MATHEMATICS OF COMPUTATION
VOLUME 62, NUMBER 206
APRIL 1994, PAGES 601–617

BACKWARD ERROR ANALYSIS OF CYCLIC REDUCTION FOR THE SOLUTION OF TRIDIAGONAL SYSTEMS

PIERLUIGI AMODIO AND FRANCESCA MAZZIA

ABSTRACT. Tridiagonal systems play a fundamental role in matrix computation. In particular, in recent years parallel algorithms for the solution of tridiagonal systems have been developed. Among these, the cyclic reduction algorithm is particularly interesting. Here the stability of the cyclic reduction method is studied under the assumption of diagonal dominance. A backward error analysis is made, yielding a representation of the error matrix for the factorization and for the solution of the linear system. The results are compared with those for LU factorization.

1. INTRODUCTION

Tridiagonal matrices arise in a large variety of applications. It is known that for the solution of linear systems, the LU factorization is the best scalar algorithm. Since it is not efficient for parallel computation, many algorithms which may be easily parallelized have been proposed [11, 13, 15]. Among many others, the cyclic reduction algorithm appears to be the most interesting [1, 2, 3, 4, 9, 10, 14]. As a result, an increasing number of papers have been written in the last ten years which consider cyclic reduction.

So far, no backward error analysis has been made. The main result about cyclic reduction is given by Heller [6], concerning block tridiagonal systems. Heller shows that norms of the off-diagonal blocks (relative to the diagonal blocks) decrease quadratically with each reduction. This is useful when an approximate solution is desired, as in a preconditioner based on cyclic reduction.

A backward error analysis is carried out in the following sections. We use a block representation of the algorithm, which provides a simpler way to analyze the propagation of the error.

In the case of diagonal dominance it is well known that, for the LU factorization, the backward error in the solution obtained in floating-point arithmetic may be bounded from above by a quantity independent of the dimension of the matrix [8, 16]. Therefore, the algorithm is numerically stable. The results of our backward error analysis for the cyclic reduction algorithm show that, for diagonally dominant matrices, the backward error of the computed solution is

Received by the editor October 30, 1991 and, in revised form, September 25, 1992.
1991 *Mathematics Subject Classification.* Primary 65F05, 65G05, 15A23.
Key words and phrases. Tridiagonal linear systems, cyclic reduction, backward error analysis.
Work performed within the activities of the project 'Matematica computazionale' supported by MURST 40% and by "Progetto finalizzato Sistemi Informatici e Calcolo Parallelo. Sottoprogetto: Calcolo Scientifico per Grandi Sistemi" of C.N.R.

0025-5718/94 $1.00 + $.25 per page

bounded by a factor which depends on the logarithm of the dimension of the linear system. This is a satisfactory growth rate for error propagation, but in fact it is a pessimistic upper bound. In numerical tests we show that a more accurate upper bound may be computed which is comparable to that of LU factorization.

In §2 a brief description of the algorithm is presented and a block representation for the factorization of the coefficient matrix is proposed. In §§3–4 the backward error analysis for the factorization and for the solution of the linear system is presented. This analysis takes into account the special structure of the matrices involved in the propagation of the error.

2. Block representation of the cyclic reduction algorithm

Consider the following system of linear equations:

$$M\mathbf{x} = \mathbf{f}, \tag{2.1}$$

where the coefficient matrix M is tridiagonal,

$$M = \begin{pmatrix} a_1 & b_1 & & \\ c_2 & a_2 & \ddots & \\ & \ddots & \ddots & b_{n-1} \\ & & c_n & a_n \end{pmatrix}. \tag{2.2}$$

We derive the cyclic reduction algorithm by considering a block factorization of M. By means of an odd-even permutation matrix P_1 (which transforms the sequence $1, \ldots, n$ into the sequence $1, 3, 5, \ldots, 2, 4, 6, \ldots$) the matrix M is expressed as a 2×2 block matrix

$$P_1 M P_1^T = \begin{pmatrix} A_1 & T_1 \\ S_1 & B_1 \end{pmatrix} \tag{2.3}$$

with A_1 and B_1 diagonal matrices

$$A_1 = \begin{pmatrix} a_1 & & \\ & a_3 & \\ & & \ddots \end{pmatrix}, \qquad B_1 = \begin{pmatrix} a_2 & & \\ & a_4 & \\ & & \ddots \end{pmatrix}$$

and S_1 and T_1 bidiagonal (not necessarily square) matrices

$$S_1 = \begin{pmatrix} c_2 & b_2 & & \\ & c_4 & b_4 & \\ & & \ddots & \ddots \end{pmatrix}, \qquad T_1 = \begin{pmatrix} b_1 & & & \\ c_3 & b_3 & & \\ & \ddots & \ddots & \end{pmatrix}.$$

The permuted matrix (2.3) is factored as

$$P_1 M P_1^T = L_1 D_1 U_1,$$

where L_1 and U_1 are block triangular

$$L_1 = \begin{pmatrix} I & \\ S_1 A_1^{-1} & I \end{pmatrix}, \qquad U_1 = \begin{pmatrix} A_1 & T_1 \\ & I \end{pmatrix}$$

and D_1 is block diagonal

$$D_1 = \begin{pmatrix} I & \\ & M_1 \end{pmatrix}$$

with a tridiagonal block $M_1 = B_1 - S_1 A_1^{-1} T_1$ of dimension $\lfloor n/2 \rfloor \times \lfloor n/2 \rfloor$ (M_1 is the Schur complement of A_1 in $P_1 M P_1^T$).

The same procedure is applied to M_1 (or equivalently to D_1, leaving unchanged the top left block). That is, we consider a $\lfloor n/2 \rfloor \times \lfloor n/2 \rfloor$ odd-even permutation matrix Q_2 and

$$P_2 = \begin{pmatrix} I & \\ & Q_2 \end{pmatrix}$$

and carry out the factorization

$$D_1 = P_2^T \begin{pmatrix} I & & \\ & A_2 & T_2 \\ & S_2 & B_2 \end{pmatrix} P_2 = P_2^T L_2 D_2 U_2 P_2$$

in which D_2 is block diagonal and contains a tridiagonal sub-block M_2 of dimension $\lfloor n/4 \rfloor \times \lfloor n/4 \rfloor$. By iterating this process of reduction, after j steps the matrix

$$D_{j-1} = \begin{pmatrix} I & \\ & M_{j-1} \end{pmatrix} \tag{2.4}$$

(where $D_0 = M$) is factored in the form

$$D_{j-1} = P_j^T L_j D_j U_j P_j ,$$

where

$$L_j = \begin{pmatrix} I & & \\ & I & \\ & S_j A_j^{-1} & I \end{pmatrix}, \qquad U_j = \begin{pmatrix} I & & \\ & A_j & T_j \\ & & I \end{pmatrix},$$

$$D_j = \begin{pmatrix} I & & \\ & I & \\ & & M_j \end{pmatrix}.$$

This new block decomposition is obtained so that the blocks on the main diagonal are square of dimension respectively $n - \lfloor n/2^{j-1} \rfloor$ on the first row, $\lfloor n/2^{j-1} \rfloor - \lfloor n/2^j \rfloor$ on the second, $\lfloor n/2^j \rfloor$ on the third. The reduction process stops after $k = \lfloor \log_2 n \rfloor$ steps when the block M_k is 1×1 and D_k is diagonal.

