Russ. J. Numer. Anal. Math. Modelling, Vol.12, No.5, pp.399–420(1997)

Generalized compensation principle in incomplete factorization methods*

V. P. IL'IN† and K. YU. LAEVSKII†

Abstract – We consider incomplete factorization methods of solving the systems of linear algebraic equations with Stieltjes matrices in which the precondition matrix B is constructed by the generalized principle of adjustment of the row sums $By_q = Ay_q$, $q = 1,\dots,m$, with a different number m of test vectors. We formulate the theorem on the conditions for the existence of these preconditioners and their positive definiteness. We give examples of numerical experiments, which demonstrate the efficiency of the algorithms proposed.

1. INTRODUCTION

The iterative incomplete factorization algorithms are now actively developed methods of solving the systems of linear algebraic high-order equations

$$Au = f \tag{1.1}$$

with sparse matrices that result from the approximation of multidimensional boundary value problems by the grid methods, viz. the finite difference method, the finite element method or the finite volume method (see [6,7] and the references therein). The main problem here is to construct factorized precondition matrices B that would be readily invertible and sufficiently close to the initial matrix of the system A in a sense (which is the subject of special discussion). The iterative process in the simplest case is realized by the formula

$$B(u^n - u^{n-1}) = f - Au^{n-1}. \tag{1.2}$$

In the most efficient algorithms, namely in Chebyshev acceleration algorithms or conjugate gradient algorithms calculations by formula (1.2) are also used. However, the vector of a successive approximation u^n is then corrected by the spectral or variational optimization of the iterative process. Provided that the matrices A, B are symmetric and positive definite, the number of iterations, which is necessary to reduce the initial error by a factor of ε^{-1}:

$$(Az^n,z^n)/(Az^0,z^0) \leqslant \varepsilon, \qquad z^n = u - u^n$$

is estimated in both cases by the value

$$n(\varepsilon) \leqslant \frac{1}{2}|\ln\varepsilon|\,æ^{-1/2} + 1$$

where $æ$ is the condition number of the matrix $B^{-1/2}AB^{-1/2}$, which is equal to the ratio of the maximum to minimum eigenvalues of the similar matrix $B^{-1}A$.

* The work was supported by the Russian Foundation for the Basic Research (96–01–01770) and by INTAS (93–0377–ext).

† Computer Center, Siberian Branch of the Russian Academy of Sciences, Novosibirsk 630090, Russia

Consider block tridiagonal matrices $A = D - L - U$ of the form

$$\begin{bmatrix} D_1 & -U_1 & & & & \mathbf{0} \\ -L_2 & D_2 & -U_2 & & & \\ & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \cdot \\ & & & -L_{M-1} & D_{M-1} & -U_{M-1} \\ \mathbf{0} & & & & -L_M & D_M \end{bmatrix}$$

where $D = \operatorname{diag}\{D_k\}$ is the diagonal or block diagonal matrix, and $L = \{L_k\}$, $U = \{U_k\}$ are the lower and upper rigorously triangular matrices (the orders N_k of diagonal blocks D_k may be different and the general order A is $N = N_1 + \ldots + N_M$). For these matrices the precondition matrix B is determined for a wide class of algorithms in the form

$$B = (G - L)G^{-1}(G - U) \tag{1.3}$$

where $G = \operatorname{diag}\{G_k\}$ is the block diagonal matrix whose blocks are found by the recursion method:

$$G_1 = D_1, \quad G_k = D_k - \overline{L_k G_{k-1}^{-1} U_{k-1}} - \theta C_k \qquad k = 2,3,\ldots,M. \tag{1.4}$$

Here $0 \leqslant \theta \leqslant 1$ is an iterative parameter. The matrix $\bar{Q}$ implies approximation (in a certain sense) to the matrix Q, and C_k is conventionally determined as a diagonal matrix whose entries are found by the equality

$$C_1 = 0, \quad C_k e = (L_k G_{k-1}^{-1} U_{k-1} - \overline{L_k G_{k-1}^{-1} U_{k-1}})e \qquad k = 2,3,\ldots,M \tag{1.5}$$

where e is the vector with unit components.

The main method of constructing symmetric matrix 'approximations' is the band method:

$$\bar{Q} = Q^{(p)} \tag{1.6}$$

where $Q^{(p)}$ implies the 'band part' of width p in the matrix Q (the entries of the matrix $Q^{(p)}$ for $|i - j| \leqslant (p-1)/2$, $p = 1,3,5,\ldots$, are the same as those of Q, and the other entries are zero ones).

N. I. Buleev was the first to develop the incomplete factorization methods. As the matrix C he took $\operatorname{diag}\{C_k\}$, and the matrices G and B were determined empirically by approximation principles, namely the condition (1.5) is that for $\theta = 1$ the vector equality holds:

$$Be = Ae. \tag{1.7}$$

This equality is called the complete compensation condition. Since the error vector $z^1 = u - u^1$ that is determined from (1.2) satisfies the relation

$$Bz^1 = (B - A)z^0$$

when the condition (1.7) is satisfied and $z^0 = e$, we obtain $z^1 = 0$, i.e. the exact solution is obtained on a single iteration.

More recently, H. L. Stone proposed the method [9] in which the compensation principle was developed so that the exact solution was obtained on a single iteration if the vector components of the initial error z^0 are the values of a linear function. However, the constructed precondition matrix B turns out to be nonsymmetric, which hampers the acceleration of the iterative process. There are other papers (see, e.g. [1]) in which attempts were made to construct symmetric preconditioners with analogous properties but the algorithms for calculating B turned out to be unstable.

In the algebraic language the condition (1.7) is called the adjustment principle of the row sums (of the initial and precondition matrices). Its development, viz. the generalized adjustment principle of the row sums is that (1.7) is replaced by the equality

$$By = Ay, \qquad y > 0.$$

This leads to the only and unessential change in the algorithm, viz. the vector e in the determination of C_k from (1.5) is replaced by the arbitrary positive vector y:

$$C_k y = (L_k G_{k-1}^{-1} U_{k-1} - \overline{L_k G_{k-1}^{-1} U_{k-1}})y, \qquad y > 0. \tag{1.8}$$

For the Stieltjes matrices (symmetric monotonic matrices with nonpositive off-diagonal entries) the incomplete factorization methods with preconditioners of the form (1.3), (1.4), (1.8) are studied as to the correctness of calculations (nonsingularity of matrices G) and the estimate of the convergence rate of the iterative process.

The extension of the algorithms considered is possible, viz. the change-over in the determination of matrices C_k from diagonal to band matrices whose entries are found by using several vectors rather than the single 'test' vector y:

$$\begin{gathered} C_k y_k^{(q)} = (L_k G_{k-1}^{-1} U_{k-1} - \overline{L_k G_{k-1}^{-1} U_{k-1}})y_k^{(q)} \\ q = 1,\dots,m, \quad k = 2,3,\dots,M. \end{gathered} \tag{1.9}$$

Note that each vector $y_k^{(q)}$ is of order N_k, which is equal to the order of the corresponding square matrices D_k, G_k, and C_k, and the band width of the matrix C_k, which is calculated by the condition (1.9), is $2m + 1$. It is not difficult to check that equalities (1.9) correspond to the relations between the initial and precondition matrices for $\theta = 1$:

$$Ay^{(q)} = By^{(q)}, \quad q = 1,\dots,m, \qquad y^{(q)} = \{y_k^{(q)},\ k = 1,\dots,M\}. \tag{1.10}$$

These relations are called the generalized compensation principle. Some numerical experiments, in which two or three test vectors $y^{(q)}$ are used, are described in [2,3]. It is shown in [4] that the tridiagonal symmetric matrix C_k for the case of two vectors of the particular form $y^{(q)}$, $q = 1,2$, is determined uniquely.

