ELSEVIER

Applied Numerical Mathematics 38 (2001) 315–345

APPLIED NUMERICAL MATHEMATICS

www.elsevier.com/locate/apnum

Error analysis for a Galerkin-spectral method with coordinate transformation for solving singularly perturbed problems

Wenbin Liu [a,*], Tao Tang [b]

[a] *Institute of Mathematics and Statistics, The University of Kent, Canterbury, CT2 7NF, England, UK*
[b] *Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong*

Abstract

In this paper, we investigate a Galerkin-spectral method, which employs coordinate stretching and a class of trial functions suitable for solving singularly perturbed boundary value problems. An error analysis for the proposed spectral method is presented. Two transformation functions are considered in detail. In solving singularly perturbed problems with conventional spectral methods, spectral accuracy can only be obtained when $N = \mathrm{O}(\varepsilon^{-\gamma})$, where ε is the singular perturbation parameter and γ is a positive constant. Our main effort is to make this γ smaller, say from $\frac{1}{2}$ to $\frac{1}{4}$ or less for Helmholtz type equations, by using appropriate coordinate stretching. Similar results are also obtained for advection–diffusion equations. Two important features of the proposed method are as follows: (a) the coordinate transformation does not involve the singular perturbation parameter ε; (b) machine accuracy can be achieved with N of the order of several hundreds, even when ε is very small. This is in contrast with conventional spectral, finite difference or finite element methods.

Keywords: Spectral methods; Error estimates; Boundary layer

1. Introduction

In this paper we consider spectral approximation for the numerical solution of the singularly perturbed convection–diffusion equation

$$\begin{cases} -\varepsilon \Delta u(\boldsymbol{x}) + \nabla u(\boldsymbol{x}) \cdot \boldsymbol{p}(\boldsymbol{x}) + q(\boldsymbol{x}) u(\boldsymbol{x}) = f(\boldsymbol{x}, \varepsilon) \quad \text{in } \Omega, \\ u|_{\partial\Omega} = 0, \end{cases} \tag{1.1}$$

where $\Omega = (-1, 1)^d$ with $d = 1$, 2 or 3, $\varepsilon > 0$ is a small parameter, $\boldsymbol{p} = (p_1, \ldots, p_d)^{\mathrm{T}}$, q and f are smooth functions on $\overline{\Omega}$, and $\|f(\cdot, \varepsilon)\|_{L^\infty(\Omega)}$ is bounded by a constant independent of ε.

The problem (1.1) is often viewed as a basic model of a steady-state convection–diffusion process. For small values of ε, this equation in general possesses a thin boundary layer; the solution u will

* Corresponding author.
E-mail addresses: w.b.liu@ukc.ac.uk (W. Liu), ttang@math.hkbu.edu.hk (T. Tang).

0168-9274/01/$ – see front matter
PII: S0168-9274(01)00036-8

vary rapidly in the layer region near the boundary. This boundary layer causes various difficulties in seeking the numerical solution of (1.1). It is well-known that conventional numerical methods to (1.1) can produce approximate solutions with oscillations that are unbounded when $\varepsilon \to 0$. Various approaches have been proposed to eliminate these oscillations. The last decades have seen substantial progress in the development of numerical methods for the solution of (1.1) and several software packages are presently available [1,6,13,14,18]. A large body of literature has been devoted to the effective resolution of the solutions of (1.1); see, e.g., books [19,22], and references therein.

Most available references analyze the convergence of finite difference or finite element schemes of fixed (usually low) polynomial degree in conjunction with various mesh refinements; see, e.g., [11,23, 27]. An alternative approach is to increase the polynomial degree, i.e., use a p version of finite element method or spectral method. In [24], the uniform approximation of boundary layers is studied by using the p and hp versions of the finite element method. For the p version with variable mesh (i.e., the hp version), it is shown that exponential convergence, uniform in the perturbation parameter ε, is achieved by taking the first element at the boundary layer to be of size $O(p\sqrt{\varepsilon})$. Discrete methods whose solutions converge independently of ε are said to be ε-uniform. If a method is ε-uniform, mesh refinement causes the error to decrease in a manner that is independent of the perturbation parameter. The examples of ε-uniform method include Shishkin's grid [27] and Schwab and Suri's grid [24]. However, these ε-uniform methods require that the size of the first (or/and last) element is of the order of boundary layer width. In other words, the information for the width of the boundary layer should be known prior to the selection of the grid points.

In [4], a Chebyshev-weighted spectral approximation is investigated for the one-dimensional version of (1.1), with $\boldsymbol{p} \equiv 0$ and $f \equiv 0$. Spectral methods for solving singularly perturbation problems can also be found in many papers such as [3,7–9,13,15]. Although the conventional spectral methods have been found attractive in solving (1.1), *spectral accuracy* cannot be observed with reasonably large N, where N is the total number of grid points/basis functions, if ε is very small (see, e.g., [5,7,10,17]). It is expected that the spectral methods together with suitable transformations will be suitable for solving the boundary layer problems such that spectral convergence can be obtained with N of the order of several hundreds, even when ε is very small. A good reference for the transformation technique is [12]. By choosing appropriate transformation functions, a boundary layer resolving spectral method is designed in [29]. With the special choice of the transformation function, the transformed coefficient functions can be generated efficiently by machine. Numerical experiments show that the boundary layer resolving spectral method is simple (the code is just a few lines longer than the standard spectral method code) and robust. The method is much more efficient and accurate than conventional spectral methods, especially when ε is very small. One of the important features of the present work is that our coordinate transformation does not involve the singular perturbation parameter, though it is essentially impossible to resolve *arbitrarily* thin boundary layers with a non-adaptive ε-independent coordinate stretching.

Although the idea and algorithm for the spectral method with coordinate transformation are quite simple, the theoretical error analysis for the approach seems rather difficult. In practice, the transformations used have to be singular (in the sense that the derivatives of the transformed functions may be zero at the end points). One of the key steps in the analysis is to handle the transformed equation that is highly degenerate due to the singularity of the transformation used. In the case of the Galerkin-spectral approximation, one of the key questions is that what trial functions should be employed to approximate the transformed equation. In this work, a Galerkin-spectral method based on a new class of trial functions will be investigated for one and higher dimensional problems. Part of the results for

1-D has been reported in a conference paper by the authors [16]. It is the purpose of this paper to give a theoretical interpretation of the high accuracy behavior of the Galerkin-spectral method involving coordinate transformations. The work is the first step towards understanding more complicated boundary layer resolving spectral methods, such as [20,21,29].

Let $x_i = g_i(y_i)$ with $g_i \in C^\infty[-1,1]$ such that

$$\begin{cases} g_i(-1) = -1, & g_i(1) = 1, \\ J_i(y_i) := g_i'(y_i) > 0, & \text{for } y_i \in (-1,1), \qquad i = 1,\dots,d. \end{cases} \tag{1.2}$$

Applying the change of variables $\boldsymbol{x} = \boldsymbol{g}(\boldsymbol{y})$ to (1.1), we obtain

$$-\varepsilon \sum_{i=1}^{d} a_i \partial_{y_i}(a_i \partial_{y_i} v) + \sum_{i=1}^{d} a_i P_i \partial_{y_i} v + \underline{Q} v = \underline{F}, \quad \text{in } \Omega, \tag{1.3}$$

where

$$\begin{cases} v(\boldsymbol{y}) = u \circ \boldsymbol{g}(\boldsymbol{y}), \quad a_i(y_i) = \dfrac{1}{J_i(y_i)}, \quad P_i(\boldsymbol{y}) = p_i \circ \boldsymbol{g}(\boldsymbol{y}), \quad i = 1,\dots,d, \\ \underline{Q}(\boldsymbol{y}) = q \circ \boldsymbol{g}(\boldsymbol{y}), \quad \underline{F}(\boldsymbol{y},\varepsilon) = f\big(\boldsymbol{g}(\boldsymbol{y}),\varepsilon\big). \end{cases} \tag{1.4}$$

Therefore, the transformed equation has variable coefficients even when the coefficients of the original equation are constants. Furthermore, in order to obtain a finer resolution near the boundary, it is often necessary to have $J_i(-1) = g_i'(-1) = 0$ and/or $J_i(1) = g_i'(1) = 0$ for at least one index i. Hence, $a_i(y_i)$ is not even bounded near the boundary. This causes several major difficulties in approximating the solutions of the transformed equation. For instance, it is not clear what trial function spaces or collocation points should be used to discretize the equation. It is also difficult to obtain error bounds due to the degenerate character of the transformed equation.

The paper is organized as follows. In Section 2, we derive a weak formulation for (1.3). Then a general Galerkin-spectral scheme for the weak formulation is introduced. Some error bounds will be obtained in Section 3. In Section 4, the theoretical results obtained in Section 3 are illustrated by some feasible transformations. Error analysis for two-dimensional problems will be given in Section 5. Numerical results will be presented in Section 6. Note that in our presentation we first consider the easiest case (Helmholtz equations in one dimension), then a more general one-dimensional equation, and later the high-dimensional Helmholtz equation. As a result, some of the technical work needs to be done more than once in different contexts. It would be possible to start off with the more complicated problems in order to shorten the presentation, but that would make the paper less readable.

2. Weak formulation

We adopt the standard notations $L^2(\Omega)$ and $H^m(\Omega)$ to denote the usual Sobolev spaces, and $H_0^m(\Omega)$ to denote the subspace of $H^m(\Omega)$ whose elements have vanishing traces. We denote by $L^2_\omega(\Omega)$ and $H^m_\omega(\Omega)$ the weighted Sobolev spaces with the weight function ω. Let $I = (-1,1)$ and denote π_N to be the space of real polynomials on I with degrees not exceeding N. We set

$$X_N = \{u_N \in \pi_N \colon u_N(\pm 1) = 0\}.$$

We shall use letters of boldface type to denote vectors and vector functions as well as product spaces such as

$$\boldsymbol{X}_N = \prod_{i=1}^{d} X_N.$$

Let

$$\begin{cases} \omega_i(y_i) = \left(1 - y_i^2\right)^{\lambda}, \quad \text{for a fixed } \lambda \text{ with } -1 < \lambda \leqslant 0, \\ \omega(\boldsymbol{y}) = \prod_{i=1}^{d} \omega_i(y_i). \end{cases} \tag{2.1}$$

It does not seem suitable to study the weak formulation of (1.3) either in $H^1_{\omega,0}(\Omega)$ or $H^1_{a\omega,0}(\Omega)$ since some integrals (e.g., $\int_{-1}^{1} a\omega v^2 \,\mathrm{d}y$ and $\int_{-1}^{1} a\omega (v')^2 \,\mathrm{d}y$ in the case $d = 1$) may not exist in these spaces.

Let us denote $J(\boldsymbol{y}) = \prod_{i=1}^{d} J_i(y_i)$ and

$$\widetilde{H}^1_{\omega,0}(\Omega) := \left\{ v \in H^1_{\omega,0}(\Omega):\ \|v\|_{L^2_{J\omega}(\Omega)} + \sum_{i=1}^{d} \|\partial_{y_i} v\|_{L^2_{a_i^2 J\omega}(\Omega)} < \infty \right\}. \tag{2.2}$$

It is noted that all the smooth functions with compact support in Ω are in this space. A weak formulation of (1.3) can be established in $\widetilde{H}^1_{\omega,\mathbf{0}}(\Omega)$, which is the image space of $H^1_{\widetilde{\omega},\mathbf{0}}(\Omega)$ under the transformation $Gu := u \circ \boldsymbol{g}$, where $\widetilde{\omega}(\boldsymbol{x}) := \omega(\mathbf{g}^{-1}(\boldsymbol{x}))$. Let

$$\begin{cases} A(v,z)_\omega = \varepsilon \sum_{i=1}^{d} \int_{\Omega} \left(a_i^2 J\right)\left(\partial_{y_i} v \partial_{y_i}(z\omega)\right) \mathrm{d}\boldsymbol{y} + \int_{\Omega} Q v z \omega \,\mathrm{d}\boldsymbol{y}, \\ B(v,z)_\omega = \sum_{i=1}^{d} \int_{\Omega} a_i J P_i(\partial_{y_i} v) z\omega \,\mathrm{d}\boldsymbol{y}, \qquad (F,z)_\omega = \int_{\Omega} F z \omega \,\mathrm{d}\boldsymbol{y}, \end{cases} \tag{2.3}$$

where $Q(\boldsymbol{y}) = \underline{Q}(\boldsymbol{y}) J(\boldsymbol{y})$ and $F(\boldsymbol{y},\varepsilon) = \underline{F}(\boldsymbol{y},\varepsilon) J(\boldsymbol{y})$. For a fixed weight ω, we multiply Eq. (1.3) by $\omega J(\boldsymbol{y})$. The weak formulation for (1.3) is as follows: find $v \in \widetilde{H}^1_{\omega,0}(\Omega)$ such that

$$A(v,z)_\omega + B(v,z)_\omega = (F,z)_\omega, \quad \forall z \in \widetilde{H}^1_{\omega,0}. \tag{2.4}$$

We now consider the approximation of (2.4) by using a Galerkin-spectral method. It is essential to find suitable trial function spaces in order to properly approximate the solution of (2.4) in $\widetilde{H}^1_{\omega,0}(\Omega)$. It is clear that $\boldsymbol{X}_N$ is not suitable. It is then natural to consider the image space of $\boldsymbol{X}_N$ under the transformation G as the trial function space. It turns out, however, that with this trial function space we would obtain the same results as those by applying the conventional Galerkin-spectral methods directly to (1.1).

