JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

ELSEVIER Journal of Computational and Applied Mathematics 147 (2002) 175–190

www.elsevier.com/locate/cam

Verified numerical computations for multiple and nearly multiple eigenvalues of elliptic operators

K. Toyonaga[a, *], M.T. Nakao[a], Y. Watanabe[b]

[a] Graduate School of Mathematics, Kyushu University, 6-10-1 Hakozaki, Higashiku, Fukuoka 812-8581, Japan
[b] Computing and Communications Center, Kyushu University, Fukuoka 812-8581, Japan

Received 20 January 2001; received in revised form 25 February 2002

Abstract

In this paper, we propose a numerical method to verify bounds for multiple eigenvalues for elliptic eigenvalue problems. We calculate error bounds for approximations of multiple eigenvalues and base functions of the corresponding invariant subspaces. For matrix eigenvalue problems, Rump (Linear Algebra Appl. 324 (2001) 209) recently proposed a validated numerical method to compute multiple eigenvalues. In this paper, we extend his formulation to elliptic eigenvalue problems, combining it with a method developed by one of the authors (Jpn. J. Indust. Appl. Math. 16 (1998) 307). © 2002 Published by Elsevier Science B.V.

1. Introduction

In [4], a method is proposed to enclose eigenvalues and eigenfunctions for elliptic eigenvalue problems that employs the numerical verification method for nonlinear elliptic problems developed in [3], etc. However, that method can only be applied to the case of simple eigenvalues, due to the property of the verification principle. Specifically, applying this method to multiple eigenvalues leads to a singularity caused by their multiplicity. For matrix eigenvalue problems, a method to compute the error bounds for approximations of multiple and nearly multiple eigenvalues and to verify the basis of the corresponding invariant subspaces was recently proposed [9].

In this paper, we attempt to extend the formulation presented in [6] to elliptic eigenvalue problems. For this purpose, we use the basic idea that has been employed previously in the numerical verification method for elliptic problems. We formulate a multiple eigenvalue problem for elliptic operator as a system of nonlinear elliptic boundary value problems with respect to the eigenvalues and the

* Corresponding author.
E-mail addresses: toyonaga@math.kyushu-u.ac.jp (K. Toyonaga), mtnakao@math.kyushu-u.ac.jp (M.T. Nakao), watanabe@cc.kyushu-u.ac.jp (Y. Watanabe).

PII: S0377-0427(02)00431-4

base functions of the corresponding invariant subspace. Then, as in [3], we devise by applying a kind of set valued version of Newton's method to enclose them numerically.

We use approximative subspaces to treat the finite-dimensional part of the problem, and using constructive error estimates, we enclose the infinite-dimensional part. In the present case, we adopt a spectral method based on the Fourier series expansion and explicit a priori error estimates.

In the following section, we give the basic formulation of the problem and describe the actual computational procedures for the self-adjoint case. However, we would like to emphasize that the present technique can also be applied to the nonself-adjoint problems. In fact, in Section 3, we give a numerical example for a nonself-adjoint problem, in addition to one for a self-adjoint problem. Furthermore, we will show an interesting example of the enclosure of two simple but nearly equal eigenvalues that cannot be verified eigenvalues using the method for simple eigenvalues.

In this paper, all computations that is based on interval arithmetic have been executed using INTLAB [8], an interval package for use under Matlab V5.3.1 [2].

We remark that there are other methods that can be applied to the problem of enclosing multiple eigenvalues for self-adjoint elliptic problems (see, e.g. [1,6,11]). However, unlike those methods, our method enables us not only to enclose eigenvalues but also verify the basis of the corresponding invariant subspaces, as well as treat nonself-adjoint problems.

2. Formulation and verification procedures

2.1. Formulation of the problem

We define Ω to be a bounded convex domain in $\mathbb{R}^2$. Let $H^m(\Omega)$ denote the L^2-Sobolev space of order m on Ω for integer m. Then, we define $H_0^1 \equiv H_0^1(\Omega) \equiv \{v \in H^1(\Omega) \mid v = 0 \text{ on } \partial\Omega\}$ and the inner product on H_0^1 as $\langle u, v\rangle_{H_0^1} \equiv (\nabla u, \nabla v)_{L^2}$ for $u, v \in H_0^1(\Omega)$, where $(\cdot,\cdot)_{L_2}$ represents the inner product on $L^2(\Omega)$. Next, let S_h be a finite-dimensional subspace of H_0^1, and let $\{\phi_i\}_{i=1\cdots N}$ be a basis in S_h. Let $P_{h1}: H_0^1(\Omega) \to S_h$ denote the H_0^1-projection for each $u \in H_0^1$ defined by

$$(\nabla u, \nabla v)_{L^2} = (\nabla P_{h1}u, \nabla v)_{L^2} \quad \text{for all } v \in S_h.$$

In this paper, we are mainly concerned with elliptic eigenvalue problems of the following self-adjoint type (see Example 2 of Section 2 for a nonself-adjoint problem):

$$\begin{aligned} -\Delta u + qu &= \lambda u \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{1}$$

where $q \in L^\infty(\Omega)$.

First, we calculate an approximate spectrum of (1). Then we compute the error bounds for the multiple eigenvalues and enclose a basis of the corresponding invariant subspace around the approximate solutions.

For matrix eigenvalue problems, Rump [9] showed that the error bounds for k-fold computed eigenvalues and the approximate basis of corresponding invariant subspace of $n \times n$ matrix A can be calculated by verifying bounds for the matrices Y, M, which satisfy the equation $AY = YM$, where Y is an $n \times k$ matrix and M a $k \times k$ matrix.

