STAN-CS-71-251

SEL-71-063

AN EFFICIENT PARALLEL ALGORITHM FOR THE SOLUTION OF A TRIDIAGONAL LINEAR SYSTEM OF EQUATIONS

by

Harold S. Stone*

NASA Ames Research Center, and

DIGITAL SYSTEMS LABORATORY
Departments of Electrical Engineering and Computer Science
Stanford University
Stanford, California

December 1971

Technical Report no. 19

* ASEE fellow, Summer 1971

This work was supported by the NASA Ames Research Center, by the Joint Services Electronics Program under contract N-00014-67-A-0112-0044, and by the National Science Foundation under grant GJ-1180.

An efficient parallel algorithm for the solution of a tridiagonal linear system of equations

by

Harold S. Stone

NASA Ames Research Center* and

Digital Systems Laboratory

Departments of Electrical Engineering and Computer Science

Stanford University

Abstract

Tridiagonal linear systems of equations can be solved on conventional serial machines in a time proportional to N, where N is the number of equations. The conventional algorithms do not lend themselves directly to parallel computation on computers of the ILLIAC IV class, in the sense that they appear to be inherently serial. An efficient parallel algorithm is presented in which computation time grows as $\log_2 N$. The algorithm is based on recursive doubling solutions of linear recurrence relations, and can be used to solve recurrence relations of all orders.

* ASEE fellow, summer, 1971

1 Introduction

The trend in large-scale high-speed computers today clearly points to the use of internal parallelism to obtain significant increases in speed. For example, the ILLIAC IV computer can perform N simultaneous computations where N = 64, 128, 256, or 512. We expect that highly efficient computations performed on a computer of the ILLIAC IV class will execute N times faster than on a serial computer of the same inherent speed. Actually, inefficiencies due to overhead and constraints on data communication among processors will reduce the speed increase to kN where k lies in the interval $0 \leq k \leq 1$. Efficient algorithms have k near unity.

Unfortunately, many parallel algorithms do not lend themselves to efficient parallel computation. We can exhibit examples of algorithms for which computation time decreases rather slowly as we increase the number of processors, and for some pathological examples the computation time is independent of the number of processors. An efficient parallel algorithm has the property that computation time decreases proportionally to 1/N as N, the parallelism factor, increases.

In this paper we examine the solution of tridiagonal systems of linear equations. It is well known that such systems can be solved using a conventional serial computer in a time proportional to N where N is the number of equations. We present an algorithm for solving the equations in a time proportional to $\log_2 N$ by using a computer with N-fold parallelism. Computation in this case grows as $(\log_2 N)/N$ which is proportional to $N^{1-\epsilon}$ for any $\epsilon > 0$ when N is sufficiently large. A different parallel algorithm for this problem which exhibits a similar time behavior has been developed by Buneman [1967], and analyzed in the literature [Buzbee, et al., 1970].

In Section II, we state the problem and indicate conventional serial methods for solution. These methods are inherently serial in that each computation depends on the result of the immediately preceding computation. In Section III we show how to perform a forward and backward sweep in $\log_2 N$ steps when given the LU decomposition of the original matrix. In Section IV we show how to obtain the LU decomposition in $\log_2 N$ steps. This particular computation is of general interest because it is an efficient method for evaluating partial fraction expansions and linear difference equations in parallel.

II Statement of the problem

We wish to solve the tridiagonal system of equations

$$\underset{\sim}{A}\ \underset{\sim}{x} = \underset{\sim}{b}$$

where

$$\underset{\sim}{A} = \begin{bmatrix} d_1 & f_1 & & & \\ e_2 & d_2 & f_2 & & \\ & e_3 & d_3 & f_3 & \\ & & \cdots & & \\ & & e_{N-1} & d_{N-1} & f_{N-1} \\ & & & e_N & d_N \end{bmatrix}$$

In the remainder of this paper we assume that N is a power of 2, but this is not an essential assumption.

There are a number of related methods for solving this system serially in a time proportional to N. The parallel algorithm presented here is based upon one such algorithm, the LU decomposition. [cf. Forsythe and Moler, 1967] In this algorithm we find two matrices, $\underset{\sim}{L}$, and $\underset{\sim}{U}$, such that

(i) $\underset{\sim}{L}\underset{\sim}{U} = \underset{\sim}{A}$

(ii) $\underset{\sim}{L}$ is a lower bidiagonal matrix with 1's on its principal diagonal.

(iii) $\underset{\sim}{U}$ is an upper bidiagonal matrix.

