Numer Algor (2017) 74:797–819
DOI 10.1007/s11075-016-0172-1

ORIGINAL PAPER

A spectral method for nonlinear elliptic equations

Kendall Atkinson[1] **· David Chien**[2] **· Olaf Hansen**[2]

Received: 1 February 2016 / Accepted: 27 June 2016 / Published online: 16 July 2016

Abstract Let Ω be an open, simply connected, and bounded region in $\mathbb{R}^d$, $d \geq 2$, and assume its boundary $\partial\Omega$ is smooth and homeomorphic to $\mathbb{S}^{d-1}$. Consider solving an elliptic partial differential equation $Lu = f(\cdot, u)$ over Ω with zero Dirichlet boundary value. The function f is a nonlinear function of the solution u. The problem is converted to an equivalent elliptic problem over the open unit ball $\mathbb{B}^d$ in $\mathbb{R}^d$, say $\widetilde{L}\widetilde{u} = \widetilde{f}(\cdot, \widetilde{u})$. Then a spectral Galerkin method is used to create a convergent sequence of multivariate polynomials $\widetilde{u}_n$ of degree $\leq n$ that is convergent to $\widetilde{u}$. The transformation from Ω to $\mathbb{B}^d$ requires a special analytical calculation for its implementation. With sufficiently smooth problem parameters, the method is shown to be rapidly convergent. For $u \in C^\infty\left(\overline{\Omega}\right)$ and assuming $\partial\Omega$ is a C^∞ boundary, the convergence of $\|\widetilde{u} - \widetilde{u}_n\|_{H^1}$ to zero is faster than any power of $1/n$. The error analysis uses a reformulation of the boundary value problem as an integral equation, and

✉ Kendall Atkinson
kendall-atkinson@uiowa.edu

Olaf Hansen
ohansen@csusm.edu

David Chien
chien@csusm.edu

[1] Departments of Mathematics & Computer Science, The University of Iowa, Iowa City, Iowa 52242, USA

[2] Department of Mathematics, California State University San Marcos, San Marcos, California 92096, USA

then it uses tools from nonlinear integral equations to analyze the numerical method. Numerical examples illustrate experimentally an exponential rate of convergence. A generalization to $-\Delta u + \gamma u = f(u)$ with a zero Neumann boundary condition is also presented.

Keywords Elliptic · Nonlinear · Spectral method

1 Introduction

Consider the nonlinear problem

$$Lu\,(s) = f\,(s, u(s))\,, \qquad s \in \Omega \tag{1}$$

$$u\,(s) = 0, \qquad s \in \partial\Omega \tag{2}$$

with L an elliptic operator over Ω and a homogeneous Dirichlet boundary condition. Let Ω be an open, simply connected, and bounded region in $\mathbb{R}^d$, and assume that its boundary $\partial\Omega$ is sufficiently differentiable and is homeomorphic to $\mathbb{S}^{d-1}$. Assume L is a strongly elliptic operator of the form

$$Lu(s) \equiv -\sum_{i,j=1}^{d} \frac{\partial}{\partial s_i}\left(a_{i,j}(s)\frac{\partial u(s)}{\partial s_j}\right) + \gamma\,(s)\,u\,(s)\,, \qquad s \in \Omega, \tag{3}$$

We present a spectral method for solving (1)–(2) based on multivariate polynomial approximation over the unit ball $\mathbb{B}^d$. Our numerical method is similar to that presented in earlier papers for linear problems; see [2, 6]. However, the nonlinearity in (1) leads to the solving of nonlinear algebraic systems. Moreover, the convergence analysis requires a new approach as the standard variational analysis applies to only the linear framework. We give a new error analysis that uses a reformulation of the problem (1)–(2) and its numerical approximation using nonlinear integral equations; see Section 3.

In (3), the functions $a_{i,j}(s)$, $1 \le i, j \le d$, are assumed to be several times continuously differentiable over $\overline{\Omega}$, and the $d \times d$ matrix $\left[a_{i,j}\,(s)\right]$ is to be symmetric and to satisfy

$$\xi^{\mathrm{T}} A(s)\xi \ge \alpha\xi^{\mathrm{T}}\xi, \qquad s \in \overline{\Omega}, \quad \xi \in \mathbb{R}^d \tag{4}$$

for some $\alpha > 0$. Also assume the coefficient $\gamma \in C\left(\overline{\Omega}\right)$. Note that because the right-hand function f is allowed to depend on u, an arbitrarily large multiple of u can be added to each side of (1), thus justifying an assumption that

$$\min_{s\in\overline{\Omega}} \gamma\,(s) > 0. \tag{5}$$

The problem (1)–(2) can be reformulated as a variational problem. Introduce

$$\begin{aligned}\mathcal{A}\,(v, w) = &\int_{\Omega}\left[\sum_{i,j=1}^{d} a_{i,j}(s)\frac{\partial v(s)}{\partial s_i}\frac{\partial w(s)}{\partial s_j}\right] ds \\ &+\int_{\Omega}\gamma\,(s)\,v\,(s)\,w\,(s)\;ds, \qquad v, w \in H_0^1\,(\Omega)\,,\end{aligned} \tag{6}$$

$$(\mathcal{F}(v))(s) = f(s, v(s)), \qquad s \in \Omega, \qquad v \in H^1(\Omega). \tag{7}$$

Note that the Sobolev space $H^m(\Omega)$ is the closure of $C^m\left(\overline{\Omega}\right)$ using the norm

$$\|g\|_{H^m(\Omega)} = \sqrt{\sum_{|i|\le m} \left\|D^i g\right\|^2_{L_2(\Omega)}}, \qquad g \in C^m\left(\overline{\Omega}\right), \quad m \ge 1$$

with i a multi-integer, $i = (i_1, \ldots, i_d)$, $|i| = i_1 + \cdots + i_d$, and

$$D^i g(s) = \frac{\partial^{|i|} g(s)}{\partial s_1^{i_1} \cdots \partial s_d^{i_d}}.$$

The space $H_0^1(\Omega)$ is the closure of $C_0^1(\Omega)$ using $\|\cdot\|_{H^1(\Omega)}$, where elements of $C_0^1(\Omega) \subseteq C^1\left(\overline{\Omega}\right)$ are zero on some open neighborhood of the boundary of Ω.

Noting (4) and (5), it can be assumed that $\mathcal{A}$ is a strongly elliptic operator on $H_0^1(\Omega)$, namely

$$\mathcal{A}(v, v) \ge c_0 \|v\|^2_{H^1(\Omega)}, \qquad \forall v \in H_0^1(\Omega)$$

for some finite $c_0 > 0$. Reformulate (1)-(2) as the following variational problem: find $u \in H_0^1(\Omega)$ for which

$$\mathcal{A}(u, w) = (\mathcal{F}(u), w), \qquad \forall w \in H_0^1(\Omega). \tag{8}$$

Throughout this paper, we assume the variational reformulation of the problem (1)–(2) has a locally unique solution $u \in H_0^1(\Omega)$. For analyses of the existence and uniqueness of a solution to (1)–(2), see Zeidler [16, Section 28.5].

In the following Section 2, we define our spectral method for the case that $\Omega = \mathbb{B}^d$; and following that we show how to reformulate the problem (1)–(2) for a general smooth region Ω as an equivalent problem over $\mathbb{B}^d$. This follows the earlier development in [2]. In Section 3, we present a convergence analysis for our numerical method, an approach using results from the numerical analysis of nonlinear integral equations. Implementation of the method is discussed in Section 4, followed by numerical examples in Section 5. An extension to a Neumann boundary value problem is given in Section 6.

2 A spectral method

Begin with the special case $\Omega = \mathbb{B}^d$, and then move to a general region Ω. Let $\mathcal{X}_n$ denote a finite-dimensional subspace of $H_0^1\left(\mathbb{B}^d\right)$, and let $\left\{\psi_1, \ldots, \psi_{N_n}\right\}$ be a basis of $\mathcal{X}_n$. Later a basis is given by using polynomials of degree $\le n$ over $\mathbb{R}^d$, denoted by Π_n^d, with N_n the dimension of Π_n^d. An approximating solution to (8) is sought by finding $u_n \in \mathcal{X}_n$ such that

$$\mathcal{A}(u_n, w) = (\mathcal{F}(u_n), w), \qquad \forall w \in \mathcal{X}_n. \tag{9}$$

More precisely, find

$$u_n(x) = \sum_{\ell=1}^{N_n} \alpha_\ell \psi_\ell(x) \tag{10}$$

that satisfies the nonlinear algebraic system

$$\begin{aligned}\sum_{\ell=1}^{N_n} \alpha_\ell \int_{\mathbb{B}^d} \left[\sum_{i,j=1}^{d} a_{i,j}(x) \frac{\partial \psi_\ell(x)}{\partial x_i} \frac{\partial \psi_k(x)}{\partial x_j} + \gamma(x)\, \psi_\ell(x)\, \psi_k(x) \right] dx \\ = \int_{\mathbb{B}^d} f\left(x, \sum_{\ell=1}^{N_n} \alpha_\ell \psi_\ell(x) \right) \psi_k(x)\, dx, \qquad k = 1, \ldots, N_n.\end{aligned} \tag{11}$$

As notation, we generally use the variable x when considering $\mathbb{B}^d$ and the variable s when considering Ω.

