BULL. AUSTRAL. MATH. SOC. 65F05, 65L10, 65N20
VOL. 29 (1984), 177-205.

THE STABILITY OF LU-DECOMPOSITIONS OF BLOCK TRIDIAGONAL MATRICES

R.M.M. MATTHEIJ

An investigation is made of the stability of block LU-decomposition of matrices A arising from boundary value problems of differential equations, in particular of ordinary differential equations with separated boundary conditions. It is shown that for such matrices the pivotal growth can be bounded by constants of the order of $\|A\|$ and, if the solution space is dichotomic, often by constants of order one. Furthermore a method to estimate the growth of the pivotal blocks is given. A number of examples support the analysis.

1. Introduction

Block tridiagonal systems of linear equations occur in a wide variety of problems, in particular in discretized differential equations. Roughly speaking, one may distinguish between systems where the block structure is induced by the ordering of the gridpoints, as in finite difference or finite element methods for partial differential equations, and systems where a block partitioning is employed just for computational reasons, usually in boundary value methods for ordinary differential equations. In the latter case the off diagonal blocks then systematically have a number

---

Received 17 October 1983. This work was supported in part by the Netherlands Organization for the Advancement of Pure Research (ZWO) and the Niels Stensen Stichting.

---

of zero rows. In solving such a system one may prefer to employ block LU-decomposition where the zero pattern is preserved. The existence and the stability of such decompositions have been investigated in Keller [4], Varah [11, 12, 13]. For the first class of block tridiagonal matrices, concepts reflecting certain analytical properties of the original partial differential equations, like positivity, positive definiteness or diagonal dominance are often sufficient in order to show this existence and stability. For the second class the aforementioned concepts usually do not make sense, because of the rather artificial partitioning. Therefore there is a need for a theory that justifies the use of block LU-decomposition in more general cases.

Although some of the estimation methods we give in this paper in principle hold for general block tridiagonal matrices, we shall concentrate most of our attention on matrices belonging to the second class. Such matrices arise in ordinary differential equations where separated boundary conditions are given and the discretization method can be described by a one step recursion. We show that the partitioning is closely related to the splitting of the fundamental solution of this ordinary differential equation into nondecreasing and nonincreasing solutions. In this way the LU-decomposition can be looked upon as a decoupling of these solutions. As a result we can show that suitable pivoting strategies, that preserve the zero pattern, lead to a stable block LU-decomposition.

In Section 2 we give an explicit formulation of the LU-decomposition. In Section 3 we consider the special type of block tridiagonal matrices as was indicated above. In Section 4 we give a number of estimation methods applicable to both classes of matrices. Finally, a number of examples in Section 5 support the theory.

2. LU-decomposition of block tridiagonal matrices

Consider the linear system

$$Ax = b \,, \tag{2.1}$$

where A is a block tridiagonal matrix with blocks of order n, that is, A can be written as

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

$$\mathbb{A} = \begin{bmatrix} B_1 & C_1 & & & \\ A_2 & B_2 & C_2 & \emptyset & \\ & \ddots & \ddots & \ddots & \\ \emptyset & & A_{N-1} & B_{N-1} & C_{N-1} \\ & & & A_N & B_N \end{bmatrix} . \tag{2.2}$$

Let l $(0 \le l \le n)$ denote the number of first rows of the C_i that are systematically zero and k $(0 \le k \le n)$ the number of last such rows of the A_i. We now look for block LU-decompositions that have as many systematically zero rows in the off diagonal blocks as $\mathbb{A}$ has. This may allow left and right multiplication of $\mathbb{A}$ by certain block diagonal matrices $\mathbb{D}$ and $\mathbb{E}$ say, which do not disturb the zero row pattern. So we investigate the factorization

$$\hat{\mathbb{A}} := \mathbb{D}\mathbb{A}\mathbb{E} = \mathbb{L}\mathbb{U} , \tag{2.3}$$

where $\mathbb{L}$ and $\mathbb{U}$ must have the form

$$\mathbb{L} := \begin{pmatrix} I & & & \emptyset \\ L_2 & I & & \\ & \ddots & \ddots & \\ \emptyset & & L_N & I \end{pmatrix} , \quad \mathbb{U} := \begin{pmatrix} U_1 & \hat{C}_1 & & \emptyset \\ & U_2 & \hat{C}_2 & \\ & & \ddots & \ddots \\ & & & \hat{C}_{N-1} \\ \emptyset & & & U_N \end{pmatrix} . \tag{2.4}$$

It can easily be checked that the proper zero pattern is preserved if Assumption 2.5 is satisfied.

ASSUMPTION 2.5. Let $k + l = n$ and $\mathbb{D}$ be a block diagonal matrix of the form $\mathbb{D} = \mathrm{diag}\big(d_1(n-k), D_2, \ldots, D_N, d_N(k)\big)$, where $D_2, \ldots, D_N$ are nth order nonsingular matrices and $d_1(n-k)$, $d_N(k)$ are $(n-k)$th order and kth order nonsingular matrices respectively. Let $\mathbb{E}$ be a block diagonal matrix of the form $\mathbb{E} = \mathrm{diag}\big(E_1, \ldots, E_N\big)$, where E_j is nonsingular and of order n.

If we partition $\hat{\mathbb{A}}$ as $\mathbb{A}$ in (2.2) then we obtain (in an obvious notation) from (2.3):

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

$$(a)\quad U_1 = \hat{B}_1 \,,$$

$$(2.6)\quad (b)\quad U_i = \hat{B}_i - \hat{A}_i U_{i-1}^{-1} \hat{C}_{i-1} \,, \quad 2 \le i \le N \,,$$

$$(c)\quad L_i = \hat{A}_i U_{i-1}^{-1} \,.$$

In the sequel we assume that $\|\cdot\|$ denotes a Hölder norm. In order to examine the stability of this LU-decomposition, we need bounds for $\|\mathbb{L}\|$ and $\|\mathbb{U}\|$, and in particular for $\|L_i\|$ and $\|U_i\|$ (*cf.* Varah [11]). From (2.6) we see that both $\|L_i\|$ and $\|U_i\|$ $(i \ge 2)$ can be estimated if $\min \operatorname{glb}(U_j)$ $\left(= \left[\max\left\|U_i^{-1}\right\|\right]^{-1}\right)$ is known. The bound $\|\mathbb{L}\|\|\mathbb{U}\|$ may be used as a stability constant in a backward error analysis. However, the nice bidiagonal form of $\mathbb{L}$ and $\mathbb{U}$ may sometimes also make a forward error analysis attractive (*cf.* Mattheij [8]), in which case we may use estimates for L_i and U_i^{-1} directly. On account of this we therefore focus our attention to finding estimates for the blocks in $\mathbb{L}$ and $\mathbb{U}$. The first method we shall deal with, is based on comparing them to a special LU-decomposition, namely,

$$(2.7)\qquad \tilde{\mathbb{D}} \mathbb{A} \tilde{\mathbb{E}} = \tilde{\mathbb{L}} \tilde{\mathbb{U}}$$

(with matrices like those in (2.3)). From (2.3) and (2.7) we obtain

$$(2.8)\qquad \mathbb{U}\mathbb{E}^{-1}\tilde{\mathbb{E}}\tilde{\mathbb{U}}^{-1} = \mathbb{L}^{-1}\mathbb{D}\tilde{\mathbb{D}}^{-1}\tilde{\mathbb{L}} \,.$$

The matrix on the left in (2.8) is block upper triangular (like $\mathbb{U}$) while the matrix on the right is block lower triangular, but with a different block structure in general (that is, like $\mathbb{D}$). In order to describe more precisely their common form, let us use the following notation for blocks

$$(2.9)\qquad P = \begin{bmatrix} P^{11} & P^{12} \\ P^{21} & P^{22} \end{bmatrix}, \quad P^{11} \text{ is a } k \times k \text{ matrix.}$$

For the matrix in (2.8) we can then write

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

(2.10) $$\mathbb{L}^{-1}\mathbb{D}\tilde{\mathbb{D}}^{-1}\tilde{\mathbb{L}} = \begin{bmatrix} H_1 & K_1 & & & \\ & H_2 & K_2 & & \emptyset \\ & & \ddots & \ddots & \\ & & & \ddots & \\ \emptyset & & & \ddots & K_{N-1} \\ & & & & H_N \end{bmatrix}$$

where

(2.11) $$H_i = \begin{bmatrix} H_i^{22} & \emptyset \\ H_i^{12} & H_i^{11} \end{bmatrix}, \quad K_i = \begin{bmatrix} \emptyset & \emptyset \\ K_i^{12} & \emptyset \end{bmatrix}.$$

By consistently denoting blocks in L_i, $\tilde{L}_i$ as

(2.12) $$L_i = \begin{bmatrix} L_i^{22} & L_i^{21} \\ \emptyset & \emptyset \end{bmatrix}, \quad \tilde{L}_i = \begin{bmatrix} \tilde{L}_i^{22} & \tilde{L}_i^{21} \\ \emptyset & \emptyset \end{bmatrix},$$

we obtain the following explicit expression for the matrix in (2.10).

