ON SUPERCONVERGENT DISCONTINUOUS GALERKIN METHODS FOR SCHRÖDINGER EQUATIONS AND SPARSE GRID CENTRAL DISCONTINUOUS GALERKIN METHOD

By

Anqi Chen

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Applied Mathematics – Doctor of Philosophy

2019

ABSTRACT

ON SUPERCONVERGENT DISCONTINUOUS GALERKIN METHODS FOR SCHRÖDINGER EQUATIONS AND SPARSE GRID CENTRAL DISCONTINUOUS GALERKIN METHOD

By

Anqi Chen

In this thesis, we design and analyze a discontinuous Galerkin (DG) method for one-dimensional Schrödinger equations under a general class of numerical fluxes, and another efficient DG method for high-dimensional hyperbolic equations.

In the first DG method, we develop an ultra-weak discontinuous Galerkin (UWDG) method to solve the one-dimensional nonlinear Schrödinger equation. Stability conditions and error estimates are derived for the scheme with a general class of numerical fluxes. The error estimates are based on detailed analysis of the projection operator associated with each individual flux choice. Depending on the parameters, we find out that in some cases, the projection can be defined element-wise, facilitating analysis. In most cases, the projection is global, and its analysis depends on the resulting 2×2 block-circulant matrix structures. For a large class of parameter choices, optimal *a priori* L^2 error estimates can be obtained. Numerical examples are provided verifying theoretical results.

In addition to the stability and error analysis, we analyze the superconvergence properties of the UWDG method for one-dimensional linear Schrödinger equation with various choices of flux parameters. Depending on the flux choices and if the polynomial degree k is even or odd, we prove $2k$ or $(2k-1)$-th order superconvergence rate for cell averages and numerical flux of the function, as well as $(2k-1)$ or $(2k-2)$-th order for numerical flux of the derivative. In addition, we prove superconvergence of $(k+2)$ or $(k+3)$-th order of the UWDG solution towards a special projection. At a class of special points, the function values and the first and

second order derivatives of the UWDG solution are superconvergent with order $k+2, k+1, k$, respectively. The proof relies on the correction function techniques initiated in [12], and applied to [10] for direct DG (DDG) methods for diffusion problems. By negative norm estimates, we apply the post-processing technique and show that the accuracy of our scheme can be enhanced to order $2k$. Theoretical results are verified by numerical experiments.

In the second DG method, we develop sparse grid central discontinuous Galerkin (CDG) scheme for linear hyperbolic systems with variable coefficients in high dimensions. The scheme combines the CDG framework with the sparse grid approach, with the aim of breaking the curse of dimensionality. A new hierarchical representation of piecewise polynomials on the dual mesh is introduced and analyzed, resulting in a sparse finite element space that can be used for non-periodic problems. Theoretical results, such as L^2 stability and error estimates are obtained for scalar problems. CFL conditions are studied numerically comparing discontinuous Galerkin (DG), CDG, sparse grid DG and sparse grid CDG methods. Numerical results including scalar linear equations, acoustic and elastic waves are provided.

ACKNOWLEDGMENTS

I would like to express my great appreciation to my advisor, Professor Yingda Cheng, for her invaluable guidance, inspiration, encouragement and patience during my five-year PhD study. Professor Cheng taught me how to think and express my ideas rigorously like a mathematician. She was always there whenever I needed advices, and she always provided me with kind and thoughtful help. I could never finish this thesis without her help.

I would also like to thank all my committee members, Professor Andrew Christlieb, Professor Chichia Chiu and Professor Jianliang Qian for their guidance and advices.

I would also like to thank my friends, Zhanjing Tao, Wei Guo, Yan Jiang, Menglun Wang, Xin Yang, Kedi Wu, Zixuan Cang, Xiao Feng, Qinfeng Gao, Chao Song and Ruochuan Zhang for their friendships and many precious moments.

Last but not least, I am grateful to my family. My father Weibing Chen and mother Manzhen Deng always encouraged me to pursue what I wanted. I can't say thanks enough for their support in my life. And I am grateful to my girlfriend Bixi Zhang for her love and support during the final stage of my PhD.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Overview

The discontinuous Galerkin (DG) method is a class of finite element methods using completely discontinuous piecewise function space for test functions and numerical solution. The first DG method was introduced by Reed and Hill in [59]. A major development of DG method is the Runge-Kutta DG (RKDG) framework introduced for solving hyperbolic conservation laws containing only first order spatial derivatives in a series of papers [25, 24, 22, 21, 26]. Because of the completely discontinuous basis, DG method has several attractive properties. It can be used on many types of meshes, even those with hanging nodes. The method has h-p adaptivity and very high parallel efficiency.

A particular type of DG methods that is related to this thesis is central DG (CDG) scheme. The CDG schemes [48, 50, 51] are a class of DG schemes on overlapping cells that combine the idea of the central schemes [56, 45, 49] with the DG weak formulation. Such methods are intrinsically Riemann solver free, therefore no costly flux evaluations are needed in the computation. It is well known that the CDG schemes allow larger CFL numbers than the standard DG methods except for piecewise constant approximations [50, 60]. This compensates the increased cost caused by duplicate representation of the solution on the dual mesh.

In this thesis, we will focus on the design of a new DG method for one-dimensional Schrödinger equations and its error analysis, and superconvergence analysis, as well as a new sparse-grid central DG method for high-dimensional hyperbolic equations to make the simulation more efficient and accurate.

1.2 UWDG method for one-dimensional Schrödinger equations

In this section, we introduce the one-dimensional time-dependent nonlinear Schrödinger (NLS) equation and review the numerical methods designed to solve this equation.

The NLS equation is written as follows:

$$iu_t + u_{xx} + f(|u|^2)u = 0, \tag{1.1}$$

where $f(u)$ is a nonlinear real function and u is a complex function. The Schrödinger equation is the fundamental equation in quantum mechanics, reaching out to many applications in fluid dynamics, nonlinear optics and plasma physics. It is also called Schrödinger wave equation as it can describe how the wave functions of a physical system evolve over time. Many numerical methods have been applied to solve NLS equations [14, 28, 42, 43, 58, 71, 74]. In [14, 74], several important finite difference schemes are implemented, analyzed and compared. In [58], the author introduced a pseudo-spectral method for general NLS equations. Many finite element methods have been tested, such as quadratic B-spline for NLS in [28, 71] and space-time DG method for nonlinear (cubic) Schrödinger equation in [42, 43].

Various types of DG schemes has been applied to solve Schrödinger equations and they have different discretization for the second order spatial derivative term. One group of such methods is the so-called local DG (LDG) method invented in [25] for convection-diffusion equations. The algorithm is based on introducing auxiliary variables and reformulating the equation into its first order form. In [80], an LDG method using alternating fluxes is developed with L^2 stability and proved $(k+\frac{1}{2})$-th order of accuracy. Later in [81], Xu and Shu proved optimal accuracy for both the solution and the auxiliary variables in the LDG method for high order wave equations based on refined energy estimates. In [47], the authors presented an LDG method with exponential time differencing Runge-Kutta scheme and investigated the energy conservation performance of the scheme. Another group of method involves treating the second order spatial derivative directly in the weak formulations, such as IPDG method [77, 30] and NIPG method [62, 63]. Those schemes enforce a penalty jump term in the weak formulation, and they have been extensively applied to acoustic and elastic wave propagation [37, 3, 61]. As for Schrödinger equations, the direct DG (DDG) method was applied to Schrödinger equation in [52] and achieved energy conservation and optimal accuracy.

In Chapter 2, we choose to discretize the second order spatial derivative term directly using UWDG method, which can be traced backed to [13], and refer to those DG methods [72] that rely on repeatedly applying integration by parts so all the spatial derivatives are shifted from the solution to test function in the weak formulation. In [18], Cheng and Shu developed UWDG methods for general time dependent problems with higher order spatial derivatives. In [6], Bona *et. al.* proposed an UWDG scheme for generalized KdV equation and performed error estimates.

We investigate a most general form of the numerical flux functions that ensures stability along with our ultra-weak formulation. To estimate the convergence rate of our scheme, we introduce special projections associated different flux parameters, and proved detailed estimates for the projections. With the results for the convergence rate of the projections, *a priori* L^2 error estimates are obtained. Numerical tests are provided to verify the theoretical results for projection operators and solution convergence rates under various flux parameters.

1.3 Superconvergence analysis of DG methods

The study of superconvergence is of importance because *a posteriori* error estimates can be derived guiding adaptive calculations. For superconvergence of DG methods, many results exist in the literature. We refer the readers to [2, 1] for ordinary differential equation results. In [19], Cheng and Shu proved that the DG and LDG solutions are $(k+3/2)$-th order superconvergent towards projections of exact solutions of hyperbolic conservation laws and convection-diffusion equations using specially designed test functions when piecewise polynomials of degree k are used. For linear hyperbolic problems, in [82], Yang and Shu proved that, under suitable initial discretization, the DG solutions of linear hyperbolic systems are convergent with optimal $(k+2)$-th order at Radau points. More recently, in [12], Cao *et al* proved the $(2k+1)$-th superconvergence rate for cell average and DG numerical fluxes by introducing a locally defined correction function. The correction function also helps simplify the proof for point wise $(k+2)$-th superconvergence rate at Radau points and prove the derivative of DG solution has $(k+1)$-th superconvergence rate at so-called "left Radau" points. Then this technique has been extended to prove the superconvergence of DG solutions for linear and nonlinear hyperbolic PDEs in [11, 9], DDG method for convection diffusion equations [10] and LDG method for linear Schrödinger equations [84]. Overall, for equations with higher order spatial derivatives, the same type of correction functions can be used for the LDG method which is based on a reformulation into a first order system of equations. For DDG method, new correction functions are needed treating the second order derivative directly [10].

Another type of superconvergence of DG methods is achieved by postprocessing the solution by convolution with a kernel function, which is a linear combination of B-spline functions. For linear hyperbolic systems, [23] provided a framework for constructing such postprocessor and proving the superconvergence of the postprocessed DG solutions. Through the analysis of negative norm estimates and divided difference estimates, they showed that the postprocessed solution is superconvergent at a rate of $2k+1$. More recently, in [41, 54] the analysis are extended to scalar nonlinear hyperbolic equations.

In Chapter 3, we study the superconvergence properties of the UWDG methods for linear Schrödinger equation with scale invariant flux parameters. Such choice include all commonly used fluxes, e.g. alternating, central, DDG and interior penalty DG (IPDG) fluxes. Using the correction function idea and negative norm estiamtes, we are able to prove superconvergence rate for solution and its derivatives at certain points, for cell averages and numerical fluxes, for solution towards a special projection, and for the postprocessed solution.

1.4 Sparse-grid DG methods for high-dimensional PDEs

It has been a challenging problem for numerical simulations in high dimensions due to the so-called *curse of dimensionality*, which means the cost of computing and storing an approximation with a prescribed accuracy increases exponentially on dimension d. Many discretization techniques and computation techniques have been developed to alleviate the problem to some extent. Among them, the sparse grid method has been a successful tool. It was originally developed to solve PDEs [83, 33] based on tensor product hierarchical basis representation. The method can reduce the full grid discretization complexity from $O(h^{-d})$ to $O(h^{-1}|\log_2 h|^{d-1})$, where h is the uniform mesh size in each dimension, and only slightly deteriorate the accuracy. In recent years, sparse grid techniques have been incorporated in collocation methods for high-dimensional stochastic differential equations [79, 78, 57, 53], finite element methods [83, 8, 66], finite difference methods [34, 36], finite volume methods [40], and spectral methods [35, 32, 68, 69] for high-dimensional PDEs.

Recently, our research group initiated a line of research on the development of sparse grid DG methods [76, 38, 39]. The sparse grid DG methods use the sparse finite element space, which has multidimensional multiwavelet bases constructed by tensor products from one-dimensional wavelet basis, in the DG framework to treat high-dimensional problems. The methods has been proven to reduce the degrees of freedom of $O(h^{-d})$ in the standard full grid approximation space to $O(h^{-1}|\log_2 h|^{d-1})$ and remain a L^2 convergence rate of $O(h^{k+1/2}|\log_2 h|^{d})$ for transport equations and a convergence rate of $O(h^{k}|\log_2 h|^{d})$ in the energy norm for elliptic equations.

In Chapter 4, we incorporate sparse grid DG method with central DG scheme to develop a class of conservative numerical schemes with high computational efficiency for high-dimensional hyperbolic equations. Our work consists of construction of the sparse finite element space , L^2 stability and error estimates, and numerical validation of the scheme.

Chapter 2

An UWDG method for Schrödinger equation in one dimension

In this chapter, we develop and analyze a new ultra-weak discontinuous Galerkin (UWDG) method for solving one-dimensional nonlinear Schrödinger (NLS) equations (1.1). The method solves the equation without introducing any auxiliary variables or rewriting the equation into a larger system.

The focus of this chapter is on the investigation of a most general form of the numerical flux functions that ensures stability along with our ultra-weak formulation. The fluxes under consideration include the alternating fluxes, and also the fluxes considered in [52], and therefore allows for flexibility for the design of the schemes. The analysis in this chapter relies on a detailed analysis of a special projection associated with different flux parameters, whose dependence on mesh size can be freely enforced. Under certain flux parameters, the projection can be defined locally. For other flux parameters, the projection is global and the projection analysis is based on a block-circulant matrix with 2×2 blocks. Our analysis reveals that under a large class of parameter choices, the UWDG method is optimally convergent in L^2 norm, which is verified by extensive numerical tests for both the projection operators and the numerical schemes for (1.1).

The remainder of this chapter is organized as follows. In Section 2.1, we introduce the

UWDG method with general flux definitions for one-dimensional nonlinear Schrödinger equations and study its stability properties. We introduce a new projection operator and analyze its properties in Section 2.2, which is later used in Section 2.3 to obtain the convergence results of the schemes. The main body of this chapter, the error estimates, is contained in Section 2.3. Numerical validations are provided in Section 2.4. Some technical details, including proof of most lemmas are collected in Appendix.

The major contents of this chapter has been published in [16].

2.1 Numerical scheme and stability

In this subsection, we formulate and discuss stability results of a DG scheme for one-dimensional NLS equation (1.1) on interval $I = [a, b]$ with initial condition $u(x, 0) = u_0(x)$ and periodic boundary conditions. Here $f(u)$ is a given real function. Our method can be defined for general boundary conditions, but the error analysis will require slightly different tools, and therefore we only consider periodic boundary conditions in this chapter.

