Parallel Algorithms for Matrix Normal Forms

Erich Kaltofen
M. S. Krishnamoorthy

Rensselaer Polytechnic Institute
Department of Computer Science
Troy, New York 12181

B. David Saunders

University of Delaware
Department of Computer and Information Sciences
Newark, Delaware 19716

ABSTRACT

Here we offer a new randomized parallel algorithm that determines the Smith normal form of a matrix with entries being univariate polynomials with coefficients in an arbitrary field. The algorithm has two important advantages over our previous one: the multipliers relating the Smith form to the input matrix are computed, and the algorithm is probabilistic of Las Vegas type, i.e., always finds the correct answer. The Smith form algorithm is also a good sequential algorithm. Our algorithm reduces the problem of Smith form computation to two Hermite form computations. Thus the Smith form problem has complexity asymptotically that of the Hermite form problem. We also construct fast parallel algorithms for Jordan normal form and testing similarity of matrices. Both the similarity and non-similarity problems are in the complexity class **RNC** for the usual coefficient fields, i.e., they can be probabilistically decided in poly-logarithmic time using polynomially many processors.

1. Introduction.

The different normal forms of matrices, Hermite, Smith and Jordan Normal Forms are widely used in many different branches of science and engineering. Sequential algorithms for computing these normal forms have been given previously. With advances in parallel hardware and software, development of parallel algorithms is not only an intellectual exercise but also a practical feasibility.

* This material is based upon work supported by the National Science Foundation under Grant No. DCR-85-04391 (first author), Grant No. MCS-83-14600 (second and third author), and by an IBM Faculty Development Award (first author). This paper appears in *Linear Algebra and Its Applications* **136**, pp. 189-208 (1990). The results in Section 2 and 4 were first announced in the *Proc. EUROCAL '87, Springer Lect. Notes Comput. Sci* **378**, pp. 317-322 (1989).

This paper is third in a series on canonical forms of matrices [13], [12]. Here we offer a new randomized parallel algorithm that determines the Smith normal form of a matrix in $F[x]^{m \times n}$. The algorithm has two important advantages over our previous one. The multipliers relating the Smith form to the input matrix are computed and the algorithm is of Las Vegas type, that is, the result is guaranteed correct, probability only enters in speed considerations. The Smith form algorithm is also a good sequential algorithm, faster than previous methods in the worst case. Its speed is that of the Hermite form algorithm on which it depends. One can use any of the algorithms by Kannan & Bachem [15], Kannan [14], Chou & Collins [2], or Iliopoulos [10].

A sequential solution to the Smith normal form problem proceeds by iterating Hermite normal form computations on the matrix (see, e.g. [15])., Although in practice usually two Hermite iterations suffice, there are input matrices for which the number of iterations is at least linear in the dimension of the matrix. Here we show that by multiplying the input matrix with a certain randomly chosen matrix, the new randomized matrix will require with high probability only two Hermite steps before the Smith normal form appears. The proof of this fact uses ideas similar to those for our Monte Carlo Smith normal form algorithm [13], but is more complicated. An "unlucky" premultiplication is discovered immediately if after two Hermite steps we do not obtain a Smith normal form. The point is now that if we do, we must have the unique Smith normal form of the input matrix together with the unimodular pre- and post-multipliers. Since the Hermite normal form algorithms are deterministic [12], the entire algorithm is Las Vegas.

In this paper we also construct fast parallel algorithms for Jordan Normal Form and testing similarity between matrices. We will show that both similarity and non-similarity can be decided in $\mathbf{RNC}^2$. We refer to [3] for the definition of the complexity classes **NC** and **RNC** of problems (probabilistically) solvable by uniform families of Boolean circuits of poly-logarithmic depth and polynomial size. We note that since the class **RNC** requires us to perform field operations on Boolean circuits, the previous claim is precise only for concrete fields such as the rationals **Q** or $\mathbf{F}_q$, the finite field with q elements. Our algorithms are randomized in the Las Vegas sense, that is they can fail but they will never give an incorrect answer.

We will also provide a parallel algorithm for computing the Jordan normal form of a given matrix $A \in F^{n \times n}$ in **RNC**. The entries of the Jordan normal form in general lie in an algebraic extension of the original field F, and we need to attach to each distinct (symbolic) eigenvalue λ_i a polynomial $h_i(x) \in F[x]$ with $h(\lambda_i) = 0$. The polynomials h_i are squarefree, identical or pairwise relatively prime, and all their roots occur among the λ_i. In fact, the Jordan block structure corresponding to different eigenvalues with identical defining equations will be the same. The construction of h_i assumes that F is perfect and that we can take p-th roots in case its characteristic is $p > 0$.

2. Echelon and Hermite Forms.

In this section we give a fast parallel algorithm to compute a canonical form for column equivalence of matrices built from our Hermite form algorithm [13]. This algorithm is needed for the parallel version of the Smith form algorithm of the next section.

Matrices, A and B, in $F[x]^{m\times n}$ are *column equivalent* if there exists unimodular Q such that $AQ = B$. A matrix in $F[x]^{n\times n}$ is *unimodular* if its determinant is a nonzero element of F. Unimodular matrices are precisely the ones with an inverse in $F[x]^{n\times n}$.

A variety of canonical or almost canonical forms for column or row equivalence have been given in the past, but we have not found one in the literature which completely meets our needs. For example, linear algebra texts often present an echelon form for matrices over a field, see [9]. An echelon form has the advantage, needed here, that for rank r, the leading r columns are independent. However, here we need the form over a PID. For matrices over a PID, Hermite presented a canonical (triangular) form for nonsingular square matrices. This has been often extended to arbitrary square matrices by allowing zeroes on the diagonal [17], [19]. One gives up uniqueness of the form in the process. For example, all strictly lower triangular matrices are in Hermite form by this definition even though large collections of them are equivalent. Though the form may be easily extended to rectangular matrices, the lack of uniqueness means that structure, such as rank, that might be revealed by a canonical form for row equivalence is not. For these reasons, we choose to extend the notion of column echelon form to matrices of arbitrary shape and rank over a PID.