PROPERTY 2.13. *Define*

$$D_i\tilde{D}_i^{-1} =: \begin{bmatrix} P_i^{11} & P_i^{12} \\ P_i^{21} & P_i^{22} \end{bmatrix}, \quad 2 \le i \le N-1,$$

$d_1(n-k)\,[\tilde{d}_1(n-k)]^{-1} =: P_1^{22}$, $d_N(k)\,[\tilde{d}_N(k)]^{-1} =: P_N^{11}$ *and* $L_1 := 0$. *Then*

$$H_i = \begin{bmatrix} -L_i^{21}P_i^{12}+P_i^{22} & \emptyset \\ P_{i+1}^{12}\tilde{L}_{i+1}^{22} & P_{i+1}^{11} + P_{i+1}^{12}\tilde{L}_{i+1}^{21} \end{bmatrix}, K_i^{12} = P_{i+1}^{12}.$$

Proof. The relations above follow from simply writing out blocks in $\mathbb{L}^{-1}\mathbb{D}\tilde{\mathbb{D}}^{-1}\tilde{\mathbb{L}}$, in which one should realize that only the first block codiagonal of $\mathbb{L}^{-1}$ is of interest; this in turn is equal to the block codiagonal of $\mathbb{L}$ but for minus signs in front of the L_i. □

PROPERTY 2.14. $U_i = H_i\tilde{U}_iE_i\tilde{E}_i^{-1}$.

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

3. One step recursions and tridiagonal systems

There exists an important class of block tridiagonal systems where $k + l = n$, namely the equations resulting from solving ordinary differential equations with separated boundary conditions. These equations consist of the boundary conditions plus the one step recursions in the multiple shooting method, the collocation relations or difference equations (as in the Box scheme), (*cf.* Ascher [1], de Boor [3], Keller [4, 5], Russell [10], Varah [11, 12, 13]). They can be written as follows:

(3.1) $S_1 x_1 = e_1$, S_1 an $(n-k) \times n$ matrix, e_1 some $(n-k)$ vector;

(3.2) $F_i x_i = G_i x_{i+1} + c_i$, $1 \le i \le N-1$, where F_i and G_i are $n \times n$ matrices of which we assume G_i to be nonsingular;

(3.3) $S_N x_N = e_N$, S_N a $k \times n$ matrix, e_N some k vector.

Hence a matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} S_1 & & & \\ F_1 & -G_1 & & \emptyset \\ & \ddots & \ddots & \\ & & F_{N-1} & -G_{N-1} \\ \emptyset & & & S_N \end{bmatrix}. \tag{3.4}$$

It can easily be seen that $\mathbf{A}$ in (3.4) can indeed be partitioned as in (2.2) and such that the C_i have $(n-k)$ zero rows and the A_i have k zero rows. In this section we would like to discuss the stability of block LU-decompositions. This discussion is different from those in Varah [11, 12, 13], in that we try to relate pivotal growth to properties of the originating boundary value problem. This relationship can best be seen from a very special LU-decomposition, which we derive first.

In Mattheij [8, 9] a method was introduced to compute solutions $\{x_i\}_{i=1}^N$ of (3.1), (3.2) by using transformed versions of the incremental matrices $G_i^{-1}F_i$, with a (block) upper triangular form. The decoupling in these (block) upper triangular incremental matrices reflects the dichotomy of the solution space (if present) and can be employed to compute the non-increasing and nondecreasing components of the solution separately by using an appropriate direction (that is, forward and backward, respectively). It goes as follows.

Let Q_1 be an orthogonal matrix such that

$$S_1 Q_1 = \left(\emptyset \mid -\tilde{S}_1^{22}\right) \tag{3.5}$$

where $\tilde{S}_1^{22}$ is an $(n-k)$th order (possibly upper triangular) matrix. Then recursively compute sequences of orthogonal matrices $\{Q_i\}$ and (block) upper triangular matrices $\{V_i\}$ such that

$$\left(G_i^{-1} F_i\right) Q_i = Q_{i+1} V_i \ , \quad i = 1, 2, \ldots, N-1 \ . \tag{3.6}$$

We use the partioned notation $\left(cf.\ (2.9)\right)$

$$V_i = \begin{pmatrix} V_i^{11} & V_i^{12} \\ \emptyset & V_i^{22} \end{pmatrix} , \quad V^{11} \ \text{a} \ k \times k \ \text{matrix.} \tag{3.7}$$

Apparently the sequence $\left\{Q_i^{-1} x_i\right\}_{i=1}^{N}$ satisfies the decoupled recursion

$$Q_{i+1}^{-1} x_{i+1} = V_i \left(Q_i^{-1} x_i\right) - Q_{i+1}^{-1} G_i^{-1} c_i \ . \tag{3.8}$$

If we partition the vectors $\tilde{x}_i := Q_i^{-1} x_i$ into $\begin{pmatrix} \tilde{x}_i^1 \\ \tilde{x}_i^2 \end{pmatrix}$ where $\tilde{x}_i^1$ has k coordinates, then the decoupled form (3.8) can be employed to compute $\left\{\tilde{x}_i^2\right\}_{i=1}^{N}$ first and then the sequence $\left\{\tilde{x}_i^1\right\}_{i=N}^{1}$ (for details see Mattheij [8]). Under fairly relaxed conditions, moreover, this method is stable.

REMARK 3.9. Although it is outside the scope of this paper to compare numerical methods with respect to efficiency, it seems that the above described algorithm is not less efficient than other LU solvers and certainly does not require any pivoting strategy.

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

We now show that the above described algorithm is related to block LU-decomposition. Define

$$\text{(3.10)}\quad \begin{aligned} &(a)\quad \tilde{D}_j := Q_j^{-1} G_{j-1}^{-1} \,,\quad \tilde{E}_j := Q_j \,, \\ &(b)\quad \tilde{S}_N = \left(\tilde{S}_N^{11} \mid \tilde{S}_N^{12}\right) := S_N Q_N \,,\quad \tilde{S}_N^{11} \ \text{a}\ k \times k \ \text{matrix}, \\ &(c)\quad \tilde{d}_1(n-k) = I_{n-k} \,,\quad \tilde{d}_N(k) = I_k \,. \end{aligned}$$

Then we can form a block tridiagonal matrix $\tilde{\mathbb{A}}$ (*cf.* (2.3), Assumption 2.5)

$$\tilde{\mathbb{A}} := \tilde{\mathbb{D}}\,\mathbb{A}\,\tilde{\mathbb{E}} = \left[\begin{array}{cc|cc|ccc} \emptyset & -\tilde{S}_1^{22} & \emptyset & \emptyset & & & \\ V_1^{11} & V_1^{12} & -I & \emptyset & & & \\ \hline \emptyset & V_1^{22} & \emptyset & -I & & & \\ & & V_2^{11} & V_2^{12} & & & \\ \hline & & & & \ddots & & \\ \hline & & & & V_{N-1}^{22} & \emptyset & -I \\ & & & & & \tilde{S}_N^{11} & \tilde{S}_N^{12} \end{array}\right] .$$

If an LU-decomposition of $\tilde{\mathbb{A}}$ exists, say $\tilde{\mathbb{A}} = \tilde{\mathbb{L}}\tilde{\mathbb{U}}$, the following explicit expressions hold for the matrices say $\tilde{U}_i$, $\tilde{L}_i$, $\tilde{C}_i$ (*cf.* (2.4), (2.6)):

$$\text{(3.11)}\quad (a)\quad \tilde{C}_i = \begin{pmatrix} \emptyset & \emptyset \\ -I & 0 \end{pmatrix} ,\quad \tilde{U}_1 = \begin{pmatrix} \emptyset & -\tilde{S}_1^{22} \\ V_1^{11} & V_1^{12} \end{pmatrix} ,$$

$$\tilde{U}_1^{-1} = \left(\begin{array}{c|c} \left[V_1^{11}\right]^{-1} V_1^{12} \left[\tilde{S}_1^{22}\right]^{-1} & \left[V_1^{11}\right]^{-1} \\ \hline -\left[\tilde{S}_1^{22}\right]^{-1} & \emptyset \end{array}\right) ;$$

$$\text{(3.11)}\quad (b)\quad \tilde{U}_i = \begin{pmatrix} \emptyset & -I \\ V_i^{11} & V_i^{12} \end{pmatrix} ,\quad \tilde{U}_i^{-1} = \left(\begin{array}{c|c} \left[V_i^{11}\right]^{-1} V_i^{12} & \left[V_i^{11}\right]^{-1} \\ \hline -I & \emptyset \end{array}\right) ,$$

$$2 \le i \le N-1 \;;$$

$$(3.11)\quad (c)\quad \tilde{L}_2 = \begin{bmatrix} -V_1^{22}\left[\tilde{S}_1^{22}\right]^{-1} & \emptyset \\ & \\ \emptyset & \emptyset \end{bmatrix}, \quad \tilde{L}_i = \begin{bmatrix} -V_{i-1}^{22} & \emptyset \\ \emptyset & \emptyset \end{bmatrix}, \quad 3 \le i \le N ;$$

$$(3.11)\quad (d)\quad \tilde{U}_N = \begin{bmatrix} \emptyset & -I \\ \tilde{S}_N^{11} & \tilde{S}_N^{12} \end{bmatrix} .$$