The following summarizes the factorization of the matrix M :

$$M = P_1^T L_1 P_2^T L_2 \cdots P_k^T L_k D_k U_k P_k \cdots U_2 P_2 U_1 P_1 . \tag{2.5}$$

Consider now a block representation of the cyclic reduction algorithm for the solution of the problem (2.1) that will be useful for the stability analysis.

Algorithm 1. Let

a) $M_0 = M$ be a tridiagonal matrix of order n, $\mathbf{y}_0 = \mathbf{f}$;

b) Q_j, for $j = 1, \ldots, k = \lfloor \log_2 n \rfloor$, be an odd-even permutation matrix of

order j such that

$$\begin{pmatrix} A_j & T_j \\ S_j & B_j \end{pmatrix} = Q_j M_{j-1} Q_j^T \qquad \text{and} \qquad \begin{pmatrix} \mathbf{y}_{j-1}^o \\ \mathbf{y}_{j-1}^e \end{pmatrix} = Q_j \mathbf{y}_{j-1}.$$

The following algorithm computes the solution $\mathbf{x}_0$ of $M\mathbf{x} = \mathbf{f}$ by cyclic reduction:

for $j = 1, k$
$V_j = S_j A_j^{-1}$
$M_j = B_j - V_j T_j$
$\mathbf{y}_j = \mathbf{y}_{j-1}^e - V_j \mathbf{y}_{j-1}^o$
end
$\mathbf{x}_k = M_k^{-1} \mathbf{y}_k$
for $j = k, 1, -1$
$$\mathbf{x}_{j-1} = Q_j^T \begin{pmatrix} A_j^{-1}(\mathbf{y}_{j-1}^o - T_j \mathbf{x}_j) \\ \mathbf{x}_j \end{pmatrix}$$
end

Denote by a_i, b_i, c_i respectively the nonzero elements on the main diagonal, the upper and the lower off-diagonal of M_{j-1}, and by a_i', b_i', c_i' the respective elements of M_j (when it is not confusing, we always simplify the notation and let a_i, b_i, c_i be the generic elements $a_i^{(r)}$, $b_i^{(r)}$, $c_i^{(r)}$ of M_r). Let further s_i and t_i denote the generic elements on the two nonzero diagonals of V_j. Then, if $m = \lfloor n/2^j \rfloor$, the elements of the matrices V_j and M_j in Algorithm 1 are computed as follows:

for $i = 1, m$
$s_i = c_{2i}/a_{2i-1}$
$t_i = b_{2i}/a_{2i+1}$ $\quad (t_m = 0 \text{ if } \lfloor n/2^{j-1} \rfloor \text{ is odd})$
$a_i' = a_{2i} - s_i b_{2i-1} - t_i c_{2i+1}$
$b_i' = -t_i b_{2i+1}$ $\quad (b_m' = 0)$
$c_i' = -s_i c_{2i-1}$ $\quad (c_1' = 0)$
end

while the algorithms for the jth step in the solution of the two linear systems are:

for $i = 1, m$
$y_i^{(j)} = y_{2i}^{(j-1)} - s_i y_{2i-1}^{(j-1)} - t_i y_{2i+1}^{(j-1)}$
end

and

for $i = 1, m$
$x_{2i-1}^{(j-1)} = (y_{2i-1}^{(j-1)} - c_{2i-1} x_{i-1}^{(j)} - b_{2i-1} x_i^{(j)})/a_{2i-1}$
$x_{2i}^{(j-1)} = x_i^{(j)}$
end
if $\lfloor n/2^{j-1} \rfloor$ *is odd* $x_{2m+1}^{(j-1)} = (y_{2m+1}^{(j-1)} - c_{2m+1} x_m^{(j)})/a_{2m+1}$

The factorization of the coefficient matrix M requires $8n$ operations while the solution of the linear systems requires $9n$ operations. The number of operations is twice the number of operations of the LU factorization algorithm, but this algorithm is easier to parallelize.

We introduce now the following convention:

$$\prod_{j=1}^{i} A_j = A_1 A_2 \cdots A_i$$

and we recall an important characterization of the factorization, also contained in [6], that will be used in the next sections.

Theorem 1. *The cyclic reduction factorization for the matrix* M,

$$M = \prod_{i=1}^{k} P_i^T L_i \; D_k \; \prod_{i=1}^{k} U_{k-i+1} P_{k-i+1},$$

may be expressed as the LU *factorization of the permuted matrix*

$$\prod_{i=1}^{k} P_i \; M \; \prod_{i=1}^{k} P_{k-i+1}^T.$$

The matrices M_i (with i varying from 1 to $k = \lfloor \log_2 n \rfloor$) enjoy special properties which are useful for studying the stability of the factorization. The two following theorems concern diagonally dominant matrices (see [17] for the proofs).

Theorem 2. *If* $M_0 = M$ *is diagonally dominant by rows (by columns), then all the matrices* M_i *are diagonally dominant by rows (by columns).*

Moreover, it has been proved that if M is strictly diagonally dominant, then the ratio between each element on the off-diagonal and the corresponding one on the main diagonal of M_i decreases quadratically as i varies from 1 to $\lfloor \log_2 n \rfloor$ [6].

Theorem 3. *If* M *is diagonally dominant by rows or by columns, then for each matrix* M_i *there holds*

$$||M_i||_\infty \leq ||M||_\infty.$$

3. Backward error analysis

Assume that computations are carried out in floating-point arithmetic following the model

$$\text{fl}(x \text{ op } y) = (x \text{ op } y)(1 + u_1) \quad \text{and} \quad \text{fl}(x \text{ op } y) = \frac{x \text{ op } y}{1 + u_2}, \qquad |u_i| \leq u, \text{ for } i = 1, 2,$$

where u is the unit roundoff and $\text{op} \in \{+, -, *, /\}$.

The implementation of the cyclic reduction algorithm on a computer for the solution of the problem (2.1) gives, instead of the exact solution $\mathbf{x}$, an approximate solution $\tilde{\mathbf{x}}$. In the following we obtain an upper bound for the infinity norm of the matrix ΔM such that $\tilde{\mathbf{x}}$ is the exact solution of the perturbed problem

$$(M + \Delta M)\tilde{\mathbf{x}} = \mathbf{f}$$

when M is diagonally dominant.

