ESAIM: M2AN 57 (2023) 367–394
https://doi.org/10.1051/m2an/2023003

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

ANY ORDER SPECTRAL VOLUME METHODS FOR DIFFUSION EQUATIONS USING THE LOCAL DISCONTINUOUS GALERKIN FORMULATION

Jing An[1] and Waixiang Cao[2,*]

Abstract. In this paper, we present and study two spectral volume (SV) schemes of arbitrary order for diffusion equations by using the local discontinuous Galerkin formulation to discretize the viscous flux. The basic idea of the scheme is to rewrite the diffusion equation into an equivalent first-order system first, and then use the SV method to solve the system. The SV scheme is designed with control volumes constructed by using the Gauss points or Radau points in subintervals of the underlying meshes, which leads to two SV schemes referred to as LSV and RSV schemes, respectively. The stability analysis for the linear diffusion equations based on alternating fluxes are provided, and optimal error estimates are established for both the exact solution and the auxiliary variable. Furthermore, a rigorous mathematical proof are given to demonstrate that the proposed RSV method is identical to the standard LDG method when applied to constant diffusion problems. Numerical experiments are presented to demonstrate the stability, accuracy and performance of the two SV schemes for both linear and nonlinear diffusion equations.

Mathematics Subject Classification. 65M15, 65M60, 65N30.

Received October 9, 2022. Accepted January 3, 2023.

1. Introduction

The spectral volume (SV) method is a class of high order Godunov-type finite volume method [10], which has been under development for several decades and is considered to be the current state-of-the-art for the numerical solution of hyperbolic conservation laws. The SV method was first proposed by Wang, Liu *et al.* and their collaborators for hyperbolic conservation laws on unstructured grids (see, *e.g.*, [29–31]). Since then it attracted a lot of attention and was successfully implemented for solving various PDEs such as Euler equation [32], Navier–Stokes equations [22, 23], and 3D Maxwell equations [17]. Similar to the discontinuous Galerkin (DG) methods [5, 6, 8, 11], SV methods adopt completely discontinuous functions as solution space and have many desired advantages such as the allowance of hanging nodes, compact stencils, easy *hp* adaptivity, high parallel efficiency, and so on. Comparisons between the SV method and other numerical methods (*e.g.*, DG, spectral difference) in terms of accuracy and stability have also been conducted in the literature, see [21, 25, 34].

For equations containing higher order spatial derivatives, such as diffusion equations, however, SV methods can not be directly applied due to the discontinuous solution space at the element interfaces, which is not regular

Keywords and phrases. Spectral Volume methods, L^2 stability, error estimates, local discontinuous Galerkin, diffusion equations.

[1] School of Mathematical Sciences, Guizhou Normal University, Guiyang 550025, P.R. China.
[2] School of Mathematical Sciences, Beijing Normal University, Beijing 100875, P.R. China.
*Corresponding author: caowx@bnu.edu.cn

enough to handle high order derivatives. To solve this problem, several DG methods have been proposed and studied in the literature to discretize the high order fluxes (*e.g.*, viscous fluxes), including the LDG method [4,7,9], the compact DG (CDG) method [19], the interior penalty (IP) method [1,28], the direct discontinuous Galerkin (DDG) method [16,33], and the method introduced by Baumann–Oden [2,18]. Inspired by the DDG, LDG and IPDG viscous flux formulation, some attempts based on SV context have been made for solving the diffusion and third order spatial derivative equations. Sun and Wang [22] were the first to implement the SV method for the Navier–Stokes equations, where the LDG method was used to discretize the viscous fluxes. Later, Kannan and Wang [13] studied SV method for a p-multigrid SV Navier–Stokes solver, by using a new penalty and BR2 viscous flux formulation. In [14,15], Kannan proposed a modified LDG (called LDG2) and a DDG viscous flux formulation in the context of SV method for the Navier–Stokes equations. Recently, a formulation using the LDG and the IPDG in the SV context has been proposed separately in [12,20] for solving equations containing third order spatial derivative.

Although different SV schemes have been proposed to solve diffusion and third-order spatial derivative equations, and numerical results indicate that the SV method are very promising and have a great potential for solving high order problems, the mathematical analysis of high order SV method in terms of stability, accuracy, and error estimates is still in poor and far from developed. Only several stability results for lower order SV schemes over uniform meshes have been obtained (see [24,26,27,34]). To the best of our knowledge, no theoretic analysis on the accuracy and optimal error estimate of the SV method has been reported yet in the literature for solving high order spatial derivative equations.

This paper is our first attempt to implement and analyze two classes of arbitrary high order SV schemes for diffusion problems when the LDG viscous flux formulation is used. This approach inherits ideas from the LDG method, where the auxiliary variable is introduced to rewrite the diffusion equation into an equivalent first-order system, and then use the SV method to solve it. Similar to the LDG method, the auxiliary variable approximating the derivatives of the solution can be locally eliminated and thus local solvable. The main contribution of the current work lies in that: on the one hand, we proposed a innovative construction of control volumes which are dependent on the choice of numerical fluxes. The special construction of control volumes finally leads to two stable SV schemes of arbitrary high order. On the other hand, we establish a unified approach to investigate the stability, accuracy and error estimates of two SV schemes, and prove that the SV schemes are energy stable and have optimal convergence rates for both the exact solution and the auxiliary variable approximation.

To establish a framework for the stability analysis and error estimates of high order SV schemes, we first construct the control volumes by using Gauss–Legendre points or right/left Radau points of the underlying meshes, which leads to two SV schemes referred to as LSV and RSV schemes, respectively. The points are specially taken according to the choice of numerical fluxes. This special choice ensures the stability of the SV scheme. Then we formulate the SV scheme into its equivalent Petrov–Galerkin method, and introduce a special mapping from the trial space to the test space for both the the exact solution and the auxiliary variable approximation. With the help of the two mappings, L^2 stability and optimal convergence rates of high order SV schemes are finally established. As a byproduct, a comparison between the proposed SV method and the LDG method is conducted. It turns out that the SV can be viewed as a discrete numerical quadrature of the LDG method. Especially, for constant diffusion equations, we prove that the discrete numerical quadrature is equivalent to the exact quadrature for RSV schemes, which indicates that the proposed RSV method is identical to the LDG method when applied to constant diffusion equations.

The rest of the paper is organized as follows. In Section 2, we present two SV schemes based on the LDG formulation for the one-dimensional diffusion equations. Then we prove that the two SV schemes are L^2 energy stable, and have optimal error estimations for both the exact solution and the auxiliary variable. In Section 3, we extend the SV schemes to the multidimensional diffusion equations and show that the proposed SV methods are also stable in the L^2 norm, and have optimal error estimations for $\mathbb{Q}_k$ elements. In Section 4, we provide some numerical examples to test the stability and accuracy of the SV method for both the linear and nonlinear diffusion equations. Concluding remarks are presented in Section 5.

2. SV method for one-dimensional diffusion equations

In this section, we present and analyze the SV method for solving the following one-dimensional diffusion equations:

$$\begin{aligned} &u_t = (\alpha u_x)_x + g(x,t), && (x,t) \in [a,b] \times (0,t_0], \\ &u(x,0) = u_0(x), && x \in R, \end{aligned} \tag{2.1}$$

where $u_0(x), g(x,t)$ are both smooth, $\alpha := \alpha(u,x) \geq 0$ is a positive diffusion coefficient. For simplicity, here we only consider the periodic boundary condition.

To construct the SV scheme, we rewrite (2.1) as a first order system

$$u_t = (\beta q)_x + g(x,t), \qquad q = \beta u_x := B_x, \tag{2.2}$$

where

$$\beta = \sqrt{\alpha}, \qquad B := B(u,x) = \int^x \beta u_x \, \mathrm{d}x. \tag{2.3}$$

2.1. SV schemes

We first divide the computational domain $\Omega = [a,b]$ into N cells with $a = x_{\frac{1}{2}} < x_{\frac{3}{2}} < \ldots < x_{N+\frac{1}{2}} = b$, and denote by $\tau_i = (x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}), x_i = \frac{1}{2}(x_{i-\frac{1}{2}} + x_{i+\frac{1}{2}})$ the cells and cell centers, respectively. Let $h_i = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$ be the length of each cell, $\bar{h}_i = h_i/2$ and $h = \max_i \ h_i$. We assume that the mesh is quasi-uniform, *i.e.*, there exists a constant c such that $h \leq c h_i$, $i \in \mathbb{Z}_N$. Here for any positive integer r, $\mathbb{Z}_r = \{1, \ldots, r\}, \mathbb{Z}_r^0 = \{0, 1, \ldots, r\}$.

Define the finite element space

$$U_h = \{v : \ v|_{\tau_i} \in \mathbb{P}_k, \ i \in \mathbb{Z}_N\},$$

where $\mathbb{P}_k$ denotes the space of polynomials of degree at most k with coefficients as functions of t. For any function v, we denote by v^+ and v^- the right and left limits of v, respectively. Define $\{v\}$ and $[v]$ the average and the jump of v, *i.e.*,

$$\{v\} = \frac{v^+ + v^-}{2}, \qquad [v] = v^+ - v^-.$$

Now let $-1 = s_0 < s_1 < \ldots < s_k < s_{k+1} = 1$, and $-1 = \bar{s}_0 < \bar{s}_1 < \ldots < \bar{s}_k < \bar{s}_{k+1} = 1$ be $k+2$ distinct points in the reference element $[-1,1]$. Define

$$x_{i,j} = x_i + \bar{h}_i s_j, \quad \bar{x}_{i,j} = x_i + \bar{h}_i \bar{s}_j, \qquad i \in \mathbb{Z}_N, j \in \mathbb{Z}_{k+1}^0.$$

Then the SV method for (2.2) reads as: Find $u_h, q_h \in U_h$ such that for all $(i,j) \in \mathbb{Z}_N \times \mathbb{Z}_k^0$

$$\begin{aligned} &\int_{x_{i,j}}^{x_{i,j+1}} \partial_t u_h(x,t) \, \mathrm{d}x - \left(\hat{\beta}\hat{q}_h\right)(x_{i,j+1}) + \left(\hat{\beta}\hat{q}_h\right)(x_{i,j}) = \int_{x_{i,j}}^{x_{i,j+1}} g(x,t) \, \mathrm{d}x, \\ &\int_{\bar{x}_{i,j}}^{\bar{x}_{i,j+1}} q_h(x,t) \, \mathrm{d}x - \hat{B}(u_h(\bar{x}_{i,j+1}), \bar{x}_{i,j+1}) + \hat{B}(u_h(\bar{x}_{i,j}), \bar{x}_{i,j}) = 0, \end{aligned} \tag{2.4}$$

where $\hat{\beta}, \hat{q}_h, \hat{B}$ denote the numerical fluxes, which are single valued functions defined on the points and should be designed based on different guiding principles for different PDEs to ensure stability. Note that at the interior points $x_{i,j}, \bar{x}_{i,j}, j \in \mathbb{Z}_k$, u_h, q_h are continuous and thus we take

$$\hat{\beta}(x_{i,j}) = \beta(u_h(x_{i,j})), \quad \hat{q}_h(x_{i,j}) = q_h(x_{i,j}), \quad \hat{B}(u_h(\bar{x}_{i,j}), \bar{x}_{i,j}) = B(u_h(\bar{x}_{i,j}), \bar{x}_{i,j}), \qquad j \in \mathbb{Z}_k.$$

While at the interface of each element (*i.e.*, $x_{i-\frac{1}{2}}, i \in \mathbb{Z}_N$), we choose the alternating fluxes. That is, $\hat{B}$ and $\hat{q}_h$ are taken from opposite sides. To be more precise,

$$\hat{\beta} = \frac{B^+ - B^-}{u_h^+ - u_h^-}, \quad \hat{q}_h = q_h^-, \quad \hat{B} = B^+, \tag{2.5}$$

or

$$\hat{\beta} = \frac{B^+ - B^-}{u_h^+ - u_h^-}, \quad \hat{q}_h = q_h^+, \quad \hat{B} = B^-. \tag{2.6}$$

We end with this subsection the discussion on specific SV schemes. By choosing different points $s_j, \bar{s}_j, j \in \mathbb{Z}_k$, we can get different SV schemes. Note that the stability and accuracy of the SV scheme are heavily dependent on the construction of control volume, *i.e.*, the choice of $s_j, \bar{s}_j, j \in \mathbb{Z}_k$. In this paper, we restrict our discussion to the following two SV schemes.

GSV: both $\{s_j\}_{j=1}^k$ and $\{\bar{s}_j\}_{j=1}^k$ are chosen as Gauss–Legendre points, *i.e.*, $s_j = \bar{s}_j, j \in \mathbb{Z}_k$ are k zeros of the Legendre polynomial L_k of degree k.

RSV: $\{s_j\}_{j=1}^k, \{\bar{s}_j\}_{j=1}^k$ are taken as k interior Radau points according to the choice of numerical fluxes, *i.e.*,

- For flux choice (2.5), $\{s_j\}_{j=1}^k$ are chosen as k interior right Radau points and $\{\bar{s}_j\}_{j=1}^k$ are chosen as k interior left Radau points.
- For flux choice (2.6), $\{s_j\}_{j=1}^k$ and $\{\bar{s}_j\}_{j=1}^k$ are taken as the k interior left and right Radau points, respectively.

Here k interior right Radau points denote the zeros of Radau polynomial $L_{k+1} - L_k$ except the point $s = 1$. Similarly, the k interior left Radau points denote the zeros of Radau polynomial $L_{k+1} + L_k$ except the point $s = -1$.

Remark 2.1. The purpose of introducing the two sets $\{s_j\}_{j=0}^{k+1}$ and $\{\bar{s}_j\}_{j=0}^{k+1}$ (separately corresponding to $x_{i,j}$ and $\bar{x}_{i,j}$) is to ensure the stability of our numerical scheme (2.4). Actually, due to the fact that the numerical fluxes $\hat{B}$ and $\hat{q}_h$ in (2.4) are taken from opposite sides, the associated points s_j and $\bar{s}_j$ (which is not necessary to be the same) should be carefully taken to match the choice of the fluxes for stability. The choice of $\{s_j\}_{j=1}^k$ and $\{\bar{s}_j\}_{j=1}^k$ in RSV scheme and our later theoretical analysis will demonstrate this point.

Remark 2.2. For convection diffusion equations, *i.e.*, $u_t + f(u)_x = (\alpha u_x)_x$, we can also design the corresponding GSV and RSV schemes by using the LDG formulation with the alternating flux for diffusion term, and the upwind flux for the convection term. For simplicity and clarity, we focus our attention on the diffusion equations in this paper. The same argument can also be applied to convection-diffusion equations.

2.2. SV methods as a Petrov–Galerkin method

In this subsection, we reformulate the SV scheme (2.4) into its equivalent Petrov–Galerkin form, and then establish the stability of SV scheme under the the framework of Glerkin method.

We begin with the construction of two types of control volumes, which are defined by

$$\mathbf{V}_{i,j} = [x_{i,j}, x_{i,j+1}], \quad \bar{\mathbf{V}}_{i,j} = [\bar{x}_{i,j}, \bar{x}_{i,j+1}], \qquad (i,j) \in \mathbb{Z}_N \times \mathbb{Z}_k^0.$$

With these control volumes $\mathbf{V}_{i,j}, \bar{\mathbf{V}}_{i,j}$, we define the piecewise constant function space associated with $\mathbf{V}_{i,j}, \bar{\mathbf{V}}_{i,j}$ as

$$V_h = \left\{w^* : w^*|_{\mathbf{V}_{i,j}} \in \mathcal{P}_0, i \in \mathbb{Z}_N, j \in \mathbb{Z}_k^0\right\}, \ \bar{V}_h = \left\{\bar{w}^* : \bar{w}^*|_{\bar{\mathbf{V}}_{i,j}} \in \mathcal{P}_0, i \in \mathbb{Z}_N, j \in \mathbb{Z}_k^0\right\}.$$

Obviously, any function $w^* \in V_h, \bar{w}^* \in \bar{V}_h$ can be represented as

$$w^*(x,t) = \sum_{i=1}^N \sum_{j=0}^k w_{i,j}^* \chi_{\mathbf{V}_{i,j}}(x), \quad \bar{w}^*(x,t) = \sum_{i=1}^N \sum_{j=0}^k \bar{w}_{i,j}^* \chi_{\bar{\mathbf{V}}_{i,j}}(x),$$

where $w_{i,j}^*, \bar{w}_{i,j}^*, (i,j) \in \mathbb{Z}_N \times \mathbb{Z}_k^0$ are coefficients as functions of t, $\chi_A, A \subset [a,b]$ is the characteristic function defined as $\chi_A = 1$ in A and $\chi_A = 0$ otherwise.

