Numer. Math. Theor. Meth. Appl.
doi: 10.4208/nmtma.2012.m1110

Vol. **5**, No. 3, pp. 493-508
August 2012

The Dissipative Spectral Methods for the First Order Linear Hyperbolic Equations

Lian Chen[1], Zhongqiang Zhang[2] and Heping Ma[1,*]

[1] *Department of Mathematics, College of Sciences, Shanghai University, Shanghai, 200444, China.*
[2] *Division of Applied Mathematics, Brown University, Providence, Rhode Island, 02912, USA.*

Received 14 May 2011; Accepted (in revised version) 1 September 2011

Available online 3 July 2012

Abstract. In this paper, we introduce the dissipative spectral methods (DSM) for the first order linear hyperbolic equations in one dimension. Specifically, we consider the Fourier DSM for periodic problems and the Legendre DSM for equations with the Dirichlet boundary condition. The error estimates of the methods are shown to be quasi-optimal for variable-coefficients equations. Numerical results are given to verify high accuracy of the DSM and to compare the proposed schemes with some high performance methods, showing some superiority in long-term integration for the periodic case and in dealing with limited smoothness near or at the boundary for the Dirichlet case.

AMS subject classifications: 65N15, 65N35

Key words: First order hyperbolic equation, dissipative spectral method, error estimate.

1. Introduction

Consider spectral methods for the following one dimensional first-order linear hyperbolic equation

$$\partial_t U + a(x)\partial_x U + b(x)U = f(x,t), \quad x \in D,\ t \in (0,T], \quad U(x,0) = U_0(x) \tag{1.1}$$

with appropriate boundary conditions.

Due to the lack of symmetry, the Galerkin spectral method (GSM) does not seem ideal for odd-order partial differential equations [2]. For better resolution, these linear hyperbolic equations are proposed to be solved with the Petrov–Galerkin spectral methods (PGSM), such as the tau method [2] and the recently developed dual-Petrov–Galerkin spectral method [21], or the skew-symmetry decomposition technique in the Galerkin method [2], the penalty method [10].

*Corresponding author. *Email addresses:* chenlianice@163.com (L. Chen), handyzang@gmail.com (Z. Zhang), hpma@shu.edu.cn (H. Ma)

Here we consider the dissipative spectral methods (DSM) for the equation (1.1), which is a streamline upwind strategy, taking $v+c\partial_x v$ (c depends on the discretization parameter in space) as its test function instead of v in the Galerkin method. This is well-known in the finite element methods (FEM) for the convection-dominated diffusion problems see e.g. [12, 23, 30]. Actually, the idea of the streamline upwind in the Galerkin framework appeared even earlier with FEM, as a stabilization technique of the Galerkin method for the first-order linear hyperbolic equations in 1970s [7, 25, 26]. Recently, the streamline upwind strategy was applied to the spectral element method for the radiative transfer problems [29] without any theoretical analysis.

Numerical analysis of the DSM is addressed for the equation (1.1) with the following conditions, respectively,

$$D=R=(-\infty,\infty),\quad U(x,t)=U(x+2\pi,t),\quad t\in[0,T],$$
$$a(\cdot),\ b(\cdot),\ U_0(\cdot) \text{ are periodic of the period } 2\pi \text{ in } x. \tag{1.2a}$$
$$D=I=(-1,1),\quad U(\pm1,t)=0,\ t\in[0,T];\quad a(-1)>0, a(1)<0. \tag{1.2b}$$

We will present some quasi-optimal error estimates of order $O(N^{\frac{1}{2}-r})$ and numerical results compared with some other high performance methods for variable-coefficient linear hyperbolic problems.

As far as the authors know, except the optimal estimate in [28] for the constant coefficient equation (1.1) with the Dirichlet boundary condition, there is no other detailed error estimate on the DSM for the first-order hyperbolic equation, despite of huge literatures on the streamline upwind (also known as streamline diffusion) FEM and their optimal error analysis for the convection-dominated diffusion problems [23, 30]. For general variable-coefficient case (1.2b), there has no better-than-sub-optimal estimate for the spectral methods yet, although two PGSMs mentioned below admit optimal error estimates for constant-coefficient equation with the Dirichlet boundary condition, see [4, 21] for details.

We briefly review two PGSMs before discussing the DSM. The *tau method*, where the test functions are always taken without any boundary condition constraints, is one of the basic forms of the spectral method [1, 14, 19, 24]. Analysis and the error estimates of the *tau method* for the equations above have been discussed in [2]. The eigenvalue problems of the first-order operator in different methods have been investigated by many authors [3, 6, 8, 9]. However, an optimal error estimate [4] is obtained until recently for the Legendre-tau method of the first order equation (1.1) with constant coefficients and the Dirichlet boundary condition. The *dual-Petrov–Galerkin spectral method* (DPGSM), which is relatively new, is originally introduced for solving the third-order and is extended to higher odd-order [20] and the first order equations [21]. The main idea is to choose the trial and the test functions satisfying the underlying boundary conditions and the "dual" boundary conditions respectively, and the key benefit is leading to a strongly coercive bilinear form for non-symmetric odd-order differential operators.

In this work, we mainly prove quasi-optimal convergence rate of the Fourier DSM and the Legendre DSM for the first-order variable-coefficient hyperbolic problems. Also numerical results are given to compare the Fourier DSM with the finite volume method (FVM)

satisfying three conservation laws proposed in [27] for the periodic case, showing their performance in long-term integration, and the Legendre DSM with the GSM and the DPGSM for the case of the Dirichlet boundary condition, showing some superiority over the GSM in dealing with solutions of limited smoothness on the boundary.

The paper is organized as follows. In Section 2, we describe the DSM for the equation (1.1) with the conditions (1.2a) and (1.2b). In the following two sections, we present the proofs of the error estimates, especially for the equations with variable-coefficients. Before concluding discussions, some numerical comparisons are presented among the DSM, the GSM and the DPGSM for the case (1.2b). Comparisons for (1.2a) are also shown with some high performance finite volume methods in long-time integration. High accuracy of the DSM is verified in several numerical results.

