Superconvergence of the Direct Discontinuous Galerkin Method for Convection-Diffusion Equations

Waixiang Cao,[1] **Hailiang Liu,**[2] **Zhimin Zhang**[1,3]
[1] *Division of Applied and Computational Mathematics, Beijing Computational Science Research Center, Beijing 100193, China*

[2] *Department of Mathematics, Iowa State University, Ames, Iowa 50011*

[3] *Department of Mathematics, Wayne State University, Detroit, Michigan 48202*

Received 12 October 2015; accepted 6 July 2016
Published online 23 July 2016 in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/num.22087

This paper is concerned with superconvergence properties of the direct discontinuous Galerkin (DDG) method for one-dimensional linear convection-diffusion equations. We prove, under some suitable choice of numerical fluxes and initial discretization, a $2k$-th and $(k+2)$ -th order superconvergence rate of the DDG approximation at nodes and Lobatto points, respectively, and a $(k+1)$ -th order of the derivative approximation at Gauss points, where k is the polynomial degree. Moreover, we also prove that the DDG solution is superconvergent with an order $k+2$ to a particular projection of the exact solution. Numerical experiments are presented to validate the theoretical results. Numer Methods Partial Differential Eq 33: 290–317, 2017

Keywords: convection-diffusion equations; direct discontinuous Galerkin methods; superconvergence

I. INTRODUCTION

In this paper, we study the superconvergence of the direct discontinuous Galerkin (DDG) method for the one-dimensional linear convection-diffusion equation

$$\begin{aligned} &\partial_t u + \partial_x f(u) = \partial_x^2 u, \quad (x,t) \in [a,b] \times [0,T], \\ &u(x,0) = u_0(x), \quad x \in [a,b], \end{aligned} \tag{1.1}$$

Correspondence to: Waixiang Cao, Beijing Computational Science Research Center, Beijing 100094, China (e-mail: wxcao@csrc.ac.cn)
Contract grant sponsor: China Postdoctoral Science Foundation; contract grant number: 2015M570026
Contract grant sponsor: National Natural Science Foundation of China (NSFC); contract grant numbers: 11501026, 11471031, and 91430216
Contract grant sponsor: US National Science Foundation (NSF); contract grant number: DMS-1419040
Contract grant sponsor: US National Science Foundation (NSF); contract grant number: DMS-1312636 and RNMS (KI-Net) 1107291 (H. Liu)

where the linear flux $f = \alpha u$. For simplicity, we only consider the case $\alpha \geq 0$ and the periodic boundary condition.

For equations containing higher-order spatial derivatives, such as the convection-diffusion equation (1.1), the discontinuous Galerkin (DG) method [1–6] cannot be directly applied, due to the discontinuous solution space at the element interfaces, which is not regular enough to handle higher derivatives. Several DG methods have been suggested in the literature to solve this problem, including the local discontinuous Galerkin (LDG) method [7], the interior penalty (IP) method [8–10], and the method introduced by Baumann-Oden [11, 12], and the direct DG discretization methods (see, e.g., [13–15]). In 2009, Liu and Yan [16] introduced a new direct DG method for convection-diffusion problems. The basic idea of this method is to directly force the weak formulation of the PDE into the DG method, without introducing any auxiliary variables or rewriting the original equation into a first-order system. It was also discussed in this paper how a careful choice of numerical fluxes can be made to ensure stability and accuracy of the DDG method. Later, Liu et al. studied a set of more effective choices of parameters in the DDG numerical fluxes [17, 18], and proved an optimal L^2 error estimates for the DDG methods solving convection-diffusion problems in [19].

The key feature of DDG methods lies in its special choice of the numerical flux for the solution derivative at cell interfaces. This gives DDG methods extra flexibility and advantage over the IPDG method [20] and the LDG method [7]. The DDG method is the only known DG method which can satisfy the maximum principle for Fokker-Planck equations up to third order [21], yet both IPDG method and LDG method can be proved to satisfy maximum principle with only up to second order of accuracy even for diffusion without drift (see e.g. [22]). The additional parameter in the DDG numerical flux also makes it possible for the DG method for Fokker–Planck type equations to satisfy some entropy dissipation law at the discrete level, see, e.g. [23, 24]. The superconvergence result obtained for a nontrivial β_1 presented in this paper provides another example of advantages of the DDG method.

The main purpose of this paper is to study and reveal the superconvergence phenomenon of the DDG method for the convection-diffusion equation (1.1). To the best of our knowledge, there was no any superconvergence result of the DDG method for these problems in the literature. Our results include the superconvergence properties of the DDG approximation at nodal and Lobatto points, and the derivative approximation at Gauss points, as well as the supercloseness between the DDG solution and a particular projection of the exact solution. To be more precise, we prove that, under a careful choice of the parameters in the DDG numerical fluxes, the DDG solution is superconvergent at all Gauss (derivative approximation) and Lobatto points, with an order $k+1$ and $k+2$, respectively; and the error between the DDG solution and the Gauss–Lobatto projection of the exact solution achieves $(k+2)$ -th order. We further prove a $2k$-th order superconvergence rate of the DDG approximation at nodes. As we may recall, all these superconvergence results are the same as the counterpart finite element methods (see. e.g., [25–27]). Moreover, the influence of the choice of the parameters in the numerical fluxes on the superconvergence is also discussed.

The analysis is based on the correction function idea, which is motivated from its successful application to the FEM and finite volume methods (FVM) for Poisson equations [28, 29], and the DG or LDG methods for hyperbolic and parabolic equations [30–32]. The essence of the correction idea is to construct a specially designed function which vanishes or is of high order at some special points. With the correction function, we first construct a special interpolation function of the exact solution, and then prove the interpolation function is superconvergent to the numerical solution, with the highest order $2k$. Finally the supercloseness between the interpolation function and the numerical solution gives the desired superconvergence results at some special points: Gauss, Lobatto, and nodal points.

The main difficulty in the superconvergence analysis lies in the construction of the correction function (or the special interpolation function) and the choice of the parameters in the numerical fluxes. The correction function here is greatly different from those in [30–32], due to the different model problems and numerical schemes. The existence of the convection-diffusion term $f(u)$ and the special choice of numerical fluxes make the construction of the correction function more sophisticated. It is also different from the correction functions of FEM or FVM for elliptic equations in [28, 29] because of the time-dependent feature. Furthermore, as the choice of the parameters in the numerical fluxes influences the superconvergence directly (a feature not shared by those methods in [28–32]), how a careful choice of the parameters in the numerical fluxes to ensure some desired superconvergence results (including superconvergence at Gauss and Lobatto points) is also a technical question needed to handle.

The rest of the paper is organized as follows. In section 2, we recall the DDG schemes and the basic stability results for linear convection-diffusion equations, following [19]. Sections 3 and 4 are the main body of the paper, where superconvergence results for zero flux and linear flux are proved separately, with suitable initial discretization and careful choice of numerical fluxes. Finally, some numerical examples are presented in section 5 to support our theoretical findings.

Throughout this paper, we adopt standard notations for Sobolev spaces such as $W^{m,p}(D)$ on subdomain $D \subset \Omega$ equipped with the norm $\|\cdot\|_{m,p,D}$ and seminorm $|\cdot|_{m,p,D}$. When $D = \Omega$, we omit the index D; and if $p=2$, we set $W^{m,p}(D) = H^m(D)$, $\|\cdot\|_{m,p,D} = \|\cdot\|_{m,D}$, and $|\cdot|_{m,p,D} = |\cdot|_{m,D}$. We use the notation $A \lesssim B$ to indicate that A can be bounded by B multiplied by a constant independent of the mesh size h. $A \simeq B$ stands for $A \lesssim B$ and $B \lesssim A$.

II. DDG SCHEMES

To discretize the weak formulation, we divide the interval $\Omega = [a, b]$ with a mesh: $a = x_{\frac{1}{2}} < x_{\frac{3}{2}} < \cdots < x_{N+\frac{1}{2}} = b$, and the mesh size $h = h_j = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}$, and the interval $I_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$. We denote by v^+ and v^- the right and left limits of any function v, and define

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

Define the k-degree discontinuous finite element space

$$V_h = \left\{ v_h : v_h|_{I_j} \in \mathbb{P}^k(I_j), j \in \mathbb{Z}_N \right\},$$

where $\mathbb{P}^k(I_j)$ denotes the set of all polynomials of degree no more than k on I_j, and $\mathbb{Z}_r = \{1, 2, \ldots, r\}$ for any positive integer r.

The DDG method for (1.1) is to find $u_h \in V_h$ such that for any $v_h \in V_h$

$$(\partial_t u_h, v_h)_j = (f(u_h) - \partial_x u_h, \partial_x v_h)_j + ((-\hat{f}(u_h) + \widehat{\partial_x u_h})v_h + (u_h - \{u_h\})\partial_x v_h)|_{\partial I_j}. \qquad (2.1)$$

Here $\hat{f}(u_h)$ and $\widehat{\partial_x u_h}$ are numerical fluxes, and

$$(\varphi, v)_j = \int_{I_j} \varphi v dx, \quad v|_{\partial I_j} = v\left(x^-_{j+\frac{1}{2}}\right) - v\left(x^+_{j-\frac{1}{2}}\right), \quad \forall \varphi, v \in L^2.$$

Following [19], we take the numerical fluxes $\hat{f}(u_h) = \alpha u_h^-$ and

$$\widehat{\partial_x u_h} = \frac{\beta_0}{h}[u_h] + \{\partial_x u_h\} + \beta_1 h[\partial_x^2 u_h] \tag{2.2}$$

with $\beta_i, i = 0, 1$ satisfying the following stability condition

$$\beta_0 \geq \Gamma(\beta_1),$$

where

$$\Gamma(\beta_1) = \sup_{v \in \mathbb{P}^{k-1}[-1,1]} \frac{2(v(1) - 2\beta_1 \partial_s v(1))^2}{\int_{-1}^{1} v^2(s)ds}.$$

Summing up all j and using the periodic boundary condition, we have from (2.1)

$$\sum_{j=1}^{N}(\partial_t u_h, v_h)_j + \sum_{j=1}^{N}(-f(u_h) + \partial_x u_h, \partial_x v_h)_j + \sum_{j=1}^{N}((-\hat{f}(u_h) + \widehat{\partial_x u_h})[v_h] + [u_h]\{\partial_x v_h\})|_{j+\frac{1}{2}} = 0,$$

or equivalently,

$$(\partial_t u_h, v_h) + A(u_h, v_h) - F(u_h, v_h) = 0, \quad \forall v_h \in V_h, \tag{2.3}$$

where $(u_h, v_h) = \sum_{j=1}^{N} (u_h, v_h)_j$, and

$$A(u_h, v_h) = \sum_{j=1}^{N} (\partial_x u_h, \partial_x v_h)_j + \sum_{j=1}^{N}(\widehat{\partial_x u_h}[v_h] + \{\partial_x v_h\}[u_h])|_{j+\frac{1}{2}}, \tag{2.4}$$

$$F(u_h, v_h) = \sum_{j=1}^{N} (f(u_h), \partial_x v_h)_j + \sum_{j=1}^{N} \hat{f}(u_h)[v_h]|_{j+\frac{1}{2}}. \tag{2.5}$$

For any function $v_h \in V_h$ satisfying the periodic boundary condition, we have (see [19])

$$F(v_h, v_h) \leq 0, \quad A(v_h, v_h) \geq \gamma \|v_h\|_E^2,$$

where $\gamma \in (0, 1)$ is some positive constant and

$$\|v_h\|_E^2 = \sum_{j=1}^{N} \int_{I_j} |\partial_x v_h|^2 dx + \sum_{j=1}^{N} \frac{\beta_0}{h}[v_h]^2|_{j+\frac{1}{2}}. \tag{2.6}$$

Then

$$\frac{1}{2}\frac{d}{dt}\|v_h\|_0^2 + \gamma\|v_h\|_E^2 \leq (\partial_t v_h, v_h) + A(v_h, v_h) - F(v_h, v_h), \quad \forall v_h \in V_h. \tag{2.7}$$

III. SUPERCONVERGENCE FOR ZERO FLUX *f* = 0

Before the study of superconvergence, we begin by introducing some notations, two important projections P_h and I_h, and the supercloseness between $I_h v$ and $P_h v$ for any smooth function v, which will be frequently used in our later superconvergence analysis.