Define the so-called broken Sobolev space as follows:

$$\mathcal{H}_h = \{v : v|_{\mathbf{V}_i} \in H^1, 1 \le i \le N\}.$$

For all $i \in \mathbb{Z}_N$ and any $(v, p, w^*, \bar{w}^*) \in \mathcal{H}_h \times \mathcal{H}_h \times V_h \times \bar{V}_h$, let

$$a_i^1(v, p; w^*) = \sum_{j=0}^{k} w_{i,j}^* \left(\int_{x_{i,j}}^{x_{i,j+1}} v_t(x,t)\,\mathrm{d}x - \hat{\beta}\hat{p}(x_{i,j+1}) + \hat{\beta}\hat{p}(x_{i,j}) \right), \tag{2.7}$$

$$a_i^2(v, p; \bar{w}^*) = \sum_{j=0}^{k} \bar{w}_{i,j}^* \left(\int_{\bar{x}_{i,j}}^{\bar{x}_{i,j+1}} p(x,t)\,\mathrm{d}x - \hat{B}(v(\bar{x}_{i,j+1}), \bar{x}_{i,j+1}) + \hat{B}(v(\bar{x}_{i,j}), \bar{x}_{i,j}) \right), \tag{2.8}$$

where the numerical fluxes $\hat{\beta} = \frac{B^+ - B^-}{v^+ - v^-}$, and $(\hat{p}, \hat{B}) = (p^-, B^+)$ or $(\hat{p}, \hat{B}) = (p^+, B^-)$. Denoting

$$a(v, p; w^*, \bar{w}^*) = \sum_{i=1}^{N} \big(a_i^1(v, p; w^*) + a_i^2(v, p; \bar{w}^*) \big). \tag{2.9}$$

Then it is easy to check that the SV solution (u_h, q_h) of (2.4) satisfies

$$a_i^1(u_h, q_h; w^*) = (g, w^*), \quad a_i^2(u_h, q_h; \bar{w}^*) = 0, \qquad \forall (w^*, \bar{w}^*) \in V_h \times \bar{V}_h, i \in \mathbb{Z}_N, \tag{2.10}$$

or equivalently,

$$a(u_h, q_h; w^*, \bar{w}^*) = (g, w^*), \qquad \forall (w^*, \bar{w}^*) \in V_h \times \bar{V}_h. \tag{2.11}$$

Conversely, if $u_h, q_h \in U_h$ satisfy (2.10) or (2.11), then by choosing $w^* = \chi_{\mathbf{V}_{i,j}}, \bar{w}^* = \chi_{\bar{\mathbf{V}}_{i,j}}$, we find that u_h, q_h satisfy (2.4). In other words, the SV method (2.4) is equivalent to the Petrov–Galerkin method (2.10) or (2.11).

2.3. Stability analysis for linear diffusion problems

In this subsection, we investigate the stability of the two SV schemes for the linear diffusion problem (2.2), *i.e.*, $\alpha = \alpha(x)$. In this case, $\beta = \sqrt{\alpha(x)}$ is continuous and thus the numerical fluxes $\hat{\beta} = \beta$ in (2.5) or (2.6).

We begin with the introduction of the special transformation from the trial space U_h to the test spaces $V_h, \bar{V}_h$.

2.3.1. Transformation from the trial space to the test space

Since the transformation we defined in this subsection is closely related to some quadratures points and their associated weights. We introduce some numerical quadratures and quadrature errors first.

Given any $f \in L^1([-1,1])$, suppose $s_j, \bar{s}_j, j \in \mathbb{Z}_{k+1}^0$ are points of some numerical quadrature to calculate $\int_{-1}^{1} f(s)\,\mathrm{d}s$ and $A_j, \bar{A}_j$ are associated weights. Define

$$Q_k(f) = \sum_{j=0}^{k+1} A_j f(s_j), \quad \bar{Q}_k(f) = \sum_{j=0}^{k+1} \bar{A}_j f(\bar{s}_j).$$

If s_j are taken as the Gauss points, then the above quadrature is referred as the standard Gauss–Legendre quadrature with $A_j = \frac{2}{(1-s_j^2)[L_k'(s_j)^2]}, j = 1, \ldots, k$ and $A_0 = A_{k+1} = 0$. Similarly, if s_j are taken as the right Radau points, we have $A_0 = 0$ and call the quadrature right Radau numerical quadrature. If s_j are taken as the left Radau points, we have $A_{k+1} = 0$. Note that the k-point Gauss quadrature and $(k+1)$-point right/left Radau quadrature is exact for polynomials of degree not more than $2k-1$ and $2k$, respectively.

Define the residual of the numerical quadrature by

$$R(f) = \int_{-1}^{1} f(s)\,\mathrm{d}s - Q_k(f), \quad \bar{R}(f) = \int_{-1}^{1} f(s)\,\mathrm{d}s - \bar{Q}_k(f).$$

By a scaling from $[-1,1]$ to τ_i, we get the residual of the numerical quadrature in each interval τ_i. That is,

$$R_i(f)=\int_{\tau_i} f(x)\,\mathrm{d}x-\sum_{j=0}^{k+1} f(x_{i,j})A_{i,j},\quad \bar{R}_i(f)=\int_{\tau_i} f(x)\,\mathrm{d}x-\sum_{j=0}^{k+1} f(\bar{x}_{i,j})\bar{A}_{i,j}, \tag{2.12}$$

where $A_{i,j}=\frac{h_i}{2}A_j, \bar{A}_{i,j}=\frac{h_i}{2}\bar{A}_j$ for all $(i,j)\in\mathbb{Z}_N\times\mathbb{Z}_{k+1}^0$.

Now we are ready to present the transformation from the trial space to the test space. for all $w\in U_h$, we define a transformation $\mathcal{F}: U_h\to V_h$ by

$$\mathcal{F}w=w^*:=\sum_{i=1}^{N}\sum_{j=0}^{k} w_{i,j}^*(t)\chi_{\mathbf{V}_{i,j}}(x), \tag{2.13}$$

where the $k+1$ coefficients $w_{i,j}^*, j\in\mathbb{Z}_k^0$ are given as

$$w_{i,0}^*=w\Big(x_{i-\frac{1}{2}}^+\Big)+A_{i,0}w_x\Big(x_{i-\frac{1}{2}}^+\Big),\qquad w_{i,j}^*-w_{i,j-1}^*=A_{i,j}w_x(x_{i,j}),\ j\in\mathbb{Z}_k. \tag{2.14}$$

By a direct calculation, we have

$$w_{i,k}^*=\sum_{j=1}^{k}\big(w_{i,j}^*-w_{i,j-1}^*\big)+w_{i,0}^*=w\Big(x_{i+\frac{1}{2}}^-\Big)-A_{i,k+1}w_x\Big(x_{i+\frac{1}{2}}^-\Big). \tag{2.15}$$

Consequently, we use the inverse inequality to get

$$\|\mathcal{F}w\|_0^2=\|w^*\|_0^2\lesssim\sum_{i=1}^{N}\sum_{j=0}^{k}h_i\big|w_{i,j}^*\big|^2\lesssim\|w\|_0^2,\qquad\forall w\in U_h. \tag{2.16}$$

Here and in the rest of this paper, by $A\lesssim B$ we mean that A can be bounded by B multiplied by a constant independent of the mesh size h.

Similarly, we can define a transformation $\bar{\mathcal{F}}: U_h\to\bar{V}_h$ by

$$\bar{\mathcal{F}}\bar{w}=\bar{w}^*:=\sum_{i=1}^{N}\sum_{j=0}^{k}\bar{w}_{i,j}^*\chi_{\mathbf{V}_{i,j}}(x) \tag{2.17}$$

with the coefficients $\bar{w}_{i,j}^*$ given as

$$\bar{w}_{i,0}^*=\bar{w}\Big(x_{i-\frac{1}{2}}^+\Big)+\bar{A}_{i,0}\bar{w}_x\Big(x_{i-\frac{1}{2}}^+\Big),\qquad \bar{w}_{i,j}^*-\bar{w}_{i,j-1}^*=\bar{A}_{i,j}\bar{w}_x(\bar{x}_{i,j}),\qquad j\in\mathbb{Z}_k. \tag{2.18}$$

Similar results are also valid for the transformation $\bar{\mathcal{F}}$, *i.e.*,

$$\bar{w}_{i,k}^*=\bar{w}\Big(x_{i+\frac{1}{2}}^-\Big)-\bar{A}_{i,k+1}\bar{w}_x\Big(x_{i+\frac{1}{2}}^-\Big),\quad\|\bar{\mathcal{F}}\bar{w}\|_0\le\|\bar{w}\|_0,\quad\forall\bar{w}\in U_h. \tag{2.19}$$

For any function $v,p\in\mathcal{H}_h$, we define

$$\partial_x^{-1}v=\int_a^x v\,\mathrm{d}x,\quad (p,v)_i=\int_{\tau_i}pv\,\mathrm{d}x,\quad (p,v)=\sum_{i=1}^{N}(p,v)_i.$$

Due to the special transformation $\mathcal{F},\bar{\mathcal{F}}$, we have the following relationship between inner product and the discrete inner product:

$$(v,\mathcal{F}w)_i=(v,w^*)_i=\sum_{j=0}^{k}w_{i,j}^*\big(\partial_x^{-1}v(x_{i,j+1})-\partial_x^{-1}v(x_{i,j})\big)$$

$$= \sum_{j=1}^{k} (w^*_{i,j-1} - w^*_{i,j}) \partial_x^{-1} v(x_{i,j}) + w^*_{i,k} \partial_x^{-1} v\Big(x^-_{i+\frac{1}{2}}\Big) - w^*_{i,0} \partial_x^{-1} v\Big(x^+_{i-\frac{1}{2}}\Big).$$

By using (2.14) and (2.15) and the integration by parts, we get

$$\begin{aligned}(v, \mathcal{F}w)_i &= -\sum_{j=0}^{k+1} A_{i,j} \big(w_x \partial_x^{-1} v\big)(x_{i,j}) + \big(w \partial_x^{-1} v\big)\Big(x^-_{i+\frac{1}{2}}\Big) - \big(w \partial_x^{-1} v\big)\Big(x^+_{i-\frac{1}{2}}\Big) \\ &= (v, w)_i + R_i\big(\partial_x^{-1} v w_x\big). \end{aligned} \tag{2.20}$$

Similarly, there holds

$$\big(v, \bar{\mathcal{F}} w\big)_i = (v, w)_i + \bar{R}_i\big(\partial_x^{-1} v w_x\big). \tag{2.21}$$

Here for any f, $R_i(f), \bar{R}_i(f)$ denote the numerical quadrature errors defined in (2.12).

2.3.2. L^2 stability

We first study the bilinear terms a_i^1, a_i^2 defined in (2.7) and (2.8).

Lemma 2.3. *For any $w, \bar{w} \in U_h$, let $w^* = \mathcal{F}w, \bar{w}^* = \bar{\mathcal{F}}\bar{w}$ with $\mathcal{F}, \bar{\mathcal{F}}$ defined in (2.13) and (2.17), respectively. Then for all $(v, p) \in \mathcal{H}_h \times \mathcal{H}_h$,*

$$a_i^1(v, p; w^*) = (v_t, w^*)_i + (\beta p, w_x)_i - \beta \hat{p} w^-|_{i+\frac{1}{2}} + \beta \hat{p} w^+|_{i-\frac{1}{2}} - R_i(w_x \beta p), \tag{2.22}$$

$$a_i^2(v, p; \bar{w}^*) = (p, \bar{w}^*)_i + (B, \bar{w}_x)_i - \hat{B} \bar{w}^-|_{i+\frac{1}{2}} + \hat{B} \bar{w}^+|_{i-\frac{1}{2}} - \bar{R}_i(\bar{w}_x B). \tag{2.23}$$

Here $v^\pm|_{i+\frac{1}{2}} = v(x^\pm_{i+\frac{1}{2}})$ and $B = B(v, x)$ is defined by (2.3) with u replaced by v.

Proof. In each element τ_i, if not otherwise stated, we use the notation $v|_{i+\frac{1}{2}}, v|_{i-\frac{1}{2}}$ to denote the value of the function v at the boundary points $x_{i+\frac{1}{2}}, x_{i-\frac{1}{2}}$, respectively. That is,

$$v|_{i+\frac{1}{2}} = v\Big(x^-_{i+\frac{1}{2}}\Big), \quad v|_{i-\frac{1}{2}} = v\Big(x^+_{i-\frac{1}{2}}\Big).$$

Recalling the definition of a_i^1 in (2.7) and using (2.14) and (2.15), we get

$$\begin{aligned} a_i^1(v, p; w^*) &= (v_t, w^*)_i - w^*_{i,0}\Big(\beta p - \hat{\beta}\hat{p}\Big)\Big(x_{i-\frac{1}{2}}\Big) + w^*_{i,k}\Big(\beta p - \hat{\beta}\hat{p}\Big)\Big(x_{i+\frac{1}{2}}\Big) - \sum_{j=0}^{k} w^*_{i,j} \int_{x_{i,j}}^{x_{i,j+1}} (\beta p)_x \\ &= (v_t, w^*)_i - ((\beta p)_x, w^*)_i + \beta w(p - \hat{p})|_{i+\frac{1}{2}} - \beta w(p - \hat{p})|_{i-\frac{1}{2}} + I \\ &= (v_t, w^*)_i - ((\beta p)_x, w)_i + \beta w(p - \hat{p})|_{i+\frac{1}{2}} - \beta w(p - \hat{p})|_{i-\frac{1}{2}} + I - R_i(\beta p w_x), \end{aligned}$$

where in the last step, we have used (2.20) and

$$I = -A_{i,0} w_x\Big(\beta p - \hat{\beta}\hat{p}\Big)|_{i-\frac{1}{2}} - A_{i,k+1} w_x\Big(\beta p - \hat{\beta}\hat{p}\Big)|_{i+\frac{1}{2}}.$$

We next estimate I. Noticing that $A_0 = A_{k+1} = 0$ for LSV schemes, we have $I = 0$. As for RSV schemes, if the fluxes are chosen as (2.5) (*i.e.*, $\hat{p} = p^-$), then $x_{i,j}, j \in \mathbb{Z}_k$ are taken as k interior right Radau points and thus,

$$A_{i,0} = 0, \quad \Big(\beta p - \hat{\beta}\hat{p}\Big)|_{i+\frac{1}{2}} = 0.$$

Similarly, if the fluxes are chosen as (2.6) (*i.e.*, $\hat{p} = p^+$), then $x_{i,j}, j \in \mathbb{Z}_k$ are chosen as interior left Radau points, which yields

$$A_{i,k+1} = 0, \quad \left(\beta p - \hat{\beta}\hat{p}\right)|_{i-\frac{1}{2}} = 0.$$

Then for both the LSV and RSV, we have $I = 0$. Consequently,

$$\begin{aligned} a_i^1(v, p; w^*) &= (v_t, w^*)_i - ((\beta p)_x, w)_i + \beta w(p - \hat{p})_{i+\frac{1}{2}} - \beta w(p - \hat{p})_{i-\frac{1}{2}} - R_i(\beta p w_x) \\ &= (v_t, w^*)_i + (\beta p, w_x)_i - \beta w^-_{i+\frac{1}{2}}\hat{p}_{i+\frac{1}{2}} + \beta w^+_{i-\frac{1}{2}}\hat{p}_{i-\frac{1}{2}} - R_i(\beta p w_x), \end{aligned}$$

where in the last step, we have used the integration by parts. Then (2.22) follows. Following the same argument, we can obtain (2.23). This finishes our proof. □

For any $v \in U_h$, we introduce equivalent L^2-norms defined by

$$(v, \mathcal{F}v) = \sum_{i=1}^{N}(v, \mathcal{F}v)_i, \quad \left(v, \bar{\mathcal{F}}v\right) = \sum_{i=1}^{N}\left(v, \bar{\mathcal{F}}v\right)_i. \tag{2.24}$$

It was proved in [3] that for GSV, there holds

$$(v, v) \le (v, \mathcal{F}v) \lesssim (v, v), \quad (v, v) \le \left(v, \bar{\mathcal{F}}v\right) \lesssim (v, v). \tag{2.25}$$

As for RSV, since both the right and left Radau numerical quadratures are exact for polynomial of degree not more than $2k$, we have from (2.20) and (2.21) that

$$(v, v) = (v, \mathcal{F}v) = \left(v, \bar{\mathcal{F}}v\right).$$

In other words, (2.25) holds true for both GSV and RRSV, which indicates $(v, \mathcal{F}v)$ and $(v, \bar{\mathcal{F}}v)$ are both equivalent to the standard L^2-norm.

