NORTH-HOLLAND

The Smith Normal Form*

Morris Newman
Department of Mathematics
University of California
Santa Barbara, California 93106

Submitted by Moshe Goldberg

SOME HISTORICAL REMARKS

Henry John Stephen Smith (1826–1883) was the Savilian Professor of Geometry at Oxford, and was regarded as one of the best number theorists of his time. His specialties were pure number theory, elliptic functions, and certain aspects of geometry. He shared a prize with H. Minkowski for a paper which ultimately led to the celebrated Hasse-Minkowski theorem on representations of integers by quadratic forms, and much of his research was concerned with quadratic forms in general. He also compiled his now famous *Report on the Theory of Numbers*, which predated L. E. Dickson's *History of the Theory of Numbers* by three-quarters of a century, and includes much of his own original work. The only paper on the Smith normal form (also known as the Smith canonical form) that he wrote [On systems of linear indeterminate equations and congruences, *Philos. Trans. Roy. Soc. London* CLI:293–326 (1861)] was prompted by his interest in finding the general solution of diophantine systems of linear equations or congruences. Matrix theory per se had not yet developed to any extent, and the numerous applications of Smith's canonical form to this subject were yet to come.

* A review article solicited by the Editors-in-Chief of *Linear Algebra and Its Applications*.

LINEAR ALGEBRA AND ITS APPLICATIONS 254:367-381 (1997)

655 Avenue of the Americas, New York, NY 10010
0024-3795/97/$17.00
PII S0024-3795(96)00163-2

Typical examples of the types of problems he considered might be to find all integral solutions of the system

$$13x - 5y + 7z = 12,$$

$$67x + 17y - 8z = 2,$$

or to find all solutions of the congruence

$$45x + 99y + 11z \equiv 7 \bmod 101.$$

The answer to the first is that in terms of an arbitrary integral parameter t,

$$x = -34 - 79t,$$

$$y = 248 + 573t,$$

$$z = 242 + 556t,$$

and the answer to the second is that if y and z are taken as arbitrary, then x is given by

$$x \equiv 18y + 2z + 63 \bmod 101.$$

These are both derivable by a systematic use of the normal form, a process which will be explained later.

MATRIX EQUIVALENCE AND INVARIANT FACTORS

The problem underlying the Smith normal form is that of matrix equivalence, which can be treated in rather general terms.

Let R be a commutative ring with an identity 1. An element a of R is a *unit* if an element b of R exists such that $ab = ba = 1$. Now let m, n be positive integers, and let R_n stand for the ring of $n \times n$ matrices over R, and $R_{m,n}$ for the ring of $m \times n$ matrices over R. An element A of R_n is *unimodular* or a *unit matrix* if an element B of R_n exists such that $AB = BA = I_n$, where I_n is the identity matrix of order n. The set of unimodular matrices of R_n will be denoted by GL(n, R), and is a multiplicative group.

Now for the definition of equivalence: two elements A, B of $R_{m,n}$ are said to be *equivalent* (written $A \sim B$) if matrices U, V exist such that U belongs to GL(m, R), V belongs to GL(n, R), and $B = UAV$. This is an equivalence relationship which partitions the set $R_{m,n}$ into disjoint equivalence classes, and the principal problem encountered here is to find a way to determine to which equivalence class an element of $R_{m,n}$ belongs.

In this general setting, the problem is too difficult, and it is necessary to make some additional assumptions about the underlying ring R. The usual one is to assume that R is a principal ideal domain, but this is unnecessarily broad for our purposes, and we shall assume instead that R is one of the three rings described below.

(1) $R = \mathbb{Z}$, the ring of integers. The units here are ± 1, and the $n \times n$ unimodular matrices over $\mathbb{Z}$ are those of determinant ± 1.

(2) $R = \mathbb{F}$, a field. Any nonzero element of $\mathbb{F}$ is a unit, and the $n \times n$ unimodular matrices over $\mathbb{F}$ are the nonsingular matrices.

(3) $R = \mathbb{F}[x]$, the ring of polynomials in x with coefficients from $\mathbb{F}$. The units (as before) are the nonzero elements of $\mathbb{F}$, and the $n \times n$ unimodular matrices over $\mathbb{F}$ are those whose determinant is a nonzero element of $\mathbb{F}$.

