Science in China Series A: Mathematics
Aug., 2008, Vol. 51, No. 8, 1391–1400
www.scichina.com math.scichina.com
www.springerlink.com

High order finite volume methods for singular perturbation problems

CHEN ZhongYing[1†], HE ChongNan[1,2] & WU Bin[1]

[1] Department of Scientific Computing and Computer Applications, Sun Yat-Sen University, Guangzhou 510275, China
[2] College of Mathematics and Computer Science, Guangxi University for Nationalities, Nanning 530006, China
(email: lnsczy@mail.sysu.edu.cn, hechn@163.com, wubin@mail.sysu.edu.cn)

Abstract In this paper we establish a high order finite volume method for the fourth order singular perturbation problems. In conjunction with the optimal meshes, the numerical solutions resulting from the method have optimal convergence order. Numerical experiments are presented to verify our theoretical estimates.

Keywords: finite volume methods, optimal meshes, singular perturbation problems.

MSC(2000): 65L10, 65L12, 65L60

1 Introduction

Numerical methods for solving singular perturbation problems have received considerable attention (see, for example, [1–11]). Difficulties in developing efficient algorithms for solving these problems are due to the effect of the boundary layers. Various numerical methods for them are presented and extensively discussed in [12]. These include finite difference and finite element methods and the use of special meshes. Recently Liu and Xu[11] proposed an optimal Galerkin method for solving the singularly perturbed high-order elliptic two-point boundary value problems of reaction-diffusion type using Hermite splines with knots adapted to the boundary layer behavior of the solution. They gave a simple sufficient condition on the mesh sizes and constructed an optimal mesh to obtain the optimal order of uniform convergence of the numerical solutions based on the Hermite splines built on the mesh.

We introduce in this paper an optimal finite volume method using Hermite elements with the optimal mesh for solving the singularly perturbed two-point boundary value problems, in which we combine ideas of mesh designs from [4, 5, 9, 11, 13, 14] and those of high order finite volume methods from [15]. A simple computing scheme is presented in this paper. The optimal order of uniform convergence is obtained under a much weaker condition than coercivity assumption, which is imposed in [11] for the analysis of Galerkin methods on the same mesh. Numerical experiments show that the finite volume method has the same accuracy as Galerkin methods.

Received April 14, 2008; accepted May 27, 2008
DOI: 10.1007/s11425-008-0120-1
† Corresponding author
This work was supported by the National Natural Science Foundation of China (Grant Nos. 10771224, 10601070), and the Guangdong Provincial Natural Science Foundation of China (Grant No. 05003308), the MOE Project of Key Research Institute of Humanities and Social Sciences at Universities, and China-France-Russia Mathematics Collaboration (Grant No. 34000-3275100).

This paper is organized into six sections. In Section 2, we outline a setting of reaction-diffusion type problems and describe the finite volume method for solving such problems. We review in Section 3 the optimal mesh which ensures the optimal order of uniform convergence of the projections. In Section 4, we prove several useful properties of the bilinear forms which are introduced for the definition of the finite volume scheme. In Section 5, these properties will be quoted for the analysis of the convergence order of the method. Finally, in Section 6, we present numerical experiments to demonstrate the performance of the numerical scheme and confirm the theoretical estimates.

2 Finite volume method based on Hermite cubic elements

In this section we describe the high-order finite volume method based on Hermite cubic elements for solving fourth-order singular perturbation problems.

Denote $I := (0,1)$ and $\mathbb{N} := \{1,2,\dots\}$. Let $L^2(I)$ be the linear space of real-value square integrable functions on I with the associated inner product $(\cdot,\cdot)$ and norm $\|\cdot\|$, $L^\infty(I)$ be the linear space of real-value essentially bounded measurable functions on I with the norm $\|\cdot\|_\infty$, $H^m(I)$ and $W^{m,\infty}(I)$, $m=1,2$ be the Sobolev spaces on I with the inner product $(\cdot,\cdot)_m$, norm $\|\cdot\|_m$ and norm $\|\cdot\|_{m,\infty}$, respectively. The semi-norms of $H^m(I)$ are denoted by $|\cdot|_m$, $m=1,2$. Moreover, let $C_0^\infty(I)$ be the linear space of infinitely differentiable functions on I with compact support, and $H_0^2(I)$ be the closure of $C_0^\infty(I)$ in the norm $\|\cdot\|_2$ (cf. [16]).

