BIT Numer Math (2009) 49: 509–526
DOI 10.1007/s10543-009-0233-0

BIT

Matrix decomposition algorithms for the C^0-quadratic finite element Galerkin method

Kui Du · Graeme Fairweather · Que N. Nguyen · Weiwei Sun

Received: 2 October 2008 / Accepted: 4 June 2009 / Published online: 7 July 2009

Abstract Explicit expressions for the eigensystems of one-dimensional finite element Galerkin (FEG) matrices based on C^0 piecewise quadratic polynomials are determined. These eigensystems are then used in the formulation of fast direct methods, matrix decomposition algorithms (MDAs), for the solution of the FEG equations arising from the discretization of Poisson's equation on the unit square subject to several

Communicated by Lars Eldén.

The work of K. Du was supported in part by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 102005).
The research of Q.N. Nguyen was supported in part by National Science Foundation grant DGE-0231611.
The work of W. Sun was supported in part by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 102005).

K. Du · W. Sun
Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong

K. Du
e-mail: kuidumath@yahoo.com

W. Sun
e-mail: maweiw@math.cityu.edu.hk

G. Fairweather (✉)
Department of Mathematical and Computer Sciences, Colorado School of Mines, Golden, CO 80401-1887, USA
e-mail: gxf@ams.org

Present address:
G. Fairweather
Mathematical Reviews, American Mathematical Society, 416 Fourth Street, Ann Arbor, MI 48103, USA

Q.N. Nguyen
Avanade Inc., 2211 Elliott Avenue, Seattle, WA 98121, USA
e-mail: quen@avanade.com

standard boundary conditions. The MDAs employ fast Fourier transforms and require $O(N^2 \log N)$ operations on an $N \times N$ uniform partition. Numerical results are presented to demonstrate the efficacy of these algorithms.

Keywords Poisson's equation · Finite element Galerkin method · Piecewise quadratic functions · Generalized eigenvalue problem · Matrix decomposition algorithms

Mathematics Subject Classification (2000) 65F05 · 65N22 · 65N30

1 Introduction

Many problems of practical interest involve the solution of Poisson's equation in the unit square; see, for example, [2, 10, 11]. Traditionally, this problem has been solved using finite difference methods and a matrix decomposition algorithm (MDA). An MDA is a direct method which employs fast Fourier transforms (FFTs) to reduce the algebraic problem to one of solving sets of independent one-dimensional systems, and requires $O(N^2 \log N)$ operations on an $N \times N$ uniform mesh of the unit square. While most attention has been devoted to finite difference methods, MDAs have also been developed for finite element Galerkin (FEG) methods [3, 7], orthogonal spline collocation methods [4, 14, 15], and modified spline collocation methods [5, 6, 8, 9]. Each of these discretizations gives rise to a linear system of the form

$$(A_1 \otimes B_2 + B_1 \otimes A_2)\mathbf{u} = \mathbf{F}, \tag{1.1}$$

where, in this paper, A_1 and B_1 are square matrices of order M_1, A_2 and B_2 are of order M_2, $\otimes$ denotes the matrix tensor product, and $\mathbf{u}$ and $\mathbf{F}$ are vectors of order $M_1 M_2$ given by

$$\begin{aligned} \mathbf{u} &= [u_{1,1}, \dots, u_{1,M_2}, \dots, u_{M_1,1}, \dots, u_{M_1,M_2}]^T, \\ \mathbf{F} &= [F_{1,1}, \dots, F_{1,M_2}, \dots, F_{M_1,1}, \dots, F_{M_1,M_2}]^T. \end{aligned} \tag{1.2}$$

To describe the MDA approach considered in this paper, let I_M denote the unit matrix of order M, and suppose that a diagonal matrix Λ and a nonsingular matrix Z are known such that

$$A_1 Z = B_1 Z \Lambda \tag{1.3}$$

and

$$Z^T B_1 Z = I_{M_1}. \tag{1.4}$$

Premultiplying (1.3) by Z^T and using (1.4), we obtain

$$Z^T A_1 Z = \Lambda. \tag{1.5}$$

The system (1.1) can then be written in the form

$$(Z^T \otimes I_{M_2})(A_1 \otimes B_2 + B_1 \otimes A_2)(Z \otimes I_{M_2})(Z^{-1} \otimes I_{M_2})\mathbf{u} = (Z^T \otimes I_{M_2})\mathbf{F}, \tag{1.6}$$

which becomes, on using (1.4) and (1.5),

$$(\Lambda \otimes B_2 + I_{M_1} \otimes A_2)(Z^{-1} \otimes I_{M_2})\mathbf{u} = (Z^T \otimes I_{M_2})\mathbf{F}. \tag{1.7}$$

From the preceding, we obtain the following algorithm for solving (1.1):

Step 1. Compute $\mathbf{g} = (Z^T \otimes I_{M_2})\mathbf{F}$.
Step 2. Solve $(\Lambda \otimes B_2 + I_{M_1} \otimes A_2)\mathbf{v} = \mathbf{g}$.
Step 3. Compute $\mathbf{u} = (Z \otimes I_{M_2})\mathbf{v}$.

In this paper, the matrix Z is a matrix whose elements are sines or cosines and consequently the matrix-vector multiplications involving Z^T and Z in steps 1 and 3, respectively, may be performed using fast Fourier transforms (FFTs) at a total cost of $O(M_1 M_2 \log M_1)$ operations. Since Λ is diagonal, the coefficient matrix in step 2 is block diagonal and the system reduces to M_1 independent systems of order M_2. As we shall see, this step requires $O(M_1 M_2)$ operations. The total cost of the algorithm is then $O(M_1 M_2 \log M_1)$ operations. Clearly, the efficacy of the algorithm depends on knowledge of explicit formulas for the matrices Λ and Z satisfying (1.3) and (1.4). In this paper, the focus is on the determination of these matrices when tensor products of C^0 quadratic piecewise polynomials are used in the FEG method for Poisson's equation in the unit square,

$$-\Delta u = f(x, y), \quad (x, y) \in \Omega, \tag{1.8}$$

where Δ denotes the Laplace operator and $\Omega = (0, 1) \times (0, 1)$ with boundary $\partial\Omega$, subject to the following boundary conditions: on the horizontal sides of $\partial\Omega$, the homogeneous Dirichlet boundary conditions

$$u(x, 0) = u(x, 1) = 0, \quad x \in (0, 1), \tag{1.9}$$

and on the vertical sides of $\partial\Omega$, that is, for $y \in [0, 1]$, one of the following:

$$u(0, y) = u(1, y) = 0 \quad \text{(Dirichlet)}; \tag{1.10}$$

$$u_x(0, y) = u_x(1, y) = 0 \quad \text{(Neumann)}; \tag{1.11}$$

$$u(0, y) = u_x(1, y) = 0 \quad \text{(Dirichlet-Neumann)}; \tag{1.12}$$

$$u_x(0, y) = u(1, y) = 0 \quad \text{(Neumann-Dirichlet)}; \tag{1.13}$$

$$u(0, y) = u(1, y), \qquad u_x(0, y) = u_x(1, y) \quad \text{(periodic)}. \tag{1.14}$$

It should be noted that the formulation of the MDA does not depend on the boundary condition (1.9); more general boundary conditions can be easily incorporated into the algorithm.