2. Numerical schemes and main results

Before addressing the numerical schemes for the aforementioned equations, we introduce some notations which will be used throughout the paper. For any non-negative integer N, let $\mathcal{V}_N = \operatorname{span}\left\{e^{ikx}, |k| \le N\right\}$, $\mathbf{P}_N$ denotes the set of all algebraic polynomials of degree not more than N, and C always denotes a positive constant independent of N. We denote by $\|\cdot\|$ the general L^2-norm in space with inner products $(\cdot,\cdot)$, and $H^r(I)$ the standard Sobolev spaces on $I=(-1,1)$, and

$$C_p^r(0,2\pi) = \left\{u|\ u \in C^r(R), u(x) = u(x+2\pi), x \in R, r \ge 0\right\},$$
$$H_p^r(0,2\pi) = \left\{u|\ u \in H_{\mathrm{loc}}^r(R), u(x) = u(x+2\pi), x \in R, r \ge 0\right\}.$$

The Fourier DSM for the equation (1.1) with (1.2a) is to find $u(\cdot,t) \in \mathcal{V}_N$ such that

$$(\partial_t u + a\partial_x u + bu, v + N^{-1}a\partial_x v) = (f, v + N^{-1}a\partial_x v), \quad \forall v \in \mathcal{V}_N. \tag{2.1}$$

The Legendre DSM for the case (1.2b) is to find $u(\cdot,t) \in \mathcal{W}_N = \mathbf{P}_N \cap H_0^1(I)$ such that

$$(\partial_t u + a\partial_x u + bu, v + N^{-1}a\partial_x v) = (f, v + N^{-1}a\partial_x v), \quad \forall v \in \mathcal{W}_N. \tag{2.2}$$

The main results of this work are the error estimates for the dissipative schemes (2.1) and (2.2).

Theorem 2.1. *Let $U \in H^1(0,T;H_p^r(0,2\pi))$ $(r \ge 1)$ be the solution to the Eq.* (1.1) *with the condition* (1.2a) *and $u \in H^1(0,T;\mathcal{V}_N)$ the solution to* (2.1). *Let $a(x), b(x) \in C_p^1(0,2\pi)$. Then there exists a constant C depending on $\|a\|_{C_p^1(0,2\pi)}$ and $\|b\|_{C_p^1(0,2\pi)}$, such that for any $t \in (0,T]$,*

$$\begin{aligned}&\|U(t)-u(t)\|^2 + N^{-2}\|a\partial_x(U(t)-u(t))\|^2\\ &\le e^{Ct}N^{1-2r}\left[\|\partial_x^r U_0\|^2 + \int_0^t \left(\|\partial_x^r U(s)\|^2 + \|\partial_x^r \partial_s U(s)\|^2\right) ds\right].\end{aligned}$$

Theorem 2.2. *Let $U \in H^1(0,T;H^r(I)\cap H_0^1(I))$ $(r \geq 1)$ be the solution to the equation* (1.1) *with the condition* (1.2b) *and $u \in H^1(0,T;\mathscr{W}_N)$ the solution to* (2.2). *Let $a(x), b(x) \in C^1(I)$. Then there exists a constant C depending on $\|a\|_{C^1(I)}$ and $\|b\|_{C^1(I)}$, such that for any $t \in (0,T]$,*

$$\begin{aligned}&\|U(t)-u(t)\|^2+N^{-2}\|a\partial_x(U(t)-u(t))\|^2\\ &\leq \mathrm{e}^{Ct}N^{1-2r}\left[\|\partial_x^r U_0\|^2+\int_0^t\left(\|\partial_x^r U(s)\|^2+\|\partial_x^r\partial_s U(s)\|^2\right)\mathrm{d}s\right].\end{aligned}$$

3. Error estimate for equation with periodic boundary condition

In this paper, we concentrate mainly on the error estimate of the corresponding scheme (2.1) for the variable-coefficient equation (1.1) with (1.2a). In this section, we denote

$$\|w\|_{1,a}=\|w\|+\|a\partial_x w\|,\quad \|w\|_{1,a,N}=\|w\|+N^{-1}\|a\partial_x w\|,$$

and define a bilinear form

$$\mathscr{F}(u,v):=(a\partial_x u, v+N^{-1}a\partial_x v)+k(u,v), \tag{3.1}$$

where $u \in H_p^1(0,2\pi)$, $v \in H_p^1(0,2\pi)$, and the constant $k \geq \frac{1}{2}\|\partial_x a\|_\infty + 1$.

Define projection $P_N^* : H_p^1(0,2\pi) \mapsto \mathscr{V}_N$, satisfying that for any $U \in H_p^1(0,2\pi)$, for any $v \in \mathscr{V}_N$,

$$\mathscr{F}(\eta,v)=0,\quad \eta = U - P_N^* U := U - u^*. \tag{3.2}$$

Before proving Theorem 2.1, we need some lemmas. Let $P_N :\ L^2(0,2\pi) \mapsto \mathscr{V}_N$ be the orthogonal projection operator, i.e.

$$(P_N u, v)=(u,v),\quad \forall v \in \mathscr{V}_N.$$

Lemma 3.1. [13] *If $0 \leq \mu \leq \sigma$ and $u \in H_p^\sigma(0,2\pi)$, then*

$$\|\partial_x^\mu(P_N u - u)\| \leq CN^{\mu-\sigma}\|\partial_x^\sigma u\|.$$

Lemma 3.2. *For the bilinear form* (3.1), *for any $u, v \in H_p^1(0,2\pi)$,*

$$\begin{aligned}&\mathscr{F}(u,u) \geq \|u\|^2+N^{-1}\|a\partial_x u\|^2 \geq \frac{1}{2}\|u\|_{1,a,N}^2,\\ &\mathscr{F}(u,v) \leq k\|u\|_{1,a}\|v\|_{1,a,N} \leq kN\|u\|_{1,a,N}\|v\|_{1,a,N}.\end{aligned}$$

Proof. For bilinear form $\mathscr{F}(\cdot,\cdot)$ and any $u \in H_p^1(0,2\pi)$,

$$\begin{aligned}\mathscr{F}(u,u)=&(a\partial_x u, u+N^{-1}a\partial_x u)+k(u,u)=(a\partial_x u,u)+N^{-1}(a\partial_x u, a\partial_x u)+k(u,u)\\ =&\frac{1}{2}\left[au^2\Big|_0^{2\pi}-(\partial_x a u,u)\right]+N^{-1}(a\partial_x u,a\partial_x u)+k(u,u)\\ =&\left(\left(k-\frac{1}{2}\partial_x a\right)u,u\right)+N^{-1}(a\partial_x u,a\partial_x u)\\ \geq&\|u\|^2+N^{-1}\|a\partial_x u\|^2.\end{aligned}$$

