On the Conditioning Analysis of Block Approximate Factorization Methods

Monga-Made Magolu* and Yvan Notay†
Service de Métrologie Nucléaire (C.P. 165)
Université Libre de Bruxelles
50, av. F.D. Roosevelt
1050 Bruxelles, Belgium

Dedicated to Gene Golub, Richard Varga, and David Young

Submitted by Owe Axelsson

ABSTRACT

The paper is devoted to the conditioning analysis of modified block incomplete factorizations of a given Stieltjes matrix. We obtain new results, improve other theories, and compare all existing upper bounds through numerical experiments. Applied to discrete elliptic PDEs, our results show that an $O(h^{-1})$ spectral bound can be achieved for a large class of problems.

1. INTRODUCTION

Concerning the solution of large sparse positive definite linear systems derived from discrete elliptic PDEs, much current research focuses on polynomially accelerated preconditioned iterative methods. In this connection, the incomplete block factorizations of the system matrix form a class of preconditioners whose robustness is familiar. Among these factorizations, "modified" versions based on some generalized row sum criterion (see Section 2) are seen to be particularly efficient (see e.g. [5], [6], and [11] for

*Research supported by the A.B.O.S. (A.G.C.D.) under project 11, within the cooperation between Belgium and Zaire.
†Research supported by the "Fonds National de la Recherche Scientifique" (Belgium)—Aspirant.

655 Avenue of the Americas, New York, NY 10010 0024-3795/91/$3.50

numerical evidence). Their conditioning analysis is however poorly developed.

Indeed, until recently, the only available works were those—nice but limited—by Beauwens and Ben Bouzid [8] and Axelsson and Eijkhout [4]. In more recent papers, upper spectral bounds improving the results of [8] have been obtained by both present authors [12, 14], but no comparative study of all these results exists within the framework of the block methods which we shall investigate here. New upper spectral bounds are presented and compared with those deduced from the former approaches.

The paper is organized as follows: in Section 2 some relevant features of the modified block incomplete factorizations we shall focus our attention on are presented, Section 3 is devoted to upper spectral bounds, and Section 4 to numerical results.

General Terminology and Notation

All vectors belong to C^n, the n-dimensional space with scalar product denoted (x, y); all matrices are $n \times n$ real matrices.

The symbols A^t and A^+ denote, respectively, the transpose and the Moore-Penrose inverse [10] of the matrix A.

The order relation between real matrices and vectors is the usual componentwise order: if $A = (a_{ij})$ and $B = (b_{ij})$, then $A \leqslant B$ $(A < B)$ if $a_{ij} \leqslant b_{ij}$ $(a_{ij} < b_{ij})$ for all i, j; A is called nonnegative (positive) if $A \geqslant 0$ $(A > 0)$. If $A = (a_{ij})$, we denote by diag(A) the (diagonal) matrix whose entries are $a_{ii}\delta_{ij}$, and we let offdiag$(A) = A -$ diag(A). Similarly, tridiag(A) denotes the tridiagonal matrix whose tridiagonal part consists of the three main diagonals of A. By e we denote the vector with all components equal to unity; by a $(0,1)$ matrix we understand a matrix whose nonzero entries are equal to unity.

Hadamard Multiplication

We recall that the Hadamard product $A * B$ of the matrices A and B of the same dimensions, with scalar entries a_{ij} and b_{ij}, is the element by element multiplication, i.e. with $(A * B)_{ij} = a_{ij} b_{ij}$, and that the unit matrix with respect to Hadamard multiplication, denoted ε, is the matrix whose entries are all equal to unity.

Standard LU Factorization

By the standard point LU factorization of a (Stieltjes) matrix A, we understand the factorization $A = LP^{-1}U$ such that U is upper triangular, $P = \text{diag}(U)$, and $L = U^t$.

Partitionings

Any partitioning of an n-vector $x = (x_I)$ into block components x_I of dimensions n_I, $I = 1, 2, \ldots, N$ (with $\sum_{I=1}^{N} n_I = n$), is uniquely determined by a partitioning $\pi = (\pi_I)_{1 \leqslant I \leqslant N}$ of the set $[1, n]$ of the first n integers. We shall assume throughout the paper that all n-vectors are partitioned in blocks according to a given such partitioning. The same partitioning π induces also a partitioning of any $n \times n$ matrix A into block components A_{IJ} of dimensions $n_I \times n_J$, and we shall similarly assume that all $n \times n$ matrices are partitioned in this way.

Lowercase indices refer to scalar entries, and capital indices to block entries. Thus scalar (block) entries of an $n \times n$ π-partitioned matrix A are denoted a_{ij} (A_{IJ}). When needed, scalar entries of block entries of A are denoted $(A_{IJ})_{ij}$, a notation which implies that $i \in \pi_I$ and $j \in \pi_J$. Similar notations are used for vector components, except that we always represent vectors by small letters.