In order to extend the method presented in [9] to the elliptic eigenvalue problem, we transform (1) into an eigenequation of the form

$$(-\Delta + q)Y = YM, \tag{2}$$

where

$$Y \equiv (y_1, y_2, \ldots, y_n) \quad \text{and} \quad M \equiv \begin{pmatrix} m_{11} & \ldots & m_{1n} \\ \vdots & & \vdots \\ m_{n1} & \ldots & m_{nn}. \end{pmatrix}.$$

The ith column in the right-hand side of (2) is interpreted as $(YM)_i \equiv \sum_{j=1}^{n} m_{ji} y_j$, $(1 \leqslant i \leqslant n)$, where n is the expected multiplicity, $y_i \in H_0^1$, and $m_{ij} \in \mathbb{R}$. Then, note that each eigenvalue of M is also an eigenvalue of (1). Here $\{y_i\}_{i=1\cdots n}$ is a basis of the corresponding invariant subspace if all y_i are linearly independent. In this paper, since we attempt to verify multiple eigenvalues together with associated base functions of the corresponding invariant subspace, when the eigenvalue λ is an n-fold eigenvalue, considering the space $V \equiv (H_0^1)^n \times (\mathbb{R})^{n^2}$, and we verify $(Y, M) \in V$ satisfying (2).

We define the inner product on V for $w_1 = (y_1^1, \ldots, y_n^1, r_1, \ldots, r_{n^2})$ and $w_2 = (y_1^2, \ldots, y_n^2, s_1, \ldots, s_{n^2})$ as

$$\langle w_1, w_2 \rangle = \langle y_1^1, y_1^2 \rangle_{H_0^1} + \cdots + \langle y_n^1, y_n^2 \rangle_{H_0^1} + r_1 s_1 + \cdots + r_{n^2} s_{n^2}.$$

Then, for $V_h \equiv (S_h)^n \times (\mathbb{R})^{n^2}$, we define the projection $P_h : V \to V_h$ by

$$P_h(u_1, \ldots, u_n, r_1, \ldots, r_{n^2}) = (P_{h1}u_1, \ldots, P_{h1}u_n, r_1, \ldots, r_{n^2}),$$

where $u_i \in H_0^1$ $(1 \leqslant i \leqslant n)$ and $r_j \in \mathbb{R}$ $(1 \leqslant j \leqslant n^2)$.

Let $\lambda_i^h \in \mathbb{R}$, $y_i^h \in S_h$ $(1 \leqslant i \leqslant n)$ be approximate solutions of (1) or (2). We now suppose that for each i, y_i^h can be represented as $y_i^h = \sum_{j=1}^{N} c_{ij}\phi_j$, $c_{ij} \in \mathbb{R}$, where N is the dimension of S_h. Then, for each i, let $\tilde{\phi}_i$ be the base function whose coefficient takes the maximal value in $\{|c_{i1}|, \ldots, |c_{iN}|\}$.

Thus, we obtain the normalized eigenvalue problem of the form

$$\begin{cases} (-\Delta + q)(y_1, y_2, \ldots, y_n) = (y_1, y_2, \ldots, y_n) \begin{pmatrix} m_{11} & \ldots & m_{1n} \\ \vdots & & \vdots \\ m_{n1} & \ldots & m_{nn} \end{pmatrix}, \\ (y_i, \tilde{\phi}_j) = (y_i^h, \tilde{\phi}_j), \quad 1 \leqslant i, \ j \leqslant n \end{cases} \tag{3}$$

which is a kind of nonlinear system with respect to y_i and m_{ij} of elliptic equations.

2.2. Transformation to fixed point form

We set $y_i = y_i^h + \tilde{y}_i$ and $m_{ij} = m_{ij}^h + \widetilde{m_{ij}}$ in (3), with $m_{ii}^h = \lambda_i^h$, $m_{ij}^h = 0$ $(i \neq j)$. Then $\tilde{y}_i$ and $\widetilde{m_{ij}}$ represent the errors of the approximate solutions y_i^h and approximate matrix elements m_{ij}^h.

We verify y_i and m_{ij} satisfying (3) by enclosing $\tilde{y}_i$ and $\widetilde{m_{ij}}$. We can rewrite (3) for $w = (\tilde{y}_1, \ldots, \tilde{y}_n, \widetilde{m_{11}}, \ldots, \widetilde{m_{nn}})$ as follows:

$$\begin{aligned}
-\Delta\widetilde{y_1} &\equiv f_1(w) = (m_{11}^h + \widetilde{m_{11}} - q)\widetilde{y_1} + (m_{21}^h + \widetilde{m_{21}})\widetilde{y_2} + \cdots \\
&\quad + (m_{n1}^h + \widetilde{m_{n1}})\widetilde{y_n} + \widetilde{m_{11}}y_1^h + \cdots + \widetilde{m_{n1}}y_n^h + v_0^1, \\
&\vdots \\
-\Delta\widetilde{y_n} &\equiv f_n(w) = (m_{1n}^h + \widetilde{m_{1n}})\widetilde{y_1} + (m_{2n}^h + \widetilde{m_{2n}})\widetilde{y_2} + \cdots \\
&\quad + (m_{nn}^h + \widetilde{m_{nn}} - q)\widetilde{y_n} + \widetilde{m_{1n}}y_1^h + \cdots + \widetilde{m_{nn}}y_n^h + v_0^n, \\
(\tilde{y}_i, \tilde{\phi}_j) &= 0 \quad (1 \leqslant i, j \leqslant n),
\end{aligned} \tag{4}$$

where we have defined the residual error v_0^i for each $1 \leqslant i \leqslant n$, by

$$v_0^i = \Delta y_i^h + (m_{ii}^h - q)y_i^h + m_{1i}^h y_1^h + m_{2i}^h y_2^h + \cdots + m_{ni}^h y_n^h.$$

Here we have assumed that $S_h \subset H_0^1 \cap H^2$. Then using the map $F(w)$ on V defined as

$$F(w) \equiv (Kf_1(w), \ldots, Kf_n(w), \widetilde{m_{11}} + (\tilde{y}_1, \tilde{\phi}_1), \ldots, \widetilde{m_{nn}} + (\tilde{y}_n, \tilde{\phi}_n)), \tag{5}$$

where K is the solution operator for the Poisson equation with homogeneous boundary condition, we have the fixed point equation

$$w = F(w). \tag{6}$$

Now, we assume the following.

Assumption 1. When we denote the Fréchet derivative of F at $\rho \in V$ by $F'(\rho)$, we assume that $[I - P_h F'(0)]_h$, which is the restriction to V_h of the operator $P_h[I - F'(0)] : V \to V_h$, is invertible,

$$[I - P_h F'(0)]_h^{-1} : V_h \to V_h.$$

The validity of this assumption can be numerically confirmed in the actual computations.