For any $u, v \in H_p^1(0, 2\pi)$, we have

$$\mathscr{F}(u, v) \le k \left(\|u\| + \|a\partial_x u\| \right) \|v\|_{1,a,N} = k\|u\|_{1,a}\|v\|_{1,a,N} \le kN\|u\|_{1,a,N}\|v\|_{1,a,N}.$$

This completes the proof. □

Remark 3.1. By this lemma, it is easy to see that the projection P_N^* defined by (3.2) exists and is unique. In fact, by (3.2), $\mathscr{F}(u^*, v) = \mathscr{F}(U, v)$ holds for any $v \in \mathscr{V}_N$. Thus taking $U = 0$ and $v = u^* \in \mathscr{V}_N$ will lead to, according to Lemma 3.2,

$$0 = \mathscr{F}(u^*, u^*) \ge \|u^*\|^2 \ge 0.$$

Then u^* must be 0.

We have the following estimate for $\eta = U - P_N^* U$.

Lemma 3.3. *For $U \in H_p^r(0, 2\pi)$, there exists a constant C depending on $\|a\|_{C_p^1(0,2\pi)}$, such that*

$$N^{-\frac{1}{2}}\|U - P_N^* U\| + N^{-1}\|a\partial_x(U - P_N^* U)\| \le CN^{-r}\|\partial_x^r U\|.$$

Proof. Firstly, let $\bar{u} = P_N U \in \mathscr{V}_N$. From Lemma 3.1, we can prove that

$$\|U - \bar{u}\|_{1,a,N} = \|U - \bar{u}\| + N^{-1}\|a\partial_x (U - \bar{u})\| \le CN^{-r}\|\partial_x^r U\|, \tag{3.3}$$

where C depends on $\|a\|_\infty$. From Lemma 3.2, we have that for any $u, v \in H_p^1(0, 2\pi)$,

$$\begin{aligned}\mathscr{F}(u, v) \le& \frac{N^{-1}}{4}\|u\|_{1,a}^2 + k^2 N\|v\|_{1,a,N}^2 \\ \le& \frac{1}{2}\left(\|u\|^2 + N^{-1}\|a\partial_x u\|^2 \right) + k^2 N\|v\|_{1,a,N}^2. \end{aligned} \tag{3.4}$$

By the definition of the projection P_N^*, since $\eta - (U - \bar{u}) = \bar{u} - u^* \in \mathscr{V}_N$, we have

$$\mathscr{F}(\eta, \eta) = \mathscr{F}(\eta, U - \bar{u}),$$

following which we can get from Lemma 3.2 and (3.3), (3.4),

$$\|\eta\|^2 + N^{-1}\|a\partial_x \eta\|^2 \le CN\|U - \bar{u}\|_{1,a,N}^2 \le CN^{1-2r}\|\partial_x^r U\|^2.$$

Thus we have

$$\|\eta\| \le CN^{\frac{1}{2}-r}\|\partial_x^r U\|, \quad \|a\partial_x \eta\| \le CN^{1-r}\|\partial_x^r U\|.$$

This completes the proof. □

In the proof of Theorem 2.1, the next lemma will be used.

Lemma 3.4. *For $U \in H_p^r(0, 2\pi)$, there exists a constant C depending on $\|a\|_{C_p^1(0,2\pi)}$, such that for any $v \in \mathscr{V}_N$,*

$$(a\partial_x(U - P_N^* U), v + N^{-1}a\partial_x v) \le CN^{\frac{1}{2}-r}\|\partial_x^r U\|\|v\|.$$

The proof of this lemma is straightforward. From (3.1), (3.2) and Lemma 3.3, for any $v \in \mathscr{V}_N$,

$$(a\partial_x \eta, v + N^{-1} a\partial_x v) = -k(\eta, v) \le C\|\eta\|\|v\| \le CN^{\frac{1}{2}-r}\|\partial_x^r U\|\|v\|.$$

Now, we are about to prove Theorem 2.1.

Proof of Theorem 2.1.

For any $t \in (0, T]$, η is defined by (3.2), U the solution to equation (1.1) with (1.2a) and denote $e = u - u^* \in \mathscr{V}_N$, then $u - U = e - \eta$.

From (1.1) and (2.1), we get the error equation

$$\left(\partial_t e + a\partial_x e + be, v + N^{-1} a\partial_x v\right) = \left(\partial_t \eta + a\partial_x \eta + b\eta, v + N^{-1} a\partial_x v\right). \tag{3.5}$$

Taking $v = 2e$ in (3.5) yields

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\|e\|^2 + 2N^{-1}\left[\|a\partial_x e\|^2 + (\partial_t e, a\partial_x e)\right] \\ =& \left(\left[\partial_x(a + N^{-1}ab) - 2b\right]e, e\right) + 2\left(\partial_t \eta + a\partial_x \eta + b\eta, e + N^{-1}a\partial_x e\right). \end{aligned}$$

By Lemmas 3.3 and 3.4, we have

$$\begin{aligned} 2\left(\partial_t \eta, e + N^{-1}a\partial_x e\right) &\le 2\|\partial_t \eta\|^2 + \|e\|^2 + N^{-2}\|a\partial_x e\|^2 \\ &\le CN^{1-2r}\|\partial_x^r \partial_t U\|^2 + \|e\|^2 + N^{-2}\|a\partial_x e\|^2, \\ 2\left(a\partial_x \eta, e + N^{-1}a\partial_x e\right) &\le 2CN^{\frac{1}{2}-r}\|\partial_x^r U\|\|e\| \le C\left[N^{1-2r}\|\partial_x^r U\|^2 + \|e\|^2\right]. \end{aligned}$$

Thus we have

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\|e\|^2 + 2N^{-1}\left[\|a\partial_x e\|^2 + \left(\partial_t e, a\partial_x e\right)\right] \\ \le& C\left(\|e\|^2 + N^{-2}\|a\partial_x e\|^2 + N^{1-2r}\left[\|\partial_x^r U\|^2 + \|\partial_x^r \partial_t U\|^2\right]\right). \end{aligned} \tag{3.6}$$