Sufficient conditions for this decomposition to exist are given in

PROPERTY 3.12. *If* $\operatorname{rank}(S_1) = n - k$ *and if, for all* i , F_i *is nonsingular then the block* LU-*decomposition of* $\tilde{\mathbb{A}}$ *exists.*

Proof. $\operatorname{rank}(S_1) = \operatorname{rank}(\tilde{S}_1) = n - k$ implies $\tilde{S}_1^{22}$ is nonsingular. Moreover F_i nonsingular implies V_i nonsingular implies V_i^{11} non-singular. Hence for $i = 1, \ldots, N-1$ we find nonsingular $\tilde{U}_i$. □

It can simply be seen from (3.11) that the decomposition $\tilde{\mathbb{A}} = \tilde{\mathbb{L}}\tilde{\mathbb{U}}$ is stable, since we have

THEOREM 3.13. *Let* $\mathbb{A}$ *be nonsingular. Then the decomposition* $\tilde{\mathbb{D}}\tilde{\mathbb{A}}\tilde{\mathbb{E}} = \tilde{\mathbb{L}}\tilde{\mathbb{U}}$ *exists. Let* $\operatorname{glb}_2\left(\tilde{S}_1^{22}\right) = \sigma$ *; then*

$$\|\tilde{L}_2\|_2 \le \left\|G_1^{-1}F_1\right\|_2/\sigma ,$$

$$\|\tilde{L}_i\|_2 \le \left\|G_{i-1}^{-1}F_{i-1}\right\|_2 , \quad 3 \le i \le N ,$$

$$\|\tilde{U}_1\|_2 \le \left\|G_1^{-1}F_1\right\|_2 + \|S_1\|_2 ,$$

$$\|\tilde{U}_i\|_2 \le \left\|G_i^{-1}F_i\right\|_2 + 1 , \quad 2 \le i \le N-1 ,$$

$$\|\tilde{U}_N\|_2 \le \|S_N\|_2 + 1 .$$

We now show that almost similar estimates hold for other block LU-decompositions, provided a (restricted) partial pivoting strategy is used (Keller [4]). We shall focus on an obvious (special) method. The basic

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

idea is that one may use LU-decomposition to compute U_{i-1}^{-1} in (2.6). However, to ensure its existence it may be necessary to permute rows of F_{i-1}. In particular, without these permutations $\hat{B}_1$ may be singular; even if $\hat{B}_1$ is nonsingular we need special precautions to make the restricted pivoting work at all. We shall describe three variants to do this.

VARIANT I. Perform a "classical" LU-decomposition (with partial pivoting) of S_1. Use the thus found pivotal rows also to produce zeros in the first l columns of F_1. Assuming this decomposition exists and neglecting permutations this can be described as

$$(3.14) \qquad \begin{bmatrix} S_1 \\ F_1 \end{bmatrix} = \left[\begin{array}{c|c} L^1 & \emptyset \\ \hline L^2 & \\ L^3 & I_{n-l} \end{array}\right] \begin{bmatrix} U^1 & U^2 \\ \emptyset & U^3 \end{bmatrix},$$

where L^1 is an $l \times l$ unit lower triangular matrix, L^2 is an $(n-l) \times l$ matrix, L^3 an $l \times l$ matrix, U^1 an $l \times l$ upper triangular matrix, U^2 an $l \times (n-l)$ matrix and U^3 an $n \times (n-l)$ matrix. After this we perform an LU-decomposition (with partial pivoting) of U^3, say (again omitting permutations)

$$(3.15) \qquad [U^3] = \begin{bmatrix} L^4 \\ L^5 \end{bmatrix} [U^4],$$

where L^4 is an $(n-l) \times (n-l)$ unit lower triangular matrix, L^5 an $l \times (n-l)$ matrix and U^4 an $(n-l) \times (n-l)$ upper triangular matrix. Combining (3.14), (3.15) we obtain

$$(3.16) \qquad \begin{bmatrix} S_1 \\ F_1 \end{bmatrix} = \left[\begin{array}{c|c} L^1 & \emptyset \\ \hline L^2 & L^4 \\ L^3 & L^5 \end{array}\right] \begin{bmatrix} U^1 & U^2 \\ \emptyset & U^4 \end{bmatrix}.$$

It is important to realize that the possible permutations do not introduce nonzero elements in the zero blocks in (3.16). Obviously we should use the factorized matrix

(3.17) $$\begin{bmatrix} L^1 & \emptyset \\ L^2 & L^4 \end{bmatrix}\begin{bmatrix} U^1 & U^2 \\ \emptyset & U^4 \end{bmatrix}$$

to compute U_1^{-1} in (2.6). Moreover we see that we can compute L_1 in (2.6) as

(3.18) $$\begin{bmatrix} L^3 & L^5 \\ \emptyset & \emptyset \end{bmatrix}\begin{bmatrix} L^1 & \emptyset \\ L^2 & L^4 \end{bmatrix}^{-1} .$$

This process can now be repeated to obtain an LU-factorization for the next pivotal block U_2 ; for this one should realize that $\hat{C}_2$ can be found from C_2 after permuting the rows of G_1 in the same way as the rows of F_1 .

The crucial point in Variant I is the existence of a nonsingular U^1 . This existence is not assured in general. The next two variants deal with this problem.

VARIANT II. Let Q_1 be an orthogonal matrix such that

(3.19) $$S_1Q_1 = [U^1 \vdots \emptyset] ,$$

where U^1 is an $l \times l$ upper triangular matrix. As was shown in Mattheij [8] the matrix Q_1 can be found as a product of l elementary hermitians. This implies that multiplication of vectors with this matrix has a complexity of $O\left(\tfrac{1}{2}\left(n^2-l^2\right)\right)$ only. Next compute

(3.20) $$\hat{F}_1 = F_1Q_1 .$$

We can now proceed as in (3.14), (3.15), by computing an LU-decomposition of $\begin{bmatrix} U^1 & \emptyset \\ & \hat{F}_1 \end{bmatrix}$. We thus obtain

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

$$(3.21) \qquad \begin{bmatrix} S_1 \\ F_1 \end{bmatrix} = \left[\begin{array}{c|c} I_l & \emptyset \\ \hline L^2 & L^4 \\ L^3 & L^5 \end{array}\right] \begin{bmatrix} U^1 & \emptyset \\ \emptyset & U^4 \end{bmatrix} Q_1 .$$

If we allow for permutations in the rows of $\hat{F}_1$ we see that this factorization always exists.

Quite often the matrices $-G_i$ are identity matrices (*cf.* the matrix $\tilde{A}$). For simplicity we assume now that S_1 has a corresponding form; so let us assume (for Variant III)

$$(3.22) \qquad S_1 = [\emptyset \quad -I_l] \ , \quad G_i = -I_n .$$

We then obtain:

VARIANT III. Let P_1 be the permutation matrix

$$(3.23) \qquad P_1 = \begin{bmatrix} \emptyset & I_{n-l} \\ I_l & \emptyset \end{bmatrix} .$$

Then, identifying P_1 with Q_1 in (3.19), (3.20) we can find a factorization like in Variant II, giving

$$(3.24) \qquad \begin{bmatrix} S_1 \\ F_1 \end{bmatrix} = \left[\begin{array}{c|c} I_l & \emptyset \\ \hline L^2 & L^4 \\ L^3 & L^5 \end{array}\right] \begin{bmatrix} -I_l & \emptyset \\ \emptyset & U^4 \end{bmatrix} P_1 .$$

Note that $\begin{bmatrix} L^2 \\ L^3 \end{bmatrix}$ are just the first l columns of $-F_1P_1$. The next step (that is, in which we determine an LU-decomposition of U_2) is preceded by permutation of the second block column of $\mathbf{A}$ as follows. Let $\hat{P}_1$ be the permutation matrix arising from permuting the *rows* of F_1P_1 . Then P_1G_1 $(= \hat{P}_1)$ and F_2 are post-multiplied by $\hat{P}_1$ first. After this we

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

have a similar form for $\hat{C}_2$ as for S_1 .