To facilitate the discussion, first we introduce some notations and definitions. For a 1-D interval $I = [a, b]$, the usual DG meshes are defined as:

$$a = x_{\frac{1}{2}} < x_{\frac{3}{2}} < \cdots < x_{N+\frac{1}{2}} = b,$$

$$I_j = (x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}), \quad x_j = \frac{1}{2}(x_{j-\frac{1}{2}} + x_{j+\frac{1}{2}}),$$

and

$$h_j = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}, \quad h = \max_j h_j,$$

with mesh regularity requirement $\frac{h}{\min h_j} < \sigma$, σ is fixed during mesh refinement.

Denote $\mathbb{Z}_N = 1, 2, \cdots, N$. The approximation space is defined as:

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

meaning v_h is a piecewise polynomial of x with degree up to k on each cell I_j. For a function $v_h \in V_h^k$, we use $(v_h)^-_{j-\frac{1}{2}}$ and $(v_h)^+_{j-\frac{1}{2}}$ to refer to the value of v_h at $x_{j-\frac{1}{2}}$ from the left cell I_{j-1} and the right cell I_j respectively. The jump and average are defined as $[v_h] = v_h^+ - v_h^-$ and $\{v_h\} = \frac{1}{2}(v_h^+ + v_h^-)$ at cell interfaces.

Throughout this chapter, we use the standard Sobolev norm notations $\|\cdot\|_{W^{s,p}(I)}$ and broken Sobolev space on mesh $\mathcal{I}_N$. We denote $\|v\|^2_{H^s(\mathcal{I}_N)} = \sum_{j=1}^N \|v\|^2_{H^s(I_j)}$ and $\|v\|_{W^{s,\infty}(\mathcal{I}_N)} = \max_j \|v\|_{W^{s,\infty}(I_j)}$. In Section 3.4, we consider negative norms and the definition is $\|v\|_{H^{-l}(I)} = \sup_{\Phi \in \mathcal{C}_0^\infty(I)} \frac{\int_I v(x)\Phi(x)dx}{\|\Phi\|_{H^l(I)}}$. Additionally, we denote by $\|v\|_{L^2(\partial\mathcal{I}_N)}$ the broken L^2 norm on cell interfaces, i.e., $\|v\|^2_{L^2(\partial\mathcal{I}_N)} = \sum_{j=1}^N \|v\|^2_{L^2(\partial I_j)}$, where$\|v\|^2_{L^2(\partial I_j)} = (v^-_{x_{j+\frac{1}{2}}})^2 + (v^+_{x_{j-\frac{1}{2}}})^2$. We also denote $\|\cdot\| = \|\cdot\|_{L^2(I)} = \|\cdot\|_{L^2(\mathcal{I}_N)}$ to shorten the notation.

Lastly, we recall inverse inequalities, $\forall v_h \in V_h^k$,

$$\begin{aligned} \|(v_h)_x\|_{L^2(\mathcal{I}_j)} \le Ch_j^{-1}\|v_h\|_{L^2(\mathcal{I}_j)}, \quad \|v_h\|_{L^2(\partial\mathcal{I}_j)} &\le Ch^{-\frac{1}{2}}\|v_h\|_{L^2(\mathcal{I}_j)}, \\ \|v_h\|_{L^\infty(\mathcal{I}_j)} &\le Ch^{-\frac{1}{2}}\|v_h\|_{L^2(\mathcal{I}_j)}, \end{aligned} \tag{2.1}$$

and trace inequalities

$$\|v\|^2_{L^2(\partial I_j)} \le Ch_j^{-1}\|v\|^2_{L^2(I_j)}, \tag{2.2}$$

here and below C is a constant independent of the function u and the mesh size h.

In this chapter, we consider a DG scheme motivated by [18] and based on integration by parts twice, or the so-called ultra-weak formulation. In particular, we look for the unique function $u_h = u_h(t) \in V_h^k$, $t \in (0, T]$, such that

$$
\begin{aligned}
i \int_{I_j} (u_h)_t v_h dx + \int_{I_j} u_h (v_h)_{xx} dx - \hat{u}_h (v_h)_x^- |_{j+\frac{1}{2}} + \hat{u}_h (v_h)_x^+ |_{j-\frac{1}{2}} \\
+ \widetilde{(u_h)_x} v_h^- |_{j+\frac{1}{2}} - \widetilde{(u_h)_x} v_h^+ |_{j-\frac{1}{2}} + \int_{I_j} f(|u_h|^2) u_h v_h dx = 0
\end{aligned} \tag{2.3}
$$

holds for all $v_h \in V_h^k$ and all $j = 1, \cdots, N$. Here, we require $k \geq 1$, because $k = 0$ yields an inconsistent scheme. Notice that (2.3) can be written equivalently in a weak formulation by performing another integration by parts back as:

$$
\begin{aligned}
i \int_{I_j} (u_h)_t v_h dx - \int_{I_j} (u_h)_x (v_h)_x dx + (u_h^- - \hat{u}_h)(v_h)_x^- |_{j+\frac{1}{2}} + (\hat{u}_h - u_h^+)(v_h)_x^+ |_{j-\frac{1}{2}} \\
+ \widetilde{(u_h)_x} v_h^- |_{j+\frac{1}{2}} - \widetilde{(u_h)_x} v_h^+ |_{j-\frac{1}{2}} + \int_{I_j} f(|u_h|^2) u_h v_h dx = 0
\end{aligned} \tag{2.4}
$$

The "hat" and"tilde" terms are the numerical fluxes we pick for u and u_x at cell boundaries, which are single valued functions defined as:

$$
\widetilde{(u_h)_x} = \{(u_h)_x\} + \alpha_1 [(u_h)_x] + \beta_1 [u_h], \quad \hat{u}_h = \{u_h\} + \alpha_2 [u_h] + \beta_2 [(u_h)_x], \tag{2.5}
$$

where $\alpha_1, \alpha_2, \beta_1, \beta_2$ are prescribed complex parameters. They may depend on the mesh parameter h. Commonly used fluxes such as the central flux (by setting $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0$) and alternating fluxes (by setting $\alpha_1 = -\alpha_2 = \pm\frac{1}{2}, \beta_1 = \beta_2 = 0$) belong to this flux family. The direct DG scheme considered in [52] is a special case of our method when $\alpha_1 = -\alpha_2, \beta_1 = \frac{c}{h}, \beta_2 = 0, c > 0, \alpha_1 \in \mathbb{R}$. The IPDG method can also be casted in this framework as $\alpha_1 = \alpha_2 = \beta_2 = 0, \beta_1 = \frac{c}{h}, c > 0$.

We write the scheme (2.3) in following short-hand notation:

$$a_j(u_h, v_h) - i\int_{I_j} f(|u_h|^2)u_h v_h dx = 0, \quad \forall j \in \mathbb{Z}_N \tag{2.6}$$

holds for all $v_h \in V_h^k$, where

$$a_j(u_h, v_h) = \int_{I_j} (u_h)_t v_h dx - iA_j(u_h, v_h),$$

with

$$A_j(u_h, v_h) = \int_{I_j} u_h(v_h)_{xx}dx - \hat{u}_h(v_h)_x^-|_{j+\frac{1}{2}} + \hat{u}_h(v_h)_x^+|_{j-\frac{1}{2}} + \widetilde{(u_h)_x}v_h^-|_{j+\frac{1}{2}} - \widetilde{(u_h)_x}v_h^+|_{j-\frac{1}{2}}$$

as the UWDG spatial discretization for the second order derivative term.

Using periodic boundary condition, we sum up on j for (2.6) and get

$$a(u_h, v_h) - i\int_I f(|u_h|^2)u_h v_h dx = 0, \tag{2.7}$$

where

$$\begin{aligned} a(u_h, v_h) &= \int_I (u_h)_t v_h dx - iA(u_h, v_h), \\ A(u_h, v_h) &= \sum_{j=1}^{N} A_j(u_h, v_h) = \int_I u_h(v_h)_{xx}dx + \sum_{j=1}^{N}\Big(\hat{u}_h[(v_h)_x] - \widetilde{(u_h)_x}[v_h]\Big)\,\Big|_{j+\frac{1}{2}}. \end{aligned}$$

The following theorem contains the results on semi-discrete L^2 stability.

Theorem 2.1.1. (Stability) *For $u, v \in H^2(\mathcal{I}_N)$ satisfying periodic boundary condition, we have $A(u, v) = A(v, u)$.*

The solution of semi-discrete UWDG scheme (2.3) *using numerical fluxes* (2.5) *satisfies L^2 stability condition*

$$\frac{d}{dt}\int_I |u_h|^2 dx \leq 0,$$

if

$$\mathrm{Im}\beta_2 \geq 0,\ \mathrm{Im}\beta_1 \leq 0,\ |\alpha_1 + \overline{\alpha_2}|^2 \leq -4\mathrm{Im}\beta_1 \mathrm{Im}\beta_2. \tag{2.8}$$

In particular, when all parameters $\alpha_1, \alpha_2, \beta_1, \beta_2$ are restricted to be real, this condition amounts to

$$\alpha_1 + \alpha_2 = 0 \tag{2.9}$$

without any requirement on β_1, β_2.

Proof. From integration by parts, we have

$$A(u,v) = -\int_I u_x v_x dx + \sum_{j=1}^{N} \left(\hat{u}[v_x] - [uv_x] - \widetilde{u_x}[v]\right)\Big|_{j+\frac{1}{2}}.$$

Similarly, $A(v,u) = -\int_I u_x v_x dx + \sum_{j=1}^{N} \left(\hat{v}[u_x] - [vu_x] - \widetilde{v_x}[u]\right)\Big|_{j+\frac{1}{2}}$. Plugging in the definition of the numerical fluxes in (2.5), we have at $x_{j+\frac{1}{2}}, \forall j \in \mathbb{Z}_N$

$$\begin{aligned}
\hat{u}[v_x] - [uv_x] - \widetilde{u_x}[v] &= \big(\{u\} - \alpha_1[u] + \beta_2[u_x]\big)[v_x] - (\{u\}[v_x] + [u]\{v_x\}) \\
&- \big(\{u_x\} + \alpha_1[u_x] + \beta_1[u]\big)[v] \\
&= [u_x]\big(\{v\} - \alpha_1[v] + \beta_2[v_x]\big) - [u]\big(\{v_x\} + \alpha_1[v_x] + \beta_1[v]\big) \\
&- \big([u_x]\{v\} + \{u_x\}[v]\big) \\
&= [u_x]\hat{v} - [u]\widetilde{v_x} - [u_x v],
\end{aligned}$$

then the proof for $A(u,v) = A(v,u)$ is complete.

From integration by parts, we have, for $\forall v_h \in V_h^k$

$$a(u_h, v_h) = \int_I (u_h)_t v_h dx + i \int_I (u_h)_x (v_h)_x dx + i \sum_{j=1}^{N} ([u_h (v_h)_x] - \hat{u}_h [(v_h)_x] + \widetilde{(u_h)_x} [v_h])|_{j+\frac{1}{2}}.$$

$$A(u_h, v_h) = -\int_I (u_h)_x (v_h)_x dx + \sum_{j=1}^{N} (\hat{u}_h [(v_h)_x] - [u_h (v_h)_x] - \widetilde{(u_h)_x} [v_h])|_{j+\frac{1}{2}}.$$

Taking $v_h = \bar{u}_h$ in (2.7) and compute its conjugate as well, we get

$$\begin{aligned} 0 &= i \int_I f(|u_h|^2)|u_h|^2 dx + \overline{i \int_I f(|u_h|^2)|u_h|^2 dx} \\ &= a(u_h, \bar{u}_h) + \overline{a(u_h, \bar{u}_h)} \\ &= \frac{d}{dt} \int_I |u_h|^2 dx - iA(u_h, \overline{u_h}) + i\overline{A(u_h, \overline{u_h})}. \end{aligned} \tag{2.10}$$

$$\begin{aligned} -iA(u_h, \overline{u_h}) + i\overline{A(u_h, \overline{u_h})} &= -2\mathrm{Im} \sum_{j=1}^{N} ([u_h (\bar{u}_h)_x] - \hat{u}_h [(\bar{u}_h)_x] + \widetilde{(u_h)_x} [\bar{u}_h])|_{j+\frac{1}{2}} \\ &= -2\mathrm{Im} \sum_{j=1}^{N} \Big(\{u_h\}[(\bar{u}_h)_x] + [u_h]\{(\bar{u}_h)_x\} - \big(\{u_h\} + \alpha_2 [u_h] + \beta_2 [(u_h)_x]\big)[(\bar{u}_h)_x] \\ &\quad + (\{(u_h)_x\} + \alpha_1 [(u_h)_x] + \beta_1 [u_h])[\bar{u}_h] \Big)|_{j+\frac{1}{2}} \\ &= -2\mathrm{Im} \sum_{j=1}^{N} \big(-\beta_2 |[(u_h)_x]|^2 + \beta_1 |[u_h]|^2 + \alpha_1 [(u_h)_x][\bar{u}_h] - \alpha_2 [u_h][(\bar{u}_h)_x] \big)|_{j+\frac{1}{2}} \\ &= 2\mathrm{Im} \sum_{j=1}^{N} (\beta_2 |[(u_h)_x]|^2 - \beta_1 |[u_h]|^2 - (\alpha_1 + \overline{\alpha_2})[\bar{u}_h][(u_h)_x])|_{j+\frac{1}{2}} \end{aligned}$$

Plug it back into (2.10):

$$\frac{d}{dt}\int_I |u_h|^2 dx + \sum_{j=1}^{N} 2\mathrm{Im}\beta_2|[(u_h)_x]|^2 - 2\mathrm{Im}\beta_1|[u_h]|^2 - 2\mathrm{Im}\{(\alpha_1+\overline{\alpha_2})[\bar{u}_h][(u_h)_x]\}|_{j+\frac{1}{2}} = 0. \tag{2.11}$$

If the stability condition (2.8) is satisfied, we have

$$\frac{d}{dt}\int_I |u_h|^2 dx \leq 0.$$

If all parameters are real and (2.9) is satisfied, then (2.11) further yields:

$$\frac{d}{dt}\int_I |u_h|^2 dx = 0, \tag{2.12}$$

which implies energy conservation. □

For simplicity of the discussion, in the contents below, we will only consider real parameters, i.e. when $\alpha_1, \alpha_2, \beta_1, \beta_2$ are real and $\alpha_1 + \alpha_2 = 0$. This property of our scheme is consistent with the energy conservation property of Schrödinger equations. It is essential to have a symmetric $A(u_h, v_h)$ for designing a finite element scheme which is energy-preserving for Schrödinger equations.

