Stationary biparametric ADI preconditioners for conjugate gradient methods

A. Hadjidimosa\textsuperscript{a,b,*}, M. Lapidakis\textsuperscript{a,2}

\textsuperscript{a}Department of Mathematics, University of Crete, GR-714 09 Heraklion, Greece
\textsuperscript{b}Institute of Applied and Computational Mathematics, Foundation of Research and Technology, GR-711 10, Heraklion, Greece

Received 8 December 2005; received in revised form 15 March 2006

Abstract

In the present article we determine optimal stationary biparametric ADI preconditioners for the conjugate gradient methods when applied for the solution of a model problem second order elliptic PDE. The PDE is approximated by 5- and 9-point stencils. As was proved in Hadjidimos and M. Lapidakis \cite{http://www.math.uoc.gr/~hadjidim/hadlap05.ps} the problem of determining the best ADI preconditioner is equivalent to that of determining the optimal extrapolated (E) ADI method. So, analytic expressions are found for the optimal acceleration and extrapolation parameters for both discretizations where those for the 9-point stencil ones are new. Finally, numerical examples run using other well-known preconditioners show that the ADI ones we propose are very competitive.

© 2006 Elsevier B.V. All rights reserved.

MSC: 65F10

Keywords: (Extrapolated) Alternating direction implicit methods; Acceleration parameters; Extrapolation parameter; Conjugate gradient methods; Preconditioners; Spectral condition number

1. Introduction

In \cite{3} a limited number of ADI preconditioners for the class of conjugate gradient (CG) methods are given. This together with the fact that a simple class of ADI preconditioners reduces by an order of magnitude the spectral condition number of the unpreconditioned matrix coefficient of the problem at hand \cite{8} made us look into the more general case of using stationary biparametric ADI preconditioners. We thus expect that the ADI preconditioners will become more effective.

It is reminded that if we are given a linear system $Au = c$, with $A \in \mathbb{C}^{n \times n}$ Hermitian positive definite, and $c \in \mathbb{C}^n$, the CG method is most suitable for its solution. However, if the spectral condition number of $A$, that is $\kappa(A) = \lambda_{\text{max}}(A)/\lambda_{\text{min}}(A)$, with $\lambda_{\text{max}}(A)$ and $\lambda_{\text{min}}(A)$ denoting the largest and the smallest eigenvalues of $A$, is

\textsuperscript{*}Corresponding author. Department of Computer and Communication Engineering, Polytechnic School, University of Thessaly, 10 Iasonos Street, GR-383 33 Volos, Greece. Tel.: +30 24210 74908, +30 24210 74732; fax: +30 24210 74923.

E-mail addresses: hadjidim@inf.uth.gr (A. Hadjidimos), mihalis@math.uoc.gr (M. Lapidakis).

\textsuperscript{1}Part of the work of this author was funded by the Program Pythagoras of the Greek Ministry of Education.

\textsuperscript{2}The work of this author was done under a scholarship from the Greek State Scholarships Foundation.

0377-0427/S - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2006.05.033
large an appropriate preconditioner $M$, with $M \in \mathbb{C}^{n,n}$ Hermitian positive definite, is used such that $\kappa(M^{-1}A) = \lambda_{\max}(M^{-1}A)/\lambda_{\min}(M^{-1}A) \ll \kappa(A)$ (see, e.g., [5]).

For our analysis we consider the Poisson equation
\begin{equation}
- au_{xx}(x, y) - bu_{yy}(x, y) = f(x, y), \quad f \in C^2
\end{equation}
defined in the rectangle $\Omega := \{(x, y) \in \mathbb{R}^2|0 < x < l_1, 0 < y < l_2\}$, where $u(x, y)$ is sufficiently continuously differentiable and is subject to Dirichlet boundary conditions $u(x, y) = \gamma(x, y)$ on $\partial \Omega$, and $a$ and $b$ are positive constants. By imposing a uniform mesh of sizes $h_1$ and $h_2$ in $x$- and $y$-directions, respectively, on $\overline{\Omega}$ we approximate (1.1) at each internal node by the difference scheme
\begin{equation}
\begin{align*}
\sqrt{\frac{a}{b}} \frac{h_2}{h_1} (-u_{i-1,j} + 2u_{ij} - u_{i+1,j}) + \sqrt{\frac{b}{a}} \frac{h_1}{h_2} (-u_{i,j-1} + 2u_{ij} - u_{i,j+1}) \\
- \theta(4u_{ij} - 2(u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1})) + u_{i-1,j-1} + u_{i+1,j-1} + u_{i-1,j+1} + u_{i+1,j+1} = \frac{h_1 h_2}{\sqrt{ab}} (f_{ij} + \phi_{ij}).
\end{align*}
\end{equation}

(Note: It is natural to assume that $\sin(\pi h_i) < \cos(\pi h_i)$, $i = 1, 2$, since we always have in mind that $h_i \to 0$, $i = 1, 2$.) The parameters $\theta$ and $\phi$ in (1.2) take the values
\begin{equation}
(\theta, \phi) = \begin{cases} 
(0, 0), \\
(\theta^*, \phi^*) = \left( \frac{1}{12} \left( \sqrt{\frac{a}{b}} \frac{h_2}{h_1} + \sqrt{\frac{b}{a}} \frac{h_1}{h_2} \right), \frac{1}{12} \left( a h_1^2 f_{xx} + b h_2^2 f_{yy} \right) \right),
\end{cases}
\end{equation}
where for $\theta = 0, (1.2)$ gives the 5-point difference formula and for $\theta = \theta^*$ the difference formula is the 9-point one. The latter approximates (1.1) with an accuracy of $O(h^4)$, when $h_1 = h_2 = h$, see, e.g., [13,4]. (Note: In the case of Laplace’s equation the accuracy of the 9-point scheme is of order $O(h^6)$.) It is also pointed out that the discretized linear operator of the 9-point formula is positive definite if
\begin{equation}
\frac{1}{5} \leq \frac{bh_2}{ah_1} \leq 5,
\end{equation}
see [13], and so $\theta^* \in \left[ \frac{1}{5}, \frac{1}{2} \right]$. Adopting a natural ordering of the nodes starting from the bottom left corner and going eastwards the linear system obtained is of the form
\begin{equation}
Au = c,
\end{equation}
where, from the general formula (1.2), $A$ can be written as follows :
\begin{equation}
A = \sqrt{\frac{a}{b}} \frac{h_2}{h_1} (I_{n_2} \otimes T_{n_1}) + \sqrt{\frac{b}{a}} \frac{h_1}{h_2} (T_{n_2} \otimes I_{n_1}) - \theta \left[ \sqrt{\frac{a}{b}} \frac{h_2}{h_1} (I_{n_2} \otimes T_{n_1}) \cdot \sqrt{\frac{b}{a}} \frac{h_1}{h_2} (T_{n_2} \otimes I_{n_1}) \right].
\end{equation}
In (1.6) $n_1$ and $n_2$ are the numbers of internal nodes in $x$- and $y$-coordinate directions, $T_{n_1} \in \mathbb{R}^{n_1 \times n_1}$ and $T_{n_2} \in \mathbb{R}^{n_2 \times n_2}$ are of the form tridiag($-1, 2, -1$) and hence they are symmetric positive definite. Putting
\begin{equation}
A_1 := \sqrt{\frac{a}{b}} \frac{h_2}{h_1} (I_{n_2} \otimes T_{n_1}) \quad \text{and} \quad A_2 := \sqrt{\frac{b}{a}} \frac{h_1}{h_2} (T_{n_2} \otimes I_{n_1}),
\end{equation}
formula (1.6) becomes
\begin{equation}
A = A_1 + A_2 - \theta A_1 A_2,
\end{equation}
and a simple tensor product manipulation shows that $A_1$ and $A_2$ commute (see [9]).
2. Biparametric EADI schemes

Peaceman and Rachford [12] introduced the ADI methods while the extrapolated (E) ADI ones were introduced basically by Guittet [6]. An excellent account of the two-dimensional ADI methods can be found in Wachspers [16] (see also [17]).