Let

$$Y_N^i = \left\{ v \in H_0^1(I):\ v' = J_i P,\ P \in \pi_N \right\}, \quad i = 1, \ldots, d. \tag{2.5}$$

It can be verified that Y_N^i is an N-dimensional subspace of $\widetilde{H}^1_{\omega,0}(I)$ with $G_i u := u \circ g_i$. It follows from (2.5) that for every element $v \in Y_N^i$ there is a unique $P_v \in \pi_N$ such that

$$v(y) = \int_{-1}^{y} J_i(t) P_v(t) \,\mathrm{d}t, \tag{2.6}$$

but the choice of P_v has to satisfy the following requirement:

$$\int_{-1}^{1} J_i(t) P_v(t)\,\mathrm{d}t = 0. \tag{2.7}$$

Observe that for any $P \in \pi_N$ there is a unique constant α_P such that

$$\int_{-1}^{1} J_i(t)\big(\alpha_P + P(t)\big)\,\mathrm{d}t = 0.$$

Thus every element $v \in Y_N^i$ can be further represented as

$$v(y) = \int_{-1}^{y} J_i(t)\big(\alpha_P + P(t)\big)\,\mathrm{d}t, \quad \text{with } P \in \pi_N,\ \alpha_P = -\frac{1}{2}\int_{-1}^{1} J_i(t) P(t)\,\mathrm{d}t, \tag{2.8}$$

where we have used the fact that

$$\int_{-1}^{1} J_i(t)\,\mathrm{d}t = 2.$$

Of course, P in (2.8) is no longer required to satisfy (2.7). It turns out that the space

$$\boldsymbol{Y}_N = \prod_{i=1}^{d} Y_N^i$$

is a good choice as the trial function space. Therefore, the proposed Galerkin-spectral approximation for (2.4) reads: find $v_N \in \boldsymbol{Y}_N$ such that

$$A(v_N, z)_\omega + B(v_N, z)_\omega = (F, z)_\omega, \quad \forall z \in \boldsymbol{Y}_N. \tag{2.9}$$

We give two sample transformations that will be used to demonstrate our theoretical results in Section 4. We consider one-dimensional case only, i.e., $d = 1$. The first one was proposed in [29],

$$x = g(y) = \sin\left(\frac{\pi y}{2}\right). \tag{2.10}$$

The second one is

$$x = g(y) = -1 + \kappa \int_{-1}^{y} \big(1 - \eta^2\big)^k\,\mathrm{d}\eta, \tag{2.11}$$

where $k > 1$ is a given constant and κ is chosen such that $g(1) = 1$. It can be shown that both transformations satisfy the following inequalities: there exist positive constants β, C_1 and C_2 such that

$$C_1 \leqslant J(y)\big(1 - y^2\big)^{-\beta} \leqslant C_2, \quad \frac{J\omega(a\omega')'}{(\omega')^2} > 2, \quad \text{for } y \in I := (-1, 1), \tag{2.12}$$

where $\omega(y) = (1 - y^2)^\lambda$, $-1 < \lambda \leqslant 0$. It will be seen in the next section that these inequalities play important roles in studying (2.9). Here we only give a brief proof for the first transformation. The first

inequality in (2.12) holds for $\beta = 1$. Direct calculation gives that (noting that $\omega(y) = (1-y^2)^\lambda$ with $\lambda \in (-1, 0]$)

$$\frac{J\omega(a\omega')'}{(\omega')^2} = -\frac{1}{2\lambda y^2}\left(1 - y^2 + 2(1-\lambda)y^2 + (1-y^2)\frac{\pi y}{2}\tan\left(\frac{\pi y}{2}\right)\right) \geqslant -\frac{1}{2\lambda y^2}2(1-\lambda)y^2 > 2.$$

This indicates that the second inequality in (2.12) holds.

3. Error analysis for 1-D

Throughout this paper, C denotes a positive constant independent of ε, N, but possibly with different values at different places. We will establish error bounds for the Galerkin method (2.9). In this section, we assume $d = 1$ and begin with the one-dimensional Helmholtz type equation.

3.1. Helmholtz type equation

The one-dimensional Helmholtz problem is as follows:

$$-\varepsilon u''(x) + q(x)u(x) = f(x,\varepsilon), \quad x \in I := (-1,1), \qquad u(\pm 1) = 0. \tag{3.1}$$

Here we assume that q is positive on $[-1, 1]$. After the transformation $x = g(y)$, problem (3.1) becomes

$$-\varepsilon\big(a(y)v'(y)\big)' + Q(y)v(y) = F(y), \quad y \in I, \qquad v(\pm 1) = 0. \tag{3.2}$$

Lemma 3.1. *Assume that there exist positive constants β, C_1 and C_2 such that* (2.12) holds. Then for any $z, v \in \widetilde{H}^1_{\omega,0}(I)$,

$$A(v,v)_\omega \geqslant C\varepsilon \int_{-1}^{1} a\omega\big(v'\big)^2\,\mathrm{d}y + \int_{-1}^{1} \omega Q v^2\,\mathrm{d}y, \tag{3.3}$$

$$\big|A(v,z)_\omega\big| \leqslant C\varepsilon\big\|v'\big\|_{L^2_{a\omega}(I)}\big\|z'\big\|_{L^2_{a\omega}(I)} + C\|v\|_{L^2_Q(I)}\|z\|_{L^2_Q(I)}, \tag{3.4}$$

where $A(\cdot,\cdot)_\omega$ is defined by

$$A(v,z)_\omega = \varepsilon \int_{-1}^{1} av'(\omega z)'\,\mathrm{d}y + \int_{-1}^{1} \omega Q v z\,\mathrm{d}y. \tag{3.5}$$

Proof. The proof follows a similar idea used in [5]. For any $v \in \widetilde{H}^1_{\omega,0}(I)$,

$$\int_{-1}^{1} av'(v\omega)'\,\mathrm{d}y = \int_{-1}^{1} a\omega\big(v'\big)^2\,\mathrm{d}y + \int_{-1}^{1} avv'\omega'\,\mathrm{d}y = \int_{-1}^{1} a\omega\big(v'\big)^2\,\mathrm{d}y - \frac{1}{2}\int_{-1}^{1} v^2\big(a\omega'\big)'\,\mathrm{d}y. \tag{3.6}$$

On the other hand,

$$\int_{-1}^{1} av'(v\omega)'\,\mathrm{d}y = \int_{-1}^{1} \big[\big(v'\omega\big)^2 + v\omega v'\omega'\big](J\omega)^{-1}\,\mathrm{d}y$$

$$= \int_{-1}^{1} (v'\omega + v\omega')^2 (J\omega)^{-1}\,\mathrm{d}y - \int_{-1}^{1} [av v'\omega' + (v\omega')^2 (J\omega)^{-1}]\,\mathrm{d}y$$

$$= \int_{-1}^{1} [(v\omega)']^2 (J\omega)^{-1}\,\mathrm{d}y + \int_{-1}^{1} v^2 \left[\frac{1}{2}(a\omega')' - (\omega')^2 (J\omega)^{-1}\right]\mathrm{d}y$$

$$\geqslant \int_{-1}^{1} v^2 \left[\frac{1}{2}(a\omega')' - (\omega')^2 (J\omega)^{-1}\right]\mathrm{d}y. \tag{3.7}$$

Combining (3.6), (3.7) and the second inequality in (2.12) gives (3.3). Furthermore,

$$\left| \int_{-1}^{1} av'(z\omega)'\,\mathrm{d}y \right| \leqslant \left| \int_{-1}^{1} a\omega v' z'\,\mathrm{d}y \right| + \left| \int_{-1}^{1} av' z\omega'\,\mathrm{d}y \right|. \tag{3.8}$$

Let $\omega_\lambda(y) = (1-y)^\lambda$. We have

$$\left| \int_{0}^{1} av' z\omega'\,\mathrm{d}y \right|^2 \leqslant C \left(\int_{0}^{1} a\omega_\lambda |v' z\omega'_\lambda/\omega_\lambda|\mathrm{d}y \right)^2$$

$$\leqslant C \int_{0}^{1} a\omega_\lambda (v')^2\,\mathrm{d}y \int_{0}^{1} a\omega_\lambda \left(\frac{z\omega'_\lambda}{\omega_\lambda}\right)^2 \mathrm{d}y$$

$$\leqslant C \left(\int_{0}^{1} \omega_{\lambda-\beta} (v')^2\,\mathrm{d}y \right) \left(\int_{0}^{1} \omega_{\lambda-\beta-2} z^2\,\mathrm{d}y \right),$$

where we have used the first inequality of (2.12) that gives $0 < a\omega_\lambda < C\omega_{\lambda-\beta}$ for $y \in (0,1)$. It then follows from Hardy's inequality (see, e.g., [5]) that

$$\left| \int_{0}^{1} av' z\omega'\,\mathrm{d}y \right|^2 \leqslant C \left(\int_{0}^{1} \omega_{\lambda-\beta} (v')^2\,\mathrm{d}y \right) \left(\int_{0}^{1} \omega_{\lambda-\beta} (z')^2\,\mathrm{d}y \right) \leqslant C \|v'\|^2_{L^2_{a\omega}(I)} \|z'\|^2_{L^2_{a\omega}(I)}.$$

Similarly, we can show that

$$\left| \int_{-1}^{0} av' z\omega'\,\mathrm{d}y \right|^2 \leqslant C \|v'\|^2_{L^2_{a\omega}(I)} \|z'\|^2_{L^2_{a\omega}(I)}. \tag{3.9}$$

Combining (3.8)–(3.9) yields (3.4). □

The weak formula of the transformed problem (3.2) reads: find $v \in \widetilde{H}^1_{\omega,0}(I)$ such that

$$A(v,z)_\omega = (F,z)_\omega, \qquad \forall z \in \widetilde{H}^1_{\omega,0}(I). \tag{3.10}$$

It follows from Lemma 3.1 and the Lax–Milgram theorem that the above problem has a unique solution. Our Galerkin-spectral method for the transformed problem (3.2) reads: find $v_N \in Y_N$ such that

$$A(v_N,z)_\omega = (F,z)_\omega, \qquad \forall z \in Y_N, \tag{3.11}$$

where $Y_N = \{v \in H_0^1(I)\colon\ v' = JP,\ P \in \pi_N\},\ J(y) = g'(y)$. The well-posedness of (3.11) follows from Lemma 3.1 and the Lax–Milgram theorem.

For any $\phi \in \widetilde{H}^1_{\omega,0}(I)$, we define its projection $\Pi\phi$ in Y_N such that

$$\int_{-1}^{1} a\big((\phi - \Pi\phi)\omega\big)' h'\,\mathrm{d}y = 0, \quad \text{for any } h \in Y_N. \tag{3.12}$$

It is clear that $\Pi\phi$ is uniquely defined in Y_N.

Lemma 3.2. *Let $u(x)$ be the unique solution of* (3.1) *and $v(y) = u(g(y))$ be the solution of* (3.2)*. Then*

$$\int_{-1}^{1} \omega a\big(v' - (\Pi v)'\big)^2\,\mathrm{d}y \leqslant CN^{-2}T_{\tilde{\omega},1}, \tag{3.13}$$

$$\int_{-1}^{1} \omega a\big(v' - (\Pi v)'\big)^2\,\mathrm{d}y \leqslant CN^{-4}T_{\tilde{\omega},2}, \tag{3.14}$$

where

$$T_{\tilde{\omega},1} = \int_{-1}^{1} \widetilde{\omega}\big(u'\big)^2\,\mathrm{d}x + \int_{-1}^{1} \widetilde{\omega}\underline{J}^2\big(u''\big)^2\,\mathrm{d}x, \tag{3.15a}$$

$$T_{\tilde{\omega},2} = \int_{-1}^{1} \widetilde{\omega}\big(u'\big)^2\,\mathrm{d}x + \int_{-1}^{1} \widetilde{\omega}\underline{J}^2\big(u''\big)^2\,\mathrm{d}x + \int_{-1}^{1} \widetilde{\omega}\underline{J}^4\big(u'''\big)^2\,\mathrm{d}x + \int_{-1}^{1} \widetilde{\omega}\big(\underline{J}'\big)^2\big(u''\big)^2\,\mathrm{d}x. \tag{3.15b}$$

Here $\widetilde{\omega}(x) = \omega(g^{-1}(x))$, $\underline{J}(x) = J(g^{-1}(x))$ and $\underline{J}'(x) = J'(g^{-1}(x))$.