A. Preliminaries

First, for a smooth function v, we define a global projection $P_h v \in V_h$ of v by

$$(P_h v, \varphi_h)_j = (v, \varphi_h)_j, \quad \forall \varphi_h \in \mathbb{P}^{k-2}(I_j), j \in \mathbb{Z}_N, \tag{3.1}$$

$$\widehat{\partial_x (P_h v)} := \beta_0 h^{-1}[P_h v] + \{\partial_x(P_h v)\} + \beta_1 h[\partial_x^2(P_h v)]|_{j+\frac{1}{2}} = \partial_x v(x_{j+\frac{1}{2}}), \tag{3.2}$$

$$\{P_h v\}|_{j+\frac{1}{2}} = v(x_{j+\frac{1}{2}}), \quad j \in \mathbb{Z}_N. \tag{3.3}$$

Note that we use the periodic boundary condition at $x_{N+\frac{1}{2}}$, and P_h only needs to satisfy the conditions (3.2)–(3.3) when $k = 1$. It is shown in [19] that the global projection P_h is uniquely defined, and

$$\|v - P_h v\|_0 \lesssim h^{k+1}\|v\|_{k+1}. \tag{3.4}$$

Second, denote by L_m the Legendre polynomial of degree m, and $\{\phi_m\}_0^\infty$ the series of Lobatto polynomials, on the interval $[-1, 1]$. That is, $\phi_0 = \frac{1-s}{2}, \phi_1 = \frac{s+1}{2}$, and

$$\phi_{m+1}(s) = \int_{-1}^{s} L_m(s')ds', \quad m \geq 1, s \in [-1, 1].$$

The orthogonal property of Legendre polynomials gives

$$\phi_m(-1) = \phi_m(1) = 0, \quad m \geq 2. \tag{3.5}$$

Moreover, there holds the identity

$$\phi_m = \frac{1}{2m-1}(L_m - L_{m-2}), \quad m \geq 2.$$

Therefore, we have for all $m, r \geq 2$,

$$\int_{-1}^{1} \phi_m \phi_r ds = \frac{1}{(2m-1)(2r-1)} \int_{-1}^{1} (L_m - L_{m-2})(L_r - L_{r-2})ds = 0, \quad \text{if } m - r \neq 0, \pm 2. \tag{3.6}$$

Let $L_{j,m}$ and $\phi_{j,m}$ be the Legendre and Lobatto polynomials of degree m on the interval I_j, respectively. That is,

$$L_{j,m}(x) = L_m(s), \quad \phi_{j,m}(x) = \phi_m(s), \quad s = \frac{2x - x_{j-\frac{1}{2}} - x_{j+\frac{1}{2}}}{h}, m \geq 0.$$

For any smooth function v, we have the following expansion in each element I_j,

$$v(x) = \sum_{m=0}^{\infty} v_{j,m}\phi_{j,m}(x), \quad \forall x \in I_j,$$

where

$$v_{j,0} = v\big(x_{j-\frac{1}{2}}^+\big), \quad v_{j,1} = v\big(x_{j+\frac{1}{2}}^-\big), \quad v_{j,m} = \frac{2m-1}{2}\int_{I_j} \partial_x v L_{j,m-1} dx, m \geq 2. \tag{3.7}$$

It has been proved in [29] that

$$|v_{j,m+1}| \lesssim h^{i+1-\frac{1}{p}} \|v\|_{i+1,p,I_j}, \quad 0 \le i \le m. \tag{3.8}$$

Now we define the Gauss–Lobatto projection of v by

$$I_h v|_{I_j} := \sum_{m=0}^{k} v_{j,m}\phi_{j,m}(x). \tag{3.9}$$

By (3.5)–(3.6) and a scaling from $[-1,1]$ to I_j, we have

$$\phi_{j,m}\big(x^+_{j-\frac{1}{2}}\big) = \phi_{j,m}\big(x^-_{j+\frac{1}{2}}\big) = 0, \quad \phi_{j,m+1} \perp \mathbb{P}^{m-2}(I_j), \quad \forall m \ge 2,$$

which yields

$$(v - I_h v)\big(x^-_{j+\frac{1}{2}}\big) = (v - I_h v)\big(x^+_{j-\frac{1}{2}}\big) = 0, \quad (v - I_h v) \perp \mathbb{P}^{k-2}.$$

Moreover, since

$$\partial_x (v - I_h v)(x) = \frac{2}{h} \sum_{m=k+1}^{\infty} v_{j,m} L_{j,m-1}(x), \quad x \in I_j, \tag{3.10}$$

we obtain from the orthogonality of the Legendre polynomials

$$\partial_x (v - I_h v) \perp \mathbb{P}^{k-1}. \tag{3.11}$$

To estimate the error $I_h v - P_h v$ for any smooth function v, we need the following important result.

Lemma 3.1. *Suppose M is an $N \times N$ block circulant matrix with the first row $[C B O \dots O]$ and the last row $[B O \dots O C]$, where O is an 2×2 zero matrix, and both C and B are 2×2 nonsingular matrix satisfying* $\det(C \pm B) \ne 0$. *Then*

$$|\rho(M^{-1})| \le c,$$

where $\rho(M)$ denotes the spectral radius of M, and c is some bounded constant independent of N.

Proof. For any vector or matrix X, we first denote by $\|X\|$ the L^2 norm of X. Let $\omega_j = e^{i\frac{2\pi j}{N}}$, $j = 0, \dots, N-1$ be N roots of 1 with $i^2 = -1$, and $\Omega_j = \begin{pmatrix} \omega_j & 0 \\ 0 & \omega_j \end{pmatrix}$. It is easy to verify from the properties of the circulant matrix

$$M\zeta_j = (C + B\Omega_j)\zeta_j, \quad \zeta_j = (I, \Omega_j, \Omega_j^2, \dots, \Omega_j^{N-1})^T,$$

where I is the 2×2 identity matrix. By denoting

$$\Upsilon = \left(\frac{\zeta_0}{\|\zeta_0\|}, \frac{\zeta_1}{\|\zeta_1\|}, \dots, \frac{\zeta_{N-1}}{\|\zeta_{N-1}\|} \right), \quad \Lambda = diag(C + B\Omega_0, C + B\Omega_1, \dots, C + B\Omega_{N-1}),$$

we have $M = \Upsilon\Lambda\Upsilon^{-1}$. By a direct calculation,

$$(\zeta_l)^T\zeta_j = \begin{pmatrix} \sum_{m=0}^{N-1} e^{i\frac{(j-l)2\pi m}{N}} & 0 \\ 0 & \sum_{m=0}^{N-1} e^{i\frac{(j-l)2\pi m}{N}} \end{pmatrix} = \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & \text{if} j = l, \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & \text{if} j \neq l. \end{cases}$$

Then Υ is an orthogonal matrix, which yields

$$\|\Upsilon\| = \|\Upsilon^{-1}\| = 1.$$

On the other hand, since $\det(C \pm B) \neq 0$, both M and Λ are nonsingular matrices. Then each $C + B\Omega_j$ is invertible and thus,

$$|\rho(M^{-1})| = \|M^{-1}\| \leq \|\Upsilon\|\|\Lambda^{-1}\|\|\Upsilon^{-1}\| \lesssim \max_j \det(C + B\Omega_j)^{-1} \lesssim 1.$$

Then the desired result follows. ∎

Now we are ready to show the supercloseness between $I_h v$ and $P_h v$.

Lemma 3.2. *For any function* $v \in H^{k+2}$, *if* $\beta_1 = \frac{1}{2k(k+1)}$, *then*

$$\|P_h v - I_h v\|_0 \lesssim h^{k+2}\|v\|_{k+2}. \tag{3.12}$$

Proof. Suppose $P_h v - I_h v$ has the following formula in I_j,

$$(P_h v - I_h v)(x) = \sum_{m=0}^{k} \bar{v}_{j,m} L_{j,m}(x), \quad x \in I_j.$$

Recalling the definition of P_h and the properties of I_h, we have

$$(P_h v - I_h v, \varphi_h)_j = 0, \quad \varphi_h \in \mathbb{P}^{k-2}(I_j), \quad \{P_h v - I_h v\}|_{j+\frac{1}{2}} = 0,$$
$$(\widehat{\partial_x(P_h v)} - \widehat{\partial_x(I_h v)})|_{j+\frac{1}{2}} = (\partial_x v - \widehat{\partial_x(I_h v)})|_{j+\frac{1}{2}},$$

which yields $\bar{v}_{j,m} = 0, m \leq k-2$, and

$$\sum_{m=k-1}^{k} (L_m(1)\bar{v}_{j,m} + L_m(-1)\bar{v}_{j+1,m}) = 0,$$
$$\sum_{m=k-1}^{k} (g_0(m)\bar{v}_{j,m} + g_1(m)\bar{v}_{j+1,m}) = h(\partial_x v - \widehat{\partial_x(I_h v)})|_{j+\frac{1}{2}} = b_j,$$

where

$$g_0(m) = -\beta_0 L_m(1) + L'_m(1) - 4\beta_1 L''_m(1), \quad g_1(m) = \beta_0 L_m(-1) + L'_m(-1) + 4\beta_1 L''_m(1). \tag{3.13}$$

Let M be an $N \times N$ block circulant matrix with the first row $[C B O \dots O]$ and the last row $[B O \dots O C]$, where O is an 2×2 zero matrix, and

$$C = \begin{pmatrix} L_{k-1}(1) & L_k(1) \\ g_0(k-1) & g_0(k) \end{pmatrix}, \; B = \begin{pmatrix} L_{k-1}(-1) & L_k(-1) \\ g_1(k-1) & g_1(k) \end{pmatrix}.$$

Then $\left\{\bar{v}^j\right\}_{j=1}^N$ with $\bar{v}^j = (\bar{v}_{j,k-1}, \bar{v}_{j,k})$ satisfies the linear system

$$MX = b, \quad X = \left(\bar{v}^1, \dots, \bar{v}^N\right)^T, \quad b = (0, b_1, 0, b_2, \dots, 0, b_N)^T. \tag{3.14}$$

It is proved in [19] that $\det(C \pm B) \neq 0$ when $\beta_0 > \Gamma(\beta_1)$. Then the conclusion in Lemma 3.1 gives

$$\sum_{j=1}^N \sum_{m=k-1}^k \bar{v}_{j,m}^2 \lesssim \|X\|^2 \lesssim \|M^{-1}\|^2 \|b\|^2 \lesssim \sum_{j=1}^N b_j^2.$$

Now we estimate the term $b_j, j \in \mathbb{Z}_N$. Noticing that $v - I_h v$ is continuous across the point $x_{j+\frac{1}{2}}$, we have

$$b_j = h(\partial_x v - \widehat{\partial_x (I_h v)})|_{j+\frac{1}{2}} = h\left(\{\partial_x (v - I_h v)\}|_{j+\frac{1}{2}} + \beta_1 h [\partial_x^2 (v - I_h v)]|_{j+\frac{1}{2}}\right).$$

Plugging (3.18) into the above identity and using the scaling from $[-1, 1]$ to $I_j, j \in \mathbb{Z}_N$ yields

$$\begin{aligned} b_j &= \sum_{m=k+1}^{\infty} (v_{j,m}(L_{m-1}(1) - 4\beta_1 L'_{m-1}(1)) + v_{j+1,m}(L_{m-1}(-1) + 4\beta_1 L'_{m-1}(-1))) \\ &= \sum_{m=k+1}^{\infty} (v_{j,m} + (-1)^{m-1} v_{j+1,m})(1 - 2\beta_1 m(m-1)), \end{aligned}$$

where the coefficient $v_{j,m}$ is the same as in (3.15), and in the last step, we have used the following identity

$$L_{m-1}(\pm 1) = (\pm 1)^{m-1}, \quad L'_{m-1}(\pm 1) = \frac{1}{2}(\pm 1)^m m(m-1).$$

Consequently, if $\beta_1 = \frac{1}{2k(k+1)}$, we have

$$b_j = \sum_{m=k+2}^{\infty} (v_{j,m} + (-1)^{m-1} v_{j+1,m})(1 - 2\beta_1 m(m-1)),$$

which yields, together with (3.16),

$$|b_j| \lesssim h^{k+2-\frac{1}{2}} (\|v\|_{k+2,I_j} + \|v\|_{k+2,I_{j+1}}). \tag{3.16}$$