A brief outline of the remainder of this paper is as follows. In Sect. 2, we introduce the spaces of C^0 quadratics considered in the paper, and present the FEG method for (1.8) in which tensor products of these spaces are used. We also describe previous work on the formulation of MDAs for FEG methods. In Sects. 3, 4, 5 and 6, we derive the matrices Λ and Z satisfying (1.3), (1.4) for each of the boundary conditions (1.10)–(1.14) in turn. In Sect. 7, features of the implementation of the MDA are described and results of numerical experiments are presented which demonstrate the efficacy of the method. Some concluding remarks are presented in Sect. 8.

2 C^0 piecewise quadratics in the finite element method

2.1 Preliminaries

Let $\{x_i\}_{i=0}^N$ be a uniform partition of [0, 1] such that $x_i = ih,\ i = 0, \ldots, N$, where N is a positive integer and $h = 1/N$ is the stepsize. Let S_h be the space of C^0 piecewise quadratics on [0, 1] defined by

$$S_h = \{v \in C^0[0, 1] : v|_{[x_{i-1},x_i]} \in P_2, i = 1, \ldots, N\},$$

where P_2 is the set of polynomials of degree ≤ 2, and let

$$S_h^{\mathcal{D}} = \{v \in S_h : v(0) = v(1) = 0\}, \qquad S_h^{\mathcal{DN}} = \{v \in S_h : v(0) = 0\},$$

$$S_h^{\mathcal{ND}} = \{v \in S_h : v(1) = 0\}, \qquad S_h^{\mathcal{P}} = \{v \in S_h : v(0) = v(1)\}.$$

Note that $\dim(S_h^{\mathcal{D}}) = 2N - 1$, $\dim(S_h) = 2N + 1$, $\dim(S_h^{\mathcal{DN}}) = \dim(S_h^{\mathcal{ND}}) = \dim(S_h^{\mathcal{P}}) = 2N$. With

$$p_1(x) = \begin{cases} (x+1)(2x+1), & x \in [-1, 0], \\ (1-x)(1-2x), & x \in [0, 1], \\ 0, & \text{otherwise}, \end{cases}$$

and

$$p_2(x) = \begin{cases} 4x(1-x), & x \in [0, 1], \\ 0, & \text{otherwise}, \end{cases}$$

let $\{\phi_i\}_{i=0}^{2N}$ be the basis for S_h defined by

$$\begin{aligned} \phi_{2i}(x) &= p_1\left(\frac{x - x_i}{h}\right), \quad i = 0, \ldots, N, \\ \phi_{2i-1}(x) &= p_2\left(\frac{x - x_{i-1}}{h}\right), \quad i = 1, \ldots, N. \end{aligned} \tag{2.1}$$

Then

$$\phi_i(jh/2) = \delta_{ij}, \quad i, j = 0, 1, \ldots, 2N, \tag{2.2}$$

where δ_{ij} is the Kronecker delta. Define the $(2N + 1) \times (2N + 1)$ matrices $\mathcal{A}$ and $\mathcal{B}$ by

$$\mathcal{A} = (a_{ij})_{i,j=1}^{2N+1}, \quad a_{ij} = (\phi'_{i-1}, \phi'_{j-1}), \qquad \mathcal{B} = (b_{ij})_{i,j=1}^{2N+1}, \quad b_{ij} = (\phi_{i-1}, \phi_{j-1}), \tag{2.3}$$

where

$$(\phi, \psi) = \int_0^1 \phi(s)\psi(s)ds.$$

These matrices take the form

$$\mathcal{A} = h^{-1}\Gamma(\alpha), \qquad \mathcal{B} = h\Gamma(\beta), \tag{2.4}$$

where the matrices $\Gamma(\alpha)$ and $\Gamma(\beta)$ have the structure shown for $N = 3$ in

$$\Gamma(\gamma) = \begin{bmatrix} \gamma_1/2 & \gamma_2 & \gamma_3 & & & & \\ \gamma_2 & \gamma_4 & \gamma_2 & & & & \\ \gamma_3 & \gamma_2 & \gamma_1 & \gamma_2 & \gamma_3 & & \\ & & \gamma_2 & \gamma_4 & \gamma_2 & & \\ & & \gamma_3 & \gamma_2 & \gamma_1 & \gamma_2 & \gamma_3 \\ & & & & \gamma_2 & \gamma_4 & \gamma_2 \\ & & & & \gamma_3 & \gamma_2 & \gamma_1/2 \end{bmatrix}, \tag{2.5}$$

with

$$\alpha_1 = 14/3, \qquad \alpha_2 = -8/3, \qquad \alpha_3 = 1/3, \qquad \alpha_4 = 16/3,$$
$$\beta_1 = 4/15, \qquad \beta_2 = 1/15, \qquad \beta_3 = -1/30, \qquad \beta_4 = 8/15.$$

Throughout this paper, we make use of the following. Let $\mathcal{I}$, $\mathcal{J}$, $\mathcal{M}$, and $\mathcal{N}$ be finite sets of increasing indices. Without loss of generality we assume

$$\mathcal{I} = \{1, \ldots, I'\}, \qquad \mathcal{J} = \{1, \ldots, J'\}, \qquad \mathcal{M} = \{1, \ldots, M'\}, \qquad \mathcal{N} = \{1, \ldots, N'\}.$$

Then the matrix-vector form of

$$\phi_{i,j} = \sum_{m \in \mathcal{M}} c^{(1)}_{i,m} \sum_{n \in \mathcal{N}} c^{(2)}_{j,n} \psi_{m,n}, \quad i \in \mathcal{I},\ j \in \mathcal{J}, \tag{2.6}$$

is

$$\boldsymbol{\phi} = (C_1 \otimes C_2)\boldsymbol{\psi}, \tag{2.7}$$

where

$$C_1 = (c^{(1)}_{i,m})_{i \in \mathcal{I}, m \in \mathcal{M}}, \qquad C_2 = (c^{(2)}_{j,n})_{j \in \mathcal{J}, n \in \mathcal{N}},$$

and

$$\boldsymbol{\phi} = \left[\phi_{1,1}, \ldots, \phi_{1,J'}, \ldots, \phi_{I',1}, \ldots, \phi_{I',J'}\right]^T,$$
$$\boldsymbol{\psi} = \left[\psi_{1,1}, \ldots, \psi_{1,N'}, \ldots, \psi_{M',1}, \ldots, \psi_{M',N'}\right]^T.$$

