Journal of Scientific Computing, Vol. 11, No. 4, 1996

A New Efficient Spectral Galerkin Method for Singular Perturbation Problems

W. B. Liu[1] and Jie Shen[2]

Received June 15, 1995

A new spectral Galerkin method is proposed for the convection-dominated convection-diffusion equation. This method employs a new class of trail function spaces. The available error bounds provide a clear theoretical interpretation for the higher accuracy of the new method compared to the conventional spectral methods when applied to problems with thin boundary layers. Efficient solution techniques are developed for the convection-diffusion equations by using appropriate basis functions for the new trial function spaces. The higher accuracy and the effectiveness of the new method for problems with thin boundary layers are confirmed by our numerical experiments.

KEY WORDS: Spectral-Galerkin method; boundary layer; singular perturbation; convection-diffusion.

1. INTRODUCTION

Many physical processes possess very thin boundary layers within which some concerned physical quantities vary sharply. The presence of thin boundary layers introduces a serious difficulty for their numerical simulations. Conventional numerical schemes, e.g., conventional spectral methods, finite element methods or central difference methods, usually suffer from numerical instability and/or unphysical oscillation when applied to a reasonably accurate mathematical model of such processes. Among the well studied mathematical models is the convection-diffusion equation:

$$-\varepsilon\, \Delta u(\mathbf{x}) + \nabla u(\mathbf{x}) \cdot \mathbf{p}(\mathbf{x}) + q(\mathbf{x})\, u(\mathbf{x}) = f(\mathbf{x}, \varepsilon), \qquad \text{in } \Omega \tag{1.1}$$

[1] Institute of Mathematics and Statistics, University of Kent, Canterbury, CT2 7NF, United Kingdom.

[2] Department of Mathematics, Penn State University, University Park, Pennsylvania 16802. The work of this author is partially supported by NSF grant DMS-9205300.

0885-7474/96/1200-0411$09.50/0

where Ω is a domain in R^d with $d=1, 2$ or 3, $\varepsilon>0$ is a fixed constant. We assume for convenience that $\mathbf{p}=(p_1,..., p_d)^T$, q and f are smooth functions on $\bar{\Omega}$ and $\|f(\cdot, \varepsilon)\|_{L^\infty(\Omega)}$ is bounded by a constant independent of ε. In this paper, we restrict ourselves to the cases $\Omega=(-1, 1)^d$ and only the homogeneous Dirichlet boundary condition for u is considered.

In many applications, Eq. (1.1) possesses boundary layers of width $O(\varepsilon^\gamma)$, where γ is a positive constant. When the parameter ε is very small, it is well known that the central difference schemes and conventional finite element schemes suffer from unphysical oscillation when applied to Eq. (1.1), unless very fine meshes are used. For the conventional spectral methods (cf. Gottlieb and Orszag (1977); Canuto *et al.* (1988): or Funaro (1992) for a general introduction of the spectral method), very large N (N is the number of modes of the approximate solution in each direction) is required to get acceptable resolution of the boundary layer. This causes various computational problems. For instance, the pseudo-spectral methods with large N lead to severely ill-conditioned systems, resulting a significant loss in precision. In fact, when the problem possesses a boundary layer of width $O(\varepsilon)$ with $\varepsilon \ll 1$ (e.g., $\varepsilon < 10^{-6}$), high accuracy cannot be expected by using the conventional pseudo-spectral method [cf. Eisen and Heinrichs (1992)]. There have been many attempts in searching suitable schemes for this problem. For instance, the adaptive finite element or finite difference method (cf. Ascher *et al.* (1979)], up-wind finite difference method [cf. Hughes (ed.) (1979)], the boundary layer resolving spectral methods [BLRSMs, cf. Orszag and Israeli (1974); Tang and Trummer (1993)] and others [Kalinay De Rivas (1972); Maclcenzie and Morton (1990); Oriordan and Stynes (1991)] have been successfully applied to the Eq. (1.1) in various cases. We shall focus our attention to the spectral methods. It is observed that the BLRSMs can handle very thin boundary layers and give very accurate results when the solutions are smooth [cf. Orszag and Israeli (1974); Liu and Tang (1994b); and Tang and Trummer (1996)]. The key to the success of the BLRSMs is to apply suitable transformations to the approximate equations before discretizing them with global polynomials as the trial functions. However, the transformed equations are usually rather complicated with degenerate coefficients even when the original equations are very simple. In fact, let $x_i=g_i(y_i)$ with $g_i \in C^\infty[-1, 1]$ such that

$$g_i(-1)=-1, \quad g_i(1)=1, \quad g_i'(y_i)>0, \qquad \text{for} \quad y_i \in (-1, 1) \text{ and } i=1,..., d$$

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

$$-\varepsilon \sum_{i=1}^{d} a_i \partial_{y_i}(a_i \partial_{y_i} v) + \sum_{i=1}^{d} a_i P_i \partial_{y_i} v + q(\mathbf{g}(\mathbf{y}))\, v = f(\mathbf{g}(\mathbf{y}), \varepsilon), \qquad \text{in } \Omega \quad (1.2)$$

where

$$v(\mathbf{y}) = u \circ \mathbf{g}(\mathbf{y}), \qquad a_i(y_i) = 1/g_i'(y_i), \qquad P_i(\mathbf{y}) = p_i \circ \mathbf{g}(\mathbf{y}), \qquad i = 1,..., d$$

In order to obtain a finer resolution near the boundary, it is 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. Therefore, the transformed equation has unbounded coefficients even when the coefficients of the original equation are constants. This causes several major difficulties to the analysis and to the implementation of these methods. For instance, it is very difficult to carry out a theoretical analysis for such schemes due to the degenerate character of the transformed equations. On the other hand, the complexity of the transformed equations also increases considerably the difficulty of the implementation.

One of the essential questions is what is the gain in accuracy by using the BLRSMs compared with the conventional spectral methods. This question has recently been-addressed in [Liu and Tang (1994a,b)]. Another important question which we would like to address here is how to efficiently solve the transformed Eq. (1.2). The existing procedures [cf. Orszag and Israeli (1974); and Tang and Trummer (1996)] are based on applying the conventional pseudo-spectral methods directly to the transformed equations. Therefore, to generate the discretized matrix system, one first needs to evaluate some unbounded functions at the collocation points including those near or at the boundary. This introduces extra computational difficulties and appreciable roundoff errors when the number of collocation points N in each direction is large. Furthermore, one needs to solve an ill-conditioned linear algebraic system with a full matrix. Therefore, this type of implementations is not efficient.

The main purpose of this paper is to address the difficulty of the implementation. We here propose a new spectral-Galerkin method which uses a new trial function space. In addition to the ability of resolving very thin boundary layers, the resulting linear system can also be efficiently solved in many notable cases. More precisely, when p_i $(i = 1,..., d)$ and q are constants, the resulting linear system has sparse matrix which can be efficiently inverted by a direct method. Therefore very efficient and accurate direct solvers can be developed for Eq. (1.1) in this case. A remarkable fact is that the computational complexity of the new spectral-Galerkin method is essentially the same as that l: of the very efficient conventional spectral-Galerkin method developed by Shen (1994). In other words, the ability of resolving much thinner boundary layers does not introduce extra computational expenses. Variable-coefficient or nonlinear problems can be dealt with using an iterative method with a suitable constant-coefficient problem

as the preconditioner or subdomain solver. More specifically, in solving Eq. (1.1) on a complex geometry by the domain decomposition methods or fictitious domain methods, it is essential to have a fast and highly accurate solver on a rectangular domain. Since this solver will be repeatedly used in the iterative process, the efficiency and accuracy of those methods will largely depend on that of the solver. The method developed here can also be used in solving time-dependent problems with thin boundary layers such as Navier-Stokes equations with high Reynolds number, since at each time step an equation like the form of Eq. (1.1) need to be solved.

Theoretical analysis for the one-dimensional case and numerical experiments for one and two dimensional cases indicate that this new method is more efficient and more accurate than the existing spectral methods. The idea in this work can be easily adopted to other singularly perturbed elliptic equations, including the fourth-order elliptic equations.

We now briefly describe some of the notations used in this paper. 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 $I=(-1, 1)$, we denote π_N to be the space of real polynomials on I with degrees not exceeding N. We set $X_N=\{u_N \in \pi_N: u_N(\pm 1)=0\}$. We shall use letters of boldface type to denote vectors and vector functions as well as product spaces such as $\mathbf{X}_N=\prod_{i=1}^d X_N$.

The rest of the paper is organized as follows. In Section 2, we introduce the new spectral method. In Section 3, we develop efficient solution techniques for the new Legendre-Galerkin method applied to the convection-diffusion equations in one-and two-dimensional domains. In Section 4, we present numerical experiments, by using both the new and conventional Legendre-Galerkin methods, on several typical examples with thin boundary layer. Some error analysis for the one-dimensional case is presented in the Appendix.