A matrix H is in *column echelon form* if

(1) nonzero columns precede zero columns,

(2) the leading nonzero element in a nonzero column is below the leading nonzero element in preceding columns and above leading nonzero element of succeeding columns.

(3) the leading nonzero element in each column is monic,

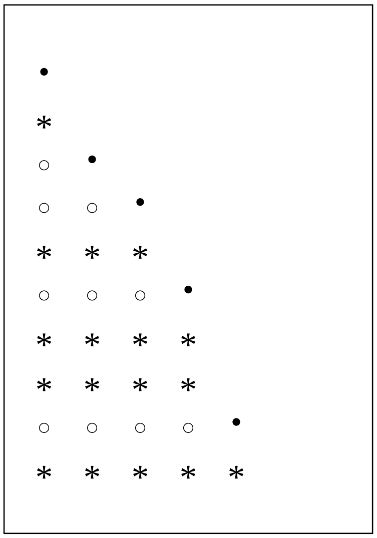
(4) in each row which contains the leading nonzero element of some column, the entries preceding that entry are of lower degree.

We denote by C_i^n the set of all length i subsequences of (1, ..., n) and by $A_{I,J}$, $I \in C_i^n$, $J \in C_i^n$, the $i \times i$ determinant of the submatrix in the rows I and columns J.

2.1 Theorem.

(1) Column equivalent matrices have the same left kernel (row dependencies).

Figure 1: Layout of a column echolon form of a matrix of rank 5. The • entries are monic, the ○ entries are residues with respect to them, and the $*$ entries are the remaining possibly non-zero entries.



$$\begin{array}{ccccc}
\bullet &&&& \\
* &&&& \\
\circ & \bullet &&& \\
\circ & \circ & \bullet && \\
* & * & * && \\
\circ & \circ & \circ & \bullet & \\
* & * & * & * & \\
* & * & * & * & \\
\circ & \circ & \circ & \circ & \bullet \\
* & * & * & * & *
\end{array}$$

(2) Let I be a fixed sequence of i rows. Column equivalent matrices have the same greatest common divisor of all $i \times i$ minors in the rows I, i.e., $d_I := \mathrm{GCD}_{J \in C_i^n} (\det(A_{I,J}))$ is an invariant.

(3) Each matrix A in $F[x]^{m \times n}$ is column equivalent to a unique matrix H in column echelon form. If the rank of A is m, the unimodular cofactor Q, such that $AQ = H$, is also unique.

Proof. These are standard results, cf [16],. [19]. In the literature, we have not found a unique canonical form over a PID and including the singular matrices, but see [21], Chapter 6. Therefore, we offer a proof of the uniqueness of the echelon form. It suffices to show that if $HQ = K$, H and K are in column echelon form, and Q is unimodular, then $H = K$ (and $Q = I$, when the rank is m). By the invariance of row dependencies, the echelon patterns of H and K are the same. Suppose the rank is r. We first permute the rows so that the rows containing the leading nonzero entries of the first r columns are at the top, thus we premultiply by permutation matrix P so that

$$PH = \begin{bmatrix} H_1 & 0 \\ H_2 & 0 \end{bmatrix}, \; PK = \begin{bmatrix} K_1 & 0 \\ K_2 & 0 \end{bmatrix},$$

where H_1 and K_1 are $r \times r$ nonsingular matrices in Hermite form. If we conformally block Q, we have

$$\begin{bmatrix} K_1 & 0 \\ K_2 & 0 \end{bmatrix} = PK = PHQ = \begin{bmatrix} H_1 & 0 \\ H_2 & 0 \end{bmatrix} \begin{bmatrix} Q_1 & Q_3 \\ Q_2 & Q_4 \end{bmatrix} = \begin{bmatrix} H_1Q_1 & H_1Q_3 \\ H_2Q_1 & H_2Q_3 \end{bmatrix}$$

Looking at the first row, by the uniqueness result for the Hermite form and unimodular cofactor in the square nonsingular case [19], we have $H_1 = K_1$, and $Q_1 = I_r$. Since H_1 is nonsingular, Q_3 is 0. Q_2 is arbitrary and Q_4 is arbitrary unimodular, but when $r = m$, $Q = Q_1$. □

Most algorithms for the Hermite form have been described for the nonsingular case but extend naturally to echelon form algorithms for the general case. However, our deterministic parallel algorithm (in $\mathbf{NC}^2$, cf. Cook [3] for a description of this parallel computation model) requires a bit more effort. We offer an extended algorithm here.

2.2 Algorithm $(H, Q) \leftarrow CEF(A)$.
[Column Echelon Form. This is a fast parallel algorithm when F is a finite algebraic extension of a prime field.]
Input: $A \in F[x]^{m \times n}$, F a field.
Output: Unimodular $Q \in F[x]^{n \times n}$ and column echelon form $H \in F[x]^{m \times n}$ such that $AQ = H$.

(1) [Find leading independent rows:]
$A' \leftarrow$ the $(m+n) \times n$ matrix $[A^T \; I_n]^T$.
$A'' \leftarrow$ the first n independent rows of A'.
(Compute in parallel the rank [18] of each matrix consisting of the first i rows of A'. Then include the i-th row in A'' if the i-th rank is greater than the $i-1$st rank (the 0-th rank is 0)).
[A'' is $n \times n$. If r is the rank of A, the first r rows of A'' are from A and the remaining $n-r$ from I_n.]

(2) [Hermite form - nonsingular column echelon form:]
$(H'', Q) \leftarrow$ the Hermite form of A'' and the corresponding unimodular cofactor. (Computed by the parallel algorithm of [13]).

(3) [Column echelon form:] $H \leftarrow AQ$. Return Q and H. □

2.3 Theorem. Algorithm CEF to compute the column echelon form and associated unimodular cofactor of a matrix is correct and is in $\mathbf{NC}^2$ when F is a finite algebraic extension of a prime field.

