SMITH NORMAL FORM OVER THE INTEGERS

Contents

1. Introduction

In these notes, we explain how to put any $k \times n$ matrix with integer entries into Smith normal form. The same result is true over an arbitrary principal ideal domain. The same proof works over any Euclidean domain.

2. Smith normal form

Let $A = (a_{ij})$ be a $k \times n$ matrix with entries in the ring $R = \mathbf{Z}$ of integers.

We say that the matrix A is in *Smith normal form* if

(1) $a_{ij} = 0$ for $i \neq j$,

(2) For some m with $0 \leq m \leq k$, $a_{ii} \neq 0$ for $i \leq m$, and $a_{ii} = 0$ for $i > m$.

(3) Let $a_i = a_{ii}$ for $1 \leq i \leq m$. Then $a_1 / a_2 / \ldots / a_m$.

If $m = 0$ in the above definition, then the matrix A in Smith normal form is the zero matrix.

Pictorially, a matrix in Smith normal form looks like

$$A = \begin{bmatrix} a_1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & a_2 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & a_3 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & a_m & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \end{bmatrix}$$

For a nonzero matrix $A = (a_{ij})$ with coefficients in R, we let $d(A)$ be the greatest positive number dividing all coefficients, so $d(A) = gcd\{a_{ij} : 1 \leq i \leq k, 1 \leq j \leq n\}$.

Let r_i be the ith row of A and let c_j be the jth column of A. Recall the row and column operations:

R_1: switches the ith and jth row of A.

R_2: multiplies the ith row of A by -1.

R_3: replaces the ith row of A by $r_i + ar_j$ for some $a \in R$.

C_1: switches the ith and jth columns of A.

C_2: multiplies the ith column of A by -1.

C_3: replaces the ith column of A by $c_i + ac_j$ for some $a \in \mathbf{Z}$.

Lemma 2.1. *Let B be the result of applying a row or column operation to A. Then $d(A) = d(B)$.*

Proof : For operations R_1, R_2, C_1, and C_2 this is trivial. For operations R_3 and C_3, the claim can be proved in the same way as one proves that $gcd(a, b) = gcd(a + sb, b)$ for $a, s, b \in R$..

Q.E.D.

Proposition 2.2. *Let A be a nonzero matrix with integer coefficients and let $d = d(A)$. Then we can use row and column operations to transform A into a matrix B with $t(B) = d$.*

Proof : We use induction on the minimum element $t = t(A)$ of the set $\{|a_{ij}| : a_{ij} \neq 0, 1 \leq i \leq k, 1 \leq j \leq n\}$. Clearly, $t \geq d$. If $t = d$, we're done. If not, we may assume the result is known for matrices C with integer coefficients such that $d(C) = d$ and $t(C) < t$.

Let $|a_{uv}| = t$, and note that by an operation of type R_2 by -1, we can assume that $a_{uv} = t$. Since $t > d$, there exists a coefficient a_{lm} of A such that t does not divide a_{lm} (otherwise, t would divide d). We have to consider two cases:

Case 1: There exists an entry a_{lm} such that t does not divide a_{lm} with a_{lm} in either the same column or row as a_{uv} (i.e., $u = l$ or $v = m$). By the Euclidean algorithm, we can write $a_{lm} = qa_{uv} + r$ for some integer r such that $0 < r < a_{uv}$. If $v = m$, so a_{uv} and $a_{lm} = a_{lv}$ are in the same column, then by replacing the lth row of A by $r_l - qr_u$ (lth row - q times uth row), we get a new matrix B. By Lemma 2.1, $d(A) = d(B)$. By construction the lm entry b_{lm} of B is r. Hence, $t(B) = r < t$, so by the inductive assumption, we can perform row and column operations on B to transform B to a new matrix C with $t(C) = d$.

Case 2: a_{uv} divides every entry a_{uj} and a_{iv} in the same row or column as a_{uv}. Then note that we can do row and column operations to transform A so that the uth row and vth column are 0 outside of the a_{uv}-entry. Indeed, we begin with the entries of the form a_{iv}. Since a_{uv} divides a_{iv} by assumption, there exists $z \in R$ such that $a_{iv} = za_{uv}$. Then a row operation replacing the row r_i by $r_i - zr_u$ has the effect of making the a_{iv} entry equal to 0. None of these operations change the uth row. Now each entry $a_{uj} = za_{uv}$ for some $z \in R$, and a column operation replacing the jth column by $c_j - zc_v$ makes the a_{uj} entry

equal to 0. Call this new matrix B. Note that $d(B) = d$ by Lemma 2.1, and $t(B) \leq t$ since $b_{uv} = a_{uv} = t$. If $t(B) = d$, we're done. If not, $t(B) > d$, and it follows that there is an entry b_{lm} such that b_{uv} does not divide b_{lm}. By construction, $u \neq l$ and $v \neq m$. We replace the uth row of B by $r_l + r_u$ and call this new matrix C. Since $b_{lv} = 0$ by construction, this does not change the entry b_{uv}, so $t(C) \leq t$. However, it makes the entry $c_{um} = b_{um} + b_{lm} = 0 + b_{lm} = b_{lm}$, since $b_{um} = 0$ by construction. Now we reason as in case 1, to show that we can change C by a column operation to get a new matrix D so that $t(D) < t(C) \leq t$, but $d(D) = d$. Hence, by induction, we can transform D by column operations to a new matrix X with $t(X) = d$.

Q.E.D.

Theorem 2.3. *If A is a matrix with integer coefficients, we can transform A into Smith normal form using row and column operations.*

Proof : If $A = 0$, the result is trivial. If $A \neq 0$, then we can use Proposition 2.2 to transform A by row and column operations into a matrix with entry a_{uv} with $a_{uv} = d = d(A)$. By definition, this means that a_{uv} divides all entries of A. By using operations of type R_1 and C_1, transform the matrix A into a new matrix A with $a_{11} = d$. Perform row and column operations of type R_3 and C_3. to zero out the first row and column. Note that since d divides all entries of A, then d divides all entries of this new matrix, still denoted A. The new matrix has the form

$$A = \begin{bmatrix} d & 0 \\ 0 & B \end{bmatrix},$$

where B is a $k-1$ by $n-1$ matrix with integer entries, all divisible by d, and the 2 $0's$ denote zero vectors of the appropriate size.. Let $a_1 = d$, and perform row and column operations to the matrix B (not involving the first row or column) to put B into similar form. The remainder of the theorem follows by an easy induction.

Q.E.D.

The proof gives a procedure for reducing a matrix into Smith normal form. One can compute $t(A)$ and $d(A)$. If $d(A) < t(A)$, then one can use the argument from the proof of the Proposition to transform A into a new matrix B with $t(B) < t(A)$. Keep doing this until we get to a matrix C with $t(C) = d(C) = d(A)$. Then it's easy to zero out all entries in the first row and column aside from the a_{11}.

I suspect there are much more efficient algorithms, but anyway this argument gives a proof of our result on Smith normal form.