In order to simplify our discussion we will assume that the matrices under consideration are all square (an unimportant restriction).

It is straightforward that equivalent matrices must have the same rank, and Smith's theorem states the following [5, p. 26]:

Every matrix A of R_n which is of *rank* r is equivalent to a diagonal matrix D given by

$$D = \operatorname{diag}(s_1, s_2, \ldots, s_r, 0, \ldots, 0),$$

where the entries s_i are different from 0 and form a divisibility sequence; that is,

$$s_1 | s_2 | \cdots | s_r.$$

Furthermore, the s_i are unique, apart from possible unit multipliers belonging to R.

The s_i [also denoted by $s_i(A)$] are known as the *invariant factors* of A, and are basic to the problem of determining when two matrices of R_n are equivalent. The matrix D is then the Smith normal form of A, and is denoted by $S(A)$.

It follows from Smith's theorem that two matrices of R_n are equivalent if and only if they are of the same rank and have the same invariant factors.

It is important to notice that in each of the three cases considered here, the ring R is a euclidean ring, so that constructive algorithms exist for the determination of the Smith normal form.

Along with the invariant factors of A, there are two other sets of invariants, known as the determinantal divisors of A and the elementary divisors of A. These are of equal importance and will be discussed individually.

DETERMINANTAL DIVISORS

Let A be any matrix of R_n. Let k be any integer such that $1 \leqslant k \leqslant n$. Choose (in all possible $\binom{n}{k}^2$ ways) k row subscripts and k column subscripts, and compute all the determinants of the submatrices constructed from these choices. Finally, take the greatest common divisor of all of these determinants. This number will be denoted by $d_k(A)$, and is known as the *kth determinantal divisor* of A. Notice that if A is of rank r, then only the first r such numbers will be different from zero. For completeness, define $d_0(A)$ to be 1. Then the relevant criterion [5, p. 28] is that *two matrices A, B of R_n are equivalent if and only if they have the same determinantal divisors, up to unit multipliers*. Note that this implies that A and B have the same rank.

Another characterization for $d_k(A)$ is that it is the greatest common divisor of the entries in the kth compound of A [3, p. 87].

Historically, this is the way that Smith approached the problem of a canonical form for equivalence.

The relationship between the determinantal divisors and the invariant factors is quite simple:

$$d_k(A) = s_1(A)s_2(A)\cdots s_k(A), \qquad 1 \leqslant k \leqslant n,$$

or written the other way around,

$$s_k(A) = \frac{d_k(A)}{d_{k-1}(A)}, \qquad 1 \leqslant k \leqslant n.$$

ELEMENTARY DIVISORS

Over the euclidean ring R, unique factorization exists, and so we can talk about the primes of R. For example, if $R = \mathbb{F}$, a field, there are no primes (every element is either a unit or 0); if $R = \mathbb{Z}$, the primes have their usual meaning; and if $R = \mathbb{F}[x]$, the primes are the irreducible polynomials of $\mathbb{F}[x]$. Any one of the invariant factors is then uniquely expressible as the product of distinct prime powers. The total set of such prime powers, for all of the r invariant factors, is then another invariant. Any such prime power is called an *elementary divisor*. Here, the relevant result is that *two matrices of R_n are equivalent if and only if they have the same elementary divisors* [5, p. 30].

The simplest (and best-known) situation where the Smith normal form comes into play is when R is a field. Since here all nonzero elements of $\mathbb{F}$ are units, the nonzero invariant factors are all 1. Hence two matrices A, B of R_n are equivalent if and only if they have the same rank r. The Smith normal form then becomes $I_r \dotplus O_{n-r}$, where I_r is the identity matrix of order r, and O_{n-r} is the $(n-r) \times (n-r)$ zero matrix.

For a thorough discussion of this material, see [5, Chapters I and II].

SOME INTERESTING FACTS

Let A, B be nonsingular $n \times n$ matrices over R. Then the following hold:

(1) [5, p. 33] $s_k(AB)$ is divisible by $s_k(A)$ and by $s_k(B)$, $1 \leqslant k \leqslant n$.