Proof. Each of the three steps is in $\mathbf{NC}^2$. The first can be done because rank is in $\mathbf{NC}^2$. The second step is by [13], and the third, matrix multiplication, is in $\mathbf{NC}^1$. It remains to show correctness.

Let r be the rank of A, and let k_1 , ..., k_r be the indices of the first r independent rows of A. Then for $i \in \{1, ..., r\}$, the row $A_{k_i,*} = A''_{i,*}$, and

$$H_{k_i,*} = A_{k_i,*}Q = H''_{i,*} = (h_{k_i,1}, \ldots, h_{k_i,i}, 0, \ldots, 0).$$

Then H is in echelon form unless some row before the i-th has its last nonzero entry in a column i or greater. Suppose k is the index of such a row, then the i rows of A numbered k_1 , ..., k_{i-1}, and k are independent, contradicting the fact that the rows numbered k_1, ..., k_i are the first i independent rows. □

Row echelon forms are defined by transposing everything in the above. Specifically, the row echelon form and unimodular $m \times m$ cofactor may be computed as follows:

2.4 Algorithm $(Q, H) \leftarrow REF(A)$.
[Row Echelon Form]

(1) $(Q', H') \leftarrow CEF(A^T)$.

(2) $H \leftarrow H'^T$. $Q \leftarrow Q'^T$. [$QA = H$] Return Q and H. □

3. A Smith Form Algorithm.

A matrix in $F[x]^{m \times n}$ is in *Smith normal form* if it is diagonal, the diagonal entries are monic or zero, and each divides the next. Matrices A and B in $F[x]^{m \times n}$ are *equivalent* if there exist unimodular matrices U in $F[x]^{n \times n}$ and V in $F[x]^{m \times m}$ such than $A = UBV$.

3.1 Theorem.

(1) Equivalent matrices have the same determinantal divisors. The i-th *determinantal divisor* of a matrix is the greatest common divisor of all $i \times i$ minors of the matrix. We denote it by s_i^*.

(2) The Smith normal form is a canonical form for equivalence, that is, there is one and only one matrix in Smith form equivalent to a given matrix. The diagonal entries of the Smith form are called the *invariant factors* of the matrix. The i-th invariant factor is $s_i = s_i^* / s_{i-1}^*$ ($s_1 = s_1^*$).

Proof. See, for example, Newman [19], Section 15. □

3.2 Algorithm. $(U, S, V) \leftarrow SNF(A)$.
[Smith Normal Form. Randomizing algorithm.]
Input: A, a matrix in $F[x]^{m \times n}$, where F is a field.
Output: U, S, and V, such that $UAV = S$, S in $F[x]^{m \times n}$ is in Smith form, U in $F[x]^{m \times m}$ is unimodular, and V in $F[x]^{n \times n}$ is unimodular.
Constant: ε, $0 < \varepsilon < 1$, the probability of failing on one try.

(1) [Randomize:] $d \leftarrow \max_{i,j} \deg(a_{i,j})$.
$R' \leftarrow$ a strictly lower triangular $n \times n$ matrix whose entries are chosen at random from C, a subset of F of size $c = 2d \min(m, n)^3/\varepsilon$. [If F has characteristic 0, C may be the integers 1 to c. The size c guarantees that the probability of having to repeat the algorithm is less

than ε. If F is finite of insufficient size, C may be a subset of an algebraic extension of F.]
$R \leftarrow I + R'$ [an invertible matrix].
$A' \leftarrow AR$.

(2) [Row operations:] $(U, H) \leftarrow REF(A')$ [row echelon form: $H = UA'$ $(= UAR)$.]
[The diagonal entries of H are now almost surely the invariant factors sought.]

(3) [Column operations:] $(S, V') \leftarrow CEF(H)$ [column echelon form: $S = HV'$ $(= UARV')$].
[This is expected to be an especially simple echelon form computation. For the most part exact divisions are needed, not GCD's. V' will be very nearly unit upper triangular.]

(4) If S is in Smith form (that with probability $\geq 1 - \varepsilon$), $V \leftarrow RV'$. Return U, S, and V.
[One could repeat with S as the input to take advantage of progress made. However, our point is that repetition will not be necessary.] □

Notice that the choice of a unit lower triangular random multiplier R makes the proof of the following theorem substantially more complicated. However, this choice is preferable, since then one never needs to check R for invertibility and one needs fewer random elements.

3.3 Theorem. Algorithm SNF is correct. It requires repetition only with probability $< \varepsilon$. Hence it is in Las Vegas $\mathbf{RNC}^2$ and runs sequentially in expected time O(CEF time) when F is a finite algebraic extension of a prime field.

Proof. It is clear from the construction that the output conditions are satisfied when the algorithm terminates. The algorithm terminates in k or fewer repetitions with probability $1 - \varepsilon^k$, which converges to 1 exponentially fast in k.

It remains to show that the probability that S is not in Smith form is less than ε. We do this with the aid of some lemmas.

From the first, we see that S, computed in step (3), will be in Smith form if H has the property that its first $r - 1$ diagonal entries are the first $r - 1$ invariant factors of A. The remaining lemmas enable us to conclude that H, computed in step (2), has that property unless the random entries of R, chosen in step 1, form a root of a certain polynomial π. By a lemma of Schwartz [20], the probability that we pick such an unlucky root is $\deg(\pi)/c$.

A suitable π is the product of the polynomials π_i of lemma 3.7, for $1 < i < r$. Each π_i is of degree bounded by $2i^2 d + i$. Thus we may bound the degree of π by $2N^3 d$ for $N = \min(n, m)$. Since we choose $c = 2N^3 d/\varepsilon$ in the algorithm, we obtain the desired probability, ε.

Since the expected number of repetitions is $\varepsilon + 2\varepsilon^2 + 3\varepsilon^3 + \cdots = \varepsilon/(1 - \varepsilon)^2$, a constant, and the time of one repetition is dominated by the time for echelon form computation, the parallel and sequential running times are those of CEF. □

3.4 Lemma (A condition under which one more echelon form suffices). Let H be a row echelon form of rank r with $h_{i,i}$ the leading nonzero entry of row i, for $i = 1, ..., r$. Let s_i be the i-th invariant factor of H. If $h_{i,i} = s_i$ for $i = 1, ..., r-1$, then the column echelon form of H is the Smith normal form of H.