In Section 2 we describe some properties of the algorithms, which are obtained by the generalized compensation principle for the Stieltjes systems of equations. We prove, in particular, that if m vectors $y_k^{(q)}$, $q = 1,\dots,M$, are strongly linearly independent (the definition is given below), equalities (1.10) uniquely define the symmetric band $(2m+1)$-diagonal matrices C_k of dimension N_k, which give the matrix $C = \operatorname{diag}\{C_k\}$ that is not positive definite for the general case. We show in Section 3 that the precondition matrix B is a positive definite one for some special cases. In Section 4 we give examples of the solutions of methodical grid equations with five-diagonal matrices. In Section 5 we discuss some results obtained and the new arising problems, which invite special investigation.

2. GENERALIZED COMPENSATION ALGORITHM

In this section we dwell on the incomplete factorization algorithm that is based, first, on the use of the band approximations of the matrices [in accordance with (1.6)] in the determination of G_k and C_k in (1.4) and (1.10) and, second, on the use of some number $m > 1$ of the test vectors $y^{(q)}$, $q = 1,\dots,m$, in (1.10).

Denoting by Y_k the rectangular $N_k \times m$ matrix whose columns are vectors $y_k^{(q)}$, we rewrite relation (1.9) in the matrix form

$$C_k Y_k = R_k Y_k = V_k, \qquad k = 2,3,\dots,M \tag{2.1}$$

where

$$R_k = L_k G_{k-1}^{-1} U_{k-1} - (L_k G_{k-1}^{-1} U_{k-1})^{(p)}$$

and V_k is a rectangular matrix of the same structure as Y_k [its columns are $v_k^{(q)} = R_k y_k^{(q)}$]. The recursion relations (1.4) are also rewritten, respectively:

$$G_1 = D_1, \qquad G_k = D_k - (L_k G_{k-1}^{-1} U_{k-1})^{(p)} - \theta C_k \\ k = 2,3,\dots,M. \tag{2.2}$$

When realizing formulae (2.1), (2.2) we face two nonconventional algebraic problems. The first problem is to find the band part of the matrix product one of whose cofactor is a matrix inverse to the band one. The main problem here is to calculate the band part of the matrix inverse to the band one. This is the only problem if the matrices L_k and U_k are diagonal ones, i.e.

$$(L_k G_{k-1}^{-1} U_{k-1})^{(p)} = L_k (G_{k-1}^{-1})^{(p)} U_{k-1}$$

The algorithm for solving this problem, which needs $O(Nm^2)$ operations, is described in [6,7].

The second problem to be discussed below is to find the band matrix C_k that satisfies the condition (2.1). It is necessary to emphasize three principal points here, viz. the derivation of recursion relations for the matrix entries $C = \{c_{i,j}\}$ (for brevity, we omit the indices k), the determination of conditions for the vectors $y^{(q)}$, which are necessary to solve this problem, and clearing up the question of whether the obtained matrix C is symmetric when the specified matrix R is symmetric.

We represent the sought matrix C as

$$\begin{bmatrix} c_{1,1} & \cdots & c_{1,m} & & & & \\ \vdots & \ddots & & c_{2,m+1} & & \mathbf{0} & \\ c_{1,m} & & \ddots & & & & \\ & c_{2,m+1} & & \ddots & & & \\ & & & & c_{N-m+1,N-m+1} & \cdots & c_{N-m+1,N} \\ & \mathbf{0} & & & \vdots & & \vdots \\ & & & & c_{N,N-m+1} & \cdots & c_{N,N} \end{bmatrix} \tag{2.3}$$

For now, we assume that the matrix C is 'quasisymmetric', i.e. its entries satisfy the condition $c_{i,j} = c_{j,i}$ only if $i,j \leqslant N-m$. In other words, only the main submatrix $C_m = \{c_{i,j};\ i,j \geqslant N-m+1\}$ of order m in the lower right corner of the matrix C is assumed to be nonsymmetric. We have chosen this representation because the number of the unknown $c_{i,j}$ in this case is equal to the number Nm of equations of the system

$$Cy^{(q)} = v^{(q)}, \qquad v^{(q)} = Ry^{(q)}, \qquad q = 1,\dots,m \tag{2.4}$$

where N is the order of the vectors $y^{(q)}$, $v^{(q)}$ and the sought matrix C.

We denote by c_l the column vector of order m whose entries for $l = 1,\dots,N-m$, are the entries $c_{l,l}, c_{l,l+1}, \dots, c_{l,l+m-1}$ of the l-th row of the matrix C, which are in its upper triangular part, and for $l = N-m+1,\dots,N$ are the entries $c_{l,N-m+1}, c_{l,N-m+2},\dots,c_{l,N}$ of the l-th row of the matrix C, which are in its last m columns. We denote by Y_l a square submatrix of order m, which is the transpose of the matrix that consists of the rows of the rectangular matrix Y with numbers from l to $l+m-1$ inclusive. Then we rewrite equations (2.4) as

$$\begin{aligned} &Y_1 c_1 = v_1 \\ &Y_l c_l = v_l - c_{1,l} y_1 - \dots - c_{l-1,l} y_{l-1}, \qquad l = 2,3,\dots,m-1 \\ &Y_l c_l = v_l - c_{m-l+1,l} y_{m-l+1} - \dots - c_{l-1,l} y_{l-1}, \qquad l = m,m+1,\dots,N-m \\ &Y_{N-m+1} c_l = v_l - c_{m-l+1,l} y_{m-l+1} - \dots - c_{N-m,l} y_{N-m}, \qquad l = N-m+1,\dots,N-1 \\ &Y_{N-m+1} c_N = v_N \end{aligned} \tag{2.5}$$

where y_l and v_l denote the column vectors of order m whose entries consist of the l-th rows of the matrices Y and V, respectively.

If we denote the right-hand sides in equations (2.5) by w_l, the evaluation of the unknown vectors c_l reduces to a successive solution of the systems

$$Y_l c_l = w_l, \qquad l = 1,\dots,N$$

where the components w_l are expressed recurrently in terms of the calculated values of the vector components c_k, $k < l$. It is evident that all the matrices Y_l must be nonsingular for the problem (2.5) to be solvable uniquely, given the 'quasisymmetric' structure of the matrix C.

Definition 2.1. The rectangular $N \times m$ matrix Y is called the matrix of strong rank m if all of its submatrices Y_l of order m (which consist of the successive rows y_l, $y_{l+1}, \ldots, y_{l+m-1}$ of the matrix Y) are nonsingular.

The above concept may be expressed in terms of the properties of the column vectors of the matrix, i.e. $y^{(q)}$, $q = 1, \ldots, m$.

Definition 2.2. The vectors $y^{(q)}$ of order N are called strongly linearly independent vectors if they form the rectangular $N \times m$ matrix $(m < N)$ of strong rank m.