Now the numerical fluxes are defined by three parameters as,

$$\widetilde{(u_h)_x} = \{(u_h)_x\} + \alpha_1[(u_h)_x] + \beta_1[u_h], \quad \hat{u}_h = \{u_h\} - \alpha_1[u_h] + \beta_2[(u_h)_x], \quad \alpha_1, \beta_1, \beta_2 \in \mathbb{R}. \tag{2.13}$$

Note we can rewrite the flux definition in a matrix form

$$\begin{bmatrix} \hat{u}_h \\ \widetilde{(u_h)_x} \end{bmatrix} = G \begin{bmatrix} u_h^- \\ (u_h)_x^- \end{bmatrix} + H \begin{bmatrix} u_h^+ \\ (u_h)_x^+ \end{bmatrix}, \quad G = \begin{bmatrix} \frac{1}{2}+\alpha_1 & -\beta_2 \\ -\beta_1 & \frac{1}{2}-\alpha_1 \end{bmatrix}, \; H = \begin{bmatrix} \frac{1}{2}-\alpha_1 & \beta_2 \\ \beta_1 & \frac{1}{2}+\alpha_1 \end{bmatrix}, \tag{2.14}$$

where I_2 denotes the 2×2 identity matrix. Note that $G + H = I_2, \det G = \det H = -(\alpha_1^2 + \beta_1\beta_2 - \frac{1}{4})$ and $GH = -(\det G)I_2$.

2.2 Projection $P_h^\star$

In this section, we perform detailed studies of a projection operator that is key to the analysis of the UWDG scheme.

Definition 2.2.1. *For the UWDG scheme with flux choice* (2.13), *we define the associated projection operator $P_h^\star$ for any periodic function $u \in W^{1,\infty}(I)$ to be the unique polynomial $P_h^\star u \in V_h^k$ (when $k \geq 1$) satisfying*

$$\int_{I_j} P_h^\star u \, v_h dx = \int_{I_j} u \, v_h dx \qquad \forall v_h \in P^{k-2}(I_j), \tag{2.15a}$$

$$\widehat{P_h^\star u} = \{P_h^\star u\} - \alpha_1[P_h^\star u] + \beta_2[(P_h^\star u)_x] = u \qquad at \quad x_{j+\frac{1}{2}}, \tag{2.15b}$$

$$\widetilde{(P_h^\star u)_x} = \{(P_h^\star u)_x\} + \alpha_1[(P_h^\star u)_x] + \beta_1[P_h^\star u] = u_x \qquad at \quad x_{j+\frac{1}{2}}, \tag{2.15c}$$

for all j. When $k = 1$, only conditions (2.15b)-(2.15c) *are needed.*

This definition is to ensure $u - \widehat{P_h^\star u} = 0$ and $u_x - \widetilde{(P_h^\star u)_x} = 0$ at cell boundaries, which will be used in error estimates for the scheme. In the following, we analyze the projection when the parameter choice reduces it to a local projection in Section 2.2.1, and then we consider the more general global projection in Section 2.2.2.

We can write (2.15b)-(2.15c) in vector form as

$$\left.\begin{bmatrix} \widehat{P_h^\star u} \\ \widetilde{(P_h^\star u)_x} \end{bmatrix}\right|_{x_{j+\frac{1}{2}}} = G \left.\begin{bmatrix} P_h^\star u \\ (P_h^\star u)_x \end{bmatrix}\right|^{-}_{x_{j+\frac{1}{2}}} + H \left.\begin{bmatrix} P_h^\star u \\ (P_h^\star u)_x \end{bmatrix}\right|^{+}_{x_{j+\frac{1}{2}}} = \left.\begin{bmatrix} u \\ u_x \end{bmatrix}\right|_{x_{j+\frac{1}{2}}}. \tag{2.16}$$

2.2.1 Local projection condition

In general, the projection $P_h^\star$ is globally defined, and its existence, uniqueness and approximation properties are quite complicated mathematically. However, with some special parameter choices, $P_h^\star$ can be reduced to a local projection, meaning that it can be solved element-wise, and hence the analysis can be greatly simplified.

For example, with the alternating fluxes $\alpha_1 = \pm\frac{1}{2}, \beta_1 = \beta_2 = 0$, $P_h^\star$ can be reduced to P_h^1 and P_h^2 defined below. $P_h^\star = P_h^1$ for parameter choice $\alpha_1 = \frac{1}{2}, \beta_1 = \beta_2 = 0$ is formulated as: for each cell I_j, we find the unique polynomial of degree k, $P_h^1 u$, satisfying

$$\int_{I_j} P_h^1 u \, v_h dx = \int_{I_j} u \, v_h dx \qquad \forall v_h \in P^{k-2}(I_j), \tag{2.17a}$$

$$(P_h^1 u)^- = u \qquad \text{at } x_{j+\frac{1}{2}}, \tag{2.17b}$$

$$(P_h^1 u)_x^+ = u_x \qquad \text{at } x_{j-\frac{1}{2}}. \tag{2.17c}$$

When $k = 1$, only conditions (2.17b)-(2.17c) are needed.

Similarly, we can define $P_h^\star = P_h^2$ for parameter choice $\alpha_1 = -\frac{1}{2}, \beta_1 = \beta_2 = 0$ as: for each cell I_j, we find the unique polynomial of degree k, $P_h^2 u$, satisfying

$$\int_{I_j} P_h^2 u\, v_h dx = \int_{I_j} u\, v_h dx \qquad \forall v_h \in P^{k-2}(I_j), \tag{2.18a}$$

$$(P_h^2 u)^+ = u \qquad \text{at } x_{j-\frac{1}{2}}, \tag{2.18b}$$

$$(P_h^2 u)_x^- = u_x \qquad \text{at } x_{j+\frac{1}{2}}. \tag{2.18c}$$

When $k = 1$, only conditions (2.18b)-(2.18c) are needed.

Similar local projections have been introduced and considered in [18]. It is obvious that $P_h^1 u, P_h^2 u$ can be solved element-wise, and their existence, uniqueness are straightforward. From a standard scaling argument by Bramble-Hilbert lemma in [20], P_h^1 and P_h^2 have the following error estimates: let $u \in W^{k+1,p}(I_j)(p = 2, \infty)$, then

$$\begin{aligned} \|u - P_h^\nu u\|_{L^p(I_j)} &\le C h_j^{k+1} |u|_{W^{k+1,p}(I_j)}, \quad p = 2, \infty,\ \nu = 1, 2, \\ \|u_x - P_h^\nu u_x\|_{L^p(I_j)} &\le C h_j^{k} |u|_{W^{k+1,p}(I_j)}, \quad p = 2, \infty,\ \nu = 1, 2, \end{aligned} \tag{2.19}$$

where here and below, C is a generic constant that is independent of the mesh size h_j, the parameters $\alpha_1, \beta_1, \beta_2$ and the function u, but may take different value in each occurrence.

Naturally, the immediate question is that if there are other parameter choices such that $P_h^\star$ can be reduced to a local projection. The following lemma addresses this issue.

Lemma 2.2.1 (The condition for reduction to a local projection)**.** *If $\alpha_1^2 + \beta_1\beta_2 = \frac{1}{4}$, $P_h^\star$ is* *a local projection. Moreover,* (2.15b) *and* (2.15c) *is equivalent to*

$$G\begin{bmatrix} P_h^\star u \\ (P_h^\star u)_x \end{bmatrix}\Bigg|^{-}_{x_{j+\frac{1}{2}}} + H\begin{bmatrix} P_h^\star u \\ (P_h^\star u)_x \end{bmatrix}\Bigg|^{+}_{x_{j-\frac{1}{2}}} = G\begin{bmatrix} u \\ u_x \end{bmatrix}\Bigg|_{x_{j+\frac{1}{2}}} + H\begin{bmatrix} u \\ u_x \end{bmatrix}\Bigg|_{x_{j-\frac{1}{2}}}. \tag{2.20}$$

Proof. The definition (2.15a) provides $k-1$ linearly independent equations for solving $P_h^\star u$ on each cell. If (2.15b) and (2.15c) can be locally decoupled, $P_h^\star$ is a local projection. By assumption $\alpha_1^2+\beta_1\beta_2=\frac{1}{4}$, if $\beta_1=\beta_2=0$, then $\alpha_1=\pm\frac{1}{2}$ and $P_h^\star u = P_h^1$ or P_h^2, and (2.20) holds. The rest of the cases are

- if $\beta_1\neq 0$, left multiply (2.16) by a matrix, we have

$$\begin{bmatrix} \beta_1 & \frac{1}{2}+\alpha_1 \\ \beta_1 & -(\frac{1}{2}-\alpha_1) \end{bmatrix}\begin{bmatrix} u \\ u_x \end{bmatrix}\Bigg|_{x_{j+\frac{1}{2}}} = \begin{bmatrix} 0 & 0 \\ \beta_1 & -(\frac{1}{2}-\alpha_1) \end{bmatrix}\begin{bmatrix} P_h^\star u \\ (P_h^\star u)_x \end{bmatrix}\Bigg|^{-}_{x_{j+\frac{1}{2}}} + \begin{bmatrix} \beta_1 & \frac{1}{2}+\alpha_1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} P_h^\star u \\ (P_h^\star u)_x \end{bmatrix}\Bigg|^{+}_{x_{j+\frac{1}{2}}},$$

  which implies the following decoupled relations

$$\begin{aligned} (P_h^\star u)^+ + \frac{\frac{1}{2}+\alpha_1}{\beta_1}(P_h^\star u)_x^+ &= u + \frac{\frac{1}{2}+\alpha_1}{\beta_1}u_x \quad \text{at } x_{j-\frac{1}{2}}, \\ (P_h^\star u)^- - \frac{\frac{1}{2}-\alpha_1}{\beta_1}(P_h^\star u)_x^- &= u - \frac{\frac{1}{2}-\alpha_1}{\beta_1}u_x \quad \text{at } x_{j+\frac{1}{2}}. \end{aligned} \tag{2.21}$$

- if $\beta_2 \neq 0$, by similar linear transformation, we have

$$\begin{bmatrix} \frac{1}{2}-\alpha_1 & \beta_2 \\ -(\frac{1}{2}+\alpha_1) & \beta_2 \end{bmatrix} \begin{bmatrix} u \\ u_x \end{bmatrix}\Bigg|_{x_{j+\frac{1}{2}}} = \begin{bmatrix} 0 & 0 \\ -(\frac{1}{2}+\alpha_1) & \beta_2 \end{bmatrix} \begin{bmatrix} P_h^\star u \\ (P_h^\star u)_x \end{bmatrix}\Bigg|_{x_{j+\frac{1}{2}}}^{-} + \begin{bmatrix} \frac{1}{2}-\alpha_1 & \beta_2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P_h^\star u \\ (P_h^\star u)_x \end{bmatrix}\Bigg|_{x_{j+\frac{1}{2}}}^{+},$$

which implies

$$\begin{aligned} (P_h^\star u)_x^+ + \frac{\frac{1}{2}-\alpha_1}{\beta_2}(P_h^\star u)^+ &= u_x + \frac{\frac{1}{2}-\alpha_1}{\beta_2}u \quad \text{at } x_{j-\frac{1}{2}}, \\ (P_h^\star u)_x^- - \frac{\frac{1}{2}+\alpha_1}{\beta_2}(P_h^\star u)^- &= u_x - \frac{\frac{1}{2}+\alpha_1}{\beta_2}u \quad \text{at } x_{j+\frac{1}{2}}. \end{aligned} \tag{2.22}$$

(2.21), (2.22) are the desired decoupled conditions on each cell I_j, and it's easy to verify that (2.21), (2.22) are equivalent to (2.20). Therefore the proof is complete. □

This lemma implies that for any parameter satisfying $\alpha_1^2+\beta_1\beta_2 = \frac{1}{4}$, $P_h^\star$ is locally defined. We remark that this condition turns out to be the same as the optimally convergent numerical flux families in [17] for two-way wave equations, although they arise in different contexts. Unfortunately, for the general definition of $P_h^\star$, unlike P_h^1 and P_h^2, we cannot directly use the Bramble-Hilbert lemma and the standard scaling argument to obtain optimal approximation property, since the second and third relations in (2.21) and (2.22) may break the scaling.

2.2.2 $P_h^\star$ properties

Before moving on to detailed discussion on $P_h^\star$, we introduce some notations to facilitate the discussion. We define the Legendre expansion of a function $u \in L^2(I)$ on cell I_j as follows,

$$u|_{I_j} = \sum_{m=0}^{\infty} u_{j,m} L_{j,m}(x), \tag{2.23}$$

where $L_{j,m}(x) := L_m(\xi), \xi = \frac{x-x_j}{h_j/2}$, and $L_m(\cdot)$ is the standard Legendre polynomial of degree m on $[-1,1]$. In what follows, we write $L_{j,m}(x)$ as $L_{j,m}$, and $L_m(\xi)$ as L_m for notational convenience. We can compute $u_{j,m}$ using orthogonality of Legendre polynomials and Rodrigues' formula,

$$\begin{aligned}
u_{j,m} &= \frac{2m+1}{h_j} \int_{I_j} u(x) L_{j,m} dx = \frac{2m+1}{2} \int_{-1}^{1} \hat{u}_j(\xi) L_m d\xi \\
&= \frac{2m+1}{2} \frac{1}{2^m m!} \int_{-1}^{1} \hat{u}_j(\xi) \frac{d}{d\xi^m} (\xi^2-1)^m d\xi \\
&= \frac{2m+1}{2} \frac{(-1)^l}{2^m m!} \int_{-1}^{1} \frac{d}{d\xi^l} \hat{u}_j(\xi) \frac{d}{d\xi^{m-l}} (\xi^2-1)^m d\xi,
\end{aligned} \tag{2.24}$$

where $\hat{u}_j(\xi) = u(x(\xi))$ is defined as the function $u|_{I_j}$ transformed to the reference domain $[-1,1]$. By Holder's inequality, if $u \in W^{l,p}(I)$,

$$|u_{j,m}| \le C h_j^{l-\frac{1}{p}} |u|_{W^{l,p}(I_j)}, \quad 0 \le l \le m. \tag{2.25}$$

The L^2 projection P_h^0 is closely related to $u_{j,m}$. By orthogonality of Legendre polynomials, we have

$$P_h^0 u = \sum_{m=0}^{k} u_{j,m} L_{j,m}.$$

We collect some frequently used notations in Table 2.1 for quick reference.

Table 2.1: Notations for some frequently used quantities. Subscript j will be dropped for uniform mesh.