Proof. We show that a unit upper triangular matrix, V, exists such that HV is a matrix which is zero everywhere except in the first $r-1$ diagonal positions and on and to the right of the diagonal in the r-th row, namely

$$HV = \begin{bmatrix} s_1 & & & & & \\ & \cdot & & & & \\ & & \cdot & & & \\ & & & s_{r-1} & & \\ & & & & h_{r,r} & \cdots & h_{r,n} \\ & & & & & & \end{bmatrix}.$$

We proceed by induction. Since $s_1 = h_{1,1}$ is the GCD of all entries of H, the entries off diagonal in the first row are zeroed by subtracting multiples of the first row, a unit upper triangular operation. We proceed by induction for $1 < r < i$. Assume that the $(i-1)$ rows have been put in column echelon form by upper triangular elementary column operations, and consider the entry $h_{i,j}$, $j > i$. Then since $s_1 \cdots s_i = h_{1,1} \cdots h_{i,i}$ divides all $i{\times}i$ minors (Theorem 2.1), it divides the minor on columns $(1, ..., i-1, j)$ and rows $(1, ..., i)$, which is just the product $h_{1,1} \cdots h_{(i-1),(i-1)} h_{i,j}$. Thus $h_{i,i}$ divides $h_{i,j}$, and hence a upper triangular column operation suffices to zero $h_{i,j}$. Noting that the hypotheses on H imply that $h_{i,j} = 0$ whenever $i > r$, we conclude that HV has the desired form.

Now when HV is brought into column echelon form, it is easy to see that it will be diagonal. The r-th (and last nonzero) diagonal element will be $\mathrm{GCD}_{j=r,...,n} h_{r,j}$. Then since $s_1 \cdots s_r$ is the GCD of all $r{\times}r$ minors, or what is the same, $\mathrm{GCD}_{j=r,...,n}(s_1 \cdots s_{r-1} h_{r,j})$, we must obtain s_r as the r-th diagonal element. □

3.5 Substitution Lemma. Let $f_1, ..., f_t$ be polynomials in $F[\bar{\rho}, x]$, $\bar{\rho}$ a list of new variables, with $\deg(f_i) \le e$. Then for some $\bar{e} \le 2e$, there exists an $\bar{e}{\times}\bar{e}$ determinant Δ in $F[\bar{\rho}]$, whose entries are coefficients of f_i, such that for any evaluation $\bar{\rho} \longrightarrow \bar{r}$, where $\bar{r}$ a list of corresponding field elements that are not a root of Δ, $\mathrm{GCD}_{i=1,...,t}(f_i(\bar{r})) = (\mathrm{GCD}_{i=1,...,t}(f_i))(\bar{r})$. (Cf [13],. Proof of Lemma 4.1.) □

3.6 Irreducibility Lemma. Let $n \ge 2$, and let

$$R = \begin{bmatrix} 1 & & & \\ \rho_{2,1} & 1 & & \\ \vdots & & & \\ \rho_{n,1} & \cdots & \rho_{n,n-1} & 1 \end{bmatrix} \in F[\bar{\rho}]^{n \times n},$$

where $\bar{\rho} = (\rho_{j,k})_{j>k}$ is a vector of indeterminants and F is a field. Then for all i, $1 \le i \le n$, $\mathbf{G} \subset C_i^n$, $\mathbf{G} \ne \emptyset$, and for all families of polynomials $f_J(x) \in F[x] \setminus \{0\}$, $J \in \mathbf{G}$,

$$\sum_{J\in\mathbf{G}} f_J R_{J,I} = \operatorname*{GCD}_{J\in\mathbf{G}}(f_J)\; p,$$

where $I = \{1,\ldots,i\} \in C_i^n$ and where $p \in F[x,\bar\rho]$ is either an irreducible polynomial in $F[\bar\rho, x] \setminus F[x]$ or is 1.

Proof: By induction on i. For $i = 1$, $\sum_{J\in\mathbf{G}} f_J R_{J,I}$ is a linear form in some of the indeterminants $\bar\rho$ over $F[x]$ plus possibly an element in $F[x]$, and the statement is immediate. Now let $i \ge 2$, and let $\mathbf{G}$ be fixed. By computing $R_{J,I}$ by minor expansion along the i-th column we get

$$R_{J,I} = \sum_{j\in J,\, j\ge i} \pm R_{J\setminus\{j\},I'}\,\rho_{j,i}, \quad \text{where } I' = I\setminus\{i\} \text{ and } \rho_{i,i} \text{ is 1 (not a variable).}$$

Define for $1 \le j \le n$

$$\mathbf{G}_j = \{J' \in C_{i-1}^n \mid j \notin J',\ J'\cup\{j\} \in \mathbf{G}\}.$$

Then

$$\sum_{J\in\mathbf{G}} f_J R_{J,I} = \sum_{J'\in\mathbf{G}_i} \pm f_{J'\cup\{i\}} R_{J',I'} + \sum_{j=i+1}^{n} \left(\sum_{J'\in\mathbf{G}_j} \pm f_{J'\cup\{j\}} R_{J',I'} \right) \rho_{j,i}.$$

For all $\mathbf{G}_j \ne \varnothing$ the induction hypothesis applies to the inner sums, that is

$$\sum_{J'\in\mathbf{G}_j} \pm f_{J'\cup\{j\}} R_{J',I'} = d_j\, p_j, \quad j \ge i,$$

where $d_j = \mathrm{GCD}_{J'\in\mathbf{G}_j}(f_{J'\cup\{j\}})$ and p_j is irreducible in $F[\bar\rho, x] \setminus F[x]$, or is 1. Since $\sum_{J\in\mathbf{G}} f_J R_{J,I}$ is now an (inhomogeneous) linear form in $\rho_{j,i}$ over a subring of $F[\bar\rho, x]$ not depending on the $\rho_{j,i}$, by Gauss's lemma

$$\sum_{J\in\mathbf{G}} f_J R_{J,I} = \operatorname*{GCD}_{\substack{i\le j\le n\\ \mathbf{G}_j\ne\varnothing}}(d_j)\ \operatorname*{GCD}_{\substack{i\le j\le n\\ \mathbf{G}_j\ne\varnothing}}(p_j)\ p,$$

where p is an irreducible polynomial in $F[\bar\rho, x] \setminus F[x]$, or is 1. It remains to show that $\mathrm{GCD}_{i\le j\le n,\, \mathbf{G}_j\ne\varnothing}(p_j) = 1$.