When $\underset{\sim}{A}$ is non-singular, its LU decomposition is unique. In fact, it is easily shown that

$$\underset{\sim}{U} = \begin{bmatrix} u_1 & f_1 & & & \\ & u_2 & f_2 & & \\ & & u_3 & f_3 & \\ & & & \cdots & \\ & & & u_{N-1} & f_{N-1} \\ & & & & u_N \end{bmatrix}$$

where f_i, $1 \le i \le N-1$, is the upper diagonal of A, and

$$\begin{aligned} u_1 &= d_1 \\ u_i &= d_i - \frac{e_i\, f_{i-1}}{u_{i-1}}, \text{ for } i > 1. \end{aligned} \tag{1}$$

The lower bidiagonal matrix, $\underset{\sim}{L}$, is then given by

$$\underset{\sim}{L} = \begin{bmatrix} 1 & & & & \\ m_2 & 1 & & & \\ & m_3 & 1 & & \\ & & \cdots & & \\ & & & m_{N-1} & 1 \\ & & & & m_N & 1 \end{bmatrix}$$

where

$$\begin{aligned} m_2 &= e_2/d_1 \\ m_i &= \frac{e_i}{d_{i-1} - f_{i-2} m_{i-1}}, \text{ for } i > 2 \\ &= \frac{e_i}{u_{i-1}}, \text{ for } i \ge 2 \end{aligned} \tag{2}$$

After computing L and U, it is relatively straight forward to solve the system of equations. The solution is a two-step process.

Letting $\underset{\sim}{y} = \underset{\sim}{U}\underset{\sim}{x}$, we have

$$\underset{\sim}{A}\,\underset{\sim}{x} = \underset{\sim}{L}\,\underset{\sim}{U}\,\underset{\sim}{x} = \underset{\sim}{L}\,\underset{\sim}{y} = \underset{\sim}{b}$$

The equation $\underset{\sim}{L}\,\underset{\sim}{y} = \underset{\sim}{b}$ is easily solved for $\underset{\sim}{y}$ since

$$\begin{aligned} y_1 &= b_1 \\ y_i &= b_i - m_i y_{i-1} \quad \text{for } 2 \le i \le N \end{aligned} \tag{3}$$

Then we solve $\underset{\sim}{U}\,\underset{\sim}{x} = \underset{\sim}{y}$ for $\underset{\sim}{x}$. This equation is solved by a backward sweep since

$$\begin{aligned} x_N &= y_N / u_N \\ x_i &= \frac{y_i - x_{i+1} f_i}{u_i} \end{aligned} \tag{4}$$

Note that the recurrence formulae (1), (2), (3) and (4) constitute a complete algorithm for the solution of $\underset{\sim}{A}\,\underset{\sim}{x} = \underset{\sim}{b}$. Since each computation in this algorithm depends on the results of the previous computation, the algorithm is satisfactory for serial computation but quite unsatisfactory for parallel computation. In the following sections we derive equivalent formulae that are well-suited for parallel computation.

III Parallel evaluation of the forward and backward sweeps

The model of a parallel processor that lies behind the development of these parallel algorithms is based upon the ILLIAC IV computer. In this computer there are N processors with independent memories, but only one instruction stream. All of the processors operate synchronously, executing the same instruction on N different operand pairs, where N can be 64, 128, 256, or 512. For added flexibility, there is a mask associated with each processor that enables or disables the processor. Hence, if a processor's mask is on, the processor executes the current instruction, otherwise the processor remains idle.

Data can be communicated among the processors in one of two ways. One datum can be broadcast to all processors simultaneously, or a vector of N items can be shifted cyclically among the processors. As an example of the latter case, suppose that the vector $\underset{\sim}{b} = (b_1, b_2, b_3, \ldots b_N)$ is stored with b_i in the i^{th} processor. Then the vector can be shifted j places cyclically so that b_i is routed to processor (i+j) mod N for all i.

In this section, we shall show how to solve (3) by a technique called recursive doubling. The idea is to rewrite (3) so that y_{2i} is a function of y_i. Thus, in successive iterations we can compute y_1, y_2, y_4, y_8, etc., and y_N can be computed in $\log_2 N$ iterations. Since (4) is of the same form as (3), the backward sweep can be done using the same algorithm, and it also requires $\log_2 N$ iterations.

To begin the derivation, we rewrite (3) in the form

$$\begin{aligned} y_1 &= b_1 \\ y_i &= b_i + (-m_i) y_{i-1} \end{aligned} \tag{3'}$$

This change is necessary because we shall make use of the associativity of addition.