In the sequel a slight modification of Guittet’s scheme to accommodate also for the biparametric EADI method is considered. Specifically,

\[(I + r_1 A_1)u^{(m+1/2)} = [(I + r_2 A_2)(I + r_1 A_1) - \omega A]u^{(m)} + \omega c,\]

\[(I + r_2 A_2)u^{(m+1)} = u^{(m+1/2)},\]

where \(A, A_1, A_2\) are given in (1.8) and (1.7), \(r_1, r_2 > 0\) are the acceleration parameters and \(\omega\) is the extrapolation parameter. (Note: Although the general \(\theta\), from (1.3), is used in (2.1) after our analysis the results for \(\theta = 0\) can be readily retrieved.) Eliminating \(u^{(m+1/2)}\) from the equations of (2.1) the following iterative scheme is obtained:

\[u^{(m+1)} = T_{EADI}u^{(m)} + c_{EADI},\]  (2.2)

where

\[T_{EADI} = I - \omega(I + r_1 A_1)^{-1}(I + r_2 A_2)^{-1} A, \quad c_{EADI} = (I + r_1 A_1)^{-1}(I + r_2 A_2)^{-1} \omega c.\]  (2.3)

Assuming that the eigenvalues \(\lambda_i\) of \(A_i, \ i = 1, 2\,\) belong to the rectangle

\[S := \{\lambda_1, \lambda_2 \in \mathbb{R}_+|\xi_1 \leq \lambda_1 \leq \beta_1, \xi_2 \leq \lambda_2 \leq \beta_2\},\]

where \(\xi_i, \beta_i \in \mathbb{R}_+, \ i = 1, 2\), then due to the commutativity property of \(A_1\) and \(A_2\), the eigenvalues of \(T_{EADI}\) are given by the expressions

\[\lambda_{T_{EADI}} = 1 - \omega\frac{\lambda_1 + \lambda_2 - \theta \lambda_1 \lambda_2}{(1 + r_1 \lambda_1)(1 + r_2 \lambda_2)}.\]  (2.4)

Denoting the fraction in (2.4) by \(f\) as a function of \(\lambda_1\) and \(\lambda_2\) we have

\[f \equiv f(\lambda_1, \lambda_2) := \frac{\lambda_1 + \lambda_2 - \theta \lambda_1 \lambda_2}{(1 + r_1 \lambda_1)(1 + r_2 \lambda_2)}.\]  (2.5)

Note that due to the positive definiteness assumption on \(A\) in (1.4) the numerator of \(f\) is positive. From (2.4) and (2.5) there holds

\[\rho(T_{EADI}) \leq \sup_{\lambda_1, \lambda_2 \in S} |1 - \omega f|.\]  (2.6)

To determine the maximum and the minimum values of \(f\), let them be denoted by

\[G := \max_{\lambda_1, \lambda_2 \in S} f \quad \text{and} \quad g := \min_{\lambda_1, \lambda_2 \in S} f,\]  (2.7)

we compute \(\partial f/\partial \lambda_i, \ i = 1, 2\), and obtain

\[\frac{\partial f}{\partial \lambda_i} = \frac{\lambda_i (1/\lambda_i - \theta - r_i)}{(1 + r_i \lambda_i)^2 (1 + r_j \lambda_j)}, \ i \neq j = 1, 2.\]  (2.8)

The expressions in (2.8) for the partial derivatives are of constant sign independent of the corresponding \(\lambda_i\) therefore, \(G\) and \(g\) are assumed at vertices of the rectangle \(S\). Hence among \(f(\xi_1, \xi_2), f(\xi_1, \beta_2), f(\beta_1, \xi_2)\) and \(f(\beta_1, \beta_2)\), \(G\) and \(g\) are to be sought.

It is obvious that for the solution of (1.5) by using the EADI method (2.1), the preconditioner

\[M = \frac{1}{\omega}(I + r_2 A_2)(I + r_1 A_1)\]  (2.9)
is used. Therefore, the optimal values for the parameters of $r_1$, $r_2$, $\omega$, let them be $r^*_1$, $r^*_2$, $\omega^*$ (see, e.g., [6]) will be found by minimizing the ratio $G/g$. Let $G^*$ and $g^*$ be the corresponding optimal values for $G$ and $g$, so that
\[
\frac{G^*}{g^*} = \min_{r_1, r_2 \in (0, \infty)} \frac{G}{g},
\] (2.10)
in which case the optimal value for $\omega$ will be
\[
\omega^* = \frac{2}{G^* + g^*}.
\] (2.11)
For the corresponding preconditioned CG method, the optimal preconditioner will be that for which $\kappa(M^{-1}A) = \lambda_{\text{max}}(M^{-1}A)/\lambda_{\text{min}}(M^{-1}A)$ is minimized. However, since the extrapolation parameter $\omega$ is simplified and thus has no effect on the preconditioned CG, this happens when the ratio $G/g$ is minimized. Hence, the solution to the optimal EADI problem will also give the solution to the optimal preconditioned CG one. Specifically we have proved that:

**Theorem 2.1.** Under the notation and the assumptions so far, the optimal (smallest) spectral condition number of the discretized Poisson problem (1.5), using the ADI preconditioner (2.9) for the CG method, is obtained for the (optimal) values of the acceleration parameters $r_1 = r^*_1$ and $r_2 = r^*_2$ that optimize the corresponding EADI problem in (2.10). Hence it is given by
\[
\kappa^*(M^{-1}A) = \frac{G^*}{g^*}.
\] (2.12)

3. Determination of the expressions for $G$ and $g$

To simplify our analysis, we assume that $1/\beta_1$, $1/\beta_2 > \theta$ so that
\[
0 < \frac{1}{\beta_i} - \theta < \frac{1}{x_i} - \theta, \quad i = 1, 2.
\] (3.1)
Obviously, (3.1) are satisfied for $\theta = 0$ while they may be not for all $\theta = \theta^*$. Cases that may arise where (3.1) are not satisfied will be examined in Section 5. Let $V_{x_1,z_2}$, $V_{\beta_1\beta_2}$, $V_{\beta_1\beta_2}$, $V_{x_1\beta_2}$ be the four vertices of the rectangle $S$. Since the extreme values of $f$ are assumed at vertices of $S$, to find them, we differentiate $f$ along its sides. The signs of these partial derivatives along each side of $S$ are shown in Table 1. For example, in case $r_j \in (0, 1/\beta_j - \theta)$, $i \neq j = 1, 2$, it is readily found out from Table 1 that $f$ increases on the sides $V_{x_1\beta_2}V_{\beta_1\beta_2}$ and $V_{\beta_1\beta_2}V_{\beta_1\beta_2}$ as well as on $V_{x_1\beta_2}V_{x_1\beta_2}$ and $V_{x_1\beta_2}V_{\beta_1\beta_2}$. As a result we have that $G = f(\beta_1, \beta_2)$ and $g = f(x_1, x_2)$. This is shown in the bottom left cell of Table 2. In the same way all other eight cases are examined, using the signs of the partial derivatives in Table 1, and the final results are illustrated in Table 2.

The only case that needs a further investigation is the one where $(r_1, r_2)$ belongs to the region
\[
(r_1, r_2) \in ABCD := \left[ \frac{1}{\beta_2} - \theta, \frac{1}{x_2} - \theta \right] \times \left[ \frac{1}{\beta_1} - \theta, \frac{1}{x_1} - \theta \right],
\] (3.2)
which will be examined a little later.

Before we go on with our analysis we introduce the symbol “$\sim$” and write
\[
E_1 \sim E_2
\] (3.3)
to denote that the two expressions $E_1$ and $E_2$ are of the same sign.