Now, following the method of [4], we decompose (6) into its finite-dimensional part and infinite-dimensional part as follows:

$$P_h w = P_h F(w),$$

$$(I - P_h)w = (I - P_h)F(w).$$

Next, we use a Newton-like method for the finite-dimensional part

$$N_h(w) := w_h - [I - P_h F'(0)]_h^{-1}(w_h - P_h F(w)).$$

Then defining $T(w) := N_h(w) + (I - P_h)F(w)$ the following equivalence relation holds:

$$w = T(w) \quad \Leftrightarrow \quad w = F(w).$$

In what follows, for the sake of simplicity, but without loss of generality, we consider only the case of $n=2$, i.e., an eigenvalue of multiplicity two.

As in [10], we use Banach's fixed point theorem to verify the solution of $w=T(w)$. We attempt to find a set W, referred as a 'candidate set', that satisfies the condition of the fixed point theorem.

We decompose a candidate set as $W = W_h \oplus W_\perp$, where $W_h \subset V_h$ and $W_\perp \subset V_h^\perp$. Here, $V_h^\perp$ is the orthogonal complement of V_h in the space V. We consider a candidate set of the form $W = W_h \oplus W_\perp$, $W_h = (\sum_{i=1}^{N} \mathscr{W}_i\phi_i, \sum_{i=N+1}^{2N} \mathscr{W}_i\phi_i, \mathscr{W}_{2N+1}, \mathscr{W}_{2N+2}, \mathscr{W}_{2N+3}, \mathscr{W}_{2N+4})$, with interval coefficient of basis $\mathscr{W}_i=[-W_i, W_i]$ $(1 \leqslant i \leqslant 2N+4)$, $W_\perp=([\alpha],[\beta],0,0,0,0)$, where for $\alpha\in\mathbb{R}^+$, $[\alpha]\equiv\{v\in S_h^\perp \mid \|v\|_{H_0^1}\leqslant\alpha\}$. Here, the linear combination of base functions with interval coefficients is interpreted as $\sum_{i=1}^{N}\mathscr{W}_i\phi_i \equiv \{\phi\in S_h \mid \phi=\sum_{i=1}^{N} a_i\phi_i,\ a_i\in\mathscr{W}_i,\ 1\leqslant i\leqslant N\}$, and, $S_h^\perp$ denotes the orthogonal complement of S_h in H_0^1.

Let T' be the Fréchet derivative of T. Then, we conceptually describe the verification condition employing the Banach fixed point theorem as

$$T(0)+T'(W)W\subset W,$$

where

$$T'(W)W:=\{v\in V \mid v=T'(\tilde{w})w,\ \tilde{w},w\in W\}.$$

We now present a computable verification condition.

We denote $(I-P_h)T(0)$ and $(I-P_h)T'(W)W$ by $T_\perp(0)$ and $T'_\perp(W)W$, respectively, and for an element $w_\perp=(w_1,w_2,0,0,0,0)\in V_h^\perp$, we set $(w_\perp)_i:=w_i$ $(i=0,1)$. Also, for an element $\Phi_h\in V_h$ or a set $\Phi_h\subset V_h$ of the form

$$\Phi_h=\left(\sum_{i=1}^{N}\mathscr{A}_i\phi_i, \sum_{i=N+1}^{2N}\mathscr{A}_i\phi_i, \mathscr{A}_{2N+1},\ldots,\mathscr{A}_{2N+4}\right),$$

we define $(\Phi_h)_i\equiv\mathscr{A}_i$ $(1\leqslant i\leqslant 2N+4)$, which is sometimes called a 'coefficient vector' for Φ_h. Then, we attempt to find the $(2N+6)$-dimensional vectors X and Z whose components $X_i>0$ and $Z_i\geqslant 0$ $(1\leqslant i\leqslant 2N+6)$ satisfy

$$(P_hT(0))_i\in\mathscr{X}_i,$$

$$\|(T_\perp(0))_1\|_{H_0^1}\leqslant X_{2N+5},$$

$$\|(T_\perp(0))_2\|_{H_0^1}\leqslant X_{2N+6} \tag{7}$$

and

$$(P_hT'(W)W)_i\subset\mathscr{Z}_i,$$

$$\|(T'_\perp(W)W)_1\|_{H_0^1}\leqslant Z_{2N+5},$$

$$\|(T'_\perp(W)W)_2\|_{H_0^1}\leqslant Z_{2N+6}, \tag{8}$$

where $\mathscr{X}_i = [-X_i, X_i]$ and $\mathscr{Z}_i = [-Z_i, Z_i]$ $(1 \leqslant i \leqslant 2N+4)$. Then for a set Φ, define $\|\Phi\|_{H_0^1} \equiv \sup_{\phi \in \Phi} \|\phi\|_{H_0^1}$, and with this, define the set

$$\Theta(W) = \{v \in V \mid (P_h v)_i \leqslant X_i + Z_i,\ 1 \leqslant i \leqslant 2N+4,\ \|((I-P_h)v)_1\|_{H_0^1} \leqslant X_{2N+5} + Z_{2N+5},$$
$$\|((I-P_h)v)_2\|_{H_0^1} \leqslant X_{2N+6} + Z_{2N+6}\}.$$

Then, we present our verification conditions as follows.

Theorem 1. *If $\Theta(W) \subset W$ holds for a candidate set W that is, if for X_i and Z_i satisfying* (7) *and* (8),

$$X_i + Z_i \leqslant W_i, \quad 1 \leqslant i \leqslant 2N+6$$

hold (where $W_{2N+5} = \alpha, W_{2N+6} = \beta$), then there exists a solution to (6) *in $\Theta(W)$. Moreover, this solution is unique in the set W.*

The proof of this theorem is almost the same as the corresponding proof in [10].

2.3. Verification procedures by numerical computation

In this subsection, we describe the actual computational procedures undertaken for the verification condition described in Theorem 1.