Assume that $p \in W^{1,\infty}(I), p(x) \geqslant p_{\min} > 0$ for all $x \in I$, $q,r \in L^\infty(I), f \in L^2(I)$, and $\varepsilon \in (0,1]$ is a perturbation parameter. We introduce differential operator L_ε by

$$(L_\varepsilon u)(x) := \varepsilon^2 u^{(4)}(x) - (p(x)u'(x))' + q(x)u'(x) + r(x)u(x),$$

and consider the boundary value problem of the fourth-order differential equation

$$\begin{cases} (L_\varepsilon u)(x) = f(x), & x \in (0,1), \\ u^{(j)}(0) = u^{(j)}(1) = 0, & j = 0,1. \end{cases} \tag{2.1}$$

By defining the bilinear forms $a(u,v) := (u'',v'')$, $b(u,v) := (-(pu')' + qu' + ru, v)$ and $A_\varepsilon(u,v) := \varepsilon^2 a(u,v) + b(u,v)$, the variational problem corresponding to problem (2.1) is identified as: Find $u \in U := H_0^2(I)$, such that for all $v \in U$,

$$A_\varepsilon(u,v) = (f,v). \tag{2.2}$$

We now consider numerically solving the variational problem (2.2) by the finite volume method. Let T_N be a mesh of I with the nodes $0 = x_0 < x_1 < \cdots < x_N = 1$, where $N \in \mathbb{N}$. For any $i \in \mathbb{Z}_N := \{1,2,\dots,N\}$, the length of the element $I_i := [x_{i-1}, x_i]$ is denoted by $h_i := x_i - x_{i-1}$.

In order to discretize (2.2), we choose the trial space U_N as the Hermite cubic element space with respect to T_N, which consists of the functions u_N with the properties

(i) $u_N \in C^1(I), u_N^{(j)}(0) = u_N^{(j)}(1) = 0, j = 0,1;$

(ii) u_N is a cubic polynomial on each I_i, $i \in \mathbb{Z}_N$.

It is easily seen that dim $U_N = 2(N-1)$. Introduce the basis $\{\varphi_{i,0}, \varphi_{i,1} : i \in \mathbb{Z}_{N-1}\}$ of U_N by

$$\varphi_{i,0}(x) := \begin{cases} (1 - h_i^{-1}|x - x_i|)^2(2h_i^{-1}|x - x_i| + 1), & x \in [x_{i-1}, x_i], \\ (1 - h_{i+1}^{-1}|x - x_i|)^2(2h_{i+1}^{-1}|x - x_i| + 1), & x \in (x_i, x_{i+1}], \\ 0, & x \notin [x_{i-1}, x_{i+1}], \end{cases}$$

and

$$\varphi_{i,1}(x) := \begin{cases} (x - x_i)(h_i^{-1}|x - x_i| - 1)^2, & x \in [x_{i-1}, x_i], \\ (x - x_i)(h_{i+1}^{-1}|x - x_i| - 1)^2, & x \in (x_i, x_{i+1}], \\ 0, & x \notin [x_{i-1}, x_{i+1}]. \end{cases}$$

With the Hermite interpolatory basis functions defined above, any $u_N \in U_N$ can be written uniquely as

$$u_N = \sum_{i \in \mathbb{Z}_{N-1}} (u_i \varphi_{i,0} + u_i' \varphi_{i,1}), \tag{2.3}$$

where $u_i = u_N(x_i)$ and $u_i' = u_N'(x_i)$.

The test space $V_N \subset L^2(I)$ is associated with the dual mesh T_N^* of T_N with the nodes $0 = x_0 < x_{1/2} < x_{3/2} < \cdots < x_{N-1/2} < x_N = 1$, where $x_{j-1/2} = (x_{j-1} + x_j)/2,\ j \in \mathbb{Z}_N$. We define V_N as the piecewise linear polynomial space with respect to T_N^*. The basis $\{\psi_{j,0}, \psi_{j,1} : j \in \mathbb{Z}_{N-1}\}$ of V_N is chosen by defining for $j \in \mathbb{Z}_{N-1}$,

$$\psi_{j,0}(x) := \begin{cases} 1, & x \in [x_{j-1/2}, x_{j+1/2}], \\ 0, & x \notin [x_{j-1/2}, x_{j+1/2}], \end{cases} \qquad \psi_{j,1}(x) := \begin{cases} x - x_j, & x \in [x_{j-1/2}, x_{j+1/2}], \\ 0, & x \notin [x_{j-1/2}, x_{j+1/2}]. \end{cases}$$

We note that the bilinear form $a(u, v)$ is not defined for $v \in V_N$. In order to establish the finite volume methods for solving (2.1), we require another bilinear form to lead to a weak formulation of (2.1). To this end, we introduce the bilinear form $\tilde{a}(\cdot, \cdot) : (U_N \cup H^4(I)) \times V_N$ by setting for $u \in U_N \cup H^4(I)$

$$\tilde{a}(u, \psi_{j,0}) = u'''_{j+1/2} - u'''_{j-1/2},$$
$$\tilde{a}(u, \psi_{j,1}) = u''_{j-1/2} - u''_{j+1/2} + \frac{h_j}{2} u'''_{j-1/2} + \frac{h_{j+1}}{2} u'''_{j+1/2}.$$