A matrix which is block diagonal (triangular) relative to a π-partitioning will be referred to as π-diagonal (π-triangular). In order to avoid confusion, we also write sometimes diag$_p(A)$ for diag(A) and offdiag$_p(A)$ for offdiag(A), the subindex p stressing that these notions refer to the point partitioning.

2. MODIFIED INCOMPLETE BLOCK FACTORIZATIONS

Let $A = D - E - F$ be a Stieltjes matrix such that D is π-diagonal, F is strictly π-upper triangular, and $E = F^t$. For simplicity, we shall only consider factorizations with no fill-in allowed outside the block diagonal, i.e., approximate factorizations of the form

$$B = (P - E) P^{-1} (P - F), \tag{2.1}$$

where P denotes a π-diagonal Stieltjes matrix whose entries are computed from the relation

$$P = D - \beta * (EKF) - \Omega, \tag{2.2}$$

where β stands for some selected symmetric π-diagonal $(0,1)$ matrix, K represents a nonnegative π-diagonal symmetric matrix which is an approximate inverse to P, and Ω is the diagonal (pointwise) matrix determined by

$$\Omega x = \left[EP^{-1}F - \beta * (EKF) \right] x, \tag{2.3}$$

x being a positive vector such that $Ax \geqslant 0$. Therefore,

$$Bx = Ax, \tag{2.4}$$

which means that we satisfy a generalized row sum criterion.

Regarding the determination of K, various techniques have been proposed in the literature; a summary may be found in Section 3 of [7]. A widespread choice that has proved its robustness consists in choosing K as a principal band portion (often diagonal, tridiagonal, or pentadiagonal) of P^{-1}, in which case one has $0 \leqslant K \leqslant P^{-1}$, whence

$$\text{offdiag}_p(A - B) \leqslant 0. \tag{2.5}$$

The latter inequality together with (2.4) leads, for the lowest eigenvalue of $B^{-1}A$, to

$$v_{\min} = 1, \tag{2.6}$$

so that its spectral condition number $\kappa(B^{-1}A)$ coincides with its largest eigenvalue.

From an implementation point of view, it is worthwhile to mention that relations (2.2) and (2.3) may be expanded in the more detailed form

$$P_{11} = D_{11},$$

$$P_{II} = D_{II} - \beta_{II} * \left(\sum_{s=1}^{I-1} E_{IS} K_{SS} F_{SI} \right) - \Omega_{II} \tag{2.7}$$

$$\Omega_{II} x_I = \left[\sum_{s=1}^{I-1} E_{IS} P_{SS}^{-1} F_{SI} - \beta_{II} * \left(\sum_{s=1}^{I-1} E_{IS} K_{SS} F_{SI} \right) \right] x_I \qquad \text{for} \quad 2 \leqslant I \leqslant N.$$

As usual, the presence of P_{SS}^{-1} has to be understood as the solution of a linear system with P_{SS} as matrix coefficient.

From a practical point of view, carrying out the algorithm (2.7) should be guaranteed, which amounts to proving the nonsingularity of P_{II} for $1 \leqslant I \leqslant N-1$. By Theorem 2.1 of [14] the following conditions are sufficient: For $I = 1, 2, \ldots, N-1$,

(1) D_{II} is irreducible;
(2) there exists some J, $I < J \leqslant N$, such that $F_{IJ} \neq 0$.

Further, one easily deduces from $Bx = Ax$ that

$$((P - F)x)_I = (Ax)_I + \sum_{s=1}^{I-1} E_{IS} P_{SS}^{-1} ((P - F)x)_S \qquad \text{for} \quad 1 \leqslant I \leqslant N; \tag{2.8}$$

therefore (readily by induction) $(P - F)x \geqslant 0$; and finally, assuming that D_{NN} is also irreducible, one has that P_{NN} is also nonsingular unless $Ax = 0$, i.e. A is singular, which is not our concern here. Other existence criteria may be found in Axelsson [1], Axelsson and Polman [6], Beauwens and Ben Bouzid [8], and Concus, Golub, and Meurant [11].

3. UPPER EIGENVALUE BOUNDS

In this section we shall concentrate to a major extent on upper eigenvalue bounds for the pencil of matrices $A - vB$, where A is positive definite and B is of the form $(P - E)P^{-1}(P - F)$ with P positive definite, F π-upper triangular, and $E = F^t$. For completeness and for the purpose of comparison, in addition to the statement of new upper bounds, we also include results that have been obtained elsewhere.

The first result we mention follows from [12, Theorem 3.1], which represents an improved version of a similar result by Beauwens and Ben Bouzid [8, Theorem 4.2].