Here we use the notation $I_{N+1} = I_1$. Note that

$$\|P_h v - I_h v\|_0^2 \lesssim h \sum_{j=1}^N \sum_{m=k-1}^k \bar{v}_{j,m}^2 \lesssim h \sum_{j=1}^N b_j^2.$$

Then the desired result (3.20) follows. ■

We end this subsection with an integral projection D_x^{-1}, which is defined as, for any function v,

$$D_x^{-1} v|_{I_j} := \int_{x_{j-\frac{1}{2}}}^{x} v dx.$$

Apparently,

$$\|D_x^{-1} v\|_0 \lesssim h\|v\|_0. \tag{3.15}$$

Moreover, noticing that

$$D_x^{-1}(v - I_h v)(x) = \sum_{m=k+1}^{\infty} v_{j,m} \int_{x_{j-\frac{1}{2}}}^{x} \phi_{j,m} dx = \sum_{m=k+1}^{\infty} \frac{v_{j,m}}{2m-1} \int_{x_{j-\frac{1}{2}}}^{x} (L_{j,m} - L_{j,m-2}) dx, \quad x \in I_j,$$

we have all for $k \geq 2$,

$$D_x^{-1}(v - I_h v)\big(x_{j+\frac{1}{2}}^-\big) = D_x^{-1}(v - I_h v)\big(x_{j-\frac{1}{2}}^+\big) = 0, \quad \forall j \in \mathbb{Z}_N. \tag{3.16}$$

Similarly, since

$$(v - P_h v) \perp \mathbb{P}^0, \quad k \geq 2,$$

we have

$$(v - P_h v)|_{I_j} = \sum_{m=1}^{\infty} \tilde{v}_{j,m} L_{j,m}(x), \quad \tilde{v}_{j,m} = \frac{2m+1}{h}(v - P_h v, L_{j,m})_j.$$

Then

$$D_x^{-1}(v - P_h v)(x_{j+\frac{1}{2}}^-) = D_x^{-1}(v - P_h v)(x_{j-\frac{1}{2}}^+) = 0, \quad \forall j \in \mathbb{Z}_N, \quad k \geq 2. \tag{3.17}$$

The identities (3.16) and (3.17) indicate that both $D_x^{-1}(v - I_h v)$ and $D_x^{-1}(v - P_h v)$ are continuous on the whole domain Ω when $k \geq 2$.

B. Superconvergence at Gauss and Lobatto Points

To study the superconvergence properties of the DDG solution u_h at Gauss and Lobatto points, we first analyze the supercloseness of u_h toward the Gauss–Lobatto projection $I_h u$ of the exact solution u, and then use the supercloseness result between u_h and $I_h u$ to obtain the superconvergence at Gauss and Lobatto points. Our analysis is along this line.

Define

$$H_h^1 = \left\{ v : v|_{I_j} \in H^1(I_j), j \in \mathbb{Z}_N \right\},$$

and for all $\varphi, v \in H_h^1$, let

$$a(\varphi, v) = A(\varphi, v) + (\varphi, v)$$

and

$$|||v|||_E = \|v\|_E + \|v\|_0,$$

where $\|\cdot\|_E$ is defined in (2.6). For any $v_h \in V_h$, the inverse inequality holds. Then

$$\{\partial_x v_h\}^2|_{j+\frac{1}{2}} \lesssim \|\partial_x v_h\|^2_{0,\infty,I_j} + \|\partial_x v_h\|^2_{0,\infty,I_{j+1}} \lesssim h^{-1}(\|\partial_x v_h\|^2_{0,I_j} + \|\partial_x v_h\|^2_{0,I_{j+1}}),\ j \in \mathbb{Z}_N.$$

Similarly, we get

$$h[\partial_x^2 v_h]^2|_{j+\frac{1}{2}} \lesssim h\|\partial_x^2 v_h\|^2_{0,\infty,I_j} + h\|\partial_x^2 v_h\|^2_{0,\infty,I_{j+1}} \lesssim h^{-1}(\|\partial_x v_h\|^2_{0,I_j} + \|\partial_x v_h\|^2_{0,I_{j+1}}).$$

Consequently,

$$h\sum_{j=1}^{N}(\{\partial_x v_h\} + \beta_1 h[\partial_x^2 v_h])^2|_{j+\frac{1}{2}} \lesssim \sum_{j=1}^{N}\int_{I_j}|\partial_x v_h|^2 dx.$$

Recalling the definition of $A(\cdot,\cdot)$ in (2.4), we have for all $\varphi_h, v_h \in V_h$,

$$\begin{aligned} A(\varphi_h, v_h) &\lesssim \|\varphi_h\|_E\left(\sum_{j=1}^{N}\int_{I_j}|\partial_x v_h|^2 dx + h\sum_{j=1}^{N}\{\partial_x v_h\}^2|_{j+\frac{1}{2}} + h^{-1}\sum_{j=1}^{N}[v_h]^2|_{j+\frac{1}{2}}\right)^{\frac{1}{2}} \\ &\lesssim \|\varphi_h\|_E\|v_h\|_E. \end{aligned}$$

Then

$$a(v_h, v_h) \geq \gamma|||v_h|||_E^2, \quad a(\varphi_h, v_h) \lesssim |||\varphi_h|||_E|||v_h|||_E, \quad \forall \varphi_h, v_h \in V_h. \tag{3.18}$$

By the Lax–Milgram lemma, there exists a unique function $w_u \in V_h$ such that

$$a(w_u, v_h) = -(\partial_t(u - P_h u), v_h), \quad \forall v_h \in V_h. \tag{3.19}$$

Choosing $v_h = w_u$ in the above equation and using the coercivity of $a(\cdot,\cdot)$, we obtain for any $k \geq 2$,

$$\begin{aligned} \gamma|||w_u|||_E^2 &\leq |(\partial_t(u - P_h u), w_u)| = |(D_x^{-1}\partial_t(u - P_h u), \partial_x w_u)| \\ &\lesssim h^{k+2}\|\partial_t u\|_{k+1}|||w_u|||_E, \end{aligned}$$

where we have used the integration by parts and (3.17) in the second step, and (3.4), (3.15) in the last step. Noticing $u_t = u_{xx}$, then

$$|||w_u|||_E \lesssim h^{k+2}\|u_t\|_{k+1} \lesssim h^{k+2}\|u\|_{k+3}. \tag{3.20}$$

Similarly, taking time derivative on both sides of (3.19) and choosing $v = \partial_t w_u$ yields

$$|||\partial_t w_u|||_E \lesssim h^{k+2}\|u_{tt}\|_{k+1} \lesssim h^{k+2}\|u\|_{k+5}. \tag{3.21}$$

We have the following superconvergence results for $u_h - I_h u$ and $u_h - P_h u$.

Theorem 3.3. *Let $u \in H^{k+5}$ and u_h be the solution of (1.1) and (2.1), respectively. Suppose the initial value $u_h(x,0)$ is chosen such that*

$$\|u_h(\cdot,0) - I_h u_0\|_0 \lesssim h^{k+2}\|u_0\|_{k+3}. \tag{3.22}$$

Then there holds for all $k \geq 2$ and $t > 0$,

$$\|(u_h - P_h u)(\cdot,t)\|_0 + \|(u_h - I_h u)(\cdot,t)\|_0 \lesssim h^{k+2} \sup_{\tau\in[0,t]} \|u(\cdot,\tau)\|_{k+5}. \tag{3.23}$$

Denote

$$e = u - u_h, \quad \xi = u_h - P_h u + w_u, \quad \eta = u - P_h u + w_u,$$

where w_u is the solution of (3.19). Noticing that the exact solution u also satisfies (2.3), we obtain, by choosing $v_h = \xi$ in (2.7) and using the orthogonality $(\partial_t e, v_h) + A(e, v_h) = 0, v_h \in V_h$,

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|\xi\|_0^2 &\leq |(\partial_t \eta, \xi) + A(\eta,\xi)|\\ &= |(\partial_t(u - P_h u), \xi) + (\partial_t w_u, \xi) + A(w_u,\xi) + A(u - P_h u, \xi)|\\ &= |(\partial_t w_u, \xi) - (w_u, \xi) + A(u - P_h u, \xi)|.\end{aligned}$$

Here in the last step, we have used the identity (3.19). By integration by parts, there holds for all $\varphi \in H_h^1$ and $v_h \in V_h$

$$\begin{aligned}A(\varphi, v_h) &= -(\varphi, \partial_x^2 v_h) + \sum_{j=1}^{N} (-[\varphi\partial_x v_h] + \widehat{\partial_x\varphi}[v_h] + \{\partial_x v_h\}[\varphi])|_{j+\frac{1}{2}}\\ &= -(\varphi, \partial_x^2 v_h) + \sum_{j=1}^{N} (\widehat{\partial_x\varphi}[v_h] - \{\varphi\}[\partial_x v_h])|_{j+\frac{1}{2}}.\end{aligned} \tag{3.24}$$

By the properties of P_h, we obtain

$$A(u - P_h u, v_h) = 0, \quad \forall v_h \in V_h. \tag{3.25}$$

Consequently,

$$\frac{1}{2}\frac{d}{dt}\|\xi\|_0^2 \leq |(\partial_t w_u, \xi) - (w_u,\xi)| \lesssim (\|\partial_t w_u\|_0 + \|w_u\|_0)\|\xi\|_0.$$

In light of (3.20)–(3.21), we have

$$\frac{d}{dt}\|\xi\|_0 \lesssim h^{k+2}\|u\|_{k+5},$$

which yields

$$\|\xi(\cdot,t)\|_0 \lesssim \|\xi(\cdot,0)\|_0 + h^{k+2}\int_0^t ||u(\cdot,\tau)\|_{k+5} d\tau \lesssim \|\xi(\cdot,0)\|_0 + Th^{k+2}\sup_{\tau\in[0,t]}\|u(\cdot,\tau)\|_{k+5}.$$

Due to the special choice of the initial solution, (3.12) and (3.20), we get

$$\|\xi(\cdot,0)\|_0 \lesssim h^{k+2}\|u_0\|_{k+2} + \|w_u(\cdot,0)\|_0 \lesssim h^{k+2}\|u_0\|_{k+3},$$

and thus,

$$\|\xi(\cdot,t)\|_0 \lesssim h^{k+2}\sup_{\tau\in[0,t]}\|u(\cdot,\tau)\|_{k+5}. \tag{3.26}$$

Then the desired result (3.23) follows from (3.12) and (3.20).

Remark 3.4. The result of Theorem 3.3 indicates that the DDG solution u_h is superclose with an order k + 2 to the global projection $P_h u$ and the Gauss–Lobatto projection $I_h u$ of the exact solution.

We denote by $g_{j,m}, (j,m) \in \mathbb{Z}_N \times \mathbb{Z}_k$ and $l_{j,m}, (j,m) \in \mathbb{Z}_N \times \mathbb{Z}_{k+1}$ the Gauss points of degree k and Lobatto points of degree $k+1$ on the interval I_j, respectively. That is, $g_{j,m}$ and $l_{j,m}$ are separately zeros of $L_{j,k}$ and $\phi_{j,k+1}$. As a direct consequence of Theorem 3.3, we have the following superconvergence results of the derivative approximation at Gauss points and the function value approximation at Lobatto points.