The evaluation of the bilinear form for any $v \in V_N$ is done by combining linearly the above expressions. It is easily observed that for $u \in H^4(I)$ and $v \in V_N$, $\tilde{a}(u, v) = (u^{(4)}, v)$. Note that the bilinear form $b(\cdot, \cdot)$ can be identified as defined on $(U_N \cup H^4(I)) \times V_N$, thus $\tilde{A}_\varepsilon(u, v) := \varepsilon^2 \tilde{a}(u, v) + b(u, v)$ is a bilinear form on $(U_N \cup H^4(I)) \times V_N$. Then the Hermite cubic element finite volume method is: Find $u_N \in U_N$ such that for all $v \in V_N$

$$\tilde{A}_\varepsilon(u_N, v) = (f, v). \tag{2.4}$$

3 Optimal meshes on I

It is well known that the accuracy of the solution u_N of (2.4) is constrained by the approximation accuracy of the trial space U_N. Thus a good trial space is the first step of good numerical schemes. For singular perturbation problems, small ε will produce boundary layers near the end points, which brings difficulties to the construction of meshes to guarantee good approximation to the true solutions of the problems. Specifically, according to [17], the solution u of (2.1) can be written as

$$u = E + F + G, \tag{3.1}$$

where the functions E, F, G are sufficiently differentiable such that for all $x \in I$ and $j \in \{0, 1, 2, \ldots\}$,

$$|G^{(j)}(x)| \leqslant c, \quad |E^{(j)}(x)| \leqslant c\varepsilon^{1-j}\mathrm{e}^{-\alpha x/\varepsilon}, \quad |F^{(j)}(x)| \leqslant c\varepsilon^{1-j}\mathrm{e}^{-\alpha(1-x)/\varepsilon} \tag{3.2}$$

for some constants $c > 0$ and $\alpha > 0$. In the whole paper we use c to denote a generic positive constant. Many authors have tried to overcome the difficulties induced by this type of boundary layers. Bakhvalov[18] introduced special grids based on mesh generating functions. The Shishkin mesh is one of the simplest piecewise equidistant meshes and was discussed by many authors (cf. [2, 4–7]). A recent development in designing meshes for singular perturbation problems was made in [11]. The authors introduced the so-called optimal mesh to ensure the uniform convergence in optimal order of the projection errors of u on U_N. For the numerical solution of (2.1) by cubic element Galerkin methods, the optimal mesh is constructed by defining a generating function $h^0(x) := \frac{\varepsilon}{N}\mathrm{e}^{\frac{\alpha x}{4\varepsilon}}$, and then for $i \in \mathbb{Z}_{\tilde{N}}$, letting the length h_i satisfy

$$h_i \leqslant \min\{h^0(x_{i-1}), h^0(1 - x_i), 1/N\}, \tag{3.3}$$

where the number of elements $\tilde{N} = \mathcal{O}(N)$. The interested readers can find the concrete construction to satisfy the above conditions in [11].

For any $N \in \mathbb{N}$, we choose T_N as the optimal mesh described above and denote by T_N^* the dual mesh of T_N. For $u \in U$, we let $\Pi_N u$ and $\Pi_N^* u$ be the Hermite interpolants of u on U_N and V_N respectively, i.e.,

$$\Pi_N u := \sum_{i \in \mathbb{Z}_{N-1}} [u(x_i)\varphi_{i,0} + u'(x_i)\varphi_{i,1}], \quad \text{and} \quad \Pi_N^* u := \sum_{i \in \mathbb{Z}_{N-1}} [u(x_i)\psi_{i,0} + u'(x_i)\psi_{i,1}].$$

Let $|v|_{k,\infty,I_i}$ denote the maximum norm of $v^{(k)}$ on I_i. Then for $j = 0, 1, 2, 3$,

$$|E - \Pi_N E|_{j,\infty,I_i} \leqslant h_i^{4-j}|E|_{4,\infty,I_i} \leqslant c\varepsilon^{-3}\mathrm{e}^{\frac{\alpha x_{i-1}}{\varepsilon}}\left(\frac{\varepsilon}{N}\mathrm{e}^{\frac{\alpha x_{i-1}}{4\varepsilon}}\right)^{4-j} \leqslant c\varepsilon^{1-j}N^{-4+j}.$$

Similar inequalities also hold for F and G. Hence we obtain the following estimates:

Lemma 3.1. *Let u be the solution of* (2.1). *Then there exists a positive constant c independent of ε and N such that*

$$\|(u - \Pi_N u)^{(j)}\| \leqslant c\varepsilon^{1-j}N^{-4+j}, \qquad j = 0, 1, 2, 3, \tag{3.4}$$

and

$$\|u - \Pi_N u\|_\varepsilon \leqslant cN^{-2}, \tag{3.5}$$

where the energy norm $\|\cdot\|_\varepsilon$ is defined by $\|u\|_\varepsilon := (\varepsilon^2|u|_2^2 + \|u\|_1^2)^{1/2}$.

4 Properties of bilinear forms

In this section we prove several properties of the bilinear forms we define in the previous section, which help establish convergence analysis in the next section.