THEOREM 3.1. *Let $A = D - E_0 - F_0$ be a symmetric positive definite matrix such that D is π-diagonal and symmetric, F_0 is strictly π-upper triangular, and $E_0 = F_0^t$. Let P be a π-diagonal Stieltjes matrix such that $\text{offdiag}_p(P) \leqslant \text{offdiag}_p(D)$, and let $P = L_p P_p^{-1} U_p$ be the standard point LU factorization of P. Let F be a nonnegative strictly π-upper triangular matrix such that $F \geqslant F_0$. Set $E = F^t$ and $B = (P - E)P^{-1}(P - F)$. Set further $\mathscr{L} = [1, N-1]$ if P_{NN} is diagonal or tridiagonal and $\mathscr{L} = [1, N]$ otherwise.*

If there exists some positive vector x such that $((P_p^{-1}U_p - L_p^{-1}F)x)_I > 0$ for all $I \in \mathscr{L}$ with

$$Bx \geqslant (1 - \tau_0) Ax \tag{3.1}$$

for some $0 \leqslant \tau_0 < 1$, then, setting

$$\tau_I = \inf\left\{ t > 0;\, tx_I \geqslant \left((I - P_p^{-1}U_p + L_p^{-1}F)x \right)_I \right\} \qquad \textit{for} \quad I \in \mathscr{L} \tag{3.2}$$

and

$$\tau = \max\Big(\tau_0, \max_{I \in \mathscr{L}} \tau_I\Big), \tag{3.3}$$

the largest eigenvalue of $B^{-1}A$ *satisfies*

$$v_{\max} \leqslant \frac{1}{1-\tau}. \tag{3.4}$$

The next result is already stated in [14] and shown to always give better upper bounds than (3.4). The scope of this improvement is however not discussed there in the context of line partitioning, as we shall do in Section 4.

THEOREM 3.2. *Let* $A = D - E_0 - F_0$ *be a symmetric positive definite matrix such that* D *is* π*-diagonal and symmetric,* F_0 *is strictly* π*-upper triangular, and* $E_0 = F_0^t$. *Let* P *be a* π*-diagonal Stieltjes matrix such that* $\text{offdiag}_p(P) \leqslant \text{offdiag}_p(D)$. *Let* F *be a nonnegative strictly* π*-upper triangular matrix such that* $F \geqslant F_0$. *Set* $E = F^t$ *and* $B = (P-E)P^{-1}(P-F)$.

If there exists some positive vector x *such that* $(I - P^{-1}F)x \geqslant 0$ *with* $((I-P^{-1}F)x)_I > 0$ *for all* I, $1 \leqslant I \leqslant N-1$, *and*

$$Bx \geqslant (1-\tau_0)Ax \tag{3.5}$$

for some $0 \leqslant \tau_0 < 1$, *then, setting*

$$\tau_I = \inf\{t > 0;\ tx_I \geqslant (P^{-1}Fx)_I\} \qquad \text{for} \quad 1 \leqslant I \leqslant N-1 \tag{3.6}$$

and

$$\tau = \max\Big(\tau_0, \max_{1 \leqslant I \leqslant N-1} \tau_I\Big), \tag{3.7}$$

the largest eigenvalue of $B^{-1}A$ *satisfies*

$$v_{\max} \leqslant \frac{1}{1-\tau}. \tag{3.8}$$

The following result is due to Axelsson and Eijkhout [4, Lemma 4.5].

THEOREM 3.3. *Let* $A = D - E - F$ *be a symmetric positive definite matrix such that* D *is symmetric and* $E = F^t$. *Let* P *be a symmetric positive definite matrix, and set* $B = (P-E)P^{-1}(P-F)$.

If $2P - D$ *is positive definite, then the largest eigenvalue of* $B^{-1}A$ *satisfies*

$$v_{\max} \leqslant \lambda_0 = \min\{\lambda > 0;\ (2 - 1/\lambda)P - D \text{ is nonnegative definite}\}. \tag{3.9}$$

No rule is however given in [4] for a practical estimation of this upper bound. In the following corollary, the nonnegative definiteness of $(2-1/\lambda)P - D$ is guaranteed by application of Gerschgorin's theorem [15].