In order now to give estimates for the L_i and U_i we use

LEMMA 3.25. *In* (3.16), (3.21) *and* (3.24) *the following estimates hold:*

$$\|[L^1]^{-1}\|_2 \le 2^{l-1} , \quad \|[L^4]^{-1}\|_2 \le 2^{n-l-1} , \quad \|L^5\|_2 \le \sqrt{l(n-l)} .$$

Proof. Since we used partial pivoting we know that the multipliers in L^1, L^4 and L^5 are bounded in modulus by 1 . Straightforward computation reveals that

$$|[L^1]^{-1}| \le \begin{bmatrix} 1 & & & & \\ 1 & \ddots & & \emptyset & \\ 2 & \ddots & \ddots & & \\ \vdots & & \ddots & \ddots & \\ 2^{l-2} & \cdots & 2 & 1 & 1 \end{bmatrix} ,$$

whence

$$\|[L^1]^{-1}\|_2 \le \left(\|[L^1]^{-1}\|_1 \|[L^1]^{-1}\|_\infty\right)^{\frac{1}{2}} = 2^{l-1} .$$

Likewise we find $\|[L^3]^{-1}\|_2 \le 2^{n-l-1}$. Trivially also $\|L^5\|_2 \le \sqrt{l(n-l)}$. □

THEOREM 3.26. *(i) If in Variant* I *the permutation matrices in the (restricted) pivoting are such that the block diagonal matrix* $\mathbb{D}\tilde{\mathbb{D}}^{-1}$ (*cf.* (2.3), (2.7)) *has the same block structure as* $\mathbb{E}$ *, that is, if for all* i *,* $K_i = 0$ *in* (2.10), *we can guarantee stability; we then have* (*cf. Theorem* 3.13), *with* $\gamma = 0$ *,*

$$\|L_2\|_2 \le \|F_1\|_2/\mathrm{glb}_2\left(S_i^T\right)(1+\gamma) + \gamma ,$$

$$\|L_i\|_2 \le \|F_{i-1}\|_2\left\|G_{i-2}^{-1}\right\|_2(1+\gamma) + \gamma , \quad 3 \le i \le N ,$$

$$\|U_1\|_2 \le \|F_1\|_2 + \|S_1\|_2 ,$$

$$\|U_i\|_2 \le \|F_i\|_2 + \|G_{i-1}\|_2(1+\gamma) , \quad 2 \le i \le N-1 ,$$

$$\|U_N\|_2 \le \|S_N\|_2 + \|G_{N-1}\|_2(1+\gamma) .$$

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

(ii) In Variant II *the estimates above hold with*
$\gamma = 2^{n-l-1}\sqrt{l(n-l)}$.

(iii) In Variant III *the estimates above hold with*
$\gamma = 2^{n-l-1}\sqrt{l(n-l)}$ *and* $\left\|G_i^{-1}\right\| = 1$.

Proof. *(i)* From Property 2.13 we see that

$$K_i = 0 \Rightarrow P_{i+1}^{12} = 0 \Rightarrow H_i = \begin{bmatrix} P_i^{22} & \emptyset \\ \emptyset & P_{i+1}^{11} \end{bmatrix} .$$

Since $E_i\tilde{E}_i^{-1}$ is an orthogonal matrix, we have $\left\|E_i\tilde{E}_i^{-1}\right\|_2 = 1$. Also, for all orthogonal matrices Z, Z^* we have $\|Z\cdot Z^*\|_2 = \|\cdot\|_2$. Now

$$H_iU_i = \begin{bmatrix} \emptyset & -P_i^{22} \\ P_{i+1}^{11}V_i^{11} & P_{i+1}^{11}V_i^{12} \end{bmatrix} \quad (i \geq 2) .$$

Apparently

$$\left\|\left[P_{i+1}^{11}V_i^{11} \mid P_{i+1}^{11}V_i^{12}\right]\right\|_2 \leq \|G_iQ_{i+1}V_i\|_2 = \|F_iQ_i\|_2 = \|F_i\|_2 ,$$

whereas $\left\|P_i^{22}\right\|_2 \leq \|G_{i-1}\|_2$. If $i = 1$ the upper right block in U^1 equals $-P_1^{22}\tilde{S}_1^{22}$, which is bounded in norm by $\|S_1\|_2$. The estimates for $\|U_i\|_2$ now follow immediately from Property 2.14. The bound for L_i follows from (2.10), namely, from the relation

$$L_i = P_iL_i\tilde{P}_{i-1}^{-1} \Rightarrow \|L_i\|_2 = \left\|(G_{i-1}Q_i)^{22}V_{i-1}^{22}Q_{i-1}G_{i-2}^{-1}\right\|_2$$
$$\leq \|F_{i-1}\|_2\left\|G_{i-2}^{-1}\right\|_2 \quad \text{for } i \geq 3$$

(for $i = 2$ similarly). If we drop the assumption that $K_i = 0$, then this bound for $\|L_i\|_2$ does not hold. As we remarked before, Variant I may not work, which in particular means that $\|L_i\|_2$ may be unbounded. Note

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

that $\left\|L_i^{21}\right\|_2 \le \gamma$ (see (3.8) and Lemma 3.25) but that L_i^{22} may not exist.

For Variants II and III we obtain (*cf.* Property 2.14)

$$H_i\tilde{U}_i = \begin{bmatrix} \emptyset & L_i^{21}P_i^{12}-P_i^{22} \\ P_{i+1}^{11}V_i^{11} & P_{i+1}^{12}V_i^{22}+P_{i+1}^{11}V^{12} \end{bmatrix}.$$

We now have $\left\|\begin{bmatrix} P_{i+1}^{11}V_i^{11} & P_i^{12}V_i^{22}+P_{i+1}V_i^{12} \end{bmatrix}\right\|_2 \le \|G_iQ_{i+1}V_i\|_2 = \|F_i\|_2$.

Moreover $\left\|P_i^{12}\right\|_2$, $\left\|P_i^{22}\right\|_2 \le \|G_{i-1}\|_2$, whilst $\left\|L_i^{21}\right\|_2$ can be estimated from (3.18), using Lemma 3.25. We find $\left\|L_2^{21}\right\|_2 = \|L^5[L^4]^{-1}\|_2 \le \gamma$; the same bound holds for $\left\|L_i^{21}\right\|_2$. Again using the fact that $\left\|E_i\tilde{E}_i^{-1}\right\|_2 = 1$ the bounds for $\|U_i\|_2$ now follow simply. We next estimate $\left\|L_2^{22}\right\|_2$. From (3.18) (with $L^1 = I_l$) we see that $L_2^{22} = L^3 - L^5[L^4]^{-1}L^2$. By construction we have $\|L^3U^1\|_2 \le \|F_1\|_2 \Rightarrow \|L^3\|_2 \le \|F_1\|_2/\mathrm{glb}_2(U^1)$ and likewise $\|L^2\|_2 \le \|F_1\|_2/\mathrm{glb}_2(U^1)$. Hence using Lemma 3.25 we obtain $\left\|L_2^{22}\right\|_2 \le \|F_1\|_2/\mathrm{glb}_2\left(S_1^T\right)[1+\gamma]$. A similar estimate follows for $\left\|L_i^{22}\right\|_2$ if we replace $\mathrm{glb}_2\left(S_1^T\right)$ by $\mathrm{glb}_2(G_{i-2})$ $\left(= \left\|G_{i-2}^{-1}\right\|_2^{-1}\right)$.

(iii) If, in particular for all i , $G_i = 1$, then we may replace $\left\|G_i^{-1}\right\|_2$ by 1 in the estimates above. □

REMARK 3.27. The term γ in Theorem 3.26 *(ii)*, *(iii)* is usually a severe overestimate. In fact 2^{n-l-1} is the familar upperbound for the growth factor encountered in a classical backward LU-decomposition analysis and is very likely to be a fairly small number; also the factor $\sqrt{l(n-l)}$ is a result of taking a worst case.

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

REMARK 3.28. Qualitatively the bounds in Theorem 3.26 show that $\|\mathbb{L}\|\|\mathbb{U}\| = O(\|\mathbb{A}\|)$. For the estimates of the $\|L_i\|$ are just moderate numbers, more or less independent of the scaling. The estimates of the $\|U_i\|$ (and trivially of the $\|\hat{C}_i\|$) are of the same order as the blocks in $\mathbb{A}$.