Let us see now how the ratio $G/g$ behaves when we are in one of the remaining eight cells defined in Table 2 and moving in either $r_1$- or $r_2$-direction keeping the other parameter fixed. For example, let us consider the cell
\[
r_1 \in \left( 0, \frac{1}{\beta_2} - \theta \right], \quad r_2 \in \left[ \frac{1}{x_1} - \theta, \infty \right).
\]
Table 1
Signs of $\frac{\partial f}{\partial r_i}$, $i = 1, 2$, along the sides of the rectangle $S$

<table>
<thead>
<tr>
<th>$r_1$</th>
<th>$\frac{1}{\beta_2} - \theta$</th>
<th>$\frac{1}{\alpha_2} - \theta$</th>
<th>$r_2$</th>
<th>$\frac{1}{\alpha_1} - \theta$</th>
<th>$\frac{1}{\alpha_1} - \theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{\partial f(V_{\alpha_1\beta_2}V_{\beta_1\beta_2})}{\partial r_1}$</td>
<td>$+$</td>
<td>$-$</td>
<td>$\frac{\partial f(V_{\alpha_1\alpha_2}V_{\beta_1\beta_2})}{\partial r_2}$</td>
<td>$+$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\frac{1}{\beta_2} - \theta$</td>
<td>$+$</td>
<td>$-$</td>
<td>$\frac{1}{\alpha_2} - \theta$</td>
<td>$+$</td>
<td>$-$</td>
</tr>
</tbody>
</table>

Table 2
Maximum and minimum of $f$ in each subregion of the first quadrant where $(r_1, r_2)$ lies

<table>
<thead>
<tr>
<th>$r_2$</th>
<th>$+\infty$</th>
<th>$\frac{1}{\alpha_1} - \theta$</th>
<th>$\frac{1}{\beta_1} - \theta$</th>
<th>$0$</th>
<th>$\frac{1}{\beta_2} - \theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G$</td>
<td>$f(\beta_1, \alpha_2)$</td>
<td>$f(\beta_1, \alpha_2)$</td>
<td>$f(\alpha_1, \beta_2)$</td>
<td>$g = f(\alpha_1, \alpha_2)$</td>
<td>$g = f(\alpha_1, \beta_2)$</td>
</tr>
<tr>
<td>$g$</td>
<td>$f(\alpha_1, \beta_2)$</td>
<td>$f(\beta_1, \beta_2)$</td>
<td>$g = f(\beta_1, \beta_2)$</td>
<td>$g = f(\alpha_1, \alpha_2)$</td>
<td>$g = f(\beta_1, \beta_2)$</td>
</tr>
<tr>
<td>$A$</td>
<td>$B$</td>
<td>$D$</td>
<td>$C$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Then we will have

$$G = f(\beta_1, \alpha_2) = \frac{(\beta_1 + \alpha_2 - \theta \alpha_1 \beta_2)(1 + \alpha_1 \beta_1)(1 + \beta_2 r_2)}{(\alpha_1 + \beta_2 - \theta \beta_1 \alpha_2)(1 + \beta_1 \beta_1)(1 + \alpha_2 r_2)}.$$

Taking partial derivatives first with respect to $r_1$ and then with respect to $r_2$ and exploiting the notation introduced in (3.3) we have

$$\frac{\partial(G/g)}{\partial r_1} \sim \alpha_1 - \beta_1 < 0, \quad \frac{\partial(G/g)}{\partial r_2} \sim \beta_2 - \alpha_2 > 0.$$

The first result shows that if we keep $r_2$ fixed then increasing $r_1$ the ratio $G/g$ is minimized when $r_1$ is maximized that is when $r_1 = 1/\beta_2 - \theta$. On the other hand, the second result implies that as $r_2$ increases the ratio in question increases and so it is minimized for $r_2 = 1/\alpha_1 - \theta$. Therefore, for all pairs $(r_1, r_2)$ in the cell considered, the ratio we are interested in is minimized at the right bottom corner of the region and this is the vertex $A$ of the region $ABCD$ in (3.2).

A similar examination in the other three cells (top right, bottom right and bottom left) defined in Table 2 show that the minimum of $G/g$ takes place at the points $B$, $C$, $D$, respectively. For the four cells that share a common boundary
with \( \text{ABCD} \) things are a little different. For example, let us consider the cell

\[
\begin{align*}
    r_1 &\in \left[ \frac{1}{\beta_2} - \theta, \frac{1}{\beta_2} - \theta \right], \\
    r_2 &\in \left[ \frac{1}{\alpha_2} - \theta, \infty \right].
\end{align*}
\]

Taking partial derivatives of

\[
\frac{G}{g} = \frac{f(\beta_1, \alpha_2)}{f(\beta_1, \beta_2)} = \frac{(\beta_1 + \alpha_2 - \theta \beta_1 \alpha_2)(1 + \beta_2 r_2)}{(\beta_1 + \beta_2 - \theta \beta_1 \beta_2)(1 + \alpha_2 r_2)},
\]

with respect to \( r_2 \) only, since the ratio in question is independent of \( r_1 \), we have

\[
\frac{\partial (G/g)}{\partial r_2} \sim \beta_2 - \alpha_2 > 0,
\]

and so the ratio \( G/g \) decreases as \( r_2 \) decreases and its minimum takes place for \( r_2 = 1/\alpha_1 - \theta \), that is on \( AB \). In the same way we can find that the minimum of this ratio when the cell is one of the other three adjacent to \( \text{ABCD} \) is minimized on \( BC, CD, DA \), respectively.

Consequently, the conclusion we arrive at is that the overall minimum of \( G/g \) is in \( \text{ABCD} \). To find it we have to decide which of the two expressions in each pair of braces, in Table 2, represents \( G \) and \( g \). For this we form the differences

\[
Q(r_1, r_2) = f(\alpha_1, \beta_2) - f(\alpha_1, \alpha_2) \quad \text{and} \quad q(r_1, r_2) = f(\beta_1, \beta_2) - f(\alpha_1, \alpha_2)
\]

and study the sign of each one in turn.

We begin with

\[
Q(r_1, r_2) \sim r_1 r_2 [\beta_2 \alpha_2 (\beta_1 - \alpha_1) - \beta_1 \alpha_1 (\beta_2 - \alpha_2)] + (r_1 - r_2)[\beta_1 (\beta_2 - \alpha_2) + \alpha_2 (\beta_1 - \alpha_1)]
\]

\[
- \theta r_1 \beta_1 \alpha_1 (\beta_2 - \alpha_2) + \theta r_2 \beta_2 \alpha_2 (\beta_1 - \alpha_1)
\]

\[
- \beta_1 (\beta_2 - \alpha_2) + \beta_2 (\beta_1 - \alpha_1) + (\beta_2 - \alpha_2) - (\beta_1 - \alpha_1).
\]

Assuming that the coefficient of the product \( r_1 r_2 \) is different from zero, the function in (3.5) represents a part of a one-sheet hyperboloid with level curves hyperbolas.

**Note:** If the coefficient of \( r_1 r_2 \) is zero, that is if \( \beta_2 \alpha_2 (\beta_1 - \alpha_1) = \beta_1 \alpha_1 (\beta_2 - \alpha_2) \), or \( 1/\alpha_1 - 1/\beta_1 = 1/\alpha_2 - 1/\beta_2 \) or, equivalently,

\[
\frac{\cot(\pi h_1) \sin(\pi h_2)}{\cot(\pi h_2) \sin(\pi h_1)} = \frac{a^2 h_4^4}{b^2 h_3^4},
\]

as for example in the case \( \alpha_2 = \alpha_1 \) and \( \beta_2 = \beta_1 \), then the hyperboloid is a plane and the level curves are straight lines. Otherwise nothing changes in the analysis that follows. So, from now on when we use the term hyperbola we will include the case of the straight line as well.