COROLLARY 3.3. *Let* A, $D = (d_{ij})$, $P = (p_{ij})$, $1 \leqslant i, j \leqslant n$, *and* B *be given as in Theorem 3.3. Assume in addition that*

$$|d_{ij}| < 2|p_{ij}| \qquad \text{for all } i \neq j \text{ such that } d_{ij} \neq 0. \tag{3.10}$$

If

$$2p_{ii} - d_{ii} - \sum_{j \neq i} |2p_{ij} - d_{ij}| > 0 \qquad \text{for} \quad 1 \leqslant i \leqslant n, \tag{3.11}$$

and we define

$$\lambda_i = \max\left(\frac{p_{ii} - \Sigma_{j \neq i}|p_{ij}|}{2p_{ii} - d_{ii} - \Sigma_{j \neq i}|2p_{ij} - d_{ij}|},\ \max_{\substack{j \neq i \\ d_{ij} \neq 0}} \frac{1}{|2 - d_{ij}/p_{ij}|} \right) \qquad \text{for} \quad 1 \leqslant i \leqslant n, \tag{3.12}$$

then the largest eigenvalue of $B^{-1}A$ *satisfies*

$$v_{\max} \leqslant \lambda = \max_{1 \leqslant i \leqslant n} \lambda_i. \tag{3.13}$$

Proof. Application of Gerschgorin's theorem to prove that $(2-1/\lambda)P - D$ is nonnegative definite leads to the conditions

$$\left(2-\frac{1}{\lambda}\right)p_{ii} - d_{ii} - \sum_{j\neq i}|2p_{ij} - d_{ij} - \lambda^{-1}p_{ij}| \geqslant 0 \qquad \text{for} \quad 1 \leqslant i \leqslant n. \quad (3.14)$$

On the other hand, for fixed i, j, the assumption (3.10) implies that $2p_{ij} - d_{ij}$ and p_{ij} have like signs. Hence, since by (3.12) $\lambda^{-1}|p_{ij}| \leqslant |2p_{ij} - d_{ij}|$, one has that

$$|2p_{ij} - d_{ij} - \lambda^{-1}p_{ij}| = |2p_{ij} - d_{ij}| - \lambda^{-1}|p_{ij}|;$$

(3.14) then becomes

$$\lambda^{-1}\left(p_{ii} - \sum_{j\neq i}|p_{ij}|\right) \leqslant 2p_{ii} - d_{ii} - \sum_{j\neq i}|2p_{ij} - d_{ij}| \qquad \text{for} \quad 1 \leqslant i \leqslant n,$$

which yields the required result. ■

Now, we improve Theorem 3.3 in the following way.

THEOREM 3.4. *Let $A = D - E - F$ be a symmetric positive definite matrix such that D is π-diagonal and symmetric, F is strictly π-upper triangular, and $E = F^t$. Let P be a π-diagonal symmetric positive definite matrix, and set $B = (P-E)P^{-1}(P-F)$.*

If $2P_{II} - D_{II}$ is positive definite for $1 \leqslant I \leqslant N-1$, then the largest eigenvalue of $B^{-1}A$ satisfies

$$v_{\max} \leqslant \lambda_0 = \min\{\lambda > 0;\ (2-1/\lambda)P_{II} - D_{II} \text{ is nonnegative definite for } 1 \leqslant I \leqslant N-1 \text{ and } \lambda P_{NN} - D_{NN} \text{ is nonnegative definite}\}. \quad (3.15)$$

Proof. Let Q be the π-diagonal matrix defined by

$$Q_{II} = \begin{cases} P_{II} & \text{for} \quad I = 1, \dots, N-1, \\ 0 & \text{for} \quad I = N, \end{cases}$$

and let $\tilde{B} = [(1-1/\lambda)Q - E]Q^+[(1-1/\lambda)Q - F]$. Then, since $EQ^+Q = E$

and $QQ^+F = F$, we have

$$\tilde{B} = B + \lambda^{-1}(E+F) + \left(1-\frac{1}{\lambda}\right)^2 Q - P,$$

whence

$$B - \lambda^{-1}A = \tilde{B} + P - \left(1-\frac{1}{\lambda}\right)^2 Q - \lambda^{-1}D,$$

and the conclusion readily follows from the nonnegative definiteness of $\tilde{B}$ and the definition of Q. ■

The following corollary also follows then from the application of Gerschgorin's theorem.

COROLLARY 3.4. *Let A, $D = (d_{ij})$, $P = (p_{ij})$, $1 \leqslant i, j \leqslant n$, and B be given as in Theorem 3.4. Assume in addition that*

$$|d_{ij}| < 2|p_{ij}| \qquad \text{for all } i \neq j \text{ such that } d_{ij} \neq 0. \quad (3.16)$$