Substituting for y_{i-1} in (3') we find

$$\begin{aligned} y_2 &= b_2 + (-m_2) \cdot b_1 \\ y_3 &= b_3 + (-m_3) \cdot b_2 + (-m_3) \cdot (-m_2) \cdot b_1 \\ y_i &= \sum_{j=1}^{i} b_j \prod_{k=j+1}^{i} (-m_k) \end{aligned} \tag{5}$$

where a vacuous product of m_k's is interpreted as the constant 1.

The last formula in (5) shows the explicit dependence of y_i on each of the coefficients of $\underset{\sim}{m}$ and $\underset{\sim}{b}$. Our goal is to derive a recurrence in which y_{2i} is a function of y_i. To anticipate the answer, momentarily consider

what happens when all of the components of $\underset{\sim}{m}$ are equal to -1. In this case y_i is the sum of the first i components of $\underset{\sim}{b}$. Then if $y_i(b_j, b_{j-1}, \ldots, b_{j-i+1})$ is defined to be the sum of b_j through b_{j-i+1}, we have

$$y_{2i}(b_{2i}, b_{2i-1}, \ldots, b_1) = y_i(b_{2i}, b_{2i-1}, \ldots, b_{i+1}) + y_i(b_i, b_{i-1}, \ldots, b_1) \tag{6}$$

Equation (6) holds for all $i \geq 1$. This recurrence has the recursive doubling form that we seek, and therefore is the basis for a parallel algorithm. The recursive doubling relation above suggests that we look for a general solution in terms of functions $Y_1, Y_2, \ldots, Y_N$ where each Y_i is a function of i components of $\underset{\sim}{b}$ and $\underset{\sim}{m}$. We shall use the notation $Y_i(j)$ as an abbreviation for the more cumbersome notation

$$Y_i(b_j, b_{j-1}, \ldots, b_{j-i+1}, m_j, m_{j-1}, \ldots, m_{j-i+1})$$

That is, $Y_i(j)$ is a function of i consecutive components of $\underset{\sim}{b}$ and $\underset{\sim}{m}$, with the j^{th} component being the highest component.

The following theorem establishes the relation we desire.

Theorem 1: Let $Y_i(j)$ satisfy the recurrence relation

$$Y_{i+1}(j) = Y_1(j) + Y_i(j-1) \cdot (-m_j) \qquad \text{for } i, j \geq 1 \tag{7}$$

with the boundary conditions

$$Y_1(j) = b_j \quad \text{for } j \geq 1$$

$$Y_i(j) = 0 \quad \text{for } j \leq 0$$

$$Y_i(j) = 0 \quad \text{for } i \leq 0$$

Then

(i) for $s \geq 2$, $Y_i(j)$ satisfies the recurrence relation

$$Y_{i+s}(j) = Y_s(j) + Y_i(j-s) \prod_{k=j-s+1}^{i} (-m_k) \quad \text{for } i \geq 1, j \geq s. \tag{8}$$

(ii) $$Y_i(j) = \sum_{k=1}^{j} Y_1(k) \prod_{s=k+1}^{j} (-m_s) \quad \text{for } i \geq j \geq 1 \tag{9}$$

(iii) for $i \geq j \geq 1$, $Y_i(j) = y_j$, where y_j is the j^{th} component of the unique solution of (3).

Proof:

To prove part (i), we use induction on s.

Basis step, s = 2.

From (7) we have

$$Y_{i+2}(j) = Y_1(j) + Y_{i+1}(j-1)\cdot(-m_j)$$
$$= Y_1(j) + Y_1(j-1)\cdot(-m_j) + Y_i(j-2)\cdot(-m_j)\cdot(-m_{j-1})$$

But using (7) again we also have

$$Y_2(j) = Y_1(j) + Y_1(j-1)\cdot(-m_j)$$

Hence,

$$Y_{i+2}(j) = Y_2(j) + Y_i(j-2)\cdot(-m_j)\cdot(-m_{j-1})$$

which is recurrence relation (8) with s = 2. This proves the basis step.

Induction step. We assume that (8) hold for all s in the interval $2 \leq s \leq n-1$, and we show it holds for s = n.

