Numer. Math. 48, 479–498 (1986)

Numerische Mathematik

On the Eigenvalue Distribution of a Class of Preconditioning Methods

Owe Axelsson[1] and Gunhild Lindskog[2]

[1] Department of Mathematics, University of Nijmegen, The Netherlands
[2] Department of Computer Sciences, Chalmers University of Technology, Göteborg, Sweden

Summary. A class of preconditioning methods depending on a relaxation parameter is presented for the solution of large linear systems of equations $Ax=b$, where A is a symmetric positive definite matrix. The methods are based on an incomplete factorization of the matrix A and include both pointwise and blockwise factorizations. We study the dependence of the rate of convergence of the preconditioned conjugate gradient method on the distribution of eigenvalues of $C^{-1}A$, where C is the preconditioning matrix. We also show graphic representations of the eigenvalues and present numerical tests of the methods.

Subject Classifications: AMS(MOS): 65F10; CR: G1.3.

1. Introduction

Discretization by the finite element method or a finite difference method of a selfadjoint partial differential equation of second order gives a system of linear equations

$$Ax=b, \tag{1.1}$$

where $x, b \in R^N$ and A is a symmetric, positive definite, sparse matrix of order $N \times N$.

We shall consider the iterative solution of the system (1.1) by the preconditioned conjugate gradient method. In [2] it is shown that this method converges to a relative error ε in the energy norm $\|x\|_{A^{1/2}}=\{x^TAx\}^{1/2}$, $x \in R^N$ in at most

$$\text{int}[\tfrac{1}{2}\{\kappa(C^{-1}A)\}^{1/2}\ln 2/\varepsilon+1] \tag{1.2}$$

iterations, where C is a symmetric, positive definite preconditioning matrix and $\kappa(C^{-1}A)$ is the spectral condition number of $C^{-1/2}AC^{-1/2}$.

A convenient choice of the matrix C is as a product of a lower and an upper triangular matrix i.e. $C=LL^T$ or $C=LU$, $U=DL^T$, where in the latter

case $\text{diag}(L)=I$. C can also be written $C=A+R$, where R is the so called defect matrix.

When $R=0$, i.e. $C=A$ we have a complete factorization of A, while $R\neq 0$ implies an incomplete factorization of A.

At each iteration step we have to solve the preconditioning system $Cx=y$, y a given vector. As we want the cost of this to be small the factor L is chosen as a sparse matrix. Examples of such incomplete factorizations are the pointwise Incomplete Cholesky (IC) method [15, 14, 12], and the Modified Incomplete Cholesky (MIC) methods [12, 8] and [4]. For the latter methods it has been shown that $\kappa(C^{-1}A)=O(h^{-1})$ for a wide class of problems, while $\kappa(A)=O(h^{-2})$. Here h is the mesh parameter defined in [6].

Incomplete block matrix factorization methods, BIC and MBIC (unmodified and modified, respectively) are presented and analysed in [10, 5] and [3]. There is evidence that they are more efficient than the pointwise methods.

Here we shall present and study a new class of modified incomplete factorization methods, the relaxed incomplete factorization methods. In the pointwise version we denote them RIC, Relaxed Incomplete Cholesky factorization and in the block matrix version, RBIC, the Relaxed Block Incomplete Cholesky factorization. This class contains a parameter ω and for $\omega=0$ the method reduces to the IC and BIC methods while for $\omega=1$ the method reduces to the MIC method without parameter and the MBIC method.

We shall in particular study the dependence of the rate of convergence on the distribution of eigenvalues of the matrix $C^{-1}A$ in the iterative solution of (1.1). The estimate (1.2) of the number of iterations is realistic when the eigenvalues of $C^{-1}A$ are relatively uniformly distributed over an interval $[a,b]$. In the case of isolated eigenvalues, however, it is sometimes possible to improve this bound, see [2, 1, 11, 13] and [16]. Here we present an alternative estimate of the number of iterations in the case of small isolated eigenvalues. A proof of this estimate is given in [7].

As is well known, the number of iterations depends also on the initial error. In all these estimates we assume the worst possible initial error.

We also study the dependence of the distribution of eigenvalues on the relaxation parameter ω and we present graphic representations of this dependence.

We also indicate why, in particular for the block methods, there is a tendency to a superlinear rate of convergence when ε decreases.

We finally show some numerical tests of the methods and make a comparison of the pointwise and blockwise relaxed methods. The improved performance of the block methods is again evidenced. Further we find that the relaxed methods may perform slightly better than the unrelaxed but that the optimal value of ω is always close to 1.

2. The Relaxed Incomplete Factorization Method

The relaxed (pointwise) incomplete Cholesky, RIC, factorization is a class of factorization methods based on the MIC factorization method.

Given a matrix A of order $N \times N$ the RIC method is as follows (we follow the notations in [4]):

Let J denote the set of indices (i,j), where fill-ins are permitted in the factorization.

For $r=1,2,\ldots,N-1$ we perform the following operations

$$l_{ir}=a_{ir}^{(r)}/a_{rr}^{(r)}$$

$$a_{ij}^{(r+1)}=\begin{cases} a_{ij}^{(r)}-l_{ir}a_{rj}^{(r)}, & (r+1\leqq j\leqq N)\cap((i,j)\in J)\cap i\neq j \\ 0, & (r+1\leqq j\leqq N)\cap((i,j)\notin J), \\ a_{ii}^{(r)}-l_{ir}a_{ri}^{(r)}+\omega\sum\limits_{\substack{p=r+1\\(i,p)\notin J}}^{N}(a_{ip}-l_{ir}a_{rp}^{(r)}), & j=i \end{cases} \tag{2.1}$$

where $i=r+1,r+2,\ldots,N$, $a_{ij}^{(1)}=a_{ij}$ and $0\leqq\omega\leqq 1$.

It is readily established (compare [14] where the case $\omega=0$ is treated) that a sufficient condition for existence of this factorization, i.e. that $a_{rr}^{(r)}\neq 0$, $r=1,2,\ldots,N-1$, is that for $\omega<1$, A is an M-matrix. For a sufficient condition for $\omega=1$, see [4].

The matrices L and $A^{(r+1)}$, $r=1,2,\ldots,N-1$ are completely defined when we add

$$l_{ij}=\begin{cases}0, & j>i\\ 1, & j=i\end{cases}$$

$$a_{ij}^{(r+1)}=\begin{cases}0, & j=1,\ldots,r;\ i=j+1,\ldots,N\\ a_{ij}^{(i)}, & i=1,\ldots,r;\ j=i,\ldots,N,\end{cases}$$

and the matrix U is defined by

$$u_{ij}=\begin{cases}0, & \text{for } j<i\\ a_{ij}^{(i)}, & \text{for } i=1,2,\ldots,N;\ j=i,i+1,\ldots,N.\end{cases}$$

The RIC factorization of A is then given by

$$C=LU, \tag{2.2}$$

where L and U are lower and upper triangular matrices respectively. (2.2) can also be written

$$C=A+R,$$

where the defect matrix R is given by

$$R=R_0+(1-\omega)\operatorname{diag}(-R_0),$$

where $R_0=\{r_{ij}^0\}$ depends on ω, and $\sum\limits_{j=1}^{N} r_{ij}^0=0$, $i=1,2,\ldots,N$. For diagonally dominant L-matrices A, R_0 is a negative semidefinite matrix.

It is easily seen that $\omega=0$ gives the IC factorization method and $\omega=1$ the MIC factorization method without parameter.