Recall that the necessary and sufficient condition for the set of the vectors $y^{(q)}$, $q = 1, \ldots, m$, to be linearly independent (in the ordinary sense) is that the matrix Y must have at least a single nonzero minor of order m (the matrix Y is of rank m).

The matrix C, which is determined by relations (2.4), (2.5), in the block second-order representation

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_m \end{pmatrix} \tag{2.6}$$

has the 'quasisymmetry' properties, viz. the conditions are satisfied:

$$C_{11} = C'_{11}, \qquad C_{21} = C'_{12}. \tag{2.7}$$

We can formulate the result obtained as follows.

Theorem 2.1. If the vectors $y^{(q)}$, $q = 1, \ldots, m$, are strongly linearly independent, the band matrix C of the form (2.3) with properties (2.6), (2.7) is determined uniquely by relations (2.4), (2.5).

We emphasize that the assertion does not require that $v^{(q)}$ be equal to $Ry^{(q)}$ in relations (2.4), i.e. the entries of the matrix C are formally determined in (2.5) for any set of the vectors $v^{(q)}$. In particular, for $v^{(q)} = 0$ we have $C = 0$.

It is evident that the strongly linearly independent vectors $y^{(q)}$ can be constructed in a variety of ways. One of them is, for example, to choose a nonsingular matrix Y_1 with rows $y_1, y_2, \ldots, y_m$ and determine the other rows successively: $y_l = y_{l-m}$ for $l > m$. We now show that the constructed matrix C is actually symmetric.

Theorem 2.2. Let the hypotheses of Theorem 2.1 be satisfied and let the vectors $v^{(q)}$ be determined by relations (2.4) with symmetric matrix R. Then the matrix C is symmetric.

Proof. We represent the matrix C as the sum of its symmetric and skew-symmetric parts:

$$C = \hat{C} + \check{C}$$

where

$$\hat{C} = \frac{1}{2}(C + C'), \qquad \check{C} = \frac{1}{2}(C - C').$$

Using them, we can rewrite equalities (2.4) as

$$\check{C}Y = (R - \hat{C})Y \tag{2.8}$$

where in view of (2.7) only the lower right block of the skew-symmetric matrix $\check{C}$ can be nonzero:

$$\check{C} = \begin{bmatrix} 0 & 0 \\ 0 & \check{C}_m \end{bmatrix}, \qquad \check{C}_m = \frac{1}{2}(C_m - C_m'). \tag{2.9}$$

If the matrix equality (2.8) is premultiplied by the transpose Y' and (2.9) is taken into account, we get

$$Y'_{N-m+1}\check{C}_m Y_{N-m+1} = Y'(R - \hat{C})Y.$$

Hence, because Y_{N-m+1} is nonsingular, we obtain the expression

$$\check{C}_m = (Y'_{N-m+1})^{-1}Y'(R - \hat{C})YY^{-1}_{N-m+1}.$$

Because the matrices $\check{C}_m$, $R - \hat{C}$ are congruent and $R - \hat{C}$ is symmetric, this expression implies that the skew-symmetric matrix $\check{C}_m$ can be only a zero one, i.e. $C = \hat{C}$ is symmetric.

3. SOME PROBLEMS OF THE ALGORITHM JUSTIFICATION

Theorems 2.1, 2.2 actually establish the minimum conditions for the generalized compensation algorithm to exist, i.e. for the symmetric matrix C to be calculated by the conditions (1.9).

Let H_k be a square nonsingular matrix of order m and let Y_k be a rectangular $N_k \times m$-matrix of strong rank m. Then it is evident that the matrix

$$\bar{Y}_k = Y_k H_k \tag{3.1}$$

is also of strong rank m. Hence the matrix equation

$$C_k \bar{Y}_k = R_k \bar{Y}_k \tag{3.2}$$

is solvable uniquely and its solution is the matrix C_k from equation (2.1). We can formulate the assertion obtained (omitting the index k for brevity) as follows.

Theorem 3.1. When the hypotheses of Theorem 2.1 are true, the matrix C is invariant under any nonsingular transformation of vectors $y^{(q)}$ of the form (3.1)

In order to understand this it is sufficient to check that if $\bar{y}^{(q)}$, $q = 1,\dots,m$, are the columns of the matrix $\bar{Y}$ and h_{pq} are the entries of the matrix H, the matrix relation $\bar{Y} = YH$ is equivalent to the linear transformation of the test vectors

$$\bar{y}^{(q)} = h_{1,q}y^{(1)} + \dots + h_{m,q}y^{(m)}.$$

Thus, if the precondition matrix B is constructed by formulae (1.3), (1.4), (1.9) for some set of test vectors $y^{(q)}$, $q = 1,\dots,m$, its form does not change under any nonsingular linear transformation of these vectors.

In order to justify the generalized compensation principle, one should first of all prove that the matrix G is positive definite and hence the matrix B.

To prove that the precondition matrix B is positive definite we need the following assertion.

Lemma 3.1. Let $A = D - L - U$ be the Stieltjes block tridiagonal matrix ($L = U'$) that satisfies the conditions

$$D_1 e_1 \geqslant U_1 e_2, \qquad D_k e_k \geqslant L_k e_{k-1} + U_k e_{k+1}, \qquad k = 2,3,\dots,M-1 \tag{3.3}$$

$$D_k e_k > L_k e_{k-1}, \qquad k = 2,3,\dots,M \tag{3.4}$$

where e_k is the vector of dimension N_k with unit components. Then if the matrices G_k calculated by formulae (2.1), (2.2) have nonpositive off-diagonal entries, they are the Stieltjes matrices.

Proof. We establish the validity of the lemma by induction, taking into account that $G_1 = D_1$ is the Stieltjes matrix and $G_1 c_1 \geqslant U_1 c_2$ in accordance with (3.3). Let G_{k-1} be the Stieltjes matrix and $G_{k-1} e_{k-1} \geqslant U_{k-1} e_k$. Then

$$G_{k-1}^{-1} \geqslant 0, \qquad R_k = L_k G_{k-1}^{-1} U_{k-1} - L_k (G_{k-1}^{-1})^{(p)} U_{k-1} \geqslant 0 .$$

Since $R_k e_k = C_k e_k$, the chain of inequalities holds:

$$G_k e_k = D_k e_k - L_k G_{k-1}^{-1} U_{k-1} e_k + (1-\theta) R_k e_k$$

$$\geqslant D_k e_k - L_k e_{k-1} + L_k G_{k-1}^{-1}(G_{k-1} e_{k-1} - U_{k-1} e_k) \geqslant D_k e_k - L_k e_{k-1} .$$

In accordance with (3.4) $G_k e_k > 0$, and since under the condition of the lemma the off-diagonal entries of the matrix G_k are nonpositive, according to Lemma 6.4 from [4] G_k is the Stieltjes matrix. Finally, the inequality $G_k e_k \geqslant U_k e_{k+1}$ follows from (3.3), which completes the induction step.

From the representation (2.2) it follows that if the matrix G_{k-1} is of the Stieltjes character, in order for the off-diagonal entries of the matrix G_k to be nonpositive it is sufficient for the off-diagonal entries of the matrix $L_k G_{k-1}^{-1} U_{k-1} + C_k$ to be nonnegative for $\theta \in [0,1]$ and, in particular, it is sufficient for the off-diagonal entries of the matrix C_k to be nonnegative.