Notation	Definition	Notation	Definition
G	$\begin{bmatrix} \frac{1}{2}+\alpha_1 & -\beta_2 \\ -\beta_1 & \frac{1}{2}-\alpha_1 \end{bmatrix}$	H	$\begin{bmatrix} \frac{1}{2}-\alpha_1 & \beta_2 \\ \beta_1 & \frac{1}{2}+\alpha_1 \end{bmatrix}$
Γ_j	$\beta_1 + \frac{\beta_2}{h_j^2}k^2(k^2-1) - \frac{2k^2}{h_j}(\alpha_1^2+\beta_1\beta_2+\frac{1}{4})$	Λ_j	$-\frac{2k}{h_j}(\alpha_1^2+\beta_1\beta_2-\frac{1}{4})$
$L^-_{j,m}$	$\begin{bmatrix} L_{j,m}(x_{j+\frac{1}{2}}) \\ \frac{2}{h_j}\frac{d}{dx}L_{j,m}(x_{j+\frac{1}{2}}) \end{bmatrix}$	$L^+_{j,m}$	$\begin{bmatrix} L_{j,m}(x_{j-\frac{1}{2}}) \\ \frac{2}{h_j}\frac{d}{dx}L_{j,m}(x_{j-\frac{1}{2}}) \end{bmatrix}$
A_j	$G[L^-_{j,k-1}, L^-_{j,k}]$	B_j	$H[L^+_{j,k-1}, L^+_{j,k}]$
Q	$-A^{-1}B$	r_l	$Q^l(I_2-Q^N)^{-1}$
$\mathcal{M}_{j,m}$	$(A_j+B_j)^{-1}(GL^-_{j,m}+HL^+_{j,m})$		

Next, we will write the explicit formula of $P_h^\star$ in order to get a clear view of the existence and uniqueness condition, as well as the error estimates. Suppose

$$P_h^\star u = \sum_{m=0}^{k} \acute{u}_{j,m} L_{j,m}.$$

By the definition (2.15a), $\acute{u}_{j,m} = u_{j,m}, m \le k-2$, i.e.,

$$P_h^\star u = \sum_{m=0}^{k-2} u_{j,m} L_{j,m} + \sum_{m=k-1}^{k} \acute{u}_{j,m} L_{j,m}.$$

In what follows, we analyze the existence and uniqueness of $P_h^\star$, i.e., the existence and uniqueness of $\acute{u}_{j,k-1}, \acute{u}_{j,k}$ based on the following assumptions on parameters:

- A0. (Local projection) $\alpha_1^2 + \beta_1\beta_2 = \frac{1}{4}$ and $\Gamma_j \neq 0$.
- A1. (Global projection) uniform mesh ($h_j = h, \forall j$), $\alpha_1^2 + \beta_1\beta_2 \neq \frac{1}{4}$ and $\left|\frac{\Gamma}{\Lambda}\right| > 1$.
- A2. (Global projection) uniform mesh ($h_j = h, \forall j$), $\alpha_1^2 + \beta_1\beta_2 \neq \frac{1}{4}$, $\left|\frac{\Gamma}{\Lambda}\right| = 1$. $\left((-1)^{k+1}\frac{\Gamma}{\Lambda}\right)^N \neq 1$. If N is odd, and if k is odd, we require $\Gamma = -\Lambda$; if k is even,

we require $\Gamma = \Lambda$.

- A3. (Global projection) uniform mesh ($h_j = h, \forall j$), $\alpha_1^2 + \beta_1\beta_2 \neq \frac{1}{4}$, $\left|\frac{\Gamma}{\Lambda}\right| < 1$, $\left((-1)^{k+1}\frac{\Gamma}{\Lambda} + \sqrt{\left(\frac{\Gamma}{\Lambda}\right)^2 - 1}\right)^N \neq 1$.

Lemma 2.2.2 ($P_h^\star$ existence, uniqueness and formula)**.** *If any of the assumptions above is satisfied, $P_h^\star$ exists and is uniquely defined. Furthermore, if assumption A0 is satisfied, then*

$$\begin{bmatrix} \acute{u}_{j,k-1} \\ \acute{u}_{j,k} \end{bmatrix} = \begin{bmatrix} u_{j,k-1} \\ u_{j,k} \end{bmatrix} + \sum_{m=k+1}^{\infty} u_{j,m}\mathcal{M}_{j,m}. \tag{2.26}$$

If any of the assumptions A1/A2/A3 is satisfied, then

$$\begin{bmatrix} \acute{u}_{j,k-1} \\ \acute{u}_{j,k} \end{bmatrix} = \begin{bmatrix} u_{j,k-1} \\ u_{j,k} \end{bmatrix} + \sum_{m=k+1}^{\infty} \big(u_{j,m}V_{1,m} + \sum_{l=0}^{N-1} u_{j+l,m}r_l V_{2,m}\big), \tag{2.27}$$

where $V_{1,m} = [L_{k-1}^+, L_k^+]^{-1}L_m^+$, $V_{2,m} = [L_{k-1}^-, L_k^-]^{-1}L_m^- - [L_{k-1}^+, L_k^+]^{-1}L_m^+$, $u_{j+l} = u_{j+l-N}$ when $j + l \geq N$, and r_l is defined in Table 2.1.

Proof. The proof of this lemma can be found in Appendix. □

Lemma 2.2.3. *Suppose any of the assumptions A0/A1/A2/A3 holds and u satisfies the condition in Definition 2.2.1. For $p = 2, \infty$, if assumption A0 is satisfied,*

$$\|P_h^\star u - u\|_{L^p(I)} \leq Ch^{k+1}|u|_{W^{k+1,\infty}(I)} \frac{\max\left(|\beta_1|, \frac{|\frac{1}{2}-\alpha_1|}{h}, \frac{|\frac{1}{2}+\alpha_1|}{h}, \frac{|\beta_2|}{h^2}\right)}{\min_j |\Gamma_j|}. \tag{2.28}$$

If assumption A1 is satisfied,

$$\|P_h^\star u - u\|_{L^p(I)} \le Ch^{k+1}|u|_{W^{k+1,\infty}(I)}(1 + \frac{\|Q_1\|_\infty}{|1-|\lambda_1||} + \frac{\|I_2 - Q_1\|_\infty}{|1-|\lambda_2||}), \tag{2.29}$$

where λ_1, λ_2 *are defined in* (49), Q_1 *is defined in* (56) *and* (57).

If assumption A2 is satisfied,

$$\|P_h^\star u - u\|_{L^p(I)} \le Ch^{k+1}\|u\|_{W^{k+4,\infty}(I)}(1 + \frac{\|Q_2\|_\infty}{|\Gamma|}), \tag{2.30}$$

where Q_2 *is defined in* (59).

If assumption A3 is satisfied, and assuming $\left|1-\lambda_1^N\right| = O(h^{\delta'})$, $|1-\lambda_1| = O(h^{\delta/2})$ *with* $0 \le \delta \le 2$,

$$\|P_h^\star u - u\|_{L^p(I)} \le Ch^{k+1}\|u\|_{W^{k+3,\infty}(I)}(1 + h^{-\delta'-\delta/2}(\|Q_1\|_\infty + \|I_2 - Q_1\|_\infty)). \tag{2.31}$$

Proof. Proof is given in Appendix. □

Above estimates provides error bound that can be computed once the parameters are given, yet its dependence on the mesh size h is not fully revealed, particularly when the parameters $\alpha_1, \beta_1, \beta_2$ also have h-dependence. To clarify such relations, next we will interpret (2.29) when considering the following common choice of parameters, where α_1 has no dependence on h, $\beta_1 = \tilde{\beta}_1 h^{p_1}, \beta_2 = \tilde{\beta}_2 h^{p_2}, \tilde{\beta}_1, \tilde{\beta}_2$ are nonzero constants that do not depend on h. If indeed β_1 or β_2 is zero, it is equivalent to let $p_1, p_2 \to +\infty$ in the discussions below. We will discuss whether the parameter choice yields optimal $(k+1)$-th order accuracy.

To distinguish different cases, we illustrate the choice of parameters p_1, p_2 in Figure 2.1.

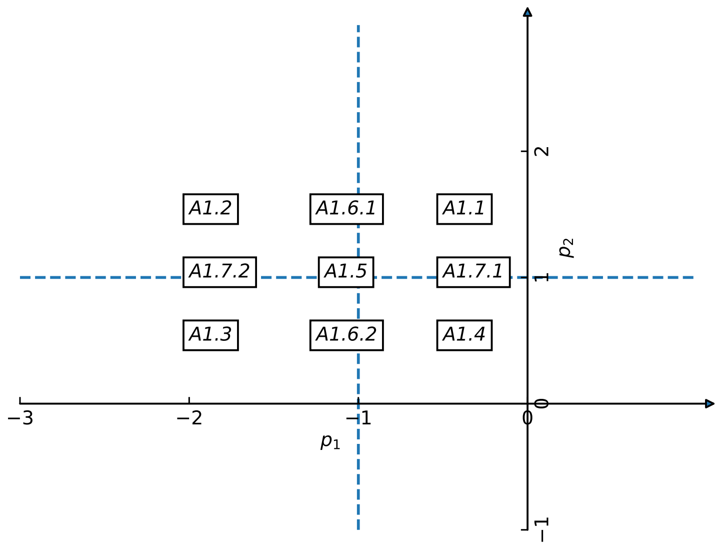


Figure 2.1: A sketch to illustrate the different cases parametrized by the values of p_1, p_2.

For example, A1.1 means $p_1 > -1, p_2 > 1$, A1.5 means $p_1 = -1, p_2 = 1$ and A1.7.1 means $p_1 > -1, p_2 = 1$. The main results are summarized in Table 2.2.

Table 2.2: Interpretation of error estimate (2.29).

1	**If** $k = 1$ and $p_2 < 1$, **then**
2	$P_h^\star$ is suboptimal and is $(k + p_2)$-th order accurate,
3	**else**
4	**if** $\lim_{h\to 0} \lvert\lambda_1, \lambda_2\rvert = 1$ with $\lvert\lambda_1, \lambda_2\rvert = 1 + O(h^{\delta/2})$, **then**
5	$P_h^\star$ is suboptimal and is $(k + 1 - \delta)$-th order accurate,
6	**else**
7	$P_h^\star$ has optimal $(k + 1)$-th order error estimates.
8	**end**
9	**end**

The main reason of order reduction for $k = 1, p_2 < 1$ in line 2 of Table 2.2 is that the term such as $\frac{1}{|\lambda_1|-1}\|Q_1\|_\infty$ is of $O(h^{p_2-1})$ instead of $O(1)$, and this will cause $(1 - p_2)$-th order reduction. The situation happens for A1.3, A1.4 and A1.6.2 when $k = 1$.

The main reason of order reduction in line 5 is because of the terms such as $\frac{1}{1-|\lambda_1|}, \frac{|1}{1-|\lambda_2|}$ in (2.29). The fractions $\frac{1}{1-|\lambda_1|}, \frac{1}{1-|\lambda_2|}$ cannot be bounded by a constant if $\lim_{h\to 0}|\lambda_2| = 1$. By definition of λ_1, λ_2 in (49), we know that $\left|\frac{\Gamma}{\Lambda}\right| \to 1 \Leftrightarrow |\lambda_1, \lambda_2| \to 1$. More precisely, if $\left|\frac{\Gamma}{\Lambda}\right| = 1 + O(h^\delta), \delta > 0$, then $|\lambda_1, \lambda_2| = 1 + O(h^{\delta/2})$, then $\frac{1}{1-|\lambda_1|}, \frac{1}{1-|\lambda_2|} = O(h^{-\delta/2})$. The relation $\Gamma^2 - \Lambda^2 = (b_1 - b_2)(b_1 + b_2) + c_2^2$ also indicates that there is some cancellation of leading terms in $b_1 - b_2$ or $b_1 + b_2$, making $\|Q_1\|_\infty \sim O(h^{-\delta/2})$, multiplying these factors together will result in δ-th order reduction in the error estimation of $P_h^\star$. Note that b_1, b_2, c_2 and Q_1 are defined in (47), (48), (45) and (56).

Then we look at what parameter choices make $\left|\frac{\Gamma}{\Lambda}\right| \to 1$. Since

$$\frac{\Gamma}{\Lambda} = \begin{cases} k + \frac{\beta_1 + \frac{k^2(k^2-1)}{h^2}\beta_2 - \frac{k^2}{h}}{\Lambda} & k > 1, \\ 1 + \frac{\beta_1 - \frac{1}{h}}{\Lambda} & k = 1, \end{cases}$$

we have

1. <u>A1.1</u> ($p_1 > -1, p_2 > 1$) with $k = 1, \alpha_1 = 0, \left|\frac{\Gamma}{\Lambda}\right| \to \left|\frac{\frac{1}{2}+2\alpha_1^2}{\frac{1}{2}-2\alpha_1^2}\right| = 1$.

2. <u>A1.6.1</u> ($p_1 = -1, p_2 > 1$) $\tilde{\beta}_1 = \frac{k(k\pm1)}{2} + 2\alpha_1^2 k(k \mp 1), \left|\frac{\Gamma}{\Lambda}\right| \to \left|k + \frac{\beta_1 - \frac{k^2}{h}}{\Lambda}\right| \to 1$.

3. <u>A1.6.2</u> ($p_1 = -1, p_2 < 1$) with $k > 1$, $\tilde{\beta}_1 = \frac{k(k\pm1)}{2}, \left|\frac{\Gamma}{\Lambda}\right| \to \left|k + \frac{\frac{k^2(k^2-1)\beta_2}{h^2}}{\Lambda}\right| \to 1$.

4. <u>A1.7.1</u> ($p_1 > -1, p_2 = 1$) $\tilde{\beta}_2 = \frac{1}{2k(k\mp1)} + \frac{2\alpha_1^2}{k(k\pm1)}, \left|\frac{\Gamma}{\Lambda}\right| \to \left|k + \frac{\frac{k^2(k^2-1)\beta_2}{h^2} - \frac{k^2}{h}}{\Lambda}\right| \to 1$.

5. <u>A1.7.2</u> ($p_1 < -1, p_2 = 1$) $\tilde{\beta}_2 = \frac{1}{2k(k\pm1)}, \left|\frac{\Gamma}{\Lambda}\right| \to \left|k + \frac{\beta_1}{\Lambda}\right| \to 1$.