In the backward error analysis we introduce a matrix δM containing the error due to the factorization. Let

$$\mathbf{L} = \prod_{j=1}^{k} P_j^T \tilde{L}_j \quad \text{and} \quad \mathbf{U} = \tilde{D}_k \prod_{j=1}^{k} \tilde{U}_{k-j+1} P_{k-j+1},$$

where $\tilde{D}_k$, $\tilde{L}_j$, $\tilde{U}_j$ are the computed matrices D_k, L_j, and U_j of (2.5). Then

$$\delta M = \mathbf{LU} - M. \tag{3.1}$$

Moreover, we consider two further error matrices $\delta\mathbf{L}$ and $\delta\mathbf{U}$ arising in the numerical solution of the two triangular systems with matrices $\mathbf{L}$ and $\mathbf{U}$, so that $\tilde{\mathbf{x}}$ is the exact solution of the system

$$(\mathbf{L} + \delta\mathbf{L})(\mathbf{U} + \delta\mathbf{U})\tilde{\mathbf{x}} = \mathbf{f}.$$

In first-order approximation we obtain (for any compatible norm)

$$||\Delta M|| \leq ||\delta M + \delta\mathbf{L}\,\mathbf{U} + \mathbf{L}\,\delta\mathbf{U}||, \tag{3.2}$$

while the relative error of the solution may be bounded by

$$\frac{||\tilde{\mathbf{x}} - \mathbf{x}||}{||\tilde{\mathbf{x}}||} \leq ||M^{-1}||\,||\Delta M||. \tag{3.3}$$

To obtain the error matrices δM, $\delta\mathbf{L}$, and $\delta\mathbf{U}$, we use an approach which permits exploiting the structure of sparsity of the error matrix.

The factorization of the matrix M is obtained by means of the k steps of factorization previously considered. At each step j, a new block-diagonal matrix D_j, see (2.4), (with the last block being tridiagonal of dimension $\lfloor n/2^j \rfloor$) is obtained from the matrix D_{j-1}.

The relations between D_j and D_{j-1} are expressed by means of the following matrix difference equation:

$$\begin{cases} D_0 = M, \\ D_j = L_j^{-1} P_j D_{j-1} P_j^T U_j^{-1}. \end{cases}$$

In floating-point arithmetic, we have

$$\tilde{D}_j = \tilde{L}_j^{-1} P_j \tilde{D}_{j-1} P_j^T \tilde{U}_j^{-1} + \delta D_j, \tag{3.4}$$

where δD_j contains the errors committed in the jth step of the factorization. In order to evaluate the error matrix δD_j, we consider the floating-point operations (see the description of Algorithm 1):

$$(\tilde{V}_j)_{r,i} = \begin{cases} \tilde{c}_{2i}/\tilde{a}_{2i-1}(1+u_1) = \tilde{s}_i & \text{if } r = i, \\ \tilde{b}_{2i}/\tilde{a}_{2i+1}(1+u_2) = \tilde{t}_i & \text{if } r = i-1, \end{cases}$$

and

$$(\tilde{M}_j)_{r,i} = \begin{cases} -\tilde{t}_i \tilde{b}_{2i+1}(1+u_3) = \tilde{b}'_i & \text{if } r = i-1, \\ \tilde{a}_{2i}(1+u_7) - \tilde{s}_i \tilde{b}_{2i-1}(1+u_4+u_6+u_7) & \\ \qquad -\tilde{t}_i \tilde{c}_{2i+1}(1+u_5+u_6+u_7) = \tilde{a}'_i & \text{if } r = i, \\ -\tilde{s}_i \tilde{c}_{2i-1}(1+u_8) = \tilde{c}'_i & \text{if } r = i+1, \end{cases}$$

where $|u_i| \leq u$, or in block form:

$$\tilde{V}_j = \tilde{S}_j \tilde{A}_j^{-1} + \delta V_j, \qquad \tilde{M}_j = \tilde{B}_j - \tilde{V}_j \tilde{T}_j + \delta M_j,$$

where δV_j is lower bidiagonal and δM_j is tridiagonal. There follows

$$|\delta V_j| \leq |\tilde{S}_j \tilde{A}_j^{-1}|\, u, \qquad |\delta M_j| \leq |\tilde{B}_j - \tilde{V}_j \tilde{T}_j| u + 2\, \mathrm{diag}(|\tilde{V}_j|\, |\tilde{T}_j|)\, u,$$

where $|A|$ denotes the matrix of the absolute values of the elements of A, $\mathrm{diag}(A)$ the diagonal matrix containing the main diagonal of A, and inequalities are understood componentwise.

Let

$$\tilde{L}_j = \begin{pmatrix} I & & \\ & I & \\ & \tilde{V}_j & I \end{pmatrix}, \quad \tilde{D}_j = \begin{pmatrix} I & & \\ & I & \\ & & \tilde{M}_j \end{pmatrix}, \quad \tilde{U}_j = \begin{pmatrix} I & & \\ & \tilde{A}_j & \tilde{T}_j \\ & & I \end{pmatrix};$$

then from (3.4) one gets

$$\delta D_j = \begin{pmatrix} 0 & & \\ & 0 & \\ & -\delta V_j & \delta V_j \tilde{T}_j - \delta M_j \end{pmatrix}.$$

Theorem 4. *The error matrix δM defined in* (3.1) *satisfies*

$$\delta M = \sum_{j=1}^{k} \prod_{i=1}^{j} P_i^T H_j^F \prod_{i=1}^{j} P_{j-i+1}, \tag{3.5}$$

where

$$H_j^F = \begin{pmatrix} 0 & & \\ & 0 & \\ & \delta V_j \tilde{A}_j & \delta M_j \end{pmatrix}. \tag{3.6}$$

Proof. To show how the error propagates, we define

$$\begin{aligned} \epsilon_0 &= \tilde{D}_0 - D_0 = \tilde{M}_0 - M = 0, \\ \epsilon_j &= \tilde{D}_j - D_j = \tilde{L}_j^{-1} P_j \tilde{D}_{j-1} P_j^T \tilde{U}_j^{-1} + \delta D_j - L_j^{-1} P_j D_{j-1} P_j^T U_j^{-1}. \end{aligned} \tag{3.7}$$

