Information Processing Letters 18 (1984) 147–150
North-Holland

30 March 1984

ON COMPUTING THE DETERMINANT IN SMALL PARALLEL TIME USING A SMALL NUMBER OF PROCESSORS

Stuart J. BERKOWITZ *

Department of Computer Science, University of Toronto, Toronto, Canada M5S17V

Communicated by K. Mehlhorn
Received 20 August 1983

The determinant, characteristic polynomial and adjoint over an arbitrary commutative ring with unity can be computed by a circuit with size $O(n^{3.496})$ and depth $O(\log^2 n)$. Furthermore, the circuits can be constructed uniformly (by a log space bounded Turing machine).

Keywords: Parallel algebraic circuit complexity, upper bounds, algebraic determinant

1. Introduction

While it has been known since 1976 [5] that small size and depth circuits exist for computing the adjoint, the determinant, and the characteristic polynomials over fields that contain the integers, it has been an open question until recently whether a similar result holds for arbitrary commutative rings with unity. Borodin, Von zur Gathen and Hopcroft [3] showed that such a result was possible using the parallelization technique in [9] and by extending the technique of Strassen fro eliminating division, but the method is not explicit. The size was $O(n^{15})$.

The problem of reducing the size depended on finding an algorithm that computed these polynomials without first using divisions as those methods do. Csanky's method seems to require a division by n! and therefore does not apply to finite fields. In this paper we present a method, based on Samuelson's method [8,6], which uses no divisions. We then efficiently parallelize it to get our result.

One of the extra advantages of our method is that it relates the adjoint and the determinant in a very efficient way. We also present what appears to be the shortest and simplest proof of Samuelson's method yet.

2. The model and definitions

For arbitrary rings T, we will describe a circuit which computes polynomial in the ring T, and uses only constants from T. Except for the fast matrix multiplication [4] subcircuit, the only constants that we use are $\{-1, 0, 1\}$.

Our model of computation is a straight line program model. Let X be a set of indeterminates. T[X] is a polynomial ring. A program over T[X] is a list of instructions $\{v_i\}_{i=1}^{C}$ such that $v_i = v_i' \circ v_i''$ and $\circ \in \{\oplus, \otimes\}$ and $v_i', v_i'' \in \{v_1, \ldots, v_C\} \cup X \cup T$. Size and depth are defined in the obvious manner.

We will state some algebraic definition that we require.

Let A be an $n \times n$ matrix, with R, S, M as $1 \times (n-1)$, $(n-1) \times 1$ and $(n-1) \times (n-1)$ submatrices. The relationship is given as follows:

$$A = \begin{pmatrix} a_{1,1} & R \\ S & M \end{pmatrix}.$$

* Present address: Array Systems Computing, Downsview, Ontario M3H 5T5, Canada.

Let

$R_i = (a_{i,i+1}, \ldots, a_{i,n})$,

$S_i^t = (a_{i+1,i}, \ldots, a_{n,i})$,

$$M_i = \begin{bmatrix} a_{i,i} & \cdots & a_{i,n} \\ \vdots & \ddots & \vdots \\ a_{n,i} & \cdots & a_{n,n} \end{bmatrix}.$$

Let $p(\lambda)$, $q(\lambda)$ be characteristic polynomials of A and M:

$$p(\lambda) = \sum_{i=0}^{n} p_{n-i}\lambda^i = \det(A - \lambda * I),$$

$$q(\lambda) = \sum_{i=0}^{n-1} q_{n-1-i}\lambda^i = \det(M - \lambda * I).$$

The determinant can be defined inductively by the method of cofactor expansion on rows or columns. We give an expansion by column j and one by row i:

$$\det(A) = \bigoplus_{i=1}^{n} A_{i,j} \otimes \det(A(i \mid j))(-1)^{i+j}$$

$$= \bigoplus_{j=1}^{n} A_{i,j} \otimes \det(A(i \mid j))(-1)^{i+j},$$

where $A(i \mid j)$ is the $(n-1)\times(n-1)$ matrix obtained from A by deleting the *ith* row and the *jth* column.

The classical adjoint is defined by

$$\operatorname{adj}(A)_{i,j} = \det(A(j \mid i))(-1)^{i+j}.$$

It follows that if $\det(A) \neq 0$, then adj(A) is the unique matrix which satisfies

$$\operatorname{adj}(A) * A = A * \operatorname{adj}(A) = I * \det(A).$$

We state four claims which are used in the main theorem, but are independently interesting.