As we noted in Section 2 we might as well consider a forward error analysis necessitating us to find bounds for $\|L_i\|$ and $\left\|U_i^{-1}\right\|$ (or rather for the solutions of the forward and backward recursion). As was shown in [7] this can be done quite conveniently for the decomposition (2.7) with $\tilde{\mathbb{D}}$ and $\tilde{\mathbb{E}}$ as defined in (3.10). In particular, if the solution space of the underlying ordinary differential equation is dichotomic it follows that

$$\left\|V_i^{22}\right\|_2 \lesssim 1 \quad \text{and} \quad \left\|\left[V_i^{22}\right]^{-1}\right\|_2 , \ \left\|\left[V_i^{11}\right]^{-1} V_i^{12}\right\|_2 \lesssim 1 .$$

It follows then that $\|\mathbb{A}^{-1}\|_2$ mainly depends on the proper "choice" of S_1 and S_N (in particular $\left\|\left[\tilde{S}_1^{22}\right]^{-1}\right\|_2$ should not be large; *cf.* Theorem 3.13). Below we give a sort of reverse result of this: if $\|\mathbb{A}^{-1}\|_2$ is not large then we have a kind of dichotomy of the solution space. Theorem 3.29 can also be seen as an analogue to [2] for block tridiagonal matrices.

THEOREM 3.29. *Let* $\|S_1\|_2 = \|S_N\|_2 = 1$ *and, for all* i, $\|G_i\|_2 = 1$. *Define* $\kappa := \|\mathbb{A}^{-1}\|_2$. *Then there exists a fundamental solution* $\{\Phi_i\}_{i=1}^N$ *of* (3.2), *that is* $F_i\Phi_i = G_i\Phi_{i+1}$, $i = 1, \dots, N-1$, *such that*

(i) *for all* i, $\Phi_i = \big(\underbrace{\Phi_i^1}_{k} \mid \underbrace{\Phi_i^2}_{n-k}\big)$ *with, for all* i,

$$\left\|\Phi_i^1\right\|_2 / \mathrm{glb}_2\left(\Phi_N^1\right) \le \kappa \quad \textit{and, for all } i , \quad \left\|\Phi_i^2\right\|_2 / \mathrm{glb}_2\left(\Phi^2\right) \le \kappa ,$$

$$\text{(ii)} \quad \max_i \left\| \Phi_i \left[\begin{pmatrix} S_1 \\ \emptyset \end{pmatrix} \Phi_1 + \begin{pmatrix} \emptyset \\ S_N \end{pmatrix} \Phi_N \right]^{-1} \right\|_2 \le \kappa .$$

Before proving this, we would like to remark that Theorem 3.29 *(i)* shows a kind of dichotomy of the solution space; it, for example, follows that no solution in $\operatorname{span}(\Phi^2)$ can "increase faster" than κ ; that is,

$$\phi \in \operatorname{span} \Phi^2 \left(\Leftrightarrow \exists v \in R^{n-k} \text{ such that } \forall_i \phi_i = \Phi_i^2 v\right) \Rightarrow \forall_i \|\phi_i\|_2 / \|\phi_1\| \le \kappa .$$

The result in Theorem 3.29 *(ii)* precisely means that the condition number of the boundary value problem, as was defined in [9], cannot exceed κ .

Proof of Theorem 3.29. Consider the equation $\tilde{\mathbf{A}}\mathbf{Z} = \mathbf{W}$, where $\mathbf{Z}$ and $\mathbf{W}$ are block vectors ($nN \times n$ matrices) of which $\mathbf{W}$ is given by

$$\mathbf{W} = \left(\begin{pmatrix} \emptyset & \\ \emptyset & \underbrace{\emptyset}_{n-k} \end{pmatrix}, \begin{pmatrix} \emptyset & \emptyset \\ \emptyset & \emptyset \end{pmatrix}\right), \ldots, \left(\begin{pmatrix} \emptyset & \emptyset \\ \emptyset & \emptyset \end{pmatrix}, \begin{pmatrix} \emptyset & \emptyset \\ \underbrace{I}_{k} & \emptyset \end{pmatrix}\right)^T .$$

Since $\mathbf{A}$ is nonsingular it follows that the V_i^{11} and also $\tilde{S}_1^{22}$ and $\tilde{S}_N^{11}$ are nonsingular (*cf.* (3.11) and Theorem 3.13).

By substitution it can now be checked that

$$Z_i = \left(\begin{array}{c|c} \left[\prod_{j=i}^{N-1} V_j^{11}\right]^{-1} \left[S_N^{11}\right]^{-1} & \Omega_i^N \left[\prod_{j=1}^{i-1} V_j^{22}\right] \left[\tilde{S}_1^{22}\right]^{-1} + \Delta_i^N \\ \hline \emptyset & -\left[\prod_{j=1}^{i-1} V_j^{22}\right] \left[\tilde{S}_1^{22}\right]^{-1} \end{array}\right) ,$$

where a product like $\left[\prod_{j=1}^{i-1} V_j^{22}\right]$ has to be understood as $V_{i-1}^{22} \ldots V_1^{22}$, empty products are 1 and empty sums are 0 and where we have denoted for short

$$\Omega_i^N = \sum_{j=i}^{N-1} \left[\prod_{l=i}^{j} V_l^{11}\right]^{-1} V_j^{12} \left[\prod_{l=i}^{j-1} V_l^{22}\right]$$

and

$$\Delta_i^N = \left[\prod_{j=i}^{N-1} V_j^{11}\right]^{-1} \left[\tilde{S}_N^{11}\right]^{-1} \tilde{S}_N^{12} \left[\prod_{j=1}^{N-1} V_j^{22}\right] \left[\tilde{S}_1^{22}\right]^{-1} .$$

Apparently $\{Z_i\}_{i=1}^N$ satisfies $Z_{i+1} = V_i Z_i$, $i = 1, \ldots, N-1$. Since

$Z = \tilde{A}^{-1}W$, we obtain

$$\|Z_i\|_2 \le \|\tilde{A}^{-1}\|_2 \|W\|_2 = \kappa \ .$$

Consider now the (obvious) partitioning for Z_i :

$$Z_i = \begin{bmatrix} Z_i^{11} & Z_i^{12} \\ \emptyset & Z_i^{22} \end{bmatrix} \quad \left(Z_i^{11} \text{ of order } k \right).$$

We then have

$$\text{(a)} \quad \left\| \begin{pmatrix} Z_i^{11} \\ \emptyset \end{pmatrix} \right\|_2 \Big/ \mathrm{glb}_2\left(\begin{pmatrix} Z_N^{11} \\ \emptyset \end{pmatrix} \right) \le \|Z_i\|_2 / \mathrm{glb}_2\left(\left[\tilde{S}_N^{11}\right]^{-1} \right) \le \kappa \left\| \tilde{S}_N^{11} \right\|_2 \le \kappa \ .$$

We also have

$$\text{(b)} \quad \left\| \begin{pmatrix} Z_i^{12} \\ Z_i^{22} \end{pmatrix} \right\|_2 \Big/ \mathrm{glb}_2\left(\begin{pmatrix} Z_1^{12} \\ Z_1^{22} \end{pmatrix} \right) \le \|Z_i\|_2 / \mathrm{glb}_2\left(\left[\tilde{S}_1^{22}\right]^{-1} \right) \le \kappa \left\| \tilde{S}_1^{22} \right\|_2 \le \kappa \ .$$

If we define $\Phi_i := Q_i Z_i$, then $\{\Phi_i\}_{i=1}^N$ clearly is a fundamental solution of (3.2) and statement Theorem 3.29 *(i)* directly follows from the unitary invariance of $\|\cdot\|_2$ and glb_2 and (a) and (b) above. Moreover, since

$$\begin{pmatrix} \emptyset \\ \tilde{S}_1 \end{pmatrix} Z_1 + \begin{pmatrix} \tilde{S}_N \\ \emptyset \end{pmatrix} Z_N = I \ , \text{ that is } \begin{pmatrix} \emptyset \\ S_1 \end{pmatrix} \Phi_1 + \begin{pmatrix} S_N \\ \emptyset \end{pmatrix} \Phi_N = I \ ,$$

we see that

$$\left\| \Phi_i \left[\begin{pmatrix} S^1 \\ \emptyset \end{pmatrix} \Phi_1 + \begin{pmatrix} \emptyset \\ S_N \end{pmatrix} \Phi_N \right]^{-1} \right\|_2 = \left\| \Phi_i \left[\begin{pmatrix} \emptyset \\ S_1 \end{pmatrix} \Phi_1 + \begin{pmatrix} S_N \\ \emptyset \end{pmatrix} \Phi_N \right]^{-1} \right\|_2 \le \kappa \qquad \square$$

4. Pivot estimates for general block tridiagonal matrices

For some types of discretizations *a priori* knowledge about a possibly dichotomic solution space and also about the conditioning of the discrete

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

problem may not be available, so that the results of Section 3 may not be directly applicable. Also if the tridiagonal matrix cannot be linked to a one step system, as in (3.1) one must try to find other ways to establish stability. In this section we therefore give some estimation methods which complement results like the ones given in Varah [11, 12].