To obtain a space for approximating the solution u of (1)–(2), proceed as follows. Denote by Π_n^d, the space of polynomials in d variables that are of degree $\le n$: $p \in \Pi_n^d$ if it has the form

$$p(x) = \sum_{|i| \le n} a_i x_1^{i_1} x_2^{i_2} \ldots x_d^{i_d},$$

$i = (i_1, \ldots, i_d)$, $|i| = i_1 + \cdots i_d$. As the approximation space over $\mathbb{B}^d$, choose

$$\mathcal{X}_n = \left\{ \left(1 - |x|^2\right) p(x) \mid p \in \Pi_n^d \right\} \subseteq H_0^1\left(\mathbb{B}^d\right) \tag{12}$$

Let $N_n = \dim \mathcal{X}_n = \dim \Pi_n^d$. For $d = 2$, $N_n = (n+1)(n+2)/2$. Practical implementation of the numerical method (9)–(11) is discussed in Section 4.

2.1 Transformation of the domain Ω

For the more general problem (1)–(2) over a general region Ω, we reformulate it as a problem over $\mathbb{B}^d$. Begin by reviewing some ideas from [2], to which the reader is referred for additional details.

Assume the existence of a function

$$\Phi : \overline{\mathbb{B}}^d \underset{onto}{\xrightarrow{1-1}} \overline{\Omega} \tag{13}$$

with Φ a twice-differentiable mapping, and let $\Psi = \Phi^{-1} : \overline{\Omega} \underset{onto}{\xrightarrow{1-1}} \overline{\mathbb{B}}^d$. For $v \in L^2(\Omega)$, let

$$\widetilde{v}(x) = v(\Phi(x)), \qquad x \in \overline{\mathbb{B}}^d \tag{14}$$

and conversely for $\widetilde{v} \in L^2\left(\mathbb{B}^d\right)$,

$$v(s) = \widetilde{v}(\Psi(s)), \qquad s \in \overline{\Omega}. \tag{15}$$

Assuming $v \in H^1(\Omega)$, it is straightforward to show

$$\nabla_x \widetilde{v}(x) = J(x)^{\mathrm{T}} \nabla_s v(s), \qquad s = \Phi(x)$$

with $J(x)$ the Jacobian matrix for Φ over the closed unit ball $\overline{\mathbb{B}}^d$,

$$J(x) \equiv (D\Phi)(x) = \left[\frac{\partial \Phi_i(x)}{\partial x_j}\right]_{i,j=1}^{d}, \qquad x \in \overline{\mathbb{B}}^d. \tag{16}$$

To use our method for problems over a region Ω, it is necessary to know explicitly the functions Φ and J. The creation of such a mapping Φ is taken up in [5] for cases in which only a boundary mapping is known, from $\mathbb{S}^{d-1} \equiv \partial\mathbb{B}^d$ to $\partial\Omega$, a common way to define the region Ω.

Assume

$$\det J(x) \neq 0, \qquad x \in \overline{\mathbb{B}}^d. \tag{17}$$

Similarly,

$$\nabla_s v(s) = K(s)^{\mathrm{T}} \nabla_x \widetilde{v}(x), \qquad x = \Psi(s)$$

with $K(s)$ the Jacobian matrix for Ψ over Ω. By differentiating the identity

$$\Psi(\Phi(x)) = x, \qquad x \in \overline{\mathbb{B}}^d$$

it follows that

$$K(\Phi(x)) = J(x)^{-1}.$$

Assumptions about the differentiability of $\widetilde{v}(x)$ can be related back to assumptions on the differentiability of $v(s)$ and $\Phi(x)$.

Lemma 1 *Let* $\Phi \in C^m\left(\overline{\mathbb{B}}^d\right)$. *If* $v \in C^k\left(\overline{\Omega}\right)$, *then* $\widetilde{v} \in C^q\left(\overline{\mathbb{B}}^d\right)$ *with* $q = \min\{k, m\}$. *Similarly, if* $v \in H^k(\Omega)$, *then* $\widetilde{v} \in H^q\left(\mathbb{B}^d\right)$.

A proof is straightforward using (14). A converse statement can be made as regards $\widetilde{v}$, v, and Ψ in (15). Moreover, the differentiability of Φ over $\mathbb{B}^d$ is exactly the same as that of Ψ over Ω.

2.2 Reformulation from Ω to $\mathbb{B}^d$

Applying this transformation to the equation (1), it follows that

$$-\sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}\left(\det(J(x))\, \widetilde{a}_{i,j}(x) \frac{\partial \widetilde{u}(x)}{\partial x_j}\right) + \widetilde{\gamma}(x)\, \widetilde{u}(x) = \widetilde{f}(x, \widetilde{u}(x)), \qquad x \in \mathbb{B}^d, \tag{18}$$

where

$$\widetilde{f}(x, \widetilde{u}(x)) = \det(J(x))\, f(\Phi(x), \widetilde{u}(x)), \qquad x \in \mathbb{B}^d \tag{19}$$

$$\widetilde{\gamma}(x) = \det(J(x))\, \gamma(\Phi(x)) \tag{20}$$

and

$$\begin{aligned}\widetilde{A}(x) &= J(x)^{-1} A(\Phi(x)) J(x)^{-\mathrm{T}} \\ &\equiv \left[\widetilde{a}_{i,j}(x)\right]_{i,j=1}^{d}.\end{aligned} \tag{21}$$

A derivation of this is given in [2, Thm. 3]. With (18), also impose the Dirichlet condition

$$\widetilde{u}(x) = 0, \qquad x \in \mathbb{B}^d. \tag{22}$$

The problem of solving (18)–(22) is completely equivalent to that of solving (1)–(2). Also, the differential operator in (18) will be strongly elliptic. As noted earlier, the creation of such a mapping Φ is discussed at length in [5] for extending a boundary mapping $\varphi : \mathbb{S}^{d-1} \to \partial\Omega$ to a mapping Φ satisfying (13) and (17).

3 Error analysis

In [14], Osborn converted a finite element method for solving an eigenvalue problem for an elliptic partial differential equation to a corresponding numerical method for approximating the eigenvalues of a compact integral operator. He then used results for the latter to obtain convergence results for his finite element method. We use his construction to convert the numerical method for (8) to a corresponding method for finding a fixed point of a completely continuous nonlinear integral operator, and this latter numerical method will be analyzed using the results given in [12, Chap. 3] and [1].

Important results about polynomial approximation have been given recently by Li and Xu [10], and they are critical to our convergence analysis.

Theorem 2 *(Li and Xu) Let $r \geq 2$. Given $v \in H^r\left(\mathbb{B}^d\right)$, there exists a sequence of polynomials $p_n \in \Pi_n^d$ such that*

$$\|v - p_n\|_{H^1(\mathbb{B}^d)} \leq \varepsilon_{n,r} \|v\|_{H^r(\mathbb{B}^d)}, \qquad n \geq 1. \tag{23}$$

The sequence $\varepsilon_{n,r} = \mathcal{O}\left(n^{-r+1}\right)$ and is independent of v.

Theorem 3 *(Li and Xu) Let $r \geq 2$. Given $v \in H_0^1\left(\mathbb{B}^d\right) \cap H^r\left(\mathbb{B}^d\right)$, there exists a sequence of polynomials $p_n \in \mathcal{X}_n$ such that*

$$\|v - p_n\|_{H^1(\mathbb{B}^d)} \leq \varepsilon_{n,r} \|v\|_{H^r(\mathbb{B}^d)}, \qquad n \geq 1. \tag{24}$$

The sequence $\varepsilon_{n,r} = \mathcal{O}\left(n^{-r+1}\right)$ and is independent of v.

These two results are Theorems 4.2 and 4.3, respectively, in [10]. For the second theorem, also see the comments immediately following [10, Thm. 4.3].