Proof. For any $\phi \in \widetilde{H}^1_{\omega,0}(I)$, let ϕ_N be the best approximation polynomial for $a\phi'$ in $L^2_{\omega J_\beta}(I)$, with $J_\beta = (1-y^2)^\beta$. Let

$$\Pi^*\phi = \int_{-1}^{y} J(\phi_N + \alpha_{\phi_N})\,\mathrm{d}y, \quad \alpha_{\phi_N} = -\frac{1}{2}\int_{-1}^{1} J\phi_N\,\mathrm{d}y. \tag{3.16}$$

Since $\int_{-1}^{1} J\,\mathrm{d}y = g(1) - g(-1) = 2$, we can easily show that $\Pi^*\phi \in Y_N$. $\Pi^*\phi$ is a modification of $\Pi\phi$ and it is the same as $\Pi\phi$ if $\omega J_\beta = 1$. However, it is easier to derive an upper bound for the modification than $\Pi\phi$ itself, since the modification is given explicitly. From the proof of Theorem 4.1 in [2] (note that the result is proved only for $|\beta + \lambda| < 1$ in [2], but it can be shown that for $\beta > 0$ and $\sigma = 2$ or $\sigma = 2m$), we have

$$\int_{-1}^{1} \omega J_\beta\big(a\phi' - \phi_N\big)^2\,\mathrm{d}y \leqslant CN^{-4}\big\|a\phi'\big\|^2_{H^2_{\omega J_\beta}(I)} \leqslant CN^{-4}\big\|a\phi'\big\|^2_{H^2_{\omega J}(I)}, \tag{3.17a}$$

$$\int_{-1}^{1} \omega J_\beta\big(a\phi' - \phi_N\big)^2\,\mathrm{d}y \leqslant CN^{-2}\big\|a\phi'\big\|^2_{H^1_{\omega J_\beta}(I)} \leqslant CN^{-2}\big\|a\phi'\big\|^2_{H^1_{\omega J}(I)}, \tag{3.17b}$$

see (4.5) of [2]. It follows from Lemma 3.1 and (3.12) that for any $\phi \in \widetilde{H}^1_{\omega,0}(I)$

$$\int_{-1}^{1} a\omega\big(\phi' - (\Pi\phi)'\big)^2\,\mathrm{d}y \leqslant C\int_{-1}^{1} a(\phi-\Pi\phi)'\big((\phi-\Pi\phi)\omega\big)'\,\mathrm{d}y = C\int_{-1}^{1} a\big(\phi-\Pi^*\phi\big)'\big((\phi-\Pi\phi)\omega\big)'\,\mathrm{d}y$$
$$\leqslant \left(\int_{-1}^{1} a\omega\big(\phi' - (\Pi\phi)'\big)^2\,\mathrm{d}y\right)^{1/2}\left(\int_{-1}^{1} a\omega\big(\phi' - \big(\Pi^*\phi\big)'\big)^2\,\mathrm{d}y\right)^{1/2},$$

which implies that

$$\int_{-1}^{1} a\omega\big(\phi' - (\Pi\phi)'\big)^2\,\mathrm{d}y \leqslant \int_{-1}^{1} a\omega\big(\phi' - \big(\Pi^*\phi\big)'\big)^2\,\mathrm{d}y. \tag{3.18}$$

Now let $w(y) := u'(g(y))$ and let w_N be the best approximation polynomial for w in $L^2_{\omega J_\beta}(-1,1)$. Thus from (3.18) and (3.16),

$$\left(\int_{-1}^{1} \omega a\big(v' - (\Pi v)'\big)^2\,\mathrm{d}y\right)^{1/2} \leqslant \left(\int_{-1}^{1} \omega J\big(w - (w_N + \alpha_{w_N})\big)^2\,\mathrm{d}y\right)^{1/2}$$
$$\leqslant C\left(\int_{-1}^{1} \omega J_\beta (w-w_N)^2\,\mathrm{d}y\right)^{1/2} + C\left|\int_{-1}^{1} J w_N\,\mathrm{d}y\right|.$$

Note that $\int_{-1}^{1} J w\,\mathrm{d}y = u(g(1)) - u(g(-1)) = 0$. Then using (3.17) and the above inequality gives

$$\left(\int_{-1}^{1} \omega a\big(v' - \big(\Pi v\big)'\big)^2\,\mathrm{d}y\right)^{1/2} \leqslant CN^{-1}\big\|av'\big\|_{H^1_{J_\omega}(I)} + C\left|\int_{-1}^{1} (Jw_N - Jw)\,\mathrm{d}y\right|$$
$$\leqslant CN^{-1}\big\|av'\big\|_{H^1_{J_\omega}(I)} + C\left(\int_{-1}^{1} \omega J_\beta (w_N - w)^2\,\mathrm{d}y\right)^{1/2}$$
$$\leqslant CN^{-1}\big\|av'\big\|_{H^1_{J_\omega}(I)}$$
$$= CN^{-1}\left(\int_{-1}^{1} \widetilde{\omega}\big(u'(x)\big)^2\,\mathrm{d}x + \int_{-1}^{1} \widetilde{\omega}\underline{J}^2\big(u''(x)\big)^2\,\mathrm{d}x\right)^{1/2}. \tag{3.19}$$

This gives (3.13). The bound (3.14) can be established in a similar way. This completes the proof of Lemma 3.2. □

Lemma 3.3. *Let $u(x)$ be the unique solution of* (3.1) *and $v(y) = u(g(y))$ be the solution of* (3.2)*. Then*

$$\int_{-1}^{1} \omega Q(v - \Pi v)^2\,\mathrm{d}y \leqslant CN^{-4}T_{\tilde{\omega},1}, \tag{3.20}$$

$$\int_{-1}^{1} \omega Q(v-\Pi v)^2\,\mathrm{d}y \leqslant CN^{-6}T_{\tilde{\omega},2}, \tag{3.21}$$

where $T_{\tilde{\omega},1}$ and $T_{\tilde{\omega},2}$ are defined by (3.15).

Proof. Observe that

$$\begin{aligned}\int_{-1}^{1} \omega Q(v-\Pi v)^2\,\mathrm{d}y \leqslant C\int_{-1}^{1} \omega J(v-\Pi v)^2\,\mathrm{d}y &= C\sup_{g\in L^2_{\omega J}(I)} \frac{(\int_{-1}^{1}\omega J(v-\Pi v)g\,\mathrm{d}y)^2}{\|g\|^2_{L^2_{\omega J}(I)}}\\ &= C\sup_{g\in L^2_{\omega J}(I)} \frac{(\int_{-1}^{1} a(\psi-\Pi\psi)'(\omega(v-\Pi v)')\,\mathrm{d}y)^2}{\|g\|^2_{L^2_{\omega J}(I)}},\end{aligned}$$

where $\psi\in \widetilde{H}^1_{\omega,0}(I)$ is defined by

$$\int_{-1}^{1} a\psi'(\omega h)'\,\mathrm{d}y = \int_{-1}^{1} J\omega g h\,\mathrm{d}y, \quad \text{for any } h\in \widetilde{H}^1_{\omega,0}(I). \tag{3.22}$$

It follows from Lemma 3.2 and inequalities similar to (3.19) that

$$\begin{aligned}\int_{-1}^{1} \omega Q(v-\Pi v)^2\,\mathrm{d}y &\leqslant C\big\|(v-\Pi v)'\big\|^2_{L^2_{\omega J}(I)} \sup_{g\in L^2_{\omega J}(I)} \frac{\|(\psi-\Pi\psi)'\|^2_{L^2_{\omega J}(I)}}{\|g\|^2_{L^2_{\omega J}(I)}}\\ &\leqslant CN^{-4}\big\|av'\big\|^2_{H^1_{\omega J}(I)} \sup_{g\in L^2_{\omega J}(I)} \|a\psi'\|^2_{H^1_{\omega J}(I)}/\|g\|^2_{L^2_{\omega J}(I)},\end{aligned}$$

and

$$\int_{-1}^{1} \omega Q(v-\Pi v)^2\,\mathrm{d}y \leqslant CN^{-6}\big\|av'\big\|^2_{H^2_{\omega J}(I)} \sup_{g\in L^2_{\omega J}(I)} \big\|a\psi'\big\|^2_{H^1_{\omega J}(I)}/\|g\|^2_{L^2_{\omega J}(I)}.$$

Using Lemma 3.1 and (3.22) gives

$$\big\|a\psi'\big\|_{H^1_{\omega J}(I)}/\|g\|_{L^2_{\omega J}(I)} \leqslant C, \quad \text{for } g\in L^2_{\omega J}(I),$$

where C is independent of g. Therefore, we have

$$\int_{-1}^{1} \omega Q(v-\Pi v)^2\,\mathrm{d}y \leqslant CN^{-4}\big\|av'\big\|^2_{H^1_{\omega J}(I)}, \qquad \int_{-1}^{1} \omega Q(v-\Pi v)^2\,\mathrm{d}y \leqslant CN^{-6}\big\|av'\big\|^2_{H^2_{\omega J}(I)}.$$

This completes the proof of Lemma 3.3. □

Having the above lemmas, we are ready to state and prove the main result of this section.

Theorem 3.1. *Let $u(x)$ be the unique solution of* (3.1) *and $v_N(y)$ be the unique solution of* (3.11). *Assume that q in* (3.1) *is positive and J satisfies* (2.12). *Then the following error bounds can be obtained*:

$$\varepsilon\|u' - \underline{v}'_N\|^2_{L^2_{\tilde{\omega}}(I)} + \|u - \underline{v}_N\|^2 \leqslant C\big(N^{-2}\varepsilon + N^{-4}\big)T_{\tilde{\omega},1}, \tag{3.23}$$

$$\varepsilon\|u' - \underline{v}'_N\|^2_{L^2_{\tilde{\omega}}(I)} + \|u - \underline{v}_N\|^2_{L^2_{\tilde{\omega}}(I)} \leqslant C\big(N^{-4}\varepsilon + N^{-6}\big)T_{\tilde{\omega},2}, \tag{3.24}$$

where $\underline{v}_N(x) = v_N(g^{-1}(x))$, and $T_{\tilde{\omega},1}$ and $T_{\tilde{\omega},2}$ are defined by (3.15).

Proof. Again let $v(y) = u(g(y))$ be the solution of (3.2). Let v_N be the solution of (3.11). It follows from the Poincare's inequality, Lemma 3.1, and the standard error estimate for the Galerkin approximation that

$$\varepsilon\|v' - v'_N\|^2_{L^2_{\tilde{\omega}a}(I)} + \|v - v_N\|^2_{L^2_{\omega Q}(-1,1)} \leqslant C \min_{h\in Y_N} \int_{-1}^{1} \big(\varepsilon a\big(v' - h'\big)^2 + Q(v-h)^2\big)\omega\,\mathrm{d}y. \tag{3.25}$$

Observe that

$$\varepsilon\|v' - v'_N\|^2_{L^2_{\tilde{\omega}a}(I)} + \|v - v_N\|^2_{L^2_{\omega Q}(I)} = \varepsilon\|u' - \underline{v}'_N\|^2_{L^2_{\tilde{\omega}}(I)} + \|u - \underline{v}_N\|^2_{L^2_{q\tilde{\omega}}(I)}, \tag{3.26a}$$

$$\|u - \underline{v}_N\|_{L^2_{\tilde{\omega},q}(I)} \geqslant C\|u - \underline{v}_N\|_{L^2_{\tilde{\omega}}(I)}. \tag{3.26b}$$

The last inequality is due to the assumption that q is positive on $[-1, 1]$. Using Lemmas 3.2 and 3.3, together with (3.26), leads to (3.23) and (3.24) . □

The above results can be generalized to higher order approximation if $u^{(m+1)}$ $(m > 2)$ exists. The dominant term on the right side of (3.23) or (3.24) will then be the integral

$$C\big(N^{-2m}\varepsilon + N^{-2(m+1)}\big)\left(\int_{-1}^{1} \widetilde{\omega}\underline{J}^{2m}\big(u^{(m+1)}\big)^2\,\mathrm{d}x\right). \tag{3.27}$$

Remark 1. The most useful feature of Theorem 3.1 is that as $\varepsilon \to 0$, the dominant terms in the right hand sides of (3.23) and (3.24) can be controlled by choosing suitable J. This is the essential difference of such estimates with the conventional ones and will be demonstrated further later. This seems to give a theoretical interpretation for the efficiency of our Galerkin-spectral scheme when $\varepsilon \ll 1$. If applying the conventional Galerkin-spectral methods to (3.1) directly (i.e., without using any transformation), then we only have

$$\varepsilon\|u' - u'_N\|^2_{L^2(I)} + \|u - u_N\|^2_{L^2_q(I)} \leqslant C\big(\varepsilon N^{-2} + N^{-4}\big)\left(\int_{-1}^{1}\big(u'\big)^2\,\mathrm{d}x + \int_{-1}^{1}\big(u''\big)^2\,\mathrm{d}x\right),$$

$$\varepsilon\|u' - u'_N\|^2_{L^2(I)} + \|u - u_N\|^2_{L^2_q(I)} \leqslant C\big(\varepsilon N^{-4} + N^{-6}\big)\left(\int_{-1}^{1}\big(u'\big)^2\,\mathrm{d}x + \int_{-1}^{1}\big(u''\big)^2\,\mathrm{d}x + \int_{-1}^{1}\big(u'''\big)^2\,\mathrm{d}x\right).$$

In the case of having boundary layers the term $\int_{-1}^{1}(u'')^2\,\mathrm{d}x$ and $\int_{-1}^{1}(u''')^2\,\mathrm{d}x$ (or $\int_{-1}^{1}\underline{J}^2(u'')^2\,\mathrm{d}x$ and $\int_{-1}^{1}\underline{J}^4(u''')^2\,\mathrm{d}x$ in (3.23)–(3.24)) are usually the dominant ones in the above error estimates (or in (3.23)–(3.24)) as $\varepsilon\to 0$. In many cases one can show that

$$\int_{-1}^{1}\underline{J}^2\big(u''\big)^2\,\mathrm{d}x \ll \int_{-1}^{1}\big(u''\big)^2\,\mathrm{d}x, \quad \int_{-1}^{1}\underline{J}^4\big(u'''\big)^2\,\mathrm{d}x \ll \int_{-1}^{1}\big(u'''\big)^2\,\mathrm{d}x, \quad \text{as } \varepsilon\to 0.$$

This seems the key gain of the present method over the conventional methods.

3.2. Advection–diffusion equation

We now consider the following perturbation problem:

$$-\varepsilon u''(x)+p(x)u'(x)+q(x)u(x)=f(x,\varepsilon),\quad x\in I,\qquad u(\pm1)=0. \tag{3.28}$$

To simplify our analysis, we will restrict ourselves to a special class of problems by assuming that

$$c(x):=-\frac{p'(x)}{2}+q(x)\geqslant 0\quad \text{for } x\in[-1,1]. \tag{3.29}$$

This assumption makes the analysis simpler and yet can cover many useful cases. Under this assumption, the well-posedness of (3.28) is standard. The transformed equation corresponding to (1.3) reads as

$$-\varepsilon\big(a(y)v'(y)\big)'+P(y)v'(y)+Q(y)v(y)=F(y,\varepsilon),\quad y\in I,\qquad v(\pm1)=0. \tag{3.30}$$

It is much more difficult to analyze our numerical schemes for (3.30) than for the Helmholtz type equations. The techniques used in [4] do not seem applicable here unless ε is not too small.