Now we are ready to present the L^2 stability for both GSV and RSV schemes.

Theorem 2.4. *Let (u_h, q_h) be the solution of (2.11). Then both GSV and RSV are L^2 stable, i.e.,*

$$\|u_h(\cdot, t)\|_0^2 + \int_0^t \|q_h\|_0^2 \, \mathrm{d}t \lesssim \|u_0\|_0^2 + \int_0^t \|g\|_0^2 \, \mathrm{d}t, \quad \forall t \in (0, t_0]. \tag{2.26}$$

Proof. By choosing $w = v, \bar{w} = p$ in (2.22), (2.23) and using the numerical quadrature error, we get for all $v, p \in U_h$ and $v^* = \mathcal{F}v, p^* = \bar{\mathcal{F}}p$ that

$$\begin{aligned} a_i^1(v, p; v^*) &= (v_t, v^*)_i + (\beta p, v_x)_i - \beta\hat{p}v^-|_{i+\frac{1}{2}} + \beta\hat{p}v^+|_{i-\frac{1}{2}} - R_i(\beta p v_x), \\ a_i^2(v, p; p^*) &= (p, p^*)_i + (B(v, x), p_x)_i - \hat{B}p^-|_{i+\frac{1}{2}} + \hat{B}p^+|_{i-\frac{1}{2}} - \bar{R}_i(Bp_x). \end{aligned}$$

Recalling the definition of B in (2.3) and using the continuity of the linear variable function $\beta = \beta(x)$, we have $B(v, x)_x = \beta v_x, [B]_{i+\frac{1}{2}} = \beta[v]_{i+\frac{1}{2}}$. Then summing up all i from 1 to N yields

$$\begin{aligned} \sum_{i=1}^{N} a_i^1(v, p; v^*) &= (v_t, v^*) + (p, B_x) + \sum_{i=1}^{N}\left(\hat{p}[B]|_{i+\frac{1}{2}} - R_i(\beta p v_x)\right), \\ \sum_{i=1}^{N} a_i^2(v, p; v^*) &= (p, p^*) + (B, p_x) + \sum_{i=1}^{N}\left(\hat{B}[p]|_{i+\frac{1}{2}} - \bar{R}_i(Bp_x)\right). \end{aligned}$$

Consequently,

$$(v_t, \mathcal{F}v) + \big(p, \bar{\mathcal{F}}p\big) = a\big(v, p; \mathcal{F}v, \bar{\mathcal{F}}p\big) + \sum_{i=1}^{N} \big(R_i(\beta p v_x) + \bar{R}_i(Bp_x)\big). \tag{2.27}$$

We next estimate the numerical quadrature errors $R_i(\beta p v_x)$ and $\bar{R}_i(Bp_x)$. Given any function f, we define $I_r f|_{\tau_i} \in \mathbb{P}_r$ the Lagrange interpolation function of f with the interpolation point $x_{i,j}, 0 \le j \le r+1$, here $r = k$ for GSV and $r = k+1$ for the RSV. Since $R_i(f) = 0, f \in \mathbb{P}_{2k-1}$ for GSV and $R_i(f) = 0, f \in \mathbb{P}_{2k}$ for RSV, then we have for both GSV and RSV that

$$|R_i(fv_x)| = 0\left|\int_{\tau_i} (f - I_r f)v_x \,\mathrm{d}x\right| \lesssim \|f - I_r f\|_{0,\tau_i}\|v_x\|_{0,\tau_i} \lesssim h^l \|v_x\|_{0,\tau_i}\|f\|_{l,\tau_i}, \qquad l \le k+1. \tag{2.28}$$

The above identity is also valid for $\bar{R}_i(fv_x)$. Consequently,

$$|R_i(\beta p v_x)| = |R_i((\beta - \beta_i)pv_x)| \lesssim \|(\beta - \beta_i)p\|_{0,\tau_i}\|v_x\|_{0,\tau_i}.$$

Here β_i denotes the cell average of β in the interval τ_i. Similarly, there holds

$$\big|\bar{R}_i(p_x B)\big| = \big|\bar{R}_i(p_x(B - I_r B))\big| \lesssim h^{k+1}\|p_x\|_{0,\tau_i}\big\|\partial_x^{k+1} B\big\|_{0,\tau_i}.$$

Using the inverse inequality and the fact that $B(v,x)_x = \beta v_x$, we have for all $v, p \in U_h$ that

$$|R_i(\beta p v_x)| \lesssim \|p\|_{0,\tau_i}\|v\|_{0,\tau_i}, \quad \big|\bar{R}_i(p_x B)\big| \lesssim h^k \|p\|_{0,\tau_i}\|v\|_{k,\tau_i} \lesssim \|p\|_{0,\tau_i}\|v\|_{0,\tau_i}.$$

Substituting the above two inequalities into (2.27), we get for both LSV and RSV that

$$(v_t, \mathcal{F}v) + \big(p, \bar{\mathcal{F}}p\big) \lesssim a\big(v, p; \mathcal{F}v, \bar{\mathcal{F}}p\big) + \|v\|_0\|p\|_0. \tag{2.29}$$

Especially, we choose $(v, p) = (u_h, q_h)$ in the above inequality and use (2.11), (2.25) and Cauchy–Schwarz inequality and then obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}(u_h, \mathcal{F}u_h) + \big(q_h, \bar{\mathcal{F}}q_h\big) \lesssim \|u_h\|_0^2 + \|g\|_0^2,$$

which yields

$$(u_h, \mathcal{F}u_h)(t) + \int_0^t \big(q_h, \bar{\mathcal{F}}q_h\big)\,\mathrm{d}t \lesssim (u_h, \mathcal{F}u_h)(0) + \int_0^t \big(\|u_h\|_0^2 + \|g\|_0^2\big)\,\mathrm{d}t.$$

Then (2.26) follows from the Gronwall inequality and the equivalence (2.25). □

We end with this subsection the discussion on the relationship between the proposed SV schemes and the LDG schemes. Recall the bilinear form of the LDG schemes for (2.2), which is defined by

$$\begin{aligned}
\bar{a}_i^1(v, p; w) &= (v_t, w)_i + (\beta p, w_x)_i - \hat{\beta}\hat{p}w^-|_{i+\frac{1}{2}} + \hat{\beta}\hat{p}w^+|_{i-\frac{1}{2}},\\
\bar{a}_i^2(v, p; w) &= (p, w)_i + (B(v,x), w_x)_i - \hat{B}w^-|_{i+\frac{1}{2}} + \hat{B}w^+|_{i-\frac{1}{2}}.
\end{aligned}$$

Denoting

$$a_{\mathrm{LDG}}(v, p; w, \bar{w}) = \sum_{i=1}^{N} \bar{a}_i(v, p; w, \bar{w}), \quad \bar{a}_i(v, p; w, \bar{w}) = \bar{a}_i^1(v, p; w) + \bar{a}_i^2(v, p; \bar{w}).$$

The LDG method for (2.2) is: Find $(u_h, q_h) \in U_h$ such that for all $w, \bar{w} \in U_h$

$$a_{\mathrm{LDG}}(u_h, q_h; w, \bar{w}) = (g, w).$$

In light of (2.22), (2.23) and (2.20), we easily obtain for all $w, \bar{w} \in U_h$

$$a(v,p;w^*,\bar{w}^*) = a_{\mathrm{LDG}}(v,p;w,\bar{w}) + \sum_{i=1}^{N}\big(R_i\big(w_x\big(\partial_x^{-1}v_t - \beta p\big)\big) + \bar{R}_i\big(\bar{w}_x(\partial_x^{-1}p - B)\big)\big), \tag{2.30}$$

where $w^* = \mathcal{F}w, \bar{w}^* = \bar{\mathcal{F}}\bar{w}$. As we may observe, the bilinear forms of the proposed LSV and RSV schemes are equivalent to those of the LDG schemes, up to some Gauss or Radau numerical quadrature errors. Especially, for constant coefficients problem (*i.e.*, $\alpha = const$), we have

$$B(v,x) = \beta v, \quad \bar{w}_x\big(\partial_x^{-1}p - B\big) \in \mathbb{P}_{2k}, \quad w_x\big(\partial_x^{-1}v_t - \beta p\big) \in \mathbb{P}_{2k}, \quad \forall v,p,w,\bar{w} \in U_h.$$

Note that the left/right Radau numerical quadrature is exact for polynomial of degree not more than $2k$. Then for RSV,

$$\bar{R}_i\big(\bar{w}_x\big(\partial_x^{-1}p - B\big)\big) = R_i\big(w_x\big(\partial_x^{-1}v_t - \beta p\big)\big) = 0.$$

Consequently,

$$a(v,p;w^*,\bar{w}^*) = a_{\mathrm{LDG}}(v,p;w,\bar{w}), \quad \forall v,p \in U_h,$$

which indicates that the RSV method is identical to the LDG method for constant diffusion problems with the source function $g(x,t) \in \mathbb{P}_{2k}$.

2.4. Error estimates for linear diffusion problems

This subsection is dedicated to the error estimates for linear diffusion problems, *i.e.*, $\alpha = \alpha(x)$. To study the convergence of SV methods, we first introduce some interpolation functions. For any function $v \in \mathcal{H}_h$, we denote by $I_h^+v, I_h^-v \in U_h$ the standard Lagrange interpolation of v, which satisfy the following $k+1$ conditions in each element τ_i:

$$\begin{aligned} &I_h^-v(x_{i,j}) = v(x_{i,j}), \quad I_h^-v\Big(x^-_{i+\frac{1}{2}}\Big) = v\Big(x^-_{i+\frac{1}{2}}\Big), \qquad \forall (i,j) \in \mathbb{Z}_N \times \mathbb{Z}_k, \\ &I_h^+v(x_{i,j}) = v(x_{i,j}), \quad I_h^+v\Big(x^+_{i-\frac{1}{2}}\Big) = v\Big(x^+_{i-\frac{1}{2}}\Big), \qquad \forall (i,j) \in \mathbb{Z}_N \times \mathbb{Z}_k. \end{aligned}$$

Similarly, we can define the interpolation functions $\bar{I}_h^+v, \bar{I}_h^-v \in U_h$ of v, with the interpolation points $\bar{x}_{i,j}$ substituting of $x_{i,j}$ in the above definition. By the standard approximation theory of the interpolation function, we have

$$\big\|I_h^-v - v\big\|_{0,m} + \big\|I_h^+v - v\big\|_{0,m} + \big\|\bar{I}_h^-v - v\big\|_{0,m} + \big\|\bar{I}_h^+v - v\big\|_{0,m} \lesssim h^r\|v\|_{r,m}, \qquad r \le k+1.$$

Define

$$(I_hu, I_hq) = \begin{cases} \big(\bar{I}_h^+u, I_h^-q\big), & \text{for flux (2.5)}, \\ \big(\bar{I}_h^-u, I_h^+q\big), & \text{for flux (2.6)}. \end{cases} \tag{2.31}$$

Due to the special construction of the interpolation function, we have for both fluxes (2.5) and (2.6) that

$$\widehat{I_hu}(\bar{x}_{i,j}) = u(\bar{x}_{i,j}), \quad \widehat{I_hq}(x_{i,j}) = q(x_{i,j}), \quad (i,j) \in \mathbb{Z}_N \times \mathbb{Z}_{k+1}^0. \tag{2.32}$$

Theorem 2.5. *Let u be the solution of (2.1) satisfying $u \in H^{k+2}, u_t \in H^{k+1}$, and (u_h, q_h) be the solution of (2.11) with the initial solution chosen as $u_h(x,0) = I_hu_0$. Then for both the flux choices (2.5) and (2.6) and both GSV and RSV,*

$$\|u - u_h\|_0 + \int_0^t \|q - q_h\|_0 \,\mathrm{d}t \lesssim h^{k+1}(\|u_t\|_{k+1} + \|u\|_{k+2}). \tag{2.33}$$

Proof. Denoting

$$\xi_u = u_h - I_h u, \quad \xi_q = q_h - I_h q.$$

By taking $(v,p) = (\xi_u, \xi_q)$ in (2.29) and the orthogonality $a(u-u_h, q-q_h; w^*, \bar{w}^*) = 0$ for all $w, \bar{w} \in U_h$, we get

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}(\xi_u, \mathcal{F}\xi_u) + (\xi_q, \bar{\mathcal{F}}\xi_q) &\lesssim a(\xi_u, \xi_q; \mathcal{F}\xi_u, \bar{\mathcal{F}}\xi_q) + \|\xi_u\|_0\|\xi_q\|_0 \\ &= a(u - I_h u, q - I_h q; \mathcal{F}\xi_u, \bar{\mathcal{F}}\xi_q) + \|\xi_u\|_0\|\xi_q\|_0. \end{aligned} \tag{2.34}$$

Recalling the definitions of a_i^1 in (2.7) and using (2.32), we easily get for all $w^* \in V_h, \bar{w}^* \in \bar{V}_h$ that

$$a_i^1(u - I_h u, q - I_h q; w^*) = (u_t - I_h u_t, w^*)_i.$$

On the other hand, by denoting $v|_{i,j} = v(\bar{x}_{i,j})$ and using the continuity of $\beta(x)$ and (2.3), we have

$$[B]|_{i,j} := B^+(\bar{x}_{i,j}) - B^-(\bar{x}_{i,j}) = \int_{\bar{x}_{i,j}^-}^{\bar{x}_{i,j}^+} \beta v_x \,\mathrm{d}x = \beta[v]|_{i,j}.$$

Consequently, for both $\hat{B} = B^+$ or $\hat{B} = B^-$,

$$\begin{aligned}\hat{B}(\bar{x}_{i,j+1}) - \hat{B}(\bar{x}_{i,j}) &= \int_{\bar{x}_{i,j}}^{\bar{x}_{i,j+1}} \beta v_x \,\mathrm{d}x + \beta(\hat{v} - v)|_{i,j+1} - \beta(\hat{v} - v)|_{i,j} \\ &= \int_{\bar{x}_{i,j}}^{\bar{x}_{i,j+1}} \beta_x v \,\mathrm{d}x + \beta\hat{v}|_{i,j+1} - \beta\hat{v}|_{i,j},\end{aligned}$$

which yields, together with the definition of a_i^2 in (2.8),

$$a_i^2(v, p; \bar{w}^*) = (p, \bar{w}^*)_i - (\beta_x v, \bar{w}^*)_i - \sum_{j=0}^{k} \bar{w}_{i,j}^*(\beta\hat{v}|_{i,j+1} - \beta\hat{v}|_{i,j}).$$

In light of (2.32), we easily obtain

$$a_i^2(u - I_h u, q - I_h q; \bar{w}^*) = (q - I_h q, \bar{w}^*)_i - (\beta_x(u - I_h u), \bar{w}^*).$$

Consequently, we have from (2.16) and the second inequality of (2.19) that

$$\begin{aligned}\big|a\big(u - I_h u, q - I_h q; \mathcal{F}\xi_u, \bar{\mathcal{F}}\xi_q\big)\big| &= |(u_t - I_h u_t, \mathcal{F}\xi_u)| + \big|\big(q - I_h q - \beta_x(u - I_h u), \bar{\mathcal{F}}\xi_q\big)\big| \\ &\lesssim h^{k+1}(\|u_t\|_{k+1} + \|u\|_{k+2})(\|\xi_u\|_0 + \|\xi_q\|_0).\end{aligned}$$

Substituting the above inequality into (2.34) and using the Cauchy–Schwarz inequality gives

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}(\xi_u, \mathcal{F}\xi_u) + (\xi_q, \bar{\mathcal{F}}\xi_q) \le C\Big(h^{2(k+1)}(\|u_t\|_{k+1} + \|u\|_{k+2})^2 + \|\xi_u\|_0^2\Big) + \frac{1}{2}\|\xi_q\|_0^2,$$

where C is a constant independent of the mesh size h. Integrating the above inequality from 0 to t and using (2.25) and the Gronwall inequality, we get

$$\|\xi_u\|_0^2 + \int_0^t \|\xi_q\|_0^2 \,\mathrm{d}t \lesssim h^{2(k+1)}(\|u_t\|_{k+1} + \|u\|_{k+2})^2.$$

Then the desired result follows from the approximation properties of $(I_h u, I_h q)$. □

Remark 2.6. The stability result and optimal error estimates established in Theorems 2.4 and 2.5 are mainly based on linear diffusion equations. Although our numerical experiments show that the stability result and optimal error estimtates hold true for nonlinear equations, its theoretical analysis is still lacking. Just as demonstrated in Theorem 2.4, the numerical quadrature errors appeared in the right hand side of (2.27) can not be bounded by the L^2 norms of v and p for nonlinear equations, which leads to the difficulty in stability. Choosing suitable SV numerical fluxes may provide a new approach to study the stability of SV method for nonlinear equations. The theoretical analysis of SV methods for nonlinear problems is interesting yet difficult, and it deserves a separate study.