Next we examine the sign of the function \( Q(r_1, r_2) \) at each vertex of the rectangle \( \text{ABCD} \). At vertex \( A \) we have

\[
\begin{align*}
    Q \left( \frac{1}{\beta_2} - \theta, \frac{1}{\alpha_1} - \theta \right) &= \frac{\alpha_1 + \beta_2 - \theta \alpha_1 \beta_2}{(1 + (1/\beta_2 - \theta) \alpha_1)(1 + (1/\alpha_1 - \theta) \beta_2)} \\
    &\sim - (\beta_1 - \alpha_1)(\beta_2 - \alpha_2) < 0.
\end{align*}
\]

At vertex \( B \) we have

\[
\begin{align*}
    Q \left( \frac{1}{\beta_2} - \theta, \frac{1}{\alpha_1} - \theta \right) &= \frac{\alpha_1 + \beta_2 - \theta \alpha_1 \beta_2}{(1 + (1/\alpha_2 - \theta) \alpha_1)(1 + (1/\alpha_1 - \theta) \beta_2)} \\
    &\sim (\alpha_1 - \alpha_2)(1 - \theta \beta_1).
\end{align*}
\]
Remembering that we examine cases where $1/\beta_1 > 0$, we have to consider three subcases depending on the order of $\alpha_1, \alpha_2$. Hence we have

$$
\begin{cases}
Q > 0 & \text{when } \alpha_1 > \alpha_2, \\
Q = 0 & \text{when } \alpha_1 = \alpha_2, \\
Q < 0 & \text{when } \alpha_1 < \alpha_2.
\end{cases}
$$

At vertex $C$ it is

$$
Q\left(\frac{1}{\alpha_2} - \theta, \frac{1}{\beta_1} - \theta\right) = \frac{\alpha_1 + \beta_2 - \theta \alpha_1 \beta_2}{(1 + (1/\alpha_2 - \theta)\alpha_1)(1 + (1/\beta_1 - \theta)\beta_2)} - \frac{\beta_1 + \alpha_2 - \theta \beta_1 \alpha_2}{(1 + (1/\alpha_2 - \theta)\beta_1)(1 + (1/\beta_1 - \theta)\alpha_2)} \sim (\beta_1 - \alpha_1)(\beta_2 - \alpha_2) > 0.
$$

Finally, at vertex $D$ the value of the function is

$$
Q\left(\frac{1}{\beta_2} - \theta, \frac{1}{\beta_1} - \theta\right) = \frac{\alpha_1 + \beta_2 - \theta \alpha_1 \beta_2}{(1 + (1/\beta_2 - \theta)\alpha_1)(1 + (1/\beta_1 - \theta)\beta_2)} - \frac{\beta_1 + \alpha_2 - \theta \beta_1 \alpha_2}{(1 + (1/\beta_2 - \theta)\beta_1)(1 + (1/\beta_1 - \theta)\alpha_2)} = 0.
$$

Based on the previous results, and depending on the order of $\alpha_1$ and $\alpha_2$, we distinguish three different cases. In case $\alpha_1 > \alpha_2$ the function $Q(r_1, r_2)$ is positive to the left of the hyperbola $DE$, vanishes on it and is negative to the right of it, as is shown in Fig. 1 on the left. This simply means that of the two expressions that form the difference $Q(r_1, r_2)$ in (3.4), the first expression, that is $f(\alpha_1, \beta_2)$, gives the maximum value for $f$ to the left of $DE$, while $f(\alpha_2, \beta_1)$ gives the corresponding maximum to the right of $DE$, and on $DE$ they are identically the same. In case $\alpha_1 < \alpha_2$, $Q(r_1, r_2)$ has the signs as these are shown in Fig. 1 on the right, these signs are interpreted in a similar way as before. The third case arises for $\alpha_1 = \alpha_2$. If this happens, then the points $E$ and $F$ of Fig. 1 coincide with the vertex $B$. Obviously, the signs of $Q(r_1, r_2)$ remain the same as were described previously.

To find the minimum of $f$ in $ABCD$ we work in the same way by using the difference $q(r_1, r_2)$ defined in (3.4). This time we have

$$
q(r_1, r_2) \sim - r_1r_2[\beta_2 \alpha_2 (\beta_1 - \alpha_1) + \beta_1 \alpha_1 (\beta_2 - \alpha_2)] + (r_1 - r_2)[\alpha_1 (\beta_2 - \alpha_2) - \alpha_2 (\beta_1 - \alpha_1)]
$$

$$
- \theta r_1 \beta_1 \alpha_1 (\beta_2 - \alpha_2) - \theta r_2 \beta_2 \alpha_2 (\beta_1 - \alpha_1)
$$

$$
- \theta \alpha_1 (\beta_2 - \alpha_2) - \theta \beta_2 (\beta_1 - \alpha_1) + (\beta_2 - \alpha_2) + (\beta_1 - \alpha_1).
$$

(3.6)

It is easy to see that the right-hand side is always part of a one-sheet hyperboloid and its level curves are always hyperbolas since the coefficient of $r_1r_2$ is negative. Examining the sign of the function $q(r_1, r_2)$, as we did before
371
A(0) B(-) D(+) C(0) + _

Fig. 2. Signs of \( q(r_1, r_2) \) in ABCD.

A B D C E q(-) Q(+) Q(+) q(+) q(-) Q(-) Q(-)

Fig. 3. Signs of \( Q(r_1, r_2) \) and \( q(r_1, r_2) \) in ABCD. (Left: \( x_1 > x_2 \), right: \( x_1 < x_2 \).)

for \( Q(r_1, r_2) \), at the vertices of ABCD we have the following results:
\[
\begin{align*}
q(r_1, r_2) &= 0 \quad \text{at vertex } A, \\
q(r_1, r_2) &< 0 \quad \text{at vertex } B, \\
q(r_1, r_2) &= 0 \quad \text{at vertex } C, \\
q(r_1, r_2) &> 0 \quad \text{at vertex } D.
\end{align*}
\]

So, \( q(r_1, r_2) \) is negative to the right of the hyperbola \( AC \), vanishes on it and is positive to its left (see Fig. 2).

If we combine the results obtained for the signs of \( Q(r_1, r_2) \) and \( q(r_1, r_2) \), we can see that the rectangle ABCD is divided by the two hyperbolas into four sectors, where we have different sign pairs, and therefore different expressions for the values of maximum and minimum of \( f \). These four sectors are illustrated in the Fig. 3, which correspond to the two main cases \( x_1 > x_2 \) and \( x_1 < x_2 \), respectively.

Based on the signs of the functions \( Q \) and \( q \), defined in (3.4), in each of the four sectors of ABCD we have the following results regarding the maximum and the minimum of \( G \) and \( g \), respectively:
\[
\begin{align*}
\text{Sector}(OAE) \text{ or } (OABF): G &= f(\beta_1, z_2), \quad g = f(\beta_1, \beta_2), \\
\text{Sector}(OEB) \text{ or } (OFC): G &= f(x_1, \beta_2), \quad g = f(\beta_1, \beta_2), \\
\text{Sector}(OC) \text{ or } (OCD): G &= f(x_1, \beta_2), \quad g = f(x_1, x_2), \\
\text{Sector}(ODA): G &= f(\beta_1, z_2), \quad g = f(x_1, x_2).
\end{align*}
\]

(3.7)

4. Optimal parameters of the EADI method

Having found the expressions for \( G \) and \( g \) in the four sectors of ABCD in (3.7) we return to our objective, as this was described in the previous section, and attack the problem of minimization of \( G/g \) in each sector separately.
Let us consider the sector $OAE$ (resp., $OABF$) shown in Fig. 3. Considering $G/g$ we have that

$$
\frac{G}{g} = \frac{f(\beta_1, x_2)}{f(\beta_1, \beta_2)} = \frac{(\beta_1 + x_2 - \theta_1 x_2)(1 + \beta_1 r_1)(1 + \beta_2 r_2)}{(\beta_1 + \beta_2 - \theta_1 x_2)(1 + \beta_1 r_1)(1 + \beta_2 r_2)}.
$$

Since the second fraction above equals 1, the ratio in question depends only on $r_2$, and so taking the partial derivative with respect to it we find that $\frac{\partial(G/g)}{\partial r_2} \sim \beta_2 - x_2 > 0$. Therefore, $G/g$ is strictly increasing with respect to $r_2$ and its minimum, since it is independent of $r_1$, is assumed at the lowest point of the sector $OAE$ (resp., $OABF$) which is the point $O$.