Inspired by (2.6) we shall use estimates based on the nonlinear recursion

$$u_i = p_i + \frac{q_i}{u_{i-1}} . \tag{4.1}$$

In [6] it was shown how one can obtain (often sharp) estimates for the u_i in terms of the fixed points of the functions ψ_i, defined by

$$\psi_i(s) := p_i + \frac{q_i}{s} . \tag{4.2}$$

Obviously ψ_i has two fixed points, say α_i, β_i with $|\alpha_i| \geq |\beta_i|$. We have the following estimation results.

THEOREM 4.3. *For all* i *let* $p_i > 0 > q_i$ *and* $p_i^2 \geq -4q_i$ (*that is,* $\alpha_i \geq \beta_i > 0$). *Assume* $\min_j \alpha_j > \max_j \beta_j$. *If* $u_l \geq \alpha_{l+1}$ *for some* l, *then* $u_i \geq \min_{i \geq j \geq l+1} \alpha_j$ *for all* $i \geq l$.

Proof. Induction. One may, for example, use the relation $u_i = \alpha_i + \beta_i - \alpha_i\beta_i/u_{i-1}$. If $u_{i-1} \geq \alpha_i$, then $u_i \geq \alpha_i$ and if $u_{i-1} \leq \alpha_i$, then $u_i \geq u_{i-1}$ (*cf.* also [6]). □

COROLLARY 4.4. *For all* i *let* $p_i \geq 1 - q_i$ (*that is* $\alpha_i \geq 1$). *If* $u_l \geq 1$ *for some* l, *then* $u_i \geq 1$ *for all* $i \geq l$.

COROLLARY 4.5. *For all* i *let* $p_i = 1$, $-\frac{1}{4} \leq q_i < 0$ (*that is* $\alpha_i = \frac{1}{2}(1+\sqrt{1+4q_i})$). *If* $u_l \geq 1$ *for some* l, *then* $u_i \geq \min_{i \geq j \geq l+1} \frac{1}{2}(1+\sqrt{1+4q_j}) \geq \frac{1}{2}$.

It is fairly straightforward that Theorem 4.3, Corollaries 4.4 and 4.5 still hold if instead of (4.1) we have the *inequality*

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

$$(4.6) \qquad u_i \geq p_i + \frac{q_i}{u_{i-1}} .$$

Upperbounds for $\left\|U_i^{-1}\right\|$, that is, lower bounds for $\mathrm{glb}(U_i)$ may then be found by substituting $u_i := \mathrm{glb}(U_i)$, $u_i := \mathrm{glb}\left(\hat{B}_i^{-1}U_i\right)$ and $u_i := \mathrm{glb}\left(U_i B_{i-1}^{-1}\right)$ respectively in the inequalities given in

PROPERTY 4.7. *(a)* $\mathrm{glb}(U_i) \geq \mathrm{glb}(\hat{B}_i) - \|\hat{A}_i\|\|\hat{C}_{i-1}\|/\mathrm{glb}(U_{i-1})$.

(b) $\mathrm{glb}\left(\hat{B}_i^{-1}U_i\right) \geq 1 - \left\|\hat{B}_i^{-1}A_i\right\|\left\|\hat{B}_{i-1}^{-1}\hat{C}_{i-1}\right\|/\mathrm{glb}(\hat{B}_{i-1}U_{i-1})$.

(c) $\mathrm{glb}\left(U_i\hat{B}_i^{-1}\right) \geq 1 - \left\|\hat{A}_i\hat{B}_{i-1}^{-1}\right\|\left\|\hat{C}_{i-1}\hat{B}_i^{-1}\right\|/\mathrm{glb}\left(U_{i-1}\hat{B}_{i-1}^{-1}\right)$.

Proof. Property 4.7 *(a)* follows directly from (2.6) (b). From (2.6) (b) we also derive

$$\hat{B}_i^{-1}U_i = I - \hat{B}_i^{-1}\hat{A}_i\left[\hat{B}_{i-1}^{-1}U_{i-1}\right]^{-1}\hat{B}_{i-1}^{-1}\hat{C}_{i-1} .$$

Taking norms and using appropriate inequalities gives Property 4.7 *(b)*. Similarly Property 4.7 *(c)* follows from estimating $U_iB_{i-1}^{-1}$. □

EXAMPLE 4.8. Let $\left\|\hat{B}_i^{-1}A_i\right\|\left\|\hat{B}_{i-1}^{-1}\hat{C}_{i-1}\right\| \leq \frac{1}{4}$, for $2 \leq i \leq N$. Since $\mathrm{glb}\left(\hat{B}_1^{-1}U_1\right) = \mathrm{glb}(I) = 1$, we deduce from Corollary 4.5, (4.6) and Property 4.7 *(b)* that $\mathrm{glb}\left(\hat{B}_i^{-1}U_i\right) \geq \frac{1}{2}$, whence $\mathrm{glb}(U_i) \geq \frac{1}{2}\,\mathrm{glb}(\hat{B}_i)$, or equivalently $\left\|U_i^{-1}\right\| \leq 2\left\|\hat{B}_i^{-1}\right\|$. This result is similar to Varah [*11*, Theorem 2.2]. The estimates in Property 4.7 all are fairly rough. With some extra effort we can sometimes give sharper results if we have a block tridiagonal matrix arising from a one step recursion. Consider the recursion (*cf.* (2.6) (b))

$$(4.9) \qquad U_i\hat{B}_i^{-1} = I - \hat{A}_i\hat{B}_{i-1}^{-1}\left[U_{i-1}\hat{B}_{i-1}^{-1}\right]^{-1}\hat{C}_{i-1}\hat{B}_i^{-1} .$$

If the first $n - k$ rows of $\hat{C}_{i-1}$ are zero then also those of $\hat{C}_{i-1}\hat{B}_i^{-1}$.

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

Similarly we see that the last k rows of $\hat{A}_i\hat{B}_{i-1}^{-1}$ are zero. Denote

$$(4.10)\quad \begin{pmatrix} \underbrace{a_i^{11}}_{n-k} & a_i^{12} \\ \emptyset & \emptyset \end{pmatrix} := \hat{A}_i\hat{B}_{i-1}^{-1}\ , \quad \begin{pmatrix} \emptyset & \emptyset \\ \underbrace{c_{i-1}^{21}}_{n-k} & c_{i-1}^{22} \end{pmatrix} := \hat{C}_{i-1}\hat{B}_i^{-1}\ , \quad \tilde{U}_i := U_i\hat{B}_i^{-1}\ .$$

We obtain then (*cf.* Varah [12, Lemma 3.1])

PROPERTY 4.11. *Assume* $\{\tilde{U}_i^{-1}\}$ *exists. Then* $\tilde{U}_i$ *is a block upper triangular matrix of the form*

$$\tilde{U}_i = \begin{pmatrix} d_i & e_i \\ \emptyset & \underbrace{I}_{k} \end{pmatrix}\ ,$$

where $\{d_i\}_{i=1}^N$ *and* $\{e_i\}_{i=1}^N$ *satisfy the recursions*

$$d_i = I + \left[a_i^{11}\left[d_{i-1}\right]^{-1}e_{i-1}-a_i^{12}\right]c_{i-1}^{21}\ , \quad d_1 = I\ ,$$

$$e_i = \left[a_i^{11}\left[d_{i-1}\right]^{-1}e_{i-1}-a_i^{12}\right]c_{i-1}\ , \quad e_1 = 0\ .$$

The proof follows from a simple induction argument $\left(\text{note } \tilde{U}_1 = I\right)$.