By adding and subtracting the expressions

$$\tilde{L}_j^{-1} P_j D_{j-1} P_j^T \tilde{U}_j^{-1} \quad \text{and} \quad L_j^{-1} P_j D_{j-1} P_j^T \tilde{U}_j^{-1}$$

we obtain

$$\begin{aligned}\epsilon_j &= \tilde{L}_j^{-1} P_j \epsilon_{j-1} P_j^T \tilde{U}_j^{-1} + \delta D_j + (\tilde{L}_j^{-1} - L_j^{-1}) P_j D_{j-1} P_j^T \tilde{U}_j^{-1} \\ &\quad + L_j^{-1} P_j D_{j-1} P_j^T (\tilde{U}_j^{-1} - U_j^{-1}) \;=\; \tilde{L}_j^{-1} P_j \epsilon_{j-1} P_j^T \tilde{U}_j^{-1} + Z_j\,,\end{aligned} \tag{3.8}$$

where

$$Z_j = \begin{pmatrix} 0 & & \\ & A_j \tilde{A}_j^{-1} - I & T_j - A_j \tilde{A}_j^{-1} \tilde{T}_j \\ \delta V_j + (V_j - \tilde{V}_j) A_j \tilde{A}_j^{-1} & \delta M_j - \delta V_j \tilde{T}_j + (V_j - \tilde{V}_j)(T_j - A_j \tilde{A}_j^{-1} \tilde{T}_j) \end{pmatrix}.$$

Solving the difference equation in (3.8) yields

$$\epsilon_k = \sum_{j=1}^{k} \prod_{i=1}^{k-j} (P_{k-i+1}^T \tilde{L}_{k-i+1})^{-1} Z_j \prod_{i=j+1}^{k} (\tilde{U}_i P_i)^{-1}.$$

For $j = 1, \ldots, k$, let $\mathbf{L}_j = \prod_{i=1}^{j} P_i^T \tilde{L}_i$ and $\mathbf{U}_j = \prod_{i=1}^{j} \tilde{U}_{j-i+1} P_{j-i+1}$ ($\mathbf{L}_k = \mathbf{L}$ and $\mathbf{U}_k = \mathbf{U}$). We multiply both members of the equation from the left by $\mathbf{L}_k$ and from the right by $\mathbf{U}_k$ to obtain

$$\mathbf{L}_k (\tilde{D}_k - D_k) \mathbf{U}_k = \sum_{j=1}^{k} \mathbf{L}_j Z_j \mathbf{U}_j$$

and hence

$$\mathbf{L}_k \tilde{D}_k \mathbf{U}_k = \mathbf{L}_k D_k \mathbf{U}_k + \sum_{j=1}^{k} \mathbf{L}_j Z_j \mathbf{U}_j = \mathbf{L}_k (D_k + Z_k) \mathbf{U}_k + \sum_{j=1}^{k-1} \mathbf{L}_j Z_j \mathbf{U}_j. \tag{3.9}$$

As to the first term on the far right of (3.9), we have

$$\mathbf{L}_{k-1} P_k^T \tilde{L}_k (D_k + Z_k) \tilde{U}_k P_k \mathbf{U}_{k-1} = \mathbf{L}_{k-1} D_{k-1} \mathbf{U}_{k-1} + \mathbf{L}_{k-1} P_k^T H_k^F P_k \mathbf{U}_{k-1}.$$

By iterating on $\mathbf{L}_{k-1} D_{k-1} \mathbf{U}_{k-1}$ and on the second term on the far right of (3.9) it follows that

$$\delta M = \mathbf{L}_k \tilde{D}_k \mathbf{U}_k - M = \sum_{j=1}^{k} \mathbf{L}_{j-1} P_j^T H_j^F P_j \mathbf{U}_{j-1}. \tag{3.10}$$

The matrix

$$P_j^T H_j^F P_j = \begin{pmatrix} 0 & \\ & Q_j^T \begin{pmatrix} 0 & \\ \delta V_j \tilde{A}_j & \delta M_j \end{pmatrix} Q_j \end{pmatrix} \tag{3.11}$$

has only a $\lfloor n/2^{j-1} \rfloor \times \lfloor n/2^{j-1} \rfloor$ block on the main diagonal different from zero. By what was said in the previous section, the matrices $\mathbf{L}_{j-1}$ and $\mathbf{U}_{j-1}$ are respectively lower and upper triangular, and they have the identity matrix in place of the nonnull block of (3.11). The result follows by simplifying the expression in (3.10). □

In order to obtain an upper bound for the norm of δM, we consider the following

Theorem 5. *If M is diagonally dominant, then the matrices H_j^F in* (3.6) *satisfy*

$$||H_j^F||_\infty \leq ||\tilde{M}_j||_\infty u + 2||\tilde{M}_{j-1}||_\infty u \leq 3\,||M||_\infty u,$$

and for the error matrix we have

$$||\delta M||_\infty \leq 3k||M||_\infty u. \tag{3.12}$$

Proof. See the Appendix. □

4. Stability of the solution of the triangular systems

Because of roundoff errors, instead of the exact solution of the lower triangular system, we obtain an approximate solution $\tilde{\mathbf{y}}$ which is the exact solution of a problem

$$(\mathbf{L} + \delta\mathbf{L})\tilde{\mathbf{y}} = \mathbf{f}.$$

To calculate $\delta\mathbf{L}$, we introduce, for each L_i, an error matrix δL_i. From Algorithm 1 we see that the error at the step j can be obtained from

$$\tilde{y}_i^{(j)} = \frac{\tilde{y}_{2i}^{(j-1)} - (\tilde{s}_i\tilde{y}_{2i-1}^{(j-1)}(1+u_1+u_3) + \tilde{t}_i\tilde{y}_{2i+1}^{(j-1)}(1+u_2+u_3)}{1+u_4}, \qquad |u_i| \leq u, \tag{4.1}$$

or, if written in block form, from

$$\tilde{\mathbf{y}}_j = (I + \delta\zeta_j)^{-1}(\tilde{\mathbf{y}}_{j-1}^e - (\tilde{V}_j + \delta\tilde{V}_j)\tilde{\mathbf{y}}_{j-1}^o), \tag{4.2}$$

where

$$|\delta\tilde{V}_j| \leq 2|\tilde{V}_j|\,u, \qquad |\delta\zeta_j| \leq Iu. \tag{4.3}$$

Therefore, the error matrix at the step j is

$$\delta L_j = \begin{pmatrix} 0 & \\ & 0 \\ \delta\tilde{V}_j & \delta\zeta_j \end{pmatrix}.$$

Our aim is to find a suitable bound for the norm in (3.2). For this, we first consider the product $\delta\mathbf{L}\mathbf{U}$.