We consider the vectors for the case $m = 2$:

$$y_k^{(1)} = e_k, \qquad y_k^{(2)} = \{y_{k,1}^{(2)},\dots,y_{k,N_k}^{(2)}\}, \qquad k = 2,3,\dots,M \tag{3.5}$$

where $y_k^{(2)}$ is a rigorously monotonic vector, i.e. for $i = 2,3,\dots,N_k$ we have either $y_{k,i}^{(2)} > y_{k,i-1}^{(2)}$ or $y_{k,i}^{(2)} < y_{k,i-1}^{(2)}$. It is not difficult to see that these vectors are strongly linearly independent. Omitting the index k for brevity, we consider the off-diagonal entries of the matrix C obtained from the conditions (2.4), (3.5).

Lemma 3.2. If the symmetric matrix R has nonnegative entries, the off-diagonal entries of the matrix C determined by formulae (2.4), (3.5) are also nonnegative.

Proof. We denote $y = Te = \{y_i = y^{(2)}_{k,i}\}$, $T = \mathrm{diag}\{y_i\}$. From relations (2.4), (3.5) it follows that

$$(CTe)_i - y_i(Ce)_i = (Se)_i, \qquad i = 1,2,\ldots,N_k \tag{3.6}$$

where $S = RT - TR$ is a skew-symmetric matrix whose entries $s_{i,j}$ are expressed in terms of the matrix $R = \{r_{i,j}\}$:

$$s_{i,j} = (y_j - y_i)r_{i,j}. \tag{3.7}$$

It is not difficult to derive from (3.6) the equations for the off-diagonal entries of the matrix C:

$$(y_2 - y_1)c_{1,2} = (Se)_1$$

$$(y_{i+1} - y_i)c_{i,i+1} - (y_i - y_{i-1})c_{i-1,i} = (Se)_i, \qquad i = 2,3,\ldots,N_k - 1.$$

The successive partial summation of these equations yields the equalities:

$$c_{i,i+1} = \frac{1}{y_{i+1} - y_i}\sum_{j=1}^{i}(Se)_j, \qquad i = 1,2,\ldots,N_k - 1.$$

Since the matrix S is skew-symmetric, there are only the entries of its upper triangular part in the last sum, i.e. according to (3.7) we have

$$c_{i,i+1} = \frac{1}{y_{i+1} - y_i}\sum_{j=1}^{i}\sum_{l=i+1}^{N_k}(y_l - y_j)r_{j,l}, \qquad i = 1,2,\ldots,N_k - 1. \tag{3.8}$$

The desired result follows from this representation, the condition $r_{j,l} \geqslant 0$, and the rigorous monotonicity of the vector y.

Thus, from Lemmas 3.1, 3.2 it follows that G is the Stieltjes matrix in the case $m = 2$ for the vectors (3.5). Thus, we have established the positive definiteness of the matrix G [4] and because of the representation $B = (G - U')G^{-1}(G - U)$ the positive definiteness of the matrix B. The result obtained can be represented as the theorem.

Theorem 3.2. If the conditions of Lemma 3.1 are satisfied for the matrix A and two test vectors are used, one of these vectors has constant components and the other has rigorously monotonic components, then the precondition matrix B is positive definite.

By Theorem 3.1 this assertion also holds when any pair of the vectors $y^{(1)}$, $y^{(2)}$, which are obtained by the nonsingular linear transformation of vectors of the form (3.5), are used.

We can carry out the analogous investigation for another pair of the test vectors:

$$y_k^{(1)} = e_k, \qquad y_k^{(2)} = (1,0,1,0,\ldots)', \qquad k = 2,3,\ldots,M \tag{3.9}$$

that are strongly linearly independent.

The tridiagonal matrix C that is determined by (2.4), (3.9) for $m=2$ has off-diagonal entries that are described by formula (3.8). However, Lemma 3.2 does not hold for this case. In order to analyse the precondition matrix (1.3) we formulate the generalized compensation condition (1.10), using the representation for the matrices G_k:

$$G_1 = D_1, \quad G_k = D_k - (1-\theta)L_k(G_{k-1}^{-1})^{(p)}U_{k-1} - \theta X_k, \quad k = 2,3,\ldots,M \tag{3.10}$$

where X_k are the tridiagonal matrices found from the conditions

$$X_k y_k^{(q)} = Q_k y_k^{(q)}, \quad q = 1,2 \tag{3.11}$$

and the matrix Q_k is determined by the equality

$$Q_k = L_k G_{k-1}^{-1} U_{k-1} = \{q_{i,j}^{(k)}\}.$$

It is evident that the matrices X_k are found from (3.11) uniquely and are related to C_k from (1.9) by

$$X_k = C_k + L_k(G_{k-1}^{-1})^{(p)}U_{k-1}$$

which ensures that the matrices G_k determined from (2.2) and (3.10) coincide.

We consider the off-diagonal entries of the matrix X_k, omitting the index k for brevity. We make an additional assumption that the entries of the matrix Q_k decrease with distance from the principal diagonal

$$q_{i,j} \geqslant q_{i,j+1} \geqslant q_{i-1,j+1} \geqslant 0, \quad j \geqslant i. \tag{3.12}$$

Remark 3.1. The assumption that these inequalities hold is reasonable since obviously they hold if G_{k-1} is the Stieltjes matrix and the entries of the matrices L_k and U_{k-1} are constants. In the general case the analysis of the additional restrictions to the initial matrix A is needed.

Lemma 3.3. If the entries of the symmetric matrix Q satisfy the conditions (3.12), the off-diagonal entries of the tridiagonal matrix X determined by formulae (3.9), (3.11) are nonnegative.

Proof. It is evident that the off-diagonal entries $x_{i,i+1}$ of the tridiagonal matrix X_k are expressed by the formula analogous to (3.8):

$$x_{i,i+1} = \frac{1}{y_{i+1} - y_i} \sum_{j=1}^{i} \sum_{l=i+1}^{N_k} s_{j,l}, \quad i = 1,2,\ldots,N_k - 1$$

where y_i are the components of the vector $y^{(2)}$ and $s_{j,l} = (y_l - y_j)q_{j,l}$ are the entries of the skew-symmetric matrix

$$S = QT - TQ, \quad T = \mathrm{diag}\{y_i\}.$$

We transform the expression for $x_{i,i+1}$ separately for odd and even i, taking into account that

$$y_{2i+1} - y_{2j+1} = y_{2i} - y_{2j} = 0, \quad y_{2i+1} - y_{2j} = 1, \quad y_{2i} - y_{2j+1} = -1.$$

For definiteness we assume that N_k is odd. The transformations for even N_k are analogous. We have for odd i:

$$x_{1,2} = \sum_{l=1}^{(N_k-1)/2} q_{1,2l}$$

$$x_{2i-1,2i} = \sum_{l=i}^{(N_k-1)/2} q_{1,2l} + \sum_{l=i}^{(N_k-1)/2} \sum_{j=1}^{i-1} (q_{2j+1,2l} - q_{2j,2l+1}), \qquad i = 2,3,\ldots,(N_k-1)/2.$$

In accordance with (3.12) we have $q_{2j+1,2l} \geqslant q_{2j+1,2l+1} \geqslant q_{2j,2l+1}$. The nonnegativity of $q_{i,j}$ entails the inequalities

$$x_{2i-1,2i} \geqslant 0, \qquad i = 1,2,\ldots,(N_k-1)/2. \tag{3.13}$$

The equality for even i follows from (3.12):

$$x_{2i,2i+1} = \sum_{j=1}^{i} q_{2j,N_k} + \sum_{l=i+1}^{(N_k-1)/2} \sum_{j=1}^{i} (q_{2j,2l-1} - q_{2j-1,2l}), \qquad i = 1,2,\ldots,(N_k-1)/2.$$

By the conditions (3.12), we have $q_{2j,2l-1} \geqslant q_{2j,2l} \geqslant q_{2j-1,2l}$, $q_{i,j} \geqslant 0$. Hence,

$$x_{2i,2i+1} \geqslant 0, \qquad i = 1,2,\ldots,(N_k-1)/2. \tag{3.14}$$

From this inequality and inequality (3.13) follows the nonnegativity of all the off-diagonal entries of the matrix X.