(2) [7] $d_k(AB)$ is divisible by $d_k(A)d_k(B)$, $1 \leqslant k \leqslant n$. If $(\det A, \det B) = 1$, then

$$d_k(AB) = d_k(A)d_k(B), \qquad 1 \leqslant k \leqslant n.$$

(3) [5, p. 28] $d_{k-1}(A)d_{k+1}(A)$ is divisible by $d_k(A)^2$, $1 \leqslant k \leqslant n-1$, where $d_0(A) = 1$.

(4) $s_{i+j-1}(AB)$ is divisible by $s_i(A)s_j(B)$, $1 \leqslant i, j \leqslant n$, $i + j \leqslant n + 1$. This is the simplest of many relationships discovered by R. C. Thompson [10].

(5) [5, p. 33] If $(\det A, \det B) = 1$, then $S(AB) = S(A)S(B)$.

(6) [5, p. 30] The prime power divisors of a diagonal matrix are in fact the elementary divisors of the matrix.

(7) [8, 9] If A is an $n \times n$ integral matrix, then $s_n(A)/\lambda$ is an algebraic integer for any nonzero eigenvalue λ of A. In the other direction,

$s_1(A)s_2(A)\cdots s_k(A)$ divides the product of any k eigenvalues of A (repetitions allowed), in the sense that the quotient is an algebraic integer, for $1 \leqslant k \leqslant n$.

(8) [4, Section 3.28] The minimal polynomial of an $n \times n$ matrix A is $s_n(A - xI)$.

AN APPLICATION TO THE SOLUTION OF LINEAR SYSTEMS

The first application we mention is to the original purpose behind the invention of this concept by Smith; namely, to the solution of systems of linear diophantine equations.

Suppose then that we are given an integral $m \times n$ matrix A and an integral $m \times 1$ vector b, and we want to find all integral solutions of the diophantine system $Ax = b$. We do this by finding an integral basis for the null space of A, and a particular integral solution (if there is one) of the system. We first find the Smith normal form $S = UAV$ of A, and replace the system by the equivalent system $Sy = c$, where $x = Vy$, and $c = Ub$.

If the rank of A is r, then

$$S = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix},$$

where D is a nonsingular diagonal $r \times r$ matrix. Put $c = (c', c'')^t$, $y = (y', y'')^t$ (t denoting the transpose), where c' and y' are $r \times 1$, and c'' and y'' are $(m - r) \times 1$. Then $Sy = c$ if and only if $Dy' = c'$, $0 = c''$. Thus the system has integral solutions if and only if $c'' = 0$, and $D^{-1}c'$ is an integral vector. Consequently, a particular solution in this case is given by $x = V(D^{-1}c', 0)^t$. It can be shown that the last $n - r$ columns of V are an integral basis for the null space of A [6].

For example, take $Ax = b$, where

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 2 & 4 & 5 & 6 & 1 & 1 & 1 \\ 1 & 4 & 2 & 5 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 & 2 & 2 & 3 \end{pmatrix}, \qquad b = \begin{pmatrix} 28 \\ 4 \\ 20 \\ 14 \\ 9 \end{pmatrix}.$$

Then the rank of A is $r = 5$, the nullity is $n - r = 2$, and the Smith normal form is

$$S(A) = UAV = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix},$$

where

$$U = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & -1 & 0 & 0 \\ 2 & 0 & -1 & 0 & -1 \\ 5 & 0 & -2 & -1 & -3 \\ -3 & -1 & 2 & 0 & 0 \end{pmatrix},$$

$$V = \begin{pmatrix} 1 & -3 & 2 & -6 & -14 & 29 & 16 \\ 0 & 0 & 0 & 7 & 15 & -34 & -18 \\ 0 & 1 & -2 & 5 & 10 & -23 & -13 \\ 0 & 0 & 1 & -7 & -14 & 33 & 18 \\ 0 & 0 & 0 & 1 & 2 & -6 & -4 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

and

$$D = \operatorname{diag}(1, 1, 1, 1, 2).$$

Hence, a particular solution of the system is given by

$$x = (D^{-1}c', 0)^t = (-44, 53, 37, -50, 11, 0, 0)^t.$$

As mentioned above, the last two columns of V form an integral basis for the null space of A.