3. SV method for multidimensional linear diffusion problems

In this section, we present the SV scheme for multidimensional linear diffusion problems and investigate its stability and convergence. The analysis for multidimensional nonlinear diffusion problems deserves a separate study and thus we skip it here. To simplify our analysis and make the idea clear, we will use the two dimensional linear diffusion problems as a model. The methodology we adopt can also be applied to linear diffusion equation in higher dimensions, *e.g.*, three dimension.

We consider the SV method for the following two dimensional linear diffusion problems

$$\begin{aligned} &u_t = \nabla\cdot(\alpha\nabla u) \qquad (x,y,t)\in[a,b]\times[c,d]\times(0,t_0],\\ &u(x,y,0)=u_0(x,y), \end{aligned} \tag{3.1}$$

where both u_0 and $\alpha := \alpha(x,y) \geq 0$ are smooth. For simplicity, we still consider the periodic boundary condition.

We rewrite the above equation into its equivalent system:

$$u_t = \nabla\cdot(\beta\mathbf{q}), \quad \mathbf{q} = \beta\nabla u := \nabla B, \quad \beta = \sqrt{\alpha}. \tag{3.2}$$

3.1. SV schemes

Let $a = x_{\frac12} < x_{\frac32} < \cdots < x_{M+\frac12} = b$ and $c = y_{\frac12} < y_{\frac32} < \cdots < y_{N+\frac12} = d$. Denote by $\mathcal{T}_h$ the rectangular partition of Ω. That is,

$$\mathcal{T}_h = \Big\{\tau_{m,n} = \Big[x_{m-\frac12}, x_{m+\frac12}\Big]\times\Big[y_{n-\frac12}, y_{n+\frac12}\Big] : (m,n)\in\mathbb{Z}_M\times\mathbb{Z}_N\Big\}.$$

For any $\tau\in\mathcal{T}_h$, we denote by h_τ^x, h_τ^y the lengths of x- and y-directional edges of τ, respectively. h is the maximal length of all edges, and $h_{\min} = \min_\tau(h_\tau^x, h_\tau^y)$. Let $\tau_m^x = [x_{m-\frac12}, x_{m+\frac12}]$, $\tau_n^y = [y_{n-\frac12}, y_{n+\frac12}]$. We assume that the mesh $\mathcal{T}_h$ is *quasi-uniform* in the sense that there exist constants $c_1, c_2 > 0$ such that

$$h \leq c_1 h_\tau^x, \quad h \leq c_2 h_\tau^y \quad \forall\tau\in\mathcal{T}_h.$$

Denote by $\mathbb{Q}_k(x,y) = \mathbb{P}_k(x)\times\mathbb{P}_k(y)$ the tensor product bi-k polynomial space, and define the finite element space

$$U_h = \{v : \ v|_\tau\in\mathbb{Q}_k, \tau\in\mathcal{T}_h\}. \ \ \mathbf{U}_h = \big\{v : \ v|_\tau\in[\mathbb{Q}_k]^2, \tau\in\mathcal{T}_h\big\}.$$

Let $-1 = s_0 < s_1 < \ldots < s_k < s_{k+1} = 1$, and $-1 = \bar{s}_0 < \bar{s}_1 < \ldots < \bar{s}_k < \bar{s}_{k+1} = 1$, and define $\hat{\tau} = [-1,1]\times[-1,1]$. Assume that F_τ is the affine mapping from $\hat{\tau}$ to τ, and

$$\begin{aligned} \mathcal{G}_\tau &= \big\{g_{i,j}^\tau = \big(g_i^{\tau,x}, g_j^{\tau,y}\big) : g_{i,j}^\tau = F_\tau(s_i,s_j), i\in\mathbb{Z}_{k+1}^0, j\in\mathbb{Z}_{k+1}^0\big\},\\ \bar{\mathcal{G}}_\tau &= \big\{\bar{g}_{i,j}^\tau = \big(\bar{g}_i^{\tau,x}, \bar{g}_j^{\tau,y}\big) : \bar{g}_{i,j}^\tau = F_\tau(\bar{s}_i,\bar{s}_j), i\in\mathbb{Z}_{k+1}^0, j\in\mathbb{Z}_{k+1}^0\big\}. \end{aligned}$$

Two types of control volumes can be constructed by using the above $(k+2)^2$ points, *i.e.*,

$$V_{i,j}^\tau = \big[g_i^{\tau,x}, g_{i+1}^{\tau,x}\big]\times\big[g_j^{\tau,y}, g_{j+1}^{\tau,y}\big], \ \bar{V}_{i,j}^\tau = \big[\bar{g}_i^{\tau,x}, \bar{g}_{i+1}^{\tau,x}\big]\times\big[\bar{g}_j^{\tau,y}, \bar{g}_{j+1}^{\tau,y}\big], \quad (i,j)\in\mathbb{Z}_k^0\times\mathbb{Z}_k^0.$$

The SV schemes for (3.2) is to find a $u_h \in U_h, \mathbf{q}_h \in \mathbf{U}_h$ such that for all $(i,j) \in \mathbb{Z}_k^0 \times \mathbb{Z}_k^0$

$$\begin{aligned} \int_{V_{i,j}^\tau} \partial_t u_h \,\mathrm{d}x\,\mathrm{d}y - \int_{\partial V_{i,j}^\tau} \beta \hat{\mathbf{q}}_h \cdot \mathbf{n}\,\mathrm{d}s = 0, \\ \int_{\bar{V}_{i,j}^\tau} \mathbf{q}_h \,\mathrm{d}x\,\mathrm{d}y - \int_{\partial \bar{V}_{i,j}^\tau} \hat{B}\mathbf{n}_0 \cdot \mathbf{n}\,\mathrm{d}s = 0. \end{aligned} \tag{3.3}$$

Here $\hat{\mathbf{q}}_h, \hat{B}$ denote the numerical fluxes, $\mathbf{n}_0 = (1,1)$, and $\mathbf{n}$ is the outward normal vector. We still take the alternating flux as our numerical fluxes, *i.e.*,

$$\left(\hat{\mathbf{q}}_h, \hat{B}\right) = \left(\mathbf{q}_h^+, B^-\right), \text{ or } \left(\hat{\mathbf{q}}_h, \hat{B}\right) = \left(\mathbf{q}_h^-, B^+\right). \tag{3.4}$$

Similar to the 1D case, we consider two classes of SV schemes here, *i.e.*, GSV schemes in which $s_j, \bar{s}_j$ are chosen as Gauss points, and RSV schemes in which $s_j, \bar{s}_j$ are chosen as right or left Radau points dependent on the choice of numerical fluxes.

We next reformulate the SV scheme (3.3) into its equivalent Petorv–Glerkin form. To this end, we first define two piecewise test spaces as follows.

$$V_h = \left\{ w^* : w^*|_{V_{i,j}^\tau} \in \mathbb{P}_0, (i,j) \in \mathbb{Z}_k^0 \times \mathbb{Z}_k^0 \right\}, \ \mathbf{V}_h = \left\{ \mathbf{w}^* : \mathbf{w}^*|_{\bar{V}_{i,j}^\tau} \in [\mathbb{P}_0]^2, (i,j) \in \mathbb{Z}_k^0 \times \mathbb{Z}_k^0 \right\}.$$

Note that any function $w^* \in V_h, \mathbf{w}^* \in \mathbf{V}_h$ can be represented as

$$w^*(x,y,t) = \sum_{\tau \in \mathcal{T}_h} \sum_{i,j=0}^{k} w_{i,j}^{\tau,*}(t) \chi_{V_{i,j}^\tau}(x,y), \quad \mathbf{w}^*(x,y,t) = \sum_{\tau \in \mathcal{T}_h} \sum_{i,j=0}^{k} \mathbf{w}_{i,j}^{\tau,*}(t) \chi_{\bar{V}_{i,j}^\tau}(x,y).$$

For all $w^* \in V_h, \mathbf{w}^* \in \mathbf{V}_h$, we define

$$\begin{aligned} a_\tau^1(v, \mathbf{p}; w^*) &= \sum_{i,j=0}^{k} w_{i,j}^{\tau,*} \left(\int_{V_{i,j}^\tau} \partial_t v dx dy - \int_{\partial V_{i,j}^\tau} \beta \hat{\mathbf{p}} \cdot \mathbf{n}\,\mathrm{d}s \right), \\ a_\tau^2(v, \mathbf{p}; \mathbf{w}^*) &= \sum_{i,j=0}^{k} \mathbf{w}_{i,j}^{\tau,*} \left(\int_{\bar{V}_{i,j}^\tau} \mathbf{p} \,\mathrm{d}x\,\mathrm{d}y - \int_{\partial \bar{V}_{i,j}^\tau} \hat{B}\mathbf{n}_0 \cdot \mathbf{n}\,\mathrm{d}s \right). \end{aligned} \tag{3.5}$$

Then the SV scheme (3.3) can be rewritten as: Find a $u_h \in U_h, \mathbf{q}_h \in \mathbf{U}_h$ such that

$$a(u_h, \mathbf{q}_h; w^*, \mathbf{w}^*) = 0, \quad \forall w^* \in V_h, \mathbf{w}^* \in \mathbf{V}_h, \tag{3.6}$$

where

$$a(u_h, \mathbf{q}_h; w^*, \mathbf{w}^*) = \sum_{\tau \in \mathcal{T}_h} \left(a_\tau^1(u_h, \mathbf{q}_h; w^*) + a_\tau^2(u_h, \mathbf{q}_h; \mathbf{w}^*) \right).$$

3.2. Transformation from the trial space to the test space

Similar to the 1D problem, we need to define a special transformation from the trial space to the test space, which is of great importance in our stability analysis.

For any $\tau = \tau_{p,q} \in \mathcal{T}_h$, we denote by $R_\tau^x(f)$ and $R_\tau^y(f)$ the corresponding residual error of numerical quadrature along x and y directions, respectively. That is,

$$R_\tau^x f(y) = \int_{x_{p-\frac{1}{2}}}^{x_{p+\frac{1}{2}}} f(x,y)\,\mathrm{d}x - \sum_{i=0}^{k+1} A_i^{\tau,x} f(g_i^{\tau,x}, y), \tag{3.7}$$

$$R_\tau^y f(x) = \int_{y_{q-\frac{1}{2}}}^{y_{q+\frac{1}{2}}} f(x,y)\,\mathrm{d}y - \sum_{j=0}^{k+1} A_j^{\tau,y} f(x, g_j^{\tau,y}), \qquad \forall f \in L^1(\tau). \tag{3.8}$$

Here $A_i^{\tau,x} = \frac{h_\tau^x}{2} A_i, A_j^{\tau,y} = \frac{h_\tau^y}{2} A_j$. Let $A_{i,j}^\tau = A_i^{\tau,x} A_j^{\tau,y}, (i,j) \in \mathbb{Z}_{k+1}^0 \times \mathbb{Z}_{k+1}^0$. The residual error $\bar{R}_\tau^x(f)$ and $\bar{R}_\tau^y(f)$ can be defined by the similar way with $g_i^{\tau,x}, g_j^{\tau,y}$ replaced by $\bar{g}_i^{\tau,x}, \bar{g}_j^{\tau,y}$. For any $w^* \in V_h$, we denote the jump of w^* at $g_{i,j}^\tau$ along x and y directions as

$$\big[w_{i,j}^{\tau,*}\big]^x = w_{i,j}^{\tau,*} - w_{i-1,j}^{\tau,*}, \quad \big[w_{i,j}^{\tau,*}\big]^y = w_{i,j}^{\tau,*} - w_{i,j-1}^{\tau,*},$$

and the double layer jump as

$$\big[w_{i,j}^{\tau,*}\big] = w_{i,j}^{\tau,*} + w_{i-1,j-1}^{\tau,*} - w_{i-1,j}^{\tau,*} - w_{i,j-1}^{\tau,*}, \quad (i,j) \in \mathbb{Z}_k \times \mathbb{Z}_k.$$

Now we define transformation $\mathcal{F} : U_h \to V_h$ by

$$\mathcal{F}w|_\tau = w^{\tau,*} := \sum_{i=0}^{k}\sum_{j=0}^{k} w_{i,j}^{\tau,*}(t)\chi_{V_{i,j}^\tau}(x,y), \qquad \forall \tau \in \mathcal{T}_h, \tag{3.9}$$

where the $(k+1)^2$ coefficients $w_{i,j}^{\tau,*}, (i,j) \in \mathbb{Z}_k^0 \times \mathbb{Z}_k^0$ are defined as

$$\begin{aligned}
w_{0,0}^{\tau,*} &= \big(w + A_{0,0}^\tau w_{xy} + A_0^{\tau,x} w_x + A_0^{\tau,y} w_y\big)\big(g_{0,0}^\tau\big),\\
\big[w_{i,j}^{\tau,*}\big] &= A_{i,j}^\tau w_{xy}\big(g_{i,j}^\tau\big),\\
\big[w_{i,0}^{\tau,*}\big]^x &= \big(A_i^{\tau,x} w_x + A_{i,0}^\tau w_{xy}\big)\big(g_{i,0}^\tau\big),\\
\big[w_{0,j}^{\tau,*}\big]^y &= \big(A_j^{\tau,y} w_y + A_{0,j}^\tau w_{xy}\big)\big(g_{0,j}^\tau\big).
\end{aligned} \tag{3.10}$$

For any $i,j \geq 1$, a direct calculation yields

$$\begin{aligned}
w_{i,j}^{\tau,*} &= \sum_{r=1}^{i}\sum_{l=1}^{j}\Big[w_{r,l}^{\tau,*}\Big] + \sum_{l=1}^{j}\Big[w_{0,l}^{\tau,*}\Big]^y + \sum_{r=1}^{i}\big[w_{r,0}^{\tau,*}\big]^x + w_{0,0}^{\tau,*}\\
&= \sum_{r=1}^{i}\sum_{l=1}^{j} A_{r,l}^\tau w_{xy}\big(g_{r,l}^\tau\big) + \sum_{l=1}^{j} A_l^{\tau,y} w_y\big(g_{0,l}^\tau\big) + \sum_{r=1}^{i} A_r^{\tau,x} w_x\big(g_{r,0}^\tau\big) + w\big(g_{0,0}^\tau\big).
\end{aligned}$$

By using the inverse inequality, we easily get

$$\|\mathcal{F}w\|_{0,\tau} \lesssim \|w\|_{0,\tau}, \qquad \forall w \in U_h.$$

Similarly, we can define a transformation $\bar{\mathcal{F}}$ from the vector trial space $\mathbf{U}_h$ to the vector test space $\mathbf{V}_h$ with the points $g_{i,j}^\tau$ replaced by $\bar{g}_{i,j}^\tau$, and there also holds

$$\|\bar{\mathcal{F}}\mathbf{w}\|_{0,\tau} \lesssim \|\mathbf{w}\|_{0,\tau}, \qquad \forall \mathbf{w} \in \mathbf{U}_h.$$

Denote

$$v(x,\cdot)|_a^b = v(b,\cdot) - v(a,\cdot), \quad (w,v)_\tau = \int_\tau (wv)(x,y)\,\mathrm{d}x\,\mathrm{d}y, \quad (w,v) = \sum_{\tau\in\mathcal{T}_h} (w,v)_\tau.$$

In each element $\tau = \tau_m^x \times \tau_n^y$, we define

$$\langle v, w\rangle_{\partial\tau_y} = \int_{\tau_n^y} \Big((vw)(x_{m+\frac{1}{2}}, y) - (vw)(x_{m-\frac{1}{2}}, y)\Big)\,\mathrm{d}y. \tag{3.11}$$

Since the transformations $\mathcal{F}$ and $\bar{\mathcal{F}}$ share the same properties, we next only consider the property of $\mathcal{F}$ and the same argument can be applied to $\bar{\mathcal{F}}$.