For the convergence analysis, we follow closely the development in Osborn [14, Section 4(a)]. We omit the details, noting only those different from [14, Section 4(a)]. Taking f to be a given function in $L_2\left(\mathbb{B}^d\right)$, the element $u \in H_0^1(\Omega)$ for which

$$\mathcal{A}(u, w) = (f, w), \qquad \forall w \in H_0^1(\Omega),$$

can be written as $u = \mathcal{T} f$ with $\mathcal{T} : L_2\left(\mathbb{B}^d\right) \to H_0^1\left(\mathbb{B}^d\right) \cap H^2\left(\mathbb{B}^d\right)$ and bounded,

$$\|\mathcal{T} f\|_{H^2(\mathbb{B}^d)} \leq C \|f\|_{L_2(\mathbb{B}^d)}, \qquad f \in L_2\left(\mathbb{B}^d\right).$$

The operator is the 'Green's integral operator' for the associated Dirichlet problem. More generally, for $r \geq 0$, $\mathcal{T} : H^r\left(\mathbb{B}^d\right) \to H_0^1\left(\mathbb{B}^d\right) \cap H^{r+2}\left(\mathbb{B}^d\right)$,

$$\|\mathcal{T} f\|_{H^{r+2}(\mathbb{B}^d)} \leq C_r \; \|f\|_{H^r(\mathbb{B}^d)}, \qquad f \in H^r\left(\mathbb{B}^d\right).$$

In addition, $\mathcal{T}$ is a compact operator on $L_2\left(\mathbb{B}^d\right)$ into $H_0^1\left(\mathbb{B}^d\right)$, and more generally, it is compact from $H^r\left(\mathbb{B}^d\right)$ into $H_0^1\left(\mathbb{B}^d\right) \cap H^{r+1}\left(\mathbb{B}^d\right)$. With our assumptions, $\mathcal{T}$ is self-adjoint on $L_2\left(\mathbb{B}^d\right)$, although Osborn allows more general non-symmetric operators L. The same argument is applied to the numerical method (9) to obtain a solution $u_n = \mathcal{T}_n f$ with $\mathcal{T}_n$ having properties similar to $\mathcal{T}$ and also having finite rank with range in $\mathcal{X}_n$.

The major assumption of Osborn is that his finite element method satisfies an approximation inequality (see [14, (4.7)]), and the above theorems of Li and Xu are the corresponding statements for our numerical method. The argument in [14, Section 4(a)] then shows

$$\|\mathcal{T} - \mathcal{T}_n\|_{L_2 \to L_2} \leq \frac{c}{n^2}. \tag{25}$$

Our variational problems (8) and (9) can now be reformulated as

$$u = \mathcal{T}\mathcal{F}(u), \tag{26}$$

$$u_n = \mathcal{T}_n \mathcal{F}(u_n), \tag{27}$$

and we regard these as equations on some subset of $L_2\left(\mathbb{B}^d\right)$, dependent on the form of the function f defining $\mathcal{F}$. The operator $\mathcal{F}$ of (7) is sometimes called the Nemytskii operator; see [12, Chap. 1, Section 2] for its properties.

It is necessary to assume that $\mathcal{F}$ is defined and continuous over some open subset $D \subseteq L_2\left(\mathbb{B}^d\right)$:

$$\begin{aligned} v \in D &\implies f(\cdot, v) \in L_2\left(\mathbb{B}^d\right), \\ v_n \to v \text{ in} L_2\left(\mathbb{B}^d\right) &\implies f(\cdot, v_n) \to f(\cdot, v) \text{ in} L_2\left(\mathbb{B}^d\right). \end{aligned} \tag{28}$$

These are somewhat restrictive. As an example in one variable, if $b(\cdot, v) = v^2$ and if $v \in L^2(0, 1)$ then $b(\cdot, v)$ may not belong to $L^2(0, 1)$. The function $v(s) \equiv 1/\sqrt[3]{s}$ is in $L^2(0, 1)$, whereas $v(s)^2 = 1/\sqrt[3]{s^2}$ does not belong to $L^2(0, 1)$. An analysis of when (28) is true can be based on [13]. Generally, if $f(\cdot, v)$ is bounded by a linear function of v, then (28) is true. Experimentally, the spectral method (9) works well for cases with $f(\cdot, v)$ increasing at greater than a linear rate in v.

The operators $\mathcal{T}$ and $\mathcal{T}_n$ are linear, and the Nemytskii operator $\mathcal{F}$ provides the nonlinearity. The reformulation (26)–(27) can be used to give an error analysis of the spectral method (9). The mapping $\mathcal{T}\mathcal{F}$ is a compact nonlinear operator on an open domain D of a Banach space $\mathcal{X}$, in this case $L_2\left(\mathbb{B}^d\right)$. Let $V \subseteq D$ be an open set containing an isolated fixed point solution u^* of (26). We can define the index of u^* (or more properly, the rotation of the vector field $v - \mathcal{T}\mathcal{F}(v)$ as v varies over the boundary of V); see [12, part II].

More generally, let $\mathcal{K}$ be a completely continuous operator, and let it have an isolated fixed point u^* of nonzero index. This fixed point is stable in the sense that small compact perturbations of $\mathcal{K}$, say $\widetilde{\mathcal{K}}$, lead to one or more fixed points for $\widetilde{\mathcal{K}}$ with those fixed points all close to u^*. For an overview of the concepts of *index* and *rotation*, see [1, Properties P1-P5, pp. 801-802]. Property P4 gives a way of computing the index of u^*, and Property P5 gives further intuition as to the stability implications of a fixed point having a nonzero index.

Theorem 4 *Assume the problem (8) with $\Omega = \mathbb{B}^d$ has a solution u^* that is unique within some open neighborhood V of u^*; further assume that u^* has nonzero index. Then for all sufficiently large n, (9) has one or more solutions u_n within V, and all such u_n converge to u^* as $n \to \infty$.*

Proof This is an application of the methods of [12, Chap. 3, Sec. 3] or [1, Thm. 3]. A sufficient requirement is the norm convergence of $\mathcal{T}_n$ to $\mathcal{T}$, given in (25); [1, Thm. 3] uses a weaker form of (25). □

The most standard case of a nonzero index involves a consideration of the Frechet derivative of $\mathcal{F}$; see [4, §5.3]. In particular, the linear operator$\mathcal{F}'(v)$ is given by

$$\left(\mathcal{F}'(v)\, w\right)(x) = \left.\frac{\partial f(x,z)}{\partial z}\right|_{z=v(x)} \times w(x)$$

Theorem 5 *Assume the problem (8) with $\Omega = \mathbb{B}^d$ has a solution u^* that is unique within some open neighborhood V of u^*; and further assume that $I - \mathcal{T}\mathcal{F}'(u^*)$ is invertible over $L_2\left(\mathbb{B}^d\right)$. Then u^* has a nonzero index. Moreover, for all sufficiently large n, there is a unique solution u_n^* to (27) within V, and u_n^* converges to u^* with*

$$\begin{aligned}\left\|u^* - u_n^*\right\|_{L_2(\mathbb{B}^d)} &\le c \left\|(\mathcal{T} - \mathcal{T}_n)\, \mathcal{F}\left(u^*\right)\right\|_{L_2(\mathbb{B}^d)} \\ &\le \frac{c}{n^2} \left\|\mathcal{F}\left(u^*\right)\right\|_{L_2(\mathbb{B}^d)} . \end{aligned} \tag{29}$$

Proof Again, this is an immediate application of results in [12, Chap. 3, Sec. 3] or [1, Thm. 4]. □

Remark To give some intuition to our assumption that $I - \mathcal{T}\mathcal{F}'(u^*)$ is invertible, consider a rootfinding problem for a real-valued function $f(x)$ with $x \in \mathbb{R}$, letting α denote the root being sought. Then our invertibility assumption is the analogue of assuming $f'(\alpha) \neq 0$.

To improve upon this last result (29), we need to bound $\|(\mathcal{T} - \mathcal{T}_n)\, g\|_{L_2(\mathbb{B}^d)}$ when $g \in H^r\left(\mathbb{B}^d\right)$ for some $r \ge 1$. Adapting the proof of [14, (4.9)] to our polynomial approximations and using Theorem 3,

$$\|(\mathcal{T} - \mathcal{T}_n)\, g\|_{H^1(\mathbb{B}^d)} \le \frac{c}{n^{r+1}} \|g\|_{H^r(\mathbb{B}^d)}.$$

Using the conservative bound

$$\|v\|_{L_2(\mathbb{B}^d)} \le \|v\|_{H^1(\mathbb{B}^d)} ,$$

we have

$$\|(\mathcal{T} - \mathcal{T}_n)\, g\|_{L_2(\mathbb{B}^d)} \le \frac{c}{n^{r+1}} \|g\|_{H^r(\mathbb{B}^d)}. \tag{30}$$

Corollary 6 *For some $r \geq 0$, assume $\mathcal{F}(u^*) \in H^r\left(\mathbb{B}^d\right)$. Then*

$$\left\|u^* - u_n^*\right\|_{L_2(\mathbb{B}^d)} \leq \mathcal{O}\left(n^{-(r+1)}\right)\left\|\mathcal{F}\left(u^*\right)\right\|_{H^r(\mathbb{B}^d)}. \tag{31}$$

We conjecture that this bound and (30) can be improved to $\mathcal{O}\left(n^{-(r+2)}\right)$. For the case $r = 0$, an improved result is given by (29).

A nonhomogeneous boundary condition Consider replacing the homogeneous boundary condition (2) with the nonhomogeneous condition

$$u(s) = g(s), \qquad s \in \partial\Omega,$$

in which g is a continuously differentiable function over $\partial\Omega$. One possible approach to solving the Dirichlet problem with this nonzero boundary condition is to begin by calculating a differentiable extension of g, call it $G : \overline{\Omega} \to \mathbb{R}$, with

$$\begin{aligned} G &\in C^2\left(\overline{\Omega}\right), \\ G(s) &= g(s), \qquad s \in \partial\Omega. \end{aligned}$$

With such a function G, introduce $v = u - G$ where u satisfies (1)–(2). Then v satisfies the equation

$$Lv(s) = f(s, v(s) + G(s)) - LG(s), \qquad s \in \Omega, \tag{32}$$

$$v(s) = 0, \qquad s \in \partial\Omega. \tag{33}$$

This problem is in the format of (1)–(2).