As preparation for the following discussion, we define Φ_i^1, Φ_i^2 and $\Phi_i \in V_h$ by $\Phi_i^1 = (\phi_i, 0, 0, 0, 0, 0)$ and $\Phi_i^2 = (0, \phi_i, 0, 0, 0, 0)$ for $1 \leqslant i \leqslant N$, and $\Phi_1 = (0, 0, 1, 0, 0, 0)$, $\Phi_2 = (0, 0, 0, 1, 0, 0)$, $\Phi_3 = (0, 0, 0, 0, 1, 0)$, and $\Phi_4 = (0, 0, 0, 0, 0, 1)$.

Now, we calculate $P_h T(0)$. The following equality holds from the definition of T:

$$P_h T(w) = [I - P_h F'(0)]_h^{-1} (P_h F(w) - P_h F'(0) P_h w).$$

If we set $w = 0$ in the above, then we have

$$[I - P_h F'(0)]_h P_h T(0) = P_h F(0). \tag{9}$$

Next, we consider the inner product on V

$$\langle [I - P_h F'(0)]_h P_h T(0), \Phi \rangle = \langle P_h F(0), \Phi \rangle, \tag{10}$$

where Φ stands for Φ_i^k or Φ_j. Now, we assume that $P_h T(0)$ can be written by

$$P_h T(0) = \left(\sum_{j=1}^{N} T_j^1 \phi_j, \sum_{j=1}^{N} T_j^2 \phi_j, T_{2N+1}, \ldots, T_{2N+4} \right).$$

Then, for each i $(1 \leqslant i \leqslant N)$, the following holds:

$$\langle [I - P_h F'(0)]_h P_h T(0), \Phi_i^1 \rangle = \sum_{j=1}^{N} T_j^1 \{(\nabla \phi_j, \nabla \phi_i) + ((q - m_{11}^h)\phi_j, \phi_i)\} - \sum_{j=1}^{N} T_j^2 (m_{21}^h \phi_j, \phi_i)$$
$$- T_{2N+1}(y_1^h, \phi_i) - T_{2N+3}(y_2^h, \phi_i).$$

Similar relation holds for Φ_i^2. Further, for Φ_1, we have

$$\langle [I - P_h F'(0)]_h P_h T(0), \Phi_1 \rangle = -\sum_{j=1}^{N} T_j^1 (\tilde{\phi}_1, \phi_j).$$

For other Φ_j, it also has similar expression as above.

Next, we consider the computation of the right-hand side of (10) for $\Phi = \Phi_i^k$ and $\Phi = \Phi_j$. For Φ_i^1, we have, by straightforward calculation,

$$\langle P_h F(0), \Phi_i^1 \rangle = (\Delta y_1^h + (m_{11}^h - q) y_1^h + m_{21}^h y_2^h, \phi_i).$$

Similarly, for Φ_i^2, we have

$$\langle P_h F(0), \Phi_i^2 \rangle = (\Delta y_2^h + m_{12}^h y_1^h + (m_{22}^h - q) y_2^h, \phi_i).$$

For Φ_j $(1 \leqslant j \leqslant 4)$, clearly $\langle P_h F(0), \Phi_j \rangle = 0$. Thus, we obtain the following $(2N+4)$-dimensional vector K_0 corresponding to the right-hand side of (11).

$$K_0 = \begin{pmatrix} (\Delta y_1^h + (m_{11}^h - q) y_1^h + m_{21}^h y_2^h, \phi_i) \\ (\Delta y_2^h + m_{12}^h y_1^h + (m_{22}^h - q) y_2^h, \phi_i) \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

To determine $P_h T(0)$, we solve the linear equation to get $P_h T(0)$.

$$G B^{(0)} = K_0, \tag{11}$$

where $B^{(0)}$ denotes the coefficient vector for $P_h T(0)$ and $G = (g_{ij})$ is a $(2N+4) \times (2N+4)$ matrix that corresponds to the left-hand side of (11) given as follows:

$$\begin{aligned} g_{ij} = {} & (\nabla\phi_j, \nabla\phi_i) + ((q - m_{11}^h)\phi_j, \phi_i), \quad (1 \leqslant i,j \leqslant N), \\ & -(m_{21}^h \phi_j, \phi_i), \quad (1 \leqslant i \leqslant N, N+1 \leqslant j \leqslant 2N), \\ & -(m_{12}^h \phi_j, \phi_i), \quad (N+1 \leqslant i \leqslant 2N, 1 \leqslant j \leqslant N), \\ & (\nabla\phi_j, \nabla\phi_i) + ((q - m_{22}^h)\phi_j, \phi_i), \quad (N+1 \leqslant i,j \leqslant 2N), \end{aligned}$$

$$\begin{aligned} g_{i,2N+1} &= -(y_1^h, \phi_i), \quad (1 \leqslant i \leqslant N), \\ g_{i,2N+2} &= -(y_1^h, \phi_i), \quad (N+1 \leqslant i \leqslant 2N), \\ g_{i,2N+3} &= -(y_2^h, \phi_i), \quad (1 \leqslant i \leqslant N), \end{aligned}$$

$$g_{i,2N+4} = -(y_2^h, \phi_i), \quad (N+1 \leqslant i \leqslant 2N),$$

$$g_{2N+1,j} = -(\phi_j, \tilde{\phi}_1), \quad (1 \leqslant j \leqslant N),$$

$$g_{2N+2,j} = -(\phi_j, \tilde{\phi}_1), \quad (N+1 \leqslant j \leqslant 2N),$$

$$g_{2N+3,j} = -(\phi_j, \tilde{\phi}_2), \quad (1 \leqslant j \leqslant N),$$

$$g_{2N+4,j} = -(\phi_j, \tilde{\phi}_2), \quad (N+1 \leqslant j \leqslant 2N),$$

$$g_{ij} = 0 \quad \text{(otherwise)}.$$

Thus, we can determine X_i in (7). In what follows, we restrict the domain Ω to the rectangle $(0,\pi)\times(0,\pi)$. However, as readily seen, our technique can be applied to more general domains by using other approximation subspaces with constructive error estimates, e.g., finite element method.