It follows from the basis spanning formula (2.3) that

$$\begin{aligned}\tilde{a}(u_N, \psi_{j,0}) = &-12h_j^{-3}u_{j-1} - 6h_j^{-2}u'_{j-1} + 12(h_j^{-3} + h_{j+1}^{-3})u_j \\ &- 6(h_j^{-2} - h_{j+1}^{-2})u'_j - 12h_{j+1}^{-3}u_{j+1} + 6h_{j+1}^{-2}u'_{j+1},\end{aligned} \tag{4.1}$$

and

$$\tilde{a}(u_N,\psi_{j,1}) = 6h_j^{-2}u_{j-1} + 2h_j^{-1}u'_{j-1} - 6(h_j^{-2} - h_{j+1}^{-2})u_j + 4(h_j^{-1} + h_{j+1}^{-1})u'_j - 6h_{j+1}^{-2}u_{j+1} + 2h_{j+1}^{-1}u'_{j+1}. \tag{4.2}$$

In a similar manner, we have

$$(-(pu'_N)',\psi_{j,0}) = \frac{3p_{j-\frac12}}{2h_j}(u_j - u_{j-1}) - \frac{3p_{j+\frac12}}{2h_j}(u_{j+1} - u_j) - \frac{p_{j-\frac12}}{4}u'_{j-1} + \frac{p_{j+\frac12} - p_{j-\frac12}}{4}u'_j + \frac{p_{j+\frac12}}{4}u'_{j+1}, \tag{4.3}$$

and

$$(-(pu'_N)',\psi_{j,1}) = -\frac{3p_{j-\frac12}}{4}(u_j - u_{j-1}) - \frac{3p_{j+\frac12}}{4}(u_{j+1} - u_j) + \frac{p_{j-\frac12}h_j}{8}u'_{j-1} + \frac{p_{j-\frac12}h_j + p_{j+\frac12}h_{j+1}}{8}u'_j + \frac{p_{j+\frac12}h_{j+1}}{8}u'_{j+1} + \int_{x_{j-\frac12}}^{x_{j+\frac12}} pu'_N dx. \tag{4.4}$$

Lemma 4.1. *For sufficiently large N, there exists a positive constant α independent of U_N, such that*

$$\tilde{a}(u_N,\Pi_N^* u_N) \geqslant \alpha|u_N|_2^2, \quad (-(pu'_N)',\Pi_N^* u_N) \geqslant \alpha|u_N|_1^2, \qquad u_N \in U_N.$$

Proof. It follows from (2.3), (4.1) and (4.2) that

$$\tilde{a}(u_N,\Pi_N^* u_N) = \sum_{j\in\mathbb{Z}_N} h_j\left[3\left(\frac{u'_{j-1} + u'_j - 2h_j^{-1}(u_j - u_{j-1})}{h_j}\right)^2 + \left(\frac{u'_j - u'_{j-1}}{h_j}\right)^2\right] \geqslant |u_N|_{2,U_N}^2,$$

where the discrete semi-norm

$$|u_N|_{2,U_N} := \left\{\sum_{j\in\mathbb{Z}_N} h_j\left[\left(\frac{u'_{j-1} + u'_j - 2h_j^{-1}(u_j - u_{j-1})}{h_j}\right)^2 + \left(\frac{u'_j - u'_{j-1}}{h_j}\right)^2\right]\right\}^{1/2}.$$

It is shown in [15] that there are positive constants c_1 and c_2 which are independent of the subspaces U_N such that $c_1|u_N|_{2,U_N} \leqslant |u_N|_2 \leqslant c_2|u_N|_{2,U_N}$, for all $u_N \in U_N$. Hence we conclude the first inequality of the lemma. The second estimate can be established in a similar way with the help of (4.3) and (4.4).

Lemma 4.2. *For $u \in W^{4,\infty}(I)$ which has the decomposition* (3.1) *with the property* (3.2), *there exists a positive constant c independent of N such that for any $w \in U_N$,*

$$|\tilde{a}(u - \Pi_N u,\Pi_N^* w)| \leqslant c\varepsilon^{-1}N^{-2}|w|_2.$$

Proof. We denote $e_N := u - \Pi_N u$, then utilize the definition of the interpolant $\Pi_N^* w$ and (4.1), (4.2) to obtain

$$\tilde{a}(e_N,\Pi_N^* w) = \sum_{j\in\mathbb{Z}_N}\left[\frac12 h_j e'''_{j-1/2}(w'_{j-1} + w'_j - 2h_j^{-1}(w_j - w_{j-1})) + e''_{j-1/2}(w'_j - w'_{j-1})\right].$$

Applying Cauchy's inequality to the above expression, we have

$$|\tilde{a}(e_N,\Pi_N^* w)| \leqslant c\left[\sum_{j\in\mathbb{Z}_N} h_j((e'''_{j-1/2})^2 h_j^2 + (e''_{j-1/2})^2)\right]^{1/2} \cdot |w|_{2,U_N}.$$

It follows that $|e'''_{j-1/2}| \leqslant |e_N|_{3,\infty,I_j} \leqslant ch_j|u|_{4,\infty,I_j}$. The property (3.2) of the decomposition of u gives

$$|u|_{4,\infty,I_j} \leqslant c[1+\varepsilon^{-3}(\mathrm{e}^{-\alpha x_{j-1}/\varepsilon}+\mathrm{e}^{-\alpha(1-x_j)/\varepsilon})].$$

Utilizing the property (3.3) of h_j, we conclude $(e'''_{j-1/2})^2h_j^2 \leqslant c\varepsilon^{-2}N^{-4}$. Similarly, $(e''_{j-1/2})^2 \leqslant c\varepsilon^{-2}N^{-4}$. Therefore,

$$\left[\sum_{j\in\mathbb{Z}_N} h_j((e'''_{j-1/2})^2h_j^2+(e''_{j-1/2})^2)\right]^{1/2} \leqslant c\varepsilon^{-1}N^{-2}$$

and the lemma is concluded.