Considering now the sector $OCD$ we have

$$
\frac{G}{g} = \frac{f(x_1, \beta_2)}{f(x_1, x_2)} = \frac{(x_1 + \beta_2 - \theta_1 x_2)(1 + x_1 r_1)(1 + x_2 r_2)}{(x_1 + \beta_2 - \theta_1 x_2)(1 + x_1 r_1)(1 + \beta_2 r_2)}.
$$

The situation is similar to the previous one except that this time we have that $\frac{\partial(G/g)}{\partial r_2} \sim \beta_2 - x_2 < 0$, as a result of which $G/g$ is decreasing with $r_2$ and its minimum is assumed at the highest point of the sector $OCD$, that is at $O$.

Combining this with the previous result we have that the overall minimum (optimal) will take place at $O$. Hence the optimal solution to our problem is given by the coordinates $(x_1, x_2)$ which solve our problem must have a unique point of intersection $O$. From the latter we can find also $\omega^*$ and $\kappa^*(M^{-1} A)$ from (2.11) and (2.12).

**Note:** It is noted that if we use the other two sectors into which the rectangle $ABCD$ is divided we end up with exactly the same conclusion.

To find the coordinates of the point $O$ we have to solve the system of the two equations $Q(r_1 r_2) = 0$ and $q(r_1, r_2) = 0$. Taking the two functions from (3.5) and (3.6) we have to solve the system below

$$
\begin{align*}
[&\beta_2 x_2(\beta_1 - x_1) - \beta_1 x_1(\beta_2 - x_2)]r_1 r_2 + [\beta_1 (\beta_2 - x_2) + x_2(\beta_1 - x_1)](r_1 - r_2) - \theta_1 x_1(\beta_2 - x_2)r_1 \\
&+ \theta_2 x_2(\beta_1 - x_1)r_2 - \theta_1 x_1(\beta_2 - x_2) + \theta_2 x_2(\beta_1 - x_1) + (\beta_2 - x_2) - (\beta_1 - x_1) = 0, \\
- [&\beta_2 x_2(\beta_1 - x_1) + \beta_1 x_1(\beta_2 - x_2)]r_1 r_2 + [x_1(\beta_2 - x_2) - x_2(\beta_1 - x_1)](r_1 - r_2) - \theta_1 x_1(\beta_2 - x_2)r_1 \\
&- \theta_2 x_2(\beta_1 - x_1)r_2 - \theta_1 x_1(\beta_2 - x_2) - \theta_2 x_2(\beta_1 - x_1) + (\beta_2 - x_2) + (\beta_1 - x_1) = 0.
\end{align*}
$$

Adding and subtracting the members of the two equations we take the equivalent system

$$
\begin{align*}
- \beta_1 x_1 r_1 r_2 + &\frac{1}{2}(x_1 + \beta_1)(r_1 - r_2) - \theta_1 x_1 r_1 - \frac{\theta}{2}(x_1 + \beta_1) + 1 = 0, \\
\beta_2 x_2 r_1 r_2 + &\frac{1}{2}(x_2 + \beta_2)(r_1 - r_2) + \theta_2 x_2 r_2 + \frac{\theta}{2}(x_2 + \beta_2) - 1 = 0.
\end{align*}
$$

Multiplying the first equation by $x_2 \beta_2$, the second by $x_1 \beta_1$ and adding the resulting equations and then solving for $r_1 - r_2$ we obtain

$$
\begin{align*}
\frac{r_1 - r_2}{r_1 r_2} = &\frac{2(\beta_1 x_1 - \beta_2 x_2) + \theta[\beta_2 x_2(x_1 + \beta_1) - \beta_1 x_1(x_2 + \beta_2)]}{\beta_2 x_2(x_1 + \beta_1) + \beta_1 x_1(x_2 + \beta_2) - 2\theta \beta_1 x_1 \beta_2 x_2} =: H(\theta) \equiv H.
\end{align*}
$$

Replacing the value of $r_1 = r_2 + H(\theta)$ into the second of (4.2) and solving the resulting quadratic equation for $r_2$ we take

$$
\begin{align*}
r_2 = &\frac{\beta_2 x_2(H + \theta) \pm [\beta_2^2 x_2^2(H + \theta)^2 - 2 \beta_2 x_2[(x_2 + \beta_2)(H + \theta) - 2]]^{1/2}}{2 \beta_2 x_2},
\end{align*}
$$

and then from $r_1 = r_2 + H$ it is

$$
\begin{align*}
r_1 = &\frac{\beta_2 x_2(H - \theta) \pm [\beta_2^2 x_2^2(H - \theta)^2 - 2 \beta_2 x_2[(x_2 + \beta_2)(H - \theta) - 2]]^{1/2}}{2 \beta_2 x_2}.
\end{align*}
$$

Note that from our analysis the two hyperbolas must have a unique point of intersection $O$ strictly within the rectangle $ABCD$. Note also that the expressions in (4.5) and (4.4) that give the pairs $(r_1, r_2)$ which solve our problem must have
Theorem 4.1. Under the notation and the assumptions so far, the optimal values of the acceleration parameters \( r_1 = r_1^* \) and \( r_2 = r_2^* \) of the discretized Poisson equation (1.5), using the ADI preconditioner (2.9), are given by the expressions in (4.6), where \( H \) is given by (4.3). The optimal values for the extrapolation parameter \( \omega^* \) and the spectral condition number \( \kappa^*(M^{-1}A) \), are obtained via (2.11) and (2.12), respectively, after having found \( G^* \) and \( g^* \), using either of the two expressions for them in Table 2.

5. Other possible cases

In the previous section the case \( \theta = \theta^* < 1/\beta_i \), \( i = 1, 2 \) was examined and solved successfully. This covers the 5-point case, since then \( \theta = 0 \) and also cases where \( \theta = \theta^* \in \left[ 0, \frac{1}{6\sqrt{3}} \right] \). However, as was found out in [8], and can be readily checked in our case too, the only other possible cases are those for which only one of the two inequalities \( \theta = \theta^* > 1/\beta_i \), \( i = 1, 2 \), holds.

We will examine very briefly one of the two possible cases, let that be the one for which \( 1/\beta_2 < \theta^* < 1/\beta_1 \). The results for the other possible case are produced in an analogous way and are therefore omitted here. Obviously, Table 1 now changes because of the restriction on \( \beta_2 \) and thus \( \partial f (V_{z_1\beta_1} \beta_1, \beta_2) / \partial r_1 < 0 \) for all \( r_1 \in (0, \infty) \). Due to this change, Table 2 is limited to a part of it where its three cells in the left do not exist any more. This is shown in Table 3, where all the results in the other corresponding cells of Table 2 remain unchanged.