Lemma 3.1. *For all $v \in \mathcal{H}_h, w \in U_h$, let $\mathcal{F}$ be defined by (3.9). Then*

$$\langle v, \mathcal{F}w\rangle_{\partial\tau_y} = \langle v, w\rangle_{\partial\tau_y} + R_\tau^y(w_y\partial_y^{-1}v)|_{g_0^{\tau,x}}^{g_{k+1}^{\tau,x}} + E_2(v,w) - E_1(v,w), \tag{3.12}$$

$$(v, \mathcal{F}w)_\tau = (v,w)_\tau + \mathcal{E}_\tau(v,w), \tag{3.13}$$

where for all $\tau = \tau_m^x \times \tau_n^y$, R_τ^x, R_τ^y are given in (3.7) and (3.8), and

$$E_1(v,w) = A_0^{\tau,x}\left(\int_{g_0^{\tau,y}}^{g_{k+1}^{\tau,y}} (w_x v)(g_0^{\tau,x}, y)\,\mathrm{d}y + R_\tau^y(w_{xy}\partial_y^{-1}v)(g_0^{\tau,x})\right), \tag{3.14}$$

$$E_2(v,w) = -A_{k+1}^{\tau,x}\left(\int_{g_0^{\tau,y}}^{g_{k+1}^{\tau,y}} (w_x v)\big(g_{k+1}^{\tau,x}, y\big)\,\mathrm{d}y + R_\tau^y\big(w_{xy}\partial_y^{-1}v\big)\big(g_{k+1}^{\tau,x}\big)\right), \tag{3.15}$$

$$\mathcal{E}_\tau(v,w) = \int_{\tau_n^y} R_\tau^x\big(w_x\partial_x^{-1}v\big)\,\mathrm{d}y + \int_{\tau_m^x} R_\tau^y\big(w_y\partial_y^{-1}v\big)\,\mathrm{d}x + R_\tau^x R_\tau^y\big(w_{xy}\partial_x^{-1}\partial_y^{-1}v\big). \tag{3.16}$$

The proof of Lemma 3.1 is given in Appendix A.

Lemma 3.2. *For all $\tau = [x_{m-\frac12}, x_{m+\frac12}] \times [y_{n-\frac12}, y_{n+\frac12}]$, there hold*

$$(\hat v_x, \mathcal{F}w)_\tau = -(v, w_x)_\tau + \langle \hat v, w\rangle_{\partial\tau_y} + \mathcal{E}_\tau^1(v,w) + \mathcal{E}_\tau(v_x, w), \tag{3.17}$$

$$(\hat v_y, \mathcal{F}w)_\tau = -(v, w_y)_\tau + \langle \hat v, w\rangle_{\partial\tau_x} + \mathcal{E}_\tau^2(v,w) + \mathcal{E}_\tau(v_y, w), \tag{3.18}$$

where $\mathcal{E}_\tau$ is define in (3.16) and

$$\mathcal{E}_\tau^1(v,w) = R_\tau^y(w_y\partial_y^{-1}(\hat v - v))|_{x_{m-\frac12}}^{x_{m+\frac12}}, \quad \mathcal{E}_\tau^2(v,w) = R_\tau^x(w_x\partial_x^{-1}(\hat v - v))|_{y_{n-\frac12}}^{y_{n+\frac12}}. \tag{3.19}$$

Consequently, for all $\mathbf{v} = (v_1, v_2) \in \mathcal{H}_h \times \mathcal{H}_h$,

$$(\nabla\cdot\hat{\mathbf{v}}, \mathcal{F}w)_\tau = -(\mathbf{v}, \nabla w)_\tau + \int_{\partial\tau} w\hat{\mathbf{v}}\cdot\mathbf{n}\,\mathrm{d}s + \mathcal{E}_\tau^1(v_1, w) + \mathcal{E}_\tau^2(v_2, w) + \mathcal{E}_\tau(\nabla\cdot\mathbf{v}, w). \tag{3.20}$$

Proof. We only prove (3.17) as the same argument can be applied to (3.18).

Let $w^* = \mathcal{F}w$. By (3.12), we have

$$\begin{aligned}(\hat v_x, w^*)_\tau &= (v_x, w^*)_\tau + \langle \hat v - v, w^*\rangle_{\partial\tau_y}\\ &= (v_x, w^*)_\tau + \langle \hat v - v, w\rangle_{\partial\tau_y} + R_\tau^y\big(w_y\partial_y^{-1}(\hat v - v)\big)|_{g_0^{\tau,x}}^{g_{k+1}^{\tau,x}} + (E_2 - E_1)(\hat v - v, w),\end{aligned}$$

where E_1, E_2 are defined in (3.14) and (3.15). Using (3.13) and the integration by parts, we get

$$(\hat v_x, w^*)_\tau = -(v, w_x)_\tau + \langle \hat v, w\rangle_{\partial\tau_y} + \mathcal{E}_\tau(v_x, w) + \mathcal{E}_\tau^1(v,w) + (E_2 - E_1)(\hat v - v, w).$$

Recalling the definition of E_1, E_2 in (3.14) and (3.15), we have $(E_2 - E_1)(\hat v - v, w) = 0$ for GSV since $A_0^{\tau,x} = A_{k+1}^{\tau,x} = 0$. As for RSV, if the flux is chosen as $\hat v = v^-$, then $(\hat v - v)(g_{k+1}^{\tau,x}, y) = 0$ and thus $E_2(\hat v - v, w) = 0$. On the other hand, the abscissas are taken as right Radau points in case $\hat v = v^-$, which yields $A_0^{\tau,x} = 0, E_1(\hat v - v, w) = 0$. Similarly, if $\hat v = v^+$, we have $(\hat v - v)(g_0^{\tau,x}, y) = 0$ and the abscissas are left Radau points and thus $A_{k+1}^{\tau,x} = 0$. Consequently, for both flux choices,

$$(E_2 - E_1)(\hat v - v, w) = 0.$$

Then the desire result (3.17) follows. Combining (3.17) and (3.18), we obtain (3.20) directly. □

We end this subsection the comparison of the SV method and the LDG method. Define

$$\bar{a}_\tau^1(v,\mathbf{p};w)=(v_t,w)_\tau+(\beta\mathbf{p},\nabla w)_\tau-\int_{\partial\tau}w\widehat{\beta\mathbf{p}}\cdot\mathbf{n}\,\mathrm{d}s,$$
$$\bar{a}_\tau^2(v,\mathbf{p};\mathbf{w})=(\mathbf{p},\mathbf{w})_\tau+(B,\nabla\cdot\mathbf{w})_\tau-\int_{\partial\tau}\widehat{B}\mathbf{w}\cdot\mathbf{n}\,\mathrm{d}s.$$

Here $B:=B(v,x,y)$ is defined in (3.2). For simplicity, we adopt the notation $B(v)$ instead of $B(v,x,y)$ if no confusion arises. Then the LDG method for (3.2) is: find $u_h\in U_h,\mathbf{q}_h\in\mathbf{U_h}$ such that

$$\bar{a}_\tau^1(u_h,\mathbf{q}_h;w)=0,\quad \bar{a}_\tau^2(u_h,\mathbf{q}_h;\mathbf{w})=0,\quad \forall\tau\in\mathcal{T}_h,w\in V_h,\mathbf{w}\in\mathbf{V}_h.$$

In light of (3.5), we get

$$a_\tau^1(v,\mathbf{p};w^*)=(v_t,w^*)_\tau-\left(\nabla\cdot\left(\widehat{\beta p}\right),w^*\right)_\tau,\quad a_\tau^2(v,\mathbf{p};\mathbf{w}^*)=(\mathbf{p},\mathbf{w}^*)_\tau-\left(\nabla\widehat{B(v)},\mathbf{w}^*\right)_\tau.$$

By using (3.13) and (3.20), we have for all $\mathbf{p}=(p_1,p_2)$

$$a_\tau^1(v,\mathbf{p};\mathcal{F}w)=\bar{a}_\tau^1(v,\mathbf{p};w)-\mathcal{E}_\tau^1(\beta p_1,w)-\mathcal{E}_\tau^2(\beta p_2,w)+\mathcal{E}_\tau(v_t-\nabla\cdot(\beta\mathbf{p}),w).\tag{3.21}$$

Similarly, there holds for all $\mathbf{w}=(w_1,w_2)$

$$a_\tau^2\big(v,\mathbf{p};\bar{\mathcal{F}}\mathbf{w}\big)=\bar{a}_\tau^2(v,\mathbf{p};\mathbf{w})-\bar{\mathcal{E}}_\tau^1(B,w_1)-\bar{\mathcal{E}}_\tau^2(B,w_2)+\bar{\mathcal{E}}_\tau(\mathbf{p}-\nabla B(v),\mathbf{w}),\tag{3.22}$$

where $\bar{\mathcal{E}}_\tau,\bar{\mathcal{E}}_\tau^i,i=1,2$ are given in (3.16) and (3.19) with $\bar{R}_\tau^x,\bar{R}_\tau^y$ substituting of R_τ^x,R_τ^y, respectively. As indicated by (3.21) and (3.22), the proposed GSV and RSV schemes are equivalent to the counterpart LDG schemes up to some Gauss or Radau numerical quadrature errors. Especially, for constant coefficients problems, *i.e.*, $\alpha=const$ in (3.2), using the left or right Radau quadrature is exact for polynomials of degree not more than $2k$, we get for all $v\in U_h,\mathbf{p}=(p_1,p_2)\in\mathbf{U}_h,\mathbf{w}=(w_1,w_2)\in\mathbf{U}_h$

$$\mathcal{E}_\tau^1(\beta p_1,w)=\mathcal{E}_\tau^2(\beta p_2,w)=\mathcal{E}_\tau(v_t-\nabla\cdot\beta\mathbf{p},w)=0,$$
$$\bar{\mathcal{E}}_\tau^1(B,w_1)=\bar{\mathcal{E}}_\tau^2(B,w_2)=\bar{\mathcal{E}}_\tau(\mathbf{p}-\nabla B(v),\mathbf{w})=0,$$

which indicates that the RSV method is exactly the same as the LDG method when applied to constant diffusion problems.

3.3. Stability

To study the stability of the SV schemes, we first estimate the numerical quadrature errors appeared in (3.21) and (3.22).

Lemma 3.3. *For any $v,p\in U_h$, let $\mathcal{E}_\tau^i(B(v),p),i=1,2$ be defined in (3.19) with $(\hat{B},p)=(B^+,p^-)$ or $(\hat{B},p)=(B^-,p^+)$, and $\mathcal{E}_\tau(B(v),p)$ be defined in (3.16). Denote*

$$I_\tau=\mathcal{E}_\tau^1(B(v),p)+\bar{\mathcal{E}}_\tau^1(\beta p,v)+\mathcal{E}_\tau((\beta p)_x,v)+\bar{\mathcal{E}}_\tau(B(v)_x,p),$$
$$J_\tau=\mathcal{E}_\tau^2(B(v),p)+\bar{\mathcal{E}}_\tau^2(\beta p,v)+\mathcal{E}_\tau\Big((\beta p)_y,v\Big)+\bar{\mathcal{E}}_\tau(B(v)_y,p).$$

Then for both GSV and RSV,

$$\left|\sum_{\tau\in\mathcal{T}_h}I_\tau\right|\lesssim\|v\|_0\|p\|_0,\quad \left|\sum_{\tau\in\mathcal{T}_h}J_\tau\right|\lesssim\|v\|_0\|p\|_0.\tag{3.23}$$

The proof of Lemma 3.3 is given in the Appendix A.

Similar to the 1D case, we define

$$(v, \mathcal{F}v) = \sum_{\tau\in\mathcal{T}_h} (v, \mathcal{F}v)_\tau.$$

In light of (3.13), we have

$$(v, \mathcal{F}v) = (v, v) + \sum_{\tau\in\mathcal{T}_h} \mathcal{E}_\tau(v, v).$$

Recalling the definition of $\mathcal{E}_\tau$ in (3.16), we have $\mathcal{E}(v, w) = 0, v, w \in U_h$ for the RSV scheme. As for the GSV scheme, by using the error of Gauss numerical quadrature in $[-1, 1]$, there holds for some $\xi \in (-1, 1)$

$$R(f) = c_k \partial_s^{2k} f(\xi), \text{ with } c_k = \frac{2^{2k+1}(k!)^4}{(2k+1)[(2k)!]^3}. \tag{3.24}$$

By a scaling from $[-1, 1]$ to τ_m^x and τ_n^y and using the fact that $\partial_x^k v, \partial_y^k v$ are both constants for all $v \in U_h$, we get for any $\tau = \tau_m^x \times \tau_n^y$ that

$$\mathcal{E}_\tau(v, v) = c\Big((h_\tau^y)^{2k}\big\|\partial_y^k v\big\|_{0,\tau}^2 + c(h_\tau^x h_\tau^y)^{2k}\big\|\partial_y^k\partial_x^k v\big\|_{0,\tau}^2 + (h_\tau^x)^{2k}\big\|\partial_x^k v\big\|_{0,\tau}^2\Big).$$

Here c is a positive constant only dependent on k. Then a direct calculation from the inverse inequality yields

$$0 \le \mathcal{E}_\tau(v, v) \lesssim \|v\|_{0,\tau}, \qquad \forall v \in U_h.$$

Consequently, for both GSV and RSV,

$$(v, v) \le (v, \mathcal{F}v) \lesssim (v, v), \qquad \forall v \in U_h. \tag{3.25}$$

Similarly, we can prove

$$(\mathbf{p}, \mathbf{p}) \le (\mathbf{p}, \bar{\mathcal{F}}\mathbf{p}) \lesssim (\mathbf{p}, \mathbf{p}), \qquad \forall \mathbf{p} \in \mathbf{U}_h. \tag{3.26}$$

Now we are ready to present the stability result for both RSV and GSV schemes.

Theorem 3.4. *Let $u_h, \mathbf{q}_h$ be the solution of the SV scheme (3.6). Then for both GSV and RSV schemes,*

$$\|u_h(\cdot, t)\|_0^2 + \int_0^t \|\mathbf{q}_h\|_0^2 \,\mathrm{d}t \lesssim \|u_0\|_0^2, \qquad \forall t \in (0, t_0]. \tag{3.27}$$

Proof. For all $v \in U_h, \mathbf{p} = (p_1, p_2) \in U_h \times U_h$, denoting $(v^*, \mathbf{p}^*) = (\mathcal{F}v, \bar{\mathcal{F}}\mathbf{p})$. Recalling the definition of a_τ^1 and using (3.17)–(3.20), we get

$$\begin{aligned} a_\tau^1(v, \mathbf{p}; v^*) &= (v_t, v^*)_\tau - \Big(\nabla\cdot\Big(\widehat{\beta\mathbf{p}}\Big), v^*\Big)_\tau \\ &= (v_t, v^*)_\tau + (\beta\mathbf{p}, \nabla v)_\tau - \int_{\partial\tau} v\widehat{\beta\mathbf{p}}\cdot\mathbf{n}\,\mathrm{d}s - \mathcal{E}_\tau^1(\beta p_1, v) - \mathcal{E}_\tau^2(\beta p_2, v) - \mathcal{E}_\tau(\nabla\cdot(\beta\mathbf{p}), v). \end{aligned}$$

Similarly, there holds

$$a_\tau^2(v, \mathbf{p}; \mathbf{p}^*) = (\mathbf{p}, \mathbf{p}^*)_\tau + (B, \nabla\cdot\mathbf{p})_\tau - \int_{\partial\tau} \widehat{B\mathbf{p}}\cdot\mathbf{n}\,\mathrm{d}s - \bar{\mathcal{E}}_\tau^1(B, p_1) - \bar{\mathcal{E}}_\tau^2(B, p_2) - \bar{\mathcal{E}}_\tau(\nabla B, \mathbf{p}).$$

Summing up all τ and using the flux choice (2.5) or (2.6) and the periodic boundary condition, we have

$$a(v, \mathbf{p}; v^*, \mathbf{p}^*) = (v_t, v^*) + (\mathbf{p}, \mathbf{p}^*) + (\beta\mathbf{p}, \nabla v) + (B, \nabla\cdot\mathbf{p}) - \sum_{\tau\in\mathcal{T}_h}\left(\int_{\partial\tau} v\widehat{\beta\mathbf{p}}\cdot\mathbf{n}\,\mathrm{d}s + \int_{\partial\tau}\widehat{B\mathbf{p}}\cdot\mathbf{n}\,\mathrm{d}s\right) + H,$$

where

$$H=\sum_{\tau\in\mathcal{T}_h}\left(\bar{\mathcal{E}}_\tau^1(B,p_1)+\bar{\mathcal{E}}_\tau^2(B,p_2)+\mathcal{E}_\tau^1(\beta p_1,v)+\mathcal{E}_\tau^2(\beta p_2,v)\right)+\bar{\mathcal{E}}_\tau(\nabla B,\mathbf{p})+\mathcal{E}_\tau(\nabla\cdot(\beta\mathbf{p}),v).$$