Let

$$2p_{ii} - d_{ii} - \sum_{j\neq i}|2p_{ij} - d_{ij}| > 0 \qquad \text{for all } i \in \pi_I \text{ with } 1 \leqslant I \leqslant N-1 \quad (3.17)$$

and

$$p_{ii} - \sum_{j\neq i}|p_{ij}| > 0 \qquad \text{for all } i \in \pi_N, \quad (3.18)$$

and define for $i \in \pi_I$, $I = 1, \dots, N-1$,

$$\lambda_i = \max\left(\frac{p_{ii} - \Sigma_{j\neq i}|p_{ij}|}{2p_{ii} - d_{ii} - \Sigma_{j\neq i}|2p_{ij} - d_{ij}|},\ \max_{\substack{j\neq i \\ d_{ij}\neq 0}} \frac{1}{|2 - d_{ij}/p_{ij}|}\right) \quad (3.19)$$

and for $i \in \pi_N$

$$\lambda_i = \max\left(\frac{d_{ii} - \Sigma_{j\neq i}\gamma_{ij}|d_{ij}|}{p_{ii} - \Sigma_{j\neq i}|p_{ij}|},\ \max_{j\neq i}\left|\frac{d_{ij}}{p_{ij}}\right|\right), \quad (3.20)$$

where for all $j \neq i$ *such that* $d_{ij} \neq 0$,

$$\gamma_{ij} = \begin{cases} 1 & \text{if } p_{ij} \text{ and } d_{ij} \text{ have like signs,} \\ -1 & \text{otherwise.} \end{cases} \tag{3.21}$$

Then the largest eigenvalue of $B^{-1}A$ *satisfies*

$$v_{\max} \leqslant \lambda = \max_{1 \leqslant i \leqslant n} \lambda_i. \tag{3.22}$$

Proof. For the nonnegative definiteness of $(2-1/\lambda)P_{II} - D_{II}$, $1 \leqslant I \leqslant N-1$, see the proof of Corollary 3.2. Application of Gerschgorin's theorem to prove that $\lambda P_{NN} - D_{NN}$ is nonnegative definite leads to the conditions

$$\lambda p_{ii} - d_{ii} - \sum_{j \neq i} |\lambda p_{ij} - d_{ij}| \geqslant 0 \qquad \text{for all} \quad i \in \pi_N.$$

Now, since by (3.20) $\lambda |p_{ij}| \geqslant |d_{ij}|$, one has that

$$|\lambda p_{ij} - d_{ij}| = \lambda |p_{ij}| - \gamma_{ij} |d_{ij}|,$$

whence the conclusion. ■

The following result is totally new.

Theorem 3.5. *Let* $A = D - E - F$ *be a symmetric positive definite matrix such that* D *is symmetric,* F *is nonnegative and strictly* π*-upper triangular, and* $E = F^t$. *Let* P *be a* π*-diagonal Stieltjes matrix and* $P = L_p P_p^{-1} U_p$ *its standard point LU factorization. Set* $B = (P-E)P^{-1}(P-F)$.
If Δ *is a nonnegative symmetric matrix such that* $2P - D + \Delta$ *is nonnegative definite, then the largest eigenvalue of* $B^{-1}A$ *satisfies*

$$v_{\max} \leqslant \max_{1 \leqslant i \leqslant n} \left\{ \left(\left[(I-E')^{-1} + (I-F')^{-1} + (I-E')^{-1} \Delta' (I-F')^{-1} \right] e \right)_i \right\}, \tag{3.23}$$

where

$$F' = P_p^{1/2} L_p^{-1} F U_p^{-1} P_p^{1/2}, \qquad E' = F'^t \tag{3.24}$$

and

$$\Delta' = P_p^{1/2} L_p^{-1} \Delta U_p^{-1} P_p^{1/2}. \tag{3.25}$$

Proof. Letting $B = LU$ with $U = P_p^{1/2} L_p^{-1}(P-F)$ and $L = U^t$, the spectrum of $B^{-1}A$ is identical to that of $L^{-1}AU^{-1}$. On the other hand, we may write $A = (D - 2P - \Delta) + \Delta + (P-E) + (P-F)$, whence

$$\begin{aligned} L^{-1}AU^{-1} &= L^{-1}(D-2P-\Delta)U^{-1} + L^{-1}\Delta U^{-1} \\ &\quad + P_p^{-1/2} U_p \left((P-F)^{-1} + (P-E)^{-1} \right) L_p P_p^{-1/2} \\ &= L^{-1}(D-2P-\Delta)U^{-1} + (I-E')^{-1}\Delta'(I-F')^{-1} \\ &\quad + (I-E')^{-1} + (I-F')^{-1}, \end{aligned}$$

and the conclusion readily follows from the non-positive-definiteness of the first right hand side term and the nonnegativity of the remaining ones. ■

Again, a practical application of Theorem 3.5 requires a criterion which guarantees the non-negative-definiteness of $2P - D + \Delta$. In the following corollary, which uses Gerschgorin's theorem, the matrix Δ is assumed to be pointwise diagonal. We omit the proof, since it is obvious.

