A Note to Hermite and Smith Normal Form Computation

Tomáš Hrúz*

June 1, 1995

Abstract

Hermite normal form and Smith normal form are canonic forms of integral matrices with respect to congruence produced by multiplication with right unimodular matrix resp. by multiplication with left and right unimodular matrix simultaneously. This lower-diagonal resp. diagonal form has many applications in various fields. Unfortunately, the computation of them is extremely resource consuming. The present note shows a construction of certain more complicated instances of integral matrices. The systematic construction of such matrices complements a research with respect to randomized versions of Hermite normal form and Smith normal form algorithms [11].

Keywords: number theory, operation research, integral optimization, Hermite normal form, Smith normal form.

1 Introduction

In many applications where computers are used to solve or to optimize problems, it has been proved useful to formulate various combinatorial relations of the problem in terms of integral matrices. It is a basic formalism for integral optimization. The same holds for many problems "inside" computer science and mathematics.

Once we have formulated the problem in terms of integral matrices it is often crucial to know a canonic form of those matrices [1]. The Hermite normal form and Smith normal form of integral matrix play an important role among various canonic forms which can be defined. The applications of Hermite normal form and Smith normal form can be traced back to various fields like integer programming, system theory, algebraic theory of control, algebraic group theory etc. For recent survey see for example [13].

The main obstacle in obtaining these canonic forms consists in extreme growth of intermediate numbers during computation. Even there was a time when it seemed that the known algorithms are exponential in *number of bits* used during the computation. The first algorithms with provable polynomial upper bound in number of bits were that of Kannan, Bachem [6] and Frumkin [5]. If the algorithm used is not highly optimized one, the intermediate results can be easily in orders larger than the determinant of a matrix, which in turn is usually much larger than the matrix entries. For example, in [12], there has been a rather spectacular example reported where a 20x20 matrix has been found with entries not exceeding 10 but with numbers of order 10^{5011} occurring during the computation. The determinant is bounded in this case by the value 10^{33}.

The present contribution is related to the following question. During the computation of Hermite normal form (HNF) resp. Smith normal form (SNF) it is worthwhile to know a gcd of a

*Slovak Technical University, Faculty of Mechanical Engineering Námestie Slobody 17, 812 31 Bratislava, Slovak Republic, tomas@sdxnet.com. This work has been supported by DAAD grant under the reference 323-gm-cm together with kind support of Professor A. Bachem and ZPR, University of Cologne.

sub-matrix of the given matrix. The question is whether we can suppose that certain "limited" computation (let us say using only few elements in matrix) can give us this number. The special construction of this note shows that this is not true in general. There are matrices for which all elements must be taken into the account when we are computing the gcd of them. On the other hand such complicated matrix instances are very large, raising a question whether we could not compute these numbers with high probability. Indeed, this is true as was shown in randomized versions of HNF and SNF algorithms [11].

We construct a matrix using a set of different primes which has a property that any gcd computation using row elementary operations between any two elements of a matrix will not produce a gcd of all elements of the matrix. Even more the gcd of all elements is not equal to the gcd of any column of the matrix. For example the following 4x4 matrix has the above property (even more, any gcd of any three elements in the same column does not equal to the gcd of the whole matrix which is 1):

$$\begin{pmatrix} 3150 & 170170 & 868434 & 2485830 \\ 4410 & 190190 & 1036518 & 2803170 \\ 11025 & 293930 & 1397046 & 3213390 \\ 7350 & 248710 & 1108002 & 3063930 \end{pmatrix}$$