Remark 2.2.1. *We only considered* T *given by* (51) *(when* Q_1 *is given by* (56)*) in the discussion above. By Appendix, we can conclude that under the parameter conditions in assumption A1,* $(b_1+b_2)(b_1-b_2)=0$ *only can happen if* $p_1=-1, p_2=1$ *with* (29) *or* (30)*. This is A1.5, for which we always have optimal error estimate.*

Remark 2.2.2. *Through numerical tests, we found that* (2.29) *is mostly sharp with two exceptions. When* $\lim_{h\to 0}|\lambda_1,\lambda_2|=1$*, the estimates show that there will be order reduction for error of* $P_h^\star$*, while in numerical experiments (see e.g. Tables 2.8, 2.9), such order reduction is observed only when* $\lim_{h\to 0}\lambda_1,\lambda_2=1$ *but not* -1*. We believe when* $\lim_{h\to 0}\lambda_1,\lambda_2=-1$*, a refined estimate can be obtained similar to* (2.30) *under assumption A1. We have not carried out this estimate in this thesis.*

Another example we find for which (2.29) *is not sharp is* $k=2$*,* $p_1=-2$ *or* -3*,* $p_2=1$*,* $(\alpha_1,\tilde{\beta_1},\tilde{\beta_2})=(0.25,-1,\frac{1}{12})$*, where parameters belong to A1.7.2,* $\tilde{\beta}_2=\frac{1}{2k(k+1)}$ *and* $\lambda_1,\lambda_2\to 1+O(h^{-(1+p_1)/2})$*. The theoretical results predict accuracy order of* $(k+2+p_1)$ *but numerical experiments in Table 2.10 show the order to be* $(k+3+p_1)$*. Our estimations can't resolve this one order difference. This special parameter may trigger a cancellation we didn't capture in analysis. We will improve this estimate in our future work.*

Remark 2.2.3. *In most cases,* (2.30) *yields optimal accuracy order, except when* $k=1,\alpha_1=0,\beta_1=0,\beta_2=O(h^{p_2}),p_2<1$*, where the* $P_h^\star$ *is only* $(k+p_2)$*-th order accurate because* $\frac{\|Q_2\|_\infty}{|\Lambda|}=\frac{|b_1+b_2|}{|\Lambda|}=\frac{|-\frac{-4}{h^2}\beta_2+\frac{1}{2h}|}{\frac{1}{2h}}\sim O(h^{p_2-1})$ *in* (2.30)*. This is verified numerically in Table 2.12.*

Remark 2.2.4. *If* $\delta/2>1$*, we can show* $\delta/2=\delta'+1$*. This is because* $|1-\lambda_1|=|1-e^{i\theta}|=2|\sin(\theta/2)|$*, and* $|1-\lambda_1^N|=|1-e^{iN\theta}|=2|\sin(N\theta/2)|$*. When* $\delta/2>1$*, one can assert that* $|1-\lambda_1|\sim\theta, |1-\lambda_1^N|\sim N\theta$*, i.e.* $\delta/2=\delta'+1$*. With this condition, we notice that* (2.31) *yields an reduction of δ-th order in convergence rate by checking the order of each term. This order reduction is consistent with numerical experiments in Example 2.4.4.*

Now we can summarize the estimation of $P_h^\star$ for some frequently used flux parameters. For IPDG scheme with $\alpha_1 = \beta_2 = 0, \beta_1 = c/h$, and DDG scheme discussed in [52] with $\alpha_1 = constant, \beta_1 = c/h, \beta_2 = 0$, and the more general scale invariant parameter choice $\alpha_1 = constant, \beta_1 = c/h, \beta_2 = ch$, $P_h^\star$ always have optimal error estimates. For those parameters, we can show that the eigenvalues λ_1, λ_2 are always constants independent of h, therefore, by Lemma 2.2.3, we will have optimal convergence rate. Corresponding numerical results are shown in Tables 2.4 and 2.7.

For a natural parameter choice where $\alpha_1, \beta_1, \beta_2$ are all real constants, if $\beta_2 \neq 0$, then $P_h^\star$ has first order convergence rate when $k = 1$ and optimal convergence rate when $k > 1$ by Lemma 2.2.3. Corresponding numerical results are shown in Tables 2.3 and 2.12. Lastly, for central flux $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0$, this parameter choice satisfies assumption A3 when $k = 1$ and assumption A2 when $k > 1$, thus we can verify that $P_h^\star$ has optimal convergence rate. Corresponding numerical results are shown in Table 2.11.

2.3 Error estimates

In this section, we will derive error estimates of the DG scheme (2.3) for the model NLS equation (1.1). We will focus on the impact of the choice of the parameters $\alpha_1, \beta_1, \beta_2$ on the accuracy of the scheme. The error estimates rely on the projection error estimates to obtain convergence result.

Theorem 2.3.1. *Assume that the exact solution u and the nonlinear term $f(|u|^2)$ of (1.1) are sufficiently smooth with bounded derivatives for any time $t \in (0, T_e]$ and that the nu-*

merical flux parameters in (2.13) *satisfy the existence conditions of* $P_h^\star$ *in Lemma 2.2.2. Furthermore, assume* $\epsilon_h = u - P_h^\star u$ *has at least first order convergence rate in* L^2 *and* L^∞ *norm from the results in Lemma 2.2.3. With periodic boundary conditions solution space* V_h^k $(k \geq 1)$*, the following error estimation holds for* u_h*, which is the numerical solution of* (2.3) *with flux* (2.13)*:*

$$\|u - u_h\|_{L^2(I)} \leq C_\star \left(\|(u - u_h)|_{t=0}\| + \|(\epsilon_h)_t\| + \|\epsilon_h\| \right), \tag{2.32}$$

where $C_\star$ *depends on* $k, \|f\|_{W^{2,\infty}}$*,* u *as well as final time* T_e*, but not on* h*.*

Proof. When $P_h^\star$ exists, we can decompose the error into two parts.

$$e = u - u_h = u - P_h^\star u + P_h^\star u - u_h := \epsilon_h + \zeta_h.$$

By Galerkin orthogonality, $\forall v_h \in V_h^k$,

$$\begin{aligned} 0 &= a(e, v_h) - i \int_I f(|u|^2) u v_h dx + i \int_I f(|u_h|^2) u_h v_h dx \\ &= a(\epsilon_h, v_h) + a(\zeta_h, v_h) - i \int_I f(|u|^2) u v_h dx + i \int_I f(|u_h|^2) u_h v_h dx. \end{aligned}$$

By letting $v_h = \overline{\zeta_h}$ and taking conjugate of above equation, we have

$$\begin{aligned} & a(\zeta_h, \overline{\zeta_h}) + \overline{a(\zeta_h, \overline{\zeta_h})} \\ = & - a(\epsilon_h, \overline{\zeta_h}) - \overline{a(\epsilon_h, \overline{\zeta_h})} - 2 \int_I f(|u|^2) \mathrm{Im}(u \overline{\zeta_h}) dx + 2 \int_I f(|u_h|^2) \mathrm{Im}(u_h \overline{\zeta_h}) dx. \end{aligned} \tag{2.33}$$

By Taylor expansion

$$f(|u_h|^2) = f(|u|^2) + f'(|u|^2)E + \frac{1}{2}\hat{f}''E^2,$$

where $\hat{f}'' = f''(c), c$ is a value between $|u_h|^2$ and $|u|^2$. $E = |u_h|^2 - |u|^2 = -2\mathrm{Re}(e\overline{u}) + |e|^2$. Therefore, the nonlinear part becomes

$$\begin{aligned}
&\int_I f(|u|^2)\mathrm{Im}(u\overline{\zeta_h})dx - \int_I f(|u_h|^2)\mathrm{Im}(u_h\overline{\zeta_h})dx \\
=&\int_I f(|u_h|^2)\mathrm{Im}\big(e\overline{\zeta_h}\big) + \big(f(|u|^2) - f(|u_h|^2)\big)\mathrm{Im}(u\overline{\zeta_h})dx \\
=&\mathcal{N}_1 + \mathcal{N}_2 + \mathcal{N}_3,
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{N}_1 &= \int_I f(|u|^2)\mathrm{Im}\big(e\overline{\zeta_h}\big) - f'(|u|^2)E\mathrm{Im}(u\overline{\zeta_h})dx, \\
\mathcal{N}_2 &= \int_I f'(|u|^2)E\mathrm{Im}\big(e\overline{\zeta_h}\big) - \frac{1}{2}\hat{f}''E^2\mathrm{Im}(u\overline{\zeta_h})dx, \\
\mathcal{N}_3 &= \int_I \frac{1}{2}\hat{f}''E^2\mathrm{Im}\big(e\overline{\zeta_h}\big)dx,
\end{aligned}$$

will be estimated separately as follows.

- $\mathcal{N}_1$ and $\mathcal{N}_2$ terms.

  Since $e\overline{\zeta_h} = \epsilon_h\overline{\zeta_h} + |\zeta_h|^2$, $\big|E\mathrm{Im}(u\overline{\zeta_h})\big| = \Big|(-2\mathrm{Re}(e\overline{u}) + |e|^2)\mathrm{Im}(u\overline{\zeta_h})\Big| \leq C(\|u\|^2_{L^\infty(I)} +$

$\|u\|_{L^\infty(I)}\|e\|_{L^\infty(I)})(|\epsilon_h|^2 + |\zeta_h|^2)$, we have

$$|\mathcal{N}_1| \le C\|f\|_{W^{1,\infty}} \left(1 + \|u\|^2_{L^\infty(I)} + \|u\|_{L^\infty(I)}\|e\|_{L^\infty(I)}\right)(\|\epsilon_h\|^2 + \|\zeta_h\|^2),$$
$$|\mathcal{N}_2| \le C\|f\|_{W^{2,\infty}}\|E\|_{L^\infty(I)} \left(1 + \|u\|^2_{L^\infty(I)} + \|u\|_{L^\infty(I)}\|e\|_{L^\infty(I)}\right)(\|\epsilon_h\|^2 + \|\zeta_h\|^2).$$

- $\mathcal{N}_3$ term.

$$|\mathcal{N}_3| \le C\|f''\|_{L^\infty}\|E\|^2_{L^\infty(I)}(\|\epsilon_h\|^2 + \|\zeta_h\|^2).$$

To conduct a proper estimate for the nonlinear part, we would like to make an *a priori* assumption that, for h small enough,

$$\|e\| = \|u - u_h\| \le h^{0.5}. \tag{2.34}$$

By our assumption on $P_h^\star$, $\|\epsilon_h\|_{L^p(I)} \le C_1 h, p = 2, \infty$, thus $\|\zeta_h\| \le C_1 h^{0.5}$ and $\|\zeta_h\|_{L^\infty(I)} \le C_1$ by inverse inequality, then $\|e\|_{L^\infty(I)} \le C_1$, $\|E\|_{L^\infty(I)} \le C_1$. Here and below, C_1 is a generic constant that has no dependence on h, but may depend on u according to the lemma used to estimate ϵ_h.

Therefore, we get the estimate:

$$|\mathcal{N}_1| + |\mathcal{N}_2| + |\mathcal{N}_3| \le C_1(\|\epsilon_h\|^2 + \|\zeta_h\|^2), \tag{2.35}$$

where C_1 depends on $\|f\|_{W^{2,\infty}}$ and u.

For linear part of the right hand side in (2.33), we have

$$\begin{aligned}
a(\epsilon_h, \overline{\zeta_h}) + \overline{a(\epsilon_h, \overline{\zeta_h})} &= \int_I (\epsilon_h)_t \overline{\zeta_h} + \overline{(\epsilon_h)_t} \zeta_h dx - i \int_I (\epsilon_h)(\overline{\zeta_h})_{xx} dx \\
&+ i \int_I \overline{(\epsilon_h)} (\zeta_h)_{xx} dx - i \sum_{j=1}^{N} (\widehat{\epsilon_h}[(\overline{\zeta_h})_x] - \widetilde{(\epsilon_h)_x}[\overline{\zeta_h}])|_{j+\frac{1}{2}} \\
&+ i \sum_{j=1}^{N} \overline{(\widehat{\epsilon_h}[(\overline{\zeta_h})_x] - \widetilde{(\epsilon_h)_x}[\overline{\zeta_h}])|_{j+\frac{1}{2}}}, \\
&= 2 \int_I \mathrm{Re}\big((\epsilon_h)_t \overline{\zeta_h}\big) dx.
\end{aligned}$$

The last equality holds because of the definition of $P_h^\star u$. For the left hand side of (2.33), by similar computation in stability analysis we have

$$a(\zeta_h, \overline{\zeta_h}) + \overline{a(\zeta_h, \overline{\zeta_h})} = \frac{d}{dt} \int_I |\zeta_h|^2 dx. \tag{2.36}$$

Combine these two equations with (2.35):

$$\frac{d}{dt} \|\zeta_h\|^2 \leq \|(\epsilon_h)_t\|^2 + \|\zeta_h\|^2 + C_1(\|\epsilon_h\|^2 + \|\zeta_h\|^2).$$

Assuming u_t, u have sufficient smoothness, then by Gronwall's inequality, we can get:

$$\|\zeta_h\|^2 \leq C_1 \left(\|\zeta_h|_{t=0}\|_{L^2(I)}^2 + \|(\epsilon_h)_t\|^2 + \|(\epsilon_h)\|^2 \right),$$

and we obtain (2.32).

To complete the proof, we shall justify the *a priori* assumption. To be more precise, we consider h_0, s.t., $\forall h < h_0, C_\star h \leq \frac{1}{2} h^{0.5}$, where $C_\star$ is defined in (2.32), dependent on T_e, but not on h. Suppose $\exists\, t^* = \sup\{t : \|u(t^*) - u_h(t^*)\| \leq h^{0.5}\}$, we would have $\|u(t^*) - u_h(t^*)\| =$

$h^{0.5}$ by continuity if t^* is finite. By (2.32), we obtain $\|e\| \leq C_\star h \leq \frac{1}{2}h^{0.5}$ if $t^* \leq T_e$, which contradicts the definition of t^*. Therefore, $t^* > T_e$ and the *a priori* assumption is justified.

$\square$

Remark 2.3.1. *If f is a constant function, we can prove the same error estimates without using the* a prixori *assumption. Therefore, the assumption that $\epsilon_h = u - P_h^\star u$ has at least first order convergence rate in L^2 and L^∞ norm is no longer needed.*

Moreover, the estimates for $\|\epsilon_h\|$ has been established in Lemma 2.2.3. In other words, the error of the DG scheme (2.3) *has the same accuracy as $P_h^\star u$, as long as $P_h^\star u$ is well-defined and the numerical initial condition is chosen sufficiently accurate.*

2.4 Numerical results

In this subsection, we present numerical experiments to validate our theoretical results. Particularly, in Section 2.4.1, we provide numerical validations of convergence rate for the projection $P_h^\star$ as discussed in Lemma 2.2.3 with focus on the dependence of the errors on parameters $\alpha_1, \beta_1, \beta_2$. Section 2.4.2 illustrates the energy conservation property and validates theoretical convergence rate of DG scheme for NLS equation (1.1).