Then taking $v = 2\partial_t e$ in (3.5) yields

$$\begin{aligned} &2\|\partial_t e\|^2 + N^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\|a\partial_x e\|^2 + 2\left(a\partial_x e, \partial_t e\right) \\ =& N^{-1}\left(\partial_x a\partial_t e, \partial_t e\right) + 2\left(\partial_t \eta + a\partial_x \eta + b(\eta - e), \partial_t e + N^{-1}a\partial_x \partial_t e\right) \\ =:& N^{-1}\left(\partial_x a\partial_t e, \partial_t e\right) + I_1 + I_2 + I_3. \end{aligned} \tag{3.7}$$

By integration by parts,

$$\begin{aligned} I_1 &= 2(\partial_t \eta, \partial_t e + N^{-1}a\partial_x \partial_t e) \\ &= 2(\partial_t \eta, \partial_t e) - 2N^{-1}(\partial_t e, \partial_x a\partial_t \eta + a\partial_t \partial_x \eta) \\ &= 2(\partial_t e, (1 - N^{-1}\partial_x a)\partial_t \eta) - 2N^{-1}(\partial_t e, a\partial_t \partial_x \eta) \\ &\le N^{-1}\|\partial_t e\|^2 + CN\|\partial_t \eta\|^2 + N^{-1}\|\partial_t e\|^2 + N^{-1}\|a\partial_t \partial_x \eta\|^2. \end{aligned}$$

It holds, by Lemma 3.4, that

$$\begin{aligned} I_2 &= 2\left(a\partial_x\eta, \partial_t e + N^{-1}a\partial_x\partial_t e\right) \\ &\le 2CN^{\frac{1}{2}-r}\|\partial_x^r U\|\|\partial_t e\| \le C\left(N^{2-2r}\|\partial_x^r U\|^2 + N^{-1}\|\partial_t e\|^2\right). \end{aligned}$$

For $I_3 = 2(b(\eta - e), \partial_t e + N^{-1}a\partial_x\partial_t e)$, we have

$$\begin{aligned} &2(b(\eta-e), \partial_t e) \le CN(\|e\|^2 + \|\eta\|^2) + CN^{-1}\|\partial_t e\|^2, \\ &2(b(\eta-e), N^{-1}a\partial_x\partial_t e) = -2N^{-1}\left(\partial_t e, \partial_x(ab)(\eta-e) + b(a\partial_x\eta - a\partial_x e)\right) \\ &\qquad\qquad \le CN^{-1}\left(\|\partial_t e\|^2 + \|e\|^2 + \|\eta\|^2 + \|a\partial_x e\|^2 + \|a\partial_x\eta\|^2\right). \end{aligned}$$

Then we have

$$\begin{aligned} &2\|\partial_t e\|^2 + N^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\|a\partial_x e\|^2 + 2\left(a\partial_x e, \partial_t e\right) \\ &\le C\left(N^{-1}\|\partial_t e\|^2 + N\|e\|^2 + N^{2-2r}\left[\|\partial_x^r\partial_t U\|^2 + \|\partial_x^r U\|^2\right] + N^{-1}\|a\partial_x e\|^2\right). \end{aligned} \tag{3.8}$$

Taking advantage of

$$2\left(a\partial_x e, \partial_t e\right) \le C_0^{-1}\|a\partial_x e\|^2 + C_0\|\partial_t e\|^2,$$

we get that while N is large enough,

$$\|\partial_t e\|^2 \le C\left(-N^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\|a\partial_x e\|^2 + \|a\partial_x e\|^2 + N\|e\|^2 + N^{2-2r}\left[\|\partial_x^r\partial_t U\|^2 + \|\partial_x^r U\|^2\right]\right).$$

Inserting this into the right-hand side of (3.8) gives

$$\begin{aligned} &2\|\partial_t e\|^2 + N^{-1}(1+N^{-1})\frac{\mathrm{d}}{\mathrm{d}t}\|a\partial_x e\|^2 + 2\left(a\partial_x e, \partial_t e\right) \\ &\le C\left(N\|e\|^2 + N^{-1}\|a\partial_x e\|^2 + N^{2-2r}\left[\|\partial_x^r\partial_t U\|^2 + \|\partial_x^r U\|^2\right]\right). \end{aligned} \tag{3.9}$$

Multiplying (3.9) by N^{-1} and adding the result to (3.6), we get

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\left(\|e\|^2 + N^{-2}\|a\partial_x e\|^2\right) + 2N^{-1}\left[\|a\partial_x e\|^2 + 2(\partial_t e, a\partial_x e) + \|\partial_t e\|^2\right] \\ &\le C\left(\|e\|^2 + N^{-2}\|a\partial_x e\|^2 + N^{1-2r}\left[\|\partial_x^r U\|^2 + \|\partial_x^r\partial_t U\|^2\right]\right). \end{aligned}$$

For $\|a\partial_x e\|^2 + 2\left(\partial_t e, a\partial_x e\right) + \|\partial_t e\|^2 \ge 0$, we have

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\left(\|e\|^2 + N^{-2}\|a\partial_x e\|^2\right) \\ &\le C\left(\|e\|^2 + N^{-2}\|a\partial_x e\|^2\right) + CN^{1-2r}\left[\|\partial_x^r U\|^2 + \|\partial_x^r\partial_t U\|^2\right]. \end{aligned} \tag{3.10}$$

From Gronwall's lemma, we have for any $t \in [0, T]$,

$$E(t) \le \left[E(0) + CN^{1-2r}\int_0^t E_c(s)\mathrm{d}s\right]\mathrm{e}^{Ct},$$

where

$$E(t) = \|e\|^2 + N^{-2}\|a\partial_x e\|^2, \;\; E_c(t) = \|\partial_x^r U\|^2 + \|\partial_x^r \partial_t U\|^2.$$

For $E(0)$, we have

$$E(0) = \|e(0)\|^2 + N^{-2}\|a\partial_x e(0)\|^2 \le CN^{-2r}\|\partial_x^r U_0\|^2.$$

Then for any $t \in (0, T]$,

$$\|e(t)\|^2 + N^{-2}\|a\partial_x e(t)\|^2 \le \mathrm{e}^{Ct} N^{1-2r}\left[\|\partial_x^r U_0\|^2 + \int_0^t \left(\|\partial_x^r U\|^2 + \|\partial_x^r \partial_s U\|^2\right) \mathrm{d}s\right].$$

Thus the proof is complete by triangle inequality and Lemma 3.3. □

4. Error estimate for variable-coefficient equation with the Dirichlet boundary condition

In this section we present some conclusions. The proofs of Theorem 2.2 can be stated by the same argument in that of Theorem 2.1 with only difference in boundary conditions. In this section, we denote

$$\|w\|_{1,a} = \|w\| + \|a\partial_x w\|, \quad \|w\|_{1,a,N} = \|w\| + N^{-1}\|a\partial_x w\|,$$

and define the bilinear form

$$\mathscr{A}(u, v) := (a\partial_x u, v + N^{-1}a\partial_x v) + k(u, v), \tag{4.1}$$

where $u \in H_0^1(I)$, $v \in H_0^1(I)$, and the constant $k \ge \frac{1}{2}\|\partial_x a\|_\infty + 1$.