Assume that a p_j explicitly depends on $\rho_{l,k}$, $k < i$, $l \ne j$. This $\rho_{l,k}$ occurs in the expansion of some $R_{J',I'}$. Consider p_l, which contains $R_{J'\cup\{j\}\setminus\{l\},I'}$. That determinant is not 0, so p_l is not either. Also p_l cannot depend on $\rho_{l,k}$, so $\mathrm{GCD}(p_j, p_l) = 1$. On the other hand, if all p_j are 1, the claim is trivial. □

Note that in this lemma it is crucial that the selected columns are the ones in I. Otherwise, the lemma is not true, and therefore the proof of lemma 3.7 must enforce the additional condition that H be triangular.

3.7 Lemma. Let A be a matrix in $F[x]^{m\times n}$ of rank r and with the degrees of the entries bounded by d, and let $i \in \{1, \ldots, r-1\}$. Then there is a polynomial π_i in $n(n-1)/2$ variables such that if

(1) R in $F[x]^{n\times n}$ is unit lower triangular,

(2) H is the row echelon form of AR,

(3) s_i^* is the i-th determinantal divisor of A,

then H is upper triangular and $s_i^* = \prod_{j=1}^{i} h_{j,j}$, unless the $n(n-1)/2$ entries below the diagonal in R form a root of π_i. The degree of π_i is no more than $2i^2 d + i$.

Proof: We first show that if R has indeterminate entries (as in Lemma 3.6) then the statement is true unconditionally over $F(\bar{\rho})[x]$. The polynomial π_i is then chosen such that the computation with a specialization of the $\rho_{j,i}$ leads to the same decisions, in particular the same GCDs.

First it easy to show that for indeterminate entries in R the first r columns of AR are linearly independent. Thus H computed over $F(\bar{\rho})[x]$ is triangular. Let $I = \{1, ..., i\}$ and let $A' = AR$. The following sequence of equalities hold, each of which will be established below. Note that all GCD's are taken in the domain of polynomials in x over the field $F(\bar{\rho})$.

$$H_{I,I} = \underset{L\in C_i^m}{\mathrm{GCD}}(H_{L,I}) = \underset{L\in C_i^m}{\mathrm{GCD}}(A'_{L,I}) \tag{A}$$

$$= \underset{L}{\mathrm{GCD}}\left(\sum_{J\in C_i^n} A_{L,J} R_{J,I} \right) \tag{B}$$

$$= \underset{L}{\mathrm{GCD}}(\underset{J}{\mathrm{GCD}}(A_{L,J})\; p_L) \tag{C}$$

$$= \underset{L,J}{\mathrm{GCD}}(A_{L,J}) \underset{L}{\mathrm{GCD}}(p_L) \tag{D}$$

$$= \underset{L,J}{\mathrm{GCD}}(A_{L,J}) = s_i^*. \tag{E}$$

(A) Since H and A' are row equivalent, Theorem 2.1, (2) applies here. (A)=(B) This is the Cauchy-Binet formula for a product of matrices:

$$(XY)_{L,I} = \sum_{J\in C_i^n} X_{L,J} Y_{J,I} \quad \text{for } X\in F^{m\times n},\ Y\in F^{n\times k},\ L\in C_i^m,\ I\in C_i^k. \tag{F}$$

(B)=(C) For each L, lemma 3.6 is applied to the sum, yielding a multiplier which we denote p_L. Note that p_L is irreducible in $F[\bar{\rho}, x] \setminus F[x]$, or is 1. (C)=(D) For L_1 and L_2 in the range of L we have

$$\mathrm{GCD}(\underset{J}{\mathrm{GCD}}(A_{L_1,J}),\ p_{L_2}) = 1,$$

again computed in $F(\bar{\rho})[x]$. This is because p_{L_2}, if it is not 1, is not an element of $F[x]$, whereas all $A_{L_1,J}$ are, and hence their GCD is as well. Therefore, the GCD of the products is the product of the GCDs of the mutually relative prime factors. (D)=(E) We claim that $\mathrm{GCD}_L(p_L) = 1$, again computed over $F(\bar{\rho})[x]$. First, we observe that since the p_L are irreducible over $F[\bar{\rho}, x]$, their GCD over $F(\bar{\rho})[x]$ is either 1 or the polynomials are all multiples by a scalar in F of one another. Now suppose to the contrary that the latter is the case. In other words,

$$\frac{\sum_J A_{L,J} R_{J,I}}{\sum_J A_{M,J} R_{J,I}} \in F, \quad L, M \in C_i^m,$$

which by the monomial structure in $\bar{\rho}$ of the two sums leads to the existence of a multiplier $g_{L,M}$ $\in F$ such that

$$A_{L,J} = g_{L,M} A_{M,J} \text{ for all } J.$$

Now let $L_0 \in C_r^m$, $J_0 \in C_r^n$, such that $A_{L_0, J_0} \neq 0$, that is L_0 and J_0 select a square non-singular matrix $\bar{A}$ of maximal rank from A. Then

$$\det\left(\Big[A_{\tilde{L}, \tilde{J}} \Big]_{\substack{\tilde{L} \subset L_0,\, \tilde{J} \subset J_0 \\ \mathrm{card}(\tilde{L}) = \mathrm{card}(\tilde{J}) = i}} \right) = 0, \tag{G}$$

provided there are at least two rows in this determinant, which are linearly dependent by the above. This is true for $r > i$. However, the matrix in (G) cannot be singular, since it is formed from the non-singular matrix $\bar{A}$ by computing all its i by i minors. To justify this we employ the Cauchy-Binet formula (F) to obtain the following identity:

$$\Big[\bar{A}_{\tilde{L}, \tilde{J}} \Big]_{\tilde{L} \in C_i^r,\, \tilde{J} \in C_i^r} \times \Big[(\bar{A}^{-1})_{\tilde{L}, \tilde{J}} \Big]_{\tilde{L} \in C_i^r,\, \tilde{J} \in C_i^r} = I_{\binom{r}{i}}.$$

Therefore, the $\binom{r}{i}$ by $\binom{r}{i}$ matrix in (G) is invertible, a contradiction to its determinant being 0.