Due to the special choice of the numerical fluxes and the periodic boundary condition, we have

$$(\beta\mathbf{p},\nabla v)+(B,\nabla\cdot\mathbf{p})-\sum_{\tau\in\mathcal{T}_h}\left(\int_{\partial\tau}v\widehat{\beta\mathbf{p}}\cdot\mathbf{n}\,\mathrm{d}s+\int_{\partial\tau}\widehat{B\mathbf{p}}\cdot\mathbf{n}\,\mathrm{d}s\right)=0.$$

On the other hand, we use the conclusion (3.23) in Lemma 3.3 and then get

$$|H|\lesssim\|v\|_0\|\mathbf{p}\|_0,$$

and thus

$$(v_t,v^*)+(\mathbf{p},\mathbf{p}^*)\le a(v,\mathbf{p};v^*,\mathbf{p}^*)+\left|\sum_{\tau\in\mathcal{T}_h}H_\tau\right|\lesssim a(v,\mathbf{p};v^*,\mathbf{p}^*)+\|v\|_0\|\mathbf{p}\|_0. \tag{3.28}$$

Choosing $(v,\mathbf{p})=(u_h,\mathbf{q}_h)$ in the above inequality and using (3.6), we get

$$\frac{\mathrm{d}}{\mathrm{d}t}(u_h,\mathcal{F}u_h)+\left(\mathbf{q}_h,\bar{\mathcal{F}}\mathbf{q}_h\right)\lesssim\|u_h\|_0\|\mathbf{q}_h\|_0.$$

Then (3.27) follows from the Gronwall inequality and the L^2 equivalences (3.25) and (3.26). □

3.4. Error estimates

This subsection is dedicated to error estimates of the SV method for 2D equations in rectangular meshes. To this end, we first introduce a suitable projection similar to the one-dimensional case, which is defined by

$$(\Pi_h u,\Pi_h\mathbf{q})=\left((I_h^x\otimes I_h^y)u,(I_h^x\otimes I_h^y)\mathbf{q}\right), \tag{3.29}$$

where the superscripts indicate the application of the one-dimensional operator I_h^x with respect to the variable x with $(I_h^x u,I_h^x\mathbf{q})$ given in (2.31). The operator $I_h^y v$ follows the same definition along the y direction. The definition of $(I_h^x u,I_h^x\mathbf{q})$ ensures that for both alternating fluxes,

$$\widehat{I_h^x u}(\bar{g}_i^{\tau,x},y)=u(\bar{g}_i^{\tau,x},y)\quad\widehat{I_h^x\mathbf{q}}(g_i^{\tau,x},y)=\mathbf{q}(g_i^{\tau,x},y),\qquad i\in\mathbb{Z}_{k+1}. \tag{3.30}$$

For any function $v\in\mathcal{H}_h$, we define

$$\mathcal{Q}^x v=v-I_h^x v,\quad\mathcal{Q}^y v=v-I_h^y v.$$

We see that $\mathcal{Q}^x v$ and $\mathcal{Q}^y v$ is continuous about y and x, respectively. By the approximation theory, we have for all $r\le l\le k+1$ that

$$\|\partial_x^r\mathcal{Q}^x v\|_{0,\tau}\lesssim h^{l-r}\|\partial_x^l v\|_{0,\tau},\quad\partial_y^r\mathcal{Q}^x v=\mathcal{Q}^x\partial_y^r v, \tag{3.31}$$

$$\|\partial_y^r\mathcal{Q}^y v\|_{0,\tau}\lesssim h^{l-r}\|\partial_y^l v\|_{0,\tau},\quad\partial_x^r\mathcal{Q}^y v=\mathcal{Q}^y\partial_x^r v. \tag{3.32}$$

Furthermore, a direct calculation yields that

$$v-\Pi_h v=(I-I_h^x)v+(I-I_h^y)v-(I-I_h^x)(I-I_h^y)v=\mathcal{Q}^x v+\mathcal{Q}^y v-\mathcal{Q}^x\mathcal{Q}^y v. \tag{3.33}$$

Now we are ready to present the error estimates for the SV method.

Theorem 3.5. *Let $u \in H^{k+3}$ be the solution of (3.1), and $\Pi_h u$ be the Lagrange interpolation function of u defined by (3.29). Assume that $u_h, \mathbf{q}_h$ are the solutions of the SV method (3.6) with the initial value chosen as $u_h(x,y,0) = \Pi_h u_0(x,y)$, Then for both GSV and RSV,*

$$\|u - u_h\|_0 + \int_0^t \|\mathbf{q} - \mathbf{q}_h\|_0 \,\mathrm{d}t \lesssim h^{k+1}\|u\|_{k+3}. \tag{3.34}$$

Proof. First, denoting

$$\xi_u = u_h - \Pi_h u, \quad \xi_{\mathbf{q}} = \mathbf{q}_h - \Pi_h \mathbf{q}, \text{ with } \mathbf{q} = (q_1, q_2).$$

Since the exact solution $(u, \mathbf{q})$ also satisfy (3.6), we take $(v, \mathbf{p}) = (\xi_u, \xi_{\mathbf{q}})$ in (3.28) and use the orthogonality to get

$$\begin{aligned}(\partial_t \xi_u, \xi_u^*) + (\xi_{\mathbf{q}}, \xi_{\mathbf{q}}^*) &\lesssim a(\xi_u, \xi_q; \xi_u^*, \xi_{\mathbf{q}}^*) + \|\xi_u\|_0 \|\xi_{\mathbf{q}}\|_0 \\ &= a(u - \Pi_h u, \mathbf{q} - \Pi_h \mathbf{q}; \xi_u^*, \xi_{\mathbf{q}}^*) + \|\xi_u\|_0 \|\xi_{\mathbf{q}}\|_0.\end{aligned} \tag{3.35}$$

On the other hand, recalling the definitions of a_τ^1, a_τ^2 in (3.5), we get for all $w^* \in V_h, \mathbf{w}^* \in \mathbf{V}_h$

$$a_\tau^1(u - \Pi_h u, \mathbf{q} - \Pi_h \mathbf{q}; w^*) = (u_t - \Pi_h u_t, w^*)_\tau - \sum_{i,j=0}^{k} w_{i,j}^{\tau,*} \int_{\partial V_{i,j}^\tau} \beta\Big(\widehat{\mathbf{q} - \Pi_h \mathbf{q}}\Big) \cdot \mathbf{n} \,\mathrm{d}s,$$

$$a_\tau^2(u - \Pi_h u, \mathbf{q} - \Pi_h \mathbf{q}; \mathbf{w}^*) = (\mathbf{q} - \Pi_h \mathbf{q}, \mathbf{w}^*)_\tau - \sum_{i,j=0}^{k} \mathbf{w}_{i,j}^{\tau,*} \int_{\partial \bar{V}_{i,j}^\tau} B\Big(\widehat{u - \Pi_h u}\Big)\mathbf{n_0} \cdot \mathbf{n} \,\mathrm{d}s.$$

Using the error decomposition (3.33), we get

$$\sum_{i,j=0}^{k} w_{i,j}^{\tau,*} \int_{\partial V_{i,j}^\tau} \beta\Big(\widehat{\mathbf{q} - \Pi_h \mathbf{q}}\Big) \cdot \mathbf{n} \,\mathrm{d}s = \sum_{i,j=0}^{k} w_{i,j}^{\tau,*} \int_{\partial V_{i,j}^\tau} \beta\Big(\widehat{\mathcal{Q}^x \mathbf{q}} + \widehat{\mathcal{Q}^y \mathbf{q}} - \widehat{\mathcal{Q}^y \mathcal{Q}^x \mathbf{q}}\Big) \cdot \mathbf{n} \,\mathrm{d}s.$$

In light of the properties of $\mathcal{Q}^x$ in (3.31) and the fact $\mathcal{Q}^x$ is continuous about y, we have

$$\begin{aligned}\sum_{i,j=0}^{k} w_{i,j}^{\tau,*} \int_{\partial V_{i,j}^\tau} \beta\widehat{\mathcal{Q}^x \mathbf{q}} \cdot \mathbf{n} \,\mathrm{d}s &= \sum_{i,j=0}^{k} w_{i,j}^{\tau,*} \int_{g_i^{\tau,x}}^{g_{i+1}^{\tau,x}} \big(\beta \mathcal{Q}^x q_2(x, g_{j+1}^{\tau,y}) - \beta \mathcal{Q}^x q_2(x, g_j^{\tau,y})\big) \,\mathrm{d}x \\ &= \Big((\beta \mathcal{Q}^x q_2)_y, w^*\Big)_\tau.\end{aligned}$$

Similarly, there holds

$$\sum_{i,j=0}^{k} w_{i,j}^{\tau,*} \int_{\partial V_{i,j}^\tau} \beta\widehat{\mathcal{Q}^y \mathbf{q}} \cdot \mathbf{n} \,\mathrm{d}s = ((\beta \mathcal{Q}^y q_1)_x, w^*)_\tau, \quad \sum_{i,j=0}^{k} w_{i,j}^{\tau,*} \int_{\partial V_{i,j}^\tau} \beta\widehat{\mathcal{Q}^y \mathcal{Q}^x \mathbf{q}} \cdot \mathbf{n} \,\mathrm{d}s = 0.$$

Substituting the last three equations into a_τ^1 yields

$$\begin{aligned}a_\tau^1(u - \Pi_h u, \mathbf{q} - \Pi_h \mathbf{q}; w^*) &= (u_t - \Pi_h u_t, w^*)_\tau - \Big((\beta \mathcal{Q}^x q_2)_y, w^*\Big)_\tau - ((\beta \mathcal{Q}^y q_1)_x, w^*)_\tau. \\ &\lesssim h^{k+1}(\|u_t\|_{k+1,\tau} + \|\mathbf{q}\|_{k+2,\tau})\|w^*\|_{0,\tau}.\end{aligned}$$

Following the same argument, we have for all $\mathbf{w}^* = (w_1^*, w_2^*)$ that

$$\big|a_\tau^2(u - \Pi_h u, \mathbf{q} - \Pi_h \mathbf{q}; \mathbf{w}^*)\big| \lesssim h^{k+1}(\|\mathbf{q}\|_{k+1,\tau} + \|u\|_{k+2,\tau})\|\mathbf{w}^*\|_{0,\tau}.$$

Therefore,

$$|a(u-\Pi_h u,\mathbf{q}-\Pi_h\mathbf{q};w^*,\mathbf{w}^*)|\lesssim h^{k+1}(\|u_t\|_{k+1}+\|u\|_{k+3})(\|w^*\|_0+\|\mathbf{w}^*\|_0),$$

which yields, together with (3.35) and (3.26) that

$$(\partial_t\xi_u,\xi_u^*)+\|\xi_{\mathbf{q}}\|_0^2\lesssim h^{2(k+1)}(\|u_t\|_{k+1}+\|u\|_{k+2})^2+\|\xi_u\|_0^2.$$

Integrating the above inequality from 0 to t, using the Gronwall inequality and the L^2 norm equivalence (3.25), we have

$$\|\xi_u\|_0^2+\int_0^t\|\xi_{\mathbf{q}}\|_0^2\lesssim h^{2(k+1)}(\|u_t\|_{k+1}+\|u\|_{k+2})^2.$$

Then the desired result follows from standard approximation property of the interpolation function and the inequality $\|u_t\|_{k+1}\lesssim\|u\|_{k+3}$ due to $u_t=-\nabla\cdot(\alpha\nabla u)$. □

4. Numerical results

In this section, we present some numerical examples to test the efficiency and accuracy of the SV method for solving the one and two dimensional diffusion equations. In our experiments, we implement the GSV and RSV numerical schemes with $k=1,2,3$ and the fluxes choice (2.5). We shall compute the L^2 error for both the exact solution u and the auxiliary variable q or $\mathbf{q}$, which is denoted by $\|e_u\|_0$ and $\|e_q\|_0$ or $(\|e_{\mathbf{q}}\|_0)$. To diminish the time discretization error, we use the fourth order Runge–Kutta method with time step $\triangle t=0.005h^2$.

Example 1. We solve the following problem

$$u_t=(\alpha u_x)_x+g(x,t),\quad u(x,0)=\sin(x),\ u(0,t)=u(2\pi,t),\ (x,t)\in[0,2\pi]\times(0,0.1].$$

The source function $g(x,t)$ is chosen such that the exact solution is

$$u(x,t)=e^{-t}\sin(x).$$

We consider two cases: the linear equation with variable coefficient $\alpha=1+\sin^2 x$ and the nonlinear equation with $\alpha=\sin^2 u$. We use non-uniform meshes of N elements in our experiment, which are obtained by randomly and independently perturbing each node of a uniform mesh by up to some percentage. That is,

$$x_j=\frac{2\pi j}{N}+\frac{0.01}{N}\sin\left(\frac{j\pi}{N}\right)\text{randn}(),\qquad 0\le j\le N,$$

where *randn*() returns a uniformly distributed random number in $(0,1)$.

Table 1 shows the L^2 errors and the corresponding convergence rate calculated from the GSV and RSV schemes for $\alpha=1+\sin^2 x$ and $k=1,2,3$. We observe that, for both GSV and RSV, the convergence rates for the errors $\|e_u\|_0$ and $\|e_q\|_0$ can achieve $\mathcal{O}(h^{k+1})$. These numerical results verify the theoretical findings (2.33) in Theorem 2.5.

Listed in Table 2 are L^2 errors and the rates of convergence of two SV schemes for the nonlinear problem with $\alpha=\sin^2 u$. Similar to the linear equation, we again observe an optimal convergence order of $\mathcal{O}(h^{k+1})$ for both the errors $\|e_u\|_0$ and $\|e_q\|_0$, which indicates that the error estimate (2.33) also holds true for the nonlinear problems. Furthermore, as demonstrated from Tables 1 and 2, the GSV method seems to have larger L^2 errors than the counterpart RSV method on the same meshes.

Example 2. We solve the following two dimensional diffusion equation

$$\begin{aligned}&u_t=\nabla\cdot(\alpha\nabla u)+g(x,y,t),&&(x,y,t)\in[0,2\pi]\times[0,2\pi]\times(0,0.1],\\&u(x,y,0)=\sin(x+y),&&(x,y)\in[0,2\pi]\times[0,2\pi]\end{aligned}$$

TABLE 1. L^2 errors and corresponding convergence rates of the SV method for Example 1 with $\alpha = 1 + \sin^2 x$ and $k = 1, 2, 3$.

		GSV				RSV			
k	N	$\|e_u\|_0$	Rate	$\|e_q\|_0$	Rate	$\|e_u\|_0$	Rate	$\|e_q\|_0$	Rate
	8	8.95e−02	–	9.48e−02	–	4.85e−02	–	6.70e−02	–
	16	2.22e−02	2.01	2.68e−02	1.82	1.20e−02	2.02	1.75e−02	1.94
1	32	5.54e−03	2.00	6.86e−03	1.97	2.98e−03	2.00	4.39e−03	1.99
	64	1.38e−03	2.00	1.73e−03	1.99	7.44e−04	2.00	1.10e−03	2.00
	128	3.46e−04	2.00	4.312-04	2.00	1.86e−04	2.00	2.75e−04	2.00
	8	6.81e−03	–	1.84e−02	–	3.82e−03	–	1.27e−02	–
	16	7.73e−04	3.14	2.33e−03	2.99	4.74e−04	3.01	1.48e−03	3.10
3	32	9.39e−05	3.04	2.94e−04	2.98	5.93e−05	3.00	1.87e−04	2.98
	64	1.17e−05	3.00	3.71e−05	2.99	7.41e−06	3.00	2.35e−05	3.00
	128	1.49e−06	2.98	4.68e−06	2.99	9.26e−07	3.00	2.93e−06	3.00
	8	3.81e−04	–	1.82e−03	–	1.83e−04	–	1.13e−03	–
	16	2.21e−05	4.11	1.69e−04	3.43	1.16e−05	3.98	1.04e−04	3.43
3	32	1.30e−06	4.09	1.08e−05	3.97	7.30e−07	3.99	6.66e−06	3.97
	64	9.21e−08	3.82	6.87e−07	3.97	4.84e−08	3.92	4.19e−07	3.99
	128	5.14e−09	4.16	4.25e−08	4.01	2.80e−09	4.11	2.61e−08	4.01

TABLE 2. L^2 errors and corresponding convergence rates of the SV method for Example 1 with $\alpha = \sin^2 u$ and $k = 1, 2, 3$.