Particularly, we choose the approximation subspace $S_h \subset H_0^1$ as

$$S_h = \operatorname{span}\{2/\pi \sin(ix)\sin(jy), 1 \leqslant i \leqslant M_0, 1 \leqslant j \leqslant N_0, \text{ for some integers } M_0 \text{ and } N_0\}.$$

Then, we can estimate $T_\perp(0)$ using the following lemma.

Lemma 1. *Defining $\mathcal{N} \equiv \min\{N_0, M_0\}$, for any $u \in H_0^1 \cap H^2$, we have*

$$\|u - P_{h1}u\|_{H_0^1} \leqslant \frac{1}{\sqrt{(\mathcal{N}+1)^2+1}} \|\Delta u\|. \tag{12}$$

Here, $\| \ \|$ represents the L^2—norm.

Proof. We use the Fourier expansion of $u \in H_0^1$ of the form

$$u = \sum_{i,j=1}^{\infty} A_{ij} \frac{2}{\pi} \sin(ix)\sin(jy),$$

where $A_{ij} \in \mathbb{R}$. Here, we have used the fact that the H_0^1-projection $P_{h1}u$ of u into S_h coincides with the truncation of u up to $i = M_0,\ j = N_0$. Therefore, we have

$$\begin{aligned}
\|u - P_{h1}u\|_{H_0^1}^2 &= \sum_{i=M_0+1}^{\infty} \sum_{j=1}^{N_0} A_{ij}^2(i^2+j^2) + \sum_{i=1}^{M_0} \sum_{j=N_0+1}^{\infty} A_{ij}^2(i^2+j^2) + \sum_{i=M_0+1}^{\infty} \sum_{j=N_0+1}^{\infty} A_{ij}^2(i^2+j^2) \\
&\leqslant \max_{i>M_0, j>N_0} \frac{1}{i^2+j^2} \sum_{i,j=1}^{\infty} A_{ij}^2(i^2+j^2)^2 \\
&= \frac{1}{(\mathcal{N}+1)^2+1} \sum_{i,j=1}^{\infty} A_{ij}^2(i^2+j^2)^2
\end{aligned}$$

Also, we have by Parseval's equality

$$\|\Delta u\|_{L_2}^2 = \sum_{i,j=1}^{\infty} A_{ij}^2 (i^2+j^2)^2.$$

Thus, we obtain

$$\|u - P_{h1}u\|_{H_0^1}^2 \leqslant \frac{1}{(\mathcal{N}+1)^2+1} \|\Delta u\|_{L_2}|^2. \qquad \square$$

Using Lemma 1 as $u = (T(0))_1$ or $(T(0))_2$, we have the following:

$$\begin{aligned}\|(T_\perp(0))_1\|_{H_0^1} = \|(I-P_h)Kf_1(0)\|_{H_0^1} &\leqslant \frac{1}{\sqrt{(\mathcal{N}+1)^2+1}} \|f_1(0)\|_{L_2} \\ &= \frac{1}{\sqrt{(\mathcal{N}+1)^2+1}} \|v_0^1\|_{L_2} \equiv X_{2N+5} \end{aligned} \tag{13}$$

and

$$\begin{aligned}\|(T_\perp(0))_2\|_{H_0^1} = \|(I-P_h)Kf_2(0)\|_{H_0^1} &\leqslant \frac{1}{\sqrt{(\mathcal{N}+1)^2+1}} \|f_2(0)\|_{L_2} \\ &= \frac{1}{\sqrt{(\mathcal{N}+1)^2+1}} \|v_0^2\|_{L_2} \equiv X_{2N+6}. \end{aligned} \tag{14}$$

Next, we calculate Z_i in (9). Observe that the following equality holds for any $w, \tilde{w} \in W$:

$$[I - P_h F'(0)]_h (P_h T'(\tilde{w})w) = P_h(F'(\tilde{w})w - F'(0)P_h w). \tag{15}$$

Now, we consider an inner product on V analogous to (11):

$$\langle [I - P_h F'(0)]_h (P_h T'(\tilde{w})w), \Phi\rangle = \langle P_h(F'(\tilde{w})w - F'(0)P_h w), \Phi\rangle. \tag{16}$$

Here, Φ stands for Φ_i^k $(1 \leqslant i \leqslant N,\ k=1,2)$ or Φ_j $(1 \leqslant j \leqslant 4)$. When we write $P_h T'(\tilde{w})w$ as

$$P_h T'(\tilde{w})w = \left(\sum_{j=1}^{N} T_j^1 \phi_j, \sum_{j=1}^{N} T_j^2 \phi_j, T_{2N+1}, \ldots, T_{2N+4} \right),$$

we obtain the equality

$$GB^{(1)} = K_1, \tag{17}$$

where $B^{(1)}$ is the coefficient vector for $P_h T'(\tilde{w})w$, and for $1 \leqslant i \leqslant N$

$$K_1 = \begin{pmatrix} (f_1'(\tilde{w})w - f_1'(0)P_h w, \phi_i) \\ (f_2'(\tilde{w})w - f_2'(0)P_h w, \phi_i) \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

By solving (17), we can determine Z_i in (9).

Next, we find bounds for $T'_\perp(W)W$. We have, for $w,\tilde{w}\in W$,

$$(T'_\perp(\tilde{w})w)_i = ((I-P_h)T'(\tilde{w})w)_i = (I-P_h)Kf'_i(\tilde{w})w, \quad 1\leqslant i\leqslant 2. \tag{18}$$

Therefore, we have, by Lemma 1 and (18)

$$\|(T'_\perp(\tilde{w})w)_1\|_{H_0^1} \leqslant \sup\nolimits_{\tilde{w},w} \frac{1}{\sqrt{(\mathcal{N}+1)^2+1}}\|f'_1(\tilde{w})w\| \equiv Z_{2N+5},$$

$$\|(T'_\perp(\tilde{w})w)_2\|_{H_0^1} \leqslant \sup\nolimits_{\tilde{w},w} \frac{1}{\sqrt{(\mathcal{N}+1)^2+1}}\|f'_2(\tilde{w})w\| \equiv Z_{2N+6}. \tag{19}$$

Collecting the above results, we can write the verification conditions in Theorem 1 as

$$X_i + Z_i \leqslant W_i, \quad 1\leqslant i\leqslant 2N+4,$$

$$\sup_{\tilde{w},w} \frac{1}{\sqrt{(\mathcal{N}+1)^2+1}}(\|v_0^1\| + \|f'_1(\tilde{w})w\|) \leqslant \alpha,$$

$$\sup_{\tilde{w},w} \frac{1}{\sqrt{(\mathcal{N}+1)^2+1}}(\|v_0^2\| + \|f'_2(\tilde{w})w\|) \leqslant \beta. \tag{20}$$

To end this section, we describe the algorithm for the case of verifying an eigenvalue of multiplicity two.