2. THE NEW SPECTRAL GALERKIN METHOD

We first examine the weak formulation for Eqs. (1.1) and (1.2). The weak formulation of Eq. (1.1) reads: Find $u \in H_0^1(\Omega)$ such that

$$\varepsilon \int_\Omega \nabla u \cdot \nabla v \, d\mathbf{x} + \int_\Omega (\nabla u \cdot \mathbf{p})\, v \, d\mathbf{x} + \int_\Omega quv \, d\mathbf{x}$$

$$= \int_\Omega fv \, d\mathbf{x}, \qquad \forall v \in H_0^1(\Omega) \tag{2.1}$$

The conventional spectral Galerkin method is: Find u_N in $\mathbf{X}_N$ such that

$$\varepsilon \int_\Omega \nabla u_N \cdot \nabla v \, d\mathbf{x} + \int_\Omega (\nabla u_N \cdot \mathbf{p})\, v \, d\mathbf{x} + \int_\Omega q u_N v \, d\mathbf{x} = \int_\Omega f v \, d\mathbf{x}, \qquad \forall v \in \mathbf{X}_N \tag{2.2}$$

As indicated in the introduction, Eq. (2.2) is not appropriate to approximate Eq. (2.1) when $\varepsilon \ll 1$. To introduce the new method, we apply the transformation $\mathbf{x} = \mathbf{g}(\mathbf{y})$ as described in Section 1 to Eq. (2.1). Let us denote $J(\mathbf{y}) = \prod_{i=1}^d J_i(y_i)$ with $J_i(y_i) = g_i'(y_i)$ and

$$GH_0^1(\Omega) := \left\{ v \in H_0^1(\Omega) \colon \|v\|_{L^2_J(\Omega)} + \sum_{i=1}^d \|\partial_{y_i} v\|_{L^2_{\omega_i}(\Omega)} < \infty \right\} \tag{2.3}$$

where $\omega_i(\mathbf{y}) = a_i^2(y_i)\, J(\mathbf{y})$. It should be noted that all the smooth functions with compact support in Ω are indeed in this space. A weak formulation of Eq. (1.2) can be established in $GH_0^1(\Omega)$ which is the image space of $H_0^1(\Omega)$ under the transformation $Gu := u \circ \mathbf{g}$. We multiply the Eq. (1.2) by $J(\mathbf{y})$ and set

$$A(v, w) = \sum_{i=1}^d \int_\Omega (a_i^2 J)(\partial_{y_i} v \, \partial_{y_i} w) \, d\mathbf{y} + \int_\Omega Q \, vw \, d\mathbf{y} \tag{2.4}$$

$$B(v, w) = \sum_{i=1}^d \int_\Omega a_i J P_i(\partial_{y_i} v)\, w \, d\mathbf{y} \tag{2.5}$$

$$(F, w) = \int_\Omega F w \, d\mathbf{y} \tag{2.6}$$

where $Q(\mathbf{y}) = q(\mathbf{g}(\mathbf{y}))\, J(\mathbf{y})$ and $F(\mathbf{y}, \varepsilon) = f(\mathbf{g}(\mathbf{y}), \varepsilon)\, J(\mathbf{y})$. Then the weak formulation for Eq. (1.2) is as follows:

Find $v \in GH_0^1(\Omega)$ such that

$$\varepsilon A(v, w) + B(v, w) = (F, w), \quad \forall w \in GH_0^1(\Omega) \tag{2.7}$$

We now consider the approximation of Eq. (2.7) by using a spectral Galerkin method. At the heart of the new spectral Galerkin method is a new trial function space. Although it is possible to present this space in the $\mathbf{x}$ variable(s) and then introduce the new method for Eq. (2.1) directly, it is more convenient to introduce the trial function space in the $\mathbf{y}$ variable(s) and introduce the scheme for the transformed Eq. (2.7).

It is essential to find suitable trial function spaces in order to properly approximate the solution of Eq. (2.7) in $GH_0^1(\Omega)$. It is clearly improper to consider the Galerkin approximation for Eq. (2.7) in $\mathbf{X}_N$. One would then

naturally consider the image space of $\mathbf{X}_N$ under the transformation G defined earlier as the trial function space. It turns out, however, that one would obtain the same results by applying the spectral Galerkin methods directly to Eq. (2.7).

Let $Y_N^i = \{v \in H_0^1(\Omega)(I): v' = J_i P, P \in \pi_N\}$, $i = 1,..., d$. It is clear that Y_N^i is a N dimensional subspace of $GH_0^1(I)$ with $Gu := u \circ g_i$. It turns out that the space $\mathbf{Y}_N = \prod_{i=1}^d Y_N^i$ is a good choice as the trial function space. Therefore, the new spectral Galerkin approximation for Eq. (2.7) reads: Find $v_N \in \mathbf{Y}_N$ such that

$$\varepsilon A(v_N, w) + B(v_N, w) = (F, w), \quad \forall w \in \mathbf{Y}_N \tag{2.8}$$

The theoretical analysis, especially the error analysis, for this scheme is not an easy task. In the Appendix, we present some results for the one dimensional case. The details of the proof can be found in Liu and Tang (1994b). The analysis for the multi-dimensional case is much more difficult and will be addressed in a future work.

3. AN EFFICIENT IMPLEMENTATION OF THE NEW GALERKIN METHOD

A new spectral Galerkin method is introduced above and the results in the Appendix indicate that the new method leads to higher accuracy when applied to Eq. (1.1) with $\varepsilon \ll 1$. However, one important question left unanswered is how to implement the new method efficiently. It is clear that the efficiency of the method depends on the choice of basis functions for Y_N. If the basis functions are not properly chosen, the resulting linear system will generally have a full matrix. Therefore, the computational work will be significantly increased compared to the conventional spectral Galerkin method [cf. Shen (1994)]. In this section, we show that for problems with constant coefficients, we can find an appropriate basis for Y_N such that the resulting linear systems have sparse matrices. Furthermore, these linear systems can be solved by an efficient direct method. The problems with variable coefficients are solved by using a preconditioned conjugate gradient type method with a suitable constant-coefficient problem as the preconditioner.

3.1. One-Dimensional Case

Let us consider first the following equation:

$$-\varepsilon v_{xx} + \beta u_x + \gamma u = f, \; x \in I; \; u(\pm 1) = 0 \tag{3.1}$$

where β and γ are some appropriate constants.

Let k be a positive integer. We consider the following transformation:

$$x = g(y) = g(y, k) = -1 + \sigma_k \int_{-1}^{y} (1 - t^2)^k \, dt$$

$$\text{with} \quad \sigma_k = 2 \left(\int_{-1}^{1} (1 - y^2)^k \, dy \right)^{-1} \tag{3.2}$$

Hence, $J(y) = J(y, k) = g'(y, k) = \sigma_k (1 - y^2)^k$ and

$$Y_N = Y_N(k) = \{ v \in H_0^1(I) : v' = JP \text{ with } P \in \pi_N \} \tag{3.3}$$

To alleviate the notations, the parameter k in the notations will be frequently dropped when no confusion is possible.

Then the new spectral Galerkin method for Eq. (3.1) is: Find $v_N \in Y_N$ such that

$$\varepsilon \int_{-1}^{1} J^{-1} v_N' w' \, dy + \beta \int_{-1}^{1} J v_N' w \, dy + \gamma \int_{-1}^{1} J v_N w \, dy$$

$$= \int_{-1}^{1} J f \circ g w \, dy, \quad \forall w \in Y_N \tag{3.4}$$

It is clear that Y_N is a N dimensional space. However, it is not clear at all how to construct an appropriate basis for Y_N such that the linear system (3.4) can be efficiently solved.

Since the Legendre polynomials form an orthogonal basis for $L^2(I)$, it is natural to construct basis functions for Y_N by using the Legendre polynomials. Let $L_i(x)$ denote the ith degree Legendre polynomial. If we set

$$\psi_i(y) = \int_{-1}^{y} J(t)(L_i(t) + \alpha_i) \, dt, \qquad i = 1, 2, \ldots, N \tag{3.5}$$

and we choose $\alpha_i = -\frac{1}{2} \int_{-1}^{1} J(t) L_i(t) \, dt$, then $\psi_i(1) = 0$ and $\psi_i \in Y_N$. It is obvious that $\{\psi_i\}_{i=1, 2, \ldots, N}$ are linearly independent and therefore form a basis for Y_N. Unfortunately, this basis leads to a full matrix for the linear system (3.4) and hence is prohibited in practice for N large. We shall construct below a more appropriate basis by exploiting the properties of the Legendre polynomials.

The following well known identities of the Legendre polynomials [see for instance Szegö (1975)] will be frequently used:

$$\int_{-1}^{1} L_i(y)\, L_j(y)\, dy = = \frac{2}{2i+1}\,\delta_{ij} \tag{3.6}$$

$$\int_{-1}^{1} L_i(t)\, dt = \frac{1}{2i+1}\,(L_{i+1}(y) - L_{i-1}(y)) \tag{3.7}$$

$$(1-y^2)\, L_i'(y) = \frac{i(i+1)}{2i+1}\,(L_{i-1}(y) - L_{i+1}(y)) \tag{3.8}$$

$$yL_i(y) = \frac{1}{2i+1}\,((i+1)\, L_{i+1}(y) + iL_{i-1}(y)) \tag{3.9}$$

Let us define

$$\phi_{i,k}(y) = \int_{-1}^{y} (1-t^2)^k\, L_{i+k+1}^{(k)}(t)\, dt, \qquad i = 0, 1, 2,..., \quad k = 1, 2,... \tag{3.10}$$

The following Lemma shows that $\phi_{i,k}$ can be expressed as compact combination of the Legendre polynomials.