Theorem 6. *The product $\delta\mathbf{L}\mathbf{U}$ introduced in* (3.2) *satisfies*

$$\delta\mathbf{L}\mathbf{U} = \sum_{j=1}^{k}\prod_{i=1}^{j} P_i^T H_j^L \prod_{i=1}^{j} P_{j-i+1}, \tag{4.4}$$

where

$$H_j^L = \begin{pmatrix} 0 & \\ & 0 \\ \delta\tilde{V}_j\tilde{A}_j & \delta\tilde{V}_j\tilde{T}_j + \delta\zeta_j\tilde{M}_j \end{pmatrix}. \tag{4.5}$$

Proof. In first-order approximation, we have

$$\prod_{j=1}^{k} P_j^T(\tilde{L}_j + \delta L_j) = \prod_{j=1}^{k} P_j^T \tilde{L}_j + \sum_{j=1}^{k} \mathbf{L}_{j-1} P_j^T \delta L_j \prod_{i=j+1}^{k} P_i^T \tilde{L}_i = \mathbf{L} + \delta \mathbf{L}.$$

Then

$$\begin{aligned} \delta \mathbf{L}\,\mathbf{U} &= \sum_{j=1}^{k} \mathbf{L}_{j-1} P_j^T \delta L_j \prod_{i=j+1}^{k} P_i^T \tilde{L}_i \tilde{D}_k \prod_{i=1}^{k} \tilde{U}_{k-i+1} P_{k-i+1} \\ &= \sum_{j=1}^{k} \mathbf{L}_{j-1} P_j^T \delta L_j \tilde{D}_j \mathbf{U}_j. \end{aligned} \tag{4.6}$$

The structures of $\mathbf{L}_{j-1}$ and $\mathbf{U}_j$ allow us to simplify the expression (4.6) and to obtain the following final expression for $\delta\mathbf{LU}$:

$$\delta \mathbf{LU} = \sum_{j=1}^{k} \prod_{i=1}^{j} P_i^T \delta L_j \tilde{D}_j \tilde{U}_j \prod_{i=1}^{j} P_{j-i+1}.$$

From this, the assertion follows. □

The next theorem establishes an upper bound for the infinity norm of $\delta\mathbf{LU}$.

Theorem 7. *If M is diagonally dominant, then the matrices H_j^L in* (4.5) *satisfy*

$$||H_j^L||_\infty \le 2(||\tilde{M}_j||_\infty + ||\tilde{M}_{j-1}||_\infty)\, u \le 4\, ||M||_\infty\, u$$

and

$$||\delta \mathbf{LU}||_\infty \le 4k\, ||M||_\infty\, u. \tag{4.7}$$

Proof. See the Appendix. □

Similar results can be obtained for the solution of the upper triangular system. The following theorems hold.

Theorem 8. *The product $\mathbf{L}\delta\mathbf{U}$ introduced in* (3.2) *satisfies*

$$\mathbf{L}\delta\mathbf{U} = \sum_{j=1}^{k} \prod_{i=1}^{j} P_i^T H_j^U \prod_{i=1}^{j} P_{k-i+1}, \tag{4.8}$$

where

$$\begin{aligned} H_j^U &= \begin{pmatrix} 0 & \\ & \delta\tilde{A}_j & \delta\tilde{T}_j \\ & \tilde{V}_j\, \delta\tilde{A}_j & \tilde{V}_j \delta\tilde{T}_j \end{pmatrix} \quad \text{for } j < k, \\ H_k^U &= \begin{pmatrix} 0 & \\ & \delta\tilde{A}_k & \delta\tilde{T}_k \\ & \tilde{V}_k\, \delta\tilde{A}_k & \tilde{V}_k \delta\tilde{T}_k + \delta\tilde{M}_k \end{pmatrix}, \end{aligned} \tag{4.9}$$

and

$$|\delta \tilde{M}_k| \leq |\tilde{M}_k|\, u, \qquad |\delta \tilde{A}_j| \leq 2\,|\tilde{A}_j|\, u, \qquad |\delta \tilde{T}_j| \leq 2\,|\tilde{T}_j|\, u.$$

Theorem 9. *If M is diagonally dominant, then the matrices H_j^U in (4.9) satisfy*

$$||H_j^U||_\infty \leq 2\,||\tilde{M}_{j-1}||_\infty\, u + ||\tilde{M}_j||_\infty\, u \leq 3\,||M||_\infty\, u$$

and

$$||\mathbf{L}\delta\mathbf{U}||_\infty \leq 3\,k\,||M||_\infty\, u. \tag{4.10}$$

The next theorem summarizes the results for the backward and forward error analyses made in this and in the previous section.

Theorem 10. *If the tridiagonal $n \times n$ matrix M is diagonally dominant by rows or by columns then, by using the cyclic reduction for solving $M\mathbf{x} = \mathbf{f}$, the computed solution $\tilde{\mathbf{x}}$ satisfies, in first-order approximation,*

$$(M + \Delta M)\tilde{\mathbf{x}} = \mathbf{f}, \qquad ||\Delta M||_\infty \leq 10\,\log_2 n ||M||_\infty\, u,$$

and

$$\frac{||\tilde{\mathbf{x}} - \mathbf{x}||_\infty}{||\tilde{\mathbf{x}}||_\infty} \leq 10\,\log_2 n\kappa(M)\, u, \tag{4.11}$$

where $\kappa(M)$ is the condition number of the matrix M in the infinity norm.

Proof. The backward error derives from (3.12), (4.7), and (4.10). The forward error is obtained from (3.3). □

5. Numerical experiments

We have carried out numerical experiments to compare our estimates of the relative error in the cyclic reduction (CR) algorithm with known estimates in the LU factorization. For the LU factorization we have used in our comparision the following upper bound (see [8, 16]):

$$\frac{||\tilde{\mathbf{x}} - \mathbf{x}||_\infty}{||\tilde{\mathbf{x}}||_\infty} \leq 8\kappa(M)\, u. \tag{5.1}$$

Numerical experiments indicate that the upper bound (4.11) is too pessimistic. In order to improve it, we have computed the following sharper bound for the norm of the error matrix, using (3.5), (4.4), and (4.8):

$$||\Delta M||_\infty \leq \left|\left| \sum_{j=1}^{k} \prod_{i=1}^{j} P_i^T \left(|H_j^F| + |H_j^L| + |H_j^U|\right) \prod_{i=1}^{j} P_{j-i+1} \right|\right|_\infty u. \tag{5.2}$$

We then have from (3.3) the following upper bound:

$$\frac{||\tilde{\mathbf{x}} - \mathbf{x}||_\infty}{||\tilde{\mathbf{x}}||_\infty} \leq \kappa(M) \frac{||\Delta M||_\infty}{||M||_\infty}, \tag{5.3}$$

where $||\Delta M||_\infty$ can be estimated by (5.2) .