Corollary 3.5. *Let* A, $D = (d_{ij})$, $P = (p_{ij})$, $1 \leqslant i, j \leqslant n$, *and* B *be given as in Theorem 3.5. Let* $\Delta = (\Delta_{ii}\ \delta_{ij})$ *be the nonnegative diagonal matrix defined by*

$$\Delta_{ii} = \max\left(-2p_{ii} + d_{ii} + \sum_{j \neq i} |2p_{ij} - d_{ij}|, 0 \right). \tag{3.26}$$

Then the largest eigenvalue of $B^{-1}A$ *satisfies*

$$v_{\max} \leqslant \max_{1 \leqslant i \leqslant n} \left\{ \left(\left[(I-F')^{-1} + (I-E')^{-1} + (I-E')^{-1} \Delta' (I-F')^{-1} \right] e \right)_i \right\}, \tag{3.27}$$

where F', E', *and* Δ *are defined from* F, E, *and* Δ *by* (3.24) *and* (3.25).

We finally recall the following result by Axelsson [2].

THEOREM 3.6. *Let $A = D - E - F$ be a symmetric positive definite matrix such that D is symmetric, F is nonnegative and strictly π-upper triangular, and $E = F^t$. Let P be a π-diagonal Stieltjes matrix. Set $B = (P-E)P^{-1}(P-F)$. Let $\mu_{\max}(A-B)$ denote the largest eigenvalue of $A-B$. Then the largest eigenvalue of $B^{-1}A$ satisfies*

$$v_{\max} \leqslant 1 + \max\{\mu_{\max}(A-B), 0\} \max_{1 \leqslant i \leqslant n} (B^{-1}e)_i \tag{3.28}$$

and

$$v_{\max} \leqslant 1 + \|A - B\|_\infty \max_{1 \leqslant i \leqslant n} (B^{-1}e)_i. \tag{3.29}$$

The ∞-norm is used in order to allow a practical use of Axelsson's result (3.28). Note that when $Ae = Be$ with $\text{offdiag}(A-B) \leqslant 0$ (as occurs when B is a modified block incomplete factorization of a Stieltjes matrix A, as defined in Section 2, with $x = e$), then $\|A-B\|_\infty$ is simply two times the largest entry of $\text{diag}_p(A-B)$.

As may be easily checked by the reader, the general assumptions of all the above theorems are met when B is a modified block incomplete factorization of a Stieltjes matrix A as defined in Section 2. It should further be noticed that these results have actually not exactly the same scope: in Theorems 3.3 and 3.4, there are no restriction on the sign of E and F as it appears in the other theorems, while it is shown in [12] that Theorem 3.1 (and hence Theorem 3.2) applies actually to the class of almost Stieltjes matrices (see [9]), which is more general than that of Stieltjes matrices.

We finally point out that the upper bounds described above may also apply to the relaxed block incomplete factorization methods introduced in [5].

4. NUMERICAL RESULTS

We present here the results of numerical experiments comparing the upper spectral bounds presented in the preceding section, on some simple but typical test problems. We consider the linear systems derived from the

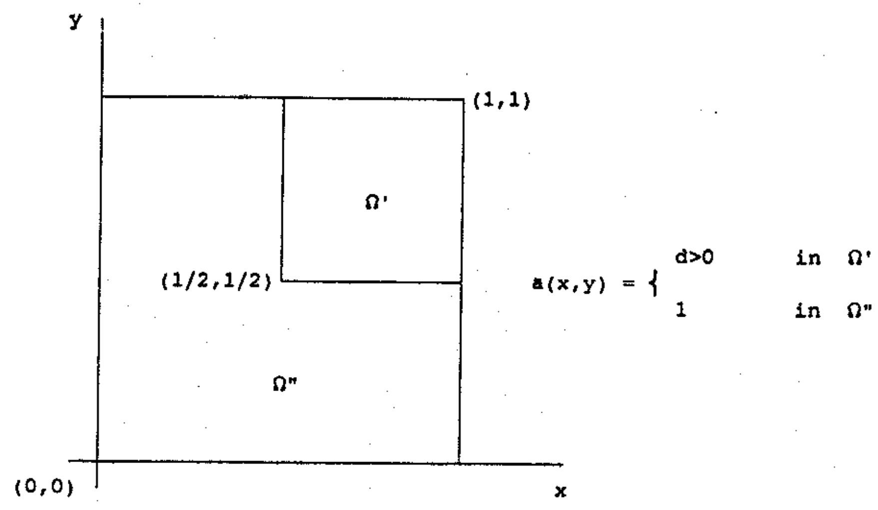

FIG. 1. Test problems 1, 2, 3. Specification of the coefficients $a(x,y)$.

five point central difference approximation of the two-dimensional PDEs (we assume a uniform grid of mesh size h in both directions)

$$-\nabla a(x,y)\nabla u(x,y) = f(x,y) \qquad \text{in } \Omega =]0,1[\times]0,1[,$$

$$u(x,y) = g(x,y) \qquad \text{on } \Gamma_0,$$

$$\frac{\partial u}{\partial n}(x,y) = h(x,y) \qquad \text{on } \Gamma_1 = \Gamma \setminus \Gamma_0,$$

with $\Gamma = \partial\Omega$, Ω being subdivided as in Figure 1, which depicts also the values of the coefficients $a(x,y)$. Three situations for the boundary conditions are investigated here: in the first (problem 1), $\Gamma_0 = \Gamma$; in the second (problem 2) Γ_0 coincides with the bottom boundary ($y = 0$); and in the third (problem 3) it corresponds to the top boundary ($y = 1$). The resulting matrices are obviously irreducibly diagonally dominant Stieltjes matrices. We use lexicographic ordering and line partitioning.