3. Main results

Claim 1

$$p(\lambda) = (a_{1,1} - \lambda) * \det(M - \lambda * I) + R * \operatorname{adj}(M - \lambda * I) * S.$$

Proof. We expand the $\det(A - \lambda * I)$ by cofactor expansion on the first row, and then on the first column. □

Claim 2

$$\operatorname{adj}(M - \lambda * I) = = -\sum_{k=2}^{n} (M^{k-2} * q_0 + \cdots + I * q_{k-2}) * \lambda^{n-k}.$$

Proof (Beame [2]). Multiply both sides by $M - \lambda * I$. The left-hand side is equal to $q(\lambda) * I$ by definition. The right-hand side is equal to the left by the Caley–Hamilton theorem and by reordering of indices.

$$-\sum_{k=2}^{n} (\cdots) * \lambda^{n-k} * (M - \lambda * I) =$$

$$= -\sum_{k=2}^{n} (M^{k-1} * q_0 + \cdots + M * q_{k-2}) * \lambda^{n-k} + \sum_{k=1}^{n-1} (M^{k-1} * q_0 + \cdots + I * q_{k-1}) * \lambda^{n-k}$$

by reordering indices of the second term.

$$= -\sum_{k=2}^{n} (M^{k-1} * q_0 + \cdots + M * q_{k-2}) * \lambda^{n-k} + \sum_{k=1}^{n} (M^{k-1} * q_0 + \cdots + I * q_{k-1}) * \lambda^{n-k}$$

by subtracting the Caley–Hamilton theorem equation $(M^{n-1} * q_0 + \cdots + I * q_{n-1} = 0)$ from the second term.

$$= +q(\lambda) * I.$$

Since $q(\lambda) \neq 0$, the claim is proved. □

Substituting Claim 2 into Claim 1 we have Samuelson's method of relating the characteristic polynomials of A and M:

$$p(\lambda) = (a_{1,1}) * \det(M - \lambda * I) - R * \left(\sum_{k=2}^{n} (M^{k-2} * q_0 + \cdots + I * q_{k-2}) * \lambda^{n-k} \right) * S. \qquad (\star)$$

Let SD(f(n), g(n)) notation stand for a circuit size of O(f(n)) and depth O(g(n)). For n, α, ϵ, $\beta > 0$ we would like to define n^{α}, n^{ϵ}, n^{β}, $n^{\beta-\epsilon}$ as integers. Let this notation stand for the smallest integer greater than or equal to its real number value.

Claim 3. *Let* R, S, M *be* $1 \times m$, $m \times m$ *and* $m \times 1$ *matrices for* $m < n$. *Let* $T = \{R * M^i * S\}_{i=0}^{m}$. *Then* T *can be computed in* $SD(n^{\alpha+\epsilon}, \log^2 n)$, *where* α *is the exponent of* n *for the size of a circuit for multiplying two matrices using* log n *depth (currently* $\alpha < 2.496$ [4]) *and* ϵ *is any positive real number greater than* 0.

Proof (using an idea suggested by S. Winograd). Define

$$U = \{R * M^i\}_{i=0}^{n^{0.5}} \quad \text{and} \quad V = \{M^{i*n^{0.5}} * S\}_{i=0}^{n^{0.5}}.$$

We can compute any element of T (uniquely) as a dot product of vectors from U and V. This is because the exponent of k of the M term in any element from T can be (uniquely) expressed in the form $k = i + j * n^{0.5}$, where $i, j < n^{0.5}$. Thus all of T can be computed from U and V using n dot products of size $n - 1$. Since each dot product can be computed in SD(n, log n), T can be computed from U and V in $SD(n^2, \log n)$.

U and V are computed similarly so we will only present the computation for U. We claim by induction on β that $Z_\beta = \{R * M^i\}_{i=0}^{n^\beta}$ can be computed in $SD(n^{\alpha+\epsilon}, \log^2 n)$ for β a constant. It is trivially true for $n^\beta - \epsilon$ equal to zero. If the hypothesis is true for $n^{\beta-\epsilon}$, then it is true for n^β as follows:

Z_β can be computed from $Z_{\beta-\epsilon}$ and $Y_\beta = \{M^{j*n^\beta}\}_{j=0}^{n^\epsilon}$. We do this by multiplying the matrix $X_{\beta-\epsilon}$, whose rows are the vectors of $Z_{\beta-\epsilon}$, by the n^ϵ matrices of Y_β. This can be done in $SD(n^{\alpha+\epsilon}, \log^2 n)$. Since Y_β can be computed in $SD(n^{\alpha+\epsilon}, \log^2 n)$ by using the parallel prefix algorithm [7], our induction hypothesis is proved. Since $U = Z_{0.5}$, our claim is proved. □

We define an $n \times m$ Toeplitz matrix as an $n \times m$ matrix such that each diagonal has a value to which all the elements in the diagonal are equal. More precisely,

$$B_{i,j} = B_{i-\min(i,j)+1,\, j-\min(i,j)+1}.$$

A lower triangular matrix has all elements above the central diagonal equal to zero:

$$B_{i,j} = 0 \quad \text{if } i < j.$$

Claim 4. *Let* A, B *be* $n \times m$ *and* $m \times p$ *lower triangular Toeplitz matrices. Then* $C = A * B$ *is lower triangular and Toeplitz and can be computed in* $SD(n^2, \log n)$.