For simplicity throughout this paper the set J is chosen as $\{(i,j);\, a_{ij}\neq 0\}$.

We now describe the relaxed block incomplete Cholesky, RBIC, factorization method. This study will be limited to matrices of tridiagonal block matrix structure.

Hence let

$$A=\begin{bmatrix} D_1 & U_1 & & & 0 \\ L_1 & D_2 & U_2 & & \\ & \ddots & \ddots & \ddots & \\ & & & & U_{M-1} \\ 0 & & & L_{M-1} & D_M \end{bmatrix}. \tag{2.3}$$

A blockfactorization of A is given by

$$C=\begin{bmatrix} G_1 & & & 0 \\ L_1 & G_2 & & \\ & \ddots & \ddots & \\ 0 & & L_{M-1} & G_M \end{bmatrix}\begin{bmatrix} G_1^{-1} & & & 0 \\ & G_2^{-1} & & \\ & & \ddots & \\ 0 & & & G_M^{-1} \end{bmatrix}\begin{bmatrix} G_1 & U_1 & & 0 \\ & G_2 & U_2 & \\ & & \ddots & U_{M-1} \\ 0 & & & G_M \end{bmatrix}$$

$$=\begin{bmatrix} G_1 & U_1 & & 0 \\ L_1 & G_2+L_1 G_1^{-1}U_1 & U_2 & \\ & \ddots & \ddots & U_{M-1} \\ 0 & & L_{M-1} & G_M+L_{M-1}G_{M-1}^{-1}U_{M-1} \end{bmatrix} \tag{2.4}$$

for appropriate choices of the matrices G_i, $i=1,2,\ldots,M$, which are assumed to be nonsingular.

With the choice $G_1=D_1$, $G_i=D_i-L_{i-1}G_{i-1}^{-1}U_{i-1}$, $i=2,\ldots,M$, (2.4) is a complete (exact) blockfactorization of A.

Let $(G_i^{-1})^{(p)}$ denote the bandmatrix of G_i^{-1} with bandwidth p located symmetrically about the main diagonal. Further let $E_i=L_i(G_i^{-1}-(G_i^{-1})^{(3)})U_i$, $i=1,2,\ldots,M-1$. The RBIC method we shall study is then defined by

$$\begin{aligned} &G_1=D_1 \\ &G_i=D_i-L_{i-1}(G_{i-1}^{-1})^{(3)}U_{i-1}-\omega\hat{D}_{i-1}, \qquad i=2,3,\ldots,M, \end{aligned} \tag{2.5}$$

where $0\leqq\omega\leqq 1$ and $\hat{D}_{i-1}$ is a diagonal matrix satisfying

$$\hat{D}_{i-1}e=E_{i-1}e, \tag{2.6}$$

where $e=(1,1,\ldots 1)^T$.

Existence of the factorization for M-matrices follows from the general result in [3].

With the choice G_i, $i=1,2,\ldots,M$ in (2.5) the matrix C can be written as

$$C=A+R,$$

where the defect matrix R is given by

$$R=\begin{bmatrix} 0 & & & 0 \\ & \tilde{E}_1 & & \\ & & \tilde{E}_2 & \\ 0 & & & \ddots \\ & & & \tilde{E}_{M-1} \end{bmatrix},$$

and $\tilde{E}_i=E_i-\omega\hat{D}_i$. Hence, $\tilde{E}_i e=(1-\omega)\hat{D}_i e$ and $\omega=1$ gives a defect matrix with rowsums equal to zero.

The methods obtained by $\omega=0$ and $\omega=1$ are here also denoted by BIC and MBIC analogous to the pointwise incomplete factorizations. The case $\omega=1$ is treated in [10] under the name MINV(1).

More general choices of the vector e in (2.6) are discussed in [7].

3. On the Rate of Convergence of the Conjugate Gradient Method for the Case of Isolated Eigenvalues

Consider the solution of the system of linear equations (1.1). Let $S(C^{-1}A)$ denote the subset of disjoint eigenvalues of the positive definite matrix $C^{-1}A$ i.e. $S(C^{-1}A)=\bigcup_{i=1}^{\tilde{N}}\lambda_i$, where $0<\lambda_1<\lambda_2<\ldots<\lambda_{\tilde{N}}$.

Let $x^{(k)}$ denote the approximation of x calculated in the k-th iteration of the preconditioned conjugate gradient method.

An upper bound for the smallest number of iterations k such that

$$\|x-x^{(k)}\|_{A^{\frac{1}{2}}}\leqq\varepsilon\|x-x^{(0)}\|_{A^{\frac{1}{2}}} \tag{3.1}$$

for $\varepsilon>0$ (small) and an arbitrary initial approximation $x^{(0)}$ is given in (1.2), where $\kappa(C^{-1}A)=\lambda_{\tilde{N}}/\lambda_1$.

As was mentioned above (1.2) gives a realistic upper bound if the eigenvalues are relatively uniformly distributed over an interval $[a,b]$, $b>a>0$. For certain distributions of eigenvalues, however, the bound in (1.2) can be reduced.

Here we consider the case of isolated eigenvalues and we can distinguish two cases:

(a) $S(C^{-1}A)\in[a,b]\cup\left(\bigcup_{i=\tilde{N}-p+1}^{\tilde{N}}\lambda_i\right)$ where $1<p<\tilde{N}$ and $\lambda_{\tilde{N}-p}\leqq b<\lambda_{\tilde{N}-p+1}$,

(b) $S(C^{-1}A)\in\left(\bigcup_{i=1}^{p}\lambda_i\right)\cup[a,b]$ where $1<p<\tilde{N}$ and $\lambda_p<a\leqq\lambda_{p+1}$.

We may also consider a combination of these cases i.e.

(c) $S(C^{-1}A)\in\left(\bigcup_{i=1}^{p}\lambda_i\right)\cup[a,b]\cup\left(\bigcup_{i=\tilde{N}-p'+1}^{\tilde{N}}\lambda_i\right)$ where $1<p<\tilde{N}$, $1<p'<\tilde{N}$ and $\lambda_p<a\leqq\lambda_{p+1}<\lambda_{p+2}<\ldots<\lambda_{\tilde{N}-p'}\leqq b<\lambda_{\tilde{N}-p'+1}$.

Here it is assumed that p and p' are small and that the few largest (a), lowest (b) or both largest and lowest (c) eigenvalues are well separated from the remaining eigenvalues.

As we shall see, case (a) occurs for the MIC(MBIC) methods, case (b) occurs for IC(BIC) methods while case (c) occurs for RIC(RBIC) methods with $0<\omega<1$.

For the case (a) it is shown in [2] that the necessary number of iterations such that (3.1) is satisfied is at most

$$k=\text{int}\left[\frac{1}{2}\sqrt{\frac{b}{a}}\ln 2/\varepsilon+p+1\right]. \tag{3.2}$$

Thus if p and b are sufficiently small (3.2) gives a better bound than (1.2). It is easily seen that this estimate is close to a sharp one.

For the case (b) a less sharp estimate was given in [2]. An improved estimate of the rate of decay of the residual was given in [16] but not for the actual number of iterations.