		GSV				RSV			
k	N	$\|e_u\|_0$	Rate	$\|e_q\|_0$	Rate	$\|e_u\|_0$	Rate	$\|e_q\|_0$	Rate
	4	3.77e−01	–	4.20e−02	–	2.15e−01	–	1.76e−01	–
	8	9.26e−02	2.02	7.06e−02	−7.49	5.27e−02	2.03	8.26e−02	1.09
1	16	2.28e−02	2.02	3.28e−02	1.10	1.24e−02	2.08	2.52e−02	1.72
	32	5.53e−03	2.04	9.18e−03	1.84	3.02e−03	2.04	6.43e−03	1.97
	64	1.36e−03	2.02	2.32e−03	1.98	7.47e−04	2.01	1.61e−03	2.00
	4	4.53e−02	–	2.21e−02	–	3.06e−02	–	2.40e−02	–
	8	7.54e−03	2.59	1.31e−02	0.76	4.10e−03	2.90	1.16e−02	1.05
2	16	9.82e−04	2.94	2.00e−03	2.71	5.09e−04	3.01	1.63e−03	2.83
	32	1.20e−04	3.03	2.93e−04	2.77	6.23e−05	3.03	2.18e−04	2.90
	64	1.40e−05	3.10	4.08e−05	2.85	7.63e−06	3.03	2.83e−05	2.95
	4	8.70e−03	–	1.92e−02	–	3.35e−03	–	2.23e−02	–
	8	6.26e−04	3.80	1.47e−03	3.71	2.02e−04	4.05	1.44e−03	3.95
3	16	3.35e−05	4.22	1.33e−04	3.47	1.20e−05	4.08	8.99e−05	4.00
	32	1.61e−06	4.38	9.16e−06	3.86	7.26e−07	4.04	5.74e−06	3.97
	64	8.15e−08	4.30	5.76e−07	3.99	4.49e−08	4.02	3.60e−07	4.00

with the periodic boundary condition

$$u(0, y, t) = u(2\pi, y, t) \quad \text{and} \quad u(x, 0, t) = u(x, 2\pi, t).$$

We consider the constant coefficient case $\alpha = 1$ and the variable coefficient case $\alpha = 1 + \sin^2 x$. The source function $g(x, y, t)$ is chosen such that the exact solution is

$$u(x, y, t) = e^{-2t} \sin(x + y).$$

TABLE 3. L^2 errors and corresponding convergence rates of the SV method for Example 2 with $\alpha = 1$ and $k = 1, 2, 3$.

		GSV				RSV			
k	N	$\|e_u\|_0$	Rate	$\|e_q\|_0$	Rate	$\|e_u\|_0$	Rate	$\|e_q\|_0$	Rate
	8	2.97e−01	–	5.58e−01	–	2.38e−01	–	4.89e−01	–
	16	8.05e−02	1.89	1.35e−01	2.04	6.02e−02	1.98	1.21e−01	2.01
1	32	2.01e−02	2.00	3.32e−02	2.03	1.51e−02	2.00	3.02e−02	2.01
	64	5.01e−03	2.00	8.25e−03	2.01	3.77e−03	2.00	7.55e−03	2.00
	128	1.25e−03	2.00	2.06e−03	2.00	9.43e−04	2.00	1.88e−03	2.00
	8	1.67e−02	–	3.82e−02	–	1.22e−02	–	2.40e−02	–
	16	2.41e−03	2.80	4.82e−03	2.99	1.53e−03	3.00	3.05e−03	2.98
2	32	3.00e−04	3.00	6.06e−04	2.99	1.90e−04	3.00	3.80e−04	3.00
	64	3.74e−05	3.01	7.48e−05	3.02	2.38e−05	3.00	4.75e−05	3.00
	128	4.67e−06	3.00	9.35e−06	3.00	2.97e−06	3.00	5.94e−06	3.00
	8	9.31e−04	–	1.87e−03	–	5.84e−04	–	1.16e−03	–
	16	5.95e−05	3.97	1.19e−04	3.98	3.67e−05	3.99	7.33e−05	3.99
3	32	3.70e−06	4.01	7.41e−06	4.01	2.29e−06	4.00	4.59e−06	4.00
	64	2.31e−07	4.00	4.63e−07	4.01	1.43e−07	4.00	2.87e−07	4.00
	128	1.45e−08	4.00	2.89e−08	4.00	8.96e−09	4.00	1.79e−08	4.00

Nonuniform meshes of $N \times N$ rectangles are obtained by randomly and independently perturbing each node in the x and y axes of a uniform mesh by up to some percentage. That is, for all $(i, j) \in \mathbb{Z}_N^0 \times \mathbb{Z}_N^0$,

$$x_i = \frac{2\pi i}{N} + \frac{0.01}{N} \sin\left(\frac{i\pi}{N}\right) \text{randn}(), \quad y_j = \frac{2\pi j}{N} + \frac{0.01}{N} \sin\left(\frac{j\pi}{N}\right) \text{randn}().$$

We compute the numerical solution at $t = 0.1$. The computational results are given in Tables 3 and 4, from which we observe a convergence rate of $\mathcal{O}(h^{k+1})$ for both $\|e_u\|_0$ and $\|e_{\mathbf{q}}\|_0$ in cases $\alpha = 1$ and $\alpha = 1 + \sin^2 x$. These results are consistent with the error estimates (3.34) in Theorem 3.5, which indicates that the error bound given in (3.34) is sharp.

5. Concluding remarks

In this work, we present and develop two classes of arbitrary high order SV schemes for solving diffusion equations by using the LDG formulation to discretize the viscous flux. We prove that both the proposed GSV and RSV schemes are energy stable when alternating fluxes are used. Optimal error estimates for both the exact solution and the auxiliary variable are also established. The comparison between the proposed SV method and the standard LDG method is given. A interesting discovery is that: the RSV is identical to the LDG method when applied to constant diffusion equations. Generally speaking, the SV method shares similar behaviors on stability, accuracy and convergence, from the perspective of theoretical analysis. While in terms of computation complexity, LDG requires both volume and surface integrations. In contrast, SV requires only surface integrations and thus requires less computational efforts or CPU time to reach the same accuracy than the counterpart LDG method. This advantage in computation efficiency makes SV method more appealing for solving some comparatively complicated nonlinear diffusion equations, *i.e.*, the porous medium equation (PME). The current study is our first attempt to investigate the difference between the LDG method and SV method, more theoretical and numerical investigations are needed to find advantages of SV methods. Our ongoing works include the study of SV methods for nonlinear diffusion problems and their superconvergence properties.

TABLE 4. L^2 errors and corresponding convergence rates of the SV method for Example 2 with $\alpha = 1 + \sin^2 x$ and $k = 1, 2, 3$.

		GSV				RSV			
k	N	$\|e_u\|_0$	Rate	$\|e_q\|_0$	Rate	$\|e_u\|_0$	Rate	$\|e_q\|_0$	Rate
	4	8.30e−01	–	2.54e−00	–	6.71e−01	–	1.03e−00	–
	8	3.09e−01	1.43	4.97e−01	2.35	1.57e−01	2.09	2.81e−01	1.87
1	16	7.99e−02	1.95	1.23e−01	2.01	3.85e−02	2.03	7.07e−02	1.99
	32	2.00e−02	2.00	3.06e−02	2.01	9.56e−03	2.01	1.77e−02	2.00
	64	4.99e−03	2.00	7.62e−03	2.00	2.39e−03	2.00	4.43e−03	2.00
	4	1.00e−01	–	3.80e−01	–	9.68e−02	–	2.53e−01	–
	8	1.90e−02	2.40	5.00e−02	2.92	1.22e−02	2.99	3.58e−02	2.82
2	16	2.43e−03	2.97	6.60e−03	2.92	1.53e−03	3.00	4.57e−03	2.97
	32	3.00e−04	3.02	8.30e−04	2.99	1.92e−04	2.99	5.79e−04	2.98
	64	3.88e−05	2.95	1.08e−04	2.94	2.46e−05	2.96	7.39e−05	2.97
	4	1.79e−02	–	7.75e−02	–	9.68e−03	–	6.05e−02	–
	8	1.19e−03	3.91	6.10e−03	3.67	5.90e−04	4.03	4.36e−03	3.79
3	16	6.48e−05	4.20	4.13e−04	3.89	3.67e−05	4.00	2.95e−04	3.89
	32	3.79e−06	4.10	2.61e−05	3.98	2.29e−06	4.00	1.87e−05	3.98
	64	2.33e−07	4.03	1.63e−06	4.00	1.44e−07	3.99	1.17e−06	4.00

APPENDIX A.

We begin with some preliminaries. For any function $w \in U_h$, let $w^* = \mathcal{F}w$ be defined in (3.9) with the coefficients given by (3.10). Then a direct calculation from (3.10) yields

$$\begin{aligned}
\left[w_{k,j}^{\tau,*}\right]^y &= A_j^{\tau,y}\left(w_y - A_{k+1}^{\tau,x} w_{xy}\right)\left(g_{k+1,j}^{\tau}\right),\\
\left[w_{i,k}^{\tau,*}\right]^x &= A_i^{\tau,x}\left(w_x - A_{k+1}^{\tau,y} w_{xy}\right)\left(g_{i,k+1}^{\tau}\right),\\
w_{0,k}^{\tau,*} &= \left(w + A_0^{\tau,x} w_x - A_{k+1}^{\tau,y} w_y - A_{0,k+1}^{\tau} w_{xy}\right)\left(g_{0,k+1}^{\tau}\right),\\
w_{k,0}^{\tau,*} &= \left(w + A_0^{\tau,y} w_y - A_{k+1}^{\tau,x} w_x - A_{k+1,0}^{\tau} w_{xy}\right)\left(g_{k+1,0}^{\tau}\right),\\
w_{k,k}^{\tau,*} &= \left(w - A_{k+1}^{\tau,x} w_x - A_{k+1}^{\tau,y} w_y + A_{k+1,k+1}^{\tau} w_{xy}\right)\left(g_{k+1,k+1}^{\tau}\right).
\end{aligned} \tag{A.1}$$

A.1. Proof of Lemma 3.1

Proof. Let $w^* = \mathcal{F}w$. Recalling the definition of $\langle , \rangle_{\partial\tau_y}$ in (3.11), we have

$$\langle v, w^*\rangle_{\partial\tau_y} = \sum_{j=0}^{k} w_{k,j}^{\tau,*} \int_{g_j^{\tau,y}}^{g_{j+1}^{\tau,y}} v\left(g_{k+1}^{\tau,x}, y\right) \mathrm{d}y - \sum_{j=0}^{k} w_{0,j}^{\tau,*} \int_{g_j^{\tau,y}}^{g_{j+1}^{\tau,y}} v(g_0^{\tau,x}, y)\, \mathrm{d}y := I_2 - I_1. \tag{A.2}$$

Then a direct calculation from (3.10) and (A.1) leads to

$$\begin{aligned}
I_1 &= -\sum_{j=1}^{k} \left[w_{0,j}^{\tau,*}\right]^y \partial_y^{-1} v\left(g_0^{\tau,x}, g_j^{\tau,y}\right) + w_{0,k}^{\tau,*} \partial_y^{-1} v\left(g_0^{\tau,x}, g_{k+1}^{\tau,y}\right) - w_{0,0}^{\tau,*} \partial_y^{-1} v\left(g_0^{\tau,x}, g_0^{\tau,y}\right)\\
&= -\sum_{j=0}^{k+1} \left(\left(A_j^{\tau,y} w_y + A_{0,j}^{\tau} w_{xy}\right)\partial_y^{-1} v\right)\left(g_{0,j}^{\tau}\right) + \left(w \partial_y^{-1} v\right)\Big|_{g_{0,0}^{\tau}}^{g_{0,k+1}^{\tau}} + A_0^{\tau,x} \left(w_x \partial_y^{-1} v\right)\Big|_{g_{0,0}^{\tau}}^{g_{0,k+1}^{\tau}}.
\end{aligned}$$

By using (3.7), (3.8) and the integration by parts,

$$I_1 = \int_{g_0^{\tau,y}}^{g_{k+1}^{\tau,y}} (wv)(g_0^{\tau,x}, y)\,\mathrm{d}y + R_\tau^y\big(w_y\partial_y^{-1}v\big)\big(g_0^{\tau,x}\big) + E_1(v,w).$$

Following the same argument, we can prove that

$$I_2 = \int_{g_0^{\tau,y}}^{g_{k+1}^{\tau,y}} (wv)\big(g_{k+1}^{\tau,x}, y\big)\,\mathrm{d}y + R_\tau^y\big(w_y\partial_y^{-1}v\big)\big(g_{k+1}^{\tau,x}\big) - E_2(v,w).$$

Substituting the last two equations into (A.2) yields the desired result (3.12).

We next prove (3.13). Denoting $\mu = \partial_x^{-1}\partial_y^{-1}v$. A direct calculation yields

$$(v,w^*)_\tau = \sum_{i=1}^{k}\sum_{j=1}^{k}\big[w_{i,j}^{\tau,*}\big]\mu\big(g_i^{\tau,x}, g_j^{\tau,y}\big) + \big\langle\partial_x^{-1}v, w^*\big\rangle_{\partial\tau_y} + F_\tau,$$

where

$$F_\tau = \sum_{i=1}^{k}\Big(-\Big[w_{i,k}^{\tau,*}\Big]^x\mu\big(g_i^{\tau,x}, g_{k+1}^{\tau,y}\big) + \big[w_{i,0}^{\tau,*}\big]^x\mu\big(g_i^{\tau,x}, g_0^{\tau,y}\big)\Big).$$

By using (3.10) and (A.1), we have

$$F_\tau = \sum_{j=0,k+1}\sum_{i=1}^{k} A_{i,j}^\tau(w_{xy}\mu)\big(g_{i,j}^\tau\big) - \sum_{i=1}^{k} A_i^{\tau,x}(w_x\mu)\big|_{g_{i,0}^\tau}^{g_{i,k+1}^\tau}.$$

Consequently,

$$\begin{aligned}(v,w^*) &= \sum_{i=1}^{k}\sum_{j=0}^{k+1} A_{i,j}^\tau(w_{xy}\mu)\big(g_{i,j}^\tau\big) - \sum_{i=1}^{k} A_i^{\tau,x}(w_x\mu)\big|_{g_{i,0}^\tau}^{g_{i,k+1}^\tau} + \langle\partial_x^{-1}v, w^*\rangle_{\partial\tau_y}\\ &= -\sum_{i=1}^{k} A_i^{\tau,x}\left(\int_{g_0^{\tau,y}}^{g_{k+1}^{\tau,y}}(w_x\mu_y)(g_i^{\tau,x}, y)\,\mathrm{d}y + R_\tau^y(w_{xy}\mu)(g_i^{\tau,x})\right) + \big\langle\partial_x^{-1}v, w^*\big\rangle_{\partial\tau_y}.\end{aligned}$$

Substituting (3.12) into the above inequality, using $\mu_y = \partial_x^{-1}v$ and the integration by parts again yields

$$\begin{aligned}(v,w^*) &= -\sum_{i=0}^{k+1} A_i^{\tau,x}\left(\int_{g_0^{\tau,y}}^{g_{k+1}^{\tau,y}}(w_x\mu_y)(g_i^{\tau,x}, y)\,\mathrm{d}y + R_\tau^y(w_{xy}\mu)(g_i^{\tau,x})\right) + \langle\partial_x^{-1}v, w\rangle_{\partial\tau_y} + R_\tau^y(w_y\mu)\big|_{g_0^{\tau,x}}^{g_{k+1}^{\tau,x}}\\ &= (w,\mu_{xy})_\tau + \int_{g_0^{\tau,y}}^{g_{k+1}^{\tau,y}} R_\tau^x(w_x\mu_y)\,\mathrm{d}y + \int_{g_0^{\tau,x}}^{g_{k+1}^{\tau,x}} R_\tau^y(w_y\mu_x)\,\mathrm{d}x + R_\tau^xR_\tau^y(w_{xy}\mu).\end{aligned}$$

Then (3.13) follows from the identities $\mu_{xy} = v, \mu_x = \partial_y^{-1}v, \mu_y = \partial_x^{-1}v$. □

A.2. Proof of Lemma 3.3

Proof. We only prove (3.23) for GSV in case $(\hat{B},\hat{p}) = (B^+, p^-)$, since the same argument can be applied directly to the other case $(\hat{B},\hat{p}) = (B^-, p^+)$ and to the other RSV schemes.