We now examine the weak formulation of (3.30). Let

$$\begin{cases} A(v,z)_\omega=\varepsilon\int_{-1}^{1}av'(\omega z)'\,\mathrm{d}y+\int_{-1}^{1}\omega Qvz\,\mathrm{d}y, & \text{for } v,z\in\widetilde{H}^1_{\omega,0}(I),\\ B(v,z)_\omega=\int_{-1}^{1}Pv'\omega z\,\mathrm{d}y & \text{for } v,z\in\widetilde{H}^1_{\omega,0}(I). \end{cases} \tag{3.31}$$

The weak formulation is now stated as follows: find $v\in\widetilde{H}^1_{\omega,0}(I)$ such that

$$A(v,z)_\omega+B(v,z)_\omega=\int_{-1}^{1}\omega Fz\,\mathrm{d}y,\quad \forall z\in\widetilde{H}^1_{\omega,0}(-1,1). \tag{3.32}$$

The Galerkin approximation of (3.30) reads as follows: find $v_N\in Y_N$ such that

$$A(v_N,h)_\omega+B(v_N,h)_\omega=\int_{-1}^{1}\omega Fh\,\mathrm{d}y,\quad \forall h\in Y_N. \tag{3.33}$$

To illustrate the main idea for establishing the error estimates for the above Galerkin approximation, we will only consider (3.33) in the case $\omega\equiv 1$.

Theorem 3.2. *Let $c(x)\geqslant 0$ on $[-1,1]$ and $\omega\equiv 1$. Eq.* (3.33) *is well posed in Y_N. Let $u(x)$ be the unique solution for* (3.28) *and $v_N(y)$ be that of* (3.33). *Assume that there exist $C_1, C_2, \beta>0$ such that*

$$C_1\leqslant J(y)\big(1-y^2\big)^{-\beta}\leqslant C_2\quad \forall\, y\in[-1,1].$$

Then the following error bounds can be obtained:

$$\varepsilon\|u' - \underline{v}'_N\|^2_{L^2(I)} + \|u - \underline{v}_N\|^2_{L^2_{\tilde{c}}(I)} \leqslant C\big(N^{-2}\varepsilon + N^{-4}\varepsilon^{-1}\big)T_1, \tag{3.34}$$

$$\varepsilon\|u' - \underline{v}'_N\|^2_{L^2(I)} + \|u - \underline{v}_N\|^2_{L^2_{\tilde{c}}(I)} \leqslant C\big(N^{-4}\varepsilon + N^{-6}\varepsilon^{-1}\big)T_2, \tag{3.35}$$

where

$$T_1 = \left(\int_{-1}^{1}(u')^2\,\mathrm{d}x + \int_{-1}^{1}\underline{J}^2(u'')^2\,\mathrm{d}x\right), \tag{3.36a}$$

$$T_2 = \left(\int_{-1}^{1}(u')^2\,\mathrm{d}x + \int_{-1}^{1}\underline{J}^2(u'')^2\,\mathrm{d}x + \int_{-1}^{1}\underline{J}^4(u''')^2\,\mathrm{d}x + \int_{-1}^{1}(\underline{J}')^2(u'')^2\,\mathrm{d}x\right). \tag{3.36b}$$

Here $\underline{v}_N(x) = v_N(g^{-1}(x))$, $\underline{J}(x) = J(g^{-1}(x))$ *and* $\underline{J}'(x) = J'(g^{-1}(x))$. *Moreover, if* $c(x) > 0$ *on* $[-1, 1]$, *we further have*

$$\varepsilon\|u' - \underline{v}'_N\|^2_{L^2(I)} + \|u - \underline{v}_N\|^2_{L^2_{\tilde{c}}(I)} \leqslant C\big(N^{-2} + N^{-4}\big)T_1, \tag{3.37}$$

$$\varepsilon\|u' - \underline{v}'_N\|^2_{L^2(I)} + \|u - \underline{v}_N\|^2_{L^2_{\tilde{c}}(I)} \leqslant C\big(N^{-4} + N^{-6}\big)T_2. \tag{3.38}$$

Proof. We begin with the following observations:

$$2\int_{-1}^{1}Pw'w\,\mathrm{d}y = -\int_{-1}^{1}P'w^2\,\mathrm{d}y = -\int_{-1}^{1}Jp'w^2\,\mathrm{d}y, \quad \forall\, w \in \widetilde{H}^1_0(I). \tag{3.39}$$

Then it can be shown that there exists a constant $C > 0$ independent of ε such that for any $v, w \in \widetilde{H}^1_0(I)$

$$\begin{cases} \varepsilon\|v'\|^2_{L^2_a(I)} + \|v\|^2_{L^2_{J\tilde{c}}(I)} \leqslant CB(v, v), \\ \varepsilon A(v, w) \leqslant \varepsilon\|v'\|_{L^2_a(I)}\|w'\|_{L^2_a(I)} + \|v\|_{L^2_Q(I)}\|w\|_{L^2_Q(I)} + \|v'\|_{L^2_a(I)}\|Pw\|_{L^2_J(I)}, \end{cases} \tag{3.40}$$

where $\tilde{c} = c(g(y))$. Without loss of generality, we will assume that the constant C in (3.40) equal to the unit (i.e., $C = 1$). Poincare's inequality and (3.40) imply that the well posedness of (3.33). From the second inequality in (3.40) we further have

$$B(v, w) \leqslant \frac{\varepsilon}{2}\big(\|v'\|^2_{L^2_a(I)} + \|w'\|^2_{L^2_a(I)}\big) + \frac{\varepsilon}{8}\|v\|^2_{L^2_Q(I)} + \frac{8}{\varepsilon}\|w\|^2_{L^2_Q(I)} + \frac{\varepsilon}{8}\|v'\|^2_{L^2_a(I)} + \frac{8}{\varepsilon}\|Pw\|^2_{L^2_J(I)}. \tag{3.41}$$

It follows from (3.32) and (3.33) that for any $h \in Y_N$

$$B(v - v_N, v - v_N) = B(v - v_N, v - h),$$

which leads to

$$\varepsilon\|v' - v'_N\|^2_{L^2_a(I)} + \|v - v_N\|^2_{L^2_{J\tilde{c}}(I)} \leqslant C\min_{h\in Y_N}\int_{-1}^{1}\big(\varepsilon a(v' - h')^2 + J\varepsilon^{-1}(v - h)^2 + Q\varepsilon^{-1}(v - h)^2\big)\,\mathrm{d}y.$$

Then following the same procedure as in the proof of Theorem 3.1 we can prove (3.34) and (3.35). If $c(x) \geqslant c_0 > 0$ on $[-1, 1]$, then

$$\begin{aligned} B(v, w) &\leqslant \varepsilon \|v'\|_{L^2_a(I)} \|w'\|^2_{L^2_a(I)} + \|v\|_{L^2_Q(I)} \|w\|_{L^2_Q(I)} + C\|v\|_{L^2_J(I)} \|w\|_{L^2_J(I)} + \|v\|_{L^2_J(I)} \|w'\|_{L^2_a(I)} \\ &\leqslant \frac{\varepsilon}{2} \left(\|v'\|^2_{L^2_a(I)} + \|w'\|^2_{L^2_a(I)} \right) + C\delta \|v\|^2_{L^2_J(I)} + C\delta^{-1} \|w\|^2_{L^2_J(I)} + \delta \|v\|^2_{L^2_J(I)} + \delta^{-1} \|w'\|^2_{L^2_a(I)}, \end{aligned} \tag{3.42}$$

where δ is a positive constant. Since $c(x) \geqslant c_0 > 0$, we can make δ small enough such that there exists a $L > 0$ satisfying

$$\|v\|^2_{L^2_{J\tilde{c}}(I)} - C\delta \|v\|^2_{L^2_J(I)} + \delta \|v\|^2_{L^2_J(I)} \geqslant L \|v\|^2_{L^2_{J\tilde{c}}(I)}. \tag{3.43}$$

Therefore, we have

$$\varepsilon \|v' - v'_N\|^2_{L^2_a(I)} + \|v - v_N\|^2_{L^2_{J\tilde{c}}(I)} \leqslant C \min_{h \in Y_N} \int_{-1}^{1} \left(a(v' - h')^2 + J(v - h)^2 \right) \mathrm{d}y.$$

Consequently, using Lemmas 3.2 and 3.3 we obtain the error bounds (3.37) and (3.38). □

4. Applications of the error estimates

In this section, we give some applications of Theorems 3.1 and 3.2. We consider two cases: the first case is for constant weight and the second one is for the weight $\omega(y) = (1 - y^2)^\lambda$.

4.1. Constant weight

Assumption 4.1. *Assume that there are positive constants α, ν, C such that for $x \in [-1, 1]$*

$$\left| u^{(i)}(x) \right| \leqslant C + C\varepsilon^{-i/2} \left(\mathrm{e}^{-\alpha(1-x)/\sqrt{\varepsilon}} + \mathrm{e}^{-\nu(1+x)/\sqrt{\varepsilon}} \right), \quad i = 1, 2, \ldots, \tag{4.1}$$

where u is the solution of (3.1).

The above assumption is indeed the case for many equations of the Helmholtz type [11]. Let us consider the transformation

$$x = g(y) = \sin\left(\frac{\pi}{2} y \right). \tag{4.2}$$

Straightforward calculation yields that

$$\begin{cases} J(y) = \dfrac{\pi}{2} \cos\left(\dfrac{\pi}{2} y \right), & \underline{J}(x) = J\left(g^{-1}(x) \right) = \dfrac{\pi}{2} \sqrt{1 - x^2}, \\ \underline{J}'(x) = J'\left(g^{-1}(x) \right) = -\dfrac{\pi^2}{4} x. \end{cases} \tag{4.3}$$

Therefore,

$$\begin{aligned} T_1 &\leqslant C + C\varepsilon^{-1}\int_{-1}^{1}\left(\mathrm{e}^{-2\alpha(1-x)/\sqrt{\varepsilon}} + \mathrm{e}^{-2\nu(1+x)/\sqrt{\varepsilon}}\right)\mathrm{d}x \\ &\quad + C\varepsilon^{-2}\int_{-1}^{1}\left(\mathrm{e}^{-2\alpha(1-x)/\sqrt{\varepsilon}} + \mathrm{e}^{-2\nu(1+x)/\sqrt{\varepsilon}}\right)\left(1-x^2\right)\mathrm{d}x \\ &\leqslant C + C\varepsilon^{-1/2} + C\varepsilon^{-1} \leqslant C\varepsilon^{-1}, \end{aligned}$$

where T_1 is defined by (3.36). Similarly, we have

$$T_2 \leqslant C\varepsilon^{-3/2}.$$

It follows from Theorem 3.1 that

$$\begin{cases} \varepsilon\left\|u' - \underline{v}_N'\right\|_{L^2(I)}^2 + \left\|u - \underline{v}_N\right\|_{L^2(I)}^2 \leqslant C\left(N^{-2} + N^{-4}/\varepsilon\right), \\ \varepsilon\left\|u' - \underline{v}_N'\right\|_{L^2(I)}^2 + \left\|u - \underline{v}_N\right\|_{L^2(I)}^2 \leqslant C\left(N^{-4}\varepsilon^{-1/2} + N^{-6}\varepsilon^{-3/2}\right). \end{cases} \tag{4.4}$$

It can be proved that the above estimates can be generalized to give the following result:

$$\varepsilon\left\|u' - \underline{v}_N'\right\|_{L^2(I)}^2 + \left\|u - \underline{v}_N\right\|_{L^2(I)}^2 \leqslant C(m)\left(N^{-2m}\varepsilon^{-(m-1)/2} + N^{-2m-2}\varepsilon^{-(m+1)/2}\right), \tag{4.5}$$

for all $m \geqslant 1$ (cf. (3.27)), where $C(m)$ is a constant dependent on m, but independent of N and ε.

In Fig. 1, we plot the error bound in (4.5) with $m = 10$, by assuming that $C(m) = 1$. It is seen that with N of the order of several hundreds the errors become very small, even for very small values of ε. We further give the following result.

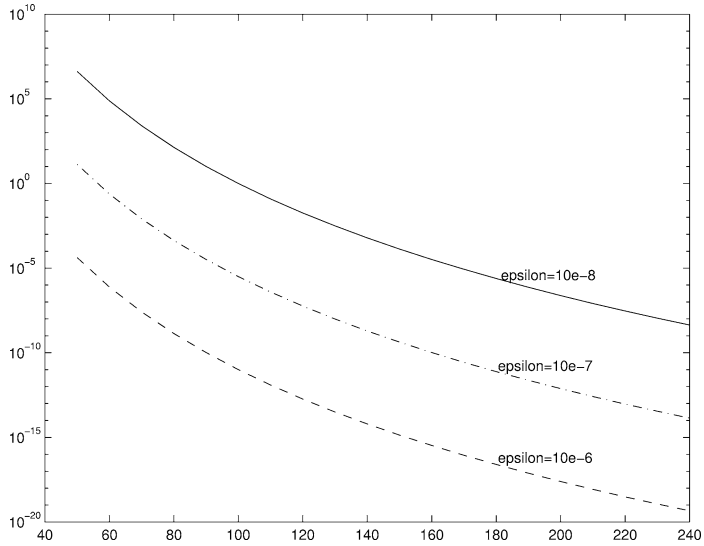

Fig. 1. Error bounds in (4.5), with $m = 10$.