Lemma 1. There exist constants $\{a_j^{(i,k)}\}$ such that

$$\phi_{i,k}(y) = \sum_{j=0}^{k+1} a_j^{(i,k)} L_{i+2j}(y), \qquad i = 0, 1, 2,..., \quad k = 1, 2,... \tag{3.11}$$

Proof. We prove the result by induction on k. For $k = 1$, we can derive Eq. (3.11) by direct computation using Eqs. (3.10), (3.7) and (3.8).

Now we assume that Eq. (3.11) holds for $k \leqslant m-1$. Then by Eq. (3.10) and integration by part,

$$\phi_{i,m}(y) = (1-y^2)^m\, L_{i+m+1}^{(m-1)}(y) + \int_{-1}^{y} 2t(1-t^2)^{m-1}\, L_{i+m+1}^{(m-1)}(t)\, dt \tag{3.12}$$

We derive from Eq. (3.10) that

$$\phi_{i+1,m-1}'(y) = (1-y^2)^{m-1}\, L_{i+m+1}^{(m-1)}(y) \tag{3.13}$$

Therefore, since $\phi_{i,k}(-1) = 0$ for any i and k, we have

$$\begin{aligned}\phi_{i,m}(y) &= (1-y^2)\, \phi_{i+1,m-1}'(y) + 2\int_{-1}^{y} t\phi_{i+1,m-1}'(t)\, dt \\ &= (1-y^2)\, \phi_{i+1,m-1}'(y) + 2y\phi_{i+1,m-1}(y) \\ &\quad - 2\int_{-1}^{y} \phi_{i+1,m-1}(t)\, dt \end{aligned} \tag{3.14}$$

By assumption, we have

$$\phi_{i+1,m-1}(y)=\sum_{j=0}^{m} a_j^{(i+1,m-1)} L_{i+1+2j}(y)$$

Hence

$$\phi'_{i+1,m-1}(y)=\sum_{j=0}^{m} a_j^{(i+1,m-1)} L'_{i+1+2j}(y)$$

Using these two relations in Eq. (3.14) and using Eq. (3.7)–(3.9), one can easily conclude that there exist constants $a_j^{(i,m)}$ such that

$$\phi_{i,m}(y)=\sum_{j=0}^{m+1} a_j^{(i,m)} L_{i+2j}(y), \qquad i=0, 1, 2,...$$

The proof is complete.

Theorem 1. $\{\phi_{i,k}\}_{i=0,1,...,N-1}$ form a basis for $Y_N(k)$, $k=1, 2,....$

Proof. By definition, $\phi_{i,k}(-1)=0$, $i=0, 1,..., N-1$, $k=1, 2,....$ On the other hand, by using Lemma 1 and the fact that $L_i(\pm 1)=(\pm 1)^i$, we derive

$$\phi_{i,k}(1)=\sum_{j=0}^{k+1} a_j^{(i,k)} L_{i+2j}(1)=\sum_{j=0}^{k+1} a_j^{(i,k)} L_{i+2j}(-1)(-1)^{i+2j}$$

$$=(-1)^i \sum_{j=0}^{k+1} a_j^{(i,k)} L_{i+2j}(-1)=(-1)^i \phi_{i,k}(-1)=0$$

Since $L_{i+k+1}^{(k)} \in \pi_N$ for $i=0, 1,..., N-1$, we conclude from Eq. (3.10) that $\phi_{i,k} \in Y_N(k)$ for $i=0, 1,..., N-1$. On the other hand, if there are constants b_i such that $\sum_{i=0}^{N-1} b_i \phi_{i,k}(y)=0$, $y \in I$, then by Eq. (3.10), we have

$$\int_{-1}^{y} (1-t^2)^k \sum_{i=0}^{N-1} b_i L_{i+k+1}^{(k)}(t)\, dt=0, \qquad y \in I$$

Taking the derivative with respect to y, we derive

$$(1-y^2)^k \sum_{i=0}^{N-1} b_i L_{i+k+1}^{(k)}(y)=0, \qquad y \in I$$

which implies

$$\sum_{i=0}^{N-1} b_i L_{i+k+1}^{(k)}(y)=0, \qquad y \in I$$

It is clear that $\{L_{i+k+1}^{(k)}(y)\}_{i=0,1,\ldots,N-1}$ are linearly independent. Therefore, $b_0 = b_1 = \cdots = b_{N-1} = 0$, and $\{\phi_{i,k}(y)\}_{i=0,1,\ldots,N-1}$ are linearly independent. Since $\dim\{Y_N(k)\} = N$, $\{\phi_{i,k}(y)\}_{i=0,1,\ldots,N-1}$ form a basis for $Y_N(k)$. The proof is complete.

Since $J(y,k) = \sigma_k(1-y^2)^k$, we conclude from Theorem 1 that for any nonzero constants $\sigma_{i,k}$, the functions

$$\psi_{i,k}(y) = \alpha_{i,k} \int_{-1}^{y} J(t,k)\, L_{i+k+1}^{(k)}(t)\, dt, \qquad i = 0, 1, \ldots, N-1 \tag{3.15}$$

also form a basis for $Y_N(k)$. The constants $\alpha_{i,k}$ are of our choice. By Lemma 1, we have

$$\psi_{i,k}(y) = \alpha_{i,k}\sigma_k \phi_{i,k}(y) = \sum_{j=0}^{k+1} \alpha_{i,k}\sigma_k a_j^{(i,k)} L_{i+2j}(y), \qquad i = 0, 1, \ldots, N-1 \tag{3.16}$$

Let us now reformulate the Eq. (3.4) under the basis functions defined by Eq. (3.15). For a fixed k, we denote

$$v_N(y) = \sum_{i=0}^{N-1} x_i \psi_{i,k}(y), \qquad \mathbf{x} = (x_0, x_1, \ldots, x_{N-1})^T$$

and

$$f_i = \int_{-1}^{1} J(y,k)\, f(g(u,k))\, \psi_{i,k}(y)\, dt, \qquad \mathbf{f} = (f_0, f_1, \ldots, f_{N-1})^T$$

We also denote

$$a_{ij} = a_{ij}(k) = \int_{-1}^{1} J^{-1}(y,k)\, \psi'_{i,k}(y)\, \psi'_{j,k}(y)\, dy$$

$$A = A(k) = (a_{ij}(k))_{i,j=0,1,\ldots,N-1} \tag{3.17}$$

$$b_{ij} = b_{ij}(k) = \int_{-1}^{1} J(y,k)\, \psi_{i,k}(y)\, \psi_{j,k}(y)\, dy$$

$$B = B(k) = (b_{ij}(k))_{i,j=0,1,\ldots,N-1} \tag{3.18}$$

$$c_{ij} = c_{ij}(k) = \int_{-1}^{1} \psi'_{j,k}(y)\, \psi_{i,k}(y)\, dy$$

$$C = C(k) = (c_{ij}(k))_{i,j=0,1,\ldots,N-1} \tag{3.19}$$

By using these notations, we find that Eq. (3.4) is equivalent to the following linear system:

$$(\varepsilon A + \beta C + \gamma B)\,\mathbf{x} = \mathbf{f} \tag{3.20}$$

The next Lemma shows that the matrices A, B, and C are banded with band-lengths independent of N. Thus, the system Eq. (3.20) can be solved in $O(N)$ operations for any k.

Lemma 2. Let k be a positive integer. Then $A(k)$ and $B(k)$ are symmetric positive definite and $C(k)$ is skew-symmetric. Furthermore,

$$\begin{aligned} a_{ij}(k) &= 0 \quad \text{for} \quad i \neq j \\ b_{ij}(k) &= 0 \quad \text{for} \quad i \neq j,\ j \pm 2, \ldots,\ j \pm (4k+2) \\ c_{ij} &= 0 \quad \text{for} \quad i \neq j \pm 1,\ j \pm 3, \ldots,\ j \pm (2k+1) \end{aligned}$$

Proof. We observe immediately that A and B are symmetric and C is skew-symmetric. Furthermore, for any $\mathbf{x} = (x_0, x_1, \ldots, x_{N-1})^T$, let $w(y) = \sum_{i=0}^{N-1} x_i \psi_{i,k}(y)$. Then we have

$$\mathbf{x}^T A \mathbf{x} = \int_{-1}^{1} J^{-1}(y)\, w'(y)\, w'(y)\, dy > 0 \qquad \text{if} \quad \mathbf{x} \neq 0$$

and

$$\mathbf{x}^T B \mathbf{x} = \int_{-1}^{1} J(y)\, w(y)\, w(y)\, dy > 0 \qquad \text{if} \quad \mathbf{x} \neq 0$$

Therefore A and B are positive definite.