To find the maximum and the minimum values of the function \( f \) in the middle rectangle \( A'BCD' \) we work in the same way as in the previous sections. So, we compute the expressions for \( Q(r_1, r_2) = f(x_1, \beta_2) - f(\beta_1, x_2) \) and \( q(r_1, r_2) = f(\beta_1, \beta_2) - f(x_1, x_2) \) in (3.4) and find the signs of their values at the four vertices \( A', B, C, D' \), whose coordinates of \( A' \) and \( D' \) are now \( (0, 1/x_1 - \theta^*) \) and \( (0, 1/\beta_1 - \theta^*) \), respectively, while those of the other two vertices remain unchanged. So, we end up with the following results:

\[
\begin{align*}
Q(r_1, r_2) &< 0 & \text{at vertex } A', \\
Q(r_1, r_2) &> 0 & \text{at vertex } B \text{ when } x_1 > x_2, \\
Q(r_1, r_2) &= 0 & \text{at vertex } B \text{ when } x_1 = x_2, \\
Q(r_1, r_2) &< 0 & \text{at vertex } B \text{ when } x_1 < x_2, \\
Q(r_1, r_2) &> 0 & \text{at vertex } C, \\
Q(r_1, r_2) &> 0 & \text{at vertex } D',
\end{align*}
\]

and

\[
\begin{align*}
q(r_1, r_2) &< 0 & \text{at vertex } A', \\
q(r_1, r_2) &< 0 & \text{at vertex } B, \\
q(r_1, r_2) &= 0 & \text{at vertex } C, \\
q(r_1, r_2) &> 0 & \text{at vertex } D'.
\end{align*}
\]
Table 3
Maximum and minimum of \( f \) in each subregion of the first quadrant where \((r_1, r_2)\) lies

<table>
<thead>
<tr>
<th>( r_2 )</th>
<th>( A' )</th>
<th>( B )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+\infty)</td>
<td>( G = f(\beta_1, \alpha_2) )</td>
<td>( G = f(\alpha_1, \alpha_2) )</td>
</tr>
<tr>
<td>( g = f(\beta_1, \beta_2) )</td>
<td>( g = f(\beta_1, \beta_2) )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \frac{1}{\alpha_1} - \theta^* )</th>
<th>( A' )</th>
<th>( B )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( G = \max{f(\alpha_1, \beta_2), f(\alpha_2, \beta_1)} )</td>
<td>( G = f(\alpha_1, \beta_2) )</td>
<td></td>
</tr>
<tr>
<td>( g = \min{f(\alpha_1, \alpha_2), f(\beta_1, \beta_2)} )</td>
<td>( g = f(\beta_1, \beta_2) )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \frac{1}{\beta_1} - \theta^* )</th>
<th>( D' )</th>
<th>( C )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( G = f(\alpha_1, \beta_2) )</td>
<td>( G = f(\alpha_1, \beta_2) )</td>
<td></td>
</tr>
<tr>
<td>( g = f(\alpha_1, \alpha_2) )</td>
<td>( g = f(\beta_1, \alpha_2) )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( 0 )</th>
<th>( A' )</th>
<th>( B )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{1}{\alpha_2} - \theta^* )</td>
<td>(+\infty)</td>
<td>( r_1 )</td>
</tr>
</tbody>
</table>

It should be pointed out that the only differences with the main cases examined before are that only the sign of the function \( Q(r_1, r_2) \) at the vertex \( D' \) becomes negative, instead of being 0 as it was at \( D \), and that of \( q(r_1, r_2) \) at the vertex \( A' \) becomes negative, instead of being 0 as it was at \( A \) (see Figs. 4 and 5, respectively).

This simply means that the two hyperbolas \( Q(r_1, r_2) = 0 \) and \( q(r_1, r_2) = 0 \), defined in (3.5) and (3.6), respectively, have an intersection point each with the side \( A'D' \) of the rectangle \( A'BCD' \). Let these points be \( J \), with coordinates \((r_1, r_2) = (0, 0)\), and \( K \), with coordinates \((r_1, r_2) = (0, r_2)\), respectively, which are clearly shown in Figs. 4 and 5.

It remains to be shown that the two hyperbolas intersect each other at a point \( O \) always strictly within the rectangle in question. More specifically:

**Theorem 5.1.** Under the notation and the assumptions so far and the additional assumption that \( 1/\beta_2 - \theta^* < 0 \), the two hyperbolas defined by the functions \( Q(r_1, r_2) = 0 \) and \( q(r_1, r_2) = 0 \), defined in (3.5) and (3.6), intersect each other at a point \( O \) strictly within the rectangle \( A'BCD' \), as in Fig. 6.

**Proof.** We begin with the main case of the previous Section 3, where it was \( 1/\beta_i - \theta^* > 0 \), \( i = 1, 2 \), and the point \( O \) was strictly within the rectangle \( ABCD \), as in Fig. 3. Suppose now that the quantity \( 1/\beta_2 - \theta^* \) decreases continuously going from positive values to zero and then to negative ones. First we examine the case of the zero value, that is when \( 1/\beta_2 = \theta^* \) and \( ABCD \equiv A'BCD' \). Let \( r_{2J} \) and \( r_{2K} \) be the ordinates \( r_2 \) for \( r_1 = 0 = 1/\beta_2 - \theta^* \) of the points \( J \) and \( K \),
respectively. Setting in both equations (4.1), \( r_1 = 0 \) and \( \theta = \theta^* = 1/\beta_2 \) and solving the first equation for \( r_2 = r_{2J} \) and the second for \( r_2 = r_{2K} \), we obtain

\[
\begin{align*}
 r_{2J} &= \frac{\beta_2 - \beta_1}{\beta_1 \beta_2}, & r_{2K} &= \frac{\beta_2 - \beta_1}{\alpha_1 \beta_2} \quad \text{and} \quad r_{2J} - r_{2K} &= \frac{(\beta_2 - \beta_1)(\alpha_1 - \beta_1)}{\alpha_1 \beta_1 \beta_2} < 0, \\
\end{align*}
\]  

(5.3)

since \( \alpha_1 < \beta_1 < \beta_2 \). This simply means that \( O \) will still lie strictly within \( A'B'C'D' \).

Assume now that the quantity \( 1/\beta_2 - \theta^* \) goes on decreasing continuously from its zero value. We claim that during this continuous decrease the point \( O \) always lies to the right of \( A'D' \). If for a negative value of the quantity in question \( O \) lies to the left of \( A'D' \) there will be a value of this quantity such that \( O \) will have a zero abscissa. In such a case, obviously, \( r_1^* = 0 \). Then, from (4.3) it will be \(-r_2^* = H =: H_0\), hence \( H_0 < 0 \). On the other hand, from the first of (4.6), for \( r_1^* = 0 \), we have after some algebra, solving for \( H = H_0 \), that

\[
H_0 = \frac{2 - (\alpha_2 + \beta_2)\theta^*}{\alpha_2 + \beta_2 - 2\theta^*\alpha_2 \beta_2}.
\]  

(5.4)

Observing that the minimum value of the numerator of (5.4) is attained for the maximum of \( (\alpha_2 + \beta_2)\theta^* \), which is equal to

\[
\max \left\{ \frac{\sqrt{b} h_2}{a h_1} \frac{1}{12} \left( \frac{b}{a h_1} + \sqrt{\frac{a}{b} h_1} + \sqrt{\frac{a h_1}{b h_2}} \right) \right\} = 4\sqrt{\frac{5}{12}} \left( \frac{1}{\sqrt{5}} + \frac{1}{\sqrt{5}} \right) = 2,
\]

we have that the numerator is always nonnegative. Consider the denominator of (5.4), which can be written as

\[
4\sqrt{\frac{b h_2}{a h_1}} \left\{ 1 - 2\sqrt{\frac{b h_2}{a h_1}} \sin^2(\pi h_2)\theta^* \right\}.
\]
The second term into the braces of the above expression has an upper bound equal to

\[ 2\sqrt{5} \left( \frac{\sqrt{2}}{2} \right)^2 \frac{1}{12} \left( \sqrt{5} + \frac{1}{\sqrt{5}} \right) = \frac{1}{2}, \]

making the expression in the denominator be always positive. The last two results regarding the signs of the terms of the fraction in (5.4) imply that \( H_0 \) can never be negative. This proves our claim and concludes the proof of the theorem. \( \square \)

Theorem 5.1 has the following consequence.