The polynomial π_i is now derived first from lemma 3.5 such that the relationship

$$\underset{L}{\mathrm{GCD}}(A'_{L,I}) = \underset{L,J}{\mathrm{GCD}}(A_{L,J})$$

is preserved by evaluation, and second that the first r columns of A' remain linearly independent. □

Incidentally, we have resolved a question on the coefficient size of multipliers.

3.8 Corollary. For polynomial matrices over the rational numbers, there exist unimodular pre- and post multipliers for the Smith normal form, whose entries have coefficients of binary length polynomial in the dimensions and coefficient lengths of the input matrices. □

4. Rational Canonical Forms and Parallel Similarity Testing

We first introduce the rational canonical form of a square matrix. The *companion matrix* $C_{f(x)}$ of

$$f(x) = x^d + c_{d-1} x^{d-1} + \cdots + c_0 \in F[x].$$

is defined as

$$C_{f(x)} = \begin{bmatrix} 0 & \cdots & & & & -c_0 \\ 1 & 0 & \cdots & & & -c_1 \\ 0 & 1 & 0 & \cdots & & -c_2 \\ \vdots & & & & & \vdots \\ 0 & \cdots & & 1 & 0 & -c_{d-2} \\ 0 & \cdots & & 0 & 1 & -c_{d-1} \end{bmatrix} \in F^{d\times d}.$$

A matrix C is in *rational canonical form* if C is block-diagonal with companion matrices on its diagonal blocks,

$$C = diag(C_{f_1(x)}, \ldots, C_{f_m(x)})$$

and $f_i(x)$ divides $f_{i+1}(x)$ for all $1 \le i \le m-1$. We have the following lemma, cf [5],. Chapter VI, or [8], Chapter S1.

4.1 Lemma. Let $A, B \in F^{n\times n}$.

(1) A is similar to B if and only if $xI - A$ and $xI - B$ are equivalent, they must have the same Smith normal forms.

(2) Let $diag(s_1(x), \ldots, s_n(x))$ be the Smith normal form of $xI - A$. Then $C_A = diag(C_{s_1(x)}, \ldots, C_{s_n(x)})$ is a rational canonical form similar to A.

(3) C_A is the only matrix in rational canonical form that is similar to A. In particular, A is similar to B if and only if $C_A = C_B$.

The non-constant invariant polynomials of $xI - A$, $s_{n-m+1}(x), \ldots, s_n(x)$, $m \le n$, are called the *invariant factors* of A. The above lemma implies that two matrices are similar if and only if the have the same set of invariant factors.

We construct the rational canonical forms C_A and C_B via the parallel algorithm for Smith normal forms. A is not similar to B if $C_A \ne C_B$. We have established the following theorem.

4.2 Theorem. Similarity and non-similarity of matrices in $F^{n\times n}$ is for $F = \mathbf{Q}$ and $F = \mathbf{F}_q$ in (properly Las-Vegas) $\mathbf{RNC}^2$.

If A is proven similar to B, it is sometimes desired to obtain a transforming matrix T such that $B = T^{-1}AT$. Rather than trying to solve the n^2 by n^2 system $AT = TB$ in T, which with sparse methods [22], still requires $O(n^5)$ field operations, our Smith form algorithm provides a better approach. For we also obtain the multipliers, namely

$$U_A(x)(xI - A)V_A(x) = U_B(x)(xI - B)V_B(x).$$

Then $T = V_B(B)V_A^{-1}(B)$ [5], Chapter VI, §5, where $V_B(x)$ and $V_A^{-1}(x)$ are interpreted as polynomials in x with matrix coefficients to the left of x. Notice that $V_A^{-1}(B) = V_A(B)^{-1}$,[@] which

[@] George Labahn (4 August 1994) points to an error in this argument. Instead, $T = P(B)$ where $P(x) = V_B(x)V_A(x)^{-1}$, as Gantmacher states. The sequential running time of $O(n^4)$ state at the end of the parareduces the computation via a matrix multiplication and inverse to evaluating V_A and V_B at B. It can be shown that the Smith normal form algorithm in this case produces multipliers of degree $O(n)$, so from the multipliers we can obtain T sequentially in $O(n^4)$ field operations, or in parallel in $O(\log(n)^2)$ time.

Finally, we wish to mention a corollary to our theorem that answers the sequential complexity of similarity and is a consequence of the above algorithm and the deterministic polynomial-time construction of Smith normal forms over $\mathbf{Q}[x]$ [13], Theorem 4.1.

4.3 Corollary. The problem of similarity of matrices in $\mathbf{Q}^{n\times n}$ is in sequential polynomial-time.

5. Parallel Jordan Normal Form Computation

We now consider the parallel construction of the Jordan normal form of a matrix $A \in F^{n\times n}$. That form is a block-diagonal matrix similar to A whose diagonal blocks are one-sided band matrices of the form

$$\begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & & \vdots \\ & & & & \\ \vdots & & & & 1 \\ 0 & 0 & 0 & \cdots & \lambda_i \end{bmatrix},$$

where λ_i is an eigenvalue of A. The Jordan normal form is unique up to permutation of the diagonal blocks. Different blocks may have the same eigenvalue and/or the same size. In fact, each n_i by n_i block corresponds to an *elementary divisor* $(x - \lambda_i)^{n_i}$ of A. The elementary divisors are simply the maximal powers of linear factors of the invariant factors of A. We refer, e.g., to [5], Chapter VI, §6, or [8], Chapter S1, for proofs of these facts. The only complication in formulating an algorithm for finding the Jordan normal form is that λ_i can lie in an algebraic extension of F and there is no unique way to represent λ_i. If we assume that F already contains the eigenvalues of A and that the distinct eigenvalues are also given as input, then we can in parallel find the elementary divisors by polynomial division from the invariant factors of A. Notice that the invariant factors are already known to be correct via the verification of C_A. We have the following lemma.