We derive from Eq. (3.15) that

$$\psi'_{i,k}(y) = \alpha_{i,k} J(y, k)\, L^{(k)}_{i+k+1}(y) \tag{3.21}$$

Using Eqs. (3.17) and (3.21), and integration by part,

$$a_{ij}(k) = \alpha_{i,k} \int_{-1}^{1} L^{(k)}_{i+k+1}(y)\, \psi'_{j,k}(y)\, dy = -\alpha_{i,k} \int_{-1}^{1} L^{(k+1)}_{i+k+1}(y)\, \psi_{j,k}(y)\, dy$$

Since $L^{(k+1)}_{i+k+1} \in \pi_i$, we derive from this relation and Eq. (3.16) that $a_{ij}(k) = 0$ for $i < j$. By symmetry, we have also $a_{ij}(k) = 0$ for $i > j$.

Since $J(y, k) = \sigma_k (1 - y^2)^k$, we have

$$b_{ij}(k) = \sigma_k \int_{-1}^{1} (1 - y^2)^k\, \psi_{i,k}(y)\, \psi_{j,k}(y)\, dy$$

We observe from Eq. (3.16) that $(1-y^2)^k \psi_{i,k} \in \pi_{i+2(k+1)+2k}$. Therefore, $b_{ij}(k)=0$ for $i+4k+2<j$. By symmetry, we have also $b_{ij}(k)=0$ for $j+4k+2<i$. On the other hand, it is easy to realize by using repeatedly Eq. (3.9) that Legendre expansion of $(1-y^2)^k L_i(y)$ is of the form $\sum_{l=-k}^{k} \beta_l^{(k,i)} L_{i+2l}(y)$. Therefore, $b_{ij}(k)=0$ if i, j are not of the same parity. We then conclude that

$$b_{ij}(k)=0 \qquad \text{for} \quad i \neq j, j \pm 2, \ldots, j \pm (4k+2)$$

By using Eq. (3.7), it is easy to derive that

$$L'_m(x) = \sum_{l=0;\, l+m \text{ odd}}^{m-1} (2l+1) L_l(x)$$

By using this relation and Eq. (3.15) in Eq. (3.19), we can conclude that

$$c_{ij}(k)=0 \qquad \text{for} \quad i \neq j \pm 1, j \pm 3, \ldots, j \pm (2k+1)$$

The proof is complete.

The entries of A, B and C can be explicitly computed by using properties of the Legendre polynomials. The following lemma provides explicit formulae for a_{ij}, b_{ij} and c_{ij} in the case of $k=1$. Note that in the following lemma, the parameter k is dropped from most of the notations, and a_i, b_i, c_i are not related to a_{ij}, b_{ij}, c_{ij}.

Lemma 3. Let $k=1$, then $J(y)=J(y,1)=\frac{3}{2}(1-y^2)$. We set $\alpha_{i,1} = -(2(2i+3)(2i+5)/3(i+2)(i+3))$ and denote

$$a_i = -\frac{2(2i+5)}{2i+7}, \quad b_i = \frac{2i+3}{2i+7}, \quad c_i = \frac{i}{2i+1}, \quad d_i = \frac{-2i^2-i+3}{(2i+1)(2i+7)}$$

$$e_i = -\frac{2i^2+19i+42}{(2i+7)(2i+9)}, \quad f_i = \frac{(2i+3)(2i+5)}{(2i+7)(2i+9)}, \quad g_i = \frac{2}{2i+1}$$

Then we have

$$\psi_{i,1}(y) = L_i(y) + a_i L_{i+2}(y) + b_i L_{i+4}(y) \tag{3.22}$$

$$a_{ii} = \frac{4(2i+3)^2(2i+5)}{3(i+2)(i+3)}, \qquad a_{ij}=0 \qquad \text{for} \quad i \neq j \tag{3.23}$$

$$c_{i+1,i} = -c_{i,i+1} = -\frac{6(2i+5)}{2i+9}, \qquad c_{i+3,i} = -c_{i,i+3} = \frac{2(2i+3)}{2i+7} \tag{3.24}$$

$$c_{ij}=0 \qquad \text{for} \quad i\neq j\pm 1,\ j\pm 3 \tag{3.25}$$

$$b_{i,i+6}=-f_i c_{i+6} g_{i+5} \tag{3.26}$$

$$b_{i,i+4}=b_i g_{i+4}-(e_i c_{i+4} g_{i+3}+f_i d_{i+4} g_{i+5}) \tag{3.27}$$

$$b_{i,i+2}=(a_i g_{i+2}+b_i a_{i+2} g_{i+4}) - (d_i c_{i+2} g_{i+1}+e_i d_{i+2} g_{i+3}+f_i e_{i+2} g_{i+5}) \tag{3.28}$$

$$b_{i,i}=(g_i+a_i^2 g_{i+2}+b_i^2 g_{i+4}) - (c_i^2 g_{i-1}+d_i^2 g_{i+1}+e_i^2 g_{i+3}+f_i^2 g_{i+5}) \tag{3.29}$$

$$b_{i,j}=b_{j,i}, \quad b_{i,j}=0 \qquad \text{for} \quad i\neq j,\ j\pm 2,\ j\pm 4,\ j\pm 6 \tag{3.30}$$

Proof. We can derive Eq. (3.22) by direct computation using Eqs. (3.10), (3.7) and (3.8). By using Eq. (3.8) and integration by part, we obtain

$$a_{ij}=\alpha_{i,1}\alpha_{j,1}\frac{(i+2)(i+3)}{2i+5}\int_{-1}^{1}(L_{i+1}(y)-L_{i+3}(y))\,L'_{j+2}(y)\,dy$$

(since $L_i(\pm 1)=(\pm 1)^i$)

$$=-\alpha_{i,1}\alpha_{j,1}\frac{(i+2)(i+3)}{2i+5}\int_{-1}^{1}L_{j+2}(y)(L'_{i+1}(y)-L'_{i+3}(y))\,dy$$

(using Eq. (3.7))

$$=\alpha_{i,1}\alpha_{j,1}(i+2)(i+3)\int_{-1}^{1}L_{j+2}(y)\,L_{i+2}(y)\,dy$$

Equation (3.23) is then a direct consequence of the above relation and Eq. (3.6).

Similarly, using Eq. (3.10), Eq. (3.8) and integration by part,

$$c_{ij}=\int_{-1}^{1}\psi'_{j,1}(y)\,\psi_{i,1}(y)\,dy=\alpha_{j,1}\int_{-1}^{1}J(y)\,L'_{j+2}(y)\,\psi_{i,1}(y)\,dy$$

$$=\frac{3}{2}\frac{(j+2)(j+3)}{2j+5}\alpha_{j,1}\int_{-1}^{1}(L_{j+1}(y)-L_{j+3}(y))\,\psi_{i,1}(y)\,dy$$

We can then derive Eq. (3.24) and (3.25) by direct computation using Eq. (3.22) and Eq. (3.6).

The computation of b_{ij} is a little more involved. We split b_{ij} into two part as follows:

$$b_{ij} = \tfrac{3}{2}\int_{-1}^{1} \psi_{i,1}(t)\,\psi_{j,1}(t)\,dt - \tfrac{3}{2}\int_{-1}^{1} t\psi_{i,1}(t)\,t\psi_{j,1}(t)\,dt = b_{ij}^1 - b_{ij}^2$$

Then b_{ij}^1 can be easily computed by using Eq. (3.22) and Eq. (3.6). To compute b_{ij}^2, we use Eq. (3.9) and Eq. (3.22) to get

$$t\psi_i(t) = c_i L_{i-1}(t) + d_i L_{i+1}(t) + e_i L_{i+3}(t) + f_i L_{i+5}(t)$$

We can then compute b_{ij} and derive Eqs. (3.26)–(3.30) by straightforward computation using Eq. (3.6).

Remark 1. Thanks to Lemma 1, the matrix for the system Eq. (3.20) has only nine nonzero diagonals in the case $k=1$ and hence can be easily inverted in $O(N)$ operations. The matrix becomes even simpler when β or $\gamma=0$. In fact, if $\gamma=0$, the matrix is tridiagonal; if $\beta=0$, the system $(\varepsilon A+\gamma B)\,\mathbf{x}=\mathbf{f}$ can be decoupled into two subsystems with odd and even components of $\mathbf{x}$.

In the case $k>1$, it is clear that we can also derive explicit formulae for the corresponding matrices $A(k)$, $B(k)$ and $C(k)$ under the basis Eq. (3.15). We should point out that although the numbers of nonzero diagonals of $B(k)$ and $C(k)$ increase as k increases, but they remain to be independent of N (see Lemma 2). Therefore, the system Eq. (3.20) can still be solved in $O(N)$ operations.