Theorem 4.1. *Assume the Galerkin-spectral method* (3.11) *with the transformation* (4.2) *be applied to the Helmholtz type equation* (3.1). *If the Assumption* 4.1 *holds, then a spectral convergence rate can be obtained in* $L^2(I)$ *provided that*

$$N \geqslant \varepsilon^{-1/4-\delta}, \tag{4.6}$$

where δ *is an arbitrary positive constant.*

Proof. The right hand side of (4.5) can be written as

$$C(m)\big(\big(N\varepsilon^{1/4}\big)^{-2m}\sqrt{\varepsilon} + \big(N\varepsilon^{1/4}\big)^{-2m-2}\big).$$

Therefore, spectral accuracy can be obtained as long as $N \geqslant \varepsilon^{-1/4-\delta}$ with $\delta > 0$. □

Let us consider another transformation:

$$x = g(y) = -1 + \kappa \int_{-1}^{y} \big(1-\eta^2\big)^k \,\mathrm{d}\eta, \qquad k \geqslant 1 \quad \text{and} \quad \kappa = \frac{2}{\int_{-1}^{1}(1-y^2)^k \,\mathrm{d}y}, \tag{4.7}$$

which is exactly (2.11). It was pointed out in Section 2 that this transformation satisfies (2.12). Moreover, we can show that

$$J(y) = \kappa\big(1-y^2\big)^k, \qquad J\big(g^{-1}(x)\big) \leqslant C\big(1-x^2\big)^{k/(k+1)}, \qquad J'\big(g^{-1}(x)\big) \leqslant C\big(1-x^2\big)^{(k-1)/(k+1)}$$

and so on. Thus,

$$\begin{aligned} T_1 \leqslant\ & C + C\varepsilon^{-1}\int_{-1}^{1}\big(\mathrm{e}^{-2\alpha(1-x)/\sqrt{\varepsilon}} + \mathrm{e}^{-2\nu(1+x)/\sqrt{\varepsilon}}\big)\mathrm{d}x \\ & + C\varepsilon^{-2}\int_{-1}^{1}\big(\mathrm{e}^{-2\alpha(1-x)/\sqrt{\varepsilon}} + \mathrm{e}^{-2\nu(1+x)/\sqrt{\varepsilon}}\big)\big(1-x^2\big)^{2k/(k+1)}\,\mathrm{d}x \\ \leqslant\ & C + C\varepsilon^{-1/2} + C\varepsilon^{-3/2+k/(k+1)} \leqslant C\varepsilon^{-3/2+k/(k+1)}. \end{aligned}$$

Similarly, it can be shown that

$$T_2 \leqslant C\varepsilon^{-3/2+(k-1)/(k+1)}.$$

Using Theorem 3.1 we obtain the following error bounds:

$$\begin{cases} \varepsilon\big\|u' - \underline{v}_N'\big\|_{L^2(I)}^2 + \|u - \underline{v}_N\|_{L^2(I)}^2 \leqslant C\big(N^{-2}\varepsilon + N^{-4}\big)\varepsilon^{-3/2+k/(k+1)}, \\ \varepsilon\big\|u' - \underline{v}_N'\big\|_{L^2(I)}^2 + \|u - \underline{v}_N\|_{L^2(I)}^2 \leqslant C\big(N^{-4}\varepsilon + N^{-6}\big)\varepsilon^{-3/2+(k-1)/(k+1)}. \end{cases} \tag{4.8}$$

In Fig. 2, we plot the error bounds in (4.8) for $k = 3$. Without loss of generality, we choose the constant C in the bounds as one.

If we let k be sufficiently large, then the above upper bounds can be made close to

$$\begin{cases} \varepsilon\big\|u' - \underline{v}_N'\big\|_{L^2(I)}^2 + \|u - \underline{v}_N\|_{L^2(I)}^2 \approx \mathrm{O}\big(N^{-2}\varepsilon^{1/2} + N^{-4}\varepsilon^{-1/2}\big), \\ \varepsilon\big\|u' - \underline{v}_N'\big\|_{L^2(I)}^2 + \|u - \underline{v}_N\|_{L^2(I)}^2 \approx \mathrm{O}\big(N^{-4}\varepsilon^{1/2} + N^{-6}\varepsilon^{-1/2}\big). \end{cases} \tag{4.9}$$

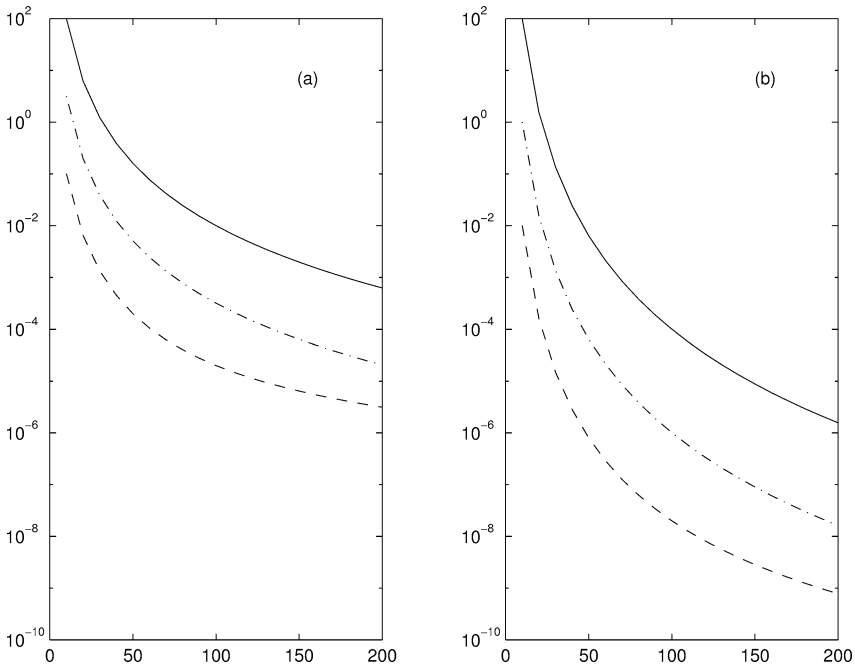

Fig. 2. Error bounds in (4.8) with $k = 3$: (a) is for the first bound, and (b) is for the second bound. The line types are as follows: dashed line for $\varepsilon = 10^{-4}$, dash-dotted line for $\varepsilon = 10^{-6}$ and solid line for $\varepsilon = 10^{-8}$.

In Fig. 3, we plot the error bounds in the right hand side of (4.9). It is observed from Figs. 2 and 3 that by suitably choosing the power k in the transformation (4.7) the spectral accuracy can be obtained with N of the order of about 100 with the present Galerkin-spectral methods.

It can be proved that the above estimates can be generalized to give the following result:

$$\varepsilon\left\|u' - \underline{v}'_N\right\|^2_{L^2(I)} + \|u - \underline{v}_N\|^2_{L^2(I)} \leqslant C(m,k)\left(N^{-2m}\varepsilon + N^{-2m-2}\right)\varepsilon^{-1/2-m/(k+1)}, \tag{4.10}$$

for any $m \geqslant 1$. By analyzing the right hand side of (4.10), we end up with the following result.

Theorem 4.2. *Assume the Galerkin-spectral method* (3.11) *with the transformation* (4.7) *be applied to the Helmholtz type equation* (3.1)*. If the Assumption* 4.1 *holds, then spectral accuracy can be obtained in* $L^2(I)$ *provided that*

$$N \geqslant \varepsilon^{-1/2(k+1)-\delta}, \tag{4.11}$$

where δ *is an arbitrary positive constant.*

Similar curves as given in Fig. 1 can be obtained. The spectral convergence properties will also be confirmed by numerical experiments in Section 6.

The third transformation is the one proposed by Orszag and Israeli [21]:

$$y = g^{-1}(x) = \left(\tan^{-1}\left(\frac{x-1}{\sqrt{\varepsilon}}\right) + \tan^{-1}\left(\frac{x+1}{\sqrt{\varepsilon}}\right)\right)\left(\tan^{-1}\left(\frac{2}{\varepsilon}\right)\right)^{-1}. \tag{4.12}$$

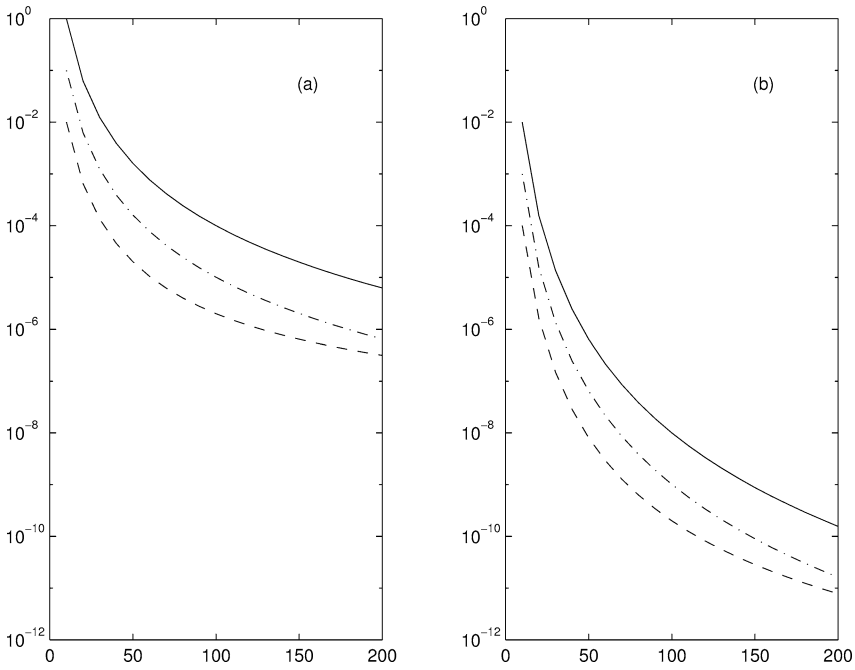

Fig. 3. Error bounds in (4.9): (a) is for the first bound, and (b) is for the second bound. The line types are as follows: dashed line for $\varepsilon = 10^{-4}$, dash-dotted line for $\varepsilon = 10^{-6}$ and solid line for $\varepsilon = 10^{-8}$.

It can be shown that

$$\int_{-1}^{1} (u')^2 \,\mathrm{d}x + \int_{-1}^{1} \underline{J}^2 (u'')^2 \,\mathrm{d}x \leqslant C\varepsilon^{-1}.$$

Therefore, the resulting error estimates are similar to those for the transformation $x = g(y) = \sin((\pi/2)y)$, see (4.4).

Applying conventional Galerkin-spectral methods to (3.1), we can only expect an upper bound of the following, see [4],

$$\varepsilon \|u' - \underline{v}'_N\|^2_{L^2(I)} + \|u - \underline{v}_N\|^2_{L^2(I)} \leqslant C\big(N^{-2}\varepsilon^{-1/2} + N^{-4}\varepsilon^{-3/2}\big). \tag{4.13}$$

This estimate is weaker than (4.4) and (4.8). Hence there is a good improvement in using our Galerkin-spectral methods with appropriate transformations.

Results similar to Theorems 4.1 and 4.2 hold for advection–diffusion equations. As an example, we consider

$$-\varepsilon u''(x) + p(x)u'(x) + q(x)u(x) = f(x,\varepsilon), \quad x \in I, \qquad u(\pm 1) = 0, \tag{4.14}$$

where

$$\begin{cases} -p'(x) + 2q(x) > 0, \quad p(x) \geqslant \alpha > 0, \quad q(x) \geqslant \beta, \qquad \forall\, x \in [-1,1], \\ \alpha^2 + 4\varepsilon\beta > 0. \end{cases} \tag{4.15}$$

It follows from [28] that for any solution u of (4.14)–(4.15)

$$\big|u^{(i)}(x)\big| \leqslant C + C\varepsilon^{-i}\big(\varepsilon^{-\alpha(1-x)/\varepsilon} + \varepsilon^{-\nu(1+x)/\varepsilon}\big), \quad i = 1, 2, \ldots, \tag{4.16}$$

where α, ν are positive constants independent of ε. The bounds in (4.16) imply that the solution of (4.14) has boundary layers of width $O(\varepsilon)$, while (4.1) suggests that width for the solution of Helmholtz equations is $O(\sqrt{\varepsilon})$.

If we use the mapping (4.7), then using (4.16) and a similar procedure for Theorem 4.2 leads to the following error bound:

$$\varepsilon\|u' - \underline{v}'_N\|^2_{L^2(I)} + \|u - \underline{v}_N\|^2_{L^2(I)} \leqslant C(\gamma, k)\big(N^{-2m} + N^{-2(m+1)}\big)\varepsilon^{-\gamma-1-2m/(k+1)}, \tag{4.17}$$

where u is the unique solution of (4.14), v_N is the solution of (3.33), $\gamma > 0$ can be arbitrarily small. The estimate (4.17) leads to the following theorem.

Theorem 4.3. *Assume the Galerkin-spectral method* (3.33) *with the transformation* (4.7) *is applied to the convection–diffusion equation* (4.14). *Then a spectral convergence rate can be observed provided that*

$$N = O\big(\varepsilon^{-1/(k+1)-\delta}\big), \tag{4.18}$$

where δ is an arbitrary positive constant.

This result is weaker than that given by Theorem 4.2, in the sense that to obtain a desired accuracy the Helmholtz equations require less number of grid points. The reason for this is that the width for the Helmholtz equations is of the order $O(\sqrt{\varepsilon})$, but for Eq. (4.14) it is $O(\varepsilon)$.