Corollary 3.5. *Suppose all the conditions of Theorem 3.3 hold. Then for any fixed* $t \in (0,T]$,

$$e_g = \left(\frac{1}{Nk}\sum_{j=1}^{N}\sum_{i=1}^{k}\partial_x(u-u_h)^2(g_{j,i},t)\right)^{\frac{1}{2}} \lesssim h^{k+1}\sup_{\tau\in[0,t]}\|u(\cdot,\tau)\|_{k+5}, \tag{3.27}$$

$$e_l = \left(\frac{1}{N(k+1)}\sum_{j=1}^{N}\sum_{i=1}^{k+1}(u-u_h)^2(l_{j,i},t)\right)^{\frac{1}{2}} \lesssim h^{k+2}\sup_{\tau\in[0,t]}\|u(\cdot,\tau)\|_{k+5}. \tag{3.28}$$

Proof. In each element I_j, $j \in \mathbb{Z}_N$, note that

$$(v - I_h v)(l_{j,i}) = \sum_{m=k+2}^{\infty} v_{j,m}\phi_{j,m}(l_{j,i}), \quad i \in \mathbb{Z}_{k+1},$$

$$(v - I_h v)_x(g_{j,i}) = \frac{2}{h}\sum_{m=k+2}^{\infty} v_{j,m}L_{j,m-1}(g_{j,i}), \quad i \in \mathbb{Z}_k,$$

where $v_{j,m}$ is defined by (3.7). Therefore, for any $v \in H^{k+2}$, we have from (3.8)

$$(v - I_h v)(l_{i,j}) \lesssim h^{k+2-\frac{1}{2}}\|v\|_{k+2,I_j}, \quad \partial_x(v - I_h v)(g_{i,j}) \lesssim h^{k+1-\frac{1}{2}}\|v\|_{k+2,I_j},$$

which yields

$$\left(\frac{1}{N(k+1)}\sum_{j=1}^{N}\sum_{i=1}^{k+1}(v - I_h v)^2(l_{j,i})\right)^{\frac{1}{2}} \lesssim h^{k+2}\|v\|_{k+2},$$

$$\left(\frac{1}{Nk}\sum_{j=1}^{N}\sum_{i=1}^{k}\partial_x(v-I_hv)^2(g_{j,i})\right)^{\frac{1}{2}}\lesssim h^{k+1}\|v\|_{k+2}.$$

On the other hand, since $I_hu-u_h\in V_h$, we have from the inverse inequality

$$\left(\frac{1}{N(k+1)}\sum_{j=1}^{N}\sum_{i=1}^{k+1}(I_hu-u_h)^2(l_{j,i})\right)^{\frac{1}{2}}\lesssim\|I_hu-u_h\|_0,$$

$$\left(\frac{1}{Nk}\sum_{j=1}^{N}\sum_{i=1}^{k}\partial_x(I_hu-u_h)^2(g_{j,i})\right)^{\frac{1}{2}}\lesssim h^{-1}\|I_hu-u_h\|_0.$$

Then the desired results (3.27)–(3.28) follow from (3.23) and the triangle inequality. ■

C. Superconvergence at Nodes

As a direct consequence of (3.23), we have the following superconvergence result at nodes.

$$\left(\frac{1}{N}\sum_{j=1}^{N}\{u-u_h\}^2|_{j+\frac{1}{2}}\right)^{\frac{1}{2}}=\left(\frac{1}{N}\sum_{j=1}^{N}\{I_hu-u_h\}^2|_{j+\frac{1}{2}}\right)^{\frac{1}{2}}$$
$$\lesssim\|I_hu-u_h\|_0\lesssim h^{k+2}\sup_{\tau\in[0,t]}\|u(\cdot,\tau)\|_{k+5}.$$

However, our numerical results show that the superconvergence rate at nodes can reach $2k$ when $k\geq 3$, which is better than the order $k+2$ we provide above. Therefore, new analysis tools are needed to prove the $2k$-th superconvergence order for $k\geq 3$.

We will adopt the idea of correction function in our superconvergence analysis. That is, we construct a specially designed function $w\in V_h$ such that w vanishes or is of high order at nodes (to guarantee that the error is small enough to be compatible with the superconvergence error estimate at nodes) and

$$\|u_h-u_I\|_0=\|u_h-P_hu+w\|_0\leq h^{2k}c(u),$$

where $c(u)$ is a constant dependent on the exact solution u (our later superconvergence analysis shows that $c(u)$ can be bounded by $\sup_{\tau\in[0,T]}\|u(\cdot,\tau)\|_{2k+2}$). In light of (2.7) and (3.25),

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|u_h-u_I\|_0^2+\gamma\|u_h-u_I\|_E^2&\leq(\partial_t(u_h-u_I),u_h-u_I)+A(u_h-u_I,u_h-u_I)\\&=(\partial_t(u-u_I),u_h-u_I)+A(u-u_I,u_h-u_I)\\&=(\partial_t(u-u_I),u_h-u_I)+A(w,u_h-u_I).\end{aligned}\tag{3.29}$$

Therefore, to achieve our superconvergence goal, we need to construct a function $w\in V_h$ such that the right hand side in (3.29) is of high order. To be more precise, we shall construct a function $w\in V_h$ satisfying

$$(\partial_t(u-u_I),u_h-u_I)+A(w,u_h-u_I)\leq h^{2k}c_1(u)(\|u_h-u_I\|_E+\|u_h-u_I\|_0).$$

Here again, $c_1(u)$ is dependent on some norm of the exact solution u.

To design w, we first denote $w_0 = u - P_h u$ and define a serial of functions $w_i, 1 \le i \le \lfloor (k-1)/2 \rfloor$ as follows.

$$(w_i, \partial_x^2 v_h) = (\partial_t w_{i-1}, v_h), \quad \forall v_h \in \mathbb{P}^k(I_j) \backslash \mathbb{P}^1(I_j), \tag{3.30}$$

$$\widehat{\partial_x w_i} := \beta_0 h^{-1}[w_i] + \{\partial_x w_i\} + \beta_1 h[\partial_x^2 w_i]|_{j+\frac{1}{2}} = 0, \tag{3.31}$$

$$\{w_i\}|_{j+\frac{1}{2}} = 0. \tag{3.32}$$

Here $\lfloor k \rfloor$ denotes the maximal integer no more than k.

Due to the existence of the global projection P_h, the function $w_i, 1 \le i \le \lfloor (k-1)/2 \rfloor$ defined in (3.30)–(3.32) is uniquely determined. Moreover, we have the following estimate for $w_i, 1 \le i \le \lfloor (k-1)/2 \rfloor$.

Lemma 3.6. *For any $k \ge 3$, suppose $w_i, 1 \le i \le \lfloor (k-1)/2 \rfloor$ is defined by (3.30)–(3.32). Then*

$$w_i|_{I_j} := \sum_{m=k-2i-1}^{k} c_{j,m}^i L_{j,m}, \tag{3.33}$$

where $L_{j,-1} = 0$ and $c_{j,m}^i$ are some bounded constants. Furthermore, if $\partial_t^r u \in H^{k+1+2i}, r = 0, 1,\ 1 \le i \le \lfloor (k-1)/2 \rfloor$, there holds

$$\|\partial_t^r w_i\|_0 \lesssim h^{k+1+2i} \|u\|_{k+1+2(i+r)}. \tag{3.34}$$

Proof. We will show (3.33) by induction. First, we suppose in each element I_j

$$w_i|_{I_j} := \sum_{m=0}^{k} c_{j,m}^i L_{j,m}.$$

For any $v_h \in \mathbb{P}^k \backslash \mathbb{P}^1$, since $\partial_x^2 v_h \in \mathbb{P}^{k-2}$, by choosing $\partial_x^2 v_h = L_{j,m}, m \in \mathbb{Z}_{k-2}$ in (3.30), we immediately obtain

$$c_{j,m}^i = \frac{2m+1}{h} \int_{I_j} \partial_t w_{i-1}(D_x^{-1} D_x^{-1} L_{j,m}) dx, \quad m \le k-2. \tag{3.35}$$

Noticing that $D_x^{-1} D_x^{-1} L_{j,m} \in \mathbb{P}^{m+2}(I_j)$ and $(u - P_h u) \perp \mathbb{P}^{k-2}$, then

$$c_{j,m}^1 = \frac{2m+1}{h} \int_{I_j} \partial_t (u - P_h u)(D_x^{-1} D_x^{-1} L_{j,m}) dx = 0, \quad \forall m \le k-4,$$

which indicates that (3.33) is valid for $i = 1$. Suppose (3.33) holds for all $i, 1 \le i \le \lfloor (k-1)/2 \rfloor - 1$. Then

$$w_i \perp \mathbb{P}^{k-2i-2}.$$

In light of (3.35), we have

$$c_{j,m}^{i+1} = \frac{2m+1}{h} \int_{I_j} \partial_t w_i (D_x^{-1} D_x^{-1} L_{j,m}) dx = 0, \quad \forall m < k-2i-3.$$

Consequently, (3.33) is also valid for $i+1$. Then the proof of (3.33) is complete.

We next estimate the coefficients $c^i_{j,m}$. By (3.35) and (3.15), we have

$$|c^i_{j,m}| \le h^{-1}\|\partial_t w_{i-1}\|_{0,I_j}\|D_x^{-1}D_x^{-1}L_{j,m}\|_{0,I_j} \lesssim h^{\frac{3}{2}}\|\partial_t w_{i-1}\|_{0,I_j}, \quad m \le k-2.$$

To estimate $c^i_{j,m}, m = k-1, k$, we obtain from (3.31)–(3.32)

$$\sum_{m=k-1}^{k}(L_m(1)c^i_{j,m} + L_m(-1)c^i_{j+1,m}) = -\sum_{m=k-2i-1}^{k-2}(L_m(1)c^i_{j,m} + L_m(-1)c^i_{j+1,m}),$$

$$\sum_{m=k-1}^{k}(g_0(m)c^i_{j,m} + g_1(m)c^i_{j+1,m}) = -\sum_{m=k-2i-1}^{k-2}(g_0(m)c^i_{j,m} + g_1(m)c^i_{j+1,m}),$$

where g_0, g_1 are the same as in (3.13). By the same argument as in Lemma 3.2, we have

$$\sum_{j=1}^{N}\sum_{m=k-1}^{k}(c^i_{j,m})^2 \lesssim \sum_{j=1}^{N}\sum_{m=k-2i-1}^{k-2}(c^i_{j,m})^2 \lesssim h^3\|\partial_t w_{i-1}\|_0^2.$$

Consequently,

$$\|w_i\|_0 \simeq \left(\sum_{j=1}^{N}\sum_{m=k-2i-1}^{k} h(c^i_{j,m})^2\right)^{\frac{1}{2}} \lesssim h^2\|\partial_t w_{i-1}\|_0, \quad 1 \le i \le \lfloor (k-1)/2 \rfloor .$$

Taking time derivative on both sides of (3.30)–(3.32), the three identities still hold. In other words, we can replace w_i by $\partial_t w_i$ in (3.30)–(3.32). Similarly, (3.30)–(3.32) are still valid if we take $\partial_t^m w_i$ instead of w_i, where $m \ge 1$. Then following the same arguments as what we did for w_i, we get

$$\|\partial_t^m w_i\|_0 \lesssim h^2\|\partial_t^{m+1} w_{i-1}\|_0, \quad \forall m \ge 1, 1 \le i \le \lfloor (k-1)/2 \rfloor .$$

By recursion formula, there holds for all $1 \le i \le \lfloor (k-1)/2 \rfloor$ and $r = 0, 1$

$$\|\partial_t^r w_i\|_0 \lesssim h^{2i}\|\partial_t^{r+i} w_0\|_0.$$

In light of (3.4), we have

$$\|\partial_t^m w_0\|_0 = \|\partial_t^m(u - P_h u)\|_0 \lesssim h^{k+1}\|\partial_t^m u\|_{k+1} \lesssim h^{k+1}\|u\|_{k+1+2m}, \quad m \ge 1.$$

Then (3.34) follows. This finishes our proof. ∎

Now we define the special correction function w as

$$w = \sum_{i=1}^{\lfloor (k-1)/2 \rfloor} w_i. \tag{3.36}$$

Theorem 3.7. *Let u and u_h be the solution of (1.1) and (2.1), respectively, and $u_I = P_h u - w$ be the special interpolation function with w defined by (3.36), (3.30)–(3.32). For all $k \ge 3$, if $u \in H^{2k+2}$, then*

$$\|(u_h - u_I)(\cdot, t)\|_0 \lesssim \|(u_h - u_I)(\cdot, 0)\|_0 + h^{2k}\sup_{\tau \in [0,t]}\|u(\cdot, \tau)\|_{2k+2}, \quad \forall t > 0. \tag{3.37}$$

Proof. By (3.33) and the orthogonality of P_{hu}, we have

$$w_i \perp \mathbb{P}^1, \quad \forall 0 \le i \le \lfloor (k-1)/2 \rfloor - 1.$$

Then for any function $v_h \in \mathbb{P}^1$,

$$(\partial_t w_i, v_h) = 0 = (w_{i+1}, \partial_x^2 v_h), \quad \forall 0 \le i \le \lfloor (k-1)/2 \rfloor - 1,$$

which gives, together with (3.30)

$$(w_{i+1}, \partial_x^2 v_h) = (\partial_t w_i, v_h), \quad \forall v_h \in V_h,\ 1 \le i \le \lfloor (k-1)/2 \rfloor - 1.$$