3.2. Two-Dimensional Case

It is well known that certain singularly perturbed problems arising in physical and engineering sciences also exhibit boundary layers of the form

$$u(\mathbf{x}, s) = a(s)\exp(-\varepsilon^{-\gamma}\rho(\mathbf{x}))$$

where s, ρ denote the arc length and the normal distance to the boundary of a point $\mathbf{x}$ within the boundary layer, $a(s)$ is a smooth function. It is clear that this type of boundary layer phenomenon is essentially one-dimensional. Hence, good approximation properties can be expected by using the tensor product of the one-dimensional approximation space. To fix the idea, we consider the model problem:

$$-\varepsilon\,\Delta u + \beta u_{x_2} + \gamma u = f, \qquad \mathbf{x}\in\Omega = I^2; \quad u|_{\partial\Omega}=0 \tag{3.31}$$

where β and γ are some appropriate constants. More general equation with additional first-order term αu_{x_1} can also be treated in similar manner.

We continue to use the notations in Section 3.1. For fixed k, Let g be the function defined in Eq. (3.2), we use the transformation $x_i = g(y_i) = g(y_i, k)$, $i = 1, 2$. Hence, $J_i(y_i) = J(y_i) = J(y_i, k) = \sigma_k(1 - y_i^2)^k$ and $a_i(y_i) = 1/J_i(y_i)$, $i = 1, 2$. We will drop the parameter k from the following notations. We set

$$\mathbf{Y}_N = Y_N \times Y_N = \text{span}\{\psi_i(y_1)\,\psi_j(y_2) : i, j = 0, 1, \ldots, N-1\}$$

Therefore, the scheme in Eq. (2.8) applied to Eq. (3.31) is as follows: Find $v_N \in \mathbf{Y}_N$ such that

$$\int_\Omega J_2 a_1\, \partial_{y_1} v_N\, \partial_{y_1} w\, d\mathbf{y} + \int_\Omega J_1 a_2\, \partial_{y_2} v_N\, \partial_{y_2} w\, d\mathbf{y} + \beta \int_\Omega J_1\, \partial_{y_2} v_N w\, d\mathbf{y}$$

$$+ \gamma \int_\Omega J_1 J_2 v_N w\, d\mathbf{y} = \int_\Omega J_1 J_2 f \circ g w\, d\mathbf{y}, \qquad \forall w \in \mathbf{Y}_N \tag{3.32}$$

Let us set

$$v_N(y_1, y_2) = \sum_{i,j=0}^{N-1} v_{ij} \psi_i(y_1)\,\psi_j(y_2) \tag{3.33}$$

$$V = (v_{ij})_{i,j=0,1,\ldots,N-1}$$

$$f_{ij} = \int_\Omega J_1(y_1)\, J_2(y_2)\, f(g(y_1), g(y_2))\, \psi_i(y_1)\, \psi_j(y_2)\, d\mathbf{y} \tag{3.34}$$

$$F = (f_{ij})_{i,j=0,1,\ldots,N-1}$$

Let A, B, and C be the matrices defined in Section 3.1, we find that Eq. (3.32) is equivalent to the following matrix equation:

$$\varepsilon(AVB + BVA) + \beta AVC + \gamma BVB = F \tag{3.35}$$

We can also rewrite this equation in the following form using the tensor product notation:

$$\{\varepsilon(A \otimes B + B \otimes A) + \beta A \otimes C + \gamma B \otimes B\}\, \mathbf{v} = \mathbf{f}$$

where $\mathbf{f}$ and $\mathbf{v}$ are respectively F and V written in the form of a column vector, i.e.

$$\mathbf{f} = (f_{0,0}, f_{1,0}, \ldots, f_{N-1,0}; f_{0,1}, \ldots, f_{N-1,1}; \ldots; f_{0,q}, \ldots, f_{N-1,N-1})^T$$

and $\otimes$ denotes the tensor product of matrices, i.e., $A \otimes B = (Ab_{ij})_{i,j=0,1,\ldots,N-1}$.

We now describe how the Eq. (3.35) can be efficiently solved by using the matrix decomposition method. Since A is symmetric positive definite, $A^{1/2}$ is well defined. We make the transformation

$$A^{1/2}V = X, \qquad \text{i.e.,} \quad V = A^{-1/2}X$$

Multiplying $A^{-1/2}$ to Eq. (3.35) and applying this transformation, we get

$$\varepsilon(XB + A^{-1/2}BA^{-1/2}XA) + \beta XC + \gamma A^{-1/2}BA^{-1/2}XB = A^{-1/2}F \tag{3.36}$$

Since the matrix $A^{-1/2}BA^{-1/2}$ is symmetric, there exist an orthogonal matrix E and a diagonal matrix Λ, consisting respectively eigenvectors and eigenvalues of $A^{-1/2}BA^{-1/2}$, such that

$$A^{-1/2}BA^{-1/2}E = E\Lambda$$

We then define $W = E^T X$ and set $X = EW$ in Eq. (3.36), obtaining

$$\varepsilon(EWB + E\Lambda WA) + \beta EWC + \gamma E\Lambda WB = A^{-1/2}F$$

Therefore, since $E^{-1} = E^T$, we find

$$\varepsilon(WB + \Lambda WA) + \beta WC + \gamma \Lambda WB = E^T A^{-1/2}F \equiv G \tag{3.37}$$

Taking the transpose of this equation, since A and B are symmetric and C is skew-symmetric, we obtain

$$\varepsilon(BW^T + AW^T\Lambda) - \beta CW^T + \gamma BW^T\Lambda = G^T \tag{3.38}$$

Let $\mathbf{w}_p = (w_{p,0}, w_{p,1}, \ldots, w_{p,N-1})^T$ and $\mathbf{g}_p = (g_{p,0}, g_{p,1}, \ldots, g_{p,N-1})^T$ for $p = 0, 1, \ldots, N-1$. Then the pth column of the Eq. (3.38) can be written as:

$$(\varepsilon B + \varepsilon\lambda_p A - \beta C + \gamma\lambda_p B)\,\mathbf{w}_p = \mathbf{g}_p, \qquad p = 0, 1, \ldots, N-1 \tag{3.39}$$

where λ_p is the pth entry of the diagonal matrix Λ. We note that for each p, Eq. (3.39) is in fact an equation of the form Eq. (3.20).

In summary, the solution of Eq. (3.35) consists of the following steps:

1. Pre-processing: compute the eigenvalues and eigenvectors of $A^{-1/2}BA^{-1/2}$;
2. Compute $G = E^T A^{-1/2}F$;
3. Obtain W by solving Eq. (3.39);
4. Set $V = A^{-1/2}EW$.

Since $A^{-1/2}BA^{-1/2}$ has only a fixed number of nonzero diagonals, the eigenvalues and eigenvectors of $A^{-1/2}BA^{-1/2}$ can be computed in $O(N^2)$ operations. Solving Eq. (3.39) for $p=0, 1,..., N_1$ requires only $O(N^2)$ operations as well. Hence, the bottleneck of the algorithm is the two matrix multiplications in Steps 2 and 4. The operation counts for the two matrix multiplications can be further reduced by half if we take into account the fact that $e_{kj}=0$ for $k+j$ odd. Consequently, Steps 2 and 4 take about $2N^3$ arithmetic operations. In short, the complexity of the new Legendre-Galerkin method is essentially the same as that of the conventional Legendre-Galerkin method [cf. Shen (1994)].

Remark 2. This algorithm can be easily extended to the three-dimensional case with $\Omega = I^3$. We refer to [Shen (1994)] for similar considerations on this aspect.

For problems with variable coefficients, the resulting discrete systems usually have full matrices. Hence, it is inefficient to evaluate these matrices and to invert them directly. However, given an equation with variable coefficients, we can choose an appropriate equation (which approximates the original equation in certain sense) with constant coefficients as a preconditioner. Then the original equation with variable coefficients can be solved by using an iterative method such as the Preconditioned Conjugate Gradient Method (see Example 3 later). The convergence rate of the iterative method varies with different equations but is usually independent of N.

4. NUMERICAL EXPERIMENTS

In this section we report on several numerical results by using the new Legendre-Galerkin method presented in the previous section. All the computations are based on the transformation of Eq. (3.2) with $k=1$. In order to demonstrate the high accuracy and the efficiency of the new method, we also make some comparisons with the conventional Legendre-Galerkin method [cf. Shen (1994)] and with the boundary layer resolving Chebyshev method [cf. Tang and Trummer (1996)].