From the induction hypothesis we have

$$Y_{i+n}(j) = Y_{n-1}(j) + Y_{i+1}(j-r+1)\cdot \prod_{k=j-n+2}^{j} (-m_k)$$

$$= Y_{n-1}(j) + Y_1(j-n+1)\cdot \prod_{k=j-n+2}^{j} (-m_k)$$

$$+ Y_i(j-n)\cdot \prod_{k=j-n+1}^{j} (-m_k)$$

But from the induction hypothesis it follows that

$$Y_n(j) = Y_{n-1}(j) + Y_1(j-n+1)\cdot \prod_{k=j-n+2}^{j} (-m_k)$$

Hence,

$$Y_{i+n}(j) = Y_n(j) + Y_i(j-n)\cdot \prod_{k=j-n+1}^{j} (-m_k)$$

which is the same recurrence as (8) with s replaced by n. This proves part (i).

To prove part (ii), we use induction on i.

Basis step. From the theorem hypothesis we have

$$Y_2(j) = Y_1(j) + Y_1(j-1)\cdot(-m_j), \quad \text{for } j \geq 1$$

Then applying the boundary condition $Y_1(0) = 0$, we obtain

$$Y_2(1) = Y_1(1)$$

$$Y_2(2) = Y_1(2) + Y_1(1)\cdot(-m_2)$$

These equations satisfy (9), thus proving the basis step.

Induction step: We assume that (9) holds for all i in the interval $2 \leq i \leq n-1$, and we prove that it holds for $i = n$. Using (8) we have

$$Y_n(j) = Y_1(j) + Y_{n-1}(j-1)\cdot(-m_j)$$

Using the induction hypothesis to substitute for $Y_{n-1}(j-1)$ yields

$$Y_n(j) = Y_1(j) + \left[\sum_{k=1}^{j-1} Y_1(k) \prod_{s=k+1}^{j-1} (-m_s)\right]\cdot(-m_j) \quad \text{for } 2 \leq j \leq n$$

$$= \sum_{k=1}^{j} Y_1(k) \prod_{s=k+1}^{j} (-m_s) \quad \text{for } 2 \leq j \leq n \qquad (10)$$

The interval $2 \leq j \leq n$ for which the equations above are valid arises from the application of the induction hypothesis to $Y_{n-1}(j-1)$ for $1 \leq j-1 \leq n-1$. Since (10) has the same form as (9), it is only necessary to show the validity of (10) for $j = 1$ to complete the proof. From the theorem hypothesis,

$$Y_n(1) = Y_1(1) + Y_{n-1}(0) = Y_1(1)$$

Since the same result is obtained by setting $j = 1$ in (10), the interval in (10) may be changed to $1 \leq j \leq n$. This proves part (ii) of the theorem.

Part (iii) is a direct consequence of the fact that with the boundary condition $Y_1(j) = b_j$, (10) is identical to the solution to (3). This completes the proof of the theorem.

Corollary:

$$Y_{2i}(j) = Y_i(j) + Y_i(j-i)\cdot \prod_{k=j-i+1}^{j} (-m_k) \qquad \text{for } i, j \geq 1 \tag{11}$$

Proof: Follows directly from part (i) of Theorem 1 by replacing s by i.

The corollary of Theorem 1 provides the recursive doubling algorithm for the solution of (3). The product term in (11) appears to be difficult to evaluate because the number of factors in the product doubles with each iteration. Fortunately, we can also use recursive doubling to compute the product term.

Let $M_i(j)$ be defined to be

$$\begin{aligned} M_i(j) &= \prod_{k=j-i+1}^{j} (-m_k) \qquad \text{for } j \geq i \\ &= \prod_{k=1}^{j} (-m_k) \qquad \text{for } j < i \end{aligned} \tag{12}$$

Then (11) can be rewritten as

$$Y_{2i}(j) = Y_i(j) + Y_i(j-i)\cdot M_i(j) \qquad \text{for } i,j \geq 1 \tag{13}$$