We consider now this latter case closer. In [7] we show the following upper bound for the necessary number of iterations such that (3.1) is satisfied. A similar analysis is made in [13].

$$k=\left\lceil\left(\ln\frac{2}{\varepsilon}+\ln\prod_{i=1}^{p}\frac{c_{\text{opt}}}{\lambda_i}\right)\Big/\ln\sigma^{-1}\right\rceil+\rho(r_{\text{opt}}+1), \tag{3.3}$$

where $r_{\text{opt}\,0}$ is the smallest nonnegative integer r for which $\dfrac{c(r)}{c(r+1)}\leqq\sigma^{-1}$ where

$$c(r)=b\left(\tan\frac{\pi}{4(r+1)}\right)\Big/(r+1),\quad c_{\text{opt}}=c(r_{\text{opt}})\quad\text{and}\quad\sigma=\left(1-\sqrt{\frac{a}{b}}\right)\Big/\left(1+\sqrt{\frac{a}{b}}\right)$$

$\lceil a\rceil$ indicates the smallest integer greater than or equal to a.

4. Eigenvalue Distribution for the Model Problem

Let the $N\times N$ matrix A be the result of discretizing

$$\begin{aligned}-\Delta u&=f,\quad (x,y)\in\Omega\\ u&=0,\quad (x,y)\in\partial\Omega,\end{aligned} \tag{4.1}$$

where $\Omega=(0,1)\times(0,1)$, by linear finite element approximations over a uniform isosceles triangulation.

Here we consider the eigenvalue distribution of the matrix $C^{-1}A$ for the relaxed incomplete factorization methods described in Sect. 2.

Table 4.1. A comparison of the number of iterations for the RIC method using the upper bounds given in Sect. 3 for various distributions of eigenvalues

N	ω	$\frac{\lambda_{\max}}{\lambda_{\min}}$	Estimated number of iterations for distribution of eigenvalues in:			Actual number of iterations
			$[a,b]$	$[a,b]\cup\left(\bigcup_{i=1}^{p}\lambda_i\right)$	$\left(\bigcup_{i=1}^{p}\lambda_i\right)\cup[a,b]$	
49	0.0	3.07	15	–	11 ($p=1$)	10
	0.3 (opt)	2.77	14	–	11 ($p=1,2$)	9
	1.0	2.24	13	14 ($p=1,2$)	–	10
225	0.0	9.96	27	–	20 ($p=1,3$)	16
	0.76 (opt)	5.60	20	–	16 ($p=1$)	14
	1.0	4.46	18	18 ($p=2$)	–	15
961	0.0	37.48	52	–	38 ($p=1,3,4$)	29
	0.875 (opt)	14.28	32	–	25 ($p=1$)	20
	1.0	9.32	26	25 ($p=2,4$)	–	21

The eigenvalues are computed on band matrices by the sequence of NAG-routines: F01BUF, F01BVF, F01BWF and F02AVF.

The graphic output of the eigenvalues is made on a Calcomp 1051 drum plotter by use of the PLAM subroutine package. The computations are made on the IBM 3033N.

In Appendix, Figs. A.1–A.6, we show the eigenvalue distributions for the RIC and the RBIC factorization methods for the number of unknowns $N=49$, 225 and 961. For each N the eigenvalues are plotted for $\omega=0$, $\omega=\omega_{\text{opt}}$ and $\omega=1$. The value ω_{opt} is chosen as follows.

For N fixed, let $W\subset[0,1]$ denote the set of values of ω which gave the least number of iterations, k, in the solution of (4.1) with the preconditioned conjugate gradient method and the relative criterion $\|r^{(k)}\|_2\leqq\varepsilon\|r^{(0)}\|_2$, $\varepsilon=10^{-7}$. Then ω_{opt} satisfies $\|r^{(k)}_{\omega_{\text{opt}}}\|_2=\min_{\omega\in W}\{\|r^{(k)}\|_2\}$, where $r^{(k)}_{\omega_{\text{opt}}}$ is the residual in the k-th iteration with $\omega=\omega_{\text{opt}}$. The values of ω_{opt} for the above model problem and $x^{(0)}=0$ are found to be for the RIC method $\omega\approx 0.3$, 0.76, 0.875 and for the RBIC method $\omega\approx 0.3$, 0.7, 0.875 for $N=49$, 225 and 961 respectively.

In Tables 4.1 and 4.2 we give upper bounds of the number of iterations k for the relative error (3.1) with $\varepsilon=10^{-7}$, based on the eigenvalue distributions shown in Figs. A.1–A.6. We use the estimates of upper bounds given in (1.2), (3.2) and (3.3) for the case of equally spaced eigenvalues, isolated large eigenvalues and isolated small eigenvalues respectively.

In the last two cases we give smallest estimates of k for $p\geqq 1$. The results are compared with the actual number of iterations for the relative error (3.1) with $\varepsilon=10^{-7}$. We use problem (4.1) with the solution $u(x,y)=x(x-1)y(y-1)e^{xy}$. The initial approximation is chosen as $x^{(0)}(i)=\text{Random}\,[-1,1]$, $i=1,2,\ldots,N$. Since the solution x of system (1.1), used in (3.1), is not explicitly known, x is computed with the accuracy $\|r^{(k)}\|_2\leqq 10^{-10}$ before starting the iterations for the relative error (3.1).

Table 4.2. A comparison of the number of iterations for the RBIC method using the upper bounds given in Sect. 3 for various distributions of eigenvalues

N	ω	$\frac{\lambda_{max}}{\lambda_{min}}$	Estimated number of iterations for distribution of eigenvalues in:			Actual number of iterations
			$[a,b]$	$[a,b]\cup\left(\bigcup_{i=1}^{p}\lambda_i\right)$	$\left(\bigcup_{i=1}^{p}\lambda_i\right)\cup[a,b]$	
49	0.0	1.26	10	–	6 $(p=1)$	5
	0.3 (opt)	1.21	10	–	6 $(p=1)$	5
	1.0	1.14	9	10 $(p=1)$	–	5
225	0.0	2.52	14	–	10 $(p=1,2)$	8
	0.7 (opt)	1.80	12	–	9 $(p=1,2)$	7
	1.0	1.60	11	12 $(p=1,2)$	–	7
961	0.0	7.66	24	–	17 $(p=1)$	13
	0.875 (opt)	3.21	16	–	13 $(p=1,2)$	9
	1.0	2.77	14	15 $(p=1,2)$	–	11

Thus, in the case of isolated *small* eigenvalues, (i.e. $\omega<1$) we get a better estimate of the number of iterations by use of the upper bound (3.3) compared to that in (1.2). This improvement is particularly evident for large N.

5. On Superlinear Rate of Convergence of the Conjugate Gradient Method in the Presence of Clusterpoints

Consider the conjugate gradient method for a selfadjoint positive definite operator $A=I+B$, where I is the identity operator and where the eigenvalues μ_i of B cluster about zero (i.e. B is a compact operator). Note that because A is positive definite, we have $\mu_i \geqq -1+\delta$ for some $\delta>0$. Let $S(A)$ be the spectrum of A and let $e^{(k)}=x-x^{(k)}$ i.e. the error in iteration k. It is easily shown that

$$\|e^{(k)}\|_{A^{1/2}} \leqq \min_{P_k\in\Pi_k^1} \max_{\lambda\in S(A)} |P_k(\lambda)| \, \|e^{(0)}\|_{A^{1/2}},$$

where Π_k^1 denotes the set of polynomials P_k of degree k such that $P_k(0)=1$. Thus we have

$$\|e^{(k)}\|_{A^{1/2}} \leqq \max_{\lambda\in S(A)} |P_k(\lambda)| \, \|e^{(0)}\|_{A^{1/2}}$$

for any polynomial $P_k(\lambda)\in\Pi_k^1$. Let $\lambda_i=1+\mu_i$ be the eigenvalues of A and let's order the eigenvalues μ_i such that $|\mu_{j+1}|\leqq|\mu_j|$, $j=1,2,\ldots$. It follows from [17] that if we choose

$$P_k(\lambda)=\prod_{j=1}^{k}\left(1-\frac{\lambda}{\lambda_j}\right)$$

then

$$\max_{\lambda\in S(A)}|P_k(\lambda)| \leqq \delta_k^k \equiv \left(\frac{2}{k}\sum_{j=1}^{k}\frac{|\mu_j|}{1+\mu_j}\right)^k.$$

Since $|\mu_j|\to 0$, $j\to\infty$ we hence find $\delta_k\to 0$, $k\to\infty$.