We note that for the transformation of Eq. (3.2) with $k=1$, the highest degree of Legendre polynomials in both $\mathbf{X}_N$ and $\mathbf{Y}_{N-3}$ is N. Hence, we shall compare the conventional Legendre-Galerkin method in $\mathbf{X}_N$ with the new Legendre-Galerkin method in $\mathbf{Y}_{N-3}$. Note however that $\mathbf{X}_N$ is a $(N-1)^d$ dimensional space, while $\mathbf{Y}_{N-3}$ is a $(N-3)^d$ dimensional space. Let $\mathcal{M}_N$ be the set of the Legendre Gauss-Lobatto collocation points with respect to $\mathbf{X}_N$. For all the examples considered below, we compute

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

and

$$\|v - v_{N-3}\|_{I^\infty} \equiv \max_{\mathbf{y} \in \mathcal{M}_N} |v(\mathbf{y}) - v_{N-3}(\mathbf{y})|$$

where $v = u(\mathbf{g}(\mathbf{y}))$, u_N and v_{N-3} are respectively the solution of the conventional Legendre-Galerkin scheme in Eq. (2.2) and the new Legendre-Galerkin scheme in Eq. (2.8). Since the collocation points in the y variable(s) in $\mathcal{M}_{N}$ are well condensed near the boundary for the x variables, the boundary layers in x variable(s) are well resolved by $\mathcal{M}_{N}$ (for N sufficiently large). Hence, $\|u - u_N\|_{I^\infty}$ (resp. $\|v - v_{N-3}\|_{I^\infty}$) is usually very close to $\|u - u_N\|_{I^\infty(\Omega)}$ (resp. $\|v - v_{N-3}\|_{I^\infty(\Omega)}$).

Example 1. Our first example is the one-dimensional diffusion equation

$$-\varepsilon u_{xx} + u = -\frac{x+1}{2}, \qquad x \in I, \quad u(\pm 1) = 0 \tag{4.1}$$

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 $O(\sqrt{\varepsilon})$.

In Tables I and II, we list the maximum pointwise error $\|u - u_N\|_{I^\infty}$ by the conventional Legendre-Galerkin method (CLGM), and maximum pointwise error $\|v - v_{N-3}\|_{I^\infty}$ by our new Legendre-Galerkin method (NLGM). For the sake of comparison, we also included in Table II the available results by the boundary layer resolving Chebyshev-collocation method with $m = 1$ which corresponds to NLGM with $k = 1$ [BLRCC, cf. Tang and Trummer (1996)]. We recall that for the same value N, the

Table I. Maximum Pointwise Errors $\|u - u_N\|_{I^\infty}$ and $\|v - v_{N-3}\|_{I^\infty}$ for Example 1

Method	N	$\varepsilon = 10E-8$	$\varepsilon = 10E-9$	$\varepsilon = 10E-10$	$\varepsilon = 10E-11$	$\varepsilon = 10E-12$
CLGM	512	3.9E−7	4.8E−3	1.6E−1		
CLGM	1024		5.5E−7	1.4E−3	1.0E−1	
CLGM	2048			3.2E−5	3.2E−4	6.1E−2
NLGM	64	4.5E−3	3.9E−2	1.4E−1		
NLGM	128	1.3E−5	4.5E−4	4.7E−3	2.3E−2	6.4E−2
NLGM	256	3.0E−12	6.6E−9	2.2E−6	1.1E−4	1.5E−3

Table II. Maximum Pointwise Errors $\|u-u_N\|_{I^\infty}$ and $\|v-v_{N-3}\|_{I\infty}$ for Example 2

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

computational complexity of CLGM and NLGM are essentially the same, while that of BLRCC is much higher.

Example 2. The second example is the one-dimensional convection equation

$$-\varepsilon u_{xx}+u_x=-\tfrac{1}{2}, \qquad x\in I, \quad u(\pm 1)=0 \tag{4.2}$$

with the exact solution

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

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

We observe that for the first two examples, the NLGM is considerable more accurate than the CLGM when ε is small. It is transparent from Tables I and II that the NLGM can resolve much finer boundary layer than the CLGM. We emphasize that for a fixed N, the computational complexities of the CLGM and the NLGM are essentially the same. It is also interesting to note that the NLGM is even a little more accurate than the BLSCC, which needs significantly more computational work than the NLGM does.

Example 3. The third example is the following one-dimensional diffusion equation with variable coefficients:

$$-\varepsilon^2 u''+(\varepsilon+x^2)\,u=2\varepsilon^2-(\varepsilon+c^2)\,x^2, \qquad x\in I; \quad u(\pm 1)=0 \tag{4.3}$$

Table III. Maximum Pointwise Errors $\|u-u_N\|_{l^\infty}$ and $\|v-v_{N-3}\|_{l^\infty}$ for Example 3

Method	N	$\varepsilon=10E-4$	$\varepsilon=10E-5$	$\varepsilon=10E-6$	$\varepsilon=10E-7$	$\varepsilon=10E-8$
CLGM	256	$8.7E-2$				
CLGM	512	$2.0E-4$	$O(1)$			
NLGM	64	$1.2E-1$				
NLGM	128	$2.8E-6$	$2.4E-2$			
NLGM	256	$3.8E-9$	$2.5E-5$	$1.6E-3$	$2.8E-2$	
NLGM	512				$5.7E-4$	$2.9E-2$

with the exact solution

$$u(x)=\exp\left(\frac{x^2-1}{2\varepsilon}\right)-x^2$$

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

In Table III, we list the maximum pointwise error $\|u-u_N\|_{l^\infty}$ by the CLGM and $\|v-v_{N-3}\|_{l^\infty}$ by the NLGM. The discrete systems for both schemes are solved by using the preconditioned conjugate gradient method with a suitable equation with constant coefficients as the preconditioner. The number of iteration used to obtain the results in Table III ranges from 30–100, indicating a good convergence behavior given the highly varying coefficient when $\varepsilon \ll 1$.

We have also used the popular adaptive collocation solver COLSYS [Ascher *et al.* (1994)] to solve the first three examples. Although it is heard to make a precise comparison due to the adaptive nature of COLSYS, we do observe that for Examples 1 and 2 which have constant-coefficients, our method is much more efficient than COLSYS while for Example 3 the efficiency of the two methods are comparable.

Example 4. The fourth example is the two-dimensional diffusion equation

$$-\varepsilon\,\Delta u+2u=F,\ (x_1, x_2)\in\Omega=I^2;\quad u|_{\partial\Omega}=0 \tag{4.4}$$

where

$$F(x_1, x_2)=-\tfrac{1}{2}((x_1+1)\,w(x_2)+(x_2+1)\,w(x_1))$$

with

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

Table IV. Maximum Pointwise Errors $\|u-u_N\|_{I^\infty}$ and $\|v-v_{N-3}\|_{I^\infty}$ for Examples 4

Method	N	$\varepsilon = 10E-8$	$\varepsilon = 10E-9$	$\varepsilon = 10E-10$	$\varepsilon = 10E-11$	$\varepsilon = 10E-12$
CLGM	256	$2.1E-2$	$2.7E-1$			
CLGM	512	$6.6E-7$	$8.0E-3$	$1.9E-1$		
NLGM	64	$9.9E-3$	$5.7E-2$			
NLGM	128	$2.6E-5$	$6.2E-4$	$7.2E-3$	$2.5E-2$	$1.2E-1$
NLGM	256	$3.5E-11$	$1.5E-8$	$4.8E-6$	$2.4E-4$	$1.9E-3$

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 the two sides $1 \times (-1, 1)$ and $(-1, 1) \times 1$.

In Table IV, we list the maximum pointwise error $u_N\|_{I^\infty}$ by the CLGM and $\|v-v_{N-3}\|_{I^\infty}$ by the NLGM.

We note that, similarly to the one-dimensional case (cf. Table I), the NLGM can resolve much finer boundary layers and is significantly more accurate than the CLGM. As described in Section 4, the computational complexities of the NLGM and the CLGM in the two-dimensional case are also essentially the same. Similar computational tests have also been carried out for two-dimensional convection-diffusion equations. Very similar results to Example 2 have been observed. We therefore do not include those tests here.

CONCLUSIONS

We have presented a new spectral Galerkin method for solving the convection-dominant convection diffusion equations in a multi-dimensional domain. The key to the success of the new method is to apply a suitable transformation to the original equation before discretizing it and to use a suitable new trial function space. The new method enjoys higher accuracy when applied to problems with thin boundary layers.

We have constructed appropriate basis functions for the new trial function space by using the Legendre polynomials. We showed that by using these basis functions, the resulting linear systems are sparse for problems with constant coefficients. We have also developed efficient solution techniques for solving these linear systems with the computational complexity similar to that of the conventional Legendre Galerkin method. We have only presented numerical results by using the transform 3.2 with $k = 1$ for several typical singular perturbation problems by using both the conventional and new Legendre Galerkin methods. These results indicate that our new method is more efficient and more accurate than the existing spectral methods for problems with thin boundary layers. Furthermore, it

is clear from the theoretical results in Section 2 that our method using a transform 3.2 with $k>1$ will perform significantly better than using the transform with $k=1$.

We note finally that the new Legendre–Galerkin method presented here can be accelerated by using the Chebyshev–Legendre method introduced in Shen (1996).