Lemma 4.3. *For any $v \in U$ and $u \in U_N$, there holds $a(u,v) = \tilde{a}(u,\Pi_N^* v)$.*

Proof. We obtain from the definition of the bilinear forms that

$$\begin{aligned} a(u,v) &= \sum_{j\in\mathbb{Z}_N}\int_{I_j} u''(x)v''(x)dx \\ &= \sum_{j\in\mathbb{Z}_N}[u''(x_j-)v'(x_j)-u''(x_{j-1}+)v'(x_{j-1})-u'''(x_{j-1/2})(v(x_j)-v(x_{j-1}))], \end{aligned}$$

$$\begin{aligned} \tilde{a}(u,\Pi_N^*v) = \sum_{j\in\mathbb{Z}_{N-1}}&\{[u'''(x_{j+1/2})-u'''(x_{j-1/2})]v(x_j)-[u''(x_{j+1/2})-u''(x_{j-1/2})]v'(x_j) \\ &+\frac{1}{2}[h_{j+1}u'''(x_{j+1/2})+h_ju'''(x_{j-1/2})]v'(x_j)\}. \end{aligned}$$

Resorting the terms in the summation leads to

$$\begin{aligned} \tilde{a}(u,\Pi_N^*v) = \sum_{j\in\mathbb{Z}_N}&\bigg[u''(x_{j-1/2})(v'(x_j)-v'(x_{j-1}))+\frac{h_j}{2}u'''(x_{j-1/2})(v'(x_j)+v'(x_{j-1})) \\ &-u'''(x_{j-1/2})(v(x_j)-v(x_{j-1}))\bigg]. \end{aligned}$$

Since u is piecewise cubic on each interval I_j, there hold

$$u''(x_j-) = u''(x_{j-1/2})+\frac{h_j}{2}u'''(x_{j-1/2}), \qquad u''(x_{j-1}+) = u''(x_{j-1/2})-\frac{h_j}{2}u'''(x_{j-1/2}).$$

We conclude the lemma by making use of the above equalities and comparing the expressions of $a(u,v)$ and $\tilde{a}(u,\Pi_N^*v)$.

5 Convergence analysis

For any $u_N \in U_N$, define

$$\|u_N\|_{L,U_N} := \sup_{w_N\in U_N,\|w_N\|_\varepsilon=1}|\tilde{A}_\varepsilon(u_N,\Pi_N^*w_N)|. \tag{5.1}$$

Lemma 5.1. *Suppose that the corresponding homogeneous equation*

$$A_\varepsilon(u,v) = 0, \qquad v \in U, \tag{5.2}$$

has only trivial solution. Then there exists a positive constant α independent of ε and U_N, such that for sufficiently large N,

$$\|u_N\|_{L,U_N} \geqslant \alpha\|u_N\|_\varepsilon, \tag{5.3}$$

for any $u_N \in U_N$.

Proof. For any given constant $\lambda > 0$, define

$$b_\lambda(u,v) := \lambda(u,v) \quad \text{and} \quad \tilde{A}_{\varepsilon,\lambda}(u,v) := \tilde{A}_\varepsilon(u,v) + b_\lambda(u,v).$$

It follows from Lemma 4.1 that there exists a positive constant β_1 such that

$$\begin{aligned}\tilde{A}_{\varepsilon,\lambda}(u_N, \Pi_N^* u_N) \geqslant &\beta_1(\varepsilon^2 |u_N|_2^2 + |u_N|_1^2) - \|q\|_\infty |u_N|_1 \cdot \|\Pi_N^* u_N\| \\ &+(\lambda - \|r\|_\infty)\|u_N\|^2 - (\lambda + \|r\|_\infty)\|u_N\| \cdot \|\Pi_N^* u_N - u_N\|.\end{aligned}$$

Note that

$$\|\Pi_N^* u_N\| \leqslant \|u_N\| \quad \text{and} \quad \|\Pi_N^* u_N - u_N\| \leqslant \beta_2 N^{-1} \|u_N\|_1$$

for some $\beta_2 > 0$. We have

$$\begin{aligned}&\|q\|_\infty |u_N|_1 \cdot \|\Pi_N^* u_N\| + (\lambda + \|r\|_\infty)\|u_N\| \cdot \|\Pi_N^* u_N - u_N\| \\ &\leqslant (\|q\|_\infty + \beta_2 N^{-1}(\lambda + \|r\|_\infty))\|u_N\|_1 \cdot \|u_N\|.\end{aligned}$$

Picking proper value of λ and sufficiently large N, then applying

$$\|u_N\|_1 \cdot \|u_N\| \leqslant \frac{\delta}{2}\|u_N\|_1^2 + \frac{1}{2\delta}\|u_N\|^2$$

with appropriate constant $\delta > 0$, we obtain

$$\tilde{A}_{\varepsilon,\lambda}(u_N, \Pi_N^* u_N) \geqslant \beta \|u_N\|_\varepsilon^2, \tag{5.4}$$

where the positive constant β is independent of ε and U_N.