Sometimes finding an extension G is straightforward; for example, $g \equiv 1$ over $\partial\Omega$ has the obvious extension $G(s) \equiv 1$. Often, however, we must compute an extension. We begin by first obtaining an extension G using a method from [5], and then we approximate it with a polynomial of some reasonably low degree. For example, see the construction of least squares approximants in [3].

4 Implementation

We consider how to set up the nonlinear system of (9)–(11) and how to solve it. Because we intend to apply the method to problems defined initially over a region Ω other than $\mathbb{B}^d$, we re-write (9)–(11) for this situation. The transformed equation we are considering is the equation (18). We look for a solution

$$\widetilde{u}_n(x) = \sum_{\ell=1}^{N_n} \alpha_\ell \psi_\ell(x),$$

and $u_n(s)$ is to be the equivalent solution considered over Ω: $\widetilde{u}_n(x) \equiv u_n(\Phi(x))$, $x \in \mathbb{B}^d$. The coefficients $\{\alpha_\ell | \ell = 1, 2, \ldots, N_n\}$ are the solutions of

$$\sum_{k=1}^{N_n} \alpha_k \int_{\mathbb{B}^d} \left[\sum_{i,j=1}^{d} \det J(x)\, \widetilde{a}_{i,j}(x) \frac{\partial \psi_k(x)}{\partial x_j} \frac{\partial \psi_\ell(x)}{\partial x_i} + \widetilde{\gamma}(x) \psi_k(x) \psi_\ell(x) \right. \\ \left. + \widetilde{\gamma}(x) \psi_k(x) \psi_\ell(x) \right] dx \\ = \int_{\mathbb{B}^d} \widetilde{f}\left(x, \sum_{k=1}^{N_n} \alpha_k \psi_k(x)\right) \psi_\ell(x)\, dx, \qquad \ell = 1, \ldots, N_n. \tag{34}$$

For the definitions of $\widetilde{\gamma}$, $\widetilde{f}$, and $\widetilde{A}(x) \equiv \left[\widetilde{a}_{i,j}(x)\right]_{i,j=1}^{d}$, recall (19)–(21).

When solving the nonlinear system (34), it is necessary to have an initial guess $\widetilde{u}_n^{(0)}(x) = \sum_{\ell=1}^{N_n} \alpha_\ell^{(0)} \psi_\ell(x)$. In our examples, we begin with a very small value for n (say $n = 1$), use $\widetilde{u}_n^{(0)} = 0$, and then solve (34) by some iterative method. Then increase n, using as an initial guess the final solution obtained with a preceding n. This has worked well in our computations, allowing us to work our way to the solution of (34) for much larger values of n. For the iterative solver, we have used the MATLAB program `fsolve`, but will work in the future on improving it.

4.1 Planar problems

The dimension of Π_n^2 is

$$N_n = \frac{1}{2}(n+1)(n+2).$$

For notation, we replace x with (x, y). We create a basis for $\mathcal{X}_n$ by first choosing an orthonormal basis for Π_n^2, say $\left\{\varphi_{m,k} | k = 0, 1, \ldots, m;\ m = 0, 1, \ldots, n\right\}$. Then define

$$\psi_{m,k}(x, y) = \left(1 - x^2 - y^2\right) \varphi_{m,k}(x, y). \tag{35}$$

How do we choose the orthonormal basis $\{\varphi_\ell(x, y)\}_{\ell=1}^{N}$ for Π_n^2? Unlike the situation for the single variable case, there are many possible orthonormal bases over $\mathbb{B}^2$, the unit disk in $\mathbb{R}^2$. We have chosen one that is convenient for our computations. These are the "ridge polynomials" introduced by Logan and Shepp [11] for solving an image reconstruction problem. A choice that is more efficient in calculational costs is given in [3], but we continue to use the ridge polynomials because we are re-using and modifying computer code written previously for use in [2, 3, 6], and [7].

We summarize here the results needed for our work. For general, $d \geq 2$, let

$$\mathcal{V}_n = \left\{ P \in \Pi_n^d \mid (P, Q) = 0 \quad \forall Q \in \Pi_{n-1}^d \right\},$$

the polynomials of degree n that are orthogonal to all elements of Π_{n-1}^d. Then

$$\Pi_n^d = \mathcal{V}_0 \oplus \mathcal{V}_1 \oplus \cdots \oplus \mathcal{V}_n \tag{36}$$

is a decomposition of Π_n^d into orthonormal subspaces. It is standard to construct orthonormal bases of each $\mathcal{V}_n$ and to then combine them to form an orthonormal basis of Π_n^d using this decomposition.

For $d = 2$, $\mathcal{V}_n$ has dimension $n + 1$, $n \geq 0$. As an orthonormal basis of $\mathcal{V}_n$, we use

$$\varphi_{n,k}(x, y) = \frac{1}{\sqrt{\pi}} U_n\left(x \cos(kh) + y \sin(kh)\right), \quad (x, y) \in D, \quad h = \frac{\pi}{n+1} \tag{37}$$

for $k = 0, 1, \ldots, n$. The function U_n is the Chebyshev polynomial of the second kind of degree n:

$$U_n(t) = \frac{\sin(n+1)\theta}{\sin\theta}, \qquad t = \cos\theta, \quad -1 \leq t \leq 1, \quad n = 0, 1, \ldots$$

The family $\left\{\varphi_{n,k}\right\}_{k=0}^{n}$ is an orthonormal basis of $\mathcal{V}_n$.

As a basis of Π_n^2, we order $\left\{\varphi_{m,k}\right\}$ lexicographically based on the ordering in (37) and (36):

$$\{\varphi_\ell\}_{\ell=1}^{N_n} = \left\{\varphi_{0,0},\ \varphi_{1,0},\ \varphi_{1,1},\ \varphi_{2,0},\ \ldots,\ \varphi_{n,0},\ \ldots,\varphi_{n,n}\right\}.$$

From (35), the family $\left\{\psi_{m,k}\right\}$ is ordered the same.

To calculate the first-order partial derivatives of $\psi_{n,k}(x, y)$, we need $U_n^{'}(t)$. The values of $U_n^{'}(t)$ and $U_n'(t)$ are evaluated using the standard triple recursion relations

$$\begin{aligned} U_{n+1}(t) &= 2tU_n(t) - U_{n-1}(t), \\ U_{n+1}^{'}(t) &= 2U_n(t) + 2tU_n^{'}(t) - U_{n-1}^{'}(t). \end{aligned}$$

For the numerical approximation of the integrals in (34), which are over $\mathbb{B}^2$, the unit disk, we use the formula

$$\int_{\mathbb{B}^2} g(x, y)\, dx\, dy \approx \frac{2\pi}{2q+1} \sum_{l=0}^{q} \sum_{m=0}^{2q} \widehat{g}\left(r_l, \frac{2\pi\, m}{2q+1}\right) \omega_l r_l \tag{38}$$

with $\widehat{g}(r, \theta) \equiv g(r\cos\theta, r\sin\theta)$. Here, the numbers r_l and ω_l are the nodes and weights of the $(q + 1)$-point Gauss-Legendre quadrature formula on [0, 1]. Note that

$$\int_0^1 p(x)dx = \sum_{l=0}^{q} p(r_l)\omega_l,$$

for all single-variable polynomials $p(x)$ with $\deg(p) \leq 2q + 1$. The formula (38) uses the trapezoidal rule with $2q + 1$ subdivisions for the integration over $\mathbb{B}^2$ in the azimuthal variable. This quadrature (38) is exact for all polynomials $g \in \Pi_{2q}^2$.

4.2 The three-dimensional case

We change our notation, replacing $x \in \mathbb{B}^3$ with (x, y, z). In $\mathbb{R}^3$, the dimension of Π_n^3 is

$$N_n = \binom{n+3}{3} = \frac{1}{6}(n+1)(n+2)(n+3).$$

Here, we choose orthonormal polynomials on the unit ball as described in [8],

$$\varphi_{n,j,k}(x) = \frac{1}{h_{n,j,k}} C_{n-j-k}^{j+k+\frac{3}{2}}(x)(1-x^2)^{\frac{j}{2}} \times$$
$$C_j^{k+1}\left(\frac{y}{\sqrt{1-x^2}}\right)(1-x^2-y^2)^{k/2} C_k^{\frac{1}{2}}\left(\frac{z}{\sqrt{1-x^2-y^2}}\right), \quad (39)$$
$$j, k = 0, \ldots, n, \quad j+k \le n, \quad n \in \mathbb{N}.$$

The function $\varphi_{n,j,k}(x)$ is a polynomial of degree n, $h_{n,j,k}$ is a normalization constant, and the functions C_i^λ are the Gegenbauer polynomials. The orthonormal base $\{\varphi_{n,j,k}\}_{n,j,k}$ and its properties can be found in [8, Chapter 2].