2 Hermite and Smith Normal Form

An integral matrix $\mathbf{U}$ with $\mid \det(\mathbf{U}) \mid= 1$ is called *unimodular.* For convenience, throughout we assume matrices are $n \times n$ and nonsingular; there is a natural generalization for arbitrary integral matrices. Every nonsingular integral matrix has a lower triangular integral canonic form called Hermite normal form (see [4], theorem 2 below). The Hermite normal form of matrix $\mathbf{A}$ is a unique representative of an equivalent class of matrices with respect to congruence produced by multiplication with right unimodular matrix: $\mathbf{A} \equiv_H \mathbf{B}$ iff $\mathbf{AR} = \mathbf{B}$ where $\mathbf{R}$ is unimodular. Another very important congruence is: $\mathbf{A} \equiv_S \mathbf{B}$ iff $\mathbf{LAR} = \mathbf{B}$ where $\mathbf{L}$,$\mathbf{R}$ are unimodular. Canonic form with respect to this congruence is called Smith normal form (see [10], theorem 3 below) and it is a diagonal integral matrix.

Definition 1 (Elementary operations) *Let us have an integer matrix* $\mathbf{A}$. *We call the following three kinds of operations on columns of the matrix* $\mathbf{A}$ *elementary operations:*

1. *Exchange of the columns i and j. We denote this operation by $\mathbf{S^c}(i,j)$ $i<j$ or $\mathbf{S}$ when only a type of the operation is of interest.*

2. *Multiplication of the column i by -1. This operation is denoted by $\mathbf{M^c}(i)$ or $\mathbf{M}$ when only a type of the operation is of interest.*

3. *Addition of k times $i-th$ column to the $j-th$ column. This operation is denoted by $\mathbf{A^c}(\overrightarrow{i,j},k)$ when $i<j$ and $\mathbf{A^c}(\overleftarrow{j,i},k)$ when $i>j$ or $\mathbf{A}$ when only a type of the operation is of interest.*

Theorem 2 (Hermite [4]) *Given a nonsingular $n \times n$ integral matrix $\mathbf{A}$, there exists a $n \times n$ unimodular matrix $\mathbf{R}$ such that $\mathbf{H} = \mathbf{AR}$ is a lower triangular with positive diagonal elements. Further, each off-diagonal element of $\mathbf{H}$ is non-positive and strictly less in absolute value than the diagonal element in its row. $\mathbf{H}$ is called the Hermite normal form of $\mathbf{A}$.*

It is known (see for example [7]) that the Hermite normal form of a matrix $\mathbf{A}$ is unique and the right unimodular matrix is unique too . Elementary operations from Definition 1 can be realized through multiplication of the matrix $\mathbf{A}$ by special matrices called *elementary matrices.* Elementary matrices are unimodular.

Theorem 3 (Smith [10]) *Given a nonsingular $n \times n$ integral matrix $\mathbf{A}$, there exists $n \times n$ unimodular matrices $\mathbf{L}$,$\mathbf{R}$ such that $\mathbf{S} = \mathbf{LAR}$ is a diagonal matrix with positive diagonal elements $d_1, \ldots, d_n$ such that d_i divides d_{i+1} $(i = 1, \ldots, n-1)$.*

The matrix $\mathbf{S}$ is unique similarly as with Hermite normal form but the matrices $\mathbf{L}$,$\mathbf{R}$ *are not unique.* The unimodular matrices $\mathbf{L}$, $\mathbf{R}$ are a product of elementary matrices. The only difference is that a left multiplication of a matrix with some unimodular matrix means a sequence of *row* elementary operations similar to those in Definition 1. We denote them $\mathbf{S^r}(i,j)$, $\mathbf{M^r}(i)$,$\mathbf{A^r}(\overrightarrow{i,j},k)$, $\mathbf{A^r}(\overleftarrow{i,j},k)$

The 2×2 matrix $\mathbf{E_x} = \begin{pmatrix} p & -A_{1,2}/g \\ q & A_{1,1}/g \end{pmatrix}$ where g is a great common divisor of $A_{1,1}$, $A_{1,2}$ and p, q are numbers satisfying $A_{1,1}p + A_{1,2}q = g$. Unimodular operation represented by matrix $\mathbf{E_x}$ embedded in $\mathbf{E_x}$ plays an important role both in theory and computation. We call this kind of operation *Euclidean* and denote it by $\mathbf{C_E}(\overrightarrow{i,j},k)$ resp. $\mathbf{C_E}(\overleftarrow{i,j},k)$. We use a word *unfolding* to denote a process between certain elements in matrix leading subsequently to their gcd. By *folding* we mean similar sequence of operations but divergent.