Consequently, for any $v_h \in V_h$, we have from (3.24) and (3.31)–(3.32),

$$\begin{aligned}
&(\partial_t(u - P_h u), v_h) + A(w, v_h) + (\partial_t w, v_h) = (\partial_t w_r, v_h), \quad \text{if } k = 2r+1,\\
&(\partial_t(u - P_h u), v_h) + A(w, v_h) + (\partial_t w, v_h) = (\partial_t w_{r-1}, v_h)\\
&\qquad = -(D_x^{-1}\partial_t w_{r-1}, \partial_x v_h), \quad \text{if } k = 2r.
\end{aligned}$$

Here in the last step for the even $k = 2r$, we have used integration by parts, and the fact that $w_{r-1} \perp \mathbb{P}^1$, which yields

$$D_x^{-1} w_{r-1}(x^-_{j+\frac{1}{2}}) = D_x^{-1} w_{r-1}(x^+_{j-\frac{1}{2}}) = 0.$$

In light of (3.15) and (3.34), we have for all $k \ge 3$

$$\begin{aligned}
|(\partial_t(u - P_h u), v_h) + A(w, v_h) + (\partial_t w, v_h)| &\le C_1 h^{2k} \|u\|_{2k+2}(\|v_h\|_0 + \|v_h\|_E)\\
&\le C_2 h^{4k}\|u\|^2_{2k+2} + C_3\|v_h\|_0^2 + \frac{\gamma}{2}\|v_h\|_E^2,
\end{aligned}$$

where $C_i, i \le 3$ are some bounded constants independent of the mesh size h. Plugging the above inequality into (3.29) yields

$$\frac{d}{dt}\|u_I - u_h\|_0^2 \lesssim h^{4k}\|u\|^2_{2k+2} + \|u_I - u_h\|_0^2,$$

and thus,

$$\|(u_I - u_h)(\cdot,t)\|_0^2 \lesssim \|(u_I - u_h)(\cdot,0)\|_0^2 + h^{4k}\sup_{\tau\in[0,t]}\|u(\cdot,\tau)\|^2_{2k+2} + \int_0^t ||(u_I - u_h)(\cdot,\tau)\|_0^2 d\tau.$$

By the Gronwall inequality (see, e.g., [33], p.9), we have

$$\|(u_I - u_h)(\cdot,t)\|_0^2 \lesssim (1 + e^t)\left(\|(u_I - u_h)(\cdot,0)\|_0^2 + h^{4k}\sup_{\tau\in[0,t]}\|u(\cdot,\tau)\|^2_{2k+2}\right).$$

Then the desired result (3.37) follows. ∎

Theorem 3.8. *Let $u \in H^{2k+2}$ be the solution of (1.1), and u_h be the solution of (2.1). Suppose $u_I = P_h u - w$ with w defined by (3.36), (3.30)–(3.32), and the initial value $u_h(x,0)$ is taken such that*

$$\|(u_h - u_I)(\cdot,0)\|_0 \lesssim h^{2k}\|u_0\|_{2k+2}. \tag{3.38}$$

Then there holds for all $k \geq 3$ and $t \in (0,T]$

$$e_n = \left(\frac{1}{N}\sum_{j=1}^{N}(u - \{u_h\})^2(x_{j+\frac{1}{2}},t)\right)^{\frac{1}{2}} \lesssim h^{2k}\sup_{\tau\in[0,t]}\|u(\cdot,\tau)\|_{2k+2}. \tag{3.39}$$

Proof. First, for any fixed $t \in (0,T]$, by the special choice of the initial value and (3.37), we have

$$\|(u_I - u_h)(\cdot,t)\|_0 \lesssim h^{2k}\sup_{\tau\in[0,t]}\|u(\cdot,\tau)\|_{2k+2}.$$

On the other hand, it is easy to obtain from (3.32) that $\{w(\cdot,t)\}\,|_{j+\frac{1}{2}} = 0$. Then

$$u(x_{j+\frac{1}{2}},t) = \{u(\cdot,t)\}\,|_{j+\frac{1}{2}} = \{(P_h u - w)(\cdot,t)\}\,|_{j+\frac{1}{2}} = \{u_I(\cdot,t)\}\,|_{j+\frac{1}{2}}.$$

Since $u_h - u_I \in V_h$, the inverse inequality holds, and thus

$$\left(\frac{1}{N}\sum_{j=1}^{N}(\{u_I\} - \{u_h\})^2(x_{j+\frac{1}{2}},t)\right)^{\frac{1}{2}} \lesssim \left(\frac{1}{N}\sum_{j=1}^{N}||(u_I - u_h)(\cdot,t)\|^2_{0,\infty,I_j}\right)^{\frac{1}{2}} \lesssim \|(u_I - u_h)(\cdot,t)\|_0.$$

Then

$$e_n \lesssim \|(u_I - u_h)(\cdot,t)\|_0 \lesssim h^{2k}\sup_{\tau\in[0,t]}\|u(\cdot,\tau)\|_{2k+2}.$$

This finishes our proof. ∎

Remark 3.9. We would like to point out that, based on the superconvergence results in Theorem 3.3, Corollary 3.5 and Theorem 3.8, the choice of β_1 in (2.1) has influence on the superconvergence of the DDG solution u_h at Gauss and Lobatto points, as well as the supercloseness between u_h and $I_h u$. However, it does not affect the superconvergence rate of u_h at nodes. We will demonstrate these points in our numerical experiments.

Remark 3.10. In our superconvergence analysis in Corollary 3.5 and Theorem 3.8, the initial discretizations (3.22) and (3.38) are of great significance. To obtain the k + 2-th order superconvergence rate at Gauss and Lobatto points, the initial value $u_h(x,0) = I_h u_0$ or $P_h u_0$ is a valid choice. However, to achieve the 2k goal at nodal points, it is indicated from Theorem 3.8 that the initial error should also reach the same superconvergence rate, which imposes a stronger condition on the initial discretization. A natural way of initial discretization to obtain (3.46) is to choose $u_h(x,0) = u_I(x,0)$, or $u_h(x,0) = \tilde{u}_I(x,0)$, where

$$\tilde{u}_I = \begin{cases} u_I & \text{if } k = 2r \\ u_I - w_r & \text{if } k = 2r+1 \end{cases} \tag{3.40}$$

The latter is a valid choice due to the fact that

$$\|w_r(\cdot,0)\|_0 \lesssim h^{2k}\|u_0\|_{2k}, \quad \text{when}\, k = 2r+1.$$

D. Initial Discretization

In this subsection, we would like to demonstrate how to implement the initial discretization as it plays an important role in our superconvergence analysis. If we take the initial data $u_h(x,0) = I_h u_0$ or $u_h(x,0) = P_h u_0$, then we can obtain $u_h(x,0)$ by only using the information of the initial value u_0 and (3.1)-(3.3) and (3.9). While if the initial data is chosen as $u_h(x,0) = u_I(x,0)$ or $u_h(x,0) = \tilde{u}_I(x,0)$, the initial discretization is somewhat complicated. We next show how to calculate $u_I(x,0)$ only using the information of u_0. The same technique can be applied to $\tilde{u}_I$. Noticing that

$$\partial_t^r u|_{t=0} = \partial_x^{2r} u_0, \quad 0 \le r \le \lfloor (k-1)/2 \rfloor,$$

then

$$\partial_t^r w_0|_{t=0} = \partial_t^r (u - P_h u)|_{t=0} = \partial_x^{2r} u_0 - P_h(\partial_x^{2r} u_0).$$

On the other hand, taking time derivative on both sides of (3.30)–(3.32), we can calculate the term $\partial_t^r w_{i+1}|_{t=0}$ from the information of $\partial_t^{r+1} w_i|_{t=0}$. Now we divide the initial discretization into the following steps:

1. Compute $P_h(\partial_x^{2r} u_0), 0 \le r \le \lfloor (k-1)/2 \rfloor$ by (3.1)–(3.3) and set $\partial_t^r w_0 = \partial_x^{2r} u_0 - P_h(\partial_x^{2r} u_0)$.
2. Calculate $\partial_t^r w_{i+1}, i \le \lfloor (k-1)/2 \rfloor, 0 \le r \le k-i$ by (3.30)–(3.32) with w_{i+1} replaced by $\partial_t^r w_{i+1}$ and $\partial_t w_i$ replaced by $\partial_t^{r+1} w_i$.
3. Figure out $u_h(x,0) = P_h u_0 - \sum_{i=1}^{\lfloor (k-1)/2 \rfloor} w_i$.

IV. SUPERCONVERGENCE FOR LINEAR FLUX $f = \alpha U$

The superconvergence analysis for the convection-diffusion equation with $f = \alpha u$ is similar to that for the zero flux $f = 0$. To simplify our proof, we only consider the case in which α is a constant in the rest of this section.

We first introduce another global projection Q_h. For a given smooth function v, the projection $Q_h v \in V_h$ is defined by

$$(Q_h v, \varphi_h)_j = (v, \varphi_h)_j, \quad \forall \varphi_h \in \mathbb{P}^{k-2}(I_j), j \in \mathbb{Z}_N, \tag{4.1}$$

$$(\widehat{\partial_x (Q_h v)} - \hat{f}(Q_h v))|_{j+\frac{1}{2}} = \partial_x v(x_{j+\frac{1}{2}}) - (\alpha v)(x_{j+\frac{1}{2}}), \tag{4.2}$$

$$\{Q_h v\}|_{j+\frac{1}{2}} = v(x_{j+\frac{1}{2}}). \tag{4.3}$$

The uniqueness and existence of the projection Q_h have been proved in [19]. Moreover, there holds

$$\|u - Q_h u\|_0 \lesssim h^n \|u\|_n, \quad 0 \le n \le k+1. \tag{4.4}$$

We have the following superconvergence result of $Q_h v - I_h v$.

Lemma 4.1. *For any function $v \in H^{k+2}$, if $\beta_1 = \frac{1}{2k(k+1)}$, then*

$$\|Q_h v - I_h v\|_0 \lesssim h^{k+2}\|v\|_{k+2}. \tag{4.5}$$

Here we omit the proof since it is similar to that for zero flux, the only difference lies in a modification of g_0

$$\tilde{g}_0(m) = g_0(m) - \alpha h,$$

where g_0 is given in (3.13).

By the definition of Q_h, we have

$$A(u - Q_h u, v_h) - F(u - Q_h u, v_h) = -(\alpha(u - Q_h u), \partial_x v_h), \quad \forall v_h \in V_h. \tag{4.6}$$

We modify the correction function $w_i \in V_h$ as follows. For all $k \geq 2$, $w_i, 1 \leq i \leq k-1$ is the function satisfying

$$(w_i, \partial_x^2 v_h)_j = (\partial_t w_{i-1}, v_h)_j - (\alpha w_{i-1}, \partial_x v_h)_j, \quad \forall v_h \in \mathbb{P}^k(I_j)\backslash\mathbb{P}^1(I_j), \tag{4.7}$$

$$(\widehat{\partial_x w_i} - \hat{f}(w_i))|_{j+\frac{1}{2}} = 0, \tag{4.8}$$

$$\{w_i\}|_{j+\frac{1}{2}} = 0, \quad j \in \mathbb{Z}_N. \tag{4.9}$$

Here $w_0 = u - Q_h u$. Note that the definition of w_i is the same as that of the global projection Q_h, except the right hand side. Therefore, $w_i, 1 \leq i \leq k-1$ defined in (4.7)–(4.9) is uniquely determined.