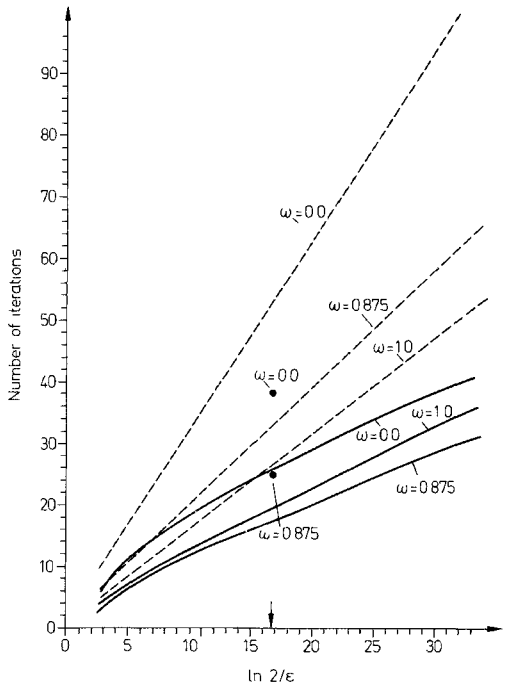

Fig. 5.1. The dependence of the number of iterations on ε, $\varepsilon = 10^{-i}$, $i = 1, 2, \ldots, 14$ for problem (4.1) with $N = 961$. The RIC method for $\omega = 0.0$, 0.875 and 1.0

But this means that

$$\left(\frac{\|e^{(k)}\|_{A^{1/2}}}{\|e^{(0)}\|_{A^{1/2}}}\right)^{1/k} \leqq \delta_k \to 0, \qquad k \to \infty,$$

which implies a superlinear rate of convergence. (For k large enough, the rate of convergence increases with k). For a finite dimensional problem, we have a finite spectrum so $k \to \infty$ doesn't make sense. However, the dimension of the matrices arising from discretized elliptic problems is frequently very large, so the arguments above make it plausible that the number of iterations should then increase less than linearly when the relative precision $\varepsilon \to 0$. To find this out consider now the problem (4.1), where $u(x, y) = x(x-1)y(y-1)e^{xy}$, solved by the RIC and RBIC preconditioned conjugate gradient methods for $N = 961$.

In the Figs. 5.1 and 5.2 we show how the number of iterations k depend on ε using the relative criterion

$$\|x - x^{(k)}\|_{A^{1/2}} \leqq \varepsilon \|x - x^{(0)}\|_{A^{1/2}}, \qquad \text{where } \varepsilon = 10^{-i},\ i = 1, 2, \ldots, 14.$$

The computations are made for $\omega = 0.0$, 0.875 and 1.0 and the initial approximation is $x^{(0)} = 0$.

We compare these results with the estimates of k given in (1.2). For $\varepsilon = 10^{-7}$ a comparison is made also with the estimate (3.3). (See also Table 4.1.)

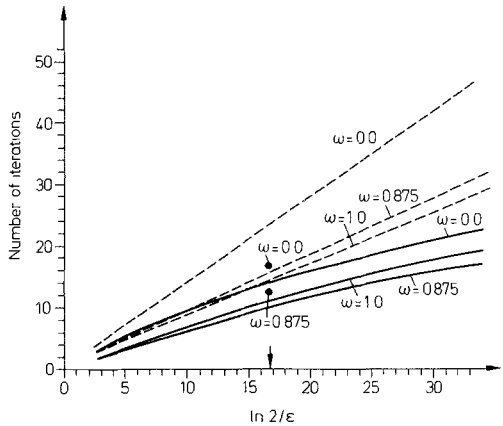

Fig. 5.2. The dependence of the number of iterations on ε, $\varepsilon = 10^{-i}$, $i = 1, 2, \ldots, 14$ for problem (4.1) with $N = 961$. The RBIC method for $\omega = 0.0$, 0.875 and 1.0

Legends Figs. 5.1, 5.2. Estimated number of iterations according to $k = \frac{1}{2}\{\kappa(C^{-1}A)\}^{\frac{1}{2}} \ln \frac{2}{\varepsilon} + 1$, dashed lines. Actual number of iterations computed by R(B)IC+CG method, bold lines. k in (3.3) for optimal value of p indicated by a point. Value of $\ln \frac{2}{\varepsilon}$ corresponding to $\varepsilon = 10^{-7}$ indicated by $\downarrow$ on the $\ln \frac{2}{\varepsilon}$ axis (compare with Tables 4.1 and 4.2).

From the graphic representations of the eigenvalues of $C^{-1}A$ (Appendix) we see that $C^{-1}A = I + B$, where B has a tendency of having a cluster point at zero. This is in particular true for the relaxed block incomplete factorization method. The Figs. 5.1 and 5.2 do indeed indicate such a superlinear rate of convergence: Calculating some extra correct digits by conjugate gradient iterations costs less and less.

Moreover for $\omega = 1.0$ we conclude that in the case $\varepsilon = 10^{-7}$, considered in Table (4.1), and even larger ε, the estimate (1.2) of k gives a realistic upper bound of k, thus explaining the absence of an improved upper bound by use of (3.2) in spite of the presence of isolated eigenvalues. For $\omega = 0.0$ and $\omega = 0.875$, however, the difference between the estimate (1.2) and the number of iterations computed by the RIC or RBIC preconditioned conjugate gradient method is greater, thus permitting an improved estimate also for $\varepsilon \geqq 10^{-7}$. With the upper bound (3.3) (see Table 4.1) and $\varepsilon = 10^{-7}$, the overshot in the estimate is about the half of that in (1.2).

6. Numerical Tests

In the testproblems we compare the RIC and RBIC preconditioning methods combined with the conjugate gradient method. The problems are discretized by linear finite element approximations and are defined on the unit square, i.e. $\Omega = \{(x, y);\ 0 < x < 1,\ 0 < y < 1\}$, where a uniform right-angled triangulation has been used. In the block methods $M = h^{-1} - 1$ where $h = (\sqrt{N} + 1)^{-1}$.

The initial approximation for the iterations is $x^{(0)}(i) = 0$, $i = 1, 2, \ldots N$ and we give the number of iterations k for the relative residual error $\|r^{(k)}\|_2 / \|r^{(0)}\|_2 \leqq 10^{-7}$.