The nonnegativity of the off-diagonal entries of the matrices X_k leads to the nonpositivity of the off-diagonal entries of the matrices G_k that are determined by formulae (2.2), (3.10) for $0 \leqslant \theta \leqslant 1$, and in view of Lemma 3.1 G_k are the Stiltjes matrices. Hence we immediately obtain the result analogous to that stated in Theorem 3.2.

Theorem 3.3. If the conditions of Lemma 3.3 are satisfied, the precondition matrix B that is determined by formulae (1.3), (3.9)–(3.11) is positive definite for $0 \leqslant \theta \leqslant 1$.

It is evident that this assertion holds when any other pair of the test vectors that are obtained from (3.9) under the nonsingular linear transformation are used.

Remark 3.2. Theorems 3.2 and 3.3 establish the convergence of the iterative processes with the precondition matrices B for two different pairs of the test vectors. However, the estimates of the convergence rate of the iterations and its optimization for the parameter θ invite further investigation.

4. EXAMPLES OF NUMERICAL EXPERIMENTS

In order to experimentally study the efficiency of the implicit incomplete factorization methods depending on the number and character of test vectors we carried out the series of methodical calculations for the systems of five-point difference equations that approximate the Dirichlet problem in a square domain on the square grid with steps $h = 1/(N+1)$. In this case the matrices $D_k = \{-1,4,-1\}$ are tridiagonal, L_k and U_k are identity matrices. The block order of the matrix A and the order of each block

Table 1. Problem 1, $p = 3$, $m = 1$, $y = e$.

$N+1$	θ					
	0.0	0.2	0.4	0.6	0.8	1.0
8	4	4	4	4	4	4
16	6	6	6	6	6	6
32	10	10	9	9	9	9
64	19	18	17	15	13	13
128	35	33	30	87	23	19

Table 2. Spectral characteristics of the matrices $B^{-1}A$ for $p = 3$, $m = 1$, $y = e$.

$N+1$		θ					
		0.0	0.2	0.4	0.6	0.8	1.0
8	$\lambda_{\max}$	1.038	1.051	1.066	1.083	1.106	1.136
	$\lambda_{\min}$	0.824	0.855	0.888	0.925	0.963	1.000
	$\varkappa$	1.259	1.230	1.200	1.172	1.149	1.136
16	$\lambda_{\max}$	1.063	1.095	1.140	1.208	1.325	1.598
	$\lambda_{\min}$	0.422	0.471	0.536	0.632	0.784	1.000
	$\varkappa$	2.516	2.326	2.125	1.910	1.690	1.598
32	$\lambda_{\max}$	1.072	1.113	1.173	1.274	1.494	2.771
	$\lambda_{\min}$	0.140	0.163	0.197	0.258	0.400	1.000
	$\varkappa$	7.664	6.844	5.945	4.933	3.734	2.771
64	$\lambda_{\max}$	1.072	1.115	1.179	1.292	1.550	5.287
	$\lambda_{\min}$	0.038	0.045	0.056	0.076	0.130	1.001
	$\varkappa$	28.162	24.862	21.223	17.067	11.959	5.283
128	$\lambda_{\max}$	1.072	1.115	1.180	1.294	1.562	10.439
	$\lambda_{\min}$	0.010	0.012	0.014	0.020	0.035	1.001
	$\varkappa$	110.123	98.865	82.235	65.514	44.879	10.427

Table 3. Problem 1, $p = 3$, $m = 2$, $y_k^{(1)} = e$, $y_k^{(2)} = \{j\}$.

$N+1$	θ					
	0.0	0.2	0.4	0.6	0.8	1.0
8	4	4	4	4	5	5
16	6	6	6	5	5	8
32	10	10	9	9	8	11
64	19	18	16	15	13	13
128	35	32	30	27	22	15

Table 4. Problem 1, $p = 3$, $m = 2$, $y_k^{(1)} = e$, $y_k^{(2)} = \{(-1)^j\}$.

$N+1$	θ					
	0.0	0.2	0.4	0.6	0.8	1.0
8	4	4	4	4	4	4
16	6	6	6	6	6	6
32	10	10	9	9	9	9
64	19	18	16	15	13	13
128	35	33	30	27	23	18

Table 5. Problem 1, $p = 3$, $m = 2$, $y_k^{(1)} = e$, $y_k^{(2)} = \{\sin(k\pi/(N+1))\sin(j\pi/(N+1))\}$.

$N+1$	θ					
	0.0	0.2	0.4	0.6	0.8	1.0
8	4	4	4	4	4	4
16	6	6	6	5	5	6
32	10	10	9	9	8	8
64	19	17	16	15	13	10
128	35	32	30	27	22	11

Table 6. The number of iterations for the optimal values of θ: Problem 1, $p = 3$, $m = 2$.

$N+1$	a	b	c
32	7 (0.900 ÷ 0.950)	8 (0.930 ÷ 0.950)	6 (0.940 ÷ 0.950)
64	8 (0.982 ÷ 0.984)	12 (0.900 ÷ 0.990)	8 (0.980 ÷ 0.990)
128	9 (0.994 ÷ 0.996)	16 (0.992 ÷ 0.994)	9 (0.992 ÷ 0.996)

are equal to N. The calculations were carried out for $N = 7, 15, 31, 63, 127$, which correspond to the sizes of the steps $h = 2^{-k}$, $k = 3, \ldots, 7$. As the characteristics of the algorithm quality we considered the boundaries of the spectrum, $\lambda_{\max}$ and $\lambda_{\min}$, of the matrix $B^{-1}A$, the condition number $æ$ (calculated by the power method [5]), and the number of iterations $n(\varepsilon)$, which is necessary to obtain the condition

$$\frac{\|r^n\|_2}{\|r^0\|_2} \leqslant \varepsilon = 10^{-5}.$$

We carried out all the calculations to a double precision.

In the main test example (Problem 1) for which the experiments were carried out, the vector in the right-hand side of equation (1.1) is obtained by the Laplace equation with unit values of the solution at the boundary when the exact solution is $u(x_i, y_j) = 1$.

Table 7. Spectral characteristics of the matrices $B^{-1}A$ for $p = 3$, $m = 2$, $y_k^{(1)} = e$, $y_k^{(2)} = \{j\}$.