We can order the basis lexicographically. To calculate these polynomials, we use a three-term recursion whose coefficients are given in [3].

For the numerical approximation of the integrals in (34), we use a quadrature formula for the unit ball $\mathbb{B}^3$,

$$\int_{\mathbb{B}^3} g(x)\,dx = \int_0^1 \int_0^{2\pi} \int_0^{\pi} \widehat{g}(r,\theta,\phi)\, r^2 \sin(\phi)\, d\phi\, d\theta\, dr \approx Q_q[g],$$
$$Q_q[g] := \sum_{i=1}^{2q} \sum_{j=1}^{q} \sum_{k=1}^{q} \frac{\pi}{q}\, \omega_j\, \nu_k \widehat{g}\left(\frac{\zeta_k+1}{2}, \frac{\pi\, i}{2q}, \arccos(\xi_j)\right).$$

Here, $\widehat{g}(r,\theta,\phi) = g(x)$ is the representation of g in spherical coordinates. For the θ integration, we use the trapezoidal rule, because the function is $2\pi-$periodic in θ. For the r direction, we use the transformation

$$\begin{aligned}\int_0^1 r^2 v(r)\,dr &= \int_{-1}^1 \left(\frac{t+1}{2}\right)^2 v\left(\frac{t+1}{2}\right)\frac{dt}{2}\\ &= \frac{1}{8}\int_{-1}^1 (t+1)^2 v\left(\frac{t+1}{2}\right)\,dt\\ &\approx \sum_{k=1}^{q} \underbrace{\frac{1}{8}\nu_k'}_{=:\nu_k} v\left(\frac{\zeta_k+1}{2}\right),\end{aligned}$$

where the ν_k' and ζ_k are the weights and the nodes of the Gauss quadrature with q nodes on $[-1, 1]$ with respect to the inner product

$$(v, w) = \int_{-1}^1 (1+t)^2 v(t) w(t)\,dt.$$

The weights and nodes also depend on q but we omit this index. For the ϕ direction, we use the transformation

$$\int_0^{\pi} \sin(\phi) v(\phi)\,d\phi = \int_{-1}^1 v(\arccos(\phi))\,d\phi \approx \sum_{j=1}^{q} \omega_j v(\arccos(\xi_j)),$$

where the ω_j and ξ_j are the nodes and weights for the Gauss–Legendre quadrature on $[-1, 1]$. For more information on this quadrature rule on the unit ball in $\mathbb{R}^3$, see [15].

Finally, we need the gradient to approximate the integral in (34). To do this, one can modify the three-term recursion in [3] to calculate the partial derivatives of $\varphi_{n,j,k}(x)$.

5 Numerical examples

We begin with a planar example. Consider the problem

$$\begin{aligned} -\Delta u\left(s,t\right) &= f\left(s,t,u\left(s,t\right)\right), \qquad && \left(s,t\right)\in\Omega, \\ u\left(s,t\right) &= 0, && \left(s,t\right)\in\partial\Omega. \end{aligned} \tag{40}$$

Note the change in notation, from $s\in\mathbb{R}^2$ to $(s,t)\in\mathbb{R}^2$.

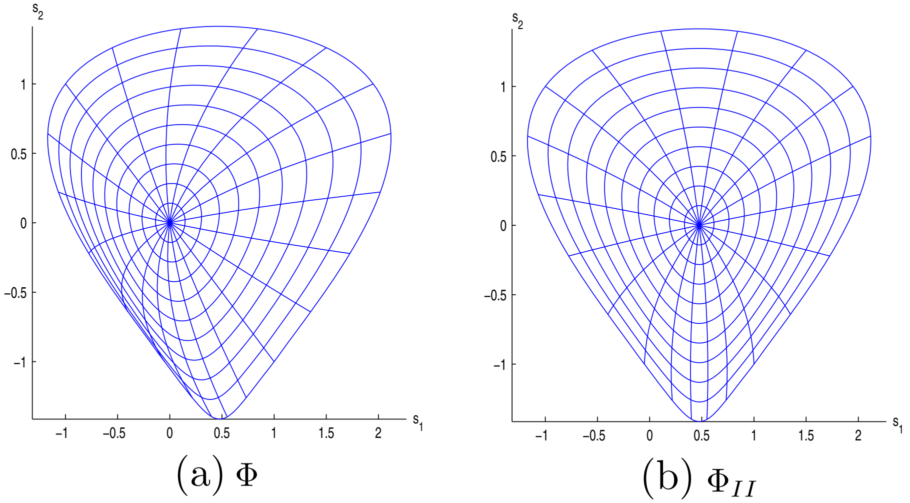
As an illustrative region Ω, we use the mapping $\Phi:\overline{\mathbb{B}}^2\rightarrow\overline{\Omega}$, $(s,t)=\Phi\left(x,y\right)$,

$$\begin{aligned} s &= x-y+ax^2, \\ t &= x+y, \end{aligned} \tag{41}$$

with $0<a<1$. It can be shown that Φ is a 1-1 mapping from the unit disk $\overline{\mathbb{B}}^2$. In particular, the inverse mapping $\Psi:\overline{\Omega}\rightarrow\overline{\mathbb{B}}^2$ is given by

$$\begin{aligned} x &= \frac{1}{a}\left[-1+\sqrt{1+a\left(s+t\right)}\right] \\ y &= \frac{1}{a}\left[at-\left(-1+\sqrt{1+a\left(s+t\right)}\right)\right] \end{aligned} \tag{42}$$

In Fig. 1a, the mapping for $a=0.95$ is illustrated by giving the images in $\overline{\Omega}$ of the circles $r=j/10$, $j=1,\ldots,10$ and the radial lines $\theta=j\pi/10$, $j=1,\ldots,20$. An alternative polynomial mapping Φ_{II} of degree 2 for this region is computed using the integration/interpolation method of [5, Section 3]; and $\Phi_{II}=\Phi$ on the boundary. $\partial\Omega$



(a) Φ

(b) Φ_{II}

Fig. 1 Illustrations of mappings on $\mathbb{B}^2$ for the region Ω given by (41)

as defined by (41). It is illustrated in Fig. 1b. This boundary mapping Φ_{II} results in better error characteristics for our spectral method as compared to the transformation Φ.

As discussed earlier, we solve the nonlinear system (34) for a lower value of the degree n, usually with an initial guess associated with $u_n^{(0)} = 0$. As we increase n, we use the approximate solution from a preceding n to generate an initial guess for the new value of n. We use the MATLAB program `fsolve` to solve the nonlinear system. In the future, we plan to look at other numerical methods that take advantage of the special structure of (34). To estimate the error, we use as a true solution a numerical solution associated with a larger value of n.

For a particular case, consider

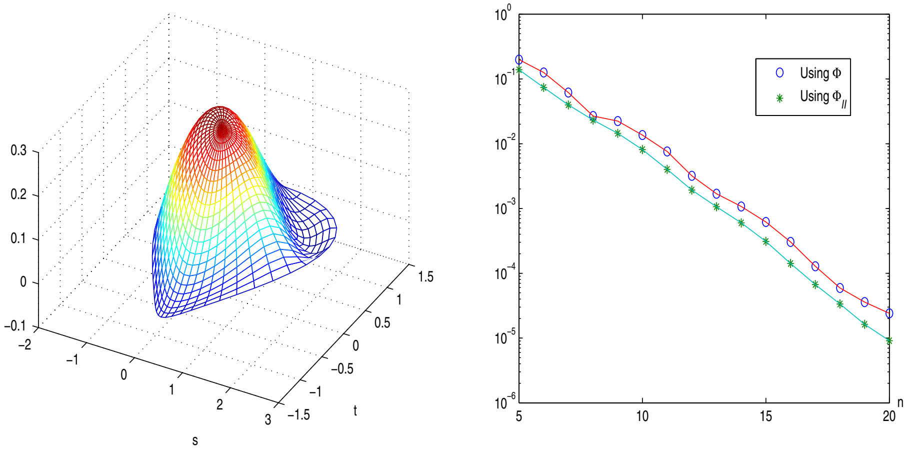
$$f\,(s,t,z) = \frac{\cos\,(\pi\; st)}{1+z^2}. \tag{43}$$

A graph of the solution is shown in Fig. 2, along with numerical results for $n = 5, 6, \ldots, 20$, with the solution u_{25} taken as the true solution. We use both the mapping Φ of (41) and the mapping Φ_{II}. Using either of the mappings, Φ or Φ_{II}, the graphs indicate an exponential rate of convergence for the mappings $\{u_n\}$. The mapping Φ_{II} is better behaved, as can be seen by visually comparing the distortion in the graphs of Fig. 1. This is the probable reason for the improved convergence of the spectral method when using Φ_{II} in comparison to Φ.

