

$$C_i = U_i \oplus V_i \oplus W_i$$

where $\partial_i(U_i) \subseteq W_{i-1}$ and $\partial_i(V_i) = \partial_i(W_i) = 0$

Step 1: let W_i be the set of all elements c of C_i such that some nonzero multiple of c is found in $\operatorname{im}\partial_{i+1}$. Check that W_i is a subgroup of C_i.

The following containments hold:

$$\operatorname{im}\partial_{i+1} \subseteq W_i \subseteq \ker\partial_i \subseteq C_i$$

The second containment holds since C_i is torsion free, from which $mc_i = \partial_{i+1}d_{i+1}$ yields $\partial_i c_i = 0$.

To show that W_i is a direct summand of $\ker \partial_i$, consider the natural projection

$$\ker \partial_i \to H_i(\mathcal{C}) \to H_i(\mathcal{C})/T_i(\mathcal{C})$$

where $T_i(\mathcal{C})$ is the torsion subgroup of $H_i(\mathcal{C})$.

An element $c \in \ker \partial_i$ is in the kernel of this projection if and only if some multiple of c is in $\operatorname{im} \partial_{i+1}$; i.e., if $c \in W_i$. Hence

$$\ker \partial_i / W_i \cong H_i(\mathcal{C})/T_i(\mathcal{C}),$$

and since $H_i(\mathcal{C})/T_i(\mathcal{C})$ is finitely generated and torsion-free, it is free.

It follows that $\ker \partial_i / W_i$ is free.

If $\{c_1 + W_i, \dots, c_k + W_i\}$ is a basis for $\ker \partial_i / W_i$ and $\{d_1, \dots, d_l\}$ is a basis for W_i, then $\{c_1, \dots, c_k, d_1, \dots, d_l\}$ is a basis for $\ker \partial_i$.

It follows that W_i is a direct summand of $\ker \partial_i$, so $\ker \partial_i = V_i \oplus W_i$, where V_i has basis $\{c_1, \dots, c_k\}$.

Step 2: Choose ordered bases $\{e_1, \ldots, e_n\}$ for C_i and $\{d_1, \ldots, d_m\}$ for C_{i-1} for which the matrix of $\partial_i \colon C_i \to C_{i-1}$ takes on the Smith normal form

$$\left(\begin{array}{ccc|c} b_1 & & & \\ & \ddots & & 0 \\ & & b_l & \\ \hline & 0 & & 0 \end{array}\right)$$

with $b_i \geq 1$ and $b_1 \mid b_2 \mid \cdots \mid b_l$.

The following hold:

1. $\{e_{l+1}, \ldots, e_n\}$ is a basis for $\ker \partial_i$.

Observe $< e_{l+1}, \ldots, e_n > \subseteq \ker \partial_i$ and if we take any element $c = \sum_{k=1}^{n} a_k e_k \in C_i$ then $\partial_i(c) = \sum_{k=1}^{l} b_k a_k d_k$.

For c to be in $\ker \partial_i$ we must have $\partial_i(c) = 0$; equivalently, $a_1, \ldots, a_l$ must be zero, and in this case $c = \sum_{k=l+1}^{n} a_k e_k$.

2. $\{d_1, \ldots, d_l\}$ is a basis for W_{i-1}.

To show the second point, observe that $d_1, \ldots, d_l \in W_{i-1}$ since $b_k d_k = \partial_i(e_k)$ for $1 \leq k \leq l$.

Conversely, let $f = \sum_{k=1}^{m} f_k d_k \in C_{i-1}$, and if $f \in W_{i-1}$, then for some $\lambda \neq 0$ and $c = \sum_{k=1}^{n} a_k e_k \in C_i$ we have

$$\lambda f = \sum_{k=1}^{m} \lambda f_k d_k = \partial_i(c) = \sum_{k=1}^{l} b_k a_k d_k.$$

It follows that $f_{l+1}, \ldots, f_m = 0$ so that $f \in < d_1, \ldots, d_l >$ as well.

3. $\{b_1 d_1, \ldots, b_l d_l\}$ is a basis for $\operatorname{im} \partial_i$.

The third point holds since any element of $\operatorname{im} \partial_i$ is in the form $\sum_{k=1}^{l} b_k a_k d_k$ which is in $< b_1 d_1, \ldots, b_l d_l >$. The reverse containment also holds, and $\{b_1 d_1, \ldots, b_l d_l\}$ is linearly independent since $b_k \geq 1$.

Step 3: To prove the theorem, choose ordered bases for C_i and C_{i-1} as in Step 2 and choose U_i to be the group generated by $\{e_1, \ldots, e_l\}$ so that

$$C_i = U_i \oplus \ker \partial_i.$$

By Step 1, decompose $\ker \partial_i$ into $V_i \oplus W_i$ so that

$$C_i = U_i \oplus V_i \oplus W_i.$$

We obtain also that $\partial_i(U_i) \subseteq W_{i-1}$ and $\partial_i(V_i) = \partial_i(W_i) = 0$. Carrying out Step 2 in full gives us the required bases for U_i and W_{i-1}.

computing homology groups

The homology groups of a finite simplicial complex K can be computed explicitly.

Orient the simplices of K and obtain the groups C_i and maps ∂_i forming a chain complex $\mathcal{C}$. Use the previous result to decompose C_p as $U_p \oplus V_p \oplus W_p$.

Then

$$\begin{aligned} H_p(\mathcal{C}) = \ker \partial_p / \operatorname{im} \partial_{p+1} &\cong (V_p \oplus W_p)/\operatorname{im} \partial_{p+1} \\ &= V_p \oplus (W_p/\operatorname{im} \partial_{p+1}) \cong (\ker \partial_p / W_p) \oplus (W_p / \operatorname{im} \partial_{p+1}). \end{aligned}$$

The first group in the direct sum is free and the second group is a torsion group.

We have thus reduced computing homology to computing these two groups.

Take the matrices of $\partial_p \colon C_p \to C_{p-1}$ and $\partial_{p+1} \colon C_{p+1} \to C_p$ (which will have entries from $\{0, 1, -1\}$) and reduce them to Smith normal forms S_p, S_{p+1}, respectively.

Let $b_1, \ldots, b_l$ be the nonzero entries appearing in the diagonal of S_{p+1}.

Then

1. The rank of $\ker \partial_p$ is equal to the number of zero *columns* of S_p.
2. The rank of W_{p-1} is equal to the number of nonzero *rows* of S_p.
3. There is an isomorphism

$$W_p / \operatorname{im} \partial_{p+1} \cong \mathbb{Z}/b_1\mathbb{Z} \oplus \mathbb{Z}/b_2\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/b_l\mathbb{Z}$$

sources

Munkres algebraic topology

Dummit and Foote algebra

Professor Speyer's (UMich LSA) worksheets from most recent Algebra class link

Edelbrunner and Harer computational topology