4.2. General weights for Helmholtz type equations

Let $\omega(y) = (1-y^2)^\lambda$ for a fixed λ with $-1 < \lambda \leqslant 0$. We still consider the transformation (4.7). For any $0 < \gamma < \frac{1}{2}$,

$$\begin{aligned} T_{\tilde{\omega},1} &\leqslant C + \int_{1-\varepsilon^\gamma}^{1} \widetilde{\omega}(u')^2\,dx + \int_{-1}^{-1+\varepsilon^\gamma} \widetilde{\omega}(u')^2\,dx + \int_{-1}^{-1+\varepsilon^\gamma} \widetilde{\omega}\underline{J}^2(u'')^2\,dx + \int_{1-\varepsilon^\gamma}^{1} \widetilde{\omega}\underline{J}^2(u'')^2\,dx \\ &\leqslant C\big(1 + \varepsilon^{\gamma-1-\gamma\lambda/(k+1)} + \varepsilon^{(3+\lambda/(k+1))\gamma-2-2\gamma/(1+k)}\big) \\ &\leqslant C\big(1 + \varepsilon^{3\gamma+\lambda\gamma/(k+1)-2-2\gamma/(1+k)}\big). \end{aligned}$$

Similarly, it can be shown that

$$T_{\tilde{\omega},2} \leqslant C\big(1 + \varepsilon^{5\gamma+\lambda\gamma/(k+1)-3-4\gamma/(1+k)}\big).$$

If γ is chosen close to $\frac{1}{2}$, then an application of Theorem 3.1 gives

$$\begin{cases} \varepsilon\|u' - \underline{v}'_N\|^2_{L^2_{\tilde{\omega}}(I)} + \|u - \underline{v}_N\|^2_{L^2_{\tilde{\omega}}(I)} \leqslant C\big(N^{-2}\varepsilon + N^{-4}\big)\varepsilon^{(\lambda-2)/(2(k+1))-1/2}, \\ \varepsilon\|u' - \underline{v}'_N\|^2_{L^2_{\tilde{\omega}}(I)} + \|u - \underline{v}_N\|^2_{L^2_{\tilde{\omega}}(I)} \leqslant C\big(N^{-4}\varepsilon + N^{-6}\big)\varepsilon^{(\lambda-4)/(2(k+1))-1/2}. \end{cases} \tag{4.19}$$

In general, we have

$$\begin{aligned} \varepsilon\|u' - \underline{v}'_N\|^2_{L^2_{\tilde{\omega}}(I)} + \|u - \underline{v}_N\|^2_{L^2_{\tilde{\omega}}(I)} &\leqslant CN^{-2m}\varepsilon^{2m\gamma+\gamma+\lambda\gamma/(k+1)-m-2m\gamma/(1+k)} \\ &\quad + CN^{-2(m+1)}\varepsilon^{2m\gamma+\gamma+\lambda\gamma/(k+1)-m-1-2m\gamma/(1+k)}. \end{aligned} \tag{4.20}$$

If γ is chosen close to $\frac{1}{2}$, then the right hand side of (4.20) gives

$$C(m,\gamma)\big(N^{-2m}\varepsilon^{1/2(1+(\lambda-2m)/(k+1))}+N^{-2m-2}\varepsilon^{1/2(-1+(\lambda-2m)/(k+1))}\big). \tag{4.21}$$

Theorem 4.4. *Let $\omega(y)=(1-y^2)^{\lambda}$, $\lambda\in(-1,0]$, $\widetilde{\omega}(x)=\omega(g^{-1}(x))$. Let the transformation function $x=g(y)$ be given by* (4.7)*. Assume the Galerkin-spectral method* (3.11) *be applied to the Helmholtz type equation* (3.1)*. If Assumption* 4.1 *holds, then a spectral convergence rate can be obtained in $L^2_{\widetilde{\omega}}(I)$ provided that*

$$N=\mathrm{O}\big(\varepsilon^{-1/2(k+1)-\delta}\big), \tag{4.22}$$

where δ is an arbitrary positive constant.

5. Error bounds for higher dimensions with regular domain

The error analysis in previous sections can be extended to higher dimensions when the solution domain is regular. In this section, we briefly describe the results and the outlines of proof. We here only consider the Helmholtz type equation. Let $A(v,z)_\omega$ and $(F,z)_\omega$ be given in (2.3), i.e.,

$$A(v,z)_\omega=\varepsilon\sum_{i=1}^{d}\int_{\Omega}(a_i^2J)\big(\partial_{y_i}v\partial_{y_i}(z\omega)\big)\mathrm{d}\boldsymbol{y}+\int_{\Omega}Qvz\omega\,\mathrm{d}\boldsymbol{y},$$

$$(F,z)_\omega=\int_{\Omega}Fz\omega\,\mathrm{d}\boldsymbol{y}.$$

Then the weak formulation for the Helmholtz type equation of (2.4) is as follows: find $v\in\widetilde{H}^1_{\omega,0}(\Omega)$ such that

$$A(v,z)_\omega=(F,z)_\omega,\quad \forall z\in\widetilde{H}^1_{\omega,0}(\Omega). \tag{5.1}$$

It will be approximated by the following Galerkin-spectral method: find $v_N\in\boldsymbol{Y}_N$ such that

$$A(v_N,z)_\omega=(F,z)_\omega,\quad \forall z\in\boldsymbol{Y}_N. \tag{5.2}$$

We wish to carry out some error analysis for this scheme. First we need results similar to Lemma 3.1. To this end, we let

$$a(v,z)_\omega=\sum_{i=1}^{d}\int_{\Omega}(a_i^2J)\big(\partial_{y_i}v\partial_{y_i}(z\omega)\big)\mathrm{d}\boldsymbol{y},\qquad c(v,z)=\int_{\Omega}Qvz\omega\,\mathrm{d}\boldsymbol{y}.$$

Lemma 5.1. *Assume that there exist positive constants β, C_1 and C_2 such that for $1\leqslant i\leqslant d$*

$$C_1\leqslant J_i(y_i)\big(1-y_i^2\big)^{-\beta}\leqslant C_2,\quad \frac{J_i\omega_i(a_i\omega_i')'}{(\omega_i')^2}>2,\quad \text{for } y_i\in I:=(-1,1). \tag{5.3}$$

Then for any $z, v\in\widetilde{H}^1_{\omega,0}(\Omega)$,

$$a(v,v)_\omega \geqslant C\sum_{i=1}^{d} \|\partial_{y_i} v\|^2_{L^2_{a_i^2 J\omega}(\Omega)}, \tag{5.4}$$

$$|a(v,z)_\omega| \leqslant C\left(\sum_{i=1}^{d} \|\partial_{y_i} v\|_{L^2_{a_i^2 J\omega}(\Omega)}\right)\left(\sum_{i=1}^{d} \|\partial_{y_i} z\|_{L^2_{a_i^2 J\omega}(\Omega)}\right). \tag{5.5}$$

Proof. The proof follows a similar idea used in Section 3.1. For the case $d=2$, for instance, we only need to note that

$$\begin{aligned} a(v,z) = {} & \int_I J_2(y_2)\left(\int_I a_1 \partial_{y_1} v \partial_{y_1}\big(z\omega_1(y_1)\big)\,\mathrm{d}y_1\right)\omega_2(y_2)\,\mathrm{d}y_2 \\ & + \int_I J_1(y_1)\left(\int_I a_2 \partial_{y_2} v \partial_{y_2}\big(z\omega_2(y_2)\big)\,\mathrm{d}y_2\right)\omega_1(y_1)\,\mathrm{d}y_1 + \int_I\int_I Q v z \omega\,\mathrm{d}\boldsymbol{y}. \end{aligned}$$

Then applying Lemma 3.1 will give (5.4) and (5.5). □

For any $\phi \in \widetilde{H}^1_{\omega,0}(\Omega)$, we define $\Pi\phi \in \boldsymbol{Y}_N$ such that

$$a(h, \phi - \Pi\phi) = 0, \quad \forall h \in \boldsymbol{Y}_N. \tag{5.6}$$

It is clear that $\Pi\phi$ is uniquely defined in $\boldsymbol{Y}_N$. Now we are able to extend the error analysis in Section 3 to the high-dimensional case:

Theorem 5.1. *Let $u(\boldsymbol{x})$ be the unique solution of* (5.1) *and $v_N(\boldsymbol{y})$ be the unique solution of* (5.2). *Assume that J_i satisfies* (5.3). *Then the following estimate holds*:

$$\begin{aligned} & \varepsilon\big\|\nabla u - \nabla \underline{v}_N\big\|^2_{L^2_{\widetilde{\omega}}(\Omega)^d} + \|u - \underline{v}_N\|^2_{L^2_{\widetilde{\omega}}(\Omega)} \\ & \quad \leqslant C\big(N^{-2}\varepsilon + N^{-4}\big)\left(\sum_{i=1}^{d}\int_\Omega \widetilde{\omega}\left(\frac{\partial u}{\partial x_i}\right)^2 \mathrm{d}\boldsymbol{x} + \sum_{i=1}^{d}\int_\Omega \widetilde{\omega}\underline{J}_i^2\left(\frac{\partial^2 u}{\partial^2 x_i}\right)^2 \mathrm{d}\boldsymbol{x}\right), \end{aligned} \tag{5.7}$$

where $\widetilde{\omega}(\boldsymbol{x}) = \omega(\boldsymbol{g}^{-1}(\boldsymbol{x}))$, $\underline{v}_N(\boldsymbol{x}) = v_N(\boldsymbol{g}^{-1}(\boldsymbol{x}))$, $\underline{J}_i(x_i) = J_i(g^{-1}(x_i))$, $\underline{J}_i'(x_i) = J_i'(g^{-1}(x_i))$.

Proof. The following results are similar to those in Lemmas 3.2 and 3.3: for any $v \in \widetilde{H}^1_{\omega,0}(\Omega)$

$$a(v - \Pi v, v - \Pi v) \leqslant CN^{-2}\left(\sum_{i=1}^{d}\int_\Omega \widetilde{\omega}\left(\frac{\partial u}{\partial x_i}\right)^2 \mathrm{d}\boldsymbol{x} + \sum_{i=1}^{d}\int_\Omega \widetilde{\omega}\underline{J}_i^2\left(\frac{\partial^2 u}{\partial^2 x_i}\right)^2 \mathrm{d}\boldsymbol{x}\right), \tag{5.8}$$

$$c(v - \Pi v, v - \Pi v) \leqslant CN^{-4}\left(\sum_{i=1}^{d}\int_\Omega \widetilde{\omega}\left(\frac{\partial u}{\partial x_i}\right)^2 \mathrm{d}\boldsymbol{x} + \sum_{i=1}^{d}\int_\Omega \widetilde{\omega}\underline{J}_i^2\left(\frac{\partial^2 u}{\partial^2 x_i}\right)^2 \mathrm{d}\boldsymbol{x}\right). \tag{5.9}$$

We briefly outline the proof for (5.8). It follows from Lemma 5.1 that

$$a(v - \Pi v, v - \Pi v) \leqslant a\big(v - \Pi^* v, v - \Pi^* v\big), \quad \forall v \in \widetilde{H}^1_{\omega,0}(\Omega),$$

where

$$\Pi^* v = \left(\prod_{i=1}^{d} \otimes \Pi_i^*\right) v,$$

and Π_i^* is defined by (3.16). Then using the same techniques as that in Lemma 3.2 and Theorem 4.4 of [2], we can obtain (5.8). The estimate (5.9) can be established by following a similar procedure. Using (5.8)–(5.9) and the techniques used in the proof of Theorem 3.1 will lead to (5.7). □

However, in general, the solutions of higher dimensional problems have much richer structures, e.g., parabolic boundary layers, which do not exist in the 1-D case. Therefore the explicit ε-energy error bounds obtained for the 1-D problems may only hold for simpler problems of a higher dimension with a single exponential boundary layer. Also the approach used in this section only works for a tensor-product type of higher dimensional problems. Much more research is still needed to deal with perturbation problems in higher dimensions.

6. Numerical experiments

In this section, we consider several numerical examples by using the Galerkin-spectral method (2.9). Thus Eq. (1.1) is first transferred into (1.3) via a suitable transformation, and then approximated using the basis functions in $\boldsymbol{Y}_N$. Then the resulting linear systems from (2.9) are solved. All of the computations are based on the transformation (2.11) with $k = 1$. In order to demonstrate the high accuracy of the proposed method in this work, we make some comparisons with the conventional Legendre–Galerkin methods, Chebyshev–Galerkin methods, and the boundary layer resolving Chebyshev-collocation method proposed in [29].

6.1. When to stop computation?

Once a coordinate transformation is chosen, we need to know when to stop the computation. *In particular, we wish to observe the spectral convergence rate*. To this end, the numerical procedure is proposed as follows.

- *Step* 1. Choose a set of points $\boldsymbol{S}$ very close to the boundaries. For example, if the boundary layer is near right boundary $x = 1$ then an example of the set is

$$\boldsymbol{S} = \{0.98 + j \cdot 0.001 \mid 0 \leqslant j \leqslant 20\}. \tag{6.1}$$

  Choose a starting number of the basis functions, $N = N_0$, and perform a computation to obtain numerical solution $\{U_j\}$.
- *Step* 2. *Interpolation.* After the approximations $\{U_j\}_{j=0}^N$ are obtained, we use the collocation idea to obtain the coefficients $\{a_j\}_{j=0}^N$ in the following expression:

$$U(x) = \sum_{j=0}^{N} a_j T_k(x), \tag{6.2}$$

  where T_k are Chebyshev/Legendre polynomials. Then use the above expression to obtain approximate solutions on the set $\boldsymbol{S}$. The interpolation values with N_0 basis functions is denoted by S_0.
- *Step* 3. For $l \geqslant 1$, use $N_l = 2^l \cdot N_0$ basis functions to obtain the numerical interpolations S_l on the set $\boldsymbol{S}$. Then compute the differences between S_l and S_{l-1}. If the difference is less than a given tolerance, then the computation is stopped.