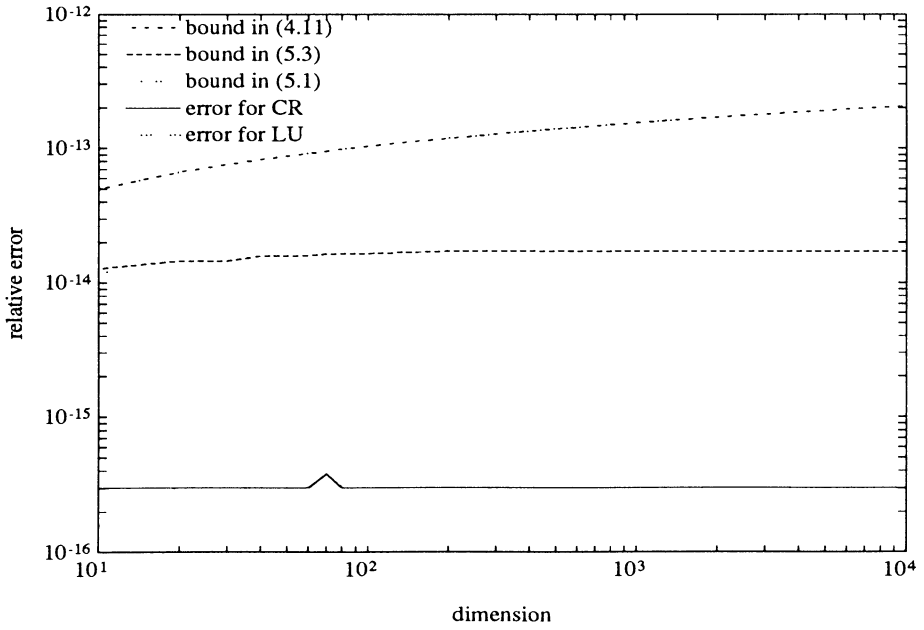

FIGURE 1. Upper bounds and relative errors for the test problem 1

We report on four numerical examples, all having different coefficient matrices. The right-hand sides $\mathbf{f}$ were always chosen in order to obtain

$$\mathbf{x} = 1.4142 \cdot (2, -1, 2, -1, \ldots 2, -1)^T$$

as the solution of problem (2.1). We used the algorithm in [7] to compute the norm of the inverse of the coefficient matrix.

Test problem 1 (see Figure 1). Consider the diagonally dominant Toeplitz matrix

$$M = \begin{pmatrix} 4 & -1 & & \\ -2 & 4 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -2 & 4 \end{pmatrix}_{n \times n},$$

with condition number $\kappa(M) \leq 7$. The solution of (2.1) by the CR algorithm and by the LU factorization is obtained to within machine precision. Therefore, the upper bounds are excessively large. In particular, there is no error growth corresponding to the term $\log_2 n$ in (4.11).

Test problem 2 (see Figure 2). Consider a symmetric weakly diagonally dominant Toeplitz matrix

$$M = \begin{pmatrix} 5 & -2.5 & & \\ -2.5 & 5 & \ddots & \\ & \ddots & \ddots & -2.5 \\ & & -2.5 & 5 \end{pmatrix}_{n \times n}.$$

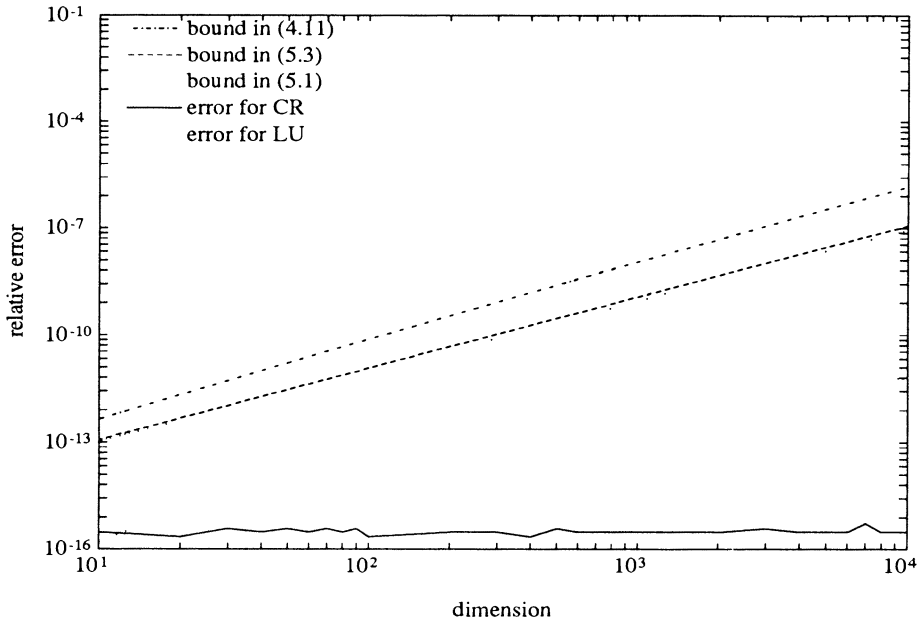

FIGURE 2. Upper bounds and relative errors for the test problem 2

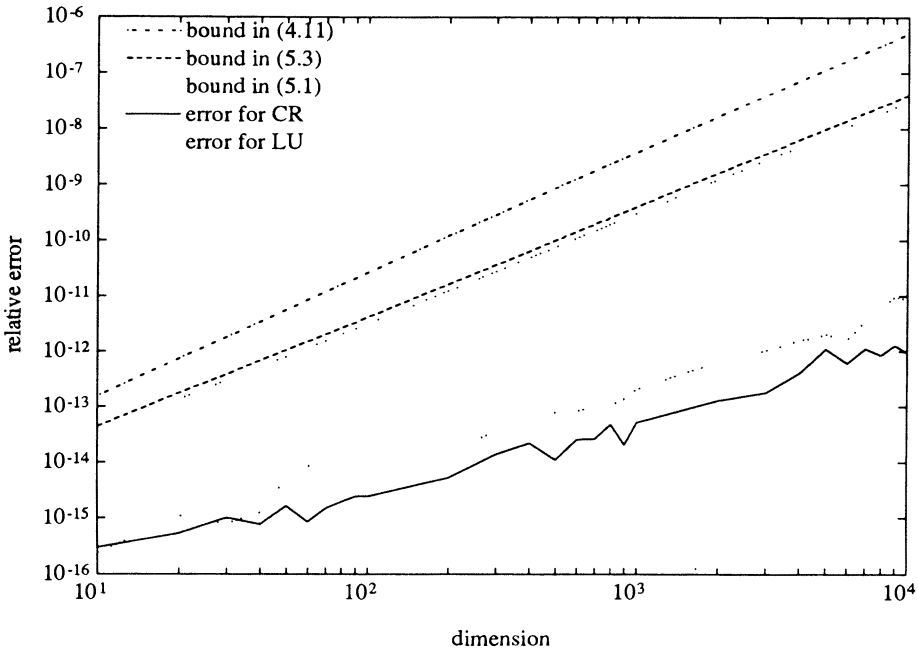

FIGURE 3. Upper bounds and relative errors for the test problem 3

For matrices in this class, operations of the CR algorithm are performed exactly. Therefore the relative error is proportional to the machine precision. The condition number and the relative error for the LU factorization algorithm grow as $O(n^2)$.