2.4.1 Numerical results of the projection operator $P_h^\star$

Example 2.4.1. *In this example, we focus on local projection where $\alpha_1^2 + \beta_1\beta_2 = \frac{1}{4}$, and verify the conclusions in Lemma 2.2.1 by considering a smooth test function $u = \cos(x), x \in [0, 2\pi]$ on a nonuniform mesh and $k = 1, 2, 3$ for various sets of parameters $(\alpha_1, \beta_1, \beta_2)$. The nonuniform mesh is generated by perturbing the nodes of a uniform mesh of N cells by at most* 10%.

We first consider two sets of parameters $(\alpha_1, \beta_1, \beta_2) = (0.3, 0.4, 0.4)$ and $(\alpha_1, \beta_1, \beta_2) = (0.3, 0.4/h, 0.4h)$. The results with $(\alpha_1, \beta_1, \beta_2) = (0.3, 0.4, 0.4)$ are listed in Table 2.3. By plugging in the parameters into (2.28), we have that when $k = 1$, the projection has sub-optimal first order convergence rate, while for $k > 1$, optimal $(k+1)$-th order convergence rate should be achieved. For $k = 1$, $\Gamma_j = \beta_1 - \frac{1}{h}$, which does not depend on β_2 any more. This technical difference cause the discrepancy of the convergence order between $k = 1$ and $k > 1$ in Table 2.3. Results in Table 2.3 agree well with the theoretical prediction. On the other hand, when we choose parameters $(\alpha_1, \beta_1, \beta_2) = (0.3, 0.4/h, 0.4h)$, by Lemma 2.2.3, we should observe optimal convergence rate for all $k \geq 1$, and this is verified by the numerical results in Table 2.4.

In [16], we also proved that $P_h^\star$ is superclose to P_h^1 when $\beta_2 = 0, \alpha_1 = \pm\frac{1}{2}$ and $\beta_2/\Gamma_j \sim o(1)$. We choose the parameters as $(\alpha_1, \beta_1, \beta_2) = (0.5, 1, 0)$ to verify this claim, i.e., the difference between $P_h^\star$ and P_h^1 can have convergence rates higher than $k+1$. The results are listed in Table 2.5. The difference of the two projections is indeed of $(k+2)$-th order for any $k \geq 1$ in all norms.

Table 2.3: Example 2.4.1. Error of local projection $P_h^\star u - u$ on a nonuniform mesh. Flux parameters: $\alpha_1 = 0.3, \beta_1 = 0.4, \beta_2 = 0.4$.

	N	L^1 error	order	L^2 error	order	L^∞ error	order
P^1	160	1.98E-02	-	1.56E-02	-	1.81E-02	-
	320	9.98E-03	0.99	7.87E-03	0.99	9.20E-03	0.97
	640	5.01E-03	0.99	3.95E-03	0.99	4.55E-03	1.02
	1280	2.51E-03	1.00	1.98E-03	1.00	2.27E-03	1.00
P^2	160	2.18E-06	-	1.91E-06	-	3.73E-06	-
	320	2.71E-07	3.01	2.39E-07	3.00	5.14E-07	2.86
	640	3.37E-08	3.01	2.97E-08	3.01	6.71E-08	2.94
	1280	4.19E-09	3.01	3.69E-09	3.01	7.99E-09	3.07
P^3	160	2.82E-09	-	2.45E-09	-	5.67E-09	-
	320	1.76E-10	4.00	1.53E-10	4.00	3.76E-10	3.92
	640	1.10E-11	4.00	9.50E-12	4.01	2.25E-11	4.06
	1280	6.86E-13	4.00	5.93E-13	4.00	1.46E-12	3.95

Table 2.4: Example 2.4.1. Error of local projection $P_h^\star u - u$ on a nonuniform mesh. Flux parameters: $\alpha_1 = 0.3, \beta_1 = 0.4/h, \beta_2 = 0.4h$.

	N	L^1 error	order	L^2 error	order	L^∞ error	order
P^1	160	3.42E-04	-	3.50E-04	-	8.62E-04	-
	320	8.55E-05	2.00	8.75E-05	2.00	2.21E-04	1.96
	640	2.14E-05	2.00	2.19E-05	2.00	5.45E-05	2.02
	1280	5.34E-06	2.00	5.47E-06	2.00	1.36E-05	2.00
P^2	160	6.36E-06	-	6.06E-06	-	2.06E-05	-
	320	8.17E-07	2.96	7.99E-07	2.92	3.09E-06	2.73
	640	1.02E-07	3.00	1.00E-07	2.99	4.51E-07	2.78
	1280	1.27E-08	3.01	1.24E-08	3.02	5.12E-08	3.14
P^3	160	3.32E-09	-	2.93E-09	-	7.58E-09	-
	320	2.08E-10	4.00	1.83E-10	4.00	5.08E-10	3.90
	640	1.30E-11	4.00	1.14E-11	4.01	3.04E-11	4.06
	1280	8.09E-13	4.00	7.12E-13	4.00	1.99E-12	3.93

Table 2.5: Example 2.4.1. Difference of local projection $P_h^\star$ with P_h^1: $P_h^\star u - P_h^1 u$ on a nonuniform mesh. Flux parameters: $\alpha_1 = 0.5, \beta_1 = 1, \beta_2 = 0$.

	N	L^1 error	order	L^2 error	order	L^∞ error	order
P^1	160	2.09E-05	-	1.96E-05	-	4.66E-05	-
	320	2.56E-06	3.03	2.40E-06	3.03	5.99E-06	2.96
	640	3.17E-07	3.01	2.96E-07	3.02	7.17E-07	3.06
	1280	3.94E-08	3.01	3.67E-08	3.01	9.11E-08	2.98
P^2	160	5.00E-09	-	5.05E-09	-	1.82E-08	-
	320	3.14E-10	3.99	3.21E-10	3.98	1.28E-09	3.83
	640	1.96E-11	4.00	2.00E-11	4.00	8.56E-11	3.90
	1280	1.22E-12	4.01	1.24E-12	4.01	5.02E-12	4.09
P^3	160	2.91E-12	-	3.38E-12	-	1.40E-11	-
	320	9.11E-14	5.00	1.06E-13	5.00	4.72E-13	4.89
	640	2.84E-15	5.00	3.27E-15	5.01	1.40E-14	5.08
	1280	8.84E-17	5.00	1.02E-16	5.00	4.63E-16	4.92

Example 2.4.2. *In this example, we consider global projection when the parameter choices satisfy assumption A1. We consider a smooth test function $u = e^{\cos(x)}$ on $[0, 2\pi]$ with a uniform mesh of size $h = 2\pi/N$ and $k = 1, 2, 3$ for various sets of parameters $(\alpha_1, \beta_1, \beta_2)$.*

We first test the situation when $\lim_{h\to 0} |\lambda_1, \lambda_2| \neq 1$ by setting the parameters $(\alpha_1, \tilde{\beta_1}, \tilde{\beta_2})$ $= (0.25, 1, 1), p_1 = -0.5, p_2 = 2$. Another example is $(\alpha_1, \beta_1, \beta_2) = (0, \frac{1}{2h}, h)$, for which the eigenvalues λ_1, λ_2 are constant dependent on k but not h. These two parameter choices belong to A1.1 and A1.5, respectively. The numerical results shown in Tables 2.6 and 2.7

verify the optimal $(k+1)$-th order convergence rate predicted by (2.29).

Then we test the situation when $\lim_{h\to 0}|\lambda_1,\lambda_2| = 1$ by using two sets of parameters $(\alpha_1,\tilde{\beta_1},\tilde{\beta_2}) = (0.25, \frac{k(k-1)}{2}+\frac{k(k+1)}{8}, 1)$, $p_1 = -1$, $p_2 = 2,3$, and $(\alpha_1,\tilde{\beta_1},\tilde{\beta_2}) = (0.25, \frac{2}{k(k-1)}, 1), p_1 = -2,-3, p_2 = 1$. The first set of parameters belongs to A1.6.1 and we can verify that $\lim_{h\to 0}\lambda_1,\lambda_2 = (-1)^k$. (2.29) and Algorithm 2.2 imply $(k+2-p_2)$-th convergence order. The numerical results listed in Table 2.8 show that the expected order reduction only happens when $\lim_{h\to 0}\lambda_1,\lambda_2 = 1$, but not for $\lim_{h\to 0}\lambda_1,\lambda_2 = -1$. The second set of parameters belongs to A1.7.2 and we can verify that $\lim_{h\to 0}\lambda_1,\lambda_2 = (-1)^{k+1}$. (2.29) and Algorithm 2.2 imply $(k+2+p_1)$-th convergence order. The numerical results listed in Table 2.9 also show that order reduction is only observed when $\lim_{h\to 0}\lambda_1,\lambda_2 = 1$.

Lastly, we test $(\alpha_1,\tilde{\beta_1},\tilde{\beta_2}) = (0.25,-1,\frac{1}{12})$ with $k=2, p_1=-2,-3, p_2=1$, where our theoretical results predict accuracy order of $(k+2+p_1)$, but numerical experiments show the order to be $(k+3+p_1)$ in Table 2.10. This is one of the exceptions that (2.29) is not sharp and has been commented in Remark 2.2.2.

Table 2.6: Example 2.4.2. Error of global projection $P_h^\star u - u$. Flux parameters: $\alpha_1 = 0.25, \tilde{\beta_1} = 1, \tilde{\beta_2} = 1$, $p_1 = -0.5, p_2 = 2$. (A1.1)

	N	L^1 error	order	L^2 error	order	L^∞ error	order
P^1	160	0.10E-03	-	0.69E-03	-	0.89E-03	-
	320	0.26E-04	1.93	0.18E-03	1.93	0.23E-03	1.94
	640	0.67E-05	1.98	0.46E-04	1.97	0.58E-04	1.98
	1280	0.17E-05	1.99	0.12E-04	1.99	0.15E-04	2.00
P^2	160	0.63E-06	-	0.52E-05	-	0.87E-05	-
	320	0.88E-07	2.85	0.71E-06	2.88	0.11E-05	2.95
	640	0.11E-07	2.95	0.91E-07	2.97	0.14E-06	3.00
	1280	0.14E-08	2.99	0.11E-07	2.99	0.17E-07	3.01
P^3	320	0.64E-10	-	0.49E-09	-	0.72E-09	-
	640	0.45E-11	3.82	0.35E-10	3.80	0.52E-10	3.79
	1280	0.29E-12	3.93	0.23E-11	3.91	0.34E-11	3.92
	2560	0.19E-13	3.97	0.15E-12	3.96	0.22E-12	3.96

Example 2.4.3. *In this example, we consider global projection when the parameter choices*

Table 2.7: Example 2.4.2. Error of global projection $P_h^\star u - u$. Flux parameters: $\alpha_1 = 0, \beta_1 = \frac{1}{2h}$, $\beta_2 = h$. (A1.5)

	N	L^1 error	order	L^2 error	order	L^∞ error	order
P^1	320	0.11E-03	-	0.63E-03	-	0.38E-03	-
	640	0.28E-04	2.00	0.16E-03	2.00	0.95E-04	2.00
	1280	0.70E-05	2.00	0.39E-04	2.00	0.24E-04	2.00
	2560	0.18E-05	2.00	0.98E-05	2.00	0.60E-05	2.00
P^2	320	0.11E-06	-	0.71E-06	-	0.62E-06	-
	640	0.14E-07	3.00	0.89E-07	3.00	0.77E-07	3.00
	1280	0.18E-08	3.00	0.11E-07	3.00	0.96E-08	3.00
	2560	0.22E-09	3.00	0.14E-08	3.00	0.12E-08	3.00
P^3	320	0.38E-10	-	0.25E-09	-	0.22E-09	-
	640	0.24E-11	4.00	0.16E-10	4.00	0.14E-10	4.00
	1280	0.15E-12	4.00	0.99E-12	4.00	0.86E-12	4.00
	2560	0.92E-14	4.00	0.62E-13	4.00	0.54E-13	3.99

Table 2.8: Example 2.4.2. Error of global projection $P_h^\star u - u$. Flux parameters: $\alpha_1 = 0.25, \tilde{\beta}_1 = \frac{k(k-1)}{2} + \frac{k(k+1)}{8}, \tilde{\beta}_2 = 1.0$, $p_1 = -1$, $p_2 = 2, 3$. Note here $\lim_{h\to 0} \lambda_1, \lambda_2 = (-1)^k$. (A1.6.1)

	N	L^1 error	order	L^2 error	order	L^∞ error	order
P^1	640	0.75E-05	-	0.52E-04	-	0.66E-04	-
$p_2 = 2$	1280	0.19E-05	1.97	0.13E-04	1.97	0.17E-04	1.97
$\tilde{\beta}_1 = \frac{1}{4}$	2560	0.48E-06	1.99	0.34E-05	1.98	0.42E-05	1.99
	5120	0.12E-06	1.99	0.84E-06	1.99	0.11E-05	1.99
P^2	640	0.15E-06	-	0.12E-05		0.23E-05	-
$p_2 = 2$	1280	0.39E-07	1.94	0.32E-06	1.93	0.61E-06	1.94
$\tilde{\beta}_1 = \frac{7}{4}$	2560	0.98E-08	1.97	0.82E-07	1.97	0.16E-06	1.97
	5120	0.25E-08	1.98	0.21E-07	1.98	0.39E-07	1.99
P^2	640	0.14E-04	-	0.12E-03	-	0.21E-03	-
$p_2 = 3$	1280	0.71E-05	1.00	0.58E-04	1.00	0.11E-03	1.00
$\tilde{\beta}_1 = \frac{7}{4}$	2560	0.35E-05	1.00	0.29E-04	1.00	0.54E-04	1.00
	5120	0.18E-05	1.00	0.15E-04	1.00	0.27E-04	1.00
P^3	320	0.12E-09	-	0.95E-09	-	0.20E-08	-
$p_2 = 2$	640	0.78E-11	3.99	0.60E-10	3.99	0.13E-09	3.99
$\tilde{\beta}_1 = \frac{9}{2}$	1280	0.49E-12	3.99	0.38E-11	3.99	0.80E-11	3.99
	2560	0.31E-13	4.00	0.24E-12	3.99	0.51E-12	3.97

are similar to central fluxes, and satisfy assumptions A1 and A2, for smooth function $u = e^{\cos(x)}$ *on* $[0, 2\pi]$ *with a uniform mesh of size* $h = 2\pi/N$ *and* $k = 1, 2, 3$.