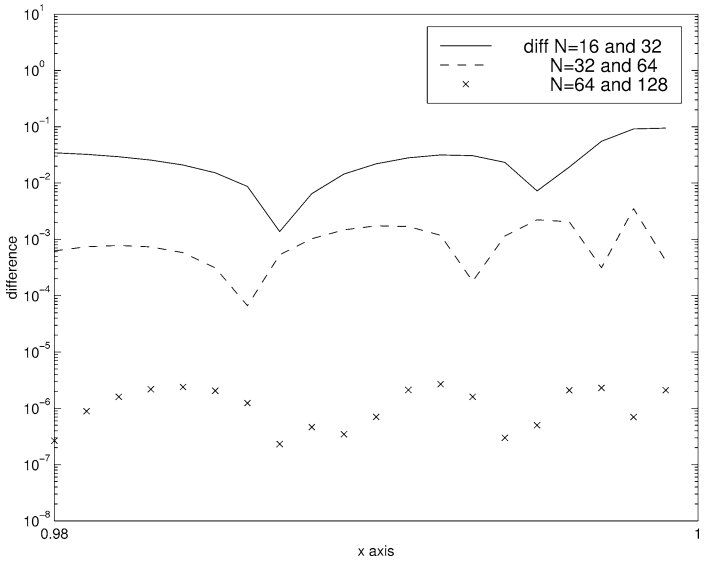

Fig. 4. The differences between consecutive grids for Example 6.1 with $\varepsilon = 10^{-6}$.

- *Step* 4. Output the numerical results obtained with the largest $N = N_l$.

Example 6.1. To illustrate the above ideas, we consider the following problem:

$$\begin{cases} -\varepsilon u''(x) + (1-x)u'(x) = f(x), \\ u(0) = \exp\left(-\dfrac{1}{\sqrt{\varepsilon}}\right), \qquad u(1) = \sin 1 + 1, \end{cases} \tag{6.3}$$

where f is chosen such that the exact solution is $u(x) = \sin x + \exp(-(1-x)/\sqrt{\varepsilon})$.

This problem does not correspond with one for which the error estimates have been derived. It is a problem with a turning point at the boundaries. One of the reasons for choosing this problem is to demonstrate numerically that our Galerkin-spectral methods can handle many types of problems. We will show that based on the above four steps the convergent solutions can be obtained. Moreover, the exponential rate of convergence will be observed without using the information of exact solution.

The procedure for solving Example 6.1 is the following. In Step 1, the transformation used is based on the transformation (2.11) with $k = 1$. The set of points for interpolation is the one given by (6.1). The starting number of basis functions is chosen as $N_0 = 16$. The numerical interpolation on the set $\boldsymbol{S}$ is performed as described in Step 2. In Step 3, three finer meshes are employed. The difference between S_l and S_{l-1}, defined by

$$d_l(s) := \left|U^{(l)}(s) - U^{(l-1)}(s)\right|$$

is computed over the boundary layer set $\boldsymbol{S}$.

We first demonstrate numerical results for $\varepsilon = 10^{-6}$. In Fig. 4, the differences between S_l and S_{l-1}, with $l = 1, 2$ and 3 are plotted, which suggests *an exponential rate of convergence* for the Galerkin-spectral methods. The numerical approximations to the unknown function u with $N = 32$ and 64 are

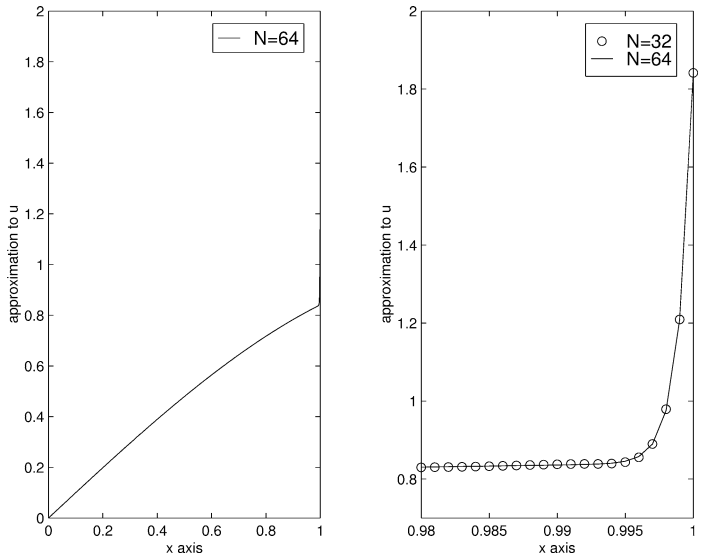

Fig. 5. The numerical approximations with 32 and 64 basis functions for Example 6.1 with $\varepsilon = 10^{-6}$.

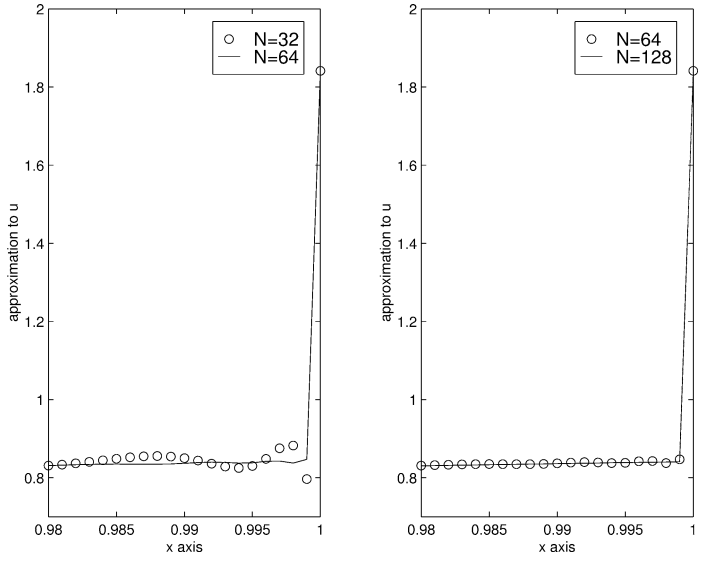

Fig. 6. The numerical approximations over the boundary point set (6.1) with 32, 64 and 128 basis functions, for Example 6.1 with $\varepsilon = 10^{-10}$.

plotted in Fig. 5. It is found that the solution curves obtained by using $N = 32$ and 64 are graphically indistinguishable. For $\varepsilon = 10^{-10}$, we repeated the previous procedure, except choosing $N_0 = 32$. Over the boundary layer set (6.1), the numerical approximations to the unknown function u with $N = 32, 64$ and 128 are plotted in Fig. 6. The agreement of solution curves obtained by using $N = 32$ and 64 seems unsatisfactory, while the curves with $N = 64$ and 128 are in good agreement.

6.2. More examples

In this subsection, we consider three examples to verify the theoretical error estimates obtained in this work. We note that for the transformation (2.11) with $k = 1$, the highest degree of Legendre polynomials in both $\boldsymbol{X}_N$ and $\boldsymbol{Y}_{N-3}$ is N. Hence, we shall compare the conventional Legendre–Galerkin method in $\boldsymbol{X}_N$ with the new Legendre–Galerkin method in $\boldsymbol{Y}_{N-3}$. Note, however, that $\boldsymbol{X}_N$ is an $(N-1)^d$-dimensional space, while $\boldsymbol{Y}_{N-3}$ is an $(N-3)^d$-dimensional space.

Let $\mathcal{M}_N$ be the set of the Legendre–Gauss–Lobatto collocation points with respect to $\boldsymbol{X}_N$. For all the examples considered below, we compute

$$\|u - u_N\|_{l^\infty} \equiv \max_{\boldsymbol{y}\in\mathcal{M}_N} \big|u\big(\boldsymbol{g}(\boldsymbol{y})\big) - u_N\big(\boldsymbol{g}(\boldsymbol{y})\big)\big|,$$

and

$$\|v - v_{N-3}\|_{l^\infty} \equiv \max_{\boldsymbol{y}\in\mathcal{M}_N} \big|v(\boldsymbol{y}) - v_{N-3}(\boldsymbol{y})\big|,$$

where $v = u(\boldsymbol{g}(\boldsymbol{y}))$, u_N and v_{N-3} are respectively the solution of the conventional Legendre–Galerkin scheme and the new Legendre–Galerkin scheme. It is clear that the collocation points in the $\boldsymbol{y}$ variable(s) in $\mathcal{M}_N$ are well condensed near the boundary for the $\boldsymbol{x}$ variables, though we are aware that these ε-independent discrete max-norms may not truly resolve very thin boundary layers.

Example 6.2. Our first example is the one-dimensional diffusion equation

$$-\varepsilon u_{xx} + u = -\frac{x+1}{2}, \quad x \in I, \qquad u(\pm 1) = 0, \tag{6.4}$$

with the exact solution

$$u(x) = \frac{\sinh((x+1)/\sqrt{\varepsilon})}{\sinh(2/\sqrt{\varepsilon})} - \frac{x+1}{2}.$$

The solution has a boundary layer at $x = 1$ of width $\mathrm{O}(\sqrt{\varepsilon})$.

In Table 1, we list the maximum pointwise error obtained by using the conventional Chebyshev–Galerkin method (CCGM), our proposed Legendre–Galerkin method (PLGM) (i.e., $\omega(y) = 1$ in (2.9)) and Chebyshev–Galerkin method (PCGM) (i.e., $\omega(y) = (1-y^2)^{-1/2}$ in (2.9)). This example is of Helmholtz type and the results in Section 4.1 should apply. In fact, the numerical results confirm that the Galerkin-spectral method (2.9) is more efficient than the conventional ones for solving problems with boundary layers. Since $k = 1$, it follows from Theorem 4.2 that spectral accuracy can be obtained for $N \geqslant \varepsilon^{-1/4-\delta}$ with $\delta > 0$. For the values of ε used in Table 1, the corresponding values of $\varepsilon^{-1/4}$ are 100, 178, 316, 562 and 1000, respectively. The errors listed in Table 1 confirm the theoretical prediction.

Example 6.3. The second example is the one-dimensional convection equation

$$-\varepsilon u_{xx} + u_x = -\frac{1}{2}, \quad x \in I, \qquad u(\pm 1) = 0, \tag{6.5}$$

with the exact solution

$$u(x) = \frac{\exp((x+1)/\varepsilon) - 1}{\exp(2/\varepsilon) - 1} - \frac{x+1}{2}.$$

Table 1
Maximum pointwise errors for Example 6.2

	N	$\varepsilon = 10^{-8}$	$\varepsilon = 10^{-9}$	$\varepsilon = 10^{-10}$	$\varepsilon = 10^{-11}$	$\varepsilon = 10^{-12}$
CCGM	512	7.9E−8	6.4E−4	4.6E−2	6.7E−1	
	1024		6.5E−8	2.4E−4	3.5E−2	
	2048			6.2E−6	7.2E−5	6.4E−3
PLGM	64	4.5E−3	3.9E−2	1.4E−1		
	128	1.3E−5	4.5E−4	4.7E−3	2.3E−2	6.4E−2
	256	3.0E−12	6.6E−9	2.2E−6	1.1E−4	1.5E−3
PCGM	64	6.7E−4	4.9E−3	3.6E−2		
	128	1.8E−6	7.5E−5	1.7E−4	3.8E−3	7.6E−3
	256	1.0E−12	1.1E−10	7.2E−7	5.4E−5	4.1E−4

Table 2
Maximum pointwise errors for Example 6.3

	N	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-5}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-7}$	$\varepsilon = 10^{-8}$
CLGM	512	2.2E−7	1.8E−1			
	1024	1.3E−9	9.5E−4	8.9E−1		
	2048		4.2E−8	5.1E−2	O(1)	
PLGM	64	3.2E−3	1.6E−1			
	128	9.7E−6	3.8E−3	3.3E−2	O(1)	
	256	2.1E−12	1.4E−6	1.6E−3	2.7E−2	O(1)
	512		6.85E−12	2.4E−7	5.1E−4	3.5E−2
BLRCC	64	O(1.0E−2)				
	128	O(1.0E−5)		O(1.0E−1)		
	256	O(1.0E−12)		O(1.0E−3)		

The solution has a boundary layer at $x = 1$ of width $O(\varepsilon)$.

In Table 2 we list the maximum pointwise error obtained by using the conventional Legendre–Galerkin method (CLGM) and the Legendre–Galerkin method (2.9) (PLGM). Table 2 also includes the results given by the boundary layer resolving Chebyshev-collocation method (BLRCC, cf. [29]). Again it is observed that the Galerkin-spectral method proposed in this work is much more accurate than the conventional spectral method. It is also noted that the PLGM is even more accurate than the BLRCC.

Table 3
Maximum pointwise errors for Example 6.4; N is the number of the basis functions used in each coordinate direction

	N	$\varepsilon = 10^{-8}$	$\varepsilon = 10^{-9}$	$\varepsilon = 10^{-10}$	$\varepsilon = 10^{-11}$
CLGM	128	3.8E−1			
	256	2.1E−2	2.7E−1		
	512	6.6E−7	8.0E−3	1.9E−1	
PLGM	64	9.9E−3	5.7E−2		
	128	2.6E−5	6.2E−4	7.2E−3	2.5E−2
	256	3.5E−11	1.5E−8	4.8E−6	2.4E−4

However, it is clear that more points are required for this problem than the previous example. This confirms the theoretical predictions in Theorems 4.2 and 4.3.

It should be pointed out that for a fixed N, the computational complexities of the conventional methods and the methods proposed in this work can be made essentially the same (see [25,26]).

Example 6.4. The last example is a two-dimensional diffusion equation:

$$\begin{cases} -\varepsilon \Delta u + 2u = F, \quad (x_1, x_2) \in \Omega = I^2; \qquad u|_{\partial\Omega} = 0, \\ F(x_1, x_2) = -\frac{1}{2}\big((x_1+1)w(x_2) + (x_2+1)w(x_1)\big) \end{cases} \tag{6.6}$$

with

$$w(x) = \frac{\sinh((x+1)/\sqrt{\varepsilon})}{\sinh(2/\sqrt{\varepsilon})} - \frac{x+1}{2}.$$

This equation has the exact solution $u(x_1, x_2) = w(x_1)w(x_2)$ which has boundary layers of width $O(\sqrt{\varepsilon})$ at $(x_1 = 1, x_2)$ and $(x_1, x_2 = 1)$.

In Table 3, we list the maximum pointwise error obtained by the CLGM and PLGM. Observations similar to those for the 1-D problem, Example 6.2, are made for this 2-D example. The computational complexities of our Galerkin methods (2.9) and the conventional methods are also essentially the same in 2-D.