$N+1$		θ					
		0.0	0.2	0.4	0.6	0.8	1.0
8	λ_{max}	1.038	1.016	1.003	1.000	1.000	1.000
	λ_{min}	0.824	0.846	0.871	0.818	0.753	0.685
	$æ$	1.259	1.201	1.151	1.223	1.328	1.461
16	λ_{max}	1.063	1.017	1.000	1.000	1.000	1.000
	λ_{min}	0.422	0.453	0.496	0.561	0.556	0.366
	$æ$	2.516	2.242	2.015	1.782	1.798	2.732
32	λ_{max}	1.072	1.015	1.000	1.000	1.000	1.000
	λ_{min}	0.140	0.154	0.176	0.213	0.298	0.181
	$æ$	7.664	6.573	5.679	4.690	3.359	5.524
64	λ_{max}	1.072	1.000	1.000	1.000	1.000	1.000
	λ_{min}	0.038	0.042	0.049	0.061	0.091	0.090
	$æ$	28.162	23.568	20.340	16.339	10.947	11.111
128	λ_{max}	1.072	1.000	1.000	1.000	1.000	1.000
	λ_{min}	0.010	0.011	0.013	0.016	0.024	0.044
	$æ$	110.123	91.887	78.941	62.892	41.253	22.559

The initial approximation is taken as the function

$$u^0(x_i, y_j) = \left(a \sin\frac{i\pi}{N+1} \sin\frac{j\pi}{N+1}\right)^\alpha + b \tag{4.1}$$

$$a = 10, \quad b = 2, \quad \alpha = 2, \quad i,j = 1,\dots,N.$$

A large number of experimental results are given for this initial data in [6,7] for different algorithms. These results may be used to compare the efficiency of the methods considered in this paper.

Table 1 gives the number of iterations for the implicit algorithm IMIF3 (as designated in [6,7]) when the tridiagonal matrices G_k are used [$p = 3$ in formulae (2.2)] for different θ and when the 'classical' diagonal matrices C_k with single test vector $y_k^{(1)} = e$, $m = 1$, are used. The table allows us to analyse the efficiency of the algorithms.

Table 2 gives the spectral characteristics of the matrices $B^{-1}A$ for the same algorithm. The maximum eigenvalue, the minimum eigenvalue, and the condition number are given in each square from top to bottom.

Tables 3–5 present the number of iterations when different kinds of tridiagonal matrices G_k and C_k are used, which are determined by the two test vectors $y_k^{(1)}$ and $y_k^{(2)}$, i.e. for $m = 2$. The vectors $y_k^{(1)}$ for all cases were defined as $y_k^{(1)} = e$. In the first version the vectors $y_k^{(2)}$ were taken 'linear', i.e. $y_k^{(2)} = \{1,\dots,N\}$, in the second version

Table 8. Spectral characteristics of the matrices $B^{-1}A$ for $p=3$, $m=2$, $y_k^{(1)}=e$, $y_k^{(2)}=\{(-1)^j\}$.

$N+1$		θ					
		0.0	0.2	0.4	0.6	0.8	1.0
8	λ_{max}	1.038	1.050	1.064	1.081	1.102	1.132
	λ_{min}	0.824	0.854	0.888	0.924	0.962	1.000
	$\ae$	1.259	1.229	1.199	1.170	1.145	1.132
16	λ_{max}	1.063	1.092	1.133	1.195	1.306	1.568
	λ_{min}	0.422	0.470	0.535	0.630	0.781	1.000
	$\ae$	2.516	2.322	2.117	1.898	1.672	1.568
32	λ_{max}	1.072	1.109	1.164	1.259	1.466	2.699
	λ_{min}	0.140	0.162	0.196	0.257	0.397	1.000
	$\ae$	7.664	6.833	5.924	4.902	3.696	2.699
64	λ_{max}	1.072	1.111	1.710	1.275	1.519	5.137
	λ_{min}	0.038	0.045	0.055	0.075	0.128	1.001
	$\ae$	28.162	24.822	21.149	16.966	11.845	5.133
128	λ_{max}	1.072	1.111	1.171	1.277	1.531	10.135
	λ_{min}	0.010	0.011	0.014	0.020	0.034	1.001
	$\ae$	110.123	96.715	81.954	65.133	44.461	10.124

they were taken oscillating with alternating signs, i.e. $y_k^{(2)}=\{-1,1,-1,1,\ldots\}$, in the third version they were taken 'sinusoidal', i.e. $y_k^{(2)}=\{\sin(k\pi/(N+1))\sin(j\pi/(N+1)),$ $j=1,\ldots,N\}$. Thus, the 'total' test vector $y^{(2)}=\{y_k^{(2)}\}$ was the first eigenvector of the initial matrix. The results show that the use of the tridiagonal matrices substantially reduces the number of iterations [except when the test vector with alternating signs or the 'high-frequency' test vector $y_k^{(2)}$ is used].

Tables 3–5 demonstrate that the number of operations decreases as the compensating parameter θ increases. To analyse this dependence in greater detail, additional experiments were carried out for values of θ in a neighbourhood of unity. The results are given in Table 6 (the columns 'a', 'b', and 'c' are complimentary to Tables 3–5, respectively). In the table the minimum number of iterations and the range of values θ for which they are obtained are presented.

As is seen from the table, the number of operations for the optimal parameters θ is some 10–30 per cent less than that for $\theta=1$, and the optimal values of θ approach unity as the system order increases.

Tables 7–10 give the spectral characteristics analogous to those in Table 2 for the matrices $B^{-1}A$ that correspond to the experimental results in Tables 3–6. (In Table 10 the optimal parameters θ imply the parameters for which the values of $\ae$ are minimum).

Table 9. Spectral characteristics of the matrices $B^{-1}A$ for $p=3$, $m=2$, $y_k^{(1)}=e$, $y_k^{(2)}=\{\sin(k\pi/(N+1))\sin(j\pi/(N+1))\}$.

$N+1$		θ					
		0.0	0.2	0.4	0.6	0.8	1.0
8	$\lambda_{\max}$	1.038	1.024	1.016	1.012	1.018	1.025
	$\lambda_{\min}$	0.824	0.849	0.877	0.875	0.833	0.790
	$\varkappa$	1.259	1.207	1.158	1.157	1.222	1.298
16	$\lambda_{\max}$	1.063	1.020	1.002	1.000	1.000	1.022
	$\lambda_{\min}$	0.422	0.454	0.499	0.568	0.691	0.458
	$\varkappa$	2.516	2.245	2.006	1.760	1.446	2.230
32	$\lambda_{\max}$	1.072	1.016	1.001	1.000	1.000	1.015
	$\lambda_{\min}$	0.140	0.155	0.176	0.214	0.300	0.235
	$\varkappa$	7.664	6.575	5.674	4.674	3.330	4.316
64	$\lambda_{\max}$	1.072	1.000	1.000	1.000	1.000	1.000
	$\lambda_{\min}$	0.038	0.042	0.049	0.061	0.092	0.118
	$\varkappa$	28.162	23.564	20.332	16.326	10.925	8.481
128	$\lambda_{\max}$	1.072	1.000	1.000	1.000	1.000	1.000
	$\lambda_{\min}$	0.010	0.011	0.013	0.016	0.024	0.059
	$\varkappa$	110.123	91.884	78.933	62.879	41.232	16.866

Table 10. Spectral characteristics of the matrices $B^{-1}A$ for the optimal values of θ ($p=3$, $m=2$).