**Theorem 5.2.** Let the notation used, the assumptions made so far and the additional assumption \( 1/\beta_1 < \theta^* \) (resp., \( 1/\beta_2 < \theta^* \)) hold. Then, the values of \( G \) and \( g \) are given by those in Table 2 without its three leftmost cells for \( r_1 \), that is by Table 3 (resp., by Table 2 without its three bottom cells for \( r_2 \)). However, the optimal values for \( r_1^* \) and \( r_2^* \) as well as for all other parameters involved are given by the same formulas and expressions as in Theorem 4.1.

6. Numerical examples

Three two-dimensional Poisson equations, with exact solutions

\[ u(x, y) = \sin \left( \frac{\pi x}{2} \right) \sin \left( \frac{\pi y}{2} \right), \quad u(x, y) = \exp(x + y), \quad u(x, y) = \exp(x + y) \sin \left( \frac{\pi x}{2} \right) \sin \left( \frac{\pi y}{2} \right) \] (6.1)

on the open unit square \( \Omega = (0, 1) \times (0, 1) \) and given Dirichlet boundary conditions based on the functions in (6.1), are considered. Uniform discretization meshes of sizes \( n_1 = 1/(n_1 + 1) \), \( n_1 = 40, 80, 160 \) and \( n_2 = n_1/2 \) as well as \( n_1 = 20, 40, 80 \) and \( n_2 = 2n_1 \), were imposed on \( \Omega \). Both the 5-point (\( \theta = 0 \)) and the 9-point difference formulas (\( \theta = \theta^* \)) were used to approximate the PDEs at each internal node. The corresponding eighteen linear systems obtained were solved by using the following seven methods: conjugate gradient (CG), Cholesky (C), optimal (E)ADI-CG, incomplete Cholesky (IC)-CG, block (B) IC-CG, modified (M) IC-CG and modified block (MB) IC-CG. FORTRAN77 with double precision arithmetic was used and the stopping criteria for the six iterative methods were

\[
\frac{\| e^{(i)} \|_A}{\| e^{(0)} \|_A} = \frac{(r^{(k)}, e^{(k)})_{1/2}}{(r^{(0)}, e^{(0)})_{1/2}} < 10^{-10} \quad \text{and} \quad \frac{\| e^{(i)} \|_{M^{-1}A}}{\| e^{(0)} \|_{M^{-1}A}} = \frac{(z^{(k)}, e^{(k)})_{1/2}}{(z^{(0)}, e^{(0)})_{1/2}} < 10^{-10}. \] (6.2)

In (6.2), the first criterion refers to the CG method and the second one to the other five preconditioned iterative methods. Also, \( u^{(k)}, e^{(k)} = u - u^{(k)}, \) and \( r^{(k)} = c - Au^{(k)} \) are the approximate solution, error and remainder vectors at the \( k \)th iteration, respectively, \( e^{(k)} \) is the solution of the system \( Mz^{(k)} = r^{(k)} \), where \( M \) is the preconditioner used, if any. The initial approximation \( u^{(0)} \) was taken to be \( [1 \; 1 \; \ldots \; 1]^T \in \mathbb{R}^{n_1.n_2} \).

The numerical results obtained for each of the seven cases for the same example and the same grid were very similar in behavior. So, only the third example of (6.1) is depicted in Figs. 7–9 and 10. Of all the methods the Cholesky method was the worst of all; the errors were “satisfactory” only for “small” values of \( n_1 \) and \( n_2 \) and even in these cases the times elapsed to solve the problem in (6.1) were up to 5 times worse than that of the worst iterative method. That is why this method is not depicted in the figures. All six iterative methods for the third example are depicted in Figs. 7 and 8 for the 5-point scheme and in Figs. 9 and 10 for the 9-point one.

First, from Figs. 7 and 8 (5-point case) it is seen that the ADI-CG method is better than the CG and the IC ones and worse than the other three methods. However, the relative errors obtained \( \| u^{(iter)} - u \|_\infty /\| u \|_\infty \), where \( u \) is the exact solution vector of the function \( u(x, y) \), with components the \( n_1 \times n_2 \) values \( u(i_1h_1, i_2h_2), \) \( i_1 = 1(1)n_1, \) \( i_2 = 1 \), \( i = 1, 2, \) at the internal nodal points and \( u^{(iter)} \) the approximate solution after the stopping criterion is satisfied. The absolute errors obtained \( (\| u^{(iter)} - u \|_\infty ) \) were more or less the same and of the order \( 10^{-8} \) for “small” values of \( n_1, n_2 \).

From Figs. 9 and 10 (9-point case) it is seen that the ADI-CG method is competitive to if not better than the best iterative methods which was compared to. The absolute errors in all cases were more or less the same and their accuracies were of the order \( 10^{-8} \) for “small” values of \( n_1, n_2 \), to \( 10^{-9} \) for larger values (Fig. 11).
The subroutines for all but the ADI-CG method were taken from the NSPCG package in [11] (see also [3]). Also, it is noted that for any grids the ADI-CG method is faster than any of the ones in the NSPCG package as this is shown from the figures and in what follows only the MBIC-CG is considered, the fastest of all in the package.

Having compared our ADI preconditioner against the IC, BIC and their modifications we also compare the ADI-CG method we propose in this paper to a variety of direct and iterative methods for the 9-point discretization scheme. Direct methods used were: (i) The fast Fourier transform (FFT9) with block cyclic reduction (BCR) of Houstis and Papatheodorou [10]. This method uses the 9-point discretization scheme for the differential operator and a certain 5-point scheme for the right-hand side of Poisson equation (see (2.1) in [10]), and (ii) the BCR from the FISHPACK package (www.cisl.ucar.edu/css/software/fishpack) of Swarztrauber and Sweet [14]. As an iterative method we use the CG as a basic solver preconditioned with MBIC taken from the NSPCG package [11] as before. Finally, our method was compared to the multigrid (MG) method (MUDPACK program) of Adams [1] (www.cisl.ucar.edu/css/software/mudpack).

All programs were written in single precision FORTRAN 77. For reasons of fair comparison double precision was not used since FFT9 [10], BCR [14] and MG [1] are all written in single precision. We restricted to grids $n_1 \times n_2 = 2^k \times 2^l$. 
since this is a requirement for FFT9 to apply, and satisfies also a corresponding requirement for MG. All schemes in the various programs used are (theoretically) of order $O(h^4)$. For the iterative methods the null vector was taken as the initial approximation. First, FFT9 and BCR were run where the expression

$$\max_{i=1(1)n_1, j=1(1)n_2} \frac{|u_{i,j}^{\text{approx}} - u_{i,j}|}{|u_{i,j}|}$$

(6.3)

was used to obtain an accuracy in the solution. In the expression above $u^{\text{approx}}$ is the approximate solution vector obtained and $u$ is the exact solution vector. Next, for the MBIC-CG and the ADI-CG a stopping criterion of the type

$$\frac{\|r^{(k)}\|_2}{\|c\|_2} < tol$$

(6.4)
was used, where \( \text{tol} \) in (6.4) was adjusted in order to obtain the same relative absolute error as in (6.3) for the FFT9 and BCR methods. Note that in NSPCG package the criterion in (6.4) is independent of the preconditioner. As a stopping criterion for the MG method the default one, \( \|u^{(k)} - u^{(k-1)}\|_2 / \|u^{(k)}\|_2 < \varepsilon \), was used, where \( u^{(k)} \) is the \( k \)th approximation to the solution vector.