The errors listed in Tables 1–3 are measured in a *discrete* maximum norm. It would be useful to see what happens to the error away from the points at which the error is sampled in the table. To this end, we plot in Figs. 7 and 8 the numerical error of Example 6.2 with $\varepsilon = 10^{-8}$ and $N = 128$, and the error of Example 6.3 with $\varepsilon = 10^{-5}$ and $N = 256$, respectively. In these two figures, we choose the following set of points

$$\boldsymbol{S} = \{-1 + 0.01j \mid 0 \leqslant j \leqslant 200\}$$

and use the idea of interpolations as described in Step 2 of the last subsection to obtain approximate values on the above set of points. Fig. 9 shows the error of Example 6.4 with $\varepsilon = 10^{-8}$ and $N = 128$. The interpolation values used for the MATLAB plot are obtained on the following point set

$$\boldsymbol{S} = \big\{(x_i, y_j) = (-1 + 0.04i, -1 + 0.04j) \mid 0 \leqslant i, j \leqslant 50\big\}.$$

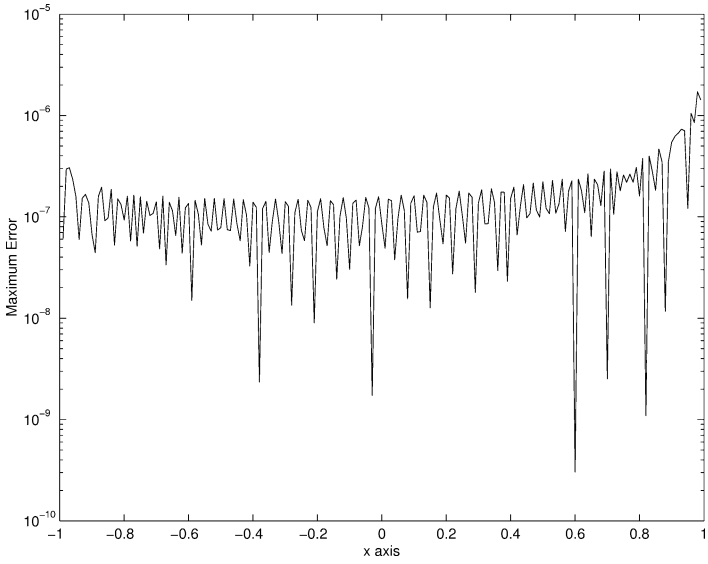

Fig. 7. Error of Example 6.2 for $\varepsilon = 10^{-8}$ and $N = 128$, obtained by using the Legendre–Galerkin method (2.9).

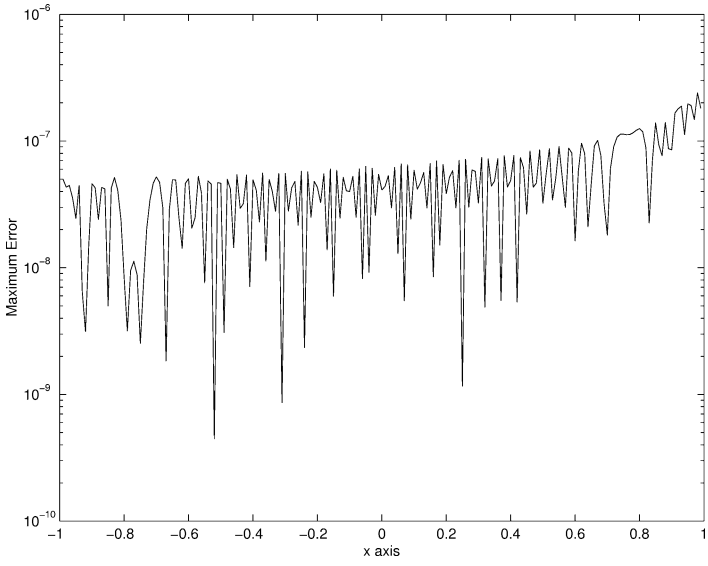

Fig. 8. Error of Example 6.3 for $\varepsilon = 10^{-5}$ and $N = 256$, obtained by using the Legendre–Galerkin method (2.9).

It may not come as a surprise to find the major portion of the errors in the three figures all located near the boundaries, and as a result it would appear natural to move more points into the boundary layers.

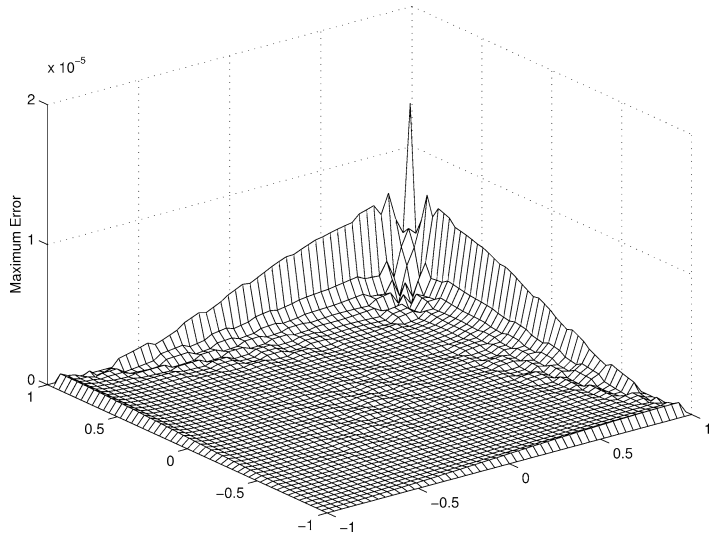

Fig. 9. Error of Example 6.4 for $\varepsilon = 10^{-8}$ and $N = 128$, obtained by using the Legendre–Galerkin method (2.9).

7. Concluding remarks

In this section, we make a number of remarks on the numerical methods and the theoretical results obtained in this work.

1. The proposed Galerkin-spectral methods involve some suitable transformation functions. In practical computations, these functions may also involve a free parameter. One family of such transformations is

$$x = g_k(y) = -1 + \kappa \int_{-1}^{y} (1-\eta^2)^k \,\mathrm{d}\eta, \qquad k \geqslant 1 \quad \text{and} \quad \kappa = \frac{2}{\int_{-1}^{1}(1-\eta^2)^k \,\mathrm{d}\eta},$$

which is exactly (4.7). Another family is proposed in [29] that is of form $x = g_k(y)$, where

$$g_0(y) = y,$$
$$g_k(y) = \sin\left(\frac{\pi}{2} g_{k-1}(y)\right), \quad k \geqslant 1.$$

Numerical experiments suggest that $k = 1$ or 2 is sufficient to speed up the spectral convergence.

2. Although it may be possible to fully resolve the boundary layer for a specified problem by applying a suitable stretching transformation in theory, in practice it is essentially impossible to resolve *arbitrarily* thin boundary layers with a non-adaptive ε-independent coordinate stretching. Consequently, the methods studied in this paper are *not* appropriate if one is interested in *details* of *arbitrarily* thin boundary layers. To fully resolve an arbitrarily small boundary layers, ε-uniform meshes such as Shishkin's grid [27] and Schwab and Suri's grid [24] should be employed.
3. On the other hand, spectral methods have the advantage that if N is reasonably large, then *exponential rate of convergence* can be obtained. For singularly perturbed problems, very large (unpractical) N is required in order to gain this spectral accuracy. The goal of this work is to show

that spectral methods plus coordinate stretching allow us to use reasonably large number of basis functions to gain the exponential rate of convergence. In solving singularly perturbed problems with conventional spectral methods, spectral accuracy can only be obtained when $N = \mathrm{O}(\varepsilon^{-\gamma})$. Our main effort is to make this γ smaller, say from $\frac{1}{2}$ to $\frac{1}{4}$ or less for Helmholtz type equations, by using appropriate coordinate stretching. If one is interested in seeing the exponential rate of convergence with practical number of basis functions then the Galerkin-spectral methods studied in this work should be one of the good choices.

4. Another objective of this work is to give a theoretical interpretation of the high accuracy behavior of the Galerkin-spectral method involving coordinate transformations. The error analysis is quite difficult, partly because the transformed equation is highly degenerate. Ideally, the error bounds should be derived in maximum-norms rather than the present energy-norms that may not truly resolve very thin boundary layers. However, there are still some technical difficulties in obtaining the error estimates with max-norm. Therefore the relation between the proved and the observed facts could be purely intuitive.

Acknowledgements

The authors would like to thank Dr. Ningning Yan of the Chinese Academy of Sciences for providing part of the numerical results. Thanks also to the referees for valuable suggestions which lead to an improved presentation of this paper. The research of the second author was supported by RGC Grant of the Hong Kong Research Grants Council and FRG Grant of the Hong Kong Baptist University.

References

[1] U. Asher, R.M. Mattheij, R.D. Russell, Numerical Solution of Boundary Value Problems for Ordinary Differential Equation, Prentice-Hall, Englewood Cliffs, NJ, 1988.
[2] C. Bernardi, Y. Maday, Properties of some weighted Sobolev spaces and applications to spectral approximations, SIAM J. Numer. Anal. 26 (1989) 769–829.
[3] J.P. Boyd, Chebyshev and Fourier Spectral Methods, Lecture Notes in Engineering, Springer, Berlin, 1989.
[4] C. Canuto, Spectral methods and a maximum principle, Math. Comp. 51 (1988) 615–629.
[5] C. Canuto, M.Y. Hussaini, A. Quarteroni, T. Zang, Spectral Methods in Fluid Dynamics, Series of Computational Physics, Springer, Berlin, 1988.
[6] J.R. Cash, M.H. Wright, A deferred correction method for nonlinear two-point boundary value problems, SIAM. J. Sci. Comput. 12 (1991) 971–989.
[7] H. Eisen, W. Heinrichs, A new method of stabilization for singular perturbation problems with spectral methods, SIAM J. Numer. Anal. 29 (1992) 107–122.
[8] L. Greengard, Spectral integration and two-point boundary value problem, SIAM J. Numer. Anal. 28 (1991) 1071–1080.
[9] W.-Z. Huang, D.M. Sloan, A new pseudospectral method with upwind features, IMA J. Numer. Anal. 13 (1993) 413–430.
[10] E. Kalinay de Rivas, On the use of nonlinear grids in finite-difference equations, J. Comput. Phys. 10 (1972) 202–210.
[11] R.B. Kellogg, A. Tsan, Analysis of some difference approximations for a singular perturbation problem without turning points, Math. Comp. 32 (1978) 1025–1039.

[12] H.-O. Kreiss, N.K. Nichols, D. Brown, Numerical methods for stiff two-point boundary value problems, SIAM J. Numer. Anal. 23 (1986) 325–368.
[13] J.-Y. Lee, L. Greengard, A fast adaptive numerical method for stiff two-point boundary value problems, SIAM J. Sci. Comput. 18 (1997) 403–429.
[14] M. Lentini, V. Peyrera, An adaptive finite difference solver for nonlinear two-point boundary problems with mild boundary layers, SIAM J. Numer. Anal. 14 (1977) 91–111.
[15] W.B. Liu, J. Shen, A new efficient spectral Galerkin methods for singular perturbation problems, J. Sci. Comput. 11 (1996) 130–145.
[16] W.B. Liu, T. Tang, A new boundary layer resolving spectral method, in: AMS Proceedings of the Conference on Mathematics of Computation 1943–1993, AMS, 1993.
[17] Y.Y. Liu, The pseudospectral Chebyshev method for two-point boundary value problems, M.Sc. thesis, Department of Mathematics and Statistics, Simon Fraser University, Burnaby, B.C., Canada, 1992.
[18] R.M.M. Mattheij, G.W. Staarink, An efficient algorithm for solving general linear two point BVP, SIAM J. Sci. Statist. Comput. 5 (1984) 745–763.
[19] J.J.H. Miller, R. O'Riordan, G.I. Shishkin, Fitted Numerical Methods for Singular Perturbation Problems, World Scientific, Singapore, 1996.
[20] L.S. Mulholland, W.-Z. Huang, D.M. Sloan, Pseudospectral solutions of near-singular problems using numerical coordinate transformations based on adaptivity, SIAM J. Sci. Comput. 19 (1998) 1261–1289.
[21] S.A. Orszag, M. Israeli, Numerical simulation of viscous incompressible flows, Ann. Rev. Fluid Mech. 6 (1974) 281–318.
[22] H.-G. Roos, M. Stynes, L. Tobiska, Numerical Methods for Singularly Perturbed Differential Equations, Springer, Berlin, 1996.
[23] A.H. Schatz, L.B. Wahlbin, On the finite element method for singularly perturbed reaction–diffusion problems in two and one dimensions, Math. Comp. 40 (1983) 47–89.
[24] C. Schwab, M. Suri, The p and hp versions of the finite element method for problems with boundary layers, Math. Comp. 65 (1996) 1403–1429.
[25] J. Shen, Efficient spectral-Galerkin method I: Direct solvers for the Helmholtz equation and the biharmonic equation using Legendre polynomials, SIAM J. Sci. Comput. 15 (1994) 1489–1505.
[26] J. Shen, Efficient spectral-Galerkin method II: Direct solvers of second- and fourth-order equations using Chebyshev polynomials, SIAM J. Sci. Comput. 16 (1995) 74–87.
[27] G.I. Shishkin, Grid approximation of singularly perturbed elliptic and parabolic equations, Second Doctoral Thesis, Keldysh Institute of Applied Mathematics, USSR Academy of Sciences, Moscow, 1990.
[28] M. Stynes, E. O'Riordan, A finite element method for a singularly perturbed boundary value problem, Numer. Math. 50 (1986) 1–15.
[29] T. Tang, M.R. Trummer, Boundary layer resolving pseudospectral method for singular perturbation problems, SIAM J. Sci. Comput. 17 (1996) 430–438.