COROLLARY 4.12. *If, for all* i , c_i^{21} *is nonsingular then* $\operatorname{glb}\left(d_i\right)$ *can be estimated from below using*

$$\operatorname{glb}\left(d_2\right) = \operatorname{glb}\left(I-a_2^{12}c_1^{21}\right)\ ,$$

$$\operatorname{glb}\left(d_i\right) \ge \operatorname{glb}\left(I-a_i^{12}c_{i-1}^{21}+a_i^{11}\left[c_{i-2}^{21}\right]^{-1}c_{i-2}^{22}c_{i-1}^{21}\right) - \left\|a_i^{11}\right\|\left\|\left[c_{i-2}^{21}\right]^{-1}c_{i-2}^{22}c_{i-1}^{21}\right\|/\operatorname{glb}\left(d_{i-1}\right)\ , \quad i \ge 3\ .$$

Consequently, $\|e_i\|$ *can be estimated from above using*

$$\|e_2\| = \left\|a_2^{12}c_1^{22}\right\|\ ,$$

$$\|e_i\| \le \left\{\left\|a_i^{11}\right\|\left\|c_{i-1}^{22}\right\|/\operatorname{glb}\left(d_{i-1}\right)\right\}\|e_{i-1}\| + \left\|a_i^{12}c_{i-1}^{22}\right\|\ , \quad i \ge 3\ .$$

Once, for example, $\left\{\operatorname{glb}_\infty\left(d_i\right)\right\}$ *and* $\left\{\|e_i\|_\infty\right\}$ *are estimated, an estimate* *for* $\|\tilde{U}_i^{-1}\|_\infty$ *is given by*

$$\|\tilde{U}_i^{-1}\|_\infty \leq \max\left(\frac{1+\|e_i\|_\infty}{\mathrm{glb}_\infty(d_i)}, 1\right) .$$

Proof. If c_i^{21} is nonsingular, then $e_i = (d_i - I)\left[c_{i-1}^{21}\right]^{-1} c_{i-1}^{22}$. □

REMARK 4.13. Apparently we need $n - k = k$ in Corollary 4.12. This corresponds to a boundary value problem where at each boundary point the same number of relations for the solution are specified. We can also give a corollary like Corollary 4.12, now assuming c_i^{22} is nonsingular (note c_i^{22} is a square matrix). We then obtain a nonlinear recursive inequality for $\|e_i\|$ which then can be used to estimate $\mathrm{glb}(d_i)$ and so on.

The $\mathrm{glb}(d_i)$ values may now be estimated using the method outlined in the beginning of this section. In the next section we give an example of this.

5. Examples

In this section we give some examples where we shall estimate the pivots.

EXAMPLE 5.1. Consider the selfadjoint ordinary differential equation

(5.2) (a) $$(r(t)y')' - s(t)y = f(t) , 0 \leq t \leq 1 ,$$

with $r > 0$ and $s \geq 0$. Let the boundary conditions be given by

(5.3) $$\alpha y(0) + \beta y'(0) = g_0 , \quad \gamma y(1) + \delta y'(1) = g_1 .$$

Suppose we want to solve this problem numerically by using the midpoint rule applied to the system

(5.2) (b) $$y' = z/r , \qquad z' = sy + f .$$

The recurrence relation then has the form (*cf.* [*12*, (2.2)])

$$(5.4)\quad \begin{pmatrix} 1 & -h/2r_j \\ -hs_j/2 & 1 \end{pmatrix}\begin{pmatrix} y_j \\ z_j \end{pmatrix} = \begin{pmatrix} 1 & h/2r_j \\ hs_j/2 & 1 \end{pmatrix}\begin{pmatrix} y_{j-1} \\ z_{j-1} \end{pmatrix} + \begin{pmatrix} 0 \\ hf\left((j-\frac{3}{2})h\right) \end{pmatrix} ,$$

where y_j and z_j are approximations for $y\left((j-1)h\right)$ and $z\left((j-1)h\right)$ respectively, r_j and s_j denote $r\left((j-\frac{3}{2})h\right)$ and $s\left((j-\frac{3}{2})h\right)$ respectively and $h = 1/(N-1)$ for some N. We thus obtain a problem like (2.1) where the unknowns are $y_1, z_1, \ldots, y_N, z_N$:

$$(5.5)\quad \mathbf{A} = \left[\begin{array}{cc|cc|ccc} \alpha & \beta/r(0) & 0 & 0 & & & \\ 1 & h/\left(2r_2\right) & -1 & h/\left(2r_2\right) & & & \\ \hline hs_2/2 & 1 & hs_2/2 & -1 & & & \\ 0 & 0 & 1 & h/\left(2r_3\right) & & & \\ \hline & & & & \ddots & & \\ & & & & & hs_N/2 & -1 \\ & & & & & \gamma & \delta/r(1) \end{array}\right]$$

Identifying $\mathbf{A}$ with $\hat{\mathbf{A}}$, we then find

$$A_2B_1^{-1} = \frac{1}{\left(\alpha h/2r_1\right)-\left(\beta/r(0)\right)}\left(\begin{array}{c|c} \frac{h^2}{4}\frac{s_1}{r_1} - 1 & \frac{-\beta hs_1}{2r(0)} + \alpha \\ \hline 0 & 0 \end{array}\right) ,$$

$$A_iB_{i-1}^{-1} = \frac{1}{\left(h^2/4\right)\left(s_{i-2}/s_{i-1}\right)+1}\left(\begin{array}{c|c} \frac{h^2}{4}\frac{s_{i-1}}{r_{i-1}} - 1 & \frac{h}{2}\left(s_{i-1}+s_{i-2}\right) \\ \hline 0 & 0 \end{array}\right) , \quad 3 \le i \le N-1 ,$$

$$C_{i-1}B_i^{-1} = \frac{1}{\left(h^2/4\right)\left(s_{i-1}/r_i\right)+1}\left(\begin{array}{c|c} 0 & 0 \\ \hline \frac{-h}{2}\left(\frac{1}{r_{i-1}} + \frac{1}{r_i}\right) & \frac{h^2}{4}\frac{s_{i-1}}{r_{i-1}} - 1 \end{array}\right) , \quad 2 \le i \le N .$$

Now let h be such that

$$(5.6)\qquad h^2 \le 4 \min \frac{r_i}{s_i} .$$

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

Then in the inequality $\mathrm{glb}(d_i) \geq p_i + q_i/\mathrm{glb}(d_{i-1})$ for $i \geq 3$, where p_i and q_i are as in Corollary 4.12, we have $p_i = 1 + m_i + n_i$, $q_i = -n_i$, for some positive m_i and n_i; that is, $p_i \geq 1 - q_i$ for $i \geq 3$. Using Corollary 4.4 this then gives

$$(5.7)\ (a) \qquad \mathrm{glb}(d_i) \geq 1\ , \quad i \geq 3\ .$$

Also from Corollary 4.12 we obtain

$$(5.7)\ (b) \qquad \mathrm{glb}(d_2) \geq 1 + \frac{\alpha-\beta(hs_1/2r(0))}{(\alpha h/2r_1)-(\beta/r(0))}\ \frac{(h/2)\left((1/r_1)+(1/r_2)\right)}{(h^2/4)(s_1/r_2)+1}\ .$$

Hence

$$(5.8) \qquad \beta/\alpha \leq 0 \Rightarrow \mathrm{glb}(d_2) > 1\ .$$

Consequently if h is sufficiently small and $\beta/\alpha \leq 0$, we see that, for all i, $\mathrm{glb}(d_i) \geq 1$ (note $\mathrm{glb}(d_1) = 1$). It is also fairly simple to see that (*cf.* Corollary 4.12)

$$(5.9) \qquad |e_i| \leq |e_{i-1}| + \frac{h}{2}\left(s_{i-1}+s_{i-2}\right)\ , \quad i \geq 3\ .$$

So

$$(5.10) \qquad |e_i| \leq \frac{\alpha-\beta(hs_1/2r(0))}{(\alpha h/2r_1)-(\beta/r(0))} + ih \max_{j\leq i} s_j\ ;$$

and as a consequence we find the following estimate for the pivot U_i:

$$(5.11) \qquad \left\|U_i^{-1}\right\|_\infty \leq \frac{\max\left(1+h/(2r_i),1+hs_{i-1}/2\right)}{1+h^2 s_{i-1}/(4r_i)} \max\left(1+|e_i|,\ 1\right)\ ;$$

(note $\left\|U_i^{-1}\right\|_\infty \leq \left(\left\|U_i^{-1}B_i\right\|_\infty \cdot \left\|B_i^{-1}\right\|_\infty\right)$).