From Table 4 (see also Fig. 11) one can see that for small grids \( (8 \times 16, 16 \times 32, 32 \times 64) \) our method is competitive to FFT9, and better than the BCR of FISHPACK and MG of MUDPACK. For a 64 \( \times \) 128 grid it is competitive to BCR and MG but not better than FFT9. However, here it should be pointed out that, regarding the accuracy obtained in the errors, FFT9 is “worse” by a factor of 10 compared to our ADI-CG and also to BCR and MBIC-CG while MG is better than the latter three methods by a factor of 10^3; maybe this is also due to the different stopping criteria used. For a 128 \( \times \) 256 grid FFT9 seems to perform better than all the methods regarding CPU time. However, from Table 4, BCR is “worse” than all when errors in the solution are considered while the MG method is again the best by a factor of about 10^3 when compared to FFT9, MBIC-CG and ours.

The reader should bear in mind that in our programs we have not used any functions or subroutines from known packages or any special storage allocation or even any special techniques for solving linear systems (as, e.g., Red–Black ordering). In addition, it is also noted that FFT9, BCR and MG use a device to obtain a \( O(h^2) \) approximation to the solution vector by a 5-point scheme which subsequently is “corrected” (“improved”) by a “difference correction” scheme to obtain the \( O(h^4) \) accuracy. Our programs were written with the sole objective to test and check the theoretical results of this work, meaning that the ADI-CG method can be made much more efficient if a code suitable for it is written.
7. Concluding remarks and discussion

In this section we make some concluding remarks, discuss some simpler cases and try to connect our work with previous ones.

(i) From the numerical examples in Section 6 it becomes clear that the optimal ADI preconditioners can drastically improve the convergence of the conjugate gradient method, due to the reduction of the spectral condition number of the preconditioned system, and are very competitive to some of the most popular ones.

(ii) In this work (E)ADI preconditioners to the CG methods were applied for the solution of the discretized Poisson equation defined in a rectangle under Dirichlet boundary conditions. Slight modifications of them can also cover the Helmholtz equation, when the coefficients $a$ and $b$ in $u_{xx}$ and $u_{yy}$ of (1.1) are functions of $x$ and $y$, respectively, under Dirichlet, Neumann and mixed (Dirichlet/Neumann) boundary conditions. In these cases, using a similarity transformation with suitable diagonal matrices makes the new resulting discrete operators $A_1$ and $A_2$ in (1.7) commute and be symmetric and positive definite (see [15]). If the operators do not commute as in the case where $a$ and $b$ are functions of both $x$ and $y$, then the same parameters as in the commutative case can be used (see, e.g., [18]) but this time the (E)ADI preconditioners cannot be used in connection with CG but rather with a Krylov subspace method like GMRES, etc. If the region is not a rectangle it can be embedded into a rectangle and the ADI acceleration parameters for the rectangle can be used. In any case we can consider and use the (E)ADI preconditioner of a model problem as an approximation to the actual (E)ADI preconditioner for any two-dimensional second order elliptic PDE (see Section 3.6.2 of [3] and the references cited therein).

(iii) To determine the optimal ADI preconditioner one has to determine the optimal acceleration parameters of the corresponding extrapolated ADI scheme using the formulas (4.6) of Section 4.

(iv) Our analysis in this work covers the stationary biparametric two-dimensional case using low ($\theta = 0$) and high ($\theta = \theta^*$) accuracy discretization schemes. In case of non-stationary schemes one may use the classical optimal set of Jordan–Wachspress parameters [16] (also [17]) which are suitable for the 5-point scheme. The determination of the optimal parameters for the non-stationary problem for the 9-point scheme is still an open one.

(v) In the 5-point scheme the corresponding optimal acceleration and extrapolation parameters are obtained by simply putting $\theta = 0$ in (4.6) and (2.11). It can be found that

$$r_1^* = \frac{a_1b_1 - a_2b_2 + [(a_1 + a_2)(b_1 + b_2)(a_2 + b_2)(a_1 + b_1)]^{1/2}}{a_1b_1(a_2 + b_2) + a_2b_2(a_1 + b_1)},$$
$$r_2^* = \frac{a_2b_2 - a_1b_1 + [(a_1 + a_2)(b_1 + b_2)(a_2 + b_2)(a_1 + b_1)]^{1/2}}{a_1b_1(a_2 + b_2) + a_2b_2(a_1 + b_1)},$$
$$\omega^* = r_1^* + r_2^*.\quad(7.1)$$

The latter optimal results are the ones obtained in [2]. As was pointed out there, the last formula for $\omega^*$ proves that the stationary biparametric Peaceman–Rachford ADI method is already optimal. For the stationary biparametric Peaceman–Rachford ADI method the problem is treated in a nice way in Young [19]. However, the optimal $r_1 = r_1^*$ and $r_2 = r_2^*$ are not given explicitly but via a number of other parameters. The problem treated in [19] would have been solved explicitly in [16] if the stationary case had been considered.

(vi) As was pointed out in the abstract the optimal EADI parameters for the 9-point difference scheme are new and are given in (4.6) and (2.11). It should be mentioned that in this case it is in general $\omega^* \neq r_1^* + r_2^*$. For example, for $a = b = 1$, $l_1 = l_2 = 1$, $n_1 = n_2$ and $\theta = \frac{1}{2}$, it was found in [8] (see (7.4)) that

$$r_1^* = r_2^* = \sqrt{1 - \frac{1}{12}z(1 - \frac{1}{12} \beta)} - \frac{1}{12} \sqrt{z \beta}, \quad \omega^* = 2 \sqrt{1 - \frac{1}{12}z(1 - \frac{1}{12} \beta)}\sqrt{z \beta},\quad(7.2)$$

where $z = a_1 = a_2$ and $\beta = b_1 = b_2$, from which we obtain that $r_1^* + r_2^* \neq \omega^*$.

(vii) There are cases where $r_1^* = r_2^*$. For this one may see from (4.3) that equality between the two optimal acceleration parameters can hold if and only if $H(\theta) = 0$, or, equivalently, the numerator of the fraction is zero, since the denominator
is always positive because it can be written as

\[ x_1 x_2 (\beta_1 + \beta_2) + \beta_1 \beta_2 [x_1 (1 - \theta x_2) + x_2 (1 - \theta x_1)] > 0. \]

In case \( \theta = 0 \), \( H(\theta) = 0 \) if and only if \( x_1 \beta_1 = x_2 \beta_2 \). In case \( \theta = 0^\circ \) and its coefficient in the numerator is not zero, then \( r_1^* = r_2^* \) if and only if

\[ \theta = \frac{\beta_2 x_2 - \beta_1 x_1}{\beta_2 x_2 (x_1 + \beta_1) - \beta_1 x_1 (x_2 + \beta_2)} \in \left[ \frac{1}{6}, \frac{1}{2\sqrt{5}} \right]. \]

However, if the coefficient of \( \theta = 0^\circ \) is zero, then one should also have \( \beta_2 x_2 - \beta_1 x_1 = 0 \). From the two equal to zero expressions one can obtain that \( \beta_2 + x_2 = \beta_1 + x_1 \) and \( \beta_2 x_2 = \beta_1 x_1 \), and from the latter equalities and the fact that \( 0 < x_i < \beta_i \), \( i = 1, 2 \), one arrives at the trivial case \( x_2 = x_1 \) and \( \beta_2 = \beta_1 \).

(viii) If one would like to use the same acceleration parameter \( r = r_1 = r_2 \) in the two half-iterations of the Peaceman–Rachford ADI scheme then for the 5-point case the optimal results are given in [19]. However, Guittet’s EADI scheme gives, in some cases better optimal accelerated parameter than the one in [19] as was found in [7]. This parameter depends on the position of the endpoints of the two intervals \([x_1, \beta_1] \) and \([x_2, \beta_2] \) with respect to each other and also on the relation between \( x_1 \beta_1 \) and \( x_2 \beta_2 \). For the 9-point scheme the same optimal parameters (acceleration and extrapolation) in the Guittet’s 5-point case are given in [7] as “good” ones. The corresponding optimal problem has been recently solved successfully in [8].

References