Table 6.1. The number of iterations for the RIC method and various ω in the interval $[0, 1]$

N \ ω	0(=IC(0))	0.3	0.6	0.7	0.8	0.9	0.94	0.97	0.98	0.99	0.998	1(=MIC(0))
9	5	5	5	5	4	4	4	4	4	3	3	2
49	9	9	9	9	10	10	9	9	9	9	9	9
225	15	14	13	13	13	13	14	14	14	14	14	14
961	28	25	22	21	20	19	19	20	20	20	21	21
3969	54	–	–	41	–	32	30	28	27	28	30	33

Table 6.2. The number of iterations for the RIC method with $\omega = 1 - \delta h$, $\delta \geqq 0$

N \ δ	0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
9	2	4	4	4	4	5	5	5	5
49	9	9	9	10	10	10	10	10	10
225	14	14	14	14	14	14	14	14	13
961	21	20	20	20	20	19	19	19	19
3969	33	29	28	28	27	27	27	27	28

Table 6.3. Number of iterations for the RBIC method and different values of ω

N \ ω	0(=BIC)	0.3	0.6	0.8	0.9	0.93	0.95	0.98	1(=MBIC)
9	4	3	3	3	3	3	3	3	3
49	5	5	5	5	5	5	5	5	5
225	9	8	8	8	8	8	8	8	8
961	15	14	12	11	11	11	11	12	13
3969	28	26	22	19	17	17	16	16	18

A Model Dirichlet Problem

We consider the problem (4.1)

$$\begin{aligned} -\Delta u &= f, \quad (x, y) \in \Omega \\ u &= 0, \quad (x, y) \in \partial\Omega, \end{aligned}$$

where $u(x, y) = x(x-1)\, y(y-1) e^{xy}$.

In order to make a survey of the RIC method, the number of iterations for various ω in the interval $[0, 1]$ and various number of unknowns N is shown in Table 6.1. This permits a rough location of the optimum choice of ω for each N.

A detailed examination of the number of iterations for the special choice of ω, $\omega = 1 - \delta h$, $\delta \geqq 0$ is shown in Table 6.2. An optimal choice of δ is about 1.2.

In Table 6.3 we present the results for the block-factorization RBIC for various ω in the interval $[0, 1]$ and Table 6.4 show the number of iterations with the choice of ω, $\omega = 1 - \delta h$, $\delta \geqq 0$.

Table 6.4. Number of iterations for the RBIC method and $\omega = 1 - \delta h$ for various $\delta \geqq 0$

N \ δ	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
9	3	3	3	3	3	3	3	4	4
49	5	5	5	5	5	5	5	5	5
225	8	8	8	8	8	8	8	8	8
961	13	12	12	12	11	11	11	11	11
3969	18	17	16	16	16	16	16	16	17

Table 6.5. A comparison of the MIC(0), RIC and RBIC methods

N	Method						
	MIC(0) with optimal parameter	RIC			RBIC		
		$\omega = 0$ (IC)	$\omega = 1 - \delta_{opt} h$	$\omega = 1$ (MIC(0))	$\omega = 0$ (BIC)	$\omega = 1 - \delta_{opt} h$	$\omega = 1$ (MBIC)
9	5	5	5	2	4	3	3
49	10	9	10	9	5	5	5
225	14	15	14	14	9	8	8
961	20	28	19	21	15	11	13
3969	29	54	27	33	28	16	18

Thus, an optimal value of δ is about $\delta = 2.5$.

A comparison of the MIC(0) method with an optimal choice of parameter, the RIC and RBIC methods for $\omega = 0$, $\omega = 1 - \delta_{opt} h$ and $\omega = 1$ is given in Table 6.5.

Remark. For the larger values of h $\omega = 1 - \delta_{opt} h$ is greater than ω_{opt} as defined in Sect. 4.

A Problem with Discontinuous Material Coefficients

Here we have the problem

$$-\frac{\partial}{\partial x}\left(\lambda \frac{\partial u}{\partial x}\right) - \frac{\partial}{\partial y}\left(\lambda \frac{\partial u}{\partial y}\right) = 1, \quad (x, y) \in \Omega$$

$$u = 0, \quad (x, y) \in \partial\Omega$$

where

$$\lambda = \begin{cases} 1000 & \text{for } (x, y) \in \Omega_1 = \{(x, y); \tfrac{1}{3} \leqq x \leqq \tfrac{2}{3}, \tfrac{1}{3} \leqq y \leqq \tfrac{2}{3}\} \\ 1 & \text{for } (x, y) \in \Omega \setminus \Omega_1. \end{cases}$$

The number of iterations for the RIC method and various ω in the interval $[0, 1]$ is given in Table 6.6.

Table 6.6. Number of iterations for the RIC method and various ω and number of unknown N

N \ ω	0.0	0.3	0.6	0.7	0.8	0.87	0.9	0.93	0.97	0.99	0.995	1
25	9	9	10	10	9	9	9	9	9	8	8	7
121	16	16	16	15	16	15	16	17	17	17	17	14
529	27	26	24	23	23	22	22	23	23	25	25	24
2209	49	45	41	39	37	34	34	33	32	34	36	37

Table 6.7. Number of iterations for the RIC method, where $\omega = 1 - \delta h$, $\delta \geqq 0$

N \ δ	0	0.5	1	1.5	1.75	1.8	2.0	2.5	4.0
25	7	9	9	9	10	10	10	10	9
121	14	17	16	15	16	16	16	16	16
529	24	24	23	23	22	22	22	22	23
2209	37	34	32	32	32	32	33	33	33

Table 6.8. The number of iterations for the RBIC method and various ω

N \ ω	0	0.6	0.8	0.93	0.97	0.99	1
25	5	6	6	5	5	5	4
121	9	9	9	8	8	8	7
529	13	13	13	13	12	12	11
2209	24	20	18	18	18	20	17

Table 6.9. A comparison of MIC(0), RIC and RBIC

N	Method					
	MIC(0) with optimal parameter	RIC			RBIC	
		$\omega = 0$	$\omega = 1 - \delta_{opt} h$	$\omega = 1$	$\omega = 0$	$\omega = 1 (= \omega_{opt})$
25	10	9	10	7	5	4
121	16	16	16	14	9	7
529	23	27	22	24	13	11
2209	34	49	32	37	24	17

With the choice $\omega = 1 - \delta h$, $\delta \geqq 0$ we have the results shown in Table 6.7 for various number of unknowns. Hence, an optimal value of δ is about 1.8.

The number of iterations with the blockfactorization RBIC is given in Table 6.8. It shows that the optimal choice of ω in $[0, 1]$ is $\omega = 1$, i.e. the MBIC method.

A comparison of MIC(0) with optimal parameter, RIC and RBIC with $\omega = 0$, $1 - \delta_{opt} h$ and 1 is made in Table 6.9.

7. Comparisons and Conclusions

We now compare the computational complexity for the various methods in the solution of the system (1.1), arising from discretization by linear finite element approximations of a selfadjoint partial differential equation

$$-\frac{\partial}{\partial x}\left(a_1(x,y)\frac{\partial}{\partial x}u(x,y)\right)-\frac{\partial}{\partial y}\left(a_2(x,y)\frac{\partial}{\partial y}u(x,y)\right)=f(x,y), \qquad (x,y)\in\Omega$$
$$u(x,y)=g(x,y), \qquad (x,y)\in\partial\Omega$$

where $a_i(x,y)>0$, $i=1,2$ and Ω is the unit square. We assume we have a uniform triangulation, where the meshpoints are numbered in the rowwise way.

In the Tables 7.1 and 7.2 we give the factorization work for the RIC and RBIC methods counted in number of additions and multiplications. The work is presented in terms greater than or equal to $O(\sqrt{N})$.

In Table 7.3 we show the computational work per iteration in the solution by the preconditioned conjugate gradient method with the RIC and RBIC methods.