$N+1$	a		b		c	
32	1.000		2.280		1.000	
	0.400	0.894	0.908	0.978	0.414	0.898
	2.500		2.512		2.417	
64	1.000		3.869		1.000	
	0.255	0.963	0.877	0.992	0.262	0.964
	3.923		4.410		3.820	
128	1.000		7.341		1.000	
	0.161	0.987	0.911	0.998	0.162	0.987
	6.194		8.057		6.159	

We also carried out the experiments in which the five-diagonal matrices G_k and C_k were used, i.e. for $p=5$ and $m=3$. Tables 11 and 12 give the results for Problem 1 for $y_k^{(1)}=e$, $y_k^{(2)}=\{j\}$, $y_k^{(3)}=\{(-1)^j\}$ (i.e. for 'constant', 'linear', and 'high-frequency' test vectors).

Table 11. The number of iterations (Problem 1) for three test vectors ($p = 5$, $m = 3$).

$N+1$	θ					
	0.0	0.2	0.4	0.6	0.8	1.0
8	3	3	3	3	4	4
16	5	5	4	5	5	7
32	8	7	6	6	6	11
64	13	12	10	8	7	14
128	24	21	18	14	11	16

Table 12. Spectral characteristics of the matrices $B^{-1}A$ for three test vectors ($p = 5$, $m = 3$).

$N+1$		θ					
		0.0	0.2	0.4	0.6	0.8	1.0
8	λ_{max}	1.032	1.021	1.014	1.008	1.004	1.000
	λ_{min}	0.962	0.970	0.967	0.948	0.926	0.903
	$\varkappa$	1.073	1.052	1.049	1.063	1.083	1.107
16	λ_{max}	1.102	1.071	1.038	1.012	1.000	1.000
	λ_{min}	0.701	0.758	0.820	0.802	0.696	0.580
	$\varkappa$	1.571	1.412	1.265	1.262	1.436	1.725
32	λ_{max}	1.144	1.111	1.069	1.026	1.000	1.000
	λ_{min}	0.301	0.357	0.440	0.566	0.514	0.277
	$\varkappa$	3.801	3.109	2.430	1.814	1.945	3.607
64	λ_{max}	1.157	1.126	1.000	1.000	1.000	1.000
	λ_{min}	0.091	0.112	0.149	0.219	0.386	0.123
	$\varkappa$	12.777	10.019	6.713	4.557	2.592	8.114
128	λ_{max}	1.159	1.130	1.000	1.000	1.000	1.000
	λ_{min}	0.024	0.030	0.041	0.063	0.127	0.056
	$\varkappa$	48.627	37.653	24.555	15.855	7.846	17.949

Table 13. The number of iterations for optimal θ: ($p = 5$, $m = 3$).

$N+1$	16	32	64	128
$n(\varepsilon)$	4 (0.34 + 0.59)	5 (0.63 ÷ 0.75)	6 (0.82 ÷ 0.85)	8 (0.91 + 0.96)

The results of the experimental search for the optimal values of θ for the algorithm are presented in Table 13. We see that the dependence of the number of iterations on θ is about the same as that for the tridiagonal matrices C_k. However, we can see a substantial decrease in the optimal values of θ in Table 7.

Table 14. The number of iterations (Problem 1) for five-diagonal matrices G_k and the single test vector $y = e$ $(p = 5,\ m = 1)$.

$N+1$	θ					
	0.0	0.2	0.4	0.6	0.8	1.0
8	3	3	3	3	3	3
16	5	5	5	5	5	5
32	8	8	7	8	7	8
64	13	13	12	11	11	12
128	24	22	21	19	17	16

Table 15. Spectral characteristics of the matrices $B^{-1}A$ for five-diagonal matrices G_k and the single test vector $y = e$ $(p = 5,\ m = 1)$.

$N+1$		θ					
		0.0	0.2	0.4	0.6	0.8	1.0
8	$\lambda_{\max}$	1.032	1.037	1.041	1.046	1.051	1.056
	$\lambda_{\min}$	0.962	0.970	0.977	0.985	0.992	0.997
	$\varkappa$	1.073	1.069	1.066	1.062	1.059	1.059
16	$\lambda_{\max}$	1.102	1.128	1.160	1.201	1.259	1.351
	$\lambda_{\min}$	0.701	0.748	0.802	0.865	0.936	1.000
	$\varkappa$	1.571	1.509	1.447	1.389	1.345	1.351
32	$\lambda_{\max}$	1.144	1.194	1.265	1.376	1.585	2.210
	$\lambda_{\min}$	0.301	0.344	0.406	0.505	0.687	1.000
	$\varkappa$	3.801	3.468	3.113	2.724	2.309	2.210
64	$\lambda_{\max}$	1.157	1.216	1.303	1.448	1.766	4.154
	$\lambda_{\min}$	0.091	0.107	0.132	0.179	0.295	1.001
	$\varkappa$	12.777	11.389	9.857	8.110	5.990	4.152
128	$\lambda_{\max}$	1.159	1.220	1.311	1.467	1.819	8.160
	$\lambda_{\min}$	0.024	0.028	0.036	0.049	0.088	1.001
	$\varkappa$	48.627	43.023	36.817	29.680	20.791	8.155

In order to illustrate the efficiency of the three test vectors Tables 14 and 15 give the number of iterations and the spectral characteristics of the matrices $B^{-1}A$ when the five-diagonal matrices G_k and the diagonal matrices C_k, which are calculated by the single test vector $y = e$, are used.

We also carried out additional experiments to study the efficiency of the proposed algorithms in relation to the behaviour of the vector of the initial error $z^0 = u - u^0$. This dependence turns out to be sufficiently weak, which is seen from the results in Table 16 for Problem 2: the vector of the right-hand side, f, in the algebraic system is constant, its components are equal to $100h^2$ (h is a grid step), the zero vector $u^0 = 0$

Table 16. The number of iterations for Problem 2, $p=5$, $m=3$, $y_k^{(1)}=e$, $y_k^{(2)}=\{j\}$, $y_k^{(3)}=\{(-1)^j\}$.

$N+1$	θ					
	0.0	0.2	0.4	0.6	0.8	1.0
8	3	3	3	3	3	4
16	5	5	4	4	5	6
32	8	7	6	5	6	9
64	13	12	10	8	6	11
128	22	19	18	15	11	13

is taken as the initial approximation. The solution of this problem is of a quadratic character. We use here the same algorithm as that in Tables 11–15. We see that the number of iterations in this case is somewhat less than that in Tables 14 and 15.

Table 17 presents the spectral characteristics of the matrix $B^{-1}A$ when the same algorithm as that in Tables 11–16 ($p=5$, $m=3$) and the three test vectors from [3] with a cyclic variation of the numbers of unit components (set 'A') are used:

$$y^{(1)} = (1,0,0,1,0,0,\dots)'$$

$$y^{(2)} = (0,1,0,0,1,0,\dots)'$$

$$y^{(3)} = (0,0,1,0,0,1,\dots)'.$$

This table allows us to analyse the effect of the choice of different test vectors.

Finally, we give the experimental results for Problem 3 considered in [8], viz. for the difference five-point system for the Dirichlet homogeneous problem for the Laplace equation in the unit square on the square grid with step $h=1/40$. This problem has zero exact solution. The iterations were performed with the initial approximation $u^0=\{\exp(ih-jh)\}$. The results obtained for this data are given in [8]. The method of steepest descent with the preconditioner corresponding to the implicit alternating direction method is used there. The value $\|z^n\|_A/\|z^0\|_A$ was controlled, and the number of iterations required to obtain the precision $\varepsilon=10^{-4}$ turned out to be 49.