For central flux $(\alpha_1, \beta_1, \beta_2) = (0, 0, 0)$, $\Gamma = -\frac{k^2}{2h}, \Lambda = \frac{k}{2h}$. If $k > 1$, $\frac{|\Gamma|}{|\Lambda|} = k > 1$, flux parameters satisfy to assumption A1, and if $k = 1$, $\Gamma = -\Lambda$ and flux parameters satisfy

Table 2.9: Example 2.4.2. Error of global projection $P_h^\star u - u$. Flux parameters: $\alpha_1 = 0.25$, $\tilde{\beta}_1 = 1$, $\tilde{\beta}_2 = \frac{1}{2k(k-1)}, p_1 = -2, -3, p_2 = 1$. Note here $\lim_{h\to 0} \lambda_1, \lambda_2 = (-1)^{k+1}$. (A1.7.2)

	N	L^1 error	order	L^2 error	order	L^∞ error	order
P^2	320	0.28E-07	-	0.21E-06	-	0.24E-06	-
$p_1 = -3$	640	0.35E-08	3.00	0.27E-07	3.00	0.31E-07	3.00
$\tilde{\beta}_2 = \frac{1}{4}$	1280	0.44E-09	3.00	0.33E-08	3.00	0.38E-08	3.00
	2560	0.55E-10	3.00	0.41E-09	3.00	0.48E-09	3.00
P^3	320	0.70E-08	-	0.57E-07	-	0.12E-06	-
$p_1 = -2$	640	0.94E-09	2.90	0.77E-08	2.90	0.16E-07	2.91
$\tilde{\beta}_2 = \frac{1}{12}$	1280	0.12E-09	2.95	0.99E-09	2.95	0.20E-08	2.95
	2560	0.15E-10	2.98	0.13E-09	2.98	0.26E-09	2.98
P^3	320	0.16E-06	-	0.13E-05	-	0.24E-05	-
$p_1 = -3$	640	0.40E-07	2.00	0.32E-06	2.00	0.61E-06	2.00
$\tilde{\beta}_2 = \frac{1}{12}$	1280	0.10E-07	2.00	0.79E-07	2.00	0.15E-06	2.00
	2560	0.25E-08	2.00	0.20E-07	2.00	0.38E-07	2.00

Table 2.10: Example 2.4.2. Error of global projection $P_h^\star u - u$. Flux parameters: $\alpha_1 = 0.25, \tilde{\beta}_1 = -1, \tilde{\beta}_2 = \frac{1}{2k(k+1)}, p_1 = -2, -3, p_2 = 1$. Note that $\lim_{h\to 0} \lambda_1, \lambda_2 = 1$. (A1.7.2)

	N	L^1 error	order	L^2 error	order	L^∞ error	order
P^2	320	0.72E-07	2.99	0.56E-06	2.98	0.94E-06	2.97
$p_1 = -2$	640	0.90E-08	2.99	0.71E-07	2.99	0.12E-06	2.99
$\tilde{\beta}_2 = \frac{1}{12}$	1280	0.11E-08	3.00	0.89E-08	3.00	0.15E-07	2.99
	2560	0.14E-09	3.00	0.11E-08	3.00	0.19E-08	3.00
P^2	320	0.80E-06	2.01	0.63E-05	2.01	0.12E-04	2.01
$p_1 = -3$	640	0.20E-06	2.00	0.16E-05	2.00	0.30E-05	2.00
$\tilde{\beta}_2 = \frac{1}{12}$	1280	0.50E-07	2.00	0.39E-06	2.00	0.75E-06	2.00
	2560	0.13E-07	2.00	0.98E-07	2.00	0.19E-06	2.00

to assumption A2. We conclude that $P_h^\star$ exists and is unique for $k = 1$ when N is odd and $k > 1$ for arbitrary N. $P_h^\star$ has optimal error estimates as proved in Lemma 2.2.3. Our numerical test in Table 2.11 demonstrates optimal convergence rate for all k.

A similar flux is $(\alpha_1, \beta_1, \beta_2) = (0, 0, 1)$. When $k = 1$, this flux parameter set satisfies assumption A2 and (2.30) yields first order convergence rate as discussed in Remark 2.2.3. When $k = 2, 3$, similar to central flux, this parameter choice satisfies assumption A1, showing optimal convergence rate. The numerical test in Table 2.12 verifies the theoretical results.

Example 2.4.4. *In this example, we consider global projection when the parameter choices*

Table 2.11: Example 2.4.3. Error of global projection $P_h^\star u - u$. (Central flux) Flux parameters: $\alpha_1 = 0, \beta_1 = 0, \beta_2 = 0$.

	N	L^1 error	order	L^2 error	order	L^∞ error	order
P^1	93	0.12E-03	-	0.74E-03	-	0.55E-03	-
	279	0.13E-04	2.00	0.82E-04	2.00	0.61E-04	2.00
	837	0.15E-05	2.00	0.91E-05	2.00	0.68E-05	2.00
	2511	0.17E-06	2.00	0.10E-05	2.00	0.76E-06	2.00
P^2	160	0.11E-05	-	0.85E-05	-	0.10E-04	-
	320	0.14E-06	3.00	0.11E-05	3.00	0.13E-05	2.99
	640	0.17E-07	3.00	0.13E-06	3.00	0.16E-06	3.00
	1280	0.22E-08	3.00	0.17E-07	3.00	0.20E-07	3.00
P^3	160	0.11E-08	-	0.83E-08	-	0.11E-07	-
	320	0.68E-10	4.00	0.52E-09	4.00	0.68E-09	4.00
	640	0.42E-11	4.00	0.32E-10	4.00	0.42E-10	4.00
	1280	0.27E-12	4.00	0.20E-11	4.00	0.26E-11	4.00

Table 2.12: Example 2.4.3. Error of global projection $P_h^\star u - u$. Flux parameters: $\alpha_1 = 0, \beta_1 = 0, \beta_2 = 1$.

	N	L^1 error	order	L^2 error	order	L^∞ error	order
P^1	93	0.21E-01	-	0.12E+00	-	0.68E-01	-
	279	0.72E-02	1.00	0.40E-01	1.00	0.23E-01	1.00
	837	0.24E-02	1.00	0.13E-01	1.00	0.75E-02	1.00
	2511	0.80E-03	1.00	0.44E-02	1.00	0.25E-02	1.00
P^2	160	0.11E-05	-	0.86E-05	-	0.10E-04	-
	320	0.14E-06	3.00	0.11E-05	3.00	0.13E-05	3.00
	640	0.17E-07	3.00	0.13E-06	3.00	0.16E-06	3.00
	1280	0.22E-08	3.00	0.17E-07	3.00	0.20E-07	3.00
	2560	0.27E-09	3.00	0.21E-08	3.00	0.25E-08	3.00
P^3	160	0.27E-08	-	0.23E-07	-	0.36E-07	-
	320	0.17E-09	4.00	0.14E-08	4.00	0.22E-08	4.00
	640	0.11E-10	4.00	0.89E-10	4.00	0.14E-09	4.00
	1280	0.66E-12	4.00	0.55E-11	4.00	0.87E-11	4.00

satisfy assumption A3 for the smooth function $u = e^{\cos(x)}$ on $[0, 2\pi]$ with uniform mesh size $h = 2\pi/N$ and $k = 1, 2, 3$.

An example of A3 is shown in Table 2.13, where the parameters are $(\alpha_1, \tilde{\beta_1}, \tilde{\beta_2}) = (0.25, -1, \frac{1}{2k(k-1)})$, $p_1 = -2, -3, p_2 = 1$, similar to the parameters in Table 2.9. The asymptotic behavior of λ_1, λ_2 when h approaches 0 is indeed similar to Table 2.9, that is, $|\lambda_1, \lambda_2| = 1 + O(h^{-(p_1+1)/2})$ and $\lim_{h\to 0} \lambda_1, \lambda_2 = (-1)^{k+1}$. Same as previous examples, order reductions are only observed when $\lim_{h\to 0} \lambda_1, \lambda_2 = 1$, that is for $k = 3$.

We performed more numerical results under assumption A3, and all are similar to those of A1 as long as the eigenvalues λ_1, λ_2 are approaching 1 at the same rate. Hence, we will not show more examples under assumption A3.

Table 2.13: Example 2.4.4. Error of global projection $P_h^\star u - u$. Flux parameters (A3, and similar to A1.7.2 in Table 2.9): $\alpha_1 = 0.25, \tilde{\beta}_1 = -1$, $\tilde{\beta}_2 = \frac{1}{2k(k-1)}, p_1 = -2, -3, p_2 = 1$. Note here $\lim_{h\to 0} \lambda_1, \lambda_2 = (-1)^{k+1}$.

	N	L^1 error	order	L^2 error	order	L^∞ error	order
P^2	320	0.28E-07	-	0.21E-06	-	0.24E-06	-
$p_1 = -3$	640	0.35E-08	3.00	0.27E-07	3.00	0.31E-07	3.00
$\tilde{\beta}_2 = \frac{1}{4}$	1280	0.44E-09	3.00	0.33E-08	3.00	0.38E-08	3.00
	2560	0.55E-10	3.00	0.41E-09	3.00	0.48E-09	3.00
P^3	320	0.70E-08	-	0.57E-07	-	0.12E-06	-
$p_1 = -2$	640	0.94E-09	2.90	0.77E-08	2.90	0.16E-07	2.91
$\tilde{\beta}_2 = \frac{1}{12}$	1280	0.12E-09	2.95	0.99E-09	2.95	0.20E-08	2.95
	2560	0.15E-10	2.98	0.13E-09	2.98	0.26E-09	2.98
P^3	320	0.16E-06	-	0.13E-05	-	0.24E-05	-
$p_1 = -3$	640	0.40E-07	2.00	0.32E-06	2.00	0.61E-06	2.00
$\tilde{\beta}_2 = \frac{1}{12}$	1280	0.10E-07	2.00	0.79E-07	2.00	0.15E-06	2.00
	2560	0.25E-08	2.00	0.20E-07	2.00	0.38E-07	2.00

2.4.2 Numerical results of the DG scheme

In this subsection, we show the numerical results of the DG scheme applied to the NLS equation. For the time discretization, we use third order IMEX Runge-Kutta method [5] and fix $\Delta t = 1/10000$, which is small enough to guarantee that the spatial errors dominate. To be more precise, we treat the DG discretization of linear term u_{xx} implicitly and nonlinear term $f(|u|^2)u$ explicitly.

Example 2.4.5. *In this example, we verify the energy conservation property of our scheme* *by considering the following linear equation*

$$iu_t + u_{xx} = 0,$$

with the progressive plane wave solution: $u(x,t) = Aexp(i(x-t))$, *with* $A = 1$.

We use L^2 projection as the numerical initial condition. In the discussion of stability condition, we derive that when $\mathrm{Im}\beta_2 \geq 0, \mathrm{Im}\beta_1 \leq 0, |\alpha_1 + \overline{\alpha_2}|^2 \leq -4\mathrm{Im}\beta_1\mathrm{Im}\beta_2$, our scheme for Schrödinger equation is stable. Furthermore, when $\alpha_1 + \alpha_2 = 0, \beta_1, \beta_2$ are real numbers, the scheme is energy conservative. In this example, we compare two different parameter choices to verify the energy conservation property. The parameter choices are $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0.25, -0.25, 1-i, 1+i)$, and $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0.25, -0.25, 1, 1)$ when $k = 2, N = 40$, ending time $T = 100$. Both are numerically stable flux parameters. For the first set of parameters, we expect energy decay due to the contributions from the imaginary part of β_1, β_2 as in (2.11). For the second set of parameter, energy should be conserved.

In Fig. 2.2, we verify that as t increases from 0 to 100, the flux with only real parameters preserve $\|u_h\|$, while the flux with complex numbers have much larger errors. More precisely, for real parameters, $\|u_h(0,\cdot)\| - \|u_h(100,\cdot)\| = 7.9E\text{-}09$, for complex parameters, $\|u_h(0,\cdot)\| - \|u_h(100,\cdot)\| = 5.7E\text{-}04$.

Example 2.4.6. *Accuracy test for NLS equation*

$$iu_t + u_{xx} + |u|^2 u + |u|^4 u = 0, \tag{2.37}$$

which admits a progressive plane wave solution: $u(x,t) = Aexp(i(cx - \omega t))$, *where* $\omega = c^2 - |A|^2 - |A|^4$ *with* $c = 1, A = 1$.

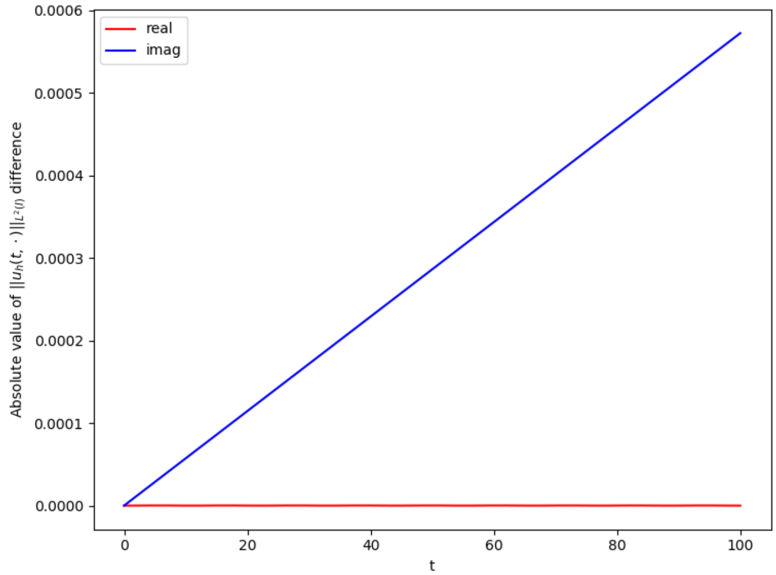


Figure 2.2: Example 2.4.5. Absolute difference of $\|u_h(t,\cdot)\|$ with $\|u_h(0,\cdot)\|$ with two sets of parameters $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0.25, -0.25, 1-i, 1+i)$ (denoted by "imag") and $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0.25, -0.25, 1, 1)$ (denoted by "real") when $k = 2, N = 40$, ending time $T_e = 100$.

For numerical initial condition, $P_h^\star$ is used when applicable, otherwise standard L^2 projection is applied. On uniform mesh, we use four sets of parameters. The numerical errors and orders are shown in Tables 2.14 - 2.19, where corresponding projection results are listed in Tables 2.3, 2.4, 2.11, 2.12, 2.8 and 2.10 respectively. Our numerical experiments show that the order of convergence for the scheme is the same as the order of error estimates for the projection $P_h^\star$.