Lemma 4.2. *Suppose $w_i \in V_h, i \in \mathbb{Z}_{k-1}$ is defined by (4.7)–(4.9), and in each element $I_j, j \in \mathbb{Z}_N$,*

$$w_i|_{I_j} := \sum_{m=0}^{k} c_{j,m}^i L_{j,m}.$$

Then there holds for any positive integer r

$$\left(h\sum_{j=1}^{N}(\partial_t^r c_{j,m}^i)^2\right)^{\frac{1}{2}} \lesssim h^l\|\partial_t^r u\|_l, \quad 0 \leq m \leq k, \tag{4.10}$$

where $0 \leq l \leq \max(2k - m, k + 1 + i)$. Consequently,

$$\|\partial_t^r w_i\|_0 \lesssim h^{k+1+i}\|\partial_t^r u\|_{k+1+i}. \tag{4.11}$$

Proof. Since (4.11) is a direct consequence of (4.10), we only prove (4.10) in the following. We will show (4.10) by induction. First, noticing that (4.7)–(4.9) still hold by taking time derivative of r-th order, then we choose $\partial_x^2 v_h = L_{j,m}, m \leq k-2$ in (4.7) to derive

$$\frac{h}{2m+1}\partial_t^r c_{j,m}^{i+1} = (\partial_t^{1+r} w_i, D_x^{-1}D_x^{-1}L_{j,m})_j - (\alpha(\partial_t^r w_i), D_x^{-1}L_{j,m})_j, \quad m \leq k-2. \tag{4.12}$$

Since $w_0 = (u - Q_h u) \perp \mathbb{P}^{k-2}$ and $D_x^{-1}D_x^{-1}L_{j,m} \in \mathbb{P}^{m+2}(I_j)$, $D_x^{-1}L_{j,m} \in \mathbb{P}^{m+1}(I_j)$, we have, by letting $i=0$ in (4.12)

$$\partial_t^r c_{j,m}^1 = 0, \quad \forall m = 0, \dots, k-4.$$

On the other hand, note that

$$\|D_x^{-1}D_x^{-1}L_{j,m}\|_{0,I_j} \lesssim h^{\frac{5}{2}}, \quad \|D_x^{-1}L_{j,m}\|_{0,I_j} \lesssim h^{\frac{3}{2}}, \quad m \geq 0.$$

By choosing $i = 0, m = k-2$ in (4.12) and using the estimate (4.4) and the fact $u_t = u_{xx} - \alpha u_x$, we obtain for all $0 \leq n \leq k+1$

$$\left(h\sum_{j=1}^{N}(\partial_t^r c_{j,k-2}^1)^2\right)^{\frac{1}{2}} \lesssim h^2\|\partial_t^{1+r}w_0\|_0 + h\|\partial_t^r w_0\|_0$$
$$\lesssim h^{n+1}(\|\partial_t^r u\|_n + \|\partial_t^{1+r}u\|_{n-1}) \lesssim h^{n+1}\|\partial_t^r u\|_{n+1}.$$

Moreover, if $k \geq 3$, we have $(\partial_t^r w_0, D_x^{-1}L_{j,k-3})_j = 0$, and thus,

$$|\partial_t^r c_{j,k-3}^1| = \frac{2m+1}{h}\left|\int_{I_j}\partial_t^{1+r}w_0(D_x^{-1}D_x^{-1}L_{j,k-3})dx\right| \lesssim h^{\frac{3}{2}}\|\partial_t^{1+r}(u - Q_h u)\|_{0,I_j},$$

which yields

$$\left(h\sum_{j=1}^{N}(\partial_t^r c_{j,k-3}^1)^2\right)^{\frac{1}{2}} \lesssim h^2\|\partial_t^{1+r}(u - Q_h u)\|_0 \lesssim h^{n+2}\|\partial_t^{1+r}u\|_n \lesssim h^{n+2}\|\partial_t^r u\|_{n+2}.$$

To estimate $\partial_t^r c_{j,m}^1, m = k-1, k$, we follow the same argument as in Lemma 3.2 to obtain

$$\left(h\sum_{j=1}^{N}\sum_{m=k-1}^{k}(\partial_t^r c_{j,m}^1)^2\right)^{\frac{1}{2}} \lesssim \left(h\sum_{j=1}^{N}\sum_{m=k-3}^{k-2}(\partial_t^r c_{j,m}^1)^2\right)^{\frac{1}{2}} \lesssim h^{n+1}\|\partial_t^r u\|_{n+1}.$$

Consequently, (4.10) is valid for all $r \geq 0$ and $m, 0 \leq m \leq k$ in case $i=1$. Now we suppose (4.10) is valid for $i, 1 \leq i \leq k-2$ and prove that the same result holds true for $i+1$. In fact, by (4.12) and the orthogonality of the Legendre polynomials,

$$|\partial_t^r c_{j,m}^{i+1}| \lesssim h^2\left(\sum_{p=0}^{m+2}|\partial_t^{r+1}c_{j,p}^i|\right) + h\left(\sum_{p=0}^{m+1}|\partial_t^r c_{j,p}^i|\right), \quad m \leq k-2.$$

For any positive integer $p, p \leq m+2$, by choosing $n' \leq \max(2k-m-2, k+i)$, we obtain from the induction hypothesis,

$$\left(h\sum_{j=1}^{N}(\partial_t^{1+r}c_{j,p}^i)^2\right)^{\frac{1}{2}} \leq h^{n'}\|\partial_t^{1+r}u\|_{n'} = h^{n'}\|\partial_t^r u\|_{n'+2},$$

which yields

$$h^2 \max_{0\le p\le m+2}\left(h\sum_{j=1}^{N}(\partial_t^{1+r}c_{j,p}^i)^2\right)^{\frac{1}{2}} \lesssim h^{n'+2}\|\partial_t^r u\|_{n'+2} = h^n\|\partial_t^r u\|_n,$$

where $n = n' + 2 \le \max(2k - m, k + 1 + (i + 1))$. Similarly, we can obtain

$$h \max_{0\le p\le m+1}\left(h\sum_{j=1}^{N}(\partial_t^{r}c_{j,p}^i)^2\right)^{\frac{1}{2}} \lesssim h^n\|\partial_t^r u\|_n.$$

Then a direct calculation indicates that (4.10) is valid for $i+1$ when $0 \le m \le k-2$. By (4.8)–(4.9) and the same argument as in Lemma 3.2, we obtain

$$\left(h\sum_{j=1}^{N}\sum_{m=k-1}^{k}(\partial_t^r c_{j,m}^{i+1})^2\right)^{\frac{1}{2}} \lesssim \left(h\sum_{j=1}^{N}\sum_{m=0}^{k-2}(\partial_t^r c_{j,m}^{i+1})^2\right)^{\frac{1}{2}} \lesssim h^{k+2+i}\|\partial_t^r u\|_{k+2+i}.$$

Therefore, (4.10) also holds true for $i+1$ when $m = k - 1, k$. Thus (4.10) is valid for $i+1$. This finishes our proof. ■

To study the superconvergence properties of u_h at Gauss and Lobatto points, we define

$$w = w_1, \quad u_I = Q_h u - w = Q_h u - w_1.$$

Then a direct calculation from (2.5), (3.24), (4.8)–(4.9) yields

$$A(w, v_h) - F(w, v_h) = -(w, \partial_x^2 v_h) - (\alpha w, \partial_x v_h). \tag{4.13}$$

On the other hand, choosing $v_h = u_h - u_I = \xi$ in (2.7) and using the orthogonal property, we get

$$\frac{1}{2}\frac{d}{dt}\|\xi\|_0^2 + \gamma\|\xi\|_E^2 \le |(\partial_t(u - u_I), \xi) + A(u - u_I, \xi) - F(u - u_I, \xi)|. \tag{4.14}$$

In light of (4.6)–(4.9), we have for all $\tilde{v} \in \mathbb{P}^k\backslash\mathbb{P}^1$,

$$(\partial_t(u - u_I), \tilde{v}) + A(u - u_I, \tilde{v}) - F(u - u_I, \tilde{v}) = (\partial_t w_1, \tilde{v}) - (\alpha w_1, \tilde{v}_x).$$

While for any $\bar{v} \in \mathbb{P}^1$, noticing that $\partial_x\bar{v} \in \mathbb{P}^0$ and $(u - Q_h u) \perp \mathbb{P}^0$, we have from (4.6)

$$A(u - Q_h u, \bar{v}) - F(u - Q_h u, \bar{v}) = 0,$$

and thus,

$$\begin{aligned}((u - u_I)_t, \bar{v}) + A(u - u_I, \bar{v}) - F(u - u_I, \bar{v}) &= (\partial_t w_1, \bar{v}) - (\alpha w_1, \partial_x\bar{v}) + (\partial_t(u - Q_h u), \bar{v})\\ &= (\partial_t w_1, \bar{v}) - (D_x^{-1}\partial_t(u - Q_h u) + \alpha w_1, \partial_x\bar{v}).\end{aligned}$$

Here in the last step, we have used the integration by parts and the fact that

$$D_x^{-1}\partial_t(u - Q_h u)(x_{j+\frac{1}{2}}^-) = D_x^{-1}\partial_t(u - Q_h u)(x_{j-\frac{1}{2}}^-) = 0.$$

Since any $v_h \in V_h$ can be decomposed into $v_h = \bar{v} + \tilde{v}$ with $\bar{v} \in \mathbb{P}^1, \tilde{v} \in \mathbb{P}^k \backslash \mathbb{P}^1$, we obtain

$$\begin{aligned}(\partial_t(u-u_I), v_h) &+ A(u-u_I, v_h) - F(u-u_I, v_h)\\ &= (\partial_t w_1, v_h) - (\alpha w_1, \partial_x v_h) - (D_x^{-1}\partial_t(u-Q_h u), \partial_x \bar{v}),\end{aligned}$$

which yields, together with (4.14) and the fact that $\|\bar{v}\|_0 + \|\tilde{v}\|_0 \lesssim \|v_h\|_0$,

$$\frac{1}{2}\frac{d}{dt}\|u_h - u_I\|_0^2 \lesssim \|\partial_t w_1\|_0^2 + \|w_1\|_0^2 + \|D_x^{-1}\partial_t(u-Q_h u)\|_0^2 + \|u_h - u_I\|^2.$$

By (4.11) and the Gronwall inequality,

$$\|(u_h-u_I)(\cdot,t)\|_0 \lesssim \|(u_h-u_I)(\cdot,0)\|_0 + h^{k+2}\sup_{\tau\in[0,t]}\|u(\cdot,\tau)\|_{k+5}, \quad \forall t > 0.$$

Consequently, if we choose the initial value $u_h(x,0) = I_h u_0(x)$ or $u_h(x,0) = Q_h u_0(x)$, then

$$\|(u_h-u_I)(\cdot,0)\|_0 \lesssim \|(I_h u - Q_h u)(\cdot,0)\|_0 + \|w_1(\cdot,0)\|_0 \lesssim \|u_0\|_{k+2}.$$

Thus,

$$\|(u_h-u_I)(\cdot,t)\|_0 \le h^{k+2}\sup_{\tau\in[0,t]}\|u(\cdot,\tau)\|_{k+5}, \quad \forall t > 0.$$

By following the same argument as in the case $f = 0$, we obtain the same superconvergence results at Gauss and Lobatto points for the convection-diffusion flux $f = \alpha u$. That is, the results in Corollary 3.5 are valid for the convection-diffusion equation $f = \alpha u$.