Algorithm.

1. Fix a maximum iteration number k.
2. Find approximate solutions λ_1^h and λ_2^h, y_1^h and y_2^h for (1) or (2).
3. Determine a candidate set W for the interval coefficient $\mathcal{W}_i$, $(1\leqslant i\leqslant 2N+4)$ with small width and $\alpha\in\mathbb{R}^+$, $\beta\in\mathbb{R}^+$.
4. Compute X, Z, $\|v_0^1\|$, $\|v_0^2\|$, $\|f'_1(\tilde{w})w\|$ and $\|f'_2(\tilde{w})w\|$ from the definition of the candidate set W using α, and W_i in (2.2).
5. Check the verification condition (20). If the condition is satisfied, the verification has succeeded. If not, we carry out an inflation (see [7]) of the candidate set W, i.e., set

$$W_i = (1+\delta)(X_i+Z_i),$$

$$\alpha = (1+\delta)\sup_{\tilde{w},w} \frac{1}{\sqrt{(\mathcal{N}+1)^2+1}}(\|v_0^1\| + \|f'_1(\tilde{w})w\|),$$

$$\beta = (1+\delta)\sup_{\tilde{w},w} \frac{1}{\sqrt{(\mathcal{N}+1)^2+1}}(\|v_0^2\| + \|f'_2(\tilde{w})w\|).$$

Repeat procedures 4 and 5 until the iteration number exceeds k.

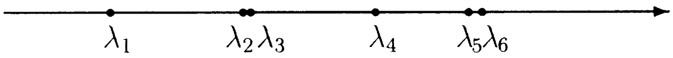

Fig. 1. Approximate eigenvalues.

Remark 1. From the arguments above, it is readily seen that we can apply our method to nonself-adjoint eigenvalue problems of the form

$$-\Delta u + k \cdot \nabla u + qu = \lambda u \quad \text{in } \Omega,$$
$$u = 0 \quad \text{on } \partial\Omega, \tag{21}$$

where $k = (k_1, k_2)$, $k_i \in L^\infty(\Omega)$ $(i = 1, 2)$.

3. Numerical examples

Example 1. We considered the following problem:

$$-\Delta u + \sin(x)\sin(y)u = \lambda u \quad \text{in } \Omega,$$
$$u = 0 \quad \text{on } \partial\Omega, \quad \Omega = (0, \pi) \times (0, \pi). \tag{22}$$

We consider the following finite-dimensional subspace, ($\mathcal{N} = 11$) in Lemma 1,

$$S_h = \operatorname{span}\{2/\pi \sin(ix)\sin(jy) \mid 1 \leqslant i, j \leqslant 11\}.$$

Let $\lambda_h \in \mathbb{R}$ and $y_h \in S_h$ be the Galerkin approximate solutions of (22) defined by

$$(\nabla y_h, \nabla v) + (\sin(x)\sin(y)y_h, v) = (\lambda_h y_h, v) \quad \text{for all } v \in S_h.$$

We numerically determined approximate eigenvalues for (22). The first eigenvalue was found to be $\lambda_1 \approx 2.71513$. This is seen to be simple. The second and third eigenvalues were found to be $\lambda_2 \approx 5.572374582086$ and $\lambda_3 \approx 5.572374582086$. These eigenvalues are depicted in Fig. 1. The numerically determined approximate eigenfunctions y_h^1 and y_h^2 are illustrated in Figs. 2 and 3, respectively.

These two eigenvalues seemed to be two-fold or clustered eigenvalues. Therefore, we verified them and the basis of the corresponding invariant subspace around the approximate solutions using the algorithm described in the previous section.

The normalized eigenequation in question is the following:

$$(-\Delta + q)(y_1, y_2) = (y_1, y_2)\begin{pmatrix} m_{11}, m_{12} \\ m_{21}, m_{22} \end{pmatrix},$$

$$(y_1, \tilde{\phi}_1) = (y_1^h, \tilde{\phi}_1),$$
$$(y_2, \tilde{\phi}_1) = (y_2^h, \tilde{\phi}_1),$$
$$(y_1, \tilde{\phi}_2) = (y_1^h, \tilde{\phi}_2),$$
$$(y_2, \tilde{\phi}_2) = (y_2^h, \tilde{\phi}_2).$$

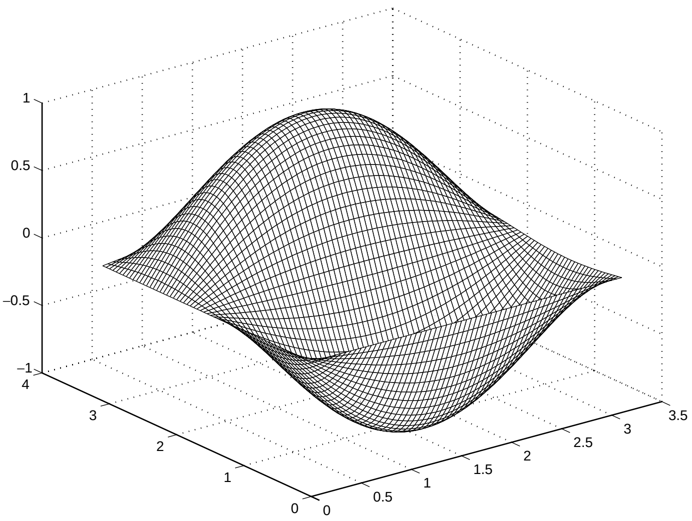

Fig. 2. Approximate eigenfunction y_1^h.