To give the reader a flavor of algorithmic approach we can think about the following simple algorithm for Hermite normal form. Take two nonzero elements in the first row, use elementary column operations to subtract the smaller number from the larger. After finite number of repetitions only one element in the first row will be nonzero. Put this element to the (1,1) position, strip out the first row and column and proceed with the resulting sub-matrix. After obtaining lower-diagonal matrix reduce under diagonal elements to the left.

Unfortunately, the above algorithm suffers from intermediate entry explosion. This phenomenon was a reason of research in this field which we survey briefly in the next section.

2.1 Algorithms for Hermite and Smith Normal Form

The computational aspects of Hermite normal form and Smith normal form begun to be more widely investigated about 1950. In 1952 Rosser [8] proposed an algorithm to compute the Hermite normal form using only elementary operations (see Definition 1). Smith normal form is treated from an algorithmic point of view in [9].

Early it was seen that attention must be given to the phenomenon associated: when unfolding operations occur in a row of the matrix (leading to the gcd of the row resp. column elements), a folding occurs at the same time in other rows of the matrix. Entries become to be very large. This problem was explicitly formulated by several authors and was called expression swell.

The next important step was made by Bradley in [2] where he introduced the operation which we call Euclidean. From that time on all proposed algorithms use Euclidean operations as a basic tool and the main attention is given to the analysis of expression swell even if the algorithm of Rosser is experimentally still investigated. But the theoretical analysis remains open.

Kannan and Bachem proposed the first algorithms with a known polynomial expression swell in 1979 [6]. They proved the upper bound on the number of bits for the largest matrix entry to be $n^3(\log_2(n) + \log_2 || \mathbf{A} ||)$. Chou and Collins in 1982 introduced a new algorithm for Hermite normal form where by reordering the computation in Kannan - Bachem process they achieved an upper bound of $n(\log_2(n) + \log_2 || \mathbf{A} ||)$.

In 1987 Domich, Kannan and Trotter introduced a modulo determinant arithmetic [7]. They proved an upper bound $n(\log_2(n) + \log_2 || \mathbf{A} ||)$ on the number of bits for their modular method. A modulo determinant computation has one very important feature, it can be combined with any other method of computing a Hermite normal form of matrix. Because Hermite normal form of an integral matrix can be computed with modulo determinant arithmetic, also Smith normal form can be computed using modulo determinant arithmetic when we use successive Hermite normal form computations with the method developed in [7]. Similar method of Smith normal form computation is also presented by Iliopoulos in [3].

Another method of Hermite normal form and Smith normal form computation consists in residual computation modulo relatively prime factorization of determinant. Computation modulo various prime factors was also considered but the problem is to choose adequately the prime moduli to obtain correct Smith normal form of a given matrix. The answer is known only for special classes of matrices.

Recently new methods have been developed to compute Smith normal form via randomization technique as described in [11].

The question whether the number of bits for the largest matrix entry is polynomialy bounded in the Bradley algorithms for Hermite normal form and Smith normal form is still open even if some experiments seem to show that it is not.

3 Construction of a Matrix with Global Gcd Different from Local Computed Gcds

In the following the symbol $A \subset B$ means that A is a proper subset of B e.g. $A \subseteq B; A \neq B$.