Now we focus on our attentions to the $2k$-th superconvergence rate of the DDG approximation at nodes for $k \ge 3$. We modify the correction function w as

$$w = \sum_{i=1}^{k-1} w_i, \tag{4.15}$$

where w_i is defined by (4.7)–(4.9). In this case, (4.13) still holds. Then for all $\tilde{v} \in \mathbb{P}^k \backslash \mathbb{P}^1$,

$$\begin{aligned}|((u-u_I)_t, \tilde{v}) + A(u-u_I, \tilde{v}) - F(u-u_I, \tilde{v})| &= |(\partial_t w_{k-1}, \tilde{v}) - (\alpha w_{k-1}, \tilde{v}_x)|\\ &\lesssim h^{2k}\|\partial_t u\|_{2k}(\|\tilde{v}\|_0 + |\tilde{v}|_1).\end{aligned}$$

For any $\bar{v} \in \mathbb{P}^1(I_j)$, we suppose $\bar{v} = v_{j,0}L_{j,0} + v_{j,1}L_{j,1}$. Then there holds for all $i, 1 \le i \le k-1$,

$$(\partial_t w_i, \bar{v})_j - \alpha(w_i, \bar{v}_x)_j = \sum_{j=1}^{N}\left(h\partial_t c_{j,0}^i v_{j,0} + \frac{h}{3}\partial_t c_{j,1}^i v_{j,1} - 2\alpha c_{j,0}^i v_{j,1}\right).$$

In light of (4.10), we have

$$\left(\sum_{j=1}^{N} h(\partial_t c_{j,0}^i)^2 + h(c_{j,0}^i)^2 + h^3(\partial_t c_{j,1}^i)^2\right)^{\frac{1}{2}} \lesssim h^{2k}\|\partial_t u\|_{2k}.$$

Then we use the Cauchy–Schwartz inequality to obtain

$$|(\partial_t w_i, \bar{v}) - (w_i, \bar{v}_x)| \lesssim h^{2k} \|\partial_t u\|_{2k} (\|\bar{v}\|_0 + |\bar{v}|_1), \quad \forall \bar{v} \in \mathbb{P}^0.$$

Noticing that $(u - Q_h u) \perp \mathbb{P}^1$ when $k \geq 3$, we get

$$\begin{aligned}(\partial_t (u - u_I), \bar{v}) + A(u - u_I, \bar{v}) - F(u - u_I, \bar{v}) &= (\partial_t w, \bar{v}) - (\alpha w, \bar{v}_x) \\ &\lesssim h^{2k} \|\partial_t u\|_{2k} (\|\bar{v}\|_0 + |\bar{v}|_1).\end{aligned}$$

Then for all $v_h = \bar{v} + \tilde{v} \in V_h$,

$$|(\partial_t (u - u_I), v_h) + A(u - u_I, v_h) - F(u - u_I, v_h)| \lesssim h^{2k} \|\partial_t u\|_{2k} (\|v_h\|_0 + |v_h|_1).$$

Consequently, we have from (4.14), the Cauchy–Schwartz inequality, and the Gronwall inequality,

$$\|(u_h - u_I)(\cdot, t)\|_0 \lesssim \|(u_h - u_I)(\cdot, 0)\|_0 + h^{2k} \sup_{\tau \in [0,t]} \|\partial_t u(\cdot, \tau)\|_{2k}, \quad \forall t > 0.$$

If the initial value is taken as $u_h(x, 0) = u_I(x, 0)$, we immediately obtain

$$\|(u_h - u_I)(\cdot, t)\|_0 \lesssim h^{2k} \sup_{\tau \in [0,t]} \|\partial_t u(\cdot, \tau)\|_{2k} \lesssim h^{2k} \sup_{\tau \in [0,t]} \|u(\cdot, \tau)\|_{2k+2}, \quad \forall t > 0.$$

Then by the same arguments as in Theorem 3.8, the superconvergence result (3.39) follows for the convection-diffusion flux $f = \alpha u$.

Remark 4.3. Following the argument as in the constant coefficient case, we can obtain the same superconvergence results for the variable coefficient α, where α is sufficiently smooth.

V. NUMERICAL EXAMPLES

In this section, we present numerical examples to verify our theoretical findings. To test the superconvergence phenomenon of u_h at Gauss, Lobatto and nodal points, as well as the supercloseness of u_h toward $I_h u$, we will measure in our numerical experiments the errors $\|u_h - I_h u\|_0$ and e_l, e_g, e_n (defined in Theorem 3.8 and Corollary 3.5).

In our experiments, we will use DDG scheme (2.1) for spatial discretization, and the fourth-order Runge–Kutta method for time discretization with the time step $\triangle t = 0.001h^2$. We obtain our meshes by equally dividing $(0, 2\pi)$ into $N(4, \ldots, 64)$ subintervals.

Example 1. We consider the following equation with the periodic boundary condition:

$$u_t = u_{xx}, \quad (x, t) \in (0, 2\pi) \times (0, T), \quad u(x, 0) = \sin(x).$$

The exact solution is

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

We take different ways of initial discretizations and different values of (β_0, β_1) to test the influence of the initial errors and the choice of parameters in the DDG numerical fluxes (2.2) on the superconvergence rate.

TABLE I. Errors and corresponding convergence rates for zero flux with $u_h(\cdot,0) = \tilde{u}_I$ and $\beta_1 = \frac{1}{2k(k+1)}$.

k	N	e_l	Order	e_n	Order	e_g	Order	$\|u_h - I_h u\|_0$	Order
	4	3.18e-03	–	2.46e-03	–	4.62e-03	–	5.38e-03	–
	8	1.94e-04	4.04	1.41e-04	4.12	4.60e-04	3.33	3.44e-04	3.97
2	16	1.20e-05	4.01	8.65e-06	4.03	5.24e-05	3.13	2.15e-05	4.00
	32	7.49e-07	4.00	5.38e-07	4.00	6.38e-06	3.04	1.35e-06	4.00
	64	4.68e-08	4.00	3.36e-08	4.00	7.91e-07	3.01	8.41e-08	4.00
	4	6.70e-05	–	2.76e-06	–	3.84e-04	–	9.71e-05	–
	8	2.36e-06	4.82	7.19e-09	8.58	2.87e-05	3.74	2.08e-06	5.55
3	16	7.70e-08	4.94	2.52e-11	8.16	1.89e-06	3.93	5.66e-08	5.20
	32	2.43e-09	4.98	9.72e-14	8.02	1.19e-07	3.98	1.70e-09	5.06
	64	7.62e-11	5.00	2.24e-16	8.76	7.49e-09	4.00	5.26e-11	5.01
	4	3.73e-06	–	4.25e-07	–	3.03e-05	–	4.11e-06	–
	8	4.75e-08	6.30	1.52e-09	8.13	8.08e-07	5.23	3.17e-08	7.01
4	16	6.90e-10	6.10	5.82e-12	8.03	2.38e-08	5.08	2.47e-10	7.00
	32	1.06e-11	6.03	2.49e-14	7.87	7.32e-10	5.02	1.93e-12	7.00
	64	1.66e-13	6.00	8.84e-17	8.13	2.28e-11	5.01	1.50e-14	7.00

TABLE II. Errors and corresponding convergence rates for zero flux with $u_h(\cdot,0) = \tilde{u}_I$ and $\beta_1 \neq \frac{1}{2k(k+1)}$.

k	N	e_l	Order	e_n	Order	e_g	Order	$\|u_h - I_h u\|_0$	Order
	4	2.76e-03	2.43	2.08e-03	–	1.04e-03	–	3.85e-03	–
	8	2.59e-04	3.41	1.12e-04	4.22	4.33e-04	1.26	3.60e-04	3.42
2	16	3.04e-05	3.09	6.73e-06	4.06	1.43e-04	1.62	4.34e-05	3.05
	32	3.75e-06	3.02	4.16e-07	4.01	3.73e-05	1.92	5.40e-06	3.01
	64	4.67e-07	3.01	2.60e-08	4.00	9.45e-06	1.98	6.75e-07	3.00
	4	8.67e-05	–	8.31e-07	–	4.86e-04	–	1.03e-04	–
	8	3.84e-06	4.50	4.33e-09	7.58	4.00e-05	3.60	4.83e-06	4.41
4	16	1.33e-07	4.85	1.86e-11	7.86	2.70e-06	3.89	1.70e-07	4.83
	32	4.26e-09	4.96	7.66e-14	7.93	1.72e-07	3.97	5.49e-09	4.95
	64	1.34e-10	4.99	3.33e-16	7.84	1.08e-08	4.00	1.73e-10	5.00

We first choose $u_h(\cdot,0) = \tilde{u}_I(\cdot,0)$ with $\tilde{u}_I(\cdot,0)$ defined in (3.40). Note that this special initial discretization satisfies both (3.22) and (3.38). We list in Table I the various errors and the corresponding convergence rates for $k = 2, 3, 4$, and $T = 1$, with the parameter (β_0, β_1) taken as $(10, \frac{1}{12}), (12, \frac{1}{24}), (16, \frac{1}{40})$, respectively. Note that the choice of (β_0, β_1) assures $\beta_1 = \frac{1}{2k(k+1)}$ for all $k = 2, 3, 4$. From Table I, we observe a convergence rate $k + 2$ for e_l and a rate $k + 1$ for e_g. These results are consistent with the superconvergence results given in Corollary 3.5, which indicates that the superconvergence of the derivative approximation at Gauss points and the function value approximation at Lobatto points exist. Moreover, the convergence rates provided in (3.27)–(3.28) are optimal. We also observe a superconvergence rate $2k$ for e_n when $k = 2, 4$, which confirms the superconvergence results in (3.39). While, it seems that the convergence rate of e_n can reach $2k+2$ when $k = 3$, two-order higher than that in (3.39). As predicted in (3.23), the error $\|u_h - I_h u\|_0$ is convergent with an order $k + 2$ for $k = 2, 3$. While, it seems that the convergence rate of $\|u_h - I_h u\|_0$ is one order better than that in (3.23) when $k = 4$.

To test the influence of β_1 on the superconvergence rate, we also consider the case in which $\beta_1 \neq \frac{1}{2k(k+1)}$. Table II demonstrates the error data and the corresponding convergence rate for $k = 2, 4$, with $(\beta_0, \beta_1) = (12, \frac{1}{24}), (10, \frac{1}{12})$ respectively. Clearly, we observe a rate $k + 1$ for e_l and $\|u_h - I_h u\|_0$, and a rate k for e_g, which means that the superconvergence properties of u_h at Gauss and Lobatto points, as well as the supercloseness between u_h and I_{hu}, disappear. Therefore,

TABLE III. Errors and corresponding convergence rates for zero flux with $u_h(\cdot, 0) = I_h u_0$ and $\beta_1 = \frac{1}{2k(k+1)}$.

k	N	e_l	Order	e_n	Order	e_g	Order	$\|u_h - I_h u\|_0$	Order
	4	1.22e-02	–	1.75e-03	–	2.34e-02	–	1.63e-02	–
	8	2.82e-04	5.43	1.39e-04	3.66	9.20e-04	4.67	4.71e-04	5.11
2	16	1.33e-05	4.41	8.62e-06	4.00	6.83e-05	3.75	2.34e-05	4.32
	32	7.69e-07	4.11	5.37e-07	4.00	6.91e-06	3.31	1.38e-06	4.01
	64	4.71e-08	4.02	3.36e-08	4.00	8.08e-07	3.01	8.46e-08	4.02
	4	1.20e-04	–	4.39e-07	–	5.76e-04	–	1.31e-04	–
	8	5.40e-06	4.47	1.13e-08	5.29	5.17e-05	3.48	5.21e-06	4.66
3	16	1.87e-07	4.85	5.51e-11	7.68	3.57e-06	3.86	1.77e-07	4.87
	32	6.00e-09	4.96	2.28e-13	7.92	2.29e-07	3.96	5.66e-09	4.97
	64	1.89e-10	4.99	9.14e-16	7.96	1.44e-08	3.99	1.78e-10	4.99
	4	1.89e-05	–	1.26e-08	–	1.15e-04	–	2.22e-05	–
	8	1.23e-07	7.27	2.19e-11	9.17	1.57e-06	6.20	1.45e-07	7.26
4	16	1.08e-09	6.83	2.31e-14	9.89	3.09e-08	5.67	1.09e-09	7.06
	32	1.23e-11	6.45	2.56e-17	9.81	7.92e-10	5.28	8.39e-12	7.02
	64	1.72e-13	6.17	2.94e-20	9.77	2.33e-11	5.09	6.71e-14	6.97

TABLE IV. Errors and corresponding convergence rates for zero flux with $u_h(\cdot, 0) = I_h u_0$ and $\beta_1 \neq \frac{1}{2k(k+1)}$.

k	N	e_l	Order	e_n	Order	e_g	Order	$\|u_h - I_h u\|_0$	Order
	4	4.14e-03	–	2.36e-03	–	2.19e-03	–	5.07e-03	–
	8	5.22e-04	2.99	1.39e-04	4.08	1.15e-03	0.93	7.07e-04	2.84
2	16	6.55e-05	2.99	8.61e-06	4.01	3.23e-04	1.83	9.32e-05	2.92
	32	8.20e-06	3.00	5.37e-07	4.00	8.29e-05	1.96	1.18e-05	2.98
	64	1.06e-06	3.00	3.36e-08	4.00	2.08e-05	1.99	1.48e-06	2.99
	4	2.56e-04	–	9.67e-07	–	1.42e-03	–	3.03e-04	–
	8	2.01e-05	3.67	1.41e-08	6.10	2.09e-04	2.76	2.53e-05	3.58
4	16	9.97e-07	4.33	9.53e-11	7.21	2.03e-05	3.36	1.28e-06	4.31
	32	3.65e-08	4.77	4.46e-13	7.74	1.47e-06	3.78	4.70e-08	4.76
	64	1.19e-09	4.94	2.21e-15	7.65	9.61e-08	3.94	1.54e-09	4.94

as we indicate in Remark 3.9, the value of β_1 affects the superconvergence of u_h at Gauss and Lobatto points. However, we can still observe the superconvergence phenomenon at nodes, which indicates that the superconvergence properties of u_h at nodes are independent of the choice of β_1.