The modified block incomplete factorization we consider is defined by (2.1)–(2.3) with $\beta = \text{tridiag}(\varepsilon)$, $K = \text{tridiag}(P^{-1})$, and $x = e$. The algorithm

(2.7) simplifies here to

$$P_{11} = D_{11}$$

and

$$K_{I-1I-1} = \text{tridiag}\left(P_{I-1I-1}^{-1}\right),$$

$$P_{II} = D_{II} - A_{II-1}K_{I-1I-1}A_{I-1I} - \Omega_{II},$$

$$\Omega_{II}x_I = A_{II-1}\left(P_{I-1I-1}^{-1} - K_{I-1I-1}\right)A_{I-1I}x_I \qquad \text{for} \quad I = 2,3,\ldots,N.$$

In Table 1 are collected the numerically computed largest eigenvalue of $B^{-1}A$ and the upper bounds displayed in Section 3.

In light of these results, the following observations can be made.

(1) As would be expected from a careful look at their derivations, the last two upper bounds, (3.27) and (3.29), appear to have the wider range of application. The second one is however clearly of no practical interest, whereas the first one behaves quite well, bringing out often the $O(h^{-1})$ dependence of $\kappa(B^{-1}A)$, even for problem 3, where all the other analyses fail.

(2) From an existence point of view, the upper bounds (3.4) and (3.8) behave very similarly, but, in compliance with the theoretical comparison done in [14], the second upper bound outperforms (generally by far) the first one. However, when finite, both exhibit the correct order of magnitude, $O(h^{-1})$, for $\kappa(B^{-1}A)$.

(3) Compared with our results, especially Equations (3.8) and (3.27), Axelsson and Eijkhout's theory [Equation (3.13)] appears (at least for the examples considered here) to perform nicely only for the "model" problem, i.e. with constant material coefficients and Dirichlet boundary conditions. Our improved version of this theory [Equation (3.22)] enlarges its range of application, which nevertheless remains somewhat limited.

(4) Of particularly motivating interest is the combined use of Theorems 3.2 and 3.5 [cf. Equations (3.8) and (3.27)], which are complementary in the sense that, in all considered situations but problem 3 with $d = 10^{-2}$, at least one of them shows the $O(h^{-1})$ behavior of $\kappa(B^{-1}A)$. Observe that in the case of problem 3 with $d = 10^{-2}$, where Theorem 3.2 fails and Theorem 3.5 gives rise to very large bounds, the spectral condition number itself is very large too (compared to all other cases) and visibly not $O(h^{-1})$.

TABLE 1

NUMERICALLY DETERMINED LARGEST EIGENVALUE AND UPPER SPECTRAL BOUNDS [a]

h^{-1}	$v_{\max}$	Upper spectral bound (3.4)	(3.8)	(3.13)	(3.22)	(3.27)	(3.29)
			$d = 0.01$				
			Problem 1				
12	1.36	9.63	6.40	—	—	9.41	95.8
24	2.18	26.67	16.4	—	—	29.4	596
48	4.09	66.20	39.2	—	—	92.2	3095
96	8.08	147.5	87.4	—	—	259	13722
			Problem 2				
12	2.32	35.38	11.3	—	—	49.9	912
24	4.41	76.15	23.5	—	—	152	3859
48	7.70	157.4	47.9	—	—	470	15129
96	16.30	318.1	96.0	—	—	1482	59232
			Problem 3				
12	8.15	∞	∞	—	—	204	1068
24	33.53	∞	∞	—	—	1291	6899
48	100.8	∞	∞	—	—	5624	37860
96	309.8	∞	∞	—	—	14635	166589
			$d = 1.0$				
			Problem 1				
12	1.35	11.96	7.0	4.00	3.93	7.44	5.12
24	2.16	30.92	18.3	9.65	9.47	14.9	26.10
48	4.02	75.52	44.9	22.9	22.5	29.9	125.6
96	7.87	159.9	94.8	47.9	47.4	59.2	535.3
			Problem 2				
12	2.46	37.4	12	—	12	50.1	36.34
24	4.66	77.5	24	—	24	151	142.7
48	8.64	157.7	48	—	48	473	565.5
96	16.73	318.1	96	—	96	1509	2246
			Problem 3				
12	2.60	∞	∞	∞	∞	13.8	42.1
24	5.02	∞	∞	∞	∞	26.7	152.8
48	9.38	∞	∞	∞	∞	52.7	583.9
96	18.34	∞	∞	∞	∞	104.8	2279