With obvious modifications, the same procedure can be used to find all solutions of the congruence $Ax \equiv b \bmod p$, where p is a prime. In fact, the discussion is quite trivial in this case, since the diagonal matrix D that arises is just the r-dimensional identity matrix.

AN APPLICATION TO PERMUTATION EQUIVALENCE

Our next application is to permutation equivalence. Although it is a finite problem to determine whether or not two matrices are permutation equivalent (i.e. whether or not one can be derived from the other by applying

suitable row and column permutations to it), in practice this is not feasible if the matrices are even moderately large. A negative criterion is available from the observation that matrices that are permutation equivalent must have the same Smith normal form. This is quite a useful criterion, and with it we can show, for example, that the following four 16×16 Hadamard matrices H_1, H_2, H_3, H_4, which were produced by Marshall Hall, Jr., and which represent (up to permutation equivalence) any 16×16 Hadamard matrix, are in fact not permutation equivalent, since they are not even equivalent (in the Smith sense):

$$H_1 = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1
\end{pmatrix},$$

$$H_2 = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1
\end{pmatrix},$$

$$H_3 = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \\
1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\
1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1
\end{pmatrix},$$

$$H_4 = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \\
1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\
1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & 1 & -1
\end{pmatrix}.$$

The invariant factors of these matrices are given by

$$H_1\colon\ 1, 2, 2, 2, 2, 4, 4, 4, 4, 4, 4, 8, 8, 8, 8, 16,$$

$$H_2\colon\ 1, 2, 2, 2, 2, 2, 4, 4, 4, 4, 8, 8, 8, 8, 8, 16,$$

$$H_3\colon\ 1, 2, 2, 2, 2, 2, 2, 4, 4, 8, 8, 8, 8, 8, 8, 16,$$

$$H_4\colon\ 1, 2, 2, 2, 2, 2, 2, 2, 8, 8, 8, 8, 8, 8, 8, 16.$$

Looking at these factors, we see, for example, that the number of 2's in each of the four lists is different. Hence our matrices are inequivalent, and therefore also permutation inequivalent.

AN APPLICATION TO ABELIAN GROUP THEORY

The Smith canonical form is a basic tool of abelian group theory. We mention only one possible application. If x is a vector whose entries are generators of an abelian group, and if A is an integral matrix (a relation matrix) such that the relations among the generators are given by $Ax = 0$ (the group of course is written additively), then we can use the Smith form to get a canonical set of generators and defining relations. Thus if $A = USV$, then $Ax = 0$ becomes $Sy = 0$, where the entries of $y = Vx$ are new generators, and the new relations are just single power relations. This idea can sometimes be used with infinite groups as well. Thus if G is a group, and if it can be shown that G/G' is infinite (G' being the commutator subgroup of G), then certainly G must also be infinite. In any case, the structure of G/G' can be worked out using the Smith form, provided that generators and relations are known for G (G/G' is just G abelianized).

As an example, let G be the group generated by x, y, z with defining relations

$$xyx = z^3,$$

$$yzy = x^3,$$

$$zxz = y^3.$$

In G/G', the relations and generators (written additively) become

$$2x + y - 3z = 0,$$

$$3x - 2y - z = 0,$$

$$x - 3y + 2z = 0,$$

or in matrix form, $Av = 0$, where

$$A = \begin{pmatrix} 2 & 1 & -3 \\ 3 & -2 & -1 \\ 1 & -3 & 2 \end{pmatrix}, \qquad v = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Now finding the Smith normal form, we get $A = USV$, where

$$U = \begin{pmatrix} 2 & 1 & 0 \\ 3 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \qquad V = \begin{pmatrix} 1 & -3 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}, \qquad S = \operatorname{diag}(1, 7, 0).$$

Then the canonical generators and relations become

$$x' = 0,$$

$$7y' = 0,$$

$$z' \text{ unrestricted},$$

where

$$x' = x - 3y + 2z,$$

$$y' = y - z,$$

$$z' = z;$$

so that G/G' is the direct product of a cyclic group of order 7 and an infinite cyclic group. Thus certainly the group G is infinite.