Define projection $P_N^* : H_0^1(I) \mapsto \mathscr{W}_N$, satisfying that for any $U \in H_0^1(I)$, for any $v \in \mathscr{W}_N$,

$$\mathscr{A}(\eta, v) = 0, \quad \eta = U - P_N^* U := U - u^*. \tag{4.2}$$

Lemma 4.1. *For the bilinear form* (4.1)*, for any* $u, v \in H_0^1(I)$*,*

$$\mathscr{A}(u, v) \le k\|u\|_{1,a}\|v\|_{1,a,N} \le kN\|u\|_{1,a,N}\|v\|_{1,a,N},$$
$$\mathscr{A}(u, u) \ge \|u\|^2 + N^{-1}\|a\partial_x u\|^2 \ge \frac{1}{2}\|u\|_{1,a,N}^2.$$

Remark 4.1. By this lemma, the existence and the uniqueness of projection P_N^* defined by (4.2) can be achieved.

Lemma 4.2. *For* $U \in H^r(I) \cap H_0^1(I)$*, there exists a constant* C *depending on* $\|a\|_{C^1(I)}$*, such that*

$$N^{-\frac{1}{2}}\|U - P_N^* U\| + N^{-1}\|a\partial_x(U - P_N^* U)\| \le CN^{-r}\|\partial_x^r U\|.$$

Remark 4.2. Proof of Lemma 4.2 is similar to that of Lemma 3.3. Optimal estimate for $\|U - P_N^* U\|$ can not be proved using Nitsche's duality argument [18] for general coefficient $a(x)$ either. In fact, when $w \in H_0^1(I) \cap H^r(I)$ satisfies that for $\eta \in H_0^r(I)$ and any $v \in H_0^r(I)$,

$$(\eta, v) = \mathscr{A}(v, w), \tag{4.3}$$

due to the property of the operator, we generally can not prove the regularity result

$$N^{-1}\|\partial_x^2 w\| \le C\|\eta\|, \tag{4.4}$$

where C depends on $a(x)$ but can only prove

$$N^{-1}\|\partial_x(a^2\partial_x w)\| \le C\|\eta\|. \tag{4.5}$$

Thus we can not facilitate the duality.

From Lemma 4.2, the following lemma, which will be used in the proof of Theorem 2.2, can be proved straightforward.

Lemma 4.3. *For $U \in H_0^1(I) \cap H^r(I)$, there exists a constant C depending on $\|a\|_{C^1(I)}$, such that for any $v \in \mathscr{W}_N$,*

$$(a\partial_x(U - P_N^* U), v + N^{-1}a\partial_x v) \le CN^{\frac{1}{2}-r}\|\partial_x^r U\|\|v\|.$$

Taking advantage of the lemmas stated in this section, taking the similar idea adopted in the proof of Theorem 2.1, we can easily obtain quasi-optimal convergence rate for the first order variable-coefficient linear hyperbolic equation with the Dirichlet boundary condition, which is the result of Theorem 2.2.

Remark 4.3. We can consider other cases of $a(\pm 1)$ with other signs. Denoting by $\Gamma_- = \{x = \pm 1,\ xa(x) < 0\}$ the inflow boundary of I, the boundary condition is

$$u(x,t) = g(x,t), \quad \forall x \in \Gamma_-.$$

Denote $\mathscr{W}_{N,\Gamma_-,\bar{g}} = \mathbf{P}_N \cap \{u|\ u(x) = \bar{g}(x), \forall x \in \Gamma_-\}$. The corresponding DSM is to find $u(\cdot,t) \in \mathscr{W}_{N,\Gamma_-,g(\cdot,t)}$ such that (2.2) holds for any $v \in \mathscr{W}_{N,\Gamma_-,0}$. We can obtain similar error estimate as Theorem 2.2. Here we should define the projection $P_{N,\Gamma_-}^* : H^1(I) \mapsto \mathbf{P}_N$ as that

$$\begin{aligned} \mathscr{A}(u - P_{N,\Gamma_-}^* u, v) &= 0, \qquad \forall v \in \mathscr{W}_{N,\Gamma_-,0}, \\ (P_{N,\Gamma_-}^* u)(x) &= u(x), \qquad \forall x \in \Gamma_-. \end{aligned}$$

Similar error estimate as in Lemma 4.2 for the projection P_{N,Γ_-}^* can be obtained. Thus we can also get the convergence of quasi-optimal rate for the scheme.

Remark 4.4. If the coefficient a depends on both spatial and temporal variables, the inflow boundary Γ_- depends also on t. The corresponding scheme is similar to that stated in Remark 4.3. But it seems difficult to get the desired error estimate in general. In fact, there will be an additional term $2N^{-1}(a\partial_x e, \partial_t a\partial_x e)$ in the right side of (3.7). However, if there exists a constant $a_0 > 0$ such that

$$|a(x,t)| \geq a_0 > 0, \quad \forall (x,t) \in I \times [0,T], \tag{4.6}$$

we can obtain optimal estimates of the projection P^*_{N,Γ_-} by using Nitsche's duality argument [18] and therefore the underlying scheme.

5. Numerical results

In this section, we first check the performance of the DSM (2.1) in long-term integration and compare it with a recently-developed finite volume scheme [5, 17, 27]. Secondly, we give some comparisons of the accuracy among the DSM (2.2), the GSM and the DPGSM. In all computations, the Crank-Nicolson scheme is used for the temporal discretization. Error in the discrete L^2-norm defined on the corresponding interval (x_L, x_R) is

$$\Big\{h\sum_{j=0}^{M}\big|U(x_j)-u(x_j)\big|^2\Big\}^{1/2},$$

where $h=(x_R-x_L)/M,\ x_j=x_L+jh,\ j=0,\cdots,M$. We take $M=20000$.