Test problem 3 (see Figure 3). Consider a weakly diagonally dominant matrix M, of order n, defined by (see (2.2))

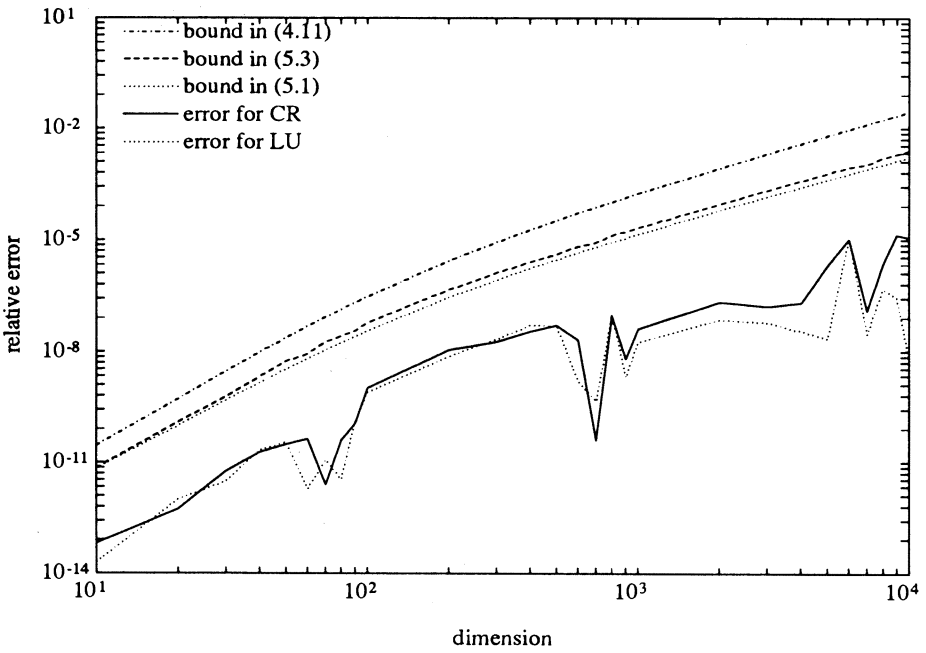

FIGURE 4. Upper bounds and relative errors for the test problem 4

$$a_i = 1, \quad b_i = -\frac{2\epsilon + (1 - i \cdot h)h}{4\epsilon}, \quad c_i = -\frac{2\epsilon - (1 - i \cdot h)h}{4\epsilon},$$

where $h = 2/(n+1)$. This matrix occurs in the discretization of a singular perturbation problem by finite differences (see [12]). We chose $\epsilon = 0.1$ in order to have diagonal dominance. Its condition number is $O(n^2)$.

Test problem 4 (see Figure 4). An ill-conditioned problem ($\kappa(M) = 1.64 \cdot 10^9$ for $n = 500$) derived from the discretization of another singular perturbation problem by finite differences given in [5, 8]. The $n \times n$ coefficient matrix M is defined by

$$c_i = \begin{cases} -\frac{\epsilon}{h^2} & \text{if } 1 \le i \le \lfloor \frac{n+1}{2} \rfloor, \\ -\frac{\epsilon}{h^2} + \frac{0.5 - i \cdot h}{h} & \text{if } \lfloor \frac{n+1}{2} \rfloor + 1 \le i \le n, \end{cases}$$

$$b_i = \begin{cases} -\frac{\epsilon}{h^2} - \frac{0.5 - i \cdot h}{h} & \text{if } 1 \le i \le \lfloor \frac{n+1}{2} \rfloor, \\ -\frac{\epsilon}{h^2} & \text{if } \lfloor \frac{n+1}{2} \rfloor + 1 \le i \le n, \end{cases}$$

$$a_i = -(b_i + c_i),$$

where $h = 1/(n+1)$ and $\epsilon = 0.009$. The numbers c_1 and b_n do not occur in the matrix M, but are only introduced to define a_1 and a_n. The matrix M is a nonsingular, row diagonally dominant M-matrix, but it becomes singular for $\epsilon \to 0$.

We conclude that the cyclic reduction algorithm is stable for tridiagonal diagonally dominant matrices. From our experiments we observe that the dependence on $\log_2 n$ is pessimistic and does not describe the actual behavior of the error (the upper bound in (5.3) has the same behavior as that in (5.1)).

6. Appendix

Proof of Theorem 5. The matrices δV_j and δM_j have the following structure:

$$\delta V_j = \begin{pmatrix} \delta s_1^{(j)} & \delta t_1^{(j)} & & \\ & \delta s_2^{(j)} & \delta t_2^{(j)} & \\ & & \ddots & \ddots \end{pmatrix} \quad \text{and} \quad \delta M_j = \begin{pmatrix} \delta a_1^{(j)} & \delta b_1^{(j)} & \\ \delta c_2^{(j)} & \delta a_2^{(j)} & \ddots \\ & \ddots & \ddots \end{pmatrix},$$

where

$$\begin{aligned} &|\delta s_i^{(j)}| \le |\tilde{c}_{2i}^{(j-1)}/\tilde{a}_{2i-1}^{(j-1)}|\, u, \qquad |\delta t_i^{(j)}| \le |\tilde{b}_{2i}^{(j-1)}/\tilde{a}_{2i+1}^{(j-1)}|\, u, \\ &|\delta a_i^{(j)}| \le |\tilde{a}_i^{(j)}|\, u + 2|\tilde{s}_i^{(j)}\tilde{b}_{2i-1}^{(j-1)}|\, u + 2|\tilde{t}_i^{(j)}\tilde{c}_{2i+1}^{(j-1)}|\, u, \\ &|\delta b_i^{(j)}| \le |\tilde{t}_i^{(j)}\tilde{b}_{2i+1}^{(j-1)}|\, u, \qquad |\delta c_i^{(j)}| \le |\tilde{s}_i^{(j)}\tilde{c}_{2i-1}^{(j-1)}|\, u. \end{aligned}$$