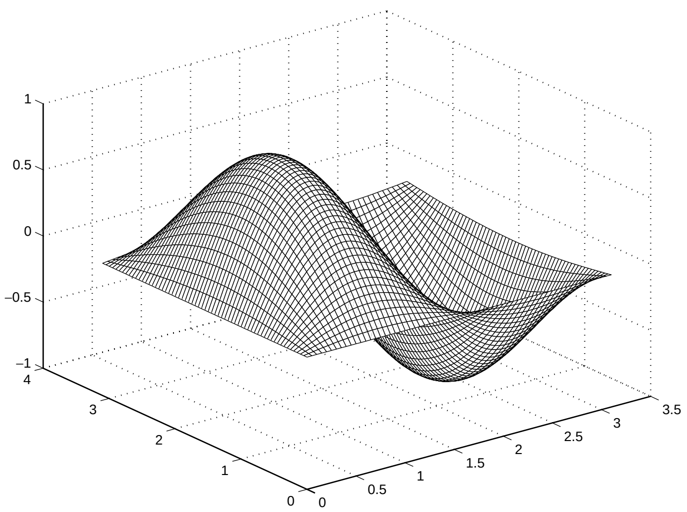

Fig. 3. Approximate eigenfunction y_2^h.

Here, $\tilde{\phi}_i$ is taken to be the base function described in Section 2.1. Then, as discussed in Section 2.2, we set $y_i = y_i^h + \tilde{y}_i$ and $m_{ij} = m_{ij}^h + \widetilde{m_{ij}}$ $(1 \leqslant i, j \leqslant 2)$ with $m_{ii}^h = 5.572374582086$, $m_{ij}^h = 0$ $(i \neq j)$.

The verification results are as follows. First, the residual errors obtained are $\|v_0^1\|_{L_2} = 0.00489$ and $\|v_0^2\|_{L_2} = 0.00489$. Eqs. (23) and (24) give the error bounds of the finite-dimensional part of the error from the base functions, i.e., the coefficient vector of $P_{h1}\tilde{y}_i$ $(i = 1, 2)$ in the corresponding invariant subspaces. Eq. (25) gives the H_0^1 error bounds of the infinite-dimensional part, i.e., $(I - P_{h1})\tilde{y}_i$ $(i = 1, 2)$:

$$\max(|(X + Z)_j|) = 0.0905 \times 10^{-4}, \quad 1 \leqslant j \leqslant 11^2, \tag{23}$$

$$\max(|(X+Z)_j|) = 0.1071 \times 10^{-4}, \quad 11^2+1 \leqslant j \leqslant 2\times 11^2, \tag{24}$$

$$\alpha = 3.6023 \times 10^{-4},$$
$$\beta = 4.3229 \times 10^{-4}. \tag{25}$$

The elements of the matrix $M=(m_{ij})$ were enclosed as described in (26). The eigenvalues of M were enclosed by using Gerschgorin circles as given in (27): In this example, verification succeeded after 4 iterations, and we used the value 0.1 for the inflation parameter δ in the algorithm.

$$m_{11} \in 5.5724 + [-0.2217, 0.2217] \times 10^{-4},$$
$$m_{12} \in [-0.2550, 0.2550] \times 10^{-4},$$
$$m_{21} \in [-0.2217, 0.2217] \times 10^{-4},$$
$$m_{22} \in 5.5724 + [-0.2550, 0.2550] \times 10^{-4}, \tag{26}$$

$$\lambda_2, \lambda_3 \in 5.5724 + [-0.4767, 0.4767] \times 10^{-4}. \tag{27}$$

We next numerically determined the fourth eigenvalue to be $\lambda_4 \approx 8.4581\ 45330119$ and found that it is simple. Then we attempt to verify two eigenvalues λ_5 and λ_6 to be close together. Approximate solutions of λ_5 and λ_6 were 10.524940396607 and 10.584986363725, respectively. In this verification procedure, we used the finite-dimensional subspace such as

$$S_h = \mathrm{span}\{2/\pi \sin(ix)\sin(jy) \mid 1 \leqslant i,j \leqslant 8\}. \tag{28}$$

The verification results are as follows. The residual errors are $\|v_0^1\| = 0.01339$, and $\|v_0^2\| = 0.01622$. Eqs. (29) and (30) give the error bounds of the finite-dimensional part of the base functions, i.e., the coefficient vector of $P_{h1}\tilde{y}_i$ $(i=1,2,)$ in the corresponding invariant subspaces. Eq. (31) gives the H_0^1 error bounds of the infinite-dimensional part, i.e., $(I-P_{h1})\tilde{y}_i$ $(i=1,2)$.

$$\max(|(X+Z)_j|) = 0.0567 \times 10^{-3}, \quad 1 \leqslant j \leqslant 8^2, \tag{29}$$

$$\max(|(X+Z)_j|) = 0.1032 \times 10^{-3}, \quad 8^2+1 \leqslant j \leqslant 2\times 8^2, \tag{30}$$

$$\alpha = 0.0020,$$
$$\beta = 0.0036. \tag{31}$$

The eigenvalues of M were enclosed by using Gerschgorin circles as follows. In this case verification succeeded after 5 iterations with inflation parameter $\delta = 0.1$. As seen in Eqs. (33) and (34), we were able to enclose two distinct eigenvalues. However, note that when we attempted to verify λ_5 and λ_6 separately as two simple eigenvalues, applying a method similar to that in [4] with the same approximation space S_h, the verification failed.

In the application of the present algorithm, the condition number of the matrix G in (11) and (17) was 121.51. Contrastingly using the method for simple eigenvalues, this quantity became as large as $\approx 3\times 10^3$. This fact demonstrates the difference between the performances of the two enclosure

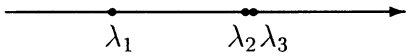

Fig. 4. Approximate eigenvalues.

methods.

$$
\begin{aligned}
&m_{11} \in 10.584986363725 + [-0.2117, 0.2117] \times 10^{-3},\\
&m_{12} \in [-0.3745, 0.3745] \times 10^{-3},\\
&m_{21} \in [-0.2053, 0.2053] \times 10^{-3},\\
&m_{22} \in 10.524940396607 + [-0.3632, 0.3632] \times 10^{-3},
\end{aligned} \tag{32}
$$

$$\lambda_5 \in 10.524940396607 + [-0.5685, 0.5685] \times 10^{-3}, \tag{33}$$

$$\lambda_6 \in 10.584986363725 + [-0.4170, 0.4170] \times 10^{-3}. \tag{34}$$

Example 2. Here we consider the following eigenvalue problem.

$$
\begin{aligned}
&-\Delta u + k \cdot \nabla u + qu = \lambda u \quad \text{in } \Omega,\\
&u = 0 \quad \text{on } \partial\Omega,
\end{aligned} \tag{35}
$$

where $k := (k_1, k_2)$ for real constants k_1 and k_2, and $q = \sin(x)\sin(y)$, $\Omega = (0,\pi) \times (0,\pi)$.