Table 7.1. The factorization work for the RIC method

Calculation	$\omega=0$		$\omega>0$	
	Add.	Mult.	Add.	Mult.
L	–	$2N-2\sqrt{N}$	–	$2N-2\sqrt{N}$
U	$2N-2\sqrt{N}$	$2N-2\sqrt{N}$	$4N-6\sqrt{N}$	$4N-6\sqrt{N}$
Total	$2N-2\sqrt{N}$	$4N-4\sqrt{N}$	$4N-6\sqrt{N}$	$6N-8\sqrt{N}$

Table 7.2. The factorization work for the RBIC method

Calculation		$\omega=0$		$\omega>0$	
		Add.	Mult.	Add.	Mult.
$(G_i^{-1})^{(3)}$, $i=1,\ldots,N-1$	Subdiagonals	$2N-6\sqrt{N}$	$6N-14\sqrt{N}$	$2N-6\sqrt{N}$	$6N-14\sqrt{N}$
	Main diagonal	$2N-4\sqrt{N}$	$3N-5\sqrt{N}$	$2N-4\sqrt{N}$	$3N-5\sqrt{N}$
$\hat{D}_i$, $i=1,\ldots,N-1$	Factorization of G_i and solution of $G_i x_1=U_i e$, where	–	–	$3N-6\sqrt{N}$	$5N-9\sqrt{N}$
	$x_1=L_i^{-1}(\hat{D}_i e+L_i(G_i^{-1})^{(3)}U_i e)$				
	$x_2=L_i x_1$	–	–	–	$N-\sqrt{N}$
	$\hat{D}_i e=x_2-L_i(G_i^{-1})^{(3)}U_i e$	–	–	$3N-5\sqrt{N}$	$6N-10\sqrt{N}$
ωD_i, $i=1,\ldots,N-1$		–	–	–	$N-\sqrt{N}$
$D_{i+1}-L_i(G_i^{-1})^{(3)}U_i-\omega\hat{D}_i$, $i=1,\ldots,N-1$		$3N-5\sqrt{N}$	$6N-10\sqrt{N}$	$4N-6\sqrt{N}$	–
Total for G_i, $i=1,\ldots,N$		$7N-15\sqrt{N}$	$15N-29\sqrt{N}$	$14N-27\sqrt{N}$	$22N-40\sqrt{N}$

Table 7.3. The computational work per iteration for the RIC and RBIC preconditioned conjugate gradient methods

Work	Method					
	RIC			RBIC		
	Computation	Add.	Mult.	Computation	Add.	Mult.
Solution of the preconditioning system	The solution of $Lx_1=y$, where $x_1=\mathrm{diag}(U)L^Tx$ $y_1=(\mathrm{diag}(U))^{-1}x_1-$	$2N-2\sqrt{N}$	$2N-2\sqrt{N}$ N	$x_2=y-Lx_1$, where $x_1=G^{-1}(G+U)x$ The solution of $Gx_1=x_2$ $x_3=x_2+Ux$	$N-\sqrt{N}$ $2N-2\sqrt{N}$ $N-\sqrt{N}$	$N-\sqrt{N}$ $3N-2\sqrt{N}$ $N-\sqrt{N}$
	The solution of $L^Tx=y_1$	$2N-2\sqrt{N}$	$2N-2\sqrt{N}$	The solution of $Gx=x_3$	$2N-2\sqrt{N}$	$3N-2\sqrt{N}$
Remaining computations in the conjugate gradient method	One matrix vector multiplication	$4N-4\sqrt{N}$	$5N+4\sqrt{N}$	One matrix vector multiplication	$4N-4\sqrt{N}$	$5N+4\sqrt{N}$
	Two scalar products and three $x+ky$, x,y vectors, k scalar	$5N$	$5N$	Two scalar products and three $x+ky$, x,y vectors, k scalar	$5N$	$5N$
	Total	$13N-8\sqrt{N}$	$15N$	Total	$15N-10\sqrt{N}$	$18N-2\sqrt{N}$

In the RIC method we solve the preconditioning system $Cx=y$, where $C=L\,\mathrm{diag}(U)L^T$.

In the RBIC method we have $C=(G+L)G^{-1}(G+U)$, where

$$G=\begin{bmatrix} G_1 & & 0 \\ & G_2 & \\ & \ddots & \\ 0 & & G_M \end{bmatrix}, \quad L=\begin{bmatrix} 0 & & 0 \\ L_1 & \ddots & \\ & \ddots & 0 \\ 0 & L_{M-1} & 0 \end{bmatrix} \quad \text{and} \quad U=\begin{bmatrix} 0 & U_1 & & 0 \\ & 0 & \ddots & \\ & & \ddots & U_{M-1} \\ 0 & & & 0 \end{bmatrix}$$

The number of multiplications per unknown to solve the model problem given in Sect. 6 for $N=225$, 961 and 3969 is given in Table 7.4. We show the number of multiplications per unknown for the factorization work, the iteration work with the relative accuracy $\varepsilon=10^{-1}$ (see Table 6.5) and for the total work. The total number of multiplications per unknown is also represented in Fig. 7.1.

The RBIC method does not only have a low computational complexity for model type problems, as is the case for multigrid methods for instance but seems in addition to be a very robust method, i.e. perform about as well on various types of problems.

Table 7.4. The number of multiplications per unknown for the various methods and number of unknowns N

Method	Work								
	Factorization			Iterations, $\varepsilon = 10^{-7}$			Total		
N	225	961	3969	225	961	3969	225	961	3969
RIC, $\omega = 0$	4	4	4	225	420	810	229	424	814
RIC, $\omega = \omega_{opt}$	6	6	6	195	285	405	201	291	411
RBIC, $\omega = 0$	13	14	15	161	269	504	174	283	519
RBIC, $\omega = \omega_{opt}$	20	21	22	143	198	288	163	219	310

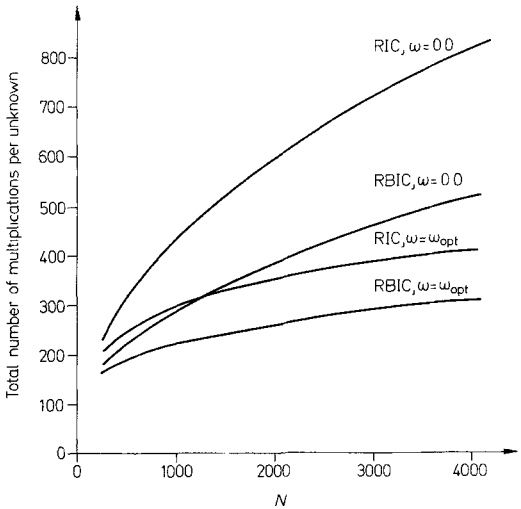

Fig. 7.1. The total number of multiplications per unknown to solve the model problem with the relative accuracy $\varepsilon = 10^{-7}$ for the various methods and number of unknowns

Note also that relaxation has the same effect as perturbation, i.e. adding a positive amount $O(h^2)$ to the diagonal of the original matrix before the incomplete factorization is started. This was used in [12] and in [4] to prove the reduction of the expression for the spectral condition number, from $O(h^{-2})$ to $O(h^{-1})$ after modification. For points on a line of discontinuity and points on a Neumann boundary condition line one adds $O(h)$. In [4] this addition is subject to a condition and is hence performed only in some cases. Beauwens [9] indicates that the perturbation may indeed not be needed at all.