The recursive doubling computation of $M_i(j)$ is provided by the formula

$$M_{2i}(j) = M_i(j)\cdot M_i(j-i) \qquad \text{for } i,j \geq 1 \tag{14}$$

with the boundary conditions

$$M_1(j) = -m_j \qquad \text{for } j \geq 1$$

$$M_i(j) = 1 \qquad \text{for } j \leq 0$$

$$M_i(j) = 1 \qquad \text{for } i \leq 0$$

The parallel algorithm for the solution of (3) is simply the iterative application of (13) and (14). It is given below in an ALGOL-like language. In the program, when an interval of the form $(1 \le j \le N)$ appears after a statement, that statement is assumed to be executed simultaneously for all indices in the interval.

```
begin
real array  Y[1:N], M[2:N];
real array  b[1:N], m[2:N];
comment Y and M are the arrays in which equations (13) and (14)
        are evaluated.  Arrays b and m are the arrays that give the
        coefficients of (3).  These arrays may utilize the same
        storage space as the arrays Y and M, respectively;
initialize:
    Y[j] : = b[j], (1 ≤ j ≤ N);
    M[j] : = -m[j], (1 ≤ j ≤ N);
for i : = 1 step i until N/2 do
    begin
    Y[j] : = Y[j] + Y[j-i] × M[j], (i+1 ≤ j ≤ N);
    M[j] : = M[j] × M[j-i], (i+1 ≤ j ≤ N);
    end;
```

At the completion of each iteration, the array Y contains $Y_i(j)$, and M contains $M_i(j)$. From Theorem 1, $Y_N(j) = y_j$ for $1 \le j \le N$, so that Y_N is the solution to (3). Since i doubles during each iteration, $\log_2 N$ iterations are required for the computation. The vector operations indicated in the program are easily carried out in an ILLIAC IV type of computer since masking operations can be used to establish the interval

for the index j, and cyclic shifting of components of a vector can be used to align Y[j] with Y[j-i]. The parallel algorithm is also suitable for efficient operation in vector processors of the pipeline class such as the CDC STAR computer.

For the solution of the backward sweep, Equation (4), the body of the iteration should be modified as indicated below:

```
begin
Y[j]: = Y[j] + Y[j+i] × M[j],    (1 ≤ j ≤ N - i);
M[j]: = M[j] × M[j+i],           (1 ≤ j ≤ N - i);
end;
```

IV Calculation of the LU decomposition by recursive doubling

We now focus attention on the efficient calculation of (1) and (2). Again we use recursive doubling to compute the coefficients $u = (u_1,u_2,\ldots,u_N)$ and $m = (m_2,m_3,\ldots,m_N)$. The approach we use is to solve (1) by recursive doubling, then compute $m_i = e_i/u_{i-1}$ simultaneously for $2 \le i \le N$ to solve (2).

Since (1) is a partial fraction expansion, it is convenient to cast it into a linear form which is suitable for a recursive doubling algorithm. It is well known [cf. Wall, 1948] that every partial fraction expansion is associated with a linear second order recurrence relation. In particular, if we define the quantities q_i, $0 \le i \le N$, by the recurrence relation

$$q_i = d_i q_{i-1} - e_i f_{i-1} q_{i-2} \qquad i \ge 2 \tag{15}$$

with the boundary conditions

$$q_0 = 1$$

$$q_1 = d_1$$

then it is easily shown that

$$u_i = q_i/q_{i-1} \qquad \text{for } i \ge 1 \tag{16}$$

or equivalently,

$$q_i = \prod_{j=1}^{i} u_j .$$

To solve (1) efficiently, we have only to solve (15) efficiently, because after calculating q_i, $0 \le i \le N$, we can evaluate (16) in a single operation carried out simultaneously on N processors. Equation (15) is somewhat more difficult to solve than (3) because it is of second order, whereas (3) is of first order. However, we can make use of an artifice to reformulate (15) as a matrix recurrence relation of first order. In particular, it follows from (15) that

$$\begin{bmatrix} q_i \\ q_{i-1} \end{bmatrix} = \begin{bmatrix} d_i & -e_i f_{i-1} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} q_{i-1} \\ q_{i-2} \end{bmatrix} = \underset{\sim}{A}_i \begin{bmatrix} q_{i-1} \\ q_{i-2} \end{bmatrix} \tag{16}$$

Note that we can substitute $\underset{\sim}{A}_{i-1} (q_{i-2}\ q_{i-3})^T$ for $(q_{i-1}\ q_{i-2})^T$ in (16), and can continue this substitution repeatedly until we obtain

$$\begin{bmatrix} q_i \\ q_{i-1} \end{bmatrix} = \underset{\sim}{A}_i \underset{\sim}{A}_{i-1} \cdots \underset{\sim}{A}_2 \begin{bmatrix} q_1 \\ q_0 \end{bmatrix} \tag{17}$$

This formulation of the problem is ideal for recursive doubling. Since matrix multiplication is associative, we can evaluate the product $\underset{\sim}{A}_i \underset{\sim}{A}_{i-1} \cdots \underset{\sim}{A}_2$ in exactly the same way that we evaluate a product of scalars. In fact, we have encountered this problem before in (12), and the recursive doubling solution is the schema of (14). Then to solve (15) for all q_i simultaneously, requires $\log_2 N$ iterations, in which the i^{th} iteration involves the 2^{N-i} simultaneous calculations of the product of two 2×2 matrices.