Lemma 4 *Let U is infinite countable set (universum). Then for every $n \in N$; $n \geq 2$ there is a system of finite nonempty sets $S_1, S_2, \ldots, S_n$; $S_i \subset U$ with the following properties:*

1.

$$\bigcap_{i=1}^{n} S_i = \emptyset$$

2. $\forall K$; $\emptyset \subset K \subset \{1, 2, \ldots, n\}$ *it holds*

$$\bigcap_{i \in K} S_i \neq \emptyset$$

Proof. We can suppose that U is enumerated in the sequence $(u_1, u_2, \ldots, u_3)$. During the construction we will subtract subsequently more and more elements from the set U and use them as elements of the constructed sets $S_1, S_2, \ldots, S_n$.

The construction proceeds by induction, let $n = 2$. Then the sets $S_1 = \{u_1, u_2\}$ and $S_2 = \{u_2, u_3\}$ fulfill the required properties. Let us suppose that we have already constructed a system of sets $S_1, S_2, \ldots, S_{n-1}$ which fulfills the both properties. The set S_n is constructed in the following way: We take a new element u_{k_n} from the sequence U. This element is not contained in any of the sets $S_1, S_2, \ldots, S_{n-1}$. Let $S'_i = S_i \cup u_{k_n}$. That means, the new element is added to each of the sets $S_1, S_2, \ldots, S_{n-1}$. Now the property 2. is clearly fulfilled for the system of sets $S'_1, S'_2, \ldots, S'_{n-1}$ but also $\bigcap_{i=1}^{n-1} S'_i \neq \emptyset$. Now we proceed by induction through the cardinality of K. At the beginning we take arbitrary element from each of the sets $S_1, S_2, \ldots, S_{n-1}$ and unify them into the new set $S_n^{(1)}$. Now the property 2. is true for all K with cardinality 1. Let us suppose that property 2.is true for all K with $\mid K \mid = l, l < n - 1$. Fix arbitrary K, that has a cardinality $l + 1$. There are two possibilities. Either, K does not contain n or it contains n. In the former case $\bigcap_{i \in K} S'_i \neq \emptyset$ is trivially valid because the intersection contains at least the element u_{k_n}. In the former case we enlarge S_n with the intersection of all S_i except the one for which $i = n$ e.g. $S_n^{(l+1)} = S_n \cup \bigcap_{i \in K \setminus \{n\}} S_i$. Now take as a new system the sets $S'_1, S'_2, \ldots, S'_{n-1}, S_n^{(n-1)}$. From the construction the system fulfills the property 2. The property 1. is also true because the intersection $\bigcap_{i=1}^{n-1} S'_i$ contains exactly the element u_{k_n} but the set $S_n^{(n-1)}$ has been exclusively constructed from the elements of the sets $S_1, S_2, \ldots, S_{n-1}$. Therefore $\bigcap_{i=1}^{n-1} S'_i \cap S_n^{(n-1)} = \emptyset$. $\boxed{\text{QED}}$

Lemma 5 *Let the symbol* $\bigwedge_{i=1}^{n} a_i$ *denotes for a moment the gcd of n elements* $\{a_1, a_2, \ldots, a-n\}$. *Then for each* $n \geq 2$ *there is an integral vector* $(a_1, a_2, \ldots, a_n)$ *for which* $\bigwedge_{i=1}^{n} a_i = 1$ *but* $\forall K$; $\emptyset \subset K \subset \{1, 2, \ldots, n\}$ $\bigwedge_{i \in K} a_i > 1$.

Proof. We can take as a universum U in Lemma 4 the set of all primes, enumerated according their size. Fix arbitrary $n \geq 2$. According to the Lemma 4 there is a system of sets $S_1, S_2, \ldots, S_n$ each consisting of different primes for which the property 1. and 2. of Lemma 4 holds. Then put $a_i = \prod_{p_k \in S_i} p_k$. The Lemma 4 now implies the required statement. $\boxed{\text{QED}}$