5.1 Lemma. Given $A \in F^{n\times n}$, F a field, and given the $k \le n$ distinct eigenvalues $\lambda_i \in F$ of A, $1 \le i \le k$, then the problem of computing the Jordan normal form J of A is in (properly Las-Vegas) **RNC** for F being an algebraic extension of the prime fields $\mathbf{Q}$ and $\mathbf{F}_p$.

The above lemma has the obvious weakness that the splitting field of the characteristic equation of A is required for the construction of J. The structure of J, that is the degrees of the elementary divisors, can be found by squarefree decomposition and GCD operations on the invariant factors. Let us make this process more formal. A *squarefree relatively prime basis* $\{h_1$

graph cannot be obtained in this way (E.K., November 17, 1997).

$,\ldots, h_l\} \subset F[x]$ for a set of polynomials $\{g_1 ,\ldots, g_m\} \subset F[x]$ satisfies

(1) h_i is squarefree for $1 \le i \le l$.

(2) $\mathrm{GCD}(h_i, h_j) = 1$ for $1 \le i < j \le l$.

(3) For all $1 \le j \le m$ there exist integers $e_{i,j} \ge 0$, $1 \le i \le l$, such that $g_j = \prod_{i=1}^{l} h_i^{e_{i,j}}$.

These bases are, of course, not unique since the refinement of a given basis by factoring some of its elements always preserves the required properties. However, the unique coarsest such basis, the *standard* basis, can be found by squarefree decomposition and iterated GCD operations as we describe below, see also [11], §3 for a sequential algorithm. We remark that over fields F of positive characteristic p the squarefree decomposition process is not purely rational. We shall assume that our fields are perfect and p-th roots can be taken. This is, of course, true for $F = \mathbf{F}_q$, the parallel cost of p-th roots being $O(\log q/p)$ arithmetic operation in $\mathbf{F}_q$. (If for $q = p^t$ we choose $\mathbf{F}_q = \mathbf{F}_p[y]/(w[y])$ with $w[y]$ irreducible in $\mathbf{F}_p[y]$ of degree t, then one can even compute p-th roots in $\log^2(t) + \log(p)$ parallel depth on a circuit over $\mathbf{F}_p$ [4].) Under these assumptions, both GCD and squarefree decomposition of polynomials is in $\mathbf{NC}^2$ [6].

We now develop the algorithm. First we show that squarefree relatively prime bases may be "merged" rapidly in parallel. Let the bases be $\{p_i\}$ and $\{q_j\}$. The entries of the merged basis are $r_{i,j} = \mathrm{GCD}(p_i, q_j)$, $p_i^* = p_i/\prod_j r_{i,j}$, and $q_j^* = q_j/\prod_i r_{i,j}$. Unit elements may be discarded. Since the given basis elements are relatively prime and squarefree, it is clear that the new polynomials are also. The $r_{i,j}$ may be computed simultaneously in $O(\log^2(n))$ time. Then the p_i^* and q_j^* are calculated, doing the multiplications in $O(\log(n))$ parallel steps, again using total parallel time $O(\log^2(n))$. Hence we have:

5.2 Lemma. The squarefree relatively prime bases for two sets of polynomials, G_1 and G_2, can be used to construct a squarefree relatively prime basis for $G_1 \cup G_2$ in $\mathbf{NC}^2$.

Now we may use this "merging" to construct the standard squarefree relatively prime basis for a given set of polynomials $\{g_1 ,\ldots, g_n\}$. First compute the squarefree factorization of each g_i. The squarefree factors are relatively prime, so this is also a squarefree relatively prime basis for $\{g_i\}$. Now the n bases may be merged in pairs to form $n/2$ bases for pairs, $\{g_i, g_{i+1}\}$. Iterating this process $\log(n)$ times yields the desired basis.

5.3 Theorem. To compute the standard squarefree relatively prime basis of polynomials $\{g_1 ,\ldots, g_n\}$ is in $\mathbf{NC}^3$.

This answers a question posed by von zur Gathen [7], Remark 6.8. A similar solution was discovered independently in [1], Section 2.1.

We need such a basis for the invariant factors $s_1 ,\ldots, s_m \in F[x]$ of A, which satisfy the additional condition that s_i divides s_{i+1}, $i < m$, so that any factor of s_i occurs to at least the same exponent in s_{i+1}. Because of this the basis construction can be streamlined somewhat. The merging can be done so that at each step bases are constructed for sets of invariant factors with adjacent indices from smaller sets with the same property. Then the divisibility property enables one to eliminate some of the computations. Specifically, if $\{p_i\}$ is the square free relatively prime basis for $\{s_{k_1},\ldots, s_{k_2}\}$ and $\{q_i\}$ is the basis for $\{s_{k_2+1},\ldots, s_{k_3}\}$, then (using the above notation) p_i^* need not be computed and $r_{i,j} = \mathrm{GCD}(p_i, q_j)$ need not be computed when the minimal exponent of p_i in $\{s_{k_1},\ldots, s_{k_2}\}$ is greater than the maximal exponent of $\{q_j\}$ in $\{s_{k_2+1},\ldots, s_{k_3}\}$. Those GCDs are necessarily units.

Let $\{h_i\}$ be the squarefree relatively prime basis constructed from the invariant factors of A. The h_i are defining polynomials for eigenvalues $\lambda_{i,\kappa}$ whose multiplicities in all invariant factors are the same. The multiplicity of $\lambda_{i,\kappa}$ in s_j is that of h_i in s_j and can be easily kept track of during the merge process. Thus we can give the Jordan form as follows.