5. APPENDIX: SOME ANALYTICAL RESULTS FOR ONE-DIMENSIONAL CASES

We first consider the Helmholtz type equation:

$$-\varepsilon u''(x)+q(x)\,u(x)=f(x,\varepsilon), \qquad x\in I, \quad u(\pm 1)=0 \tag{5.5}$$

with $q(x)\geqslant 0$, $x\in I$. Then the transformed equation becomes

$$-\varepsilon(a(y)\,v'(y))'+Q(y)\,v(y)=F(y,\varepsilon), \qquad y\in I, \quad v(\pm 1)=0 \tag{5.6}$$

where

$$v(y)=u(g(y)),\ a(y)=\frac{1}{g'(y)}=\frac{1}{J(y)}$$

$$Q(y)=q(g(y))\,J(y) \qquad \text{and} \qquad F(y,\varepsilon)=f(g(y,\varepsilon)\,J(y)$$

The weak formulation is stated as follows: Find $v\in GH_0^1(I)$ such that

$$-\varepsilon\int_{-1}^{1} av'w'\,dy+\int_{-1}^{1} Qvw\,dy=\int_{-1}^{1} Fw\,dy, \qquad \forall w\in GH_0^1(I) \tag{5.7}$$

We note that $GH_0^1(I)$ is a Hilbert space with the inner product $(v,w)_G=\int_{-1}^{1} av'w'\,dy+\int_{-1}^{1} Jvw\,dy$. From the Poincaré's inequality there is a $c_1>0$ such that

$$\|v\|_{H^1(I)}\leqslant c_1\,\|v\|_{GH_0^1(I)}, \qquad \forall v\in GH_0^1(I) \tag{5.8}$$

The spectral Galerkin approximation of Eq. (5.7) in Y_N reads: Find $v_N\in Y_N$ such that

$$\varepsilon\int_{-1}^{1} av_N'w'\,dy+\int_{-1}^{1} Qv_Nw\,dy=\int_{-1}^{1} Fw\,dy, \qquad \forall w\in Y_N \tag{5.9}$$

It follows from the Lax–Milgram Theorem that Eq. (5.9) is well posed in Y_N.

Hereafter, we use C to denote a positive constant independent of ε and N, but possibly with different values at different places.

Theorem 2. [cf. Liu and Tang (1994a,b)]. Let $u(x)$ and $v_N(y)$ be respectively the unique solution of Eq. (5.5) and of Eq. (5.9). Assume that

there exist constants $C_1, C_2, \beta > 0$ such that $C_1 \leqslant J(y)(1-y^2)^{-\beta} \leqslant C_2$ for any $y \in I$. Then the following estimates hold:

$$\varepsilon \|u' - \underline{v}'_N\|^2_{L^2(I)} + \|u - \underline{v}_N\|^2_{L^2_q(I)} \leqslant C(N^{-2}\varepsilon + N^{-4}) \left(\int_{-1}^{1} (u')^2 \, dx + \int_{-1}^{1} \underline{J}^2 (u'')^2 \, dx \right) \tag{5.10}$$

and

$$\varepsilon \|u' - \underline{v}'_N\|^2_{L^2(I)} + \|u - \underline{v}_N\|^2_{L^2_q(I)} \leqslant C(N^{-4}\varepsilon + N^{-6}) \left(\int_{-1}^{1} (u')^2 \, dx + \int_{-1}^{1} \underline{J}^2 (u'')^2 \, dx + \int_{-1}^{1} \underline{J}^4 (u''')^2 \, dx + \int_{-1}^{1} (\underline{J}')^2 (u'')^2 \, dx \right) \tag{5.11}$$

where $\underline{v}_N(x) = v_N(g^{-1}(x))$, $\underline{J}(x) = J(g^{-1}(x))$ and $\underline{J}'(x) = J'(g^{-1}(x))$.

These estimates can be generalized to higher order approximations, with the right-hand sides dominated by $N^{-2m} \int_{-1}^{1} \underline{J}^{2(m-1)} (u^{(m)})^2 \, dx$ $(m = 4, 5, \ldots)$, as $\varepsilon \to 0$. The most remarkable feature of Theorem 2 is that as $\varepsilon \to 0$ the dominant terms in the right-hand sides of Eqs. (5.10) and (5.11) can be controlled by choosing suitable J. This is the essential difference of such estimates compared to the available estimates for the conventional spectral methods. It provides a theoretical interpretation for the high accuracy of BLRSMs when $\varepsilon \ll 1$. More precisely, when the conventional spectral method (i.e., without using any transformation) is applied directly to Eq. (5.5), then one only has the following error estimates (cf. Canuto (1988)]

$$\varepsilon \|u' - v'_N\|^2_{L^2(I)} + \|u - U_N\|^2_{L^2_q(I)} \leqslant C(\varepsilon N^{-2} + N^{-4}) \left(\int_{-1}^{1} (u')^2 \, dx + \int_{-1}^{1} (u'')^2 \, dx \right) \tag{5.12}$$

and

$$\varepsilon \|u' - u'_N\|^2_{L^2(I)} + \|u - u_N\|^2_{L^2_q(I)} \leqslant C(\varepsilon N^{-4} + N^{-6}) \times \left(\int_{-1}^{1} (u')^2 \, dx + \int_{-1}^{1} \underline{J}^2 (u'')^2 \, dx + \int_{-1}^{1} (u''')^2 \, dx \right) \tag{5.13}$$

In the presence of a thin boundary layer (i.e., $\varepsilon \ll 1$), the terms $\int_{-1}^{1} (u'')^2\, dx$ and $\int_{-1}^{1} (u''')^2\, dx$ are usually the dominant ones in these error estimates. Similarly, $\int_{-1}^{1} \underline{J}^2 (u'')^2\, dx$ and $\int_{-1}^{1} \underline{J}^4 (u''')^2\, dx$ are dominant terms in Eq. (5.10) and (5.11). However, in many important cases one can show that

$$\int_{-1}^{1} \underline{J}^2 (u'')^2\, dx \ll \int_{-1}^{1} (u'')^2\, dx$$

$$\int_{-1}^{1} \underline{J}^4 (u''')^2\, ds \ll \int_{-1}^{1} (u''')^2\, dx, \qquad \text{as} \quad \varepsilon \to 0 \tag{5.14}$$

which explains why the BLRSMs are superior to the conventional spectral methods for singular perturbation problems, as is further demonstrated by the following example.

We assume that there are positive constants α, ν, C such that

$$|u^{(i)}(x)| \leqslant C + C\varepsilon^{-i/2}(e^{-\alpha(1-x)/\sqrt{\varepsilon}} + e^{-\nu(1+x)/\sqrt{\varepsilon}}), \qquad x \in I, \quad i = 1, 2, \ldots \tag{5.15}$$

where u is the solution of Eq. (5.5).

This assumption is indeed verified by many equations of Helmholtz type [see Ascher (1988)]. Let us consider the following transformation:

$$x = (g(y) = -1 + \sigma_k \int_{-1}^{y} (1 - y^2)^k\, dy \qquad \text{with} \quad k \geqslant 1$$

$$\text{and} \qquad \sigma_k = 2 \Big/ \int_{-1}^{1} (1 - y^2)^k\, dy \tag{5.16}$$

In this case we can show that $J(y) = \sigma_k (1 - y^2)^k$, $J(g^{-1}(x)) \leqslant C(1 - x^2)^{k/(k+1)}$ and $J'(g^{-1}(x)) \leqslant C(1 - x^2)^{(k-1)/(k+1)}$ and so on. Applying Theorem 2, we obtain the following estimates [see Example 2 in Liu and Tang (1994b)]:

$$\varepsilon \|u' - \underline{v}_N'\|_{L^2(I)}^2 + \|u - \underline{v}_N\|_{L_q^2(I)}^2 \leqslant C(N^{-2}\varepsilon + N^{-4})\, \varepsilon^{-3/2 + k/(k+1)}$$

$$\varepsilon \|u' - \underline{v}_N'\|_{L^2(I)}^2 + \|u - \underline{v}_N\|_{L_q^2(I)}^2 \leqslant C(N^{-4}\varepsilon + N^{-6})\, \varepsilon^{-3/2 + (k-1)/(k+1)}$$

Here the constants C may depend on k but not ε. If we choose k sufficiently large, then these upper bounds can be made arbitrarily close to $O(N^{-2}\sqrt{\varepsilon} + N^{-4}\varepsilon^{-1/2})$ and $O(N^{-4}\sqrt{\varepsilon} + N^{-6}\varepsilon^{-1/2})$ respectively. On the contrary, if the conventional spectral method is applied to Eq. (2.1), one can only expect a upper bound of $O(N^{-2}\varepsilon^{-1/2} + N^{-4}\varepsilon^{-3/2})$ [cf. Canuto

(1988)]. Hence, there is a significant improvement in using the BLRSMs when ε is sufficiently small. Moreover, it can be shown that for $m \geqslant 1$,

$$\varepsilon \|u' - \underline{v}_N'\|^2_{L^2(I)} + \|u - \underline{v}_N\|^2_{L^2_q(I)} \leqslant C(N^{-2m}\varepsilon^{1/2-\delta} + N^{-2(m+1)}\varepsilon^{-1/2-\delta})$$

where δ is a positive constant and can be made arbitrarily small if k is sufficiently large. This estimate shows that the error bound of the new scheme can be made almost independent of ε. On the other hand, one only can expect that for $m \geqslant 1$,

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

if the conventional spectral method is applied directly to Eq. (5.5).