5.1. Long-term integration

In this subsection, we mainly compare the DSM (2.1) with a high performance FVM which satisfies three conservation laws (Thr-con) designed in [27]. This FVM is designed for long-term integration and performs much better than traditional schemes along this line. See more in [5, 17, 27].

Example 5.1. The wave-packet problem has been studied as example for hyperbolic equations in many works [15, 16]. Consider the case (1.2a) of period 1 with the initial condition whose expression on interval $[0,1)$ is as follows,

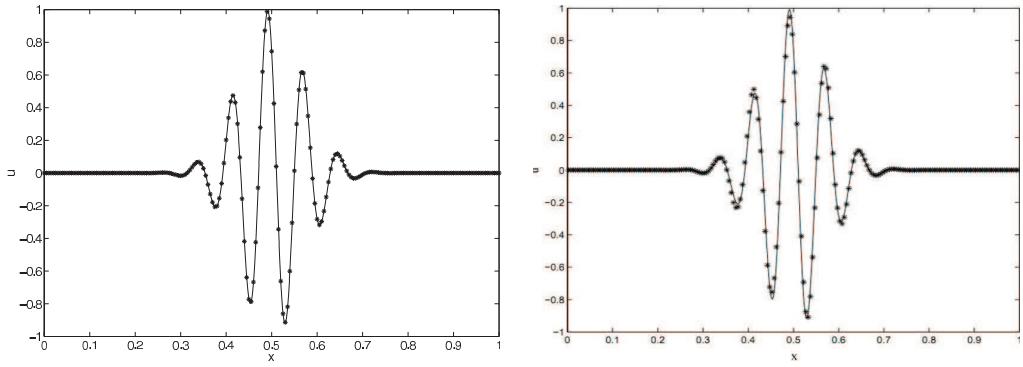
$$U(x,0)=U_0(x)=\mathrm{e}^{-p\left(x-\frac{1}{2}\right)^2}\sin\gamma x. \tag{5.1}$$

Here we take $p=100$, $\gamma=80$,

$$a(x)=\pm\frac{1+\sin 2\pi x}{2}, \quad b(x)=a'(x)=\pm\pi\cos 2\pi x,$$

and $2N=200$, $\tau=10^{-3}$. Since it is hard to obtain the exact solution of this problem, we adopt the following strategy as in Example 5 of [27] to test the scheme (2.1). We first take

$$a(x)=\frac{1+\sin 2\pi x}{2}, \quad b(x)=\pi\cos 2\pi x,$$



(a) DSM with $2N = 200$ and $T = 40000$ (b) Thr-con with $h = 0.005$ and $T = 40$

Figure 1: (a) Numerical solution of DSM for Example 5.1 with $2N = 200$ and $\tau = 10^{-3}$ agrees well with the "exact solution" at $T = 40000$. (b) Numerical solution in [27] with $h = 0.005$ at $T = 40$.

and compute up to $T = 0.2$, and then take

$$a(x) = -\frac{1+\sin 2\pi x}{2}, \quad b(x) = -\pi \cos 2\pi x,$$

up to $T = 0.4$, which is called a "bout".

Fig. 1(a) shows the good agreement between the numerical solution of the DSM and the "exact solution" (initial condition) at the end of the 10^5th bout (now the maximum errors and the L^2-errors are 1.21e-10 and 8.07e-11 respectively in fact). Compared with Fig. 8 in [27] which we take here as Fig. 1(b), our results have higher precision than that of the Thr-con with $N = 200$ at the end of the 100th bout; especially we can find obvious departure from the "exact solution" at the peak and trough near $x = 0.4$ in Fig. 1(b).

5.2. Eq. (1.1) with the Dirichlet B.C. (1.2b)

Firstly, an example is given to show the accuracy of the methods.

Example 5.2. Let $a(x) = -\sin x$ and $b(x) = 0$ in (1.1). The exact solution is

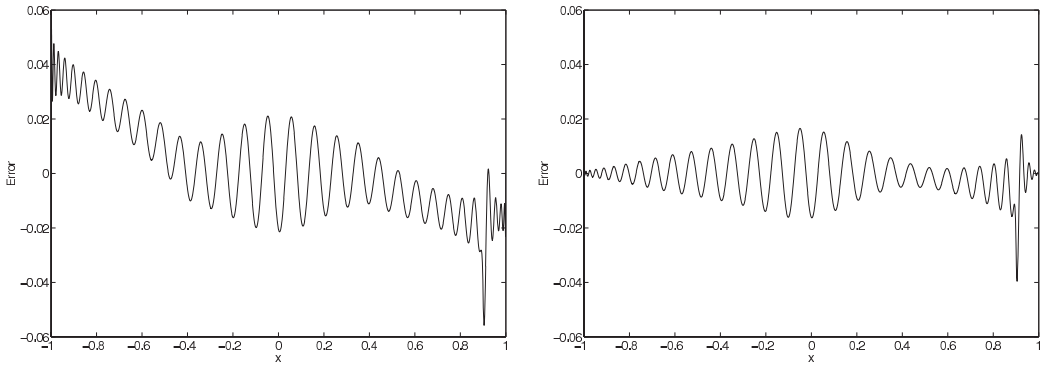
$$U(x,t) = |x|^{2p+1} \sin t.$$

We compute the case for $p = 1$ and $\tau = 10^{-3}$ up to $T = 1$ by the GSM, the DSM, and the DPGSM. In this case, $U(\cdot,t) \in H^{3.5-\varepsilon}(I)$ (ε is an arbitrary small positive number). The errors and the convergence orders in L^2-norm are reported in Table 1. From the table, we can see that the optimal convergence rate of all the three methods is observed numerically, better than our theoretical prediction $N^{-(3-\varepsilon)}$. The errors of the DSM and the DPGSM are smaller than the GSM.

Then we give two examples to compare the DSM with the GSM and the DPGSM.

Table 1: L^2-errors for Example 5.2 with $T=1, \tau=10^{-3}$, and $p=1$.

	L^2-errors			order		
N	GSM	DSM	DPGSM	GSM	DSM	DPGSM
8	1.85e-03	1.31e-03	1.33e-03			
16	1.83e-04	1.29e-04	1.29e-04	$N^{-3.34}$	$N^{-3.35}$	$N^{-3.36}$
32	1.76e-05	1.20e-05	1.21e-05	$N^{-3.38}$	$N^{-3.42}$	$N^{-3.42}$
64	1.67e-06	1.11e-06	1.11e-06	$N^{-3.39}$	$N^{-3.44}$	$N^{-3.44}$



(a) GSM (b) DSM

Figure 2: Errors of numerical solutions by the GSM and the DSM.