Now we assume that no positive constant α is available for (5.3), then we can find a sequence $\{\tilde{u}_N : \tilde{u}_N \in U_N\}$ satisfying $\|\tilde{u}_N\|_\varepsilon = 1$ and $\|\tilde{u}_N\|_{L,U_N} \to 0,\ N \to \infty$. Since U is weakly sequentially compact, we assume without loss of generality that $\{\tilde{u}_N\}$ weakly converges to some $\tilde{u} \in U$. Given any $w \in \tilde{U} := H^3(I) \cap U$, we have $\Pi_N^*(w - \Pi_N w) = 0$, thus

$$|\tilde{A}_\varepsilon(\tilde{u}_N, \Pi_N^* w)| = |\tilde{A}_\varepsilon(\tilde{u}_N, \Pi_N^* \Pi_N w)| \leqslant c\|\tilde{u}_N\|_{L,U_N} \|\Pi_N w\|_\varepsilon.$$

Meanwhile, there holds

$$\|\Pi_N w\|_\varepsilon \leqslant \|w\|_\varepsilon + \|\Pi_N w - w\|_\varepsilon \leqslant c\|w\|_3$$

for some constant $c > 0$ and sufficiently large N. Therefore, $|\tilde{A}_\varepsilon(\tilde{u}_N, \Pi_N^* w)| \leqslant \|\tilde{u}_N\|_{L,U_N} \|w\|_3 \to 0$ when $N \to \infty$. On the other hand,

$$b(\tilde{u}_N, \Pi_N^* w - w) \leqslant cN^{-2} \|\tilde{u}_N\|_2 \|w\|_2 \to 0, \qquad N \to \infty.$$

Moreover, it follows from Lemma 4.3 that

$$\tilde{A}_\varepsilon(\tilde{u}_N, \Pi_N^* w) - A_\varepsilon(\tilde{u}_N, w) = b(\tilde{u}_N, \Pi_N^* w - w).$$

Hence $A_\varepsilon(\tilde{u}_N, w) \to 0,\ N \to \infty$. Since $A_\varepsilon(\cdot, w)$ is a bounded linear functional for any fixed $w \in \tilde{U}$, we have $A_\varepsilon(\tilde{u}_N, w) \to A_\varepsilon(\tilde{u}, w),\ N \to \infty$. Therefore, $A_\varepsilon(\tilde{u}, w) = 0,\ w \in \tilde{U}$. But $\tilde{U}$ is dense in U, thus $A_\varepsilon(\tilde{u}, w) = 0$, $w \in U$. It follows from the assumption on the homogeneous equation that $\tilde{u} = 0$, hence $\{\tilde{u}_N\}$ weakly converges to 0. By noticing U compactly imbeds into $L^2(I)$, we conclude that $\{\tilde{u}_N\}$ converges to 0 in $L^2(I)$. Hence $b_\lambda(\tilde{u}_N, \Pi_N^* \tilde{u}_N) \leqslant \lambda \|\tilde{u}_N\|^2 \to 0$, $N \to \infty$. It follows from the definition (5.1) of the norm $\|\cdot\|_{L,U_N}$ and $\|\tilde{u}_N\|_\varepsilon = 1$ that $|\tilde{A}_\varepsilon(\tilde{u}_N, \Pi_N^* \tilde{u}_N)| \leqslant \|\tilde{u}_N\|_{L,U_N}$. Consequently,

$$|\tilde{A}_{\varepsilon,\lambda}(\tilde{u}_N, \Pi_N^* \tilde{u}_N)| \leqslant |\tilde{A}_\varepsilon(\tilde{u}_N, \Pi_N^* \tilde{u}_N)| + |b_\lambda(\tilde{u}_N, \Pi_N^* \tilde{u}_N)| \leqslant \|\tilde{u}_N\|_{L,U_N} + \lambda \|\tilde{u}_N\|^2 \to 0, \quad N \to \infty,$$

which contradicts with (5.4).