It is rather interesting to investigate the properties of the functions q_i because it is possible to exploit their characteristics and obtain a parallel algorithm slightly more efficient than the solution to (17) described above. Fortunately, a great deal is known about these functions. One important property is well illustrated by the first few q_i.

$$q_0 = 1$$

$$q_1 = d_1$$

$$q_2 = d_2d_1 - e_2f_1$$

$$q_3 = d_3d_2d_1 - d_3e_2f_1 - e_3f_2d_1$$

$$q_4 = d_4d_3d_2d_1 - d_3d_3e_2f_1 - d_4e_3f_2d_1 - e_4f_3d_2d_1 + e_4f_3e_2f_1$$

Knuth [1971] attributes to Euler [1748] the observation that q_i contains the term $d_id_{i-1}\ldots d_1$, together with every term that can be constructed by replacing d_jd_{j-1} by $-e_jf_{j-1}$ for all possible combinations of such pairs. This property follows directly from the recurrence relation (15). The first product in (15), d_iq_{i-1}, creates terms in q_i for which adjacent d-pairs are deleted from among only the coefficients d_1, d_2, ..., d_{i-1} in all possible ways, and thus produces every possible way there can be terms containing d_i. The second product in (15) replaces d_id_{i-1} by $-e_if_{i-1}$, and combines this with every possible way d-pairs can be eliminated among the coefficients d_1, d_2, ..., d_{i-2}. This produces every possible term without d_i.

We can obtain factorizations of the q_i functions that correspond to the intermediate results in the evaluation of (17). To arrive at these factorizations, let us define $Q_i(j)$ for $j \geq i$ to be the function q_i with the subscripts of its arguments increased systematically so that the leading subscript is j. For $j < i$, we define $Q_i(j) = Q_j(j)$. Some examples of $Q_i(j)$

should clarify ambiguities in the definition.

$$Q_1(1) = d_1$$

$$Q_2(1) = d_2$$

$$Q_3(3) = d_3d_2d_1 - d_3e_2f_1 - e_3f_2d_1$$

$$Q_3(4) = d_4d_3d_2 - d_4e_3f_2 - e_4f_3d_2$$

$$Q_3(2) = Q_2(2) = d_2d_1 - e_2f_1$$

From this definition it now follows directly that the Q_i functions satisfy the recurrence

$$Q_{i+s}(j) = Q_s(j)Q_i(j-s) - e_{j-s+1}f_{j-s}Q_{s-1}(j)Q_{i-1}(j-s-1) \qquad (18)$$

$$\text{for } j \geq s,\ i \geq 1$$

with the boundary conditions

$$Q_1(j) = d_j \qquad \text{for } j \geq 1$$

$$Q_i(j) = 1 \qquad \text{for } j \geq 0,\ i \leq 0$$

$$Q_i(j) = 1 \qquad \text{for } j \leq 0,\ i \geq 0$$

$$e_{j+1}f_j = 0 \qquad \text{for } j \leq 0$$

This recurrence formulation is also well-known, with citations in the literature at least as early as 1853. [Sylvester, 1853; Perron, 1913].

The validity of (18) can be verified by an intuitive argument. To find all possible ways of eliminating adjacent d-pairs in a sequence of i+s coefficients, combine every possible way of eliminating pairs in the first s coefficients with every possible way of eliminating pairs in the last i coefficients. This accounts for the first term of (18). However, one d pair contains the last coefficient from the set of s coefficients and the first coefficient from the set of i coefficients. The first term in (18) does not account for any of the ways this pair can be eliminated. We see that the second term in (18) accounts for all such ways, because $e_{j-s+1}f_{j-s}$ replaces the pair and this replacement is combined with every

possible way of eliminating pairs in the first s-1 coefficients and in the last i-1 coefficients. From (18) we obtain the recursive doubling formulae.

Theorem 2: $Q_i(j)$ satisfies the recurrence relations

$$Q_{2i}(j) = Q_i(j)Q_i(j-i) + (-e_{j-i+1}f_{j-i})Q_{i-1}(j)Q_{i-1}(j-i-1)$$
$$Q_{2i-1}(j) = Q_i(j)Q_{i-1}(j-i) + (-e_{j-i+1}f_{j-i})Q_{i-1}(j)Q_{i-2}(j-i-1) \qquad (19)$$
$$Q_{2i-2}(j) = Q_{i-1}(j)Q_{i-1}(j-i+1) + (-e_{j-i+2}j_{j-i+1})Q_{i-2}(j)Q_{i-2}(j-i)$$

Proof: These formulae follow directly from (18).