As a second planar example, we consider the stationary Fisher equation where the function f in (40) is given by

$$f(s,t,u) = 100u(1-u), \qquad (s,t) \in \Omega.$$

Fisher's equation is used to model the spreading of biological populations, and from f, we see that $u = 0$ and $u = 1$ are stationary points for the time-dependent equation



The solution u

The maximum error

Fig. 2 The solution u to (40) with right side (43) and its error

on an unbounded domain; see [9, Chap. 17]. The original Fisher equation does not contain the term 100, but for small domains the Fisher equation might have no nontrivial solution and the factor 100 corresponds to a scaling by a factor 10 to guarantee the existence of a nontrivial solution on the domain Ω. The domain Ω is the interior of the curve

$$\varphi(t) = (3 + \cos(t) + 2\sin(t))\,(\cos t, \sin t)\,. \tag{44}$$

We studied this domain in earlier papers (see [5]) where we called this domain a ‘Limacon domain’. In the article [5], we also describe how we use equation (44) to create a domain mapping $\Phi : \overline{\mathbb{B}}^2 \to \overline{\Omega}$ by two dimensional interpolation. Similar to the previous example, we calculate the numerical solutions u_n for $n = 1, \ldots, 40$, where we use the coefficients of u_{n-1} as a starting value $u_n^{(0)}$ for $n = 2, \ldots, 40$, and for $u_1^{(0)}$, we use coefficients which are non zero (all equal to 10), so the iteration of `fsolve` does not converge to the trivial solution. As a reference solution, we calculated u_{45}; see Fig. 3.

The shape of the solution is very much like we expect it, the function is close to 1 inside the domain Ω and drops off very steeply to the boundary value 0. By looking at the reference solution in Fig. 3, we also see that the function will be harder to approximate by polynomials than the function in the previous example, because of the sharp drop off. This becomes clear when we look at the convergence, also shown in Fig. 3. The final error is in the range of 10^{-3}–10^{-4} with a polynomial degree of 40, so the error is in the same range as in the previous example where we only used polynomials up to degree 20 for the approximation. Still the graph suggests that the convergence is exponential as predicted by (31) for the L^2 norm.

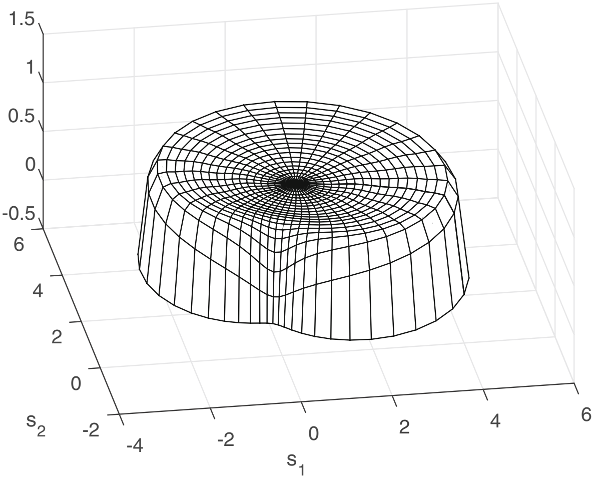


Fig. 3 The reference solution and maximum error for Fisher’s equation

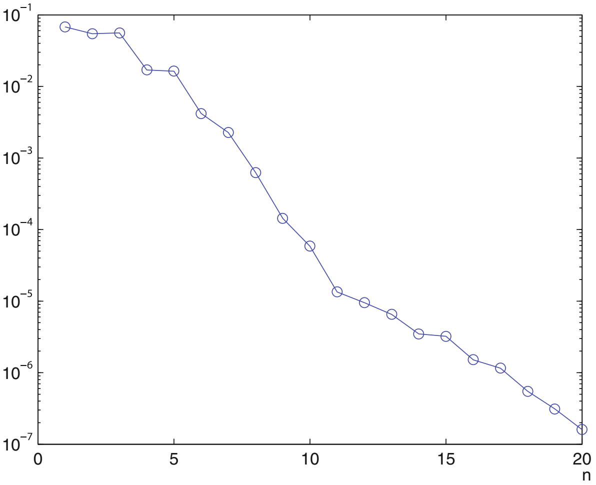


Fig. 4 For the problem (46), the convergence of the errors $\|u - u_n\|_\infty$

A three-dimensional example In the following, we present a three-dimensional example. We use the mapping $\Phi : \overline{\mathbb{B}}^3 \to \overline{\Omega}$, $(s, t, v) = \Phi(x, y, z)$, defined by

$$\begin{aligned} s &= x - y + ax^2, \\ t &= x + y, \\ v &= 2z + bz^2, \end{aligned} \tag{45}$$

where $a = b = 0.5$. We have used this mapping in a previous article, see [2], where one finds plots of the surface $\partial\Omega$. On Ω, we solve

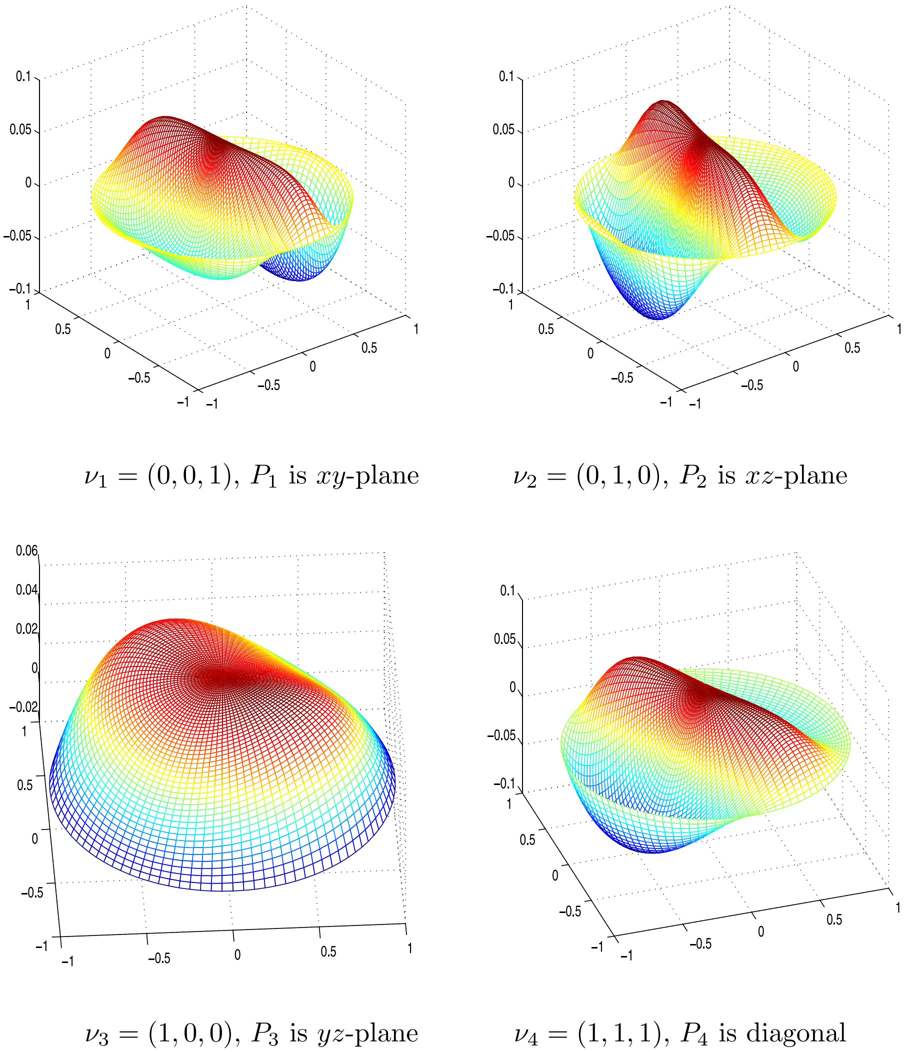
$$\begin{aligned} -\Delta u(s, t, v) &= f(s, t, v, u(s, t, v)), \qquad (s, t, v) \in \Omega \\ u(s, t, v) &= 0, \qquad\qquad (s, t, v) \in \partial\Omega \end{aligned} \tag{46}$$

where f is defined by

$$f(s, t, v, u) = \frac{\cos(6x + y + z)}{1 + u^2}, \qquad (s, t, v) \in \Omega.$$

We calculated approximate solutions $u_1, \ldots, u_{20}$ and used u_{25} as a reference solution. In Fig. 4, we see the convergence in the maximum norm on a grid in $\overline{\mathbb{B}}^3$. As in our previous examples, the graph suggests that we have exponential convergence.

In our final Fig. 5, we show the graph of the reference solution u_{25} on $\overline{\mathbb{B}}^3 \cap P_\nu$ where P_ν is a plane in $\mathbb{R}^3$ normal to the vector ν. We have used several normal vectors $\nu_1 = (0, 0, 1)^T$, so P_{ν_1} is the xy–plane, $\nu_2 = (0, 0, 1)^T$, so P_{ν_2} is the xz–plane, $\nu_3 = (1, 0, 0)^T$, so P_{ν_3} is the yz–plane, and $\nu_4 = (1, 1, 1)^T$, so P_{ν_4} is a diagonal plane. Figure 5 shows that the solution reflects the periodic character of the nonlinearity f. In the yz–plane, the oscillation of f is much slower which is also visible in the plot along the yz–plane.