As a comparison group, we also test another way of initial discretization. That is, $u_h(\cdot, 0) = I_h u_0$. Note that this choice of initial discretization satisfies the condition (3.30) while does not satisfy (3.38). Two cases, i.e. $\beta_1 = \frac{1}{2k(k+1)}$ and $\beta_1 \neq \frac{1}{2k(k+1)}$, are considered. Listed in Table III are the corresponding errors and convergence rates in case $\beta_1 = \frac{1}{2k(k+1)}$. That is, we choose $(\beta_0, \beta_1) = (2, \frac{1}{12}), (6, \frac{1}{24}), (8, \frac{1}{40})$ for $k = 2, 3, 4$, respectively. Table IV demonstrates the corresponding numerical results in case $\beta_1 \neq \frac{1}{2k(k+1)}$ for $k = 2, 4$ with $(\beta_0, \beta_1) = (6, \frac{1}{24}), (2, \frac{1}{12})$.

Again, we observe similar superconvergence phenomena as in the correction initial disrcretization case. To be more precise, when $\beta_1 = \frac{1}{2k(k+1)}$, the DDG solution is superconvergent with an order $k+2$ to the Gauss–Lobatto projection of the exact solution, as well as the function value approximation at Lobatto points; and for the error of derivative approximation at Gauss points, the convergence rate is $k+1$. While the superconvergence phenomena disappears when $\beta_1 \neq \frac{1}{2k(k+1)}$. It should be pointed out that, we still observe an order $2k$ for e_n in both cases $\beta_1 = \frac{1}{2k(k+1)}$ and $\beta_1 \neq \frac{1}{2k(k+1)}$, although the initial condition (3.38) is not satisfied. In addition, the convergence rate of e_n can reach $2k+2$ for $k = 3, 4$ when $\beta_1 = \frac{1}{2k(k+1)}$, two-order better than that in (3.39).

TABLE V. Errors and corresponding convergence rates for linear flux with $u_h(\cdot,0) = \tilde{u}_I$ and $\beta_1 = \frac{1}{2k(k+1)}$.

k	N	e_l	Order	e_n	Order	e_g	Order	$\|u_h - I_h u\|_0$	Order
	4	6.77e-03	–	5.00e-03	–	2.79e-02	–	8.43e-03	–
	8	4.08e-04	4.05	3.05e-04	4.04	3.15e-03	3.15	4.61e-04	4.19
2	16	2.34e-05	4.12	1.83e-05	4.05	3.59e-04	3.13	2.47e-05	4.23
	32	1.45e-06	4.02	1.13e-06	4.02	4.36e-05	3.04	1.47e-06	4.07
	64	9.08e-08	4.00	7.05e-08	4.00	5.40e-06	3.01	9.09e-08	4.02
	4	9.63e-04	–	1.88e-04	–	5.06e-03	–	8.36e-04	–
	8	3.26e-05	4.88	2.86e-06	6.04	3.53e-04	3.84	3.12e-05	4.74
3	16	1.11e-06	4.88	3.86e-08	6.21	2.52e-05	3.81	1.02e-06	4.94
	32	3.57e-08	4.96	5.84e-10	6.05	1.64e-06	3.94	3.22e-08	5.00
	64	1.13e-09	5.00	9.06e-12	6.01	1.04e-07	3.98	1.01e-09	5.00
	4	1.04e-04	–	1.61e-05	–	9.35e-04	–	6.32e-05	–
	8	2.28e-06	5.51	3.82e-08	8.72	3.51e-05	4.74	1.58e-06	5.32
4	16	3.10e-08	6.20	1.43e-10	8.06	9.30e-07	5.24	2.29e-08	6.11
	32	4.67e-10	6.05	5.41e-13	8.04	2.76e-08	5.07	3.52e-10	6.03
	64	7.21e-12	6.02	2.10e-15	8.01	8.51e-10	5.02	5.47e-12	6.01

TABLE VI. Errors and corresponding convergence rates for linear flux with $u_h(\cdot,0) = \tilde{u}_I$ and $\beta_1 \neq \frac{1}{2k(k+1)}$.

k	N	e_l	Order	e_n	Order	e_g	Order	$\|u_h - I_h u\|_0$	Order
	4	7.67e-03	–	4.74e-03	–	2.08e-02	–	7.98e-03	–
	8	6.75e-04	3.51	2.55e-04	4.22	2.41e-03	3.11	6.82e-04	3.55
2	16	7.14e-05	3.24	1.45e-05	4.13	4.48e-04	2.43	8.35e-05	3.03
	32	8.23e-06	3.12	8.81e-07	4.04	1.08e-04	2.05	1.07e-05	2.97
	64	9.84e-07	3.06	5.46e-08	4.01	2.70e-05	2.00	1.35e-06	2.99
	4	1.75e-04	–	1.87e-05	–	1.55e-03	–	9.56e-05	–
	8	1.04e-05	4.08	4.86e-08	8.59	1.55e-04	3.32	7.55e-06	3.66
4	16	3.85e-07	4.75	2.20e-10	7.79	1.09e-05	3.83	3.18e-07	4.57
	32	1.26e-08	4.93	8.96e-13	7.94	7.00e-07	3.96	1.09e-08	4.87
	64	4.00e-10	5.00	3.55e-15	7.98	4.41e-08	3.99	3.51e-10	4.96

Example 2. We consider the linear convection-diffusion equation

$$u_t + (\alpha u)_x = u_{xx}, \quad (x,t) \in (0,2\pi) \times (0,T), \quad u(x,0) = e^{\sin(x)/2}$$

with the periodic boundary condition and $\alpha = 1 + \frac{1}{2}\cos(x-t)$. The exact solution is

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

Listed in Tables V and VI are error data and the corresponding convergence rate for the final time $T=1$ in cases $\beta_1 = \frac{1}{2k(k+1)}$ and $\beta_1 \neq \frac{1}{2k(k+1)}$. In Table V, the parameter is chosen as $(\beta_0, \beta_1) = (10, \frac{1}{12}), (12, \frac{1}{24}), (16, \frac{1}{40})$ for $k=2, 3, 4$, respectively, and in Table VI, $(\beta_0, \beta_1) = (12, \frac{1}{24})$ for both $k=2, 4$. Similar as the zero flux case, we observe a $(k+2)$ -th order superconvergence rate for e_l and $\|u_h - I_h u\|_0$, and a $(k+1)$ -th order for e_g when $\beta_1 = \frac{1}{2k(k+1)}$. However, those superconvergence phenomena disappears when $\beta_1 \neq \frac{1}{2k(k+1)}$. Different from the errors e_l, e_g, we observe a $2k$-th superconvergent order for the error e_n in both $\beta_1 = \frac{1}{2k(k+1)}$ and $\beta_1 \neq \frac{1}{2k(k+1)}$. As we may recall, all these numerical results are consistent with our theoretical findings. Moreover, all the superconvergence rates in (3.27)–(3.28) and (3.39) are optimal, i.e., the error bounds are sharp.

References

1. B. Cockburn, S. Hou, and C.-W. Shu, The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws, IV: the multidimensional case, Math Comput 54 (1990), 545–581.
2. B. Cockburn, S. Lin, and C.-W. Shu, TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws, III: one dimensional systems, J Comput Phys 84 (1989), 90–113.
3. B. Cockburn and C.-W. Shu, TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws, II: general framework, Math Comput, 52 (1989), 411–435.
4. B. Cockburn and C.-W. Shu, The Runge-Kutta local projection P1-discontinuous Galerkin finite element method for scalar conservation laws, RAIRO Modél, Math Anal Numér 25 (1991), 337–361.
5. 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.
6. W. H. Reed and T. R. Hill, Triangular Mesh for Neutron Transport Equation, Los Alamos Scientific Laboratory Report LA-UR-73-479, Los Alamos, NM, 1973.
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. D. N. Arnold, An interior penalty finite element method with discontinuous elements, SIAM J Numer Anal 19 (1982), 742–760.
9. G. A. Baker, Finite element methods for elliptic equations using noncomforming elements, Math Comp 31 (1977), 45–59.
10. M. F. Wheeler, An elliptic collacation-finite element method with interior penalties, SIAM J Numer Anal 15 (1978), 152–161.
11. 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.
12. 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.
13. Y. Cheng and C.-W. Shu, A discontinuous Galerkin finite element method for time dependent partial differential equations with higher order derivatives, Math Comput 77 (2007), 699–730.
14. G. Gassner, F. Lörcher, and C. D. Munz, A contribution to the construction of diffusion fluxes for finite volume and discontinuous Galerkin schemes, J Comput Phys 224 (2007), 1049–1063.
15. B. van Leer and S. Nomura, Discontinuous Galerkin for diffusion, Proceeding of 17th AIAA Computational Fluid Dynamics Conference, Toronto, Ontario Canada, American Institute of Aeronautics and Astronautics, Inc., AIAA-2005-5108, 2005.
16. H. Liu and J. Yan, The direct discontinuous Galerkin (DDG) methods for diffusion problems, SIAM J Numer Anal 47 (2009), 675–698.
17. H. Liu and J. Yan, The direct discontinuous Galerkin (DDG) method for diffusion with interface corrections, Commun Comput Phys 8 (2010), 541–564.
18. Y.-Q. Huang, H. Liu, and N.-Y. Yi, Recovery of normal derivatives from the piecewise L^2 projection, J Comput Phys 231 (2012), 1230–1243.
19. H. Liu, Optimal error estimates of the direct discontinuous Galerkin method for convection-diffusion equations, Math Comput 84 (2015), 2263–2295.
20. D. N. Arnold, An interior penalty finite element method with discontinuous elements, SIAM J Numer Anal 19 (1982), 742C-760C.
21. H. Liu and H. Yu. Maximum-Principle-Satisfying third order discontinuous Galerkin schemes for Fokker-Planck equations, SIAM J Sci Comput 36 (2014), A2296C–A2325C.
22. Y.-F. Zhang, X.-X. Zhang, and C.-W. Shu, Maximum-principle-satisfying second order discontinuous Galerkin schemes for convection-diffusion equations on triangular meshes, J Comput Phys 234 (2013), 295–316.

23. H. Liu and H. Yu, The entropy satisfying dicontinuous Galerkin method for Fokker-Planck equations, J Sci Comput 62 (2015), 803C–830C.

24. H. Liu and Z.-M. Wang, An entropy satisfying discontinuous Galerkin method for nonlinear Fokker-Planck equations, J Sci Comput (2016), 1–24.

25. C. Chen, Structure Theory of Superconvergence of Finite Elements (in Chinese), Hunan Science and Technology Press, Hunan, China, 2001.

26. R. E. Ewing, R. D. Lazarov, and J. Wang, Superconvergence of the velocity along the Gauss lines in mixed finite element methods, SIAM J Numer Anal 28 (1991), 1015–1029.

27. L. B. Wahlbin, Superconvergence in Galerkin Finite Element Methods, Lecture Notes in Mathematics, Vol. 1605, Springer, Berlin, 1995.

28. W. Cao, Z. Zhang, and Q. Zou, Is $2k$-conjecture valid for finite volume methods? SIAM J Numer Anal 53 (2015), 942–962.

29. C. Chen and S. Hu, The highest order superconvergence for bi-k degree rectangular elements at nodes: a proof of $2k$-conjecture, Math Comput 82 (2013), 1337–1355.

30. W. Cao, C.-W. Shu, Y. Yang, and Z. Zhang, Superconvergence of discontinuous Galerkin methods for 2-D hyperbolic equations, SIAM J Numer Anal 53 (2015), 1651–1671.

31. W. Cao and Z. Zhang, Superconvergence of local discontinuous Galerkin methods for one-dimensional linear parabolic equations, Math Comput 85 (2016), 63–84.

32. W. Cao, Z. Zhang, and Q. Zou, Superconvergence of discontinuous Galerkin methods for linear hyperbolic equations, SIAM J Numer Anal 52 (2014), 2555–2573.

33. A. Halanay, Differential Equations: Stability, Oscillations, Time Lags, Academic Press, New York and London, 1966.