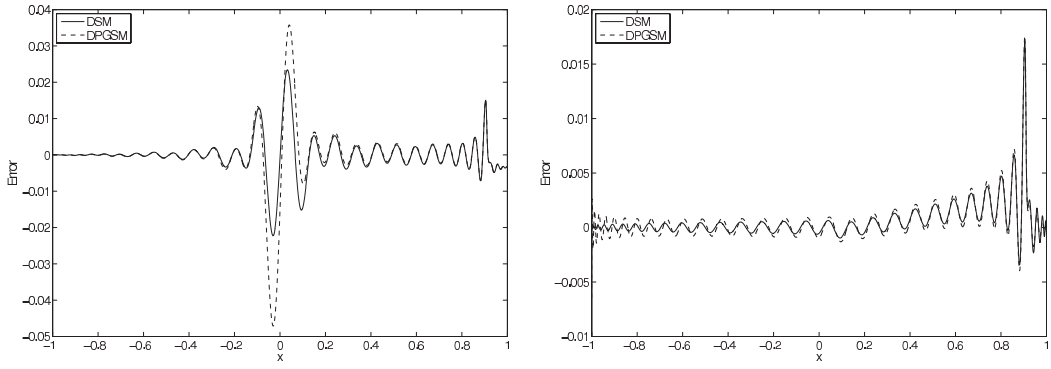
Example 5.3. Let $a(x)=-x^3$ and $b(x)=x^2$ in (1.1). The exact solution is

$$U(x,t)=\begin{cases} t, & x\in(-1,x_\gamma), \\ t\cos\pi\gamma(x-x_\gamma), & x\in(x_\gamma,1), \end{cases} \tag{5.2}$$

where $x_\gamma=1-1/\gamma$. The solution is C^1 continuous at $x=x_\gamma$ and steep in the interval $(x_\gamma,1)$ when large γ is taken.

It is computed by the GSM and the DSM (2.2) up to $T=3$ with $\gamma=10$, $N=64$, and $\tau=10^{-3}$. The errors of both methods are reported in Fig. 2 plotted with 20000 even points respectively. From the figures we find that, at the weak discontinuity $x=0.9$, the errors are both the main one due to the singularity there. In comparison, at anywhere else, the errors of the DSM are weaker than those of the GSM, especially near the left boundary. This phenomenon is similar to that in FEM. We presume from the results that, thanks to the viscosity added in the DSM, parts of the errors coming from the singularity point $x=0.9$ are absorbed during the propagating to the boundary, and the dissipative effect of the DSM is more obvious at the boundary.

Example 5.4. For Eq. (1.1) and solution (5.2), we take $b(x)=0$, $\gamma=10$, $N=64$, and $\tau=10^{-3}$.



(a) $a(x)=\sin(x)$ (b) $a(x)=-\cos(x)$

Figure 3: Errors of numerical solutions by the DSM (solid line) and the DPGSM (dash line).

It is computed by the DSM and the DPGSM up to $T=1$ with $a(x)=\sin(x)$ and $a(x)=-\cos(x)$ being taken respectively. We report the errors in Fig. 3. From the figures, we can see that when we take $a(x)=\sin(x)$, the errors of the DSM are smaller than those of the DPGSM in the center of the interval; and the errors of the DPGSM are larger near the left boundary than the DSM when we take $a(x)=-\cos(x)$. The result of the case $a(x)=-\sin(x)$ is not reported here, because the errors of both methods are almost the same.

Next we give an example whose solution has limited regularity at the boundary of the interval.

Example 5.5. Let $a(x)=-x^3$ and $b(x)=-3x^2$ in (1.1). The exact solution is

$$U(x,t)=(1-x^2)^{p+\frac{1}{2}}\mathrm{e}^{\sin t}, \tag{5.3}$$

which is singular at the boundary nodes $x=\pm1$ for an integer p.

Taking $T=1$ and $\tau=10^{-3}$, we compute the case for $p=1$. The errors and the convergence orders in L^2-norm are listed in Table 2, from which we find that, the accuracy and the convergence orders of the DSM are higher than that of the GSM. In fact, the phenomenon – the errors near the boundary of the DSM are obviously less than that of

Table 2: L^2-errors for Example 5.5 with $T=1,\tau=10^{-3}$, and $p=1$.

	L^2-errors		order	
N	GSM	DSM	GSM	DSM
32	1.498e-03	8.966e-04		
64	2.972e-04	1.411e-04	$N^{-2.33}$	$N^{-2.67}$
128	6.036e-05	2.205e-05	$N^{-2.30}$	$N^{-2.68}$
256	1.186e-05	3.695e-06	$N^{-2.35}$	$N^{-2.58}$

the GSM – presented in Fig. 2 can also be seen in the error plotting of this example. We presume that the DSM might be superior over the GSM to deal with boundary singularity.

In this example, the numerical orders reported in Table 2 are even higher than the "optimal" rate of convergence. In fact, the error estimate could be improved by using some weighted norm in analysis as has been discussed in [11, 22].

6. Discussion and conclusion

We introduce the dissipative spectral methods (DSM) for the first-order linear hyperbolic equations with periodic and the Dirichlet boundary conditions. It is shown that the schemes admit quasi-optimal convergence rate for the variable-coefficient equations, which is better than the other aforementioned spectral methods.

The numerical results show that the DSM has some superiority over the GSM and the DPGSM for solving the first-order linear hyperbolic equations. And the DSM has the ability of long-term integration for the first-order linear hyperbolic equations. However, this ability is limited. To check the potential ability in long-time integration, we apply dispersion relation analysis of the scheme (2.1) when a is a positive constant. In fact, the eigenvalues of the first-order operator with the DSM are complex, instead of pure imaginary, with non-positive real part, which means it behaves like the convection-diffusion operator. This results in damping in long-term integration.

The DSM can be extended to high dimensional problems with simple boundary conditions. However, for the cases with complex boundary conditions as discussed in Remark 4.3, it is better to use the dissipative spectral element scheme. Future work may include the DSM for nonlinear equations, extending the method to the two dimensional equations, and more comparisons among different numerical methods for long-term integration.