2.2 The finite element Galerkin method

In the FEG method with C^0 piecewise quadratics for (1.8), (1.9), and one of the boundary conditions (1.9)–(1.14), we seek $u^h \in V_h \otimes S_h^{\mathcal{D}}$, where $\otimes$ denotes the space tensor product, such that

$$\int_\Omega \left(u_x^h v_x^h + u_y^h v_y^h\right) dx\, dy = \int_\Omega f(x, y) v^h\, dx\, dy, \quad v^h \in V_h \otimes S_h^{\mathcal{D}}, \tag{2.8}$$

where $V_h = S_h^{\mathcal{D}}$ for (1.10), $V_h = S_h$ for (1.11), $V_h = S_h^{\mathcal{DN}}$ for (1.12), $V_h = S_h^{\mathcal{ND}}$ for (1.13), and $V_h = S_h^{\mathcal{P}}$ for (1.14). If $\{\psi_n\}_{n=1}^{M_1}$ is a basis for V_h, and $\{\phi_n\}_{n=1}^{M_2}$ is a basis for $S_h^{\mathcal{D}}$ with $M_2 = 2N-1$, we may write

$$u^h(x,y) = \sum_{i=1}^{M_1}\sum_{j=1}^{M_2} u_{i,j}\psi_i(x)\phi_j(y).$$

Then the Galerkin equations (2.8) with $v^h(x,y) = \psi_m(x)\phi_n(y)$ become

$$\begin{aligned}\sum_{i=1}^{M_1}\sum_{j=1}^{M_2} u_{i,j}\left[(\psi_i', \psi_m')(\phi_j, \phi_n) + (\psi_i, \psi_m)(\phi_j', \phi_n')\right]\\ = \int_\Omega f(x,y)\psi_m(x)\phi_n(y)\,dx\,dy.\end{aligned} \tag{2.9}$$

If $\mathbf{u}$ and $\mathbf{F}$ are as in (1.2) with

$$F_{m,n} = \int_\Omega f(x,y)\psi_m(x)\phi_n(y)\,dx\,dy,$$

then, using (2.6) and (2.7), we obtain the linear system (1.2) with

$$\begin{aligned} A_1 &= (a_{ij}^{(1)})_{i,j=1}^{M_1}, \quad a_{ij}^{(1)} = (\psi_i', \psi_j'),\\ B_1 &= (b_{ij}^{(1)})_{i,j=1}^{M_1}, \quad b_{ij}^{(1)} = (\psi_i, \psi_j),\end{aligned} \tag{2.10}$$

and

$$\begin{aligned} A_2 &= (a_{ij}^{(2)})_{i,j=1}^{M_2}, \quad a_{ij}^{(2)} = (\phi_i', \phi_j'),\\ B_2 &= (b_{ij}^{(2)})_{i,j=1}^{M_2}, \quad b_{ij}^{(2)} = (\phi_i, \phi_j),\end{aligned} \tag{2.11}$$

which are symmetric and positive definite or positive semi-definite matrices. The determination of the matrices Λ and Z in the FEG solution of (1.8) using piecewise linear functions is straightforward and is described in [3]. Bank [1] formulated MDA-like methods for solving the FEG linear systems (1.1) for the case of homogeneous Dirichlet boundary conditions, (1.9), (1.10), using $S_h^{\mathcal{D}}$ with the basis comprising $\{-\frac{h^2}{8}\phi_{2i-1}\}_{i=1}^{N}$ from (2.1) together with the functions

$$\phi_{2i} = p_3\left(\frac{x - x_i}{h}\right), \quad i = 1, \ldots, N-1,$$

where

$$p_3(x) = \begin{cases} 1 + x, & x \in [-1, 0],\\ 1 - x, & x \in [0, 1],\\ 0, & \text{otherwise}.\end{cases} \tag{2.12}$$

He transformed the system (1.1) in which $A_1 = A_2 = A$ and $B_1 = B_2 = B$ with $M_1 = M_2 = 2N-1$ to introduce $\bar{A} = \mathcal{S}_{2N-1}A\mathcal{S}_{2N-1}$ and $\bar{B} = \mathcal{S}_{2N-1}B\mathcal{S}_{2N-1}$, where

$\mathcal{S}_{2N-1}$ is the symmetric orthogonal matrix given by

$$\mathcal{S}_M = \left(\frac{2}{M+1}\right)^{1/2} \left(\sin \frac{ik\pi}{M+1}\right)_{i,k=1}^{M}, \tag{2.13}$$

with $M = 2N - 1$. Then the matrices $\bar{A}$, $\bar{B}$ are reordered to become block diagonal with $N - 1$ blocks of order 2 and a single 1×1 block. With these transformed matrices, the original system can be written as one in which the coefficient matrix is block diagonal with $(N-1)^2$ 4×4 blocks, $2(N-1)$ 2×2 blocks and one 1×1 block. The resulting algorithm for solving the FEG equations requires $O(N^2 \log N)$ operations but requires twice as much work as the corresponding method in the present paper because it requires twice as many FFTs. An approach in which the orthogonal transformation is applied only to A_1 and B_1 is very briefly mentioned in [1] but no details are given. Moreover, it is not clear how either of Bank's approaches would extend to other boundary conditions. No numerical results are presented in [1]. In [7], MDAs are developed for the solution of the finite element Galerkin systems when piecewise Hermite bicubics are used to solve (1.8). Using an approach which is quite different from that employed in the present paper, the matrices Λ and Z are determined for various choices of boundary conditions, and numerical results are presented to demonstrate the efficacy of the MDAs. Kaufman and Warner [12, 13] developed and implemented MDAs based on (1.4), (1.5) for the FEG method for more general elliptic problems in which the eigensystems cannot be determined explicitly. These problems are such that the matrices A_1 and B_1 are symmetric and positive definite, and hence there exist a real diagonal matrix Λ and a real nonsingular matrix Z satisfying (1.4), (1.5). However, in general, Λ and Z are not known explicitly and must be computed. Since FFTs cannot be used, the total cost of the algorithm is $O(N^3)$ operations on an $N \times N$ partition, which, however, can be nonuniform.

3 Dirichlet boundary conditions

Using the basis $\{\phi_i\}_{i=1}^{2N-1}$ for $S_h^{\mathcal{D}}$, the matrices A_1 and B_1 are obtained by deleting the first and last rows and columns of the matrices $\mathcal{A}$ and $\mathcal{B}$ of (2.4), respectively. Then, with $\mathcal{S}_{2N-1}$ defined by (2.13), we have

$$\mathcal{S}_{2N-1} A_1 \mathcal{S}_{2N-1} = h^{-1} \begin{bmatrix} D_{1,\alpha} & \mathbf{0} & D_{2,\alpha} K \\ \mathbf{0}^T & \alpha_4 & \mathbf{0}^T \\ (D_{2,\alpha} K)^T & \mathbf{0} & K D_{3,\alpha} K \end{bmatrix}, \tag{3.1}$$

and

$$\mathcal{S}_{2N-1} B_1 \mathcal{S}_{2N-1} = h \begin{bmatrix} D_{1,\beta} & \mathbf{0} & D_{2,\beta} K \\ \mathbf{0}^T & \beta_4 & \mathbf{0}^T \\ (D_{2,\beta} K)^T & \mathbf{0} & K D_{3,\beta} K \end{bmatrix}, \tag{3.2}$$