$\nu_1 = (0, 0, 1)$, P_1 is xy-plane

$\nu_2 = (0, 1, 0)$, P_2 is xz-plane

$\nu_3 = (1, 0, 0)$, P_3 is yz-plane

$\nu_4 = (1, 1, 1)$, P_4 is diagonal

Fig. 5 The solution $\widetilde{u}\,(x, y, z)$ over $P \cap \mathbb{B}^3$ with P a plane passing through the origin and orthogonal to ν

6 A Neumann boundary value problem

Consider the boundary value problem

$$-\,\Delta u\,(s) + \gamma\,(s)\,u\,(s) = f\,(s, u(s))\,, \qquad s \in \Omega, \tag{47}$$

$$\frac{\partial u\,(s)}{\partial n_s} = 0, \qquad s \in \partial\Omega, \tag{48}$$

with n_s the exterior unit normal to $\partial\Omega$ at the boundary point s. Later, we discuss an extension to a nonzero normal derivative over $\partial\Omega$. A necessary condition for the unknown function u^* to be a solution of (47)–(48) is that it satisfy

$$\int_\Omega f\left(s, u^*(s)\right)\, ds = \int_\Omega \gamma(s)\, u^*(s)\, ds. \tag{49}$$

With our assumption that (47)–(48) has a locally unique solution u^*, (49) is satisfied.

Proceed in analogy with the earlier treatment of the Dirichlet problem. Use integration by parts to show that for arbitrary functions $u \in H^2(\Omega)$, $v \in H^1(\Omega)$,

$$\begin{aligned}\int_\Omega & v(s)\left[-\Delta u(s) + \gamma(s)u\right] ds \\ &= \int_\Omega \left[\nabla u(s)\cdot\nabla v(s) + \gamma(s)u(s)v(s)\right] ds - \int_{\partial\Omega} v(s)\,\frac{\partial u(s)}{\partial n_s}\, ds.\end{aligned} \tag{50}$$

Introduce the bilinear functional

$$\mathcal{A}(v_1, v_2) = \int_\Omega \left[\nabla v_1(s)\cdot\nabla v_2(s) + \gamma(s)v_1(s)v_2(s)\right] ds.$$

The variational form of the Neumann problem (47)–(48) is as follows: find $u \in H^1(\Omega)$ such that

$$\mathcal{A}(u, v) = (\mathcal{F}(u), v), \qquad \forall v \in H^1(\Omega) \tag{51}$$

with, as before, the operator $\mathcal{F}$ defined by

$$(\mathcal{F}(u))(s) = f(s, u(s)).$$

The theory for (51) is essentially the same as for the Dirichlet problem in its reformulation (8).

Because of changes that take place in the normal derivative under the transformation $s = \Phi(x)$, we modify the construction of the numerical method. In the actual implementation, however, it will mirror that for the Dirichlet problem. For the approximating space, let

$$\mathcal{X}_n = \left\{ q \mid q \circ \Phi = p \text{ for some} p \in \Pi_n^d \right\}.$$

For the numerical method, we seek $u_n^* \in \mathcal{X}_n$ for which

$$\mathcal{A}\left(u_n^*, v\right) = \left(\mathcal{F}\left(u_n^*\right), v\right), \qquad \forall v \in \mathcal{X}_n. \tag{52}$$

A similar approach was used in [6] for the linear Neumann problem.

To carry out a convergence analysis for (52), it is necessary to compare convergence of approximants in $\mathcal{X}_n$ to that of approximants from Π_n^d. For simplicity in notation, we assume $\Phi \in C^\infty\left(\overline{\mathbb{B}}^d\right)$. Begin by referring to Lemma 1 and its discussion in Section 2.1, linking differentiability in $H^m(\Omega)$ and $H^m\left(\mathbb{B}^d\right)$. In particular, for $m \geq 0$,

$$c_{1,m}\,\|v\|_{H^m(\Omega)} \leq \|\widetilde{v}\|_{H^m(\mathbb{B}^d)} \leq c_{2,m}\,\|v\|_{H^m(\Omega)}, \qquad v \in H^m(\Omega), \tag{53}$$

with $\widetilde{v} = v \circ \Phi$, with constants $c_{1,m}, c_{2,m} > 0$.

Also recall Theorem 2 concerning approximation of functions $\widetilde{v} \in H^r\left(\mathbb{B}^d\right)$ and link this to approximation of functions $v \in H^r(\Omega)$.

Lemma 7 *Let $\Phi \in C^\infty\left(\overline{\mathbb{B}}^d\right)$. Assume $v \in H^r(\Omega)$ for some $r \geq 2$. Then there exist a sequence $q_n \in \mathcal{X}_n$, $n \geq 1$, for which*

$$\|v - q_n\|_{H^1(\Omega)} \leq \varepsilon_{n,r} \|v\|_{H^r(\Omega)}, \qquad n \geq 1. \tag{54}$$

The sequence $\varepsilon_{n,r} = \mathcal{O}\left(n^{-r+1}\right)$ and is independent of v.

Proof Begin by applying Theorem 2 to the function $\widetilde{v}(x) = v(\Phi(x))$. Then there is a sequence of polynomials $p_n \in \Pi_n^d$ for which

$$\|\widetilde{v} - p_n\|_{H^1(\mathbb{B}^d)} \leq \varepsilon_{n,r} \|\widetilde{v}\|_{H^r(\mathbb{B}^d)}, \qquad n \geq 1.$$

Let $q_n = p_n \circ \Phi^{-1}$. The result then follows by applying (53). □

The theoretical convergence analysis now follows exactly that given earlier for the Dirichlet problem. Again, we use the construction from [14, Section 4(a)], but now use the integral operator $\mathcal{T}$ arising from the zero Neumann boundary condition. As with the Dirichlet problem, it is necessary to have $\mathcal{A}$ be strongly elliptic, and for that reason and without any loss of generality, assume

$$\min_{s \in \overline{\Omega}} \gamma(s) > 0.$$

The solution of (51) can be written as $u = \mathcal{T}\mathcal{F}(u)$ with $\mathcal{T} : L_2\left(\mathbb{B}^d\right) \to H^2\left(\mathbb{B}^d\right)$ and bounded. Use Theorem 2 in place of Theorem 3 for polynomial approximation error, as in the derivation of (29). Theorems 4 and 5, along with Corollary 6 are valid for the spectral method for the Neumann problem (47)–(48).

6.1 Implementation

As in Section 4, we look for a solution to (51) by looking for

$$u_n(s) = \sum_{\ell=1}^{N_n} \alpha_\ell \psi_\ell(s) \tag{55}$$

with $\{\psi_\ell \mid 1 \leq j \leq N_n\}$ a basis for $\mathcal{X}_n$. The system associated with (51) that is to be solved is

$$\begin{aligned} &\sum_{\ell=1}^{N_n} \alpha_\ell \int_\Omega \left[\sum_{i,j=1}^{d} a_{i,j}(s) \frac{\partial \psi_\ell(s)}{\partial s_i} \frac{\partial \psi_k(s)}{\partial s_j} + \gamma(s) \psi_\ell(s) \psi_k(s) \right] ds \\ &\quad = \int_\Omega f\left(s, \sum_{\ell=1}^{N_n} \alpha_\ell \psi_\ell(s)\right) \psi_k(s)\, ds, \qquad k = 1, \ldots, N_n. \end{aligned} \tag{56}$$

For such a basis $\{\psi_\ell\}$, we begin with an orthonormal basis for Π_n, say $\{\varphi_j \mid 1 \le j \le N_n\}$, and then define

$$\psi_\ell(s) = \varphi_\ell(x) \quad with = \Phi(x), \qquad 1 \le \ell \le N.$$

The function $\widetilde{u}_n(x) \equiv u_n(\Phi(x))$, $x \in \mathbb{B}^d$,is to be the equivalent solution considered over $\mathbb{B}^d$. Using the transformation of variables $s = \Phi(x)$ in the system (56), the coefficients $\{\alpha_\ell | \ell = 1, 2, \ldots, N_n\}$ are the solutions of

$$\begin{aligned}\sum_{k=1}^{N_n} \alpha_k \int_{\mathbb{B}^d} \left[\sum_{i,j=1}^{d} \widetilde{a}_{i,j}(x) \frac{\partial \varphi_k(x)}{\partial x_j} \frac{\partial \varphi_\ell(x)}{\partial x_i} + \gamma(\Phi(x)) \varphi_k(x) \varphi_\ell(x) \right] \det J(x)\, dx \\ = \int_{\mathbb{B}^d} f\left(x, \sum_{k=1}^{N_n} \alpha_k \varphi_k(x)\right) \varphi_\ell(x) \det J(x)\, dx, \qquad \ell = 1, \ldots, N_n.\end{aligned} \tag{57}$$

For the equation (47), the matrix $A(s)$ is the identity, and therefore from (21),

$$\widetilde{A}(x) = J(x)^{-1} J(x)^{-\mathrm{T}}.$$

The system (57) is much the same as (34) for the Dirichlet problem, differing only by the basis functions being used for the solution $\widetilde{u}_n$. We use the same numerical integration as before, and also the same orthonormal basis for Π_n^d.