Acknowledgments The authors are very grateful to the anonymous referees for their valuable suggestions and helpful comments. The work is supported by National Natural Science Foundation of China (11171209), Leading Academic Discipline Project of Shanghai Municipal Education Commission (J50101), Specialized Research Fund for the Doctoral Program of Higher Education (20060280010) and Graduate Innovative Foundation of Shanghai University (SHUCX091048).

References

[1] M. H. Aliabadi and E. L. Ortiz, *Numerical treatment of moving and free boundary value problems with the tau method*, Computers Math. Applic., vol. 35, no. 8 (1998), pp. 53–61.

[2] C. Canuto, M. Y. Hussaini, A. Quartesoni, and T. A. Zang, *Spectral Methods: Fundamentals in Single Domains*, Springer, 2006.

[3] M. Charalambides and F. Waleffe, *Spectrum of the Jacobi tau approximation for the second derivative operator*, SIAM J. Numer. Anal., 46 (2008), pp. 280–294.

[4] K. Cui and H. P. Ma, *The Legendre-tau method for a first-order hyperbolic equation(In Chinese)*, J. Commun. Appl. Math. Comput., 23 (2009), pp. 42–52.

[5] Y. F. Cui and D. K. Mao, *Numerical method satisfying the first two conservation laws for the Korteweg-de Vries equation*, J. Comput. Phys., 227 (2007), pp. 376–399.
[6] P. T. Dawkins, S. R. Dunbar, and R. W. Douglass, *The origin and nature of spurious eigenvalues in the spectral tau method*, J. Comput. Phys., 147 (1998), pp. 441–462.
[7] J. E. Dendy, *Two methods of Galerkin type achieving optimum L^2 rates of convergence for first order hyperbolics*, SIAM J. Numer. Anal., 11 (1974), pp. 637–653.
[8] M. K. El-Daou and E. L. Ortiz, *Error analysis of the tau method: Dependence of the error on the degree and the length of the interval of approximation*, Computers Math.Applic., vol. 25, no. 7 (1993), pp. 33–45.
[9] M. K. El-Daou and E. L. Ortiz, *A posteriori error bounds for the approximate solution of second-order ODEs by piecewise coefficients perturbation methods*, J. Comput. Appl. Math., 189 (2006), pp. 51–66.
[10] D. Gottlieb and J. S. Hesthaven, *Spectral methods for hyperbolic problems*, J. Comput. Appl. Math., 128 (2001), pp. 83–131.
[11] W. Z. Huang, H. P. Ma, and W. W. Sun, *Convergence analysis of spectral collocation methods for a singular differential equation*, SIAM J. Numer. Anal., 41 (2004), pp. 2333–2349.
[12] T. J. Hughes and A. N. Brooks, *A multi-dimensional upwind scheme with no crosswind diffusion*, in Finite Element Methods for Convection Dominated Flows, T. J. Hughes, ed., vol. 34, ASME, New York, 1979, pp. 19–35.
[13] H. O. Kreiss and J. Oliger, *Stability of the Fourier method*, SIAM J. Numer. Anal., 16 (1979), pp. 421–433.
[14] G. S. Landriani, *Spectral tau approximation of the two-dimensional Stokes problem*, Numer. Math., 52 (1988), pp. 683–699.
[15] R. J. LeVeque, *Numerical Methods for Conservation Laws*, Birkhauser-verlag, 1992.
[16] R. J. LeVeque, *Finite Volume Methods for Hyperbolic Problems*, Cambridge Univ. Press., 2002.
[17] H. X. Li, Z. G. Wang, and D. K. Mao, *Numerically neither dissipative nor compressive scheme for linear advection equation and its application to the Euler system*, J. Sci. Comput., 36 (2008), pp. 285–331.
[18] J. A. Nitsche, *Ein kriterium für die quasi-optimalitat des Ritzchen Verfahrens*, Numer. Math., 11 (1968), pp. 346–348.
[19] J. Shen, *A spectral-tau approximation for the Stokes and Navier-Stokes equations*, Math. Model. Num. Anal., 22 (1988), pp. 677–693.
[20] J. Shen, *A new dual-Petrov–Galerkin method for third and higher odd-order differential equations: Application to the KdV equation*, SIAM J. Numer. Anal., 41 (2004), pp. 1595–1619.
[21] J. Shen and L. L. Wang, *Legendre and Chebyshev dual-Petrov–Galerkin methods for hyperbolic equations*, Comput. Methods Appl. Mech. Engrg., 196 (2007), pp. 3785–3797.
[22] T. T. Shen, Z. Q. Zhang, and H. P. Ma, *Optimal error estimates of the Legendre tau method for second-order differential equations*, J. Sci. Comput., 42 (2010), pp. 198–215.
[23] M. Stynes and L. Tobiska, *The SDFEM for a convection-diffusion problem with a boundary layer: Optimal error analysis and enhancement of accuracy*, SIAM J. Numer. Anal., 41 (2004), pp. 1620–1642.
[24] J. G. Tang and H. P. Ma, *Single and multi-interval Legendre τ-methods in time for parabolic equations*, Adv. Comput. Math., 17 (2002), pp. 349–367.
[25] L. B. Wahlbin, *A dissipative Galerkin method applied to some quasilinear hyperbolic equations*, Revue francaise d'automatique, informatique, recherche opérationelle. Analyse numérique, 8 (1974), pp. 109–117.
[26] L. B. Wahlbin, *A dissipative Galerkin method for the numerical solution of first-order hyperbolic equation*, in Mathematical Aspects of Finite Elements in Partial Differential Equations,

Academic Press, New York, 1974, pp. 147–169.
[27] Z. G. WANG AND D. K. MAO, *A finite difference scheme for linear advection equation satisfying three conservation laws (in Chinese)*, J. Shanghai Univ. Nat. Sci., vol. 12, no. 6 (2006), pp. 588–598.
[28] Z. YANG AND Z. Q. ZHANG, *An optimal error estimate of dissipative spectral tau method for a first order hyperbolic equation (in Chinese)*, Journal of Lishui University, vol. 30, no. 5 (2008), pp. 12–15.
[29] J. M. ZHAO AND L. H. LIU, *Spectral element method with adaptive artificial diffusion for solving the radiative transfer equation*, Numerical Heat Transfer, Part B: Fundamentals, 53 (2008), pp. 536–554.
[30] G. H. ZHOU, *How accurate is the streamline diffusion finite element method?* Math. Comp., 66 (1997), pp. 31–44.