The first of the equations in the corollary above is a recursive doubling formula which shows that Q_{2i} depends on both Q_i and Q_{i-1}. Hence, to compute Q_{4i} we need to compute both Q_{2i} and Q_{2i-1}. To compute Q_{4i-1} we have to compute Q_{2i-2}. Since Q_{2i-2} depends on the same quantities as Q_{2i} and Q_{2i-1}, we need only the three equations (19) in a recursive doubling algorithm. Since we have to compute Q_{2i-1} and Q_{2i-2} anyway, it is slightly more efficient to compute Q_{2i} by the formula

$$Q_{2i}(j) = b_j Q_{2i-1}(j) + (-e_j f_{j-1})Q_{2i-2}(j-2).$$

The complete algorithm to compute q_i, $1 \le i \le N$ is given below in an ALGOL-like language. The initial conditions establish the values of Q_0, Q_1, and Q_2. The first iteration computes Q_2, Q_3, and Q_4, the second iteration computes Q_6, Q_7, and Q_8, and the last iteration computes Q_{N-2}, Q_{N-1}, and Q_N.

```
begin
    real array E[2:N], F[1:N-1], D[1:N], EF[1:N],
        TEMP[1:N], QI[1:N], QIM1[0:N], QIM2[-1:N];
    comment the arrays hold the quantities indicated below.
        E       The lower diagonal of the tridiagonal matrix A.
        F       The upper diagonal of A.
        D       The major diagonal of A.
        EF      This holds products of the form -e_i f_{i-1}.
```

TEMP A temporary array.

QI Holds $Q_i(j)$.

QIM1 Holds $Q_{i-1}(j)$.

QIM2 Holds $Q_{i-2}(j)$.

The computation begins by initializing EF, QI, QIM1, and QIM2; initialize:

```
EF[i] := - E[i]×F[i-1], (2 ≤ i ≤ N);
QIM2[i] := 1, (1 ≤ i ≤ N);
QIM1[i] := D[i], (1 ≤ i ≤ N);
QI[i] := D[i]×D[i-1] + EF[i], (2 ≤ i ≤ N);
QI[1] := D[1];
```

comment the last three lines initialize the arrays to Q_0, Q_1, and Q_2, respectively;

```
for i := 2 step i until N/2 do
    begin
      TEMP[j] := QIM1[j]×QIM1[j-i+1] + EF[j-i+2]×QIM2[j]×QIM2[j-1],
                                                  (i-1 ≤ j ≤ N);
      comment TEMP contains Q2i-2. It cannot be written over Qi-2 yet
      since Qi-2 is needed in the next line;
      QIM1[j] := QI[j]×QIM1[j-i] + EF[j-i+1]×QIM1[j]×QIM2[j-i-1],
                                                  (i ≤ j ≤ N);
      QIM2[j] := TEMP[j], (i-1 ≤ j ≤ N);
      QI[j] := D[j]×QIM1[j-1] + EF[j]×QIM2[j-2], (i+1 ≤ j ≤ N);
    end;
```

At the termination of the algorithm, QI[i] will contain q_i for $1 \le i \le N$. We use (16) to compute the diagonal of U from the q_i's. This clearly can be done in parallel by dividing the vector QI by a shift of

itself. Finally, to compute the subdiagonal of L, we note that (2) indicates that this computation can be done by one parallel division.

In executing the algorithm on an ILLIAC IV class of computer, the vector alignment required for the calculation is done by cyclically shifting vectors among the processors. Since the algorithm requires that QI[j] = QIM1[j] = QIM2[j] = 1 for $j \leq 0$, we can avoid storing these quantities by changing the cyclic shift of these vectors to an end-off shift in which the integer 1 is shifted into element 1 of each of these vectors. Similarly, EF[j] = 0 for $j \leq 1$, so that 0's are always shifted into EF[2] when the EF vector is aligned.

The ranges indicated for each statement in the basic iteration show the positions of the vectors which change when that statement is executed. The algorithm will work correctly when all ranges are replaced by the full range $1 \leq i \leq N$ since values that do not change are recomputed at each step. It is somewhat more efficient to use the full range for a calculation than the ranges given, although redundant recomputation of values may be accompanied by greater round-off error.