Theorem 5.2. *Suppose that the homogeneous equation* (5.2) *has only trivial solution. Let u and u_N be the solutions of* (2.1) *and* (2.4), *respectively. Then there are a positive integer N_0 and a positive constant c, such that for any $N \geqslant N_0$, $\|u - u_N\|_\varepsilon \leqslant cN^{-2}$.*

Proof. It follows from (2.1) and (2.4) that $\tilde{A}_\varepsilon(u - u_N, \Pi_N^* w_N) = 0$, $w_N \in U_N$. Then it follows from Lemma 5.1 that for $N \geqslant N_0$

$$\begin{aligned}\|\Pi_N u - u_N\|_\varepsilon &\leqslant c \sup_{w_N \in U_N, \|w_N\|_\varepsilon = 1} |\tilde{A}_\varepsilon(\Pi_N u - u_N, \Pi_N^* w_N)| \\ &= c \sup_{w_N \in U_N, \|w_N\|_\varepsilon = 1} |\tilde{A}_\varepsilon(u - \Pi_N u, \Pi_N^* w_N)|.\end{aligned}$$

We let $e_N := u - \Pi_N u$ and decompose $\tilde{A}_\varepsilon(e_N, \Pi_N^* w_N)$ into three parts:

$$\tilde{A}_\varepsilon(e_N, \Pi_N^* w_N) = \varepsilon^2 \tilde{a}(e_N, \Pi_N^* w_N) + b(e_N, w_N) + b(e_N, \Pi_N^* w_N - w_N).$$

For the first part, we have from Lemma 4.2 that $|\tilde{a}(e_N, \Pi_N^* w_N)| \leqslant c\varepsilon^{-1} N^{-2} |w_N|_2$. For the second part, since $e_N, w_N \in H^1(I)$, there holds

$$|b(e_N, w_N)| \leqslant c\|e_N\|_1 \|w_N\|_1 \leqslant cN^{-3} \|w_N\|_1,$$

where the second inequality results from Lemma 3.1. For the third part,

$$\begin{aligned}|b(e_N, \Pi_N^* w_N - w_N)| &= \sum_{j \in \mathbb{Z}_N} \int_{I_j} [-(pe_N')' + qe_N' + re_N](x)(\Pi_N^* w_N - w_N)(x) dx \\ &\leqslant c \sum_{j \in \mathbb{Z}_N} \left(\int_{I_j} [e_N''(x)]^2 dx \right)^{1/2} \left(\int_{I_j} (\Pi_N^* w_N - w_N)^2(x) dx \right)^{1/2}.\end{aligned}$$

Similar to the proof of Lemma 4.2, we have $|e_N''(x)| \leqslant ch_j^2 |u|_{4,\infty,I_j}$ for any $x \in I_j$, hence

$$\left(\int_{I_j} [e_N''(x)]^2 dx \right)^{1/2} \leqslant ch_j^{1/2} h_j^2 |u|_{4,\infty,I_j}.$$

Moreover,

$$\left(\int_{I_j} (\Pi_N^* w_N - w_N)^2(x) dx \right)^{1/2} \leqslant ch_j |w_N|_{1,I_j},$$

where $|w_N|_{1,I_j} := (\int_{I_j} [w_N'(x)]^2 dx)^{1/2}$. It follows from (3.3) that $h_j^3 |u|_{4,\infty,I_j} \leqslant cN^{-3}$. Therefore,

$$\begin{aligned}|b(e_N, \Pi_N^* w_N - w_N)| &\leqslant cN^{-3} \sum_{j \in \mathbb{Z}_N} h_j^{1/2} |w_N|_{1,I_j} = cN^{-3} \sum_{j \in \mathbb{Z}_N} \left(\int_{I_j} 1 \cdot dx \right)^{1/2} |w_N|_{1,I_j} \\ &\leqslant cN^{-3} \left(\sum_{j \in \mathbb{Z}_N} \int_{I_j} 1 \cdot dx \right)^{1/2} \left(\sum_{j \in \mathbb{Z}_N} |w_N|_{1,I_j}^2 \right)^{1/2} = cN^{-3} |w_N|_1.\end{aligned}$$

Summing up the three parts concludes

$$|\tilde{A}_\varepsilon(u-\Pi_N u, \Pi_N^* w_N)| \leqslant cN^{-2}(\varepsilon|w_N|_2 + N^{-1}\|w_N\|_1) \leqslant cN^{-2}\|w_N\|_\varepsilon.$$

Hence the proof is complete.

We remark the assumption that the homogeneous equation has only trivial solution is very weak. In [11], the authors assumed the coercivity of the bilinear form, which is a strong condition and results to our assumption. In finite dimensional case, the coercivity corresponds to positive definiteness, while our assumption corresponds to nonsingularity of the coefficient matrix. In fact, the effectiveness of Galerkin methods, such as developed in [11], can be concluded with this much weaker condition.

6 Numerical experiments

We solve in this section with our proposed numerical scheme the reaction–diffusion problem

$$\begin{cases} \varepsilon^2 u^{(4)} + ((1+x(1-x)u')' = f(x), \quad x \in (0,1), \\ u(0) = u'(0) = u(1) = u'(1) = 0. \end{cases}$$

The right hand function f is chosen so that the exact solution of the problem

$$u(x) = \varepsilon\left(\frac{\mathrm{e}^{-x/\varepsilon} + \mathrm{e}^{-(1-x)/\varepsilon}}{1+\mathrm{e}^{-1/\varepsilon}} - 1\right) + \frac{1-\mathrm{e}^{-1/\varepsilon}}{1+\mathrm{e}^{-1/\varepsilon}} x(1-x) + x^2(1-x)^2.$$

We observe from the expression of u that there are boundary layers near 0 and 1. This equation has been used as an example in [11], which brings us convenience to compare the numerical results between our method and the corresponding Galerkin scheme. In order to create the same mesh as that in [11], we use the generating function $h^0(x) = \frac{4\varepsilon}{N}\mathrm{e}^{\frac{x}{4\varepsilon}}$.