Proof. For the proof, see [1]. □

Actually, in rings with all the n*th* roots of unity where we can do the Fast Fourier Transform, we can multiply these Toeplitz matrices using the FFT in O(n log n) size and O(log n) depth circuits. But for the general case we need $O(n^2)$ size and the same depth.

Theorem 5. *For all* $\epsilon > 0$, *the coefficients of the characteristic polynomial can be computed in*

$$SD(n^{\alpha+1+\epsilon}, \log^2 n).$$

Proof. Let C_t be $(n - t + 1) \times (n - t)$ matrices that are lower triangular and Toeplitz defined as

$$C_{t\,1,i} = \begin{cases} -1 & \text{if } i = 1, \\ \alpha_{t,t} & \text{if } i = 2, \\ -R_t * M_t^{i-3} * S_t & \text{if } i > 2. \end{cases}$$

By (★) we have a linear relation between the coefficient of $p(\lambda)$ and the coefficients of $q(\lambda)$ which can be easily expressed as

$$(p_0, p_1, \ldots, p_n)^t = C_1 * (q_0 \cdots q_{n-1})^t,$$

where $(\cdots)^t$ is a transposed matrix. Applying this recursively, we get

$$= \bigotimes_{i=1}^{n} C_i.$$

The entries to the matrices $\{C_i\}$ can be computed using n applications of Claim 3 in $SD(n^{\alpha+1+\epsilon}, \log^2 n)$. The characteristic coefficients

can be computed from $\{C_i\}$ with a balanced binary tree of matrix multiplies in $SD(n^3, \log^2 n)$ using Claim 4. □

The adjoint of A can easily be computed from the characteristic polynomial of A and $\{A^i\}$ in SD(n, log n) by Claim 2. $\{A^i\}$ can be computed from A in $SD(n^{\alpha+1}, \log^2 n)$ by the parallel prefix algorithm [7]. Thus the adjoint computation presented has the same complexity as the characteristic polynomial computation. The determinant can be computed from the characteristic polynomial by letting λ be 0.

At present $\alpha < 2.496$. Since ϵ can be chosen so that $\alpha + \epsilon < 2.496$, the claim in the abstract is proved.

4. Conclusions

By reexamining Csanky's paper [5] with the ideas of Claim 3 in mind, one can get a circuit in $SD(n^{\alpha+0.5} \log n, \log^2 n)$ [10]. Can our technique be improved to get this size exponent? The problem with this is twofold. The product of the $\{C_i\}$ can be done in $SD(n^2 \log n, \log^2 n)$ if we can use FFTs. But no way is known that is better than the way presented fro arbitrary rings. The other problem is that computing $\{M_i^j\}$ for all i and j seems just as hard as computing it n times for one i and all j. This seems that it could be improved because the M_i's are submatrices of each other.

Does a more direct way exist for avoiding some of these problems? Can Samuelson's method be modified to relate the characteristic polynomial of one matrix efficiently to the characteristic polynomial of a related matrix which is about one half the size? It is conjectured that there is a circuit for computing the characteristic polynomial over arbitrary rings in $SD(n^\alpha, \log^2 n)$.

References

[1] A.V. Aho, J. Hopcroft and J.D. Ullman, The Design and Analysis of Computer Algorithms (Addison-Wesley, Reading, MA, 1974).
[2] P. Beame, Private communication, 1982.
[3] A. Borodin, J. von zur Gathen and J. Hopcroft, Fast parallel matrix and GCD computations, Preprint, 1981.
[4] D. Coopersmith and S. Winograd, On the asymptotic complexity of matrix multiplication, Proc. 22nd Symp. on Foundations of Computer Science (1981) pp. 82–90.
[5] L. Csanky, Fast parallel inversion algorithms, SIAM J. Comput. 5 (4) (1976) 818–823.
[6] D. Faddeev and V. Faddeev, Computational Methods of Linear Algebra (Freeman, San Francisco, 1963) pp. 247–251.
[7] R. Ladner and M. Fischer, Parallel prefix computations, J. ACM 27 (1980) 831–838.
[8] P.A. Samuelson, A method for determining explicitly the coefficients of the characteristic equation, Ann. Math. Statist. 13 (1942) 424–429.
[9] L. Valiant, S. Skyum, S. Berkowitz and C. Rackoff, Fast parallel computation of polynomials using few processors, SIAM J. Comput. (1982) submitted for publication.
[10] S. Winograd, Private communication, 1982.