For GSV, since $s_j = \bar{s}_j, j \in \mathbb{Z}_k$, we have $\mathcal{E}_\tau = \bar{\mathcal{E}}_\tau, \mathcal{E}_\tau^i = \bar{\mathcal{E}}_\tau^i, i = 1,2$. Denote by $\bar{\beta}$ the cell average of β on $\tau = \tau_m^x \times \tau_n^y$, and

$$\beta_n(x) = \frac{1}{h_\tau^y}\int_{\tau_n^y}\beta(x,y)\,\mathrm{d}y, \quad B_n(v) = \int_a^x \beta_n(x)v_x\,\mathrm{d}x.$$

Since $(\hat{B},\hat{p})=(B^+,p^-)$, we have, from the definitions of $\mathcal{E}_\tau^1$ in (3.19) and $\mathcal{E}_\tau$ in (3.16),

$$I_\tau = R_\tau^y\big(p_y\partial_y^{-1}[B]\big)\Big(x_{m+\frac12}\Big) + R_\tau^y\big(v_y\partial_y^{-1}[\beta p]\big)\Big(x_{m-\frac12}\Big) + \int_{\tau_m^x} R_\tau^y\big(v_y\partial_y^{-1}(\beta p)_x + p_y\partial_y^{-1}B_x\big)\,\mathrm{d}x$$
$$+ \int_{\tau_n^y} R_\tau^x(v_x\beta p + p_xB)\,\mathrm{d}y + R_\tau^xR_\tau^y\big(v_{xy}\partial_y^{-1}(\beta p) + p_{xy}\partial_y^{-1}B\big) := \sum_{i=1}^5 I_i,$$

where

$$I_1 = R_\tau^y\big(p_y\partial_y^{-1}[B_n]\big)\Big(x_{m+\frac12}\Big) + R_\tau^y\big(v_y\partial_y^{-1}[\beta_n p]\big)\Big(x_{m-\frac12}\Big),$$
$$I_2 = \int_{\tau_m^x} R_\tau^y\big(v_y\partial_y^{-1}(\beta_n p)_x + p_y\partial_y^{-1}(B_n)_x\big)\,\mathrm{d}x,$$
$$I_3 = R_\tau^y\big(p_y\partial_y^{-1}[B-B_n]\big)\Big(x_{m+\frac12}\Big) + R_\tau^y\big(v_y\partial_y^{-1}[(\beta-\beta_n)p]\big)\Big(x_{m-\frac12}\Big),$$
$$I_4 = \int_{\tau_m^x} R_\tau^y\big(v_y\partial_y^{-1}((\beta-\beta_n)p)_x + p_y\partial_y^{-1}(B-B_n)_x\big)\,\mathrm{d}x,$$
$$I_5 = \int_{\tau_n^y} R_\tau^x(v_x\beta p + p_xB)\,\mathrm{d}y + R_\tau^xR_\tau^y\big(v_{xy}\partial_y^{-1}(\beta p) + p_{xy}\partial_y^{-1}B\big).$$

We next estimate $I_i, i\le 5$, respectively. By using the Gauss numerical quadrature error (3.24) and the identities $[B_n](x_{m+\frac12},y)=\beta_n[v](x_{m+\frac12},y)$ and $(B_n)_x=\beta_nv_x$,

$$I_1 = c(h_\tau^y)^{2k+1}\Big(\big(\beta_n\partial_y^k p[\partial_y^k v]\big)\Big(x_{m+\frac12},\theta\Big) + \big(\beta_n\partial_y^k v[\partial_y^k p]\big)\Big(x_{m-\frac12},\theta\Big)\Big),$$
$$I_2 = c(h_\tau^y)^{2k+1}\Big(\big(\beta_n\partial_y^k p\partial_y^k v\big)\Big(x_{m+\frac12},\theta\Big) - \big(\beta_n\partial_y^k p\partial_y^k v\big)\Big(x_{m-\frac12},\theta\Big)\Big),$$

where c is a constant only dependent on k and θ is a point in τ_n^y. Summing up all τ and using the periodic boundary condition, we easily get

$$\sum_{\tau\in\mathcal{T}_h}(I_1+I_2)=0.$$

To estimate I_3, I_4, we first choose $l=1$ and $f=\partial_y^{-1}(\beta-\beta_n)p$ in (2.28) to get for all $x\in\tau_m^x$

$$\big|R_\tau^y\big(v_y\partial_y^{-1}((\beta-\beta_n)p)\big)(x)\big| \lesssim h\bigg(\int_{\tau_n^y} v_y^2(x,y)\,\mathrm{d}y\bigg)^{\frac12}\bigg(\int_{\tau_n^y}((\beta-\beta_n)p)^2(x,y)\,\mathrm{d}y\bigg)^{\frac12}.$$

Then it is concluded from the inverse inequality that

$$\big|R_\tau^y\big(v_y\partial_y^{-1}((\beta-\beta_n)p)\big)(x)\big| \lesssim \|v\|_{0,\tau}\|p\|_{0,\tau}, \qquad \forall x\in\tau_m^x, \tag{A.3}$$

which yields, together with the identity $[B-B_n](x_{m+\frac12})=(\beta-\beta_n)[v](x_{m+\frac12})$,

$$\sum_{\tau\in\mathcal{T}_h}|I_3| \le \sum_{\tau\in\mathcal{T}_h}\Big|R_\tau^y\big(p_y\partial_y^{-1}[(\beta-\beta_n)v]\big)\Big(x_{m+\frac12}\Big) + R_\tau^y\big(v_y\partial_y^{-1}[(\beta-\beta_n)p]\big)\Big(x_{m-\frac12}\Big)\Big|$$
$$\lesssim \|v\|_0\|p\|_0.$$

Similarly, using (A.3) and the inverse inequality again yields

$$\sum_{\tau\in\mathcal{T}_h}|I_4| \le h\sum_{\tau\in\mathcal{T}_h}\max_{x\in\tau_m^x}\big|R_\tau^y\big(v_y\partial_y^{-1}((\beta-\beta_n)p)_x + p_y\partial_y^{-1}((\beta-\beta_n)v_x)\big)\big|$$

$$\lesssim h(\|v\|_0\|p_x\|_0 + \|v_x\|_0\|p\|_0) \lesssim \|v\|_0\|p\|_0.$$

As for I_5, we take $l = k$ and $f = B$ and use $\partial_x^{k+1}B = \partial_x^k(\beta v_x)$ in (2.28) to get

$$|R_\tau^x(p_xB)| \lesssim h^k\left(\int_{\tau_m^x} p^2(x,y)\,\mathrm{d}x\right)^{\frac{1}{2}}\left(\int_{\tau_m^x}\left(\partial_x^k v\right)^2(x,y)\,\mathrm{d}x\right)^{\frac{1}{2}},$$

which yields, together with the Cauchy–Schwartz inequality and the inverse inequality

$$\int_{\tau_n^y} R_\tau^x(p_xB)\,\mathrm{d}y \lesssim \|v\|_{0,\tau}\|p\|_{0,\tau}. \tag{A.4}$$

Similarly, by using (2.28) and the fact that the Gauss numerical quadrature is exact for polynomial of degree of not more than $2k-1$, we get

$$|R_\tau^x(v_x\beta p)| = \left|R_\tau^x\left(v_x(\beta-\bar\beta)p\right)\right| \lesssim \left(\int_{\tau_m^x} v_x^2(x,y)\,\mathrm{d}x\right)^{\frac{1}{2}}\left(\int_{\tau_m^x}\left(\beta p-\bar\beta p\right)^2(x,y)\,\mathrm{d}x\right)^{\frac{1}{2}},$$

and thus

$$\int_{\tau_n^y} R_\tau^x(v_x\beta p) \lesssim \|v_x\|_{0,\tau}\left\|(\beta-\bar\beta)p\right\|_{0,\tau} \lesssim \|v\|_{0,\tau}\|p\|_{0,\tau}. \tag{A.5}$$

On the other hand, as a direct consequence of (2.28), there holds

$$|R_\tau^x(f)| \lesssim h\|f(\cdot,y)\|_{0,\infty,\tau_m^x}.$$

Consequently, we choose $l = 1$ in (2.28) again to obtain

$$\begin{aligned}\left|R_\tau^xR_\tau^y\left(v_{xy}\partial_y^{-1}(\beta p)\right)\right| &\lesssim h\max_{x\in\tau_m^x}\left|R_\tau^y\left(v_{xy}\partial_y^{-1}\left(\beta p-\bar\beta p\right)\right)\right|\\ &\lesssim h^2\max_{x\in\tau_m^x}\left(\int_{\tau_n^y} v_{xy}^2(x,y)\,\mathrm{d}y\right)^{\frac{1}{2}}\left(\int_{\tau_n^y}\left(\beta p-\bar\beta p\right)^2(x,y)\,\mathrm{d}y\right)^{\frac{1}{2}}.\end{aligned}$$

By using the inverse inequality, we have

$$\left|R_\tau^xR_\tau^y\left(v_{xy}\partial_y^{-1}(\beta p)\right)\right| \lesssim h\|v_x\|_{0,\tau}\|p\|_{0,\tau} \lesssim \|v\|_{0,\tau}\|p\|_{0,\tau}, \qquad \forall v,p\in U_h. \tag{A.6}$$

Similarly, we can prove

$$\left|R_\tau^xR_\tau^y\left(p_{xy}\partial_y^{-1}B(v)\right)\right| \lesssim \|v\|_{0,\tau}\|p\|_{0,\tau}, \qquad \forall v,p\in U_h. \tag{A.7}$$

Combining (A.4)–(A.7) together leads to

$$|I_5| \lesssim \|v\|_{0,\tau}\|p\|_{0,\tau}.$$

Substituting the estimates of $I_i, i\le 5$ into the formula of I_τ yields

$$\left|\sum_{\tau\in\mathcal{T}_h} I_\tau\right| \lesssim \|v\|_0\|p\|_0.$$

This finishes the proof the first inequality of (3.23). The second inequality of (3.23) follows from the same argument and thus we omit it here. □

Acknowledgements. This work is supported in part by the National Natural Science Foundation of China under grants Nos. 12061023, 12271049.

References

[1] D.N. Arnold, An interior penalty finite element method with discontinuous elements. *SIAM J. Numer. Anal.* **19** (1982) 742–760.

[2] C.E. Baumann and J.T. Oden, A discontinuous *hp* finite element method for convection-diffusion problems. *Comput. Methods Appl. Mech. Eng.* **175** (1999) 311–341.

[3] W. Cao and Q. Zou, Analysis of spectral volume methods for 1D linear scalar hyperbolic equations. *J. Sci. Comput.* **90** (2022) 1–29.

[4] P. Castillo, B. Cockburn, I. Perugia and D. Schötzau, An a priori error analysis of the local discontinuous Galerkin method for elliptic problems, *SIAM J. Numer. Anal.* **38** (2000) 1676–1706.

[5] B. Cockburn and C.-W. Shu, TVB Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws, II: general framework. *Math. Comp.* **52** (1989) 411–435.

[6] B. Cockburn and C.-W. Shu, The Runge–Kutta discontinuous Galerkin method for conservation laws, V: multidimensional systems. *J. Comput. Phys.* **141** (1998) 199–224.

[7] B. Cockburn and C.-W. Shu, The local discontinuous Galerkin method for time-dependent convection-diffusion systems. *SIAM J. Numer. Anal.* **35** (1998) 2440–2463.

[8] B. Cockburn, G. Karniadakis and C.-W. Shu, The Development of Discontinuous Galerkin Methods. Springer Berlin Heidelberg (2000).

[9] B. Cockburn, G. Kanschat and D. Schotzau, A locally conservative LDG method for the incompressible Navier–Stokes equations. *Math. Comput.* **74** (2005) 1067–1095.

[10] S.K. Godunov, A finite-difference method for the numerical computation of discontinuous solutions of the equations of fluid dynamics. *Mat. Sb.* **47** (1959) 271.

[11] P. Houston, C. Schwab and E. Suli, Discontinuous *hp*-finite element methods for advection-diffusion-reaction problems. *SIAM J. Numer. Anal.* **39** (2002) 2133–2163.

[12] R. Kannan, A high order spectral volume formulation for solving equations containing higher spatial derivative terms: formulation and analysis for third derivative spatial terms using the LDG discretization procedure. *Commun. Comput. Phys.* **10** (2011) 1257–1279.

[13] P. Kannan and Z.J. Wang, A study of viscous flux formulations for a *p*-multigrid spectral volume Navier–Stokes solver. *J. Sci. Comput.* **41** (2009) 165–199.

[14] R. Kannan and Z.J. Wang, The direct discontinuous Galerkin (DDG) viscous flux scheme for the high order spectral volume method. *Comput. Fluids* **39** (2010) 2007–2021.

[15] R. Kannan and Z.J. Wang, LDG2: a variant of the LDG flux formulation for the spectral volume method. *J. Sci. Comput.* **46** (2011) 314–328.

[16] H. Liu and J. Yan, The direct discontinuous Galerkin (DDG) methods for diffusion problems. *SIAM J. Numer. Anal.* **47** (2009) 675–698.

[17] Y. Liu, M. Vinokur and Z. Wang, Spectral (finite) volume method for conservation laws on unstructured grids V: extension to three-dimensional systems. *J. Comput. Phys.* **212** (2006) 454–472.

[18] J.T. Oden, I. Babuška and C.E. Baumann, A discontinuous *hp* finite element method for convection-diffusion problems. *J. Comput. Phys.* **146** (1998) 491–519.

[19] J. Peraire and P.-O. Persson, The compact discontinuous Galerkin (CDG) method for elliptic problems. *SIAM J. Sci. Comput.* **30** (2008) 1806–1824.

[20] R. Raghavendra, A high order spectral volume method for equations containing third spatial derivatives using an Interior Penalty formulation. *CFD Lett.* **3** (2011) 74–88.

[21] Y. Sun and Z. Wang, Evaluation of discontinuous Galerkin and spectral volume methods for scalar and system conservation laws unconstructured grids. *Int. J. Numer. Meth. Fluids* **45** (2004) 819–838.

[22] Y. Sun, Z. Wang and Y. Liu, Spectral (finite) volume method for conservation laws on unstructured grids. VI. Extension to viscous flow. *J. Comput. Phys.* **215** (2006) 41–58.

[23] Y. Sun, Z. Wang and Y. Liu, High-order multidomain spectral difference method for the Navier–Stokes equations on unstructured hexahedral grids. *Commun. Comput. Phys.* **2** (2007) 310–333.

[24] K. Van den Abeele and C. Lacor, An accuracy and stability study of the 2D spectral volume method. *J. Comput. Phys.* **226** (2007) 1007–1026.

[25] K. Van den Abeele, C. Lacor and Z. Wang, On the connection btetween the spectral volume method and the spectral difference method. IV. Extension to two-dimensional systems. *J. Comput. Phys.* **227** (2007) 877–885.

[26] K. Van den Abeele, T. Broeckhoven and C. Lacor, Dispersion and dissipation properties of the 1D spectral volume method and application to a *p*-multigrid algorithm. *J. Comput. Phys.* **224** (2007) 616–636.

[27] K. Van den Abeele, G. Ghorbaniasl, M. Parsani and C. Lacor, A stability analysis for the spectral volume method on tetrahedral grids. *J. Comput. Phys.* **228** (2009) 257–265.

[28] M.F. Wheeler, An elliptic collocation-finite element method with interior penalties. *SIAM J. Numer. Anal.* **15** (1978) 152–161.

[29] Z. Wang, Spectral (finite) volume method for conservation laws on unstructured grids: basic formulation. *J. Comput. Phys.* **178** (2002) 210–251.

[30] Z. Wang and Y. Liu, Spectral (finite) volume method for conservation laws on unstructured grids. II. Extension to two-dimensional scalar equation. *J. Comput. Phys.* **179** (2002) 665–697.

[31] Z. Wang and Y. Liu, Spectral (finite) volume method for conservation laws on unstructured grids. III. One dimensional systems and partition optimization. *J. Sci. Comput.* **20** (2004) 137–157.
[32] Z. Wang, L. Zhang and Y. Liu, Spectral (finite) volume method for conservation laws on unstructured grids. IV. Extension to two-dimensional systems. *J. Comput. Phys.* **194** (2004) 716–741.
[33] J. Yan, A new nonsymmetric discontinuous Galerkin method for time dependent convection diffusion equations. *J. Sci. Comput.* **54** (2013) 663–683.
[34] M. Zhang and C.-W. Shu, An analysis of and a comparison between the discontinuous Galerkin and the spectral finite volume methods. *Comput. Fluids* **34** (2005) 581–592.