A THEORETICAL APPLICATION

A very useful result that can be derived from the Smith form is the following: Supose that A is an integral $n \times n$ matrix such that $\det A \equiv 1 \mod m$, where m is a positive integer. Then an integral $n \times n$ matrix B exists such that $\det B = 1$ and $B \equiv A \mod m$ [5, p. 36].

Without going into too much detail, this result can be used to exploit the connection between the matrix group $\mathrm{GL}(n, \mathbb{Z})$ and the matrix group $\mathrm{GL}(n, \mathbb{Z}/m)$, where $\mathbb{Z}/m$ is the ring of integers modulo m.

TWO CLASSICAL APPLICATIONS

Perhaps the best-known application of the Smith form occurs with respect to similarity. Suppose that A, B are $n \times n$ matrices over an algebraically closed field $\mathbb{F}$, and we want to know whether or not they are similar over $\mathbb{F}$;

i.e., whether or not a nonsingular $n \times n$ matrix T exists such that $B = TAT^{-1}$. Questions on similarity are difficult, but here there is an answer: A and B are similar over $\mathbb{F}$ if and only if $A - xI$ and $B - xI$ are equivalent over $\mathbb{F}[x]$ [5, p. 45].

Another application along these lines is to the proof of the fact that an $n \times n$ matrix over an algebraically closed field $\mathbb{F}$ can be diagonalized (by a similarity of course) if and only if the elementary divisors of $A - xI$ are simple. This will happen if the minimal polynomial $s_n(A - xI)$ of A has no repeated roots [5, p. 49].

As an example, choose

$$A = \begin{pmatrix} 3 & -3 & 0 \\ 2 & -2 & 0 \\ 1 & -1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 3 & -5 & 4 \\ 2 & -4 & 4 \\ 1 & -2 & 2 \end{pmatrix},$$

considered as matrices over the field of complex numbers $\mathbb{C}$. Then A and B have the same eigenvalues 0, 0, 1, but

$$S(A - xI) = \operatorname{diag}(1, x, x(x-1)),$$

$$S(B - xI) = \operatorname{diag}(1, 1, x^2(x-1)).$$

Thus A and B are not similar over $\mathbb{C}$. Also, the elementary divisors of $A - xI$ are x, x, and $x - 1$, while the elementary divisors of $B - xI$ are x^2 and $x - 1$. Thus A is similar to a diagonal matrix over $\mathbb{C}$, but B is not.

Many other applications can be given, but perhaps these will suffice. In broad terms, the utility of the Smith normal form rests in the fact that the problem under consideration is usually reduced to a number of independent linear problems.

POSSIBLE GENERALIZATIONS

When $R = \mathbb{Z}$, we can replace the group of all units $\mathrm{GL}(n, \mathbb{Z})$ by one of its subgroups G, and require that the definitions of equivalence use the elements of G, instead of all of the elements of $\mathrm{GL}(n, \mathbb{Z})$. Then the formal definition would say that two $n \times n$ matrices A and B over $\mathbb{Z}$ are *G-equivalent* if $B = UAV$, where U and V belong to G. When G is of finite index μ in $\mathrm{GL}(n, \mathbb{Z})$, it can be shown that the number of G-equivalence classes obtained in this fashion for matrices of determinant Δ is at most μ^2

times the number of ordinary equivalence classes for matrices of determinant Δ. However, the calculation of this number is still in general an open question.

If we change this problem slightly, and work only with matrices that belong to $GL(n, \mathbb{Z})$ [so that we are asking for the number of G-equivalence classes in $GL(n, \mathbb{Z})$ with respect to the elements of G], we have a fundamental group-theoretic problem at hand; namely, the double coset decomposition of a group with respect to one of its subgroups.