where K is the $(N-1)\times(N-1)$ matrix

$$K = \begin{bmatrix} & & 1 \\ & \cdot & \\ & \cdot & \\ 1 & & \end{bmatrix}, \tag{3.3}$$

and

$$D_{l,\alpha} = \operatorname{diag}(\alpha_{l,1}, \ldots, \alpha_{l,N-1}), \qquad D_{l,\beta} = \operatorname{diag}(\beta_{l,1}, \ldots, \beta_{l,N-1}), \quad l = 1, 2, 3,$$

with, for $k = 1, \ldots, N-1$,

$$\left.\begin{aligned}
\alpha_{1,k} &= (\alpha_1 + \alpha_4)/2 + 2\alpha_2\mu_{k/2} + \alpha_3\mu_k, \\
\alpha_{2,k} &= (\alpha_4 - \alpha_1)/2 - \alpha_3\mu_k, \\
\alpha_{3,k} &= (\alpha_1 + \alpha_4)/2 + 2\alpha_2\mu_{(2N-k)/2} + \alpha_3\mu_{2N-k}, \\
\beta_{1,k} &= (\beta_1 + \beta_4)/2 + 2\beta_2\mu_{k/2} + \beta_3\mu_k, \\
\beta_{2,k} &= (\beta_4 - \beta_1)/2 - \beta_3\mu_k, \\
\beta_{3,k} &= (\beta_1 + \beta_4)/2 + 2\beta_2\mu_{(2N-k)/2} + \beta_3\mu_{2N-k},
\end{aligned}\right\}$$

where

$$\mu_\ell = \cos(\ell\pi/N). \tag{3.4}$$

It follows from (3.1)–(3.2) that one of the eigenvalues, λ_0, say, is given by

$$\lambda_0 = h^{-2}\alpha_4/\beta_4. \tag{3.5}$$

The remaining eigenvalues and corresponding eigenvectors are obtained from (1.3)–(1.4) by solving $(N-1)$ generalized eigenvalue problems of order 2,

$$\begin{bmatrix} \alpha_{1,k} & \alpha_{2,k} \\ \alpha_{2,k} & \alpha_{3,k} \end{bmatrix} \begin{bmatrix} d_{1,k}^{\pm} \\ d_{2,k}^{\pm} \end{bmatrix} = h^2\lambda_k^{\pm} \begin{bmatrix} \beta_{1,k} & \beta_{2,k} \\ \beta_{2,k} & \beta_{3,k} \end{bmatrix} \begin{bmatrix} d_{1,k}^{\pm} \\ d_{2,k}^{\pm} \end{bmatrix}, \tag{3.6}$$

such that

$$h \begin{bmatrix} d_{1,k}^{\pm} \\ d_{2,k}^{\pm} \end{bmatrix}^T \begin{bmatrix} \beta_{1,k} & \beta_{2,k} \\ \beta_{2,k} & \beta_{3,k} \end{bmatrix} \begin{bmatrix} d_{1,k}^{\pm} \\ d_{2,k}^{\pm} \end{bmatrix} = 1, \tag{3.7}$$

where $k = 1, \ldots, N-1$. The eigenvalues $\lambda_k^{\pm}$ then satisfy

$$\begin{aligned}
&(\beta_{1,k}\beta_{3,k} - \beta_{2,k}^2)(h^2\lambda)^2 - (\beta_{1,k}\alpha_{3,k} + \beta_{3,k}\alpha_{1,k} - 2\beta_{2,k}\alpha_{2,k})h^2\lambda \\
&\quad + (\alpha_{1,k}\alpha_{3,k} - \alpha_{2,k}^2) = 0,
\end{aligned}$$

from which it follows that

$$\lambda_k^{\pm} = h^{-2}\Phi^{\pm}(\mu_k, \mu_{k/2}), \quad k = 1, \dots, N-1, \qquad \lambda_0 = h^{-2}\alpha_4/\beta_4,$$

where

$$\Phi^{\pm}(\mu, \nu) = (-b \pm \sqrt{b^2 - 4ac})/2a, \tag{3.8}$$

with

$$\left.\begin{aligned} a &= \beta_4(\beta_1 + 2\beta_3\mu) - 4\beta_2^2\nu^2, \\ b &= 8\alpha_2\beta_2\nu^2 - \alpha_4(\beta_1 + 2\beta_3\mu) - \beta_4(\alpha_1 + 2\alpha_3\mu), \\ c &= \alpha_4(\alpha_1 + 2\alpha_3\mu) - 4\alpha_2^2\nu^2. \end{aligned}\right\} \tag{3.9}$$

We set

$$\Lambda = \mathrm{diag}(\lambda_1^+, \dots, \lambda_{N-1}^+, \lambda_0, \lambda_{N-1}^-, \dots, \lambda_1^-).$$

Then, from (3.6),

$$(\alpha_{2,k} - h^2\lambda_k^{\pm}\beta_{2,k})d_{1,k}^{\pm} + (\alpha_{3,k} - h^2\lambda_k^{\pm}\beta_{3,k})d_{2,k}^{\pm} = 0.$$

Assuming that

$$d_{1,k}^{\pm} = -\gamma_k^{\pm}(\alpha_{3,k} - h^2\lambda_k^{\pm}\beta_{3,k}), \qquad d_{2,k}^{\pm} = \gamma_k^{\pm}(\alpha_{2,k} - h^2\lambda_k^{\pm}\beta_{2,k}), \tag{3.10}$$

substituting in (3.7) gives

$$\begin{aligned} &h\gamma_k^{\pm^2}[\beta_{3,k}(\beta_{1,k}\beta_{3,k} - \beta_{2,k}^2)(h^2\lambda_k^{\pm})^2 - 2\alpha_{3,k}(\beta_{1,k}\beta_{3,k} - \beta_{2,k}^2)h^2\lambda_k^{\pm} \\ &\quad + (\beta_{1,k}\alpha_{3,k}^2 - 2\beta_{2,k}\alpha_{2,k}\alpha_{3,k} + \beta_{3,k}\alpha_{2,k}^2)] = 1. \end{aligned}$$

Then

$$\begin{aligned} \gamma_k^{\pm} = h^{-1/2}&[\beta_{3,k}(\beta_{1,k}\beta_{3,k} - \beta_{2,k}^2)(h^2\lambda_k^{\pm})^2 - 2\alpha_{3,k}(\beta_{1,k}\beta_{3,k} - \beta_{2,k}^2)h^2\lambda_k^{\pm} \\ &+ (\beta_{1,k}\alpha_{3,k}^2 - 2\beta_{2,k}\alpha_{2,k}\alpha_{3,k} + \beta_{3,k}\alpha_{2,k}^2)]^{-1/2}. \end{aligned} \tag{3.11}$$