TABLE 1 (*Continued*)

h^{-1}	v_{max}	Upper spectral bound (3.4)	(3.8)	(3.13)	(3.22)	(3.27)	(3.29)
				$d = 100$			
				Problem 1			
12	1.39	607	241	121	121	7.06	195
24	2.31	1669	1069	535	535	14.4	1568
48	4.47	3960	2376	1188	1188	28.9	9396
96	8.89	8493	4800	2401	2401	58.5	48571
				Problem 2			
12	3.65	985	600	—	517	890	4066
24	9.63	2052	1202	—	1159	2830	15688
48	24.22	4243	2416	—	2415	9549	59132
96	54.59	8685	4848	—	4848	33056	229132
				Problem 3			
12	2.34	∞	∞	∞	∞	13.3	2749
24	4.14	∞	∞	∞	∞	26.2	13110
48	8.86	∞	∞	∞	∞	51.4	56201
96	17.08	∞	∞	∞	∞	100.9	227052

[a]Computed (whenever possible) from Equations (3.4), (3.8), (3.13), (3.22), (3.27), and (3.29) for problems 1, 2, and 3 with $d = 0.01$, 1.0, and 100; h stands for the mesh size.

5. CONCLUSION

We have applied the conditioning analyses presented in this paper to second order elliptic PDEs. For a large class of problems (wider then covered by previous theoretical approaches [4, 8, 12, 14]), we obtain, for the spectral condition numbers associated with modified block incomplete factorization methods, upper bounds that are readily computed during the factorization. Numerical experiments show that these bounds exhibit $O(h^{-1})$ behavior in most cases. A formal proof of this behavior is investigated in [13].

Useful suggestions from Professor O. Axelsson and an anonymous referee to improve the readability of the paper are gratefully acknowledged.

REFERENCES

1 O. Axelsson, A general incomplete block-matrix factorization method, *Linear Algebra Appl.* 74:179–190 (1986).
2 O. Axelsson, Condition Number Estimates for Elliptic Difference Problems with Anisotropy, Scientific Report, Dept. of Mathematics, Catholic Univ., Nijmegen, The Netherlands, 1989.
3 O. Axelsson and V. A. Barker, *Finite Element Solution of Boundary Value Problems. Theory and Computation*, Academic, New York, 1984.
4 O. Axelsson and V. Eijkhout, Robust vectorizable preconditioners for three-dimensional elliptic difference equations with anisotropy, in *Algorithms and Applications on Vector and Parallel Computers* (H. J. J. te Riele, Th. J. Dekker, and H. A. van der Vorst, Eds.), North Holland, Amsterdam, 1987.
5 O. Axelsson and G. Lindskog, On the eigenvalue distribution of a class of preconditioning methods, *Numer. Math.* 48:479–498 (1986).
6 O. Axelsson and B. Polman, On approximate factorization methods for block matrices suitable for vector and parallel processors, *Linear Algebra Appl.* 77:3–26 (1986).
7 R. Beauwens and M. Ben Bouzid, On sparse block factorization iterative methods, *SIAM J. Numer. Anal.* 24:1066–1076 (1987).
8 R. Beauwens and M. Ben Bouzid, Existence and conditioning properties of sparse approximate block factorizations, *SIAM J. Numer. Anal.* 25:941–956 (1988).
9 R. Beauwens and R. Wilmet, Conditioning analysis of positive definite matrices by approximate factorizations, *J. Comput. Appl. Math.* 26:257–269 (1989).
10 A. Ben Israel and T. N. E. Greville, *Generalized Inverses: Theory & Applications*, Wiley, 1974.
11 P. Concus, G. H. Golub, and G. Meurant, Block preconditioning for the conjugate gradient method, *SIAM J. Sci. Statist. Comput.*, 6:220–252 (1985).
12 M. M. Magolu, Conditioning Analysis of Sparse Block Approximate Factorizations. *Appl. Numer. Math.* To appear.
13 M. M. Magolu, Analytical Bounds for Block Approximate Factorization Methods, Scientific Report, Univ. Libre de Bruxelles, June 1990; *Linear Algebra Appl.*, submitted for publication.
14 Y. Notay, Conditioning analysis of modified block incomplete factorizations, *Linear Algebra Appl.* 154–156:711–722 (1991).
15 R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1962.
16 D. M. Young, *Iterative Solution of Large Linear Systems*, Academic, New York, 1971.

Received 30 March 1990; final manuscript accepted 29 November 1990