It is interesting to note that relaxation by $\omega = 1 - \delta h$ for some positive δ is optimal for the block method and the model problem, but that for problems with discontinuous coefficients $\omega = 1$ is optimal.

Acknowledgement. Financial support for this study has been received from the National Swedish Board for Technical Development (STU), which is gratefully acknowledged.

Appendix

Graphic Representations of the Distributions of Eigenvalues for $C^{-1}A$ and the Model Problem

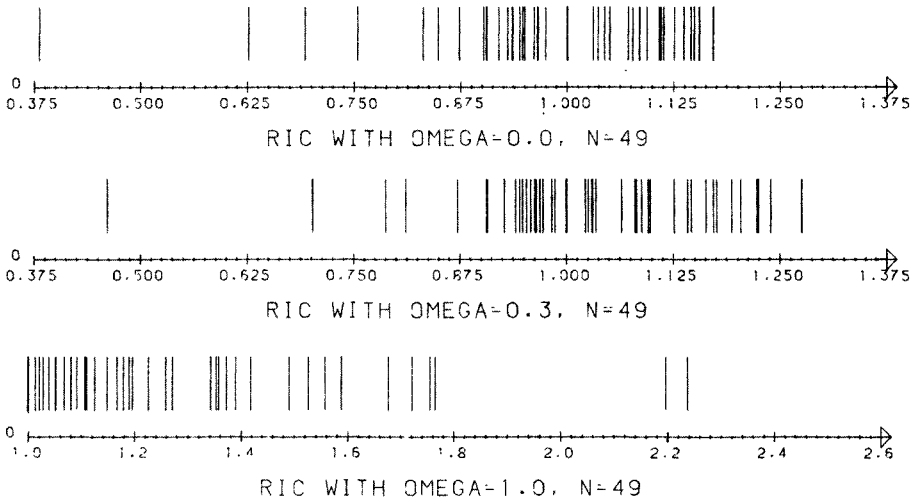

Fig. A.1. The eigenvalue distribution of $C^{-1}A$ for the RIC method and various choices of ω. $N = 49$

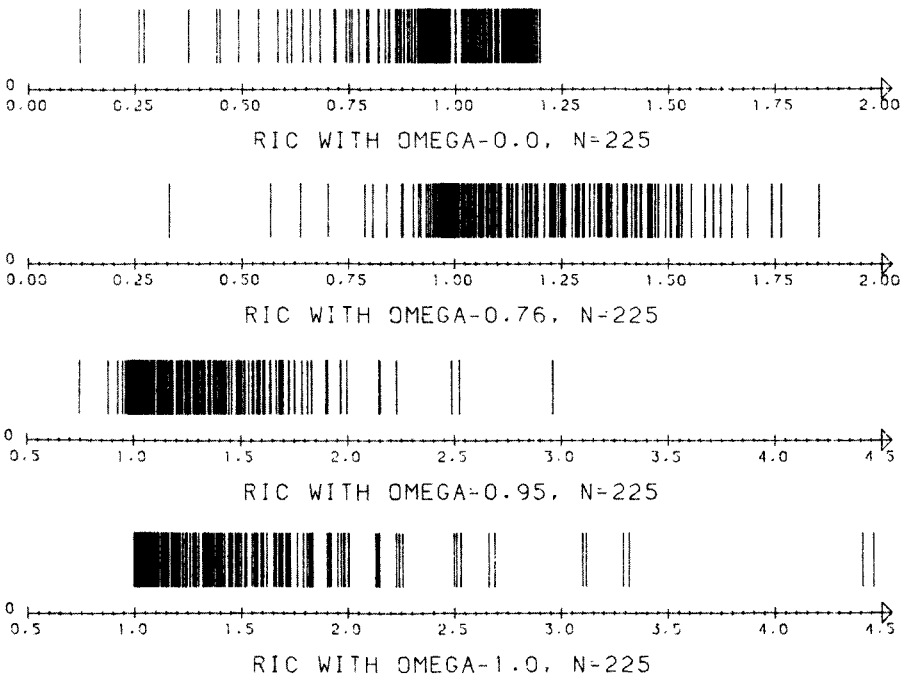

Fig. A.2. The eigenvalue distribution of $C^{-1}A$ for the RIC method and $N = 225$

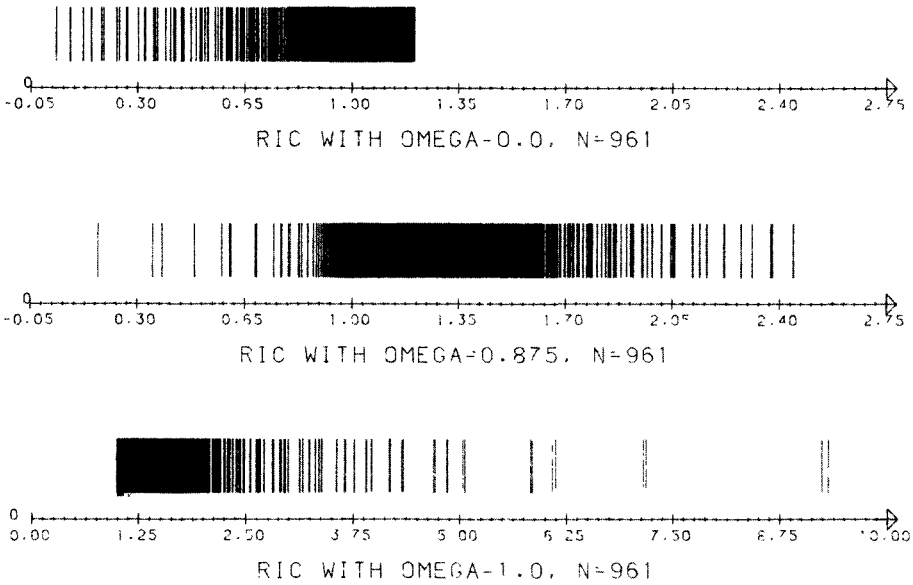

Fig. A.3. The eigenvalue distribution of $C^{-1}A$ for the RIC method and $N=961$

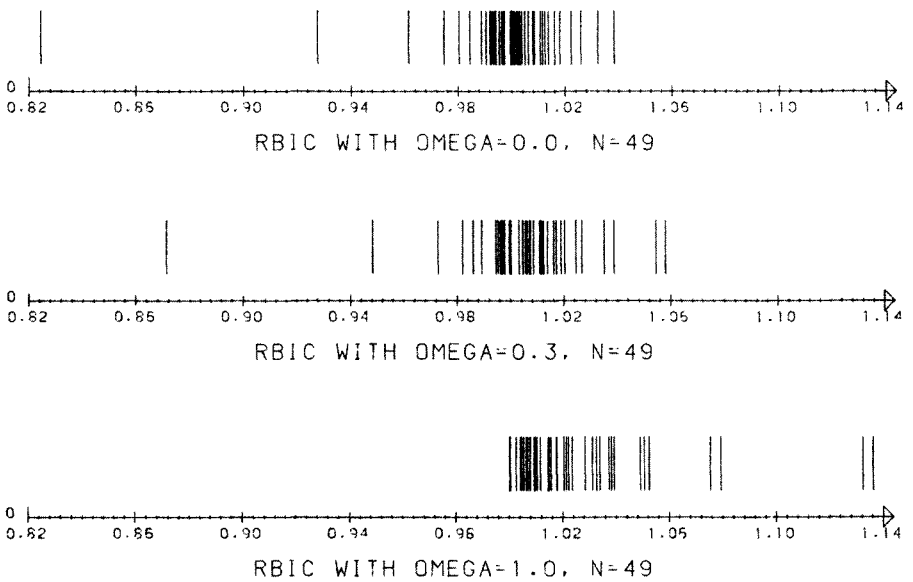

Fig. A.4. The eigenvalue distribution of $C^{-1}A$ for the RBIC method and $N=49$