Let us now examine the convection-diffusion equation:

$$-\varepsilon u''(x) + p(x)\,u'(x) + q(x)\,u(x) = f(x, \varepsilon), \qquad x \in I, \quad u(\pm 1) = 0 \tag{5.17}$$

with $c(x) = -p'(x)/2 + q(x) \geqslant 0$ for $x \in I$. In this case Eq. (1.2) reads:

$$-\varepsilon(a(y)\,v(y)')' + P(y)\,v'(y) + Q(y)\,v(y) = F(y, \varepsilon), \qquad y \in I, \quad v(\pm 1) = 0 \tag{5.18}$$

where $P(y) = p \circ g(y)$ and a, Q, F are defined as in Eq. (5.6). We use the assumption $c(x) \geqslant 0$ because it makes the analysis simpler and yet can cover many useful cases. Let us denote

$$D(v, w) = \varepsilon \int_{-1}^{1} av'w'\,dy + \int_{-1}^{1} Pv'w\,dy + \int_{-1}^{1} Qvw\,dy, \qquad \text{for } v, w \in GH_0^1(I)$$

The weak formulation then can be stated as follows:

Find $v \in GH_0^1(I)$ such that

$$D(v, w) = \int_{-1}^{1} Fw\,dy, \qquad \forall w \in GH_0^1(I) \tag{5.19}$$

The new spectral Galerkin approximation of Eq. (5.19) in Y_N is:

Find $v_N \in Y_N$ such that

$$D(v_N, w) = \int_{-1}^{1} Fw\,dy, \qquad \forall w \in Y_N \tag{5.20}$$

Theorem 3. [cf. Liu and Tang (1994a,b)]. Let $c(x) \geqslant 0$ on I. The Eqs. (5.19) and (5.20) are well posed in $GH_0^1(I)$ and Y_N respectively. Let $u(x)$ and $v_N(y)$ be respectively the solution of Eq. (5.17) and Eq. (5.20).

We assume that there exist C_1, C_2, $\beta > 0$ such that $C_1 \leqslant J(y)(1-y^2)^{-\beta} \leqslant C_2$ for $y \in I$. Then the following estimates hold:

$$\varepsilon \|u' - \underline{v}_N'\|^2_{L^2(I)} + \|u - \underline{v}_N\|^2_{L^2_c(I)} \leqslant C(N^{-2}\varepsilon + N^{-4}\varepsilon^{-1}) \left(\int_{-1}^{1} (u')^2 \, dx + \int_{-1}^{1} \underline{J}^2 (u'')^2 \, dx \right) \tag{5.21}$$

$$\varepsilon \|u' - \underline{v}_N'\|^2_{L^2(I)} + \|u - \underline{v}_N\|^2_{L^2_c(I)} \leqslant C(N^{-4}\varepsilon + N^{-6}\varepsilon^{-1}) \left(\int_{-1}^{1} (u')^2 \, dx + \int_{-1}^{1} \underline{J}^2 (u'')^2 \, dx + \int_{-1}^{1} \underline{J}^4 (u''')^2 \, dx + \int_{-1}^{1} (\underline{J}')^2 (u'')^2 \, dx \right) \tag{5.22}$$

where $\underline{v}_N(x) = v_N(g^{-1}(x))$, $\underline{J}(x) = J(g^{-1}(x))$ and $\underline{J}'(x) = J'(g^{-1}(x))$.

Furthermore, if $c(x) > 0$ on I, we have

$$\varepsilon \|u' - \underline{v}_N'\|^2_{L^2(I)} + \|u - \underline{v}_N\|^2_{L^2_c(I)} \leqslant C(N^{-2}\varepsilon + N^{-4}) \left(\int_{-1}^{1} (u')^2 \, dx + \int_{-1}^{1} \underline{J}^2 (u'')^2 \, dx \right) \tag{5.23}$$

$$\varepsilon \|u' - \underline{v}_N'\|^2_{L^2(I)} + \|u - \underline{v}_N\|^2_{L^2_c(I)} \leqslant C(N^{-4}\varepsilon + N^{-6}) \left(\int_{-1}^{1} (u')^2 \, dx + \int_{-1}^{1} \underline{J}^2 (u'')^2 \, dx + \int_{-1}^{1} \underline{J}^4 (u''')^2 \, dx + \int_{-1}^{1} (\underline{J}')^2 (u'')^2 \, dx \right) \tag{5.24}$$

These results can also be generated to higher order cases. In fact, if the boundary layers of Eq. (5.17) are of width $O(\varepsilon)$ and if the transformation in Eq. (5.16) is used, we can show that for $m \geqslant 1$ and $\gamma > 0$ (the proof for $m = 1$ and 2, can be found in Liu and Tang (1994b), and the other case can be shown in a similar way),

$$\varepsilon \|u' - \underline{v}_N'\|^2_{L^2(I)} + \|u - \underline{v}_N\|^2_{L^2(I)} \leqslant C(\gamma; k)(N^{-2m}\varepsilon + N^{-2(m+1)})\, \varepsilon^{-\gamma - 1 - 4m/(k+1)} \tag{5.25}$$

If we choose k large enough and a small enough, then the dominant term on the right-hand side of Eq. (5.25) behaves like $N^{-2m}\varepsilon^{-1}$. For the conventional spectral methods, the dominated term in the error bound can only

be shown to behave like $N^{-2(m+1)}\varepsilon^{-m}$. We note that even this last statement was only proved [cf. Canuto (1988)] for the special cases where p and q are constants. Thus, the BLRSMs do provide a substantial improvement over the conventional spectral methods when $\varepsilon \ll 1$.

We note that very recently Schwab and Suri (1996) has obtained uniform (in ε) error estimates for the *hp* version of the finite element method for the model boundary layer function $u(x) = \exp(-(ax/\varepsilon))$.

REFERENCES

Ascher, U. M., Christiansen, J., and Russell, R. D. (1994). A collocation solver for mixed order systems of boundary problems. *Math. Comp.* **33**, 659–679.

Ascher, U., Mattheij R. M., and Russell, R. (1988). *Numerical Solution of Boundary Value Problems for Ordinary Differential Equation*, Prentice-Hall, New Jersey.

Canuto, C. (1988). Spectral methods and a maximum principle. *Math. Comp.* **51**, 615–629.

Canuto, C., Hussaini, M. Y., Quarteroni, A., and Zang, T. A. (1988). Spectral methods in fluid dynamics, *Series of Computational Physics*, Springer-Verlag.

Eisen, H., and Heinrichs, W. (1992). A new method of stabilization for singular perturbation problems with spectral methods. *SIAM J. Numer. Anal.* **29**, 107–122.

Funaro, D. (1992). *Polynomial Approximation of Differential Equations*, Springer-Verlag.

Gottlieb, D. and Orszag, S. A. (1977). *Numerical Analysis of Spectral Methods: Theory and Applications*, SIAM-CBMS.

Greengard, L. (1991). Spectral integration and two-point boundary value problem. *SIAM J. Numer. Anal.* **28**, 1071–1080.

Hughes, T. J. R. (ed.) (1979). *Finite Element Method for Convection Dominated Flows*, AMD Vol. 34, *Am. Soc. Mech. Eng.*

Kalinay De Rivas, E. (1972). On the use of nonlinear grids in finite-difference equations. *J. Comput. Phys.* **10**, 202–210.

Liu, W. B., and Tang, T. (1994a). A new boundary layer resolving spectral method. *AMS Proc. Symp. in Applied Math.* **48** (Gautschi, W. (ed.)), 323–326.

Liu, W. B., and Tang, T. (1994b). Error analysis for boundary resolving spectral methods. Research Report 94-04, Dept. of Math. and Stats, Simon Fraser University B.C., Canada.

Mackenzie, J. A. and Morton, K. W. (1990). Finite volume solution of convection-diffusion test problems. *Math. Comp.* **60**, 189–220.

Oriordan, E., and Stynes, M. (1991). A Globally uniformly convergent finite element method for a singularly perturbed elliptic problem in two dimensions. *Math. Comp.* **57**, 47–62.

Orszag, S. A., and Israeli, M. (1974). Numerical simulation of viscous incompressible flows. *Ann. Rev. Fluid Mech.* **6**, 281–318.

Schwab, C., and Suri, M. (1996). The *p* and *hp* versions of the finite element method for problems with boundary layers. *Math. Comp.* **65**, 1403–1429.

Shen, J. (1994). Efficient spectral-Galerkin method I. Direct solvers for the second and fourth order equations using Legendre polynomials. *SIAM J. Sci. Comput.* **15**, 1489–1505.

Shen, J. (1996). Efficient Chebychev–Legendre–Galerkin methods for elliptic problems. Proceedings of ICOSAHOM'95, 233–239, *Houston J. Math.*

Szegö, G. (1975). *Orthogonal Polynomials*, AMS Coll. Publ. Vol. 23, Providence.

Tang, T., and Trummer, M. R. (1996). Boundary layer resolving pseudospectral method for singular perturbation problems. *SIAM J. Sci. Comput.* **17**, 430–438.

Printed in Belgium