We chose the parameters as $k_1 = k_2 = 0.1$ here. The approximate first eigenvalue of this problem is $\lambda_1 \approx 2.72013$ which is simple, and the second and the third eigenvalues are $\lambda_2 \approx 5.57737396314098$, $\lambda_3 \approx 5.57737396314351$ as depicted in Fig. 4. The eigenvalues λ_2 and λ_3 are either equal or almost equal. Therefore, we verified them together with a basis for the corresponding invariant subspace.

We used the finite-dimensional subspace

$$S_h = \text{span}\{2/\pi \sin(ix)\sin(jy) \mid 1 \leqslant i, j \leqslant 30\}.$$

The verification results are as follows. The residual errors are $\|v_0^1\| = 0.03824$, $\|v_0^2\| = 0.03809$.

Eqs. (36) and (37) give the error bounds of the finite-dimensional part of the base functions, i.e., the coefficient vector of $P_{h1}\tilde{y}_i$ $(i = 1, 2)$ in the corresponding invariant subspaces, and (38) gives the H_0^1 error bounds of the infinite-dimensional part, i.e., $(I - P_{h1})\tilde{y}_i$ $(i = 1, 2)$.

$$\max(|(X+Z)_i|) = 8.54939 \times 10^{-5}, \quad 1 \leqslant i \leqslant 30^2, \tag{36}$$

$$\max(|(X+Z)_i|) = 8.94350 \times 10^{-5}, \quad 30^2 + 1 \leqslant i \leqslant 2 \times 30^2, \tag{37}$$

$$
\begin{aligned}
&\alpha = 1.3023 \times 10^{-3},\\
&\beta = 1.3623 \times 10^{-3}.
\end{aligned} \tag{38}
$$

The eigenvalues of $M = (m_{ij})$ were enclosed by using Gerschgorin circles. In this case, verification succeeded after 10 iteration. and we used the value 0.00001 for the inflation parameter δ in the algorithm.

$$m_{11} \in 5.57737396314351 + [-3.3836, 3.3836] \times 10^{-4},$$

$$m_{12} \in [-3.5396, 3.5396] \times 10^{-4},$$

$$m_{21} \in [-3.1876, 3.1876] \times 10^{-4},$$

$$m_{22} \in 5.57737396314098 + [-3.3346, 3.3346] \times 10^{-4}.$$

Denoting the ball in the complex plane of radius ε centered at the origin by $B(\varepsilon)$, for $\varepsilon_1 = 6.5712 \times 10^{-4}$, $\varepsilon_2 = 6.5222 \times 10^{-4}$, these eigenvalues were enclosed as

$$\lambda_2, \lambda_3 \in (5.57737396314351 + B(\varepsilon_1)) \cup (5.57737396314098 + B(\varepsilon_2)). \tag{39}$$

All computations were performed using INTLAB [8], an interval package for use under Matlab V5.3.1 [2].

Remark 2. By some transformation, e.g., [5], the problem in Example 2 can be reduced to the self-adjoint eigenvalue problem. However, our verification result implies that we could also enclose eigenvalues for the problems with some small non-self-adjoint perturbation, which are no longer possible to be transformed to the self-adjoint case.

Acknowledgements

We would like to thank referees for their helpful comments.

References

[1] H. Behnke, F. Goerisch, Inclusions for eigenvalues of self-adjoint problems, in: J. Herzberger (Ed.), Topics in Validated Computations, Elsevier, Amsterdam, Lausanne, New York, Oxford, Shannon, Tokyo, 1994, pp. 277–322.
[2] MATLAB User's Guide for UNIX Workstations, The MathWorks Inc., 1996.
[3] M.T. Nakao, Solving nonlinear elliptic problems with result verification using an H^{-1} residual iteration, Computers 9 (Suppl.) (1993) 161–173.
[4] M.T. Nakao, N. Yamamoto, K. Nagatou, Numerical verifications for eigenvalues of second-order elliptic operators, Jpn. J. Ind. Appl. Math. 16 (1998) 307–320.
[5] M. Plum, Numerical existence proofs and explicit bounds for solutions of nonlinear elliptic boundary value problems, Computing 49 (1992) 25–44.
[6] M. Plum, Guaranteed numerical bounds for eigenvalues, in: D. Hinton, P.W. Schaefer (Eds.), Spectral Theory and Computational Methods of Sturm–Liouville Problems, Dekker, New York, 1997, pp. 313–332.
[7] S.M. Rump, Solving algebraic problems with high accuracy, in: U. Kulisch, W.L. Miranker (Eds.), A New Approach to Scientific Computation, Academic Press, New York, 1983.
[8] S.M. Rump, INTLAB—INTerval LABoratory, in: Tibor Cesndes (Ed.), Developments in Reliable Computing, Kluwer Academic Publishers, Dordrecht, 1999, pp. 77–104.

[9] S.M. Rump, Computational error bounds for multiple or nearly multiple eigenvalues, Linear Algebra Appl. 324 (2001) 209–226.
[10] N. Yamamoto, A numerical verification method for solutions of boundary value problems with local uniqueness by Banach's fixed point theorem, SIAM J. Numer. Anal. 35 (1998) 2004–2013.
[11] S. Zimmermann, U. Mertins, Variational bounds to eigenvalue of self-adjoint eigenvalue problems with arbitrary spectrum, Z. Anal. Anwendungen 14 (1995) 327–345.