The calculations completely correspond to the above conditions. Table 18 gives the number of iterations for different values of p (the band width of the matrix G), m, the parameter θ, and the different types of the test vectors considered above. The rows of the table correspond to the following versions of the algorithm:

(1) $p=3$, $m=1$, $y=e$;

(2) $p=3$, $m=2$, $y^{(1)}=e$, $y^{(2)}$ – 'linear';

(3) $p=3$, $m=2$, $y^{(1)}=e$, $y^{(2)}$ – 'oscillating';

(4) $p=3$, $m=2$, $y^{(1)}=e$, $y^{(2)}$ – 'sinusoidal';

(5) $p=3$, $m=2$, $y^{(1)}$ – 'linear', $y^{(2)}$ – 'oscillating';

(6) $p=5$, $m=1$, $y^{(1)}=e$;

Table 17. Spectral characteristics for the set 'A' of three test vectors ($p = 5$, $m = 3$).

$N+1$		θ					
		0.0	0.2	0.4	0.6	0.8	1.0
8	λ_{max}	1.032	1.038	1.045	1.052	1.058	1.066
	λ_{min}	0.962	0.969	0.977	0.984	0.992	1.000
	$\varkappa$	1.073	1.071	1.070	1.068	1.067	1.066
16	λ_{max}	1.102	1.129	1.162	1.205	1.266	1.365
	λ_{min}	0.701	0.746	0.800	0.862	0.033	1.000
	$\varkappa$	1.571	1.512	1.454	1.398	1.357	1.365
32	λ_{max}	1.144	1.194	1.267	1.380	1.592	2.234
	λ_{min}	0.301	0.343	0.404	0.501	0.681	1.000
	$\varkappa$	3.801	3.481	3.135	2.752	2.336	2.234
64	λ_{max}	1.157	1.217	1.304	1.450	1.769	4.200
	λ_{min}	0.091	0.106	0.131	0.177	0.290	1.001
	$\varkappa$	12.777	11.439	9.942	8.215	6.089	4.198
128	λ_{max}	1.159	1.221	1.313	1.468	1.820	8.251
	λ_{min}	0.024	0.028	0.035	0.049	0.086	1.001
	$\varkappa$	48.627	43.218	37.156	30.096	21.184	8.246

Table 18. The number of iterations for Problem 3 for a different number and different types of the test vectors.

	θ					
	0.0	0.2	0.4	0.6	0.8	1.0
1	44	39	33	27	20	9
2	44	38	33	27	19	3
3	44	39	33	27	20	8
4	44	38	33	27	19	3
5	44	39	33	28	20	14
6	21	19	16	13	11	7
7	21	17	12	9	5	3
8	21	19	16	13	11	7

(7) $p = 5$, $m = 3$, $y^{(1)} = e$, $y^{(2)}$ – 'linear', $y^{(3)}$ – 'oscillating';

(8) $p = 5$, $m = 3$, the set of vectors 'A'.

We conclude the survey of the experimental results with the remark: by and large in the generalized compensation principle the problems of algorithm stability, the positive definiteness of the preconditioner, and the convergence of the iterative process remain unsolved. For example, when using the three test vectors with 'constant' $y^{(1)}$, 'linear' $y^{(2)}$, and 'quadratic' $y^{(3)} = \{j^2\}$ the iterative process does not converge, viz. the entries of the matrix C_k, which are large in magnitude, appear, the Stieltjes character of the matrices G_k is affected, and large round-off errors occur even in the case of the double precision of the machine computation.

5. SOME CONCLUSIONS

The above experimental evaluation allows us to make the following conclusions as to the efficiency of using the additional test vectors and increasing the band width of the matrix G.

(a) The introduction of a single additional test vector into the algorithm with tridiagonal matrices G ($p = 3$, $m = 2$) allows us to substantially decrease the condition number of the matrix $B^{-1}A$ and the number of iterations, and these indicators of efficiency have a pronounced minimum for the optimal values of the parameter θ, which turn out to be somewhat less than unity.

(b) The 'linear' or 'sinusoidal' test vectors lead to approximately the same positive effect, whereas the 'oscillating' vector (which can be considered to be a high-frequency one, i.e. it corresponds to the maximum eigenvalue of the matrix of the original system) does not practically speed up the convergence (in the following we do not consider it).

(c) As the order of the system increases, the optimal value of the iterative parameter θ increases (in Table 10 for $p = 3$, $m = 2$ the value θ varies from 0.894 to 0.998 on the grids of size from 32×32 to 128×128). Variations in the condition numbers of the matrices $B^{-1}A$, which correspond to the optimal values of θ, are approximately proportional to $N^{1/2}$. Note that in the 'classical' version $p = 3$, $m = 1$, $y = e$ (see Table 2), the value $\theta = 1$ is practically optimal on all the grids, and the condition numbers linearly depend on N.

(d) If the number of the test vectors increases from two to three and the band width of the matrix G from three to five, the number of iterations decreases by about 10–20 per cent. With allowance for the higher computational cost of a single iteration, the efficiency of the algorithms for $p = 3$, $m = 2$ and $p = 5$, $m = 3$ is considered to be approximately identical (the number of arithmetic operations on a single iteration at each node is $4p + 21$). With increasing p and m the optimal values of the parameter θ somewhat decrease.

(e) Not only the change in the number of the test vectors but also the change in their character can strongly change the spectral properties of the matrix $B^{-1}A$. Thus, for $m = 2$, $\theta = 1$ the spectrum is to the left of unity when the second vector is 'linear' and to the right of unity when it is 'oscillating' (see Table 8).

Thus, we can formulate the assertion that the use of additional test vectors is a possible means of increasing the efficiency of the incomplete factorization methods. However, new theoretical problems, which invite further investigation, arise here, viz. the choice (including the adaptive one) of the best test vectors, finding the optimal iterative parameters, the justification of nonsingularity of the precondition matrices, the study of algorithm stability, and the estimate of the convergence rate of the iterations.

REFERENCES

1. Avi Lin, Towards generalization and optimization of implicit methods. *Intern. J. Numer. Methods i* *Fluids* (1985) **5**, 357–380.
2. O. Axelsson and B. Polman, Block preconditioning and domain decomposition methods, I. Cathol University, *Report No. 8735*, Nijmegen, 1987.
3. O. Axelsson and B. Polman, Block preconditioning and domain decomposition methods. Cathol University, *Report No. 8807*, Nijmegen, 1988.
4. O. Axelsson, *Iterative Solution Methods*. Cambridge University Press, 1994.
5. D. K. Faddeev and V. N. Faddeeva, *Computational Methods of Linear Algebra*. Izd. Fiz. Mat. Liter, Moscow, Leningrad, 1963 (in Russian).
6. V. P. Il'in, *Iterative Incomplete Factorization Methods*. World Scientific Publishing Co., Singapore, 199
7. V. P. Il'in, *Incomplete Factorization Methods for Solving Algebraic Systems*. Nauka, Moscow, 1995 (i Russian).
8. A. A. Samarskii and E. S. Nikolaev, *Methods of Solving the Grid Equations*. Nauka, Moscow, 1978 (i Russian).
9. H. L. Stone, Iterative solution of implicit approximations of multidimensional partial differenti equations. *SIAM J. Numer. Anal.* (1968) **5**, 530–558.