We set

$$Z = \mathcal{S}_{2N-1}\left[\begin{array}{c|c|c} \Lambda_1^+ & \mathbf{0} & \Lambda_1^- K \\ \hline \mathbf{0}^T & 1/\sqrt{\beta_4 h} & \mathbf{0}^T \\ \hline (\Lambda_2^+ K)^T & \mathbf{0} & K\Lambda_2^- K \end{array}\right]$$

where $\mathcal{S}_{2N-1}$ and K are given in (2.13) and (3.3), respectively, and

$$\Lambda_l^{\pm} = \mathrm{diag}(d_{l,1}^{\pm}, \dots, d_{l,N-1}^{\pm}), \quad l = 1, 2.$$

4 Neumann boundary conditions

In this case, with the basis $\{\phi_i\}_{i=0}^{2N}$ for S_h, the matrices A_1 and B_1 are the matrices $\mathcal{A}$ and $\mathcal{B}$, respectively, of (2.4). Let

$$\mathcal{C}_{M+1} = \left(\frac{2}{M}\right)^{1/2} \left(w_k \cos\frac{(i-1)(k-1)\pi}{M}\right)_{i,k=1}^{M+1}, \tag{4.1}$$

where

$$w_k = \begin{cases} 1/\sqrt{2}, & k = 1, M+1, \\ 1, & k = 2, \dots, M. \end{cases}$$

Then

$$\mathcal{C}_{2N+1}^T \mathcal{A} \mathcal{C}_{2N+1} = h^{-1} \begin{bmatrix} D_{1,\alpha} & \mathbf{0} & D_{2,\alpha} K \\ \mathbf{0}^T & \alpha_1 - 2\alpha_3 & \mathbf{0}^T \\ (D_{2,\alpha} K)^T & \mathbf{0} & K D_{3,\alpha} K \end{bmatrix}$$

and

$$\mathcal{C}_{2N+1}^T \mathcal{B} \mathcal{C}_{2N+1} = h \begin{bmatrix} D_{1,\beta} & \mathbf{0} & D_{2,\beta} K \\ \mathbf{0}^T & \beta_1 - 2\beta_3 & \mathbf{0}^T \\ (D_{2,\beta} K)^T & \mathbf{0} & K D_{3,\beta} K \end{bmatrix},$$

where K is an $N \times N$ matrix of the form in (3.3) and

$$D_{l,\alpha} = \text{diag}(\alpha_{l,1}, \dots, \alpha_{l,N}), \qquad D_{l,\beta} = \text{diag}(\beta_{l,1}, \dots, \beta_{l,N}), \quad l = 1, 2, 3, \tag{4.2}$$

with, for $k = 1, \dots, N$,

$$\left.\begin{aligned} \alpha_{1,k} &= (\alpha_1 + \alpha_4)/2 + 2\alpha_2 \mu_{(k-1)/2} + \alpha_3 \mu_{k-1}, \\ \alpha_{2,k} &= -(\alpha_4 - \alpha_1)/2 + \alpha_3 \mu_{k-1}, \\ \alpha_{3,k} &= (\alpha_1 + \alpha_4)/2 + 2\alpha_2 \mu_{(2N+1-k)/2} + \alpha_3 \mu_{2N+1-k}, \\ \beta_{1,k} &= (\beta_1 + \beta_4)/2 + 2\beta_2 \mu_{(k-1)/2} + \beta_3 \mu_{k-1}, \\ \beta_{2,k} &= -(\beta_4 - \beta_1)/2 + \beta_3 \mu_{k-1}, \\ \beta_{3,k} &= (\beta_1 + \beta_4)/2 + 2\beta_2 \mu_{(2N+1-k)/2} + \beta_3 \mu_{2N+1-k}, \end{aligned}\right\} \tag{4.3}$$

where μ_ℓ is given by (3.4). Similar to the Dirichlet case, the eigenvalues are

$$\lambda_k^\pm = h^{-2} \Phi^\pm(\mu_{k-1}, \mu_{(k-1)/2}) \quad k = 1, \dots, N, \qquad \lambda_0 = h^{-2}(\alpha_1 - 2\alpha_3)/(\beta_1 - 2\beta_3),$$

where $\Phi^\pm$ is defined in (3.8) with a, b and c as in (3.9). We set

$$\Lambda = \text{diag}(\lambda_1^+, \dots, \lambda_N^+, \lambda_0, \lambda_N^-, \dots, \lambda_1^-).$$

The corresponding eigenvectors satisfying (3.6)–(3.7) are given in (3.10) with $\alpha_{l,k}$ and $\beta_{l,k}$ as in (4.3). Then

$$Z = \mathcal{C}_{2N+1}\left[\begin{array}{c|c|c} \Lambda_1^+ & \mathbf{0} & \Lambda_1^- K \\ \hline \mathbf{0}^T & 1/\sqrt{(\beta_1 - 2\beta_3)h} & \mathbf{0}^T \\ \hline (\Lambda_2^+ K)^T & \mathbf{0} & K\Lambda_2^- K \end{array}\right]$$

with

$$\Lambda_l^{\pm} = \mathrm{diag}(d_{l,1}^{\pm}, d_{l,2}^{\pm}, \ldots, d_{l,N}^{\pm}), \quad l = 1, 2,$$

and $d_{l,k}^{\pm}$, $l = 1, 2$, as in (3.10) and (3.11).

5 Mixed boundary conditions

In the Dirichlet-Neumann case with the basis $\{\phi_i\}_{i=1}^{2N}$ for $S_h^{\mathcal{DN}}$, the matrices A_1 and B_1 are obtained by deleting the first rows and columns of the corresponding matrices in (2.4). Then, with

$$\bar{\mathcal{S}}_N = \left(\frac{2}{N}\right)^{1/2} \left(\sin \frac{i(2k-1)\pi}{2N}\right)_{i,k=1}^{N}, \tag{5.1}$$

we have

$$\bar{\mathcal{S}}_{2N}^T A_1 \bar{\mathcal{S}}_{2N} = h^{-1}\begin{bmatrix} D_{1,\alpha} & D_{2,\alpha}K \\ (D_{2,\alpha}K)^T & KD_{3,\alpha}K \end{bmatrix},$$

$$\bar{\mathcal{S}}_{2N}^T B_1 \bar{\mathcal{S}}_{2N} = h\begin{bmatrix} D_{1,\beta} & D_{2,\beta}K \\ (D_{2,\beta}K)^T & KD_{3,\beta}K \end{bmatrix}$$

where K is the $N \times N$ matrix of the form given in (3.3), and $D_{l,\alpha}$ and $D_{l,\beta}$, $l = 1, 2, 3$, are defined in (4.2) with, $k = 1, \ldots, N$,

$$\left.\begin{aligned} \alpha_{1,k} &= (\alpha_1 + \alpha_4)/2 + 2\alpha_2\mu_{k/2-1/4} + \alpha_3\mu_{k-1/2}, \\ \alpha_{2,k} &= (\alpha_4 - \alpha_1)/2 - \alpha_3\mu_{k-1/2}, \\ \alpha_{3,k} &= (\alpha_1 + \alpha_4)/2 + 2\alpha_2\mu_{N-k/2+1/4} + \alpha_3\mu_{2N-k+1/2}, \\ \beta_{1,k} &= (\beta_1 + \beta_4)/2 + 2\beta_2\mu_{k/2-1/4} + \beta_3\mu_{k-1/2}, \\ \beta_{2,k} &= (\beta_4 - \beta_1)/2 - \beta_3\mu_{k-1/2}, \\ \beta_{3,k} &= (\beta_1 + \beta_4)/2 + 2\beta_2\mu_{N-k/2+1/4} + \beta_3\mu_{2N-k+1/2}, \end{aligned}\right\} \tag{5.2}$$