6.2 Numerical example

Consider the problem

$$\begin{aligned} -\Delta u(s,t) + u(s,t) &= f(s,t,u(s,t)), \qquad && (s,t) \in \Omega, \\ \frac{\partial u(s)}{\partial n_s} &= 0, && (s,t) \in \partial\Omega, \end{aligned} \tag{58}$$

with Ω the elliptical region

$$\left(\frac{s}{a}\right)^2 + \left(\frac{t}{b}\right)^2 \le 1.$$

The mapping of $\mathbb{B}^2$ onto Ω is simply

$$\Phi(x,y) = (ax, by), \qquad (x,y) \in \overline{\mathbb{B}}^2.$$

As before, note the change in notation, from $s \in \Omega$ to $(s,t) \in \Omega$, and from $x \in \mathbb{B}^2$ to $(x,y) \in \mathbb{B}^2$.

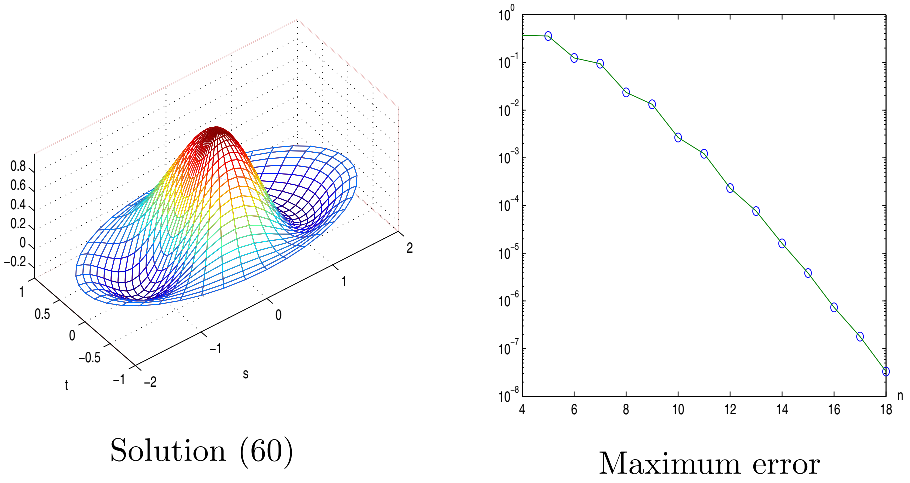
The right side f is given by

$$f(s,t,u) = -e^u + f_1(s,t) \tag{59}$$

with the function f_1 determined from the given true solution and the equation (58) to define $f(s,t,u)$. In our case,

$$u(s,t) = \left(1 - \left(\frac{s}{a}\right)^2 - \left(\frac{t}{b}\right)^2\right)^2 \cos\left(2s + t^2\right). \tag{60}$$

Easily this has a normal derivative of zero over the boundary of Ω.



Solution (60)

Maximum error

Fig. 6 The solution u to (58) with right side (59) and true solution (60)

The nonlinear system (57) was solved using `fsolve` from MATLAB, as earlier in Section 5. Our region Ω uses $(a, b) = (2, 1)$. Figure 6 contains the approximate solution for $n = 18$ and also shows the maximum error over $\overline{\Omega}$. Again, the convergence appears to be exponential.

6.3 Handling a nonzero Neumann condition

Consider the problem

$$-\Delta u(s) + \gamma(s)\, u(s) = f(s, u(s)), \qquad s \in \Omega, \tag{61}$$

$$\frac{\partial u(s)}{\partial n_s} = g(s), \qquad s \in \partial\Omega \tag{62}$$

with a nonzero Neumann boundary condition. Let $u^*(s)$ denote the solution we are seeking. A necessary condition for solvability of (61)–(62) is that

$$\int_\Omega f\left(s, u^*(s)\right) ds = \int_\Omega \gamma(s)\, u^*(s)\, ds - \int_{\partial\Omega} g(s)\, ds. \tag{63}$$

There are at least two approaches to extending our spectral method to solve this problem.

First, consider the problem

$$-\Delta v(s) = c_0, \qquad s \in \Omega, \tag{64}$$

$$\frac{\partial v(s)}{\partial n_s} = g(s), \qquad s \in \partial\Omega, \tag{65}$$

with c_0 a constant. From (63), solvability of (64)–(65) requires

$$\int_\Omega c_0\, ds = -\int_{\partial\Omega} g(s)\, ds \tag{66}$$

to be satisfied. To achieve this, choose

$$c_0 = \frac{-1}{\mathrm{Vol}\,(\Omega)} \int_{\partial\Omega} g\,(s)\;ds.$$

A solution $v^*\,(s)$ exists, although it is not unique. The solution of (64)–(65) can be approximated using the method given in [6]. Then introduce

$$w = u - v^*.$$

Substituting into (61)–(62), the new unknown function w^* satisfies

$$-\Delta w\,(s) + \gamma\,(s)\,w\,(s) = f\left(s, w(s) + v^*\,(s)\right) - \gamma\,(s)\,v^*\,(s) - c_0, \qquad s \in \Omega, \quad (67)$$

$$\frac{\partial w\,(s)}{\partial n_s} = 0, \qquad s \in \partial\Omega. \quad (68)$$

The methods of this section can be used to approximate w^*, and then use $u^* = w^* + v^*$.

A second approach is to use (50) to reformulate (61)–(62) as the problem of finding $u = u^*$ for which

$$\mathcal{A}\,(u, v) = (\mathcal{F}\,(u)\,, v) + \ell\,(v)\,, \qquad \forall v \in H^1\,(\Omega) \quad (69)$$

with

$$\ell\,(v) = \int_{\partial\Omega} v\,(s)\,g\,(s)\;ds.$$

Thus, we seek

$$u_n\,(s) = \sum_{\ell=1}^{N_n} \alpha_\ell \psi_\ell\,(s)$$

for which

$$\mathcal{A}\,(u_n, v) = (\mathcal{F}\,(u)\,, v) + \ell\,(v)\,, \qquad \forall v \in \mathcal{X}_n. \quad (70)$$

The first approach, that of (61)–(68), is usable, and the convergence analysis follows from combining this paper's analysis with that of [6]. Unfortunately, we do not have a convergence analysis for this second approach, that of (69)–(70), as the Green's function approach of this paper does not seem to extend to it.

References

1. Atkinson, K.: The numerical evaluation of fixed points for completely continuous operators. SIAM J. Num. Anal. **10**, 799–807 (1973)
2. Atkinson, K., Chien, D., Hansen, O.: A spectral method for elliptic equations The Dirichlet problem. Adv. Comput. Math. **33**, 169–189 (2010)
3. Atkinson, K., Chien, D., Hansen, O.: Evaluating polynomials over the unit disk and the unit ball. Numer. Algorithm. **67**, 691–711 (2014)
4. Atkinson, K., Han, W.: Theoretical numerical analysis: a functional analysis framework, 3rd edn. Springer-Verlag (2009)
5. Atkinson, K., Hansen, O.: Creating domain mappings. Electron. Trans. Numer. Anal. **39**, 202–230 (2012)
6. Atkinson, K., Hansen, O., Chien, D.: A spectral method for elliptic equations The Neumann problem. Adv. Comput. Math. **34**, 295–317 (2011)

7. Atkinson, K., Hansen, O., Chien, D.: A spectral method for parabolic differential equations. Numer. Algorithm. **63**, 213–237 (2013)
8. Dunkl, C., Xu, Y.: Orthogonal polynomials of several variables. Cambridge University Press (2001)
9. Kot, M.: Elements of mathematical ecology. Cambridge University Press (2001)
10. Li, H., Xu, Y.: Spectral approximation on the unit ball. SIAM J. Num Anal. **52**, 2647–2675 (2014)
11. Logan, B., Shepp, L.: Optimal reconstruction of a function from its projections. Duke Math. J. **42**, 645–659 (1975)
12. Krasnoseľskii, M.: Topological methods in the theory of nonlinear integral equations. Pergamon Press (1964)
13. Marcus, M., Mizel, V.: Absolute continuity on tracks and mappings of Sobolev spaces. Arch. Rat. Mech. Anal. **45**, 294–320 (1972)
14. Osborn, J.: Spectral approximation for compact operators. Math. Comput. **29**, 712–725 (1975)
15. Stroud, A.: Approximate calculation of multiple integrals. Prentice-Hall, Inc. (1971)
16. Zeidler, E.: Nonlinear functional analysis and its applications: II/B. Springer-Verlag (1990)