5.4 Corollary. Given $A \in F^{n\times n}$, $F = \mathbf{Q}$ or $\mathbf{F}_q$, we can compute within $\mathbf{NC}^3$ from the invariant factors of A squarefree pairwise relatively prime polynomials h_i, $\deg(h_i) = k_i$, $1 \le i \le l$, and the symbolic Jordan normal form J of A, in which $k = k_1 + \cdot\ \cdot\ \cdot + k_m$ distinct symbols $\lambda_{i,\kappa}$, $1 \le \kappa \le k_i$, take the place of the k distinct eigenvalues of A, with the understanding that $h_i(\lambda_{i,\kappa}) = 0$.

The symbolic Jordan normal form as described in the above theorem appears the best we can hope to obtain by rational operations. We would like to add that any squarefree relatively prime basis $\{h_1,\ldots, h_l\}$ gives rise to a rational form similar to A,

$$diag(C_{h_1(x)^{e_{1,1}}},\ldots, C_{h_l(x)^{e_{l,m}}}),$$

where $e_{i,j}$ is the multiplicity of h_i in s_j. If the h_i are the irreducible factors of s_m then the canonical form is known in the literature as the *primary* rational canonical form. Our standard basis gives rise to a canonical form between the rational and primary rational one. It is the finest of such forms that is obtainable by purely rational operations. Each block $C_{h_i(x)^{e_{i,j}}}$ can be replaced by an $e_{i,j}$ by $e_{i,j}$ matrix of blocks in “block-Jordan” form

$$\begin{bmatrix} C_{h_i(x)} & I & 0 & \cdots & 0 \\ 0 & C_{h_i(x)} & I & & \vdots \\ & & & & \\ \vdots & & & & I \\ 0 & 0 & 0 & \cdots & C_{h_i(x)} \end{bmatrix}.$$

Of course, if the h_i are chosen the linear factors of s_m, then we get the Jordan canonical form that way. All this follows from the fact that all these block matrices have the same invariant factors.

6. Conclusion

Similarity of matrices and the rational and Jordan canonical forms play an important roly in the study of linear operators on finite dimensional vector spaces. We have provided parallel algorithms for this theory by applying our parallel solution for the somehow lesser-known Smith normal form problem. Our algorithms are also of interest as sequential new methods to solve problems in this theory.

Acknowledgement: The authors like to express their gratitude to the anonymous referee for correcting several errors in the original submission.

References

1. Ben-Or, M., Kozen, D., and Reif, J., “The complexity of elementary algebra and geometry,” *J. Comp. Sys. Sci.*, 32, 2, pp. 251-264 (1986).

2. Chou, T. J. and Collins, G. E., “Algorithms for the solution of systems of diophantine linear equations,” *SIAM J. Comp.*, 11, pp. 687-708 (1982).

3. Cook, S. A., “A taxonomy of problems with fast parallel algorithms,” *Inf. Control,* 64, pp. 2-22 (1985).

4. Fich, F. E. and Tompa, M., “The parallel complexity of exponentiating polynomials over finite fields,” *Proc. 17th Annual ACM Symp. Theory Comp.*, pp. 38-47 (1985).

5. Gantmacher, F. R., *The Theory of Matrices, Vol. 1,* Chelsea Publ. Co., New York, N. Y. (1960).

6. Gathen, J. von zur, “Parallel algorithms for algebraic problems,” *SIAM J. Comp.*, 13, pp. 802-824 (1984).

7. Gathen, J. von zur, “Representations and parallel computations for rational functions,” *SIAM J. Comp.*, 15, pp. 432-452 (1986).

8. Gohberg, I., Lancaster, P., and Rodman, L., *Matrix Polynomials,* Academic Press, New York, NY (1982).

9. Hoffman, K. and Kunze, R., *Linear Algebra,* Prentice Hall, Englewood Cliffs, N.J. (1961).

10. Iliopoulos, C. S., “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,” Manuscript, Purdue Univ. (1986).

11. Kaltofen, E., “Sparse Hensel lifting,” *Proc. EUROCAL ’85, Vol. 2, Springer Lec. Notes Comp. Sci.*, 204, pp. 4-17 (1985).

12. Kaltofen, E., Krishnamoorthy, M. S., and Saunders, B. D., “Fast parallel algorithms for similarity of matrices,” *Proc. 1986 ACM Symp. Symbolic Algebraic Comp.*, pp. 65-70 (1986).

13. Kaltofen, E., Krishnamoorthy, M. S., and Saunders, B. D., “Fast parallel computation of Hermite and Smith forms of polynomial matrices,” *SIAM J. Alg. Discrete Meth.,* 8, pp. 683-690 (1987).

14. Kannan, R., “Polynomial-time algorithms for solving systems of linear equations over polynomials,” *Theoretical Comp. Sci.,* 39, pp. 69-88 (1985).

15. Kannan, R. and Bachem, A., “Polynomial algorithms for computing the Smith and Hermite normal forms of an integer matrix,” *SIAM J. Comp.,* 8, pp. 499-507 (1981).

16. MacDuffee, C. C., *Vectors and Matrices*, Math. Assoc. America (1943).

17. MacDuffee, C. C., *The Theory of Matrices,* Chelsea Publ. Co., New York, N. Y. (1956).

18. Mulmuley, K., “A fast parallel algorithm to compute the rank of a matrix over an arbitrary field,” *Combinatorica*, 7, pp. 101-104 (1987).

19. Newman, M., *Integral Matrices*, Pure and Applied Mathematics, 45, Academic Press, New York (1972).

20. Schwartz, J. T., “Fast probabilistic algorithms for verification of polynomial identities,” *J. ACM*, 27, pp. 701-717 (1980).

21. Sims, C. C., *Abstract Algebra, A Computational Approach,* Wiley, New York (1984).

22. Wiedemann, D., “Solving sparse linear equations over finite fields,” *IEEE Trans. Inf. Theory*, IT-32, pp. 54-62 (1986).