The serial solution of a tridiagonal system of equations, when done as outlined in Section II, requires $3(N-1)$ of each of the operations division, multiplication, and subtraction. The parallel computation has three loops, each executed $\log_2 N$ times. The loop that computes the LU decomposition requires eight multiplications and three additions per iteration, whereas the forward and back substitutions each require two multiplication and one addition per iteration. Apart from the computations within loops, there are at least four divisions, two multiplications and one addition applied to N elements simultaneously.

Hence the operation count for the parallel algorithm (exclusive of overhead computations) is

$12 \log_2 N + 2$ multiplications

$5 \log_2 N + 1$ additions

4 divisions.

The reduction in the number of divisions is particularly important for computers which take much longer to divide than to multiply. (On the ILLIAC IV computer division is approximately five times longer than multiplication).

At this writing the stability of the algorithm has not been thoroughly investigated. Clearly, the algorithm is unstable if any q_i vanishes. Since $q_i = \prod_{j=1}^{i} u_j$, q_i vanishes if and only if one of the u_i coefficients vanishes. However, if the $\underline{A}$ matrix is diagonally dominant and non-singular, every u_i is bounded away from zero [Isaacson and Keller, 1966].

We conjecture that the error bounds for the parallel algorithm are comparable to those of the serial algorithm.

Summary and conclusions

The parallel algorithm for the solution of tridiagonal systems of linear equations really consists of two different algorithms. One algorithm is the parallel evaluation of first order difference equations of the form

$$x_i = b_i x_{i-1} + c_i$$

where the b_i and c_i are constants.

The second algorithm solves second order equations of the form

$$x_i = b_i x_{i-1} + c_i x_{i-2} \qquad (20)$$

Since partial fraction expansions are associated with second order difference equations, the second algorithm may also be used to compute partial fraction expansions. The form of the solution obviously generalizes to linear recurrence relations of arbitrary m^{th} order, still requiring $\log_2 N$ iterations, where each iteration involves simultaneous multiplications of $m \times m$ matrices.

It is well known that a straightforward serial evaluation of (20) can be unstable [Gautschi, 1967], although it is not unstable when the coefficients are obtained from diagonally dominant matrices. The stability of the parallel algorithm in such cases has not been investigated, but it too is undoubtedly unstable. Since (20) can be solved by backward recursion when forward recursion is unstable, we expect that backward parallel recursion would also be stable.

Acknowledgment

The author expresses his appreciation to William Jones and David Galant of NASA Ames Research Center for their many conversations, comments, and criticisms which materially aided the research. He is also grateful to Donald Knuth of Stanford University for pointing out the early contributions to the factorization of second order recurrence relations. The recursive doubling algorithm for solving first order recurrence relations was discovered independently by Harvard Lomax of NASA Ames Research Center and by Robert Downs of Systems Control, Inc. Gene Golub of Stanford University pointed out Buneman's algorithm as an alternative method for solving tridiagonal systems in a time proportional to $\log_2 N$.

References

Buneman, Oscar, 1969. "A compact non-iterative Poisson solver," Report 294, Stanford University Institute for Plasma Research, Stanford, California, 1969.

Buzbee, B. L., G. H. Golub, and C. W. Nielson, 1970. "On direct methods for solving Poisson's equations," SIAM J. Numer. Anal., Vol.7, No. 4, December 1970.

Euler, Leonhard, 1748. Introductio in Analysin Infinitorum, Lausanne, Section 359, 1748.

Forsythe, G. E. and C. B. Moler, 1967. Computer Solution of Linear Algebraic Systems, Prentice-Hall, Englewood Cliffs, New Jersey, 1967.

Gautschi, Walter, 1967. "Computational aspects of three-term recurrence relations," SIAM Review, Vol. 9, No. 1, pp. 24-82, Jan. 1967.

Isaacson, E., and H. B. Keller, 1966. Analysis of Numerical Methods, John Wiley and Sons, New York, 1966.

Knuth, D. E., 1971. "Mathematical analysis of algorithms," Report Stan-CS-71-206, Stanford Computer Science Department, March 1971.

Perron, O., 1913. Die Lehre von den Kettenbruchen, Leipzig, 1913.

Sylvester, J. J., 1853. Philosophical Magazine, 6, pp. 297-299, 1853.

Wall, H. S., 1948. Analytic Theory of Continued Fractions, Van Nostrand, Princeton, N. J., 1948.