The difficulties that we encounter by extending the definition of equivalence are best illustrated by an example due to Shu-Chu Chang [1]. Let $\Gamma = SL(2, \mathbb{Z})$, and choose $G = \Gamma_0(p)$, the subgroup of Γ consisting of all matrices

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

of Γ for which $c \equiv 0 \bmod p$, where p is a prime. Then G is of index $p + 1$ in Γ, and a complete set of coset representatives (left or right) is given by

$$R_k = W^k, \quad 0 \leqslant k \leqslant p - 1, \qquad R_p = T,$$

where

$$W = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad T = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Now let A be any integral 2×2 matrix of determinant $\Delta \neq 0$. Then $A = USV$, where U and V are unimodular matrices, and $S = \operatorname{diag}(s_1, s_2)$ is the Smith normal form of A. Here, $s_1 \mid s_2$, and $s_1 s_2 = \Delta$. Thus $S = s_1 \operatorname{diag}(1, d)$, where $d = s_2/s_1$. Set $M = \operatorname{diag}(1, d)$. By considering U and V modulo G, we need only examine the matrices $R_i M R_j$, $0 \leqslant i, j \leqslant p$, for equivalence. It turns out that these reduce modulo G to M and TM when $(p, d) = 1$, and to M, TM, MT, and TMT when $p \mid d$. These are all in disjoint equivalence classes modulo G, and so the number of classes in this case is either 2 or 4. The first invariant factor s_1 also enters into this description, since each of the matrices above must be multiplied by s_1 to obtain the class representatives.

Work on the double coset problem resulted in a number of significant results, among which the following purely group-theoretic ones, due to Matthew Lazar [2], are perhaps the most interesting:

Let H_1 and K_1 be normal subgroups of G such that $H_1K_1 = K_1H_1 = G$. Let H and K be subgroups of G of finite index such that H contains H_1 and K contains K_1. Let $N = H \cap K$, and let x be any element of G. Then $(HxH) \cap (KxK) = NxN$.

Furthermore, assuming the conditions of the previous theorem, let n_1 equal the number of double cosets of (H, H) in G, n_2 the number of double cosets of (K, K) in G, and n_3 the number of double cosets of (N, N) in G. Then $n_3 = n_1n_2$.

These useful theorems can be used to prove, for example, the following:

As before, let $\Gamma = \mathrm{SL}(2, \mathbb{Z})$, and let $G = \Gamma_0(m)$, the subgroup of Γ consisting of all matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ for which $m \mid c$. Let $m = p_1^{e_1}p_2^{e_2} \cdots p_k^{e_k}$ be the canonical decomposition of m into prime powers. Let $f(m)$ equal the number of double cosets of (G, G) in Γ. Then if m is odd, $f(m)$ equals $\prod_{i=1}^{k} (2e_i)$; and if m is even, $f(m)$ equals $(e_1 + 1) \prod_{i=2}^{k} (2e_i)$, where $p_1 = 2$.

Similar formulas for higher dimensional groups [i.e. for subgroups of $\mathrm{SL}(n, \mathbb{Z})$] have also been derived.

Most of the applications have involved the groups $\Gamma_0(m)$. These are of the highest interest now, because of their appearance in algebraic geometry, modular functions, the proof of Fermat's theorem, etc.

We leave the discussion at this point, but the ideas given above are currently the subject of intense investigation.

REFERENCES

1 S.-C. Chang, Two-Sided Equivalence with Respect to Subgroups of the Modular Group, Ph.D. Thesis, Univ. of California, Santa Barbara, 1990, unpublished.

2 M. Lazar, Two-Sided Equivalence, Double Cosets, Ph.D. Thesis, Univ. of California, Santa Barbara, 1994, unpublished.

3 C. C. MacDuffee, *The Theory of Matrices*, Chelsea, New York, 1966.

4 M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities*, Allyn and Bacon, Boston, 1964.

5 M. Newman, *Integral Matrices*, Academic Press, New York, 1972.

6 M. Newman, *Algorithmic Matrix Theory*, Monograph, Inst. for Interdisciplinary Applications of Algebra and Combinatorics, Univ. of California, Santa Barbara, 1980.

7 M. Newman, A result about determinantal divisors, *Linear and Multilinear Algebra* 11:363–366 (1988).

8 M. Newman and R. C. Thompson, Matrices over rings of algebraic integers, *Linear Algebra Appl.* 145:1–20 (1991).
9 J. J. Rushanan, Eigenvalues and the Smith normal form, *Linear Algebra Appl.* 216:177–184 (1995).
10 R. C. Thompson, An inequality for invariant factors, *Proc. Amer. Math. Soc.* 86:8–11 (1982).

Received 2 October 1995; final manuscript accepted 9 October 1995