Now, take subsequently new primes and construct vectors $\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_n$ each consisting of n different numbers according to Lemma 5. Construct also another vector $(r_1, r_2, \ldots, r_n)$ in the same way. Then the matrix consisting of columns $(r_1 \underline{c}_1^T, r_2 \underline{c}_2^T, \ldots, r_n \underline{c}_n^T)$ has the property we are looking for. Actually, we can take as a vector $(r_1, r_2, \ldots, r_n)$ one of the vectors $\underline{c}_i$. When $n = 4$ we obtain the following matrix:

$$\begin{pmatrix} 3150 & 170170 & 868434 & 2485830 \\ 4410 & 190190 & 1036518 & 2803170 \\ 11025 & 293930 & 1397046 & 3213390 \\ 7350 & 248710 & 1108002 & 3063930 \end{pmatrix}$$

4 Conclusion

The general construction presented in this paper which leads to the matrix instances with high complexity provides a deeper insight into the structure of HNF and SNF computation. It can also serve as a test-bed for algorithms computing Hermite normal form and Smith normal form.

One interesting fact about this construction is that even in low dimension the resulting matrices are very large. This complements the research branch dealing with randomized algorithms for HNF and SNF. If the construction could be proved to be optimal, further consequences can be derived for improvements of randomized algorithms as well as general algorithms for HNF and SNF.

References

[1] M. Newman, Integral matrices, AP New York, 1972.

[2] G. H. Bradley, Algorithms for Hermite and Smith Normal Form Matrices and Linear Diophantine Equations, Math. Comp. 25, 1971, pp. 897-907.

[3] Costas S. Iliopoulos, Worst-Case Complexity Bounds on Algorithms for Computing the Canonical Structure of Finite Abelian Groups and the Hermite and Smith Normal Forms of an Integer Matrix, SIAM. J. Comput., Vol. 18, No. 4, August 1989, pp. 658-669.

[4] C. Hermite, Sur l'Introduction des Variables Continues dans la Theorie des Nombres, J. Reine Angew. Math. 41, 1851, pp. 191-216.

[5] M. A. Frumkin, *Polynomial Time Algorithms in the Theory of Linear Diophantine Equations*, in M. Karpinski (Ed.), Fundamentals of Computation Theory. Springer, Berlin and New York, 1977, Lecture Notes in Computer Sci. 56, pp. 386-392.

[6] R. Kannan and A. Bachem, Polynomial Algorithms for Computing the Smith and Hermite Normal Forms of an Integer Matrix, SIAM J. Comput., 9, 1979, pp. 499-507.

[7] P. D. Domich, R. Kannan and L. E. Trotter, Jr. Hermite normal form computation using modulo determinant arithmetic, Math. of Operating Research, Vol. 12, No. 1, February 1987, pp. 50-59.

[8] J. B. Rosser, A Method of Computing Exact Inverse of Matrices with Integer Coefficients, J. Res. Nat. Bur. Standarts, 49, 1952, pp. 349-358.

[9] D. A. Smith, A Basis Algorithm for Finitely Generated Abelian Groups. Math. Algorithms, 1, 1966, pp. 13-26.

[10] H. J. S. Smith, On systems of indeterminate equations and congruences, Philos. Trans., 151, 1861, pp. 293-326.

[11] E. Kaltofen, M.S. Krishnamoorthy and B.D. Saunders, Mr.Smith Goes to Las Vegas: Randomized Parallel Computation of The Smith Normal Form of Polynomial Matrices, EUROCAL'87 European Conference on Computer Algebra, Proceedings, Springer-Verlag 1989, pp. 317-322.

[12] J. L. Hafner, K. S. McCurley, Asymptotically Fast Triangularization of Matrices over Rings, SIAM J. Comput., Vol. 20, No. 6, December 1991, pp. 1068-1083.

[13] T.Hruz and D.Fortin, Parallelism in Hermite and Smith Normal Forms, INRIA research report RR 2077, October 1993, ftp://ftp.inria.fr/INRIA/publication/ publi-ps-gz/RR/RR-2077.ps.gz