From the hypothesis of diagonal dominance we have

$$|\tilde{s}_i^{(j)}\tilde{b}_{2i-1}^{(j-1)}| + |\tilde{t}_i^{(j)}\tilde{c}_{2i+1}^{(j-1)}| \;\le\; |\tilde{c}_{2i}^{(j-1)}| + |\tilde{b}_{2i}^{(j-1)}| \;\le\; |\tilde{a}_{2i}^{(j-1)}|$$

and therefore

$$\begin{aligned} \|H_j^F\|_\infty &\le \max_i(|\delta a_i^{(j)}| + |\delta b_i^{(j)}| + |\delta c_i^{(j)}| + |\delta s_i^{(j)} a_{2i-1}^{(j-1)}| + |\delta t_i^{(j)} a_{2i+1}^{(j-1)}|) \\ &\le \max_i(|\tilde{a}_i^{(j)}| + 2|\tilde{a}_{2i}^{(j-1)}| + |\tilde{b}_i^{(j)}| + |\tilde{c}_i^{(j)}| + |\tilde{b}_{2i}^{(j-1)}| + |\tilde{c}_{2i}^{(j-1)}|)u \\ &\le \|\tilde{M}_j\|_\infty\, u + 2\,\|\tilde{M}_{j-1}\|_\infty u. \end{aligned}$$

The assertion now follows from Theorem 3 and (3.7). □

Proof of Theorem 7. From (4.1) and (4.2) we obtain

$$\delta \tilde{V}_j = \begin{pmatrix} \delta \tilde{s}_1^{(j)} & \delta \tilde{t}_1^{(j)} & & \\ & \delta \tilde{s}_2^{(j)} & \delta \tilde{t}_2^{(j)} & \\ & & \ddots & \ddots \end{pmatrix},$$

where

$$|\delta \tilde{s}_i^{(j)}| \le 2|\tilde{c}_{2i}^{(j-1)}/\tilde{a}_{2i-1}^{(j-1)}|\, u, \qquad |\delta \tilde{t}_i^{(j)}| \le 2|\tilde{b}_{2i}^{(j-1)}/\tilde{a}_{2i+1}^{(j-1)}|\, u.$$

Moreover, from (4.3) we have

$$|\delta \zeta_j\, \tilde{M}_j| \le |\tilde{M}_j|\, u.$$

From the diagonal dominance we conclude

$$\begin{aligned}\|H_j^L\|_\infty &\le \max_i (2(|\tilde{s}_i^{(j)}\tilde{b}_{2i-1}^{(j-1)}| + |\tilde{t}_i^{(j)}\tilde{c}_{2i+1}^{(j-1)}| + |\tilde{s}_i^{(j)}\tilde{c}_{2i-1}^{(j-1)}| + |\tilde{t}_i^{(j)}\tilde{b}_{2i+1}^{(j-1)}|) \\ &\qquad + |\tilde{a}_i^{(j)}| + |\tilde{b}_i^{(j)}| + |\tilde{c}_i^{(j)}| + 2(|\tilde{s}_i^{(j)}\tilde{a}_{2i-1}^{(j-1)}| + |\tilde{t}_i^{(j)}\tilde{a}_{2i+1}^{(j-1)}|))u \\ &\le 2\max_i (|\tilde{a}_{2i}^{(j-1)}| + |\tilde{c}_i^{(j)}| + |\tilde{b}_i^{(j)}| + |\tilde{a}_i^{(j)}| + |\tilde{b}_{2i}^{(j-1)}| + |\tilde{c}_{2i}^{(j-1)}|)u \\ &\le 2\left(\|\tilde{M}_j\|_\infty + \|\tilde{M}_{j-1}\|_\infty\right) u.\end{aligned}$$

The assertion now follows from Theorem 3 and (3.7). □

Acknowledgments

We are very grateful to Professor Donato Trigiante for his helpful comments and suggestions. Moreover, we are very indebted to Professor W. Gautschi and to the referee for suggestions which have improved our paper.

Bibliography

1. P. Amodio, *Optimized cyclic reduction for the solution of linear tridiagonal systems on parallel computers*, Comput. Math Appl. **26** (1993), 45–53.
2. P. Amodio and L. Brugnano, *Parallel factorizations and parallel solvers for tridiagonal linear systems*, Linear Algebra Appl. **172** (1992), 347–364.
3. P. Amodio, L. Brugnano, and T. Politi, *Parallel factorizations for tridiagonal matrices*, SIAM J. Numer. Anal. **30** (1993), 813–823.
4. B. L. Buzbee, G. H. Golub, and C. W. Nielson, *On direct methods for solving Poisson's equations*, SIAM J. Numer. Anal. **7** (1970), 627–656.
5. F. W. Dorr, *An example of ill-conditioning in the numerical solution of singular perturbation problems*, Math. Comp. **25** (1971), 271–283.
6. D. Heller, *Some aspects of the Cyclic Reduction Algorithm for block tridiagonal linear systems*, SIAM J. Numer. Anal. **13** (1976), 484–496.
7. N. J. Higham, *Efficient algorithms for computing the condition number of a tridiagonal matrix*, SIAM J. Sci. Statist. Comput. **7** (1986), 150–165.
8. ———, *Bounding the error in Gaussian elimination for tridiagonal systems*, SIAM J. Matrix Anal. Appl. **11** (1990), 521–530.
9. S. L. Johnsson, *Solving tridiagonal systems on ensemble architectures*, SIAM J. Sci. Statist. Comput. **8** (1987), 354–392.
10. D. Kershaw, *Solution of single tridiagonal linear systems and vectorization of the ICCG algorithm on the CRAY* 1, Parallel Computations (G. Rodrigue, ed.), Academic Press, New York, 1982, pp. 85–99.
11. J. J. Lambiotte, R. G. Voigt, *The solution of tridiagonal linear systems on the CDC Star*-100 *computer*, ACM Trans. Math. Software **1** (1975), 308–329.
12. F. Mazzia and D. Trigiante, *Numerical methods for second order singular perturbation problems*, Comput. Math. Appl. **23** (1992), 81–89.
13. J. M. Ortega, *Introduction to parallel and vector solution of linear systems*, Plenum Press, New York, 1988.
14. H. S. Stone, *An efficient parallel algorithm for the solution of a tridiagonal linear system of equations*, J. Assoc. Comput. Mach. **20** (1973), 27–38.

15. H. A. Van der Vorst, *Large tridiagonal and block tridiagonal linear systems on vector and parallel computers*, Parallel Comput. **5** (1987), 45–54.

16. J. H. Wilkinson, *The algebraic eigenvalue problem*, Oxford Univ. Press, Oxford, 1965.

17. ———, *Error analysis of direct methods of matrix inversion*, J. Assoc. Comput. Mach. **8** (1961), 281–330.

DIPARTIMENTO DI MATEMATICA, UNIVERSITÀ DI BARI, VIA ORABONA, 4, I-70125 BARI, ITALY
E-mail address: 00110570@vm.csata.it