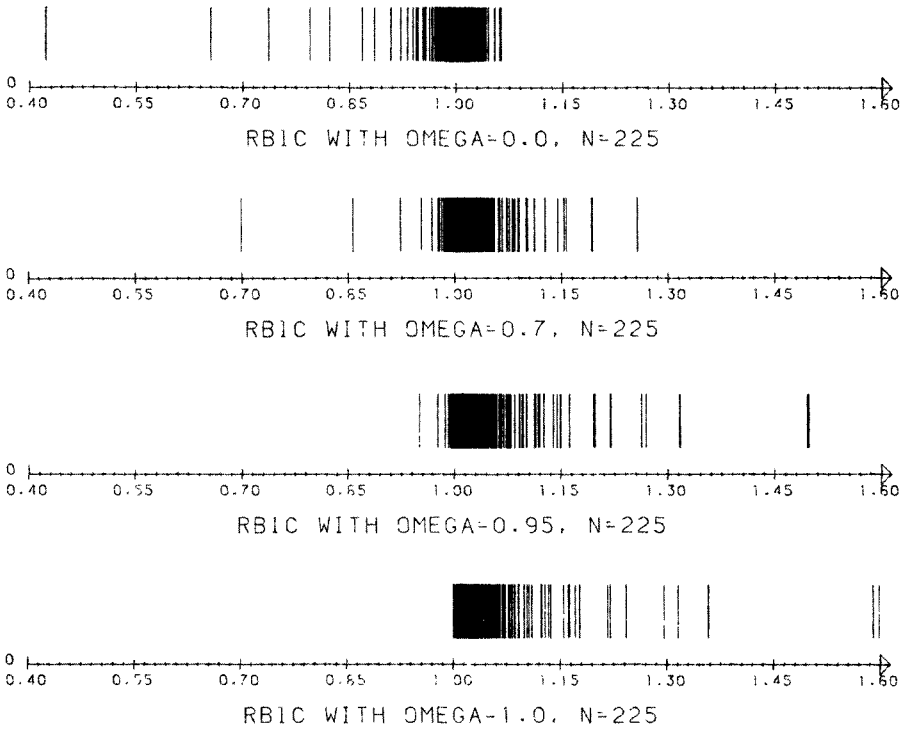

Fig. A.5. The eigenvalue distribution of $C^{-1}A$ for the RBIC method and $N=225$

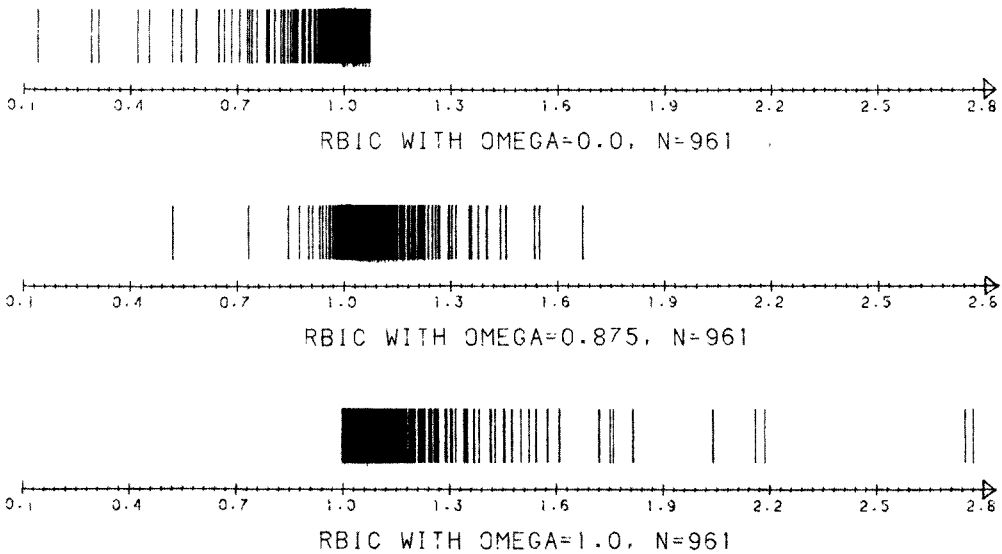

Fig. A.6. The eigenvalue distribution of $C^{-1}A$ for the RBIC method and $N=961$

References

1. Andersson, L.: SSOR preconditioning of Toeplitz matrices. Thesis, Chalmers University of Technology, Göteborg, Sweden, 1976
2. Axelsson, O.: A class of iterative methods for finite element equations. Comput. Methods Appl. Mech. Eng. **9**, 123–137 (1976)
3. Axelsson, O.: A general incomplete block-matrix factorization method. Linear Algebra Appl. (To appear)
4. Axelsson, O., Barker, V.A.: Finite Element Solutions of Boundary Value Problems. Theory and Computation. New York: Academic Press 1984
5. Axelsson, O., Brinkkemper, S., Il'in, V.P.: On some versions of incomplete block-matrix factorization iterative methods. Linear Algebra Appl. **58**, 3–15 (1984)
6. Axelsson, O., Gustafsson, I.: Preconditioning and Two-Level Multigrid Methods of Arbitrary Degree of Approximation. Math. Comput. **40**, 219–242 (1983)
7. Axelsson, O., Lindskog, G.: On the rate of convergence of the preconditioned conjugate gradient method. Numer. Math. (To appear)
8. Axelsson, O., Munksgaard, N.: A class of preconditioned conjugate gradient methods for the solution of a mixed finite element discretization of the biharmonic operator. Internat. J. Numer. Methods Eng. **14**, 1001–1019 (1979)
9. Beauwens, R.: On Axelssons perturbations. Linear Algebra Appl. **68**, 221–242 (1985)
10. Concus, P., Golub, G.H., Meurant, G.: Block preconditioning for the conjugate gradient method. SIAM J. Sci. Stat. Comput. **6**, 220–252 (1985)
11. Greenbaum, A.: Comparison of splittings used with the conjugate gradient algorithm. Numer. Math. **33**, 181–194 (1979)
12. Gustafsson, I.: Modified Incomplete Cholesky (MIC) Methods. In: Preconditioning Methods, Theory and Applications. (Evans, D.J., ed.), pp. 265–293. New York, London, Paris: Gordon and Breach Science 1983
13. Jennings, A.: Influence of the Eigenvalue Spectrum on the Convergence Rate of the Conjugate Gradient Method. J. Inst. Math. Appl. **20**, 61–72 (1977)
14. Meijerink, J.A., van der Vorst, H.A.: An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix. Math. Comput. **31**, 148–152 (1977)
15. Varga, R.S.: Factorization and normalized iterative methods. In: Boundary problems in differential equations. (R.E. Langer, ed.), pp. 121–142. Madison: Madison University of Wisconsin Press 1960
16. van der Vorst, H.A., van der Sluis, A.: The rate of convergence of conjugate gradients. (Preprint Nr. 354, November 1984, Department of Mathematics, University of Utrecht)
17. Winther, R.: Some superlinear convergence results for the conjugate gradient method. SIAM J. Numer. Anal. **17**, 14–17 (1980)

Received February 4, 1985/October 23, 1985