We compare the numerical results of our method and Galerkin method in Table 1.

Table 1 Comparison of finite volume method and Galerkin Method

ε	N	Finite Volume		Galerkin Method	
		$\|u-u_N\|_\varepsilon$	Order	$\|u-u_N\|_\varepsilon$	Order
3.905e–3	63	1.21e–004		1.13e–004	
	127	2.97e–005	2.00	2.72e–005	2.03
	255	7.18e–006	2.04	6.78e–006	1.99
	511	1.77e–006	2.02	1.68e–006	2.01
	1026	4.36e–007	2.01	4.17e–007	2.00
60104e–5	63	2.57e–005		1.79e–005	
	127	4.75e–006	2.40	3.97e–006	2.14
	255	1.06e–006	2.15	9.58e–007	2.04
	513	2.49e–007	2.07	2.34e–007	2.02
	1023	6.14e–008	2.02	5.86e–008	2.01
3.816e–6	64	1.86e–005		8.65e–006	
	126	2.43e–006	3.00	1.33e–006	2.76
	254	3.39e–007	2.80	2.61e–007	2.32
	510	6.34e–008	2.41	6.07e–008	2.09
	1023	1.42e–008	2.14	1.48e–008	2.02

We choose three small values for the perturbation parameter ε and calculate the errors of the numerical solutions u_N with different values of N. Meanwhile, the numerical errors of Galerkin method are listed for comparison. We note that the data of Galerkin method are collected from [11]. It is observed that the two methods have nearly the same accuracy. The "Order" columns present the convergence order calculated from the numerical errors, the values of which are created by the formula

$$\log\left(\frac{\|u-u_{N_1}\|_\varepsilon}{\|u-u_{N_2}\|_\varepsilon}\right)/\log\left(\frac{N_2}{N_1}\right),$$

where N_1 and N_2 are the two successive numbers of grid points in the table. It is seen that the values in the columns are even a little better than the theoretical value 2, which is claimed by Theorem 5.2.

References

1 Stynes M, O'Riordan E. A finite element method for a singularly perturbed boundary value problem. *Numer Math*, **50**: 1–15 (1986)

2 Shishkin G I. Grid approximation of singularly perturbed parabolic equations with internal layers. *Sov J Numer Anal Math Model*, **3**: 393–407 (1988)

3 Stynes M, O'Riordan E. An analysis of a singularly perturbed two-point boundary value problem using only finite element techniques. *Math Comp*, **56**: 663–675 (1991)

4 Sun G, Stynes M. Finite element methods for singularly perturbed high-order elliptic two-point boundary problems, I: Reaction-diffusion-type problems. *IMA J Numer Anal*, **15**: 117–139 (1995)

5 Sun G, Stynes M. Finite element methods for singularly perturbed high-order elliptic two-point boundary problems, II: Convection-diffusion-type problems. *IMA J Numer Anal*, **15**: 197–219 (1995)

6 Miller J J H, O'Riordan E, Shishkin G I. Fitted Numerical Methods for Singularly Perturbation Problems. Singapore: World Scientific, 1996

7 Roos H G, Stynes M, Tobiska L. Numerical Methods for Singularly Perturbed Differential Equations. Berlin: Springer-Verlag, 1996

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

9 Qiu Y, Sloan D M. Analysis of difference approximations to a singular perturbed two-point boundary value problem on an adaptively generated grid. *J Comput Appl Math*, **101**: 1–25 (1999)

10 Liu W, Tang T. Error analysis for a Galerkin-spectral method with coordinate transformation for solving singularly perturbed problems. *Appl Numer Math*, **38**: 315–345 (2001)

11 Liu S, Xu Y. Galerkin methods based on Hermite splines for singular perturbation problems. *SIAM J Number Anal*, **43**: 2607–2623 (2006)

12 Stynes M. Steady-state convection-diffusion problems. *Acta Numer*, **14**: 445–508 (2005)

13 Cao Y Z, Herdman T, Xu Y. A hybrid collocation method for Volterra integral equations with weakly singular kernels. *SIAM J Numer Anal*, **41**: 364–381 (2003)

14 Vulanovic R. Non-equidistant generalizations of the Gushchin-Shchennikov scheme. *Z Angew Math Mec*, **67**: 625–632 (1987)

15 Li R, Chen Z, Wu W. Generalized Difference Methods for Differential Equations: Numerical Analysis of Finite Volume Methods. New York-Basel: Marcel Dekker Inc, 2000

16 Adams R A. Sobolev Spaces. New York: Academic Press, 1975

17 O'Malley R E. Introduction to Singular Perturbations. New York: Academic Press, 1974

18 Bakhvalov A S. K optimizacii methdov resenia kraevyh zadac prinalicii pogranicnogo sloja. *Zh Vychisl Mat Mat Fiz*, **9**: 841–859 (1969)