We would like to make some additional comments on Tables 2.16 and 2.17, whose parameter choices satisfy assumption A2 when $k = 1$. The existence of $P_h^\star$ requires N to be odd for this parameter assumption. However, this assumption is not needed for the optimal convergence rate of the numerical scheme for (2.37) as shown in Tables 2.16 and 2.17. Similar comments have been made in [6].

Table 2.14: Example 2.4.6. Error in L^1, L^2 and L^∞ norm for solving NLS equation (2.37) on a nonuniform mesh using flux parameters (corresponding to Table 2.3) $\alpha_1 = 0.3$, $\beta_1 = \beta_2 = 0.4$, ending time $T_e = 0.3$.

	N	L^1 error	order	L^2 error	order	L^∞ error	order
P^1	40	2.86E-02	-	2.48E-02	-	3.92E-02	-
	80	1.26E-02	1.18	1.02E-02	1.28	1.56E-02	1.33
	160	6.34E-03	1.00	4.99E-03	1.03	6.77E-03	1.20
	320	3.18E-03	1.00	2.56E-03	0.96	3.47E-03	0.96
	640	1.58E-03	1.01	1.27E-03	1.01	1.85E-03	0.91
P^2	40	2.22E-04	-	2.13E-04	-	6.06E-04	-
	80	1.99E-05	3.48	2.13E-05	3.33	7.28E-05	3.06
	160	3.17E-06	2.65	3.03E-06	2.81	9.01E-06	3.02
	320	3.49E-07	3.18	3.34E-07	3.18	1.23E-06	2.87
P^3	40	1.54E-06	-	1.35E-06	-	3.29E-06	-
	80	4.96E-08	4.96	4.36E-08	4.95	1.29E-07	4.67
	160	2.81E-09	4.14	2.60E-09	4.07	8.37E-09	3.95
	320	1.61E-10	4.13	1.57E-10	4.05	7.68E-10	3.45

Table 2.15: Example 2.4.6. Error in L^1, L^2 and L^∞ norm for solving NLS equation (2.37) on a nonuniform mesh using flux parameters (corresponding to Table 2.4) $\alpha_1 = 0.3$, $\beta_1 = 0.4h, \beta_2 = 0.4/h_j$, ending time $T_e = 1$.

	N	L^1 error	order	L^2 error	order	L^∞ error	order
P^1	40	7.47E-03	-	6.50E-03	-	1.29E-02	-
	80	2.10E-03	1.83	1.76E-03	1.89	4.22E-03	1.62
	160	4.82E-04	2.12	4.18E-04	2.07	1.16E-03	1.86
	320	1.21E-04	1.99	1.05E-04	1.99	2.87E-04	2.01
	640	3.12E-05	1.96	2.71E-05	1.95	7.40E-05	1.96
P^2	40	5.14E-04	-	5.37E-04	-	1.74E-03	-
	80	6.81E-05	2.92	7.00E-05	2.94	2.99E-04	2.54
	160	8.04E-06	3.08	8.06E-06	3.12	3.58E-05	3.06
	320	9.53E-07	3.08	9.75E-07	3.05	3.92E-06	3.19
	640	1.68E-07	2.50	1.61E-07	2.60	4.90E-07	3.00
P^3	40	1.30E-06	-	1.25E-06	-	4.09E-06	-
	80	5.74E-08	4.51	6.00E-08	4.38	2.60E-07	3.98
	160	4.44E-09	3.69	4.12E-09	3.86	1.49E-08	4.13
	320	2.25E-10	4.30	2.13E-10	4.28	9.65E-10	3.94

Example 2.4.7. *A simulation for the NLS equation*

$$iu_t + u_{xx} + 2|u|^2u = 0 \tag{2.38}$$

Table 2.16: Example 2.4.6. Error in L^1, L^2 and L^∞ norm for solving NLS equation (2.37) using central flux (corresponding to A2 in Table 2.11) $\alpha_1 = \beta_1 = \beta_2 = 0$, ending time $T_e = 1$.

	N	L^1 error	order	L^2 error	order	L^∞ error	order
	40	0.28E-02	-	0.22E-02	-	0.27E-02	-
	80	0.71E-03	2.00	0.56E-03	2.00	0.67E-03	2.02
P^1	160	0.18E-03	2.00	0.14E-03	2.00	0.17E-03	2.01
	320	0.45E-04	2.00	0.35E-04	2.00	0.41E-04	2.00
	640	0.11E-04	2.00	0.88E-05	2.00	0.10E-04	2.00
	40	0.13E-03	-	0.11E-03	-	0.16E-03	-
	80	0.16E-04	2.99	0.14E-04	2.99	0.20E-04	3.00
P^2	160	0.21E-05	3.00	0.18E-05	3.00	0.25E-05	3.01
	320	0.26E-06	3.00	0.22E-06	3.00	0.31E-06	3.00
	640	0.32E-07	3.00	0.27E-07	3.00	0.39E-07	3.00
	40	0.22E-06	-	0.18E-06	-	0.24E-06	-
	80	0.16E-07	3.76	0.13E-07	3.80	0.13E-07	4.16
P^3	160	0.10E-08	4.00	0.79E-09	4.00	0.84E-09	4.00
	320	0.62E-10	4.00	0.49E-10	4.00	0.52E-10	4.00
	640	0.39E-11	3.99	0.31E-11	3.99	0.33E-11	3.96

Table 2.17: Example 2.4.6. Error in L^1, L^2 and L^∞ norm for solving NLS equation (2.37) using flux parameters (corresponding to A2 in Table 2.12): $\alpha_1 = \beta_1 = 0, \beta_2 = 1$, ending time $T_e = 1$.

	N	L^1 error	order	L^2 error	order	L^∞ error	order
	40	0.17E+00	-	0.13E+00	-	0.14E+00	-
	80	0.92E-01	0.90	0.72E-01	0.89	0.75E-01	0.87
P^1	160	0.48E-01	0.94	0.38E-01	0.94	0.38E-01	0.97
	320	0.24E-01	0.97	0.19E-01	0.97	0.19E-01	0.98
	640	0.12E-01	0.98	0.97E-02	0.98	0.98E-02	0.99
	40	0.13E-03	-	0.11E-03	-	0.17E-03	-
	80	0.16E-04	3.00	0.14E-04	3.00	0.20E-04	3.02
P^2	160	0.21E-05	3.00	0.18E-05	3.00	0.25E-05	3.01
	320	0.26E-06	3.00	0.22E-06	3.00	0.31E-06	3.01
	640	0.32E-07	3.00	0.27E-07	3.00	0.39E-07	3.00
	40	0.68E-06	-	0.56E-06	-	0.83E-06	-
	80	0.42E-07	4.00	0.35E-07	4.01	0.51E-07	4.01
P^3	160	0.26E-08	4.00	0.22E-08	4.00	0.32E-08	4.00
	320	0.16E-09	4.00	0.14E-09	4.00	0.20E-09	4.00
	640	0.10E-10	4.00	0.85E-11	4.00	0.13E-10	4.00

with double-soliton collision

$$u(x,t) = \operatorname{sech}(x+10-4t)\exp(i(2(x+10)-3t)) + \operatorname{sech}(x-10+4t)\exp(i(-2(x-10)-3t)). \tag{2.39}$$

Table 2.18: Example 2.4.6. Error in L^1, L^2 and L^∞ norm for solving NLS equation (2.37) using flux parameters (corresponding to A1.6.1 in Table 2.8): $\alpha_1 = 0.25, \tilde{\beta}_1 = \frac{k(k-1)}{2} + \frac{k(k+1)}{8}, \tilde{\beta}_2 = 1.0, p_1 = -1, p_2 = 2, 3$, ending time $T_e = 1$.

	N	L^1 error	order	L^2 error	order	L^∞ error	order
P^1, $p_2 = 2$, $\tilde{\beta}_1 = \frac{1}{4}$	40	0.41E-02	-	0.37E-02	-	0.72E-02	-
	80	0.12E-02	1.77	0.10E-02	1.82	0.21E-02	1.80
	160	0.31E-03	1.93	0.25E-03	2.05	0.39E-03	2.39
	320	0.87E-04	1.86	0.69E-04	1.87	0.10E-03	1.94
	640	0.23E-04	1.93	0.18E-04	1.94	0.26E-04	1.97
P^2, $p_2 = 2$, $\tilde{\beta}_1 = \frac{7}{4}$	40	0.49E-04	-	0.49E-04	-	0.13E-03	-
	80	0.83E-05	2.55	0.73E-05	2.74	0.14E-04	3.23
	160	0.31E-05	1.44	0.29E-05	1.32	0.65E-05	1.12
	320	0.95E-06	1.69	0.92E-06	1.69	0.20E-05	1.70
	640	0.26E-06	1.85	0.25E-06	1.86	0.55E-06	1.87
P^2, $p_2 = 3$, $\tilde{\beta}_1 = \frac{7}{4}$	40	0.36E-03	-	0.34E-03	-	0.74E-03	-
	80	0.21E-03	0.78	0.20E-03	0.76	0.43E-03	0.77
	160	0.11E-03	0.92	0.11E-03	0.92	0.23E-03	0.92
	320	0.56E-04	1.00	0.53E-04	1.00	0.11E-03	0.99
	640	0.28E-04	1.00	0.27E-04	1.00	0.58E-04	1.00
P^3, $p_2 = 2$, $\tilde{\beta}_1 = \frac{9}{2}$	40	0.19E-05	-	0.19E-05	-	0.43E-05	-
	80	0.43E-07	5.50	0.38E-07	5.65	0.84E-07	5.66
	160	0.15E-08	4.88	0.15E-08	4.68	0.26E-08	5.00
	320	0.91E-10	4.00	0.90E-10	4.02	0.17E-09	3.94
	640	0.58E-11	3.96	0.57E-11	3.99	0.11E-10	3.98

Table 2.19: Example 2.4.6. Error in L^1, L^2 and L^∞ norm for solving NLS equation (2.37) using flux parameters (corresponding to A1.7.2 in Table 2.10): $\alpha_1 = 0.25, \tilde{\beta}_1 = -1, \tilde{\beta}_2 = \frac{1}{2k(k+1)}, p_1 = -2, -3, p_2 = 1$, ending time $T_e = 1$.

	N	L^1 error	order	L^2 error	order	L^∞ error	order
P^2, $p_1 = -2$, $\tilde{\beta}_1 = \frac{1}{12}$	40	0.60E-04	-	0.54E-04	-	0.95E-04	-
	80	0.76E-05	2.99	0.68E-05	2.98	0.12E-04	2.96
	160	0.96E-06	3.00	0.85E-06	3.00	0.15E-05	2.99
	320	0.12E-06	3.00	0.11E-06	3.00	0.19E-06	2.99
	640	0.15E-07	3.00	0.13E-07	3.00	0.24E-07	3.00
P^2, $p_1 = -3$, $\tilde{\beta}_1 = \frac{1}{12}$	40	0.95E-04	-	0.85E-04	-	0.15E-03	-
	80	0.21E-04	2.22	0.18E-04	2.20	0.33E-04	2.18
	160	0.49E-05	2.08	0.44E-05	2.07	0.79E-05	2.06
	320	0.12E-05	2.02	0.11E-05	2.02	0.20E-05	2.02
	640	0.29E-06	2.02	0.27E-06	2.02	0.48E-06	2.02

We use periodic boundary condition and L^2 projection initialization to run the simulation for double-soliton collision solution. The two waves propagate in opposite directions and collide at $t = 2.5$, after that, the two waves separate. Such behaviors are accurately captured by our numerical simulations, see Figure 2.3 for details.

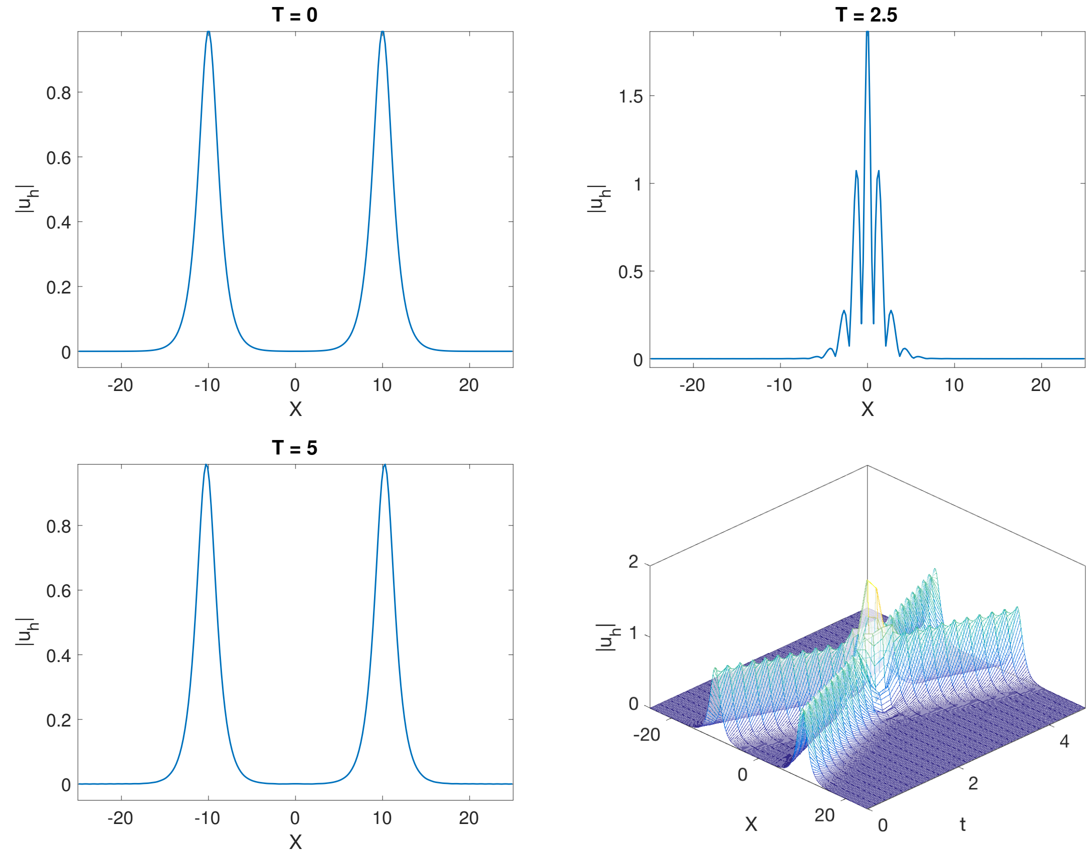


Figure 2.3: Example 2.4.7. Double soliton collision graphs at $t = 0, 2.5, 5$ and a $x