EXAMPLE 5.12. Consider the same problem and the same scheme as in Example 5.1. We would now like to investigate stability by using properties of the one step recursion. Denote for short

$$(5.13) \qquad \xi_i := \frac{h}{2}\sqrt{s_i/r_i}\ .$$

Then

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

$$(5.14) \qquad \begin{pmatrix} y_i \\ z_i \end{pmatrix} = \frac{1}{1-\xi_i^2} \begin{pmatrix} 1+\xi_i^2 & \frac{h}{r_i} \\ hs_i & 1+\xi_i^2 \end{pmatrix} \begin{pmatrix} y_{i-1} \\ z_{i-1} \end{pmatrix} + f_i \quad (\text{for some } f_i).$$

Assume $s > 0$. Then we obtain, for the eigenvalues λ_i, μ_i $(\lambda_i > \mu_i)$ of the matrix in (5.14),

$$(5.15) \qquad \lambda_i = \frac{1+\xi_i}{1-\xi_i}, \quad \mu_i = \frac{1-\xi_i}{1+\xi_i}.$$

The corresponding eigenvectors, g_i and h_i respectively, are

$$(5.16) \qquad g_i = \left(1, \sqrt{r_i s_i}\right)^T, \quad h_i = \left(1, -\sqrt{r_i s_i}\right)^T.$$

From Mattheij [6] it then follows that there exists a homogeneous solution to (5.14), $\phi^1 := \left\{\phi_i^1\right\}_{i=1}^N$ say, with $\phi_1^1 = g_1$ and

$$(5.17) \qquad \left\|\phi_i^1\right\| / \left\|\phi_j^1\right\| \le \sigma \prod_{l=i}^{j-1} \lambda_l^{-1}, \text{ for } j > i, \text{ where } \sigma = O(1).$$

Similarly there exists a solution $\phi^2 := \left\{\phi_i^2\right\}_{i=1}^N$ say, with $\phi_N^2 = h_N$ and

$$(5.18) \qquad \left\|\phi_i^2\right\| / \left\|\phi_j^2\right\| \le \sigma \prod_{l=i+1}^{j} \lambda_l^{-1}, \text{ for } j < i.$$

We may as well normalize these ϕ^1 and ϕ^2 such that for $\Phi := \left(\phi^1 \mid \phi^2\right)$ we have $\max_i \|\Phi_i\| = 1$. This gives a factor σ' rather than σ in the estimates (5.17). Suppose we use the max norm, we than obtain

$$(5.19) \qquad \begin{aligned} &(a) \quad \left\|\phi_1^1\right\| \le \varepsilon := \sigma' \exp\left[-\min\left(\lambda_i^{-1}\right)\right] \approx O\left(\exp\left(-\min\sqrt{s_i/r_i}\right)\right), \\ &\qquad \left\|\phi_N^1\right\| = 1; \\ &(b) \quad \left\|\phi_1^2\right\| = 1, \quad \left\|\phi_N^2\right\| \le \varepsilon. \end{aligned}$$

The estimates in (5.19) can now be used to find a bound for the condition number of the boundary value problem $\left(cf.\ (3.16)\right)$. To this end define

$$(5.20)\qquad R := \begin{pmatrix} \alpha & \overline{\beta} \\ 0 & 0 \end{pmatrix}\Phi_1 + \begin{pmatrix} 0 & 0 \\ \gamma & \overline{\delta} \end{pmatrix}\Phi_N = \begin{pmatrix} O(\varepsilon) & \zeta \\ \eta & O(\varepsilon) \end{pmatrix},$$

where $\overline{\beta} := \beta/r(0)$, $\overline{\delta} := \delta/r(1)$ and $\xi := (\alpha, \overline{\beta})\phi_1^2$ and $\eta := (\gamma, \overline{\delta})\phi_N^2$.

The condition number is then given by

$$(5.21)\qquad \kappa := \max_i \left\|\Phi_i R^{-1}\right\|_\infty \le \left\|\begin{pmatrix} 1/\eta & \varepsilon/\zeta\eta \\ \varepsilon/\zeta\eta & 1/\zeta \end{pmatrix}\right\|_\infty .$$

It is not restrictive to assume that $\max(|\alpha|, |\overline{\beta}|) = \max(|\gamma|, |\overline{\delta}|) = 1$. We then see

$$(5.22)\qquad \beta/\alpha \le 0 \text{ and } \delta/\gamma \ge 0 \Rightarrow |\eta|,\ |\zeta| \approx 1 .$$

In this case therefore $\kappa = O(1)$. From Mattheij [9] it then follows that $\|A^{-1}\|_\infty = O(N)$. $\|U^{-1}\| = \|A^{-1}L\|$ gives $\left\|U_i^{-1}\right\|_\infty = O(N)$. In fact even $\left\|U_i^{-1}\right\|_\infty = O(1)$ holds from what has been remarked on page 192.

EXAMPLE 5.23. In Varah [12] the problem in the previous example was considered for the case $r = 1$, $s = 0$. We then do not have a significantly dichotomic solution space. A fundamental solution is given by

$$(5.24)\qquad \Phi_i = \begin{pmatrix} 1 & hi \\ 0 & 1 \end{pmatrix}.$$

Again, if $\beta/\alpha \le 0$ we can show that we have a well conditioned problem and thus prove stability, as was also established in Varah [12]. It is not difficult to see that the bound for $\left\|U_i^{-1}\right\|_\infty$, as above, namely $\left\|U_i^{-1}\right\|_\infty \approx N$ is realistic now.

EXAMPLE 5.25. Consider the Laplace equation on a rectangular two dimensional region

$$(5.26)\qquad \Delta u = f,\quad u = u(x, y),$$

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

and let Dirichlet boundary conditions be given on three consecutive sides and Neumann boundary conditions on the other side. If we use central differences for discretization and the usual ("reading order") ordering of grid points, we obtain a matrix A which typically has the form

$$(5.2\ )\qquad \mathbf{A} = \begin{bmatrix} T & 2I & & & \emptyset \\ I & T & I & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & T & I \\ \emptyset & & & I & T \end{bmatrix}, \quad T = \begin{bmatrix} -4 & 1 & & & \emptyset \\ 1 & -4 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & -4 & 1 \\ \emptyset & & & 1 & -4 \end{bmatrix}$$

For matrices like $\mathbf{A}$ in (5.27) it is sometimes important to know whether there exists an LU-decomposition, not for computational reasons (as there exist much more efficient iterative methods) but for theoretical purposes; in particular we can then give a bound for $\|\mathbf{A}^{-1}\|$, if we can bound the $\|L_i\|$ and $\|U_i^{-1}\|$. To find a bound for $\|U_i^{-1}\|$ we use Property 4.7 *(a)*. We obtain

$$(5.28)\qquad \mathrm{glb}_2(T) \geq 2\ .$$

Hence in particular

$$(5.29)\qquad \mathrm{glb}_2(U_1) = \mathrm{glb}_2(T) \geq 2\ ,\quad \mathrm{glb}_2(U_2) \geq 2 - 2/2 = 1\ .$$

Since

$$(5.30)\qquad \mathrm{glb}_2(U_i) \geq 2 - \frac{1}{\mathrm{glb}_2(U_{i-1})}$$

we find from Corollary 4.4 that, for all i,

$$(5.31)\qquad \left\|U_i^{-1}\right\|_2 \leq 1\ .$$

A similar result follows by applying Theorem 2.1 in Varah [*11*].

References

[*1*] U. Ascher, U.S. Pruess and R.D. Russell, "On spline basis selection for solving differential equations", *SIAM J. Numer. Anal.* 20 (1983), 121-142.

https://doi.org/10.1017/S0004972700021432 Published online by Cambridge University Press

[2] Carl de Boor, "Dichotomies for band matrices", *SIAM J. Numer. Anal.* **17** (1980), 894-907.

[3] Carl de Boor and Richard Weiss, "SOLVEBLOK: a package for solving almost block diagonal linear systems", *ACM Trans. Math. Software* **6** (1980), 80-87.

[4] Herbert B. Keller, "Accurate difference methods for nonlinear two-point boundary value problems", *SIAM J. Numer. Anal.* **11** (1974), 305-320.

[5] Herbert B. Keller, *Numerical solution of two point boundary value problems* (Regional Conference Series in Applied Mathematics, **24**. Society of Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1976).

[6] R.M.M. Mattheij, "Accurate estimates of solutions of second order recursions", *Lin. Algebra Appl.* **12** (1975), 29-54.

[7] R.M.M. Mattheij, "Characterizations of dominant and dominated solutions of linear recursions", *Numer. Math.* **35** (1980), 421-442.

[8] R.M.M. Mattheij, Stable computation of solutions of unstable linear initial value recursions", *BIT* **22** (1982), 79-93.

[9] R.M.M. Mattheij, "The conditioning of linear boundary value problems", *SIAM J. Numer. Anal.* **19** (1982), 963-978.

[10] Robert D. Russell, "A comparison of collocation and finite differences for two-point boundary value problems", *SIAM J. Numer. Anal.* **14** (1977), 19-39.

[11] J.M. Varah, "On the solution of block-tridiagonal systems arising from certain finite-difference equations", *Math. Comp.* **26** (1972), 859-868.

[12] James M. Varah, "A comparison of some numerical methods for two-point boundary value problems", *Math. Comp.* **28** (1974), 743-755.

[13] J.M. Varah, "Alternate row and column elimination for solving certain linear systems", *SIAM J. Numer. Anal.* **13** (1976), 71-75.

[14] J.H. Wilkinson, *The algebraic eigenvalue problem* (Clarendon Press, Oxford, 1965).

Mathematisch Instituut,
Katholieke Universiteit,
Nijmegen,
The Netherlands;

Department of Mathematical Sciences,
Rensselaer Polytechnic Institute,
Troy,
New York 12181,
USA.