where μ_ℓ is given by (3.4). The eigenvalues are

$$\lambda_k^{\pm} = h^{-2}\Phi^{\pm}(\mu_{k-1/2}, \mu_{k/2-1/4}), \quad k = 1, \ldots, N,$$

where $\Phi^{\pm}(\mu, \nu)$ is defined in (3.8) with a, b and c as in (3.9). We set

$$\Lambda = \mathrm{diag}(\lambda_1^+, \ldots, \lambda_N^+, \lambda_N^-, \ldots, \lambda_1^-),$$

and

$$Z = \bar{\mathcal{S}}_{2N} \left[\begin{array}{c|c} \Lambda_1^+ & \Lambda_1^- K \\ \hline (\Lambda_2^+ K)^T & K\Lambda_2^- K \end{array} \right],$$

where

$$\Lambda_l^{\pm} = \mathrm{diag}(d_{l,1}^{\pm}, \ldots, d_{l,N}^{\pm}), \quad l = 1, 2,$$

with $d_{lk}^{\pm}$ as in (3.10), (3.11) with $\alpha_{l,k}$ and $\beta_{1,k}$ as in (5.2). Now consider the Neumann-Dirichlet case in which we choose the basis $\{\phi_i\}_{i=0}^{2N-1}$ for $S_h^{\mathcal{ND}}$. If the matrices A_1 and B_1 in the Dirichlet-Neumann case are denoted by $A_{\mathcal{DN}}$ and $B_{\mathcal{DN}}$, respectively, and those in the Neumann-Dirichlet case by $A_{\mathcal{ND}}$ and $B_{\mathcal{ND}}$, then it is easy to see that

$$A_{\mathcal{ND}} = K A_{\mathcal{DN}} K, \qquad B_{\mathcal{ND}} = K B_{\mathcal{DN}} K,$$

where K is given in (3.3). Thus the eigenvalue matrix is the same in both cases, and $Z_{\mathcal{ND}} = K Z_{\mathcal{DN}}$.

6 Periodic boundary conditions

In this case, the basis functions for $S_h^{\mathcal{P}}$ are taken to be

$$\{\phi_0 + \phi_{2N}, \phi_2, \ldots, \phi_{2N-2}, \phi_1, \ldots, \phi_{2N-1}\},$$

so that

$$A_1 = h^{-1} \begin{bmatrix} R(\alpha_3, \alpha_1, \alpha_3) & R(\alpha_2, \alpha_2, 0) \\ R(0, \alpha_2, \alpha_2) & \alpha_4 I \end{bmatrix},$$

$$B_1 = h \begin{bmatrix} R(\beta_3, \beta_1, \beta_3) & R(\beta_2, \beta_2, 0) \\ R(0, \beta_2, \beta_2) & \beta_4 I \end{bmatrix},$$

where

$$R(\hat{a}, \hat{b}, \hat{c}) = \begin{bmatrix} \hat{b} & \hat{c} & & & \hat{a} \\ \hat{a} & \hat{b} & \hat{c} & & \\ & \ddots & \ddots & \ddots & \\ & & \hat{a} & \hat{b} & \hat{c} \\ \hat{c} & & & \hat{a} & \hat{b} \end{bmatrix}.$$

Let F_N denote the Fourier transformation, that is,

$$F_N = N^{-1/2} (\epsilon_{j-1}^{l-1})_{j,l=1}^{N} \quad \epsilon_j = e^{\frac{\iota 2\pi j}{N}}, \ \iota = \sqrt{-1}, \tag{6.1}$$

and $\mathcal{F}_{2N} = \mathrm{diag}(F_N, F_N)$. By the basic properties of circulant matrices, we have

$$\mathcal{F}_{2N}^H A \mathcal{F}_{2N} = h^{-1} \begin{bmatrix} D_{1,\alpha} & D_{2,\alpha}^H \\ D_{2,\alpha} & D_{3,\alpha} \end{bmatrix}, \qquad \mathcal{F}_{2N}^H B \mathcal{F}_{2N} = h \begin{bmatrix} D_{1,\beta} & D_{2,\beta}^H \\ D_{2,\beta} & D_{3,\beta} \end{bmatrix},$$

where $D_{3,\alpha} = \alpha_4 I$, $D_{3,\beta} = \beta_4 I$, and $D_{l,\alpha}$ and $D_{l,\beta}$, $l = 1, 2$, are defined in (4.2) with

$$\left.\begin{aligned} \alpha_{1,k} &= \alpha_1 + 2\alpha_3 \mu_{2(k-1)}, & \alpha_{2,k} &= \alpha_2(1 + \epsilon_{k-1}), & \alpha_{3,k} &= \alpha_4, \\ \beta_{1,k} &= \beta_1 + 2\beta_3 \mu_{2(k-1)}, & \beta_{2,k} &= \beta_2(1 + \epsilon_{k-1}), & \beta_{3,k} &= \beta_4, \end{aligned}\right\} \tag{6.2}$$

where μ_ℓ is given by (3.4). The eigenvalues are given by

$$\lambda_k^\pm = h^{-2} \Phi^\pm(\mu_{2(k-1)}, \mu_{k-1}), \quad k = 1, \ldots, N, \tag{6.3}$$

where $\Phi^\pm(\mu, \nu)$ is defined in (3.8) with a, b and c as in (3.9). We set

$$\Lambda = \mathrm{diag}(\lambda_1^+, \ldots, \lambda_N^+, \lambda_1^-, \ldots, \lambda_N^-).$$

The corresponding eigenvectors can be obtained analogously to (3.6)–(3.11), so that

$$Z = \mathcal{F}_{2N} \left[\begin{array}{c|c} \Lambda_1^+ & \Lambda_1^- \\ \hline \Lambda_2^+ & \Lambda_2^- \end{array} \right], \tag{6.4}$$

where $\Lambda_l^\pm = \mathrm{diag}(d_{l,1}^\pm, \ldots, d_{l,N}^\pm)$, $l = 1, 2$,

$$\begin{bmatrix} d_{1,N/2+1}^+ & d_{1,N/2+1}^- \\ d_{2,N/2+1}^+ & d_{2,N/2+1}^- \end{bmatrix} = \begin{bmatrix} (\beta_{1,N/2+1} h)^{-1/2} & 0 \\ 0 & (\beta_{3,N/2+1} h)^{-1/2} \end{bmatrix}, \tag{6.5}$$

for even N, and $d_{l,k}^\pm$ in other cases are given in (3.10) with

$$\begin{aligned} \gamma_k^\pm = h^{-1/2} \big(&\beta_{1,k} |\alpha_{3,k} - \beta_{3,k} \lambda_k^\pm h^2|^2 \\ &- 2\,\mathrm{Re}\big(\overline{\beta}_{2,k} (\alpha_{2,k} - \beta_{2,k} \lambda_k^\pm h^2)(\alpha_{3,k} - \beta_{3,k} \lambda_k^\pm h^2)\big) \\ &+ \beta_{3,k} |\alpha_{2,k} - \beta_{2,k} \lambda_k^\pm h^2|^2\big)^{-1/2}, \quad k = 1, \ldots, N, \end{aligned}$$

with $\alpha_{l,k}$, $l = 2, 3$, and $\beta_{l,k}$, $l = 1, 2, 3$, as in (6.2).

We now present a real form of the eigenvector matrix. First, we write the matrix F_N of (6.1) in the form $F_N = F_s + \iota F_c$. Then

$$\mathcal{F}_r = \begin{bmatrix} F_s + F_c & 0 \\ 0 & R(0, 1, 1)(F_s + F_c) + E_r \end{bmatrix},$$

where $(E_r)_{ij} = \delta_N(j)(-1)^i$, $i, j = 1, \ldots, N$, with $\delta_N(m) = N^{-1/2}$ for even N and $m = N/2 + 1$ and $\delta_N(m) = 0$ otherwise. Then

$$\mathcal{F}_r^T A \mathcal{F}_r = \begin{bmatrix} D_{1,\alpha} & D_{2,\alpha}^T \\ D_{2,\alpha} & D_{3,\alpha} \end{bmatrix}, \qquad \mathcal{F}_r^T B \mathcal{F}_r = \begin{bmatrix} D_{1,\beta} & D_{2,\beta}^T \\ D_{2,\beta} & D_{3,\beta} \end{bmatrix},$$

where $D_{l,\alpha}$ and $D_{l,\beta}$, $l = 1, 2, 3$, are defined in (4.2) with, for $k = 1, \dots, N$,

$$\left.\begin{aligned}
\alpha_{1,k} &= \alpha_1 + 2\alpha_3\mu_{2(k-1)},\\
\alpha_{2,k} &= 2\alpha_2(1+\mu_{2(k-1)}),\\
\alpha_{3,k} &= 2\alpha_4(1+\mu_{2(k-1)}) + N^{1/2}\delta_N(k)\alpha_4,\\
\beta_{1,k} &= \beta_1 + 2\beta_3\mu_{2(k-1)},\\
\beta_{2,k} &= 2\beta_2(1+\mu_{2(k-1)}),\\
\beta_{3,k} &= 2\beta_4(1+\mu_{2(k-1)}) + N^{1/2}\delta_N(k)\beta_4.
\end{aligned}\right\} \tag{6.6}$$

With

$$a = \beta_{1,k}\beta_{3,k} - \beta_{2,k}^2, \qquad b = 2\alpha_{2,k}\beta_{2,k} - \alpha_{1,k}\beta_{3,k} - \alpha_{3,k}\beta_{1,k},$$
$$c = \alpha_{1,k}\alpha_{3,k} - \alpha_{2,k}^2,$$

the eigenvalues are the same as in (6.3), and the corresponding eigenvectors are given in (3.10), (3.11) with $\alpha_{l,k}$ and $\beta_{l,k}$ of (6.6). Then

$$Z = \mathcal{F}_r \left[\begin{array}{c|c} \Lambda_1^+ & \Lambda_1^- \\ \hline \Lambda_2^+ & \Lambda_2^- \end{array}\right],$$

where $\Lambda_l^{\pm}$, $l = 1, 2$, are defined in (6.4) and (6.5).

7 Numerical results

We solved (2.8) with $f(x, y)$ corresponding to the exact solution

$$u(x,y) = \begin{cases} e^{xy}(x^2-x)(y^2-y) & \text{for (1.10)},\\ e^{xy}(x^2-x)^2(y^2-y) & \text{for (1.11) and (1.12)},\\ e^{y}[1+\sin(2\pi x)](y-y^2) & \text{for (1.14)}. \end{cases} \tag{7.1}$$

First, we describe some features of the implementation of the MDA of Sect. 1. The components of the vector $\mathbf{F} = \{F_{m,n}\}$ are approximated by replacing f by its piecewise quadratic interpolant, $\tilde{f} \in S_h \otimes S_h$, where

$$\tilde{f}(x,y) = \sum_{i=0}^{2N}\sum_{j=0}^{2N} f_{i,j}\psi_i(x)\phi_j(y),$$

so that $f_{i,j} = f(ih/2, jh/2)$ from (2.2). We describe this process for the case of Dirichlet boundary conditions; the other cases are treated similarly. In this case, for $1 \le m, n \le 2N-1$,

$$F_{m,n} = \int_\Omega \tilde{f}(x,y)\psi_m(x)\phi_n(y)\,dx\,dy = \sum_{i=0}^{2N}\sum_{j=0}^{2N}\big(\psi_m,\psi_i\big)\big(\phi_n,\phi_j\big) f_{i,j}$$

$$= \sum_{i=0}^{2N}\big(\psi_m,\psi_i\big)\sum_{j=0}^{2N}\big(\phi_n,\phi_j\big) f_{i,j}.$$

Thus, on using (2.6) and (2.7), we have

$$\mathbf{F} = (C_1 \otimes C_2)\,\mathbf{f} = (C_1 \otimes I_{2N-1})\,(I_{2N+1} \otimes C_2)\,\mathbf{f},$$

where

$$C_1 = ((\psi_m,\psi_i))_{m=1,i=0}^{2N-1,2N}, \qquad C_2 = ((\phi_n,\phi_j))_{n=1,j=0}^{2N-1,2N},$$

and

$$\mathbf{f} = [f_{0,0},\ldots,f_{0,2N},\ldots,f_{2N,0},\ldots,f_{2N,2N}]^T.$$

Then we first compute

$$\mathbf{g} = (I_{2N+1} \otimes C_2)\,\mathbf{f},$$

followed by

$$\mathbf{F} = (C_1 \otimes I_{2N-1})\,\mathbf{g}.$$

Note that the elements of the matrices C_1 and C_2 are easily determined from those of the matrix $\mathcal{B}$ in (2.4). As the following numerical results demonstrate, this approximation does not degrade the accuracy of the FEG method. In step 2 of the MDA, each coefficient matrix has the block structure shown in (2.5) with the first and last rows and columns eliminated. Moreover, since the matrices A_2, B_2 and Λ are positive definite, the coefficient matrices are positive definite and the systems can be solved

Table 1 Errors and convergence rates for (1.10)

N	Max Error	Rate (h^α)	L^2 Error	Rate (h^α)	H^1 Error	Rate (h^α)
4	6.506E−006	–	1.138E−004	–	2.988E−003	–
8	4.243E−007	3.939	1.429E−005	2.993	7.430E−004	2.008
16	2.684E−008	3.982	1.789E−006	2.998	1.855E−004	2.002
32	1.705E−009	3.977	2.236E−007	3.000	4.636E−005	2.001
64	1.070E−010	3.993	2.796E−008	3.000	1.159E−005	2.000
128	6.706E−012	3.997	3.495E−009	3.000	2.897E−006	2.000

Table 2 Errors and convergence rates for (1.11)

N	Max Error	Rate (h^α)	L^2 Error	Rate (h^α)	H^1 Error	Rate (h^α)
4	5.578E−005	–	1.539E−004	–	4.100E−003	–
8	4.597E−006	3.601	2.035E−005	2.919	1.062E−003	1.949
16	3.328E−007	3.788	2.580E−006	2.980	2.678E−004	1.988
32	2.238E−008	3.895	3.236E−007	2.995	6.710E−005	1.997
64	1.451E−009	3.947	4.049E−008	2.999	1.678E−005	1.999
128	9.238E−011	3.973	5.062E−009	3.000	4.196E−006	2.000

Table 3 Errors and convergence rates for (1.12)

N	Max Error	Rate (h^α)	L^2 Error	Rate (h^α)	H^1 Error	Rate (h^α)
4	5.795E−005	–	1.531E−004	–	4.100E−003	–
8	4.710E−006	3.621	2.033E−005	2.913	1.062E−003	1.949
16	3.397E−007	3.793	2.579E−006	2.979	2.678E−004	1.988
32	2.281E−008	3.897	3.236E−007	2.995	6.710E−005	1.997
64	1.478E−009	3.948	4.049E−008	2.999	1.678E−005	1.999
128	9.406E−011	3.974	5.062E−009	3.000	4.196E−006	2.000

Table 4 Errors and convergence rates for (1.14)

N	Max Error	Rate (h^α)	L^2 Error	Rate (h^α)	H^1 Error	Rate (h^α)
4	2.483E−003	–	5.363E−003	–	1.295E−001	–
8	1.877E−004	3.726	6.462E−004	3.053	3.287E−002	1.978
16	1.307E−005	3.844	8.003E−005	3.013	8.255E−003	1.994
32	8.373E−007	3.964	9.981E−006	3.003	2.066E−003	1.998
64	5.269E−008	3.990	1.247E−006	3.001	5.167E−004	2.000
128	3.299E−009	3.997	1.558E−007	3.000	1.292E−004	2.000

using the Choleski method without fill-in at a total cost of $O(N^2)$ operations. In Tables 1–4, we present errors and the corresponding convergence rates in the maximum norm defined by

$$\text{Max Error} = \max_{i,j} |u(x_i, y_j) - u_h(x_i, y_j)|,$$

and the L^2 and H^1 norms, for the boundary conditions (1.10), (1.11), (1.12), and (1.14), respectively. Convergence rates in the various norms are determined using the formula

$$\text{Rate} = \frac{\log(e_{N/2}/e_N)}{\log 2},$$

where e_N is the error corresponding to the $N \times N$ partition of Ω. As expected, the convergence rates for the L^2 and H^1 norms are 3 and 2, respectively, whereas the fourth order convergence rate in the maximum norm demonstrates the superconvergence of the approximate solution at the nodes, where one would expect only third order accuracy.

8 Concluding remarks

Several extensions of the methods described in this paper are easily formulated. As was mentioned earlier, on the horizontal sides of the unit square, one can prescribe more general boundary conditions than (1.9), such as a Robin condition, or a nonlocal condition as in [2]. Moreover, in place of Poisson's equation (1.8), the equation

$$-u_{xx} - (a(y)u_y)_y + b(y)u_y + c(y)u = f(x, y), \quad (x, y) \in \Omega,$$

can be considered. Also, the partition in the y-direction can be non-uniform. The extension to biharmonic Dirichlet problems of the form

$$\left.\begin{aligned} &\Delta^2 u(x, y) = f(x, y), \quad (x, y) \in \Omega = (0, 1) \times (0, 1), \\ &u(x, y) = 0, \qquad \frac{\partial u}{\partial n} = 0, \quad (x, y) \in \partial\Omega, \end{aligned}\right\}$$

where $\partial/\partial n$ denotes the outward normal on the boundary $\partial\Omega$, is a topic for future research.

Acknowledgements We gratefully acknowledge the assistance of Bryan Romero, Michael McCourt and Dan Latterner during the preparation of this paper.

References

1. Bank, R.E.: Efficient algorithms for solving tensor product finite element equations. Numer. Math. **31**, 49–61 (1978)
2. Bao, G., Sun, W.: A fast algorithm for the electromagnetic scattering from a large cavity. SIAM J. Sci. Comput. **27**, 553–574 (2005)
3. Bialecki, B., Fairweather, G.: Matrix decomposition methods for separable elliptic boundary value problems in two dimensions. J. Comput. Appl. Math. **46**, 369–386 (1993)
4. Bialecki, B., Fairweather, G., Bennett, K.R.: Fast direct solvers for piecewise Hermite bicubic orthogonal spline collocation equations. SIAM J. Numer. Anal. **29**, 156–173 (1992)
5. Bialecki, B., Fairweather, G., Karageorghis, A.: Matrix decomposition algorithms for modified spline collocation for Helmholtz problems. SIAM J. Sci. Comput. **24**, 1733–1753 (2003)
6. Bialecki, B., Fairweather, G., Karageorghis, A.: Optimal superconvergent one step nodal cubic spline collocation methods. SIAM J. Sci. Comput. **27**, 575–598 (2005)
7. Bialecki, B., Fairweather, G., Knudson, D.B., Lipman, D.A., Nguyen, Q.N., Sun, W., Weinberg, G.M.: Matrix decomposition algorithms for the finite element Galerkin method with piecewise Hermite cubics. Numer. Algorithms (to appear). doi:10.1007/s11075-008-9255-y
8. Christara, C.C., Ng, K.S.: Fast Fourier transform solvers and preconditioners for quadratic spline collocation. BIT **42**, 702–739 (2002)
9. Constas, A.: Fast Fourier transform solvers for quadratic spline collocation. M.Sc. Thesis, Department of Computer Science, University of Toronto, Toronto, Canada (1996)
10. E, W., Liu, J.G.: Essentially compact schemes for unsteady viscous incompressible flows. J. Comput. Phys. **126**, 122–138 (1996)

11. Jiang, H., Shao, S., Cai, W., Zhang, P.: Boundary treatments in non-equilibrium Green's function (NEGF) methods for quantum transport in nano-MOSFETs. J. Comput. Phys. **227**, 6553–6573 (2008)
12. Kaufman, L., Warner, D.: High-order, fast-direct methods for separable elliptic equations. SIAM J. Numer. Anal. **21**, 672–694 (1984)
13. Kaufman, L., Warner, D.: Algorithm 685: a program for solving separable elliptic equations. ACM Trans. Math. Softw. **16**, 325–351 (1990)
14. Sun, W.: Fast algorithms for high-order spline collocation systems. Numer. Math. **81**, 143–160 (1998)
15. Sun, W., Zamani, N.G.: A fast algorithm for solving the tensor product collocation equations. J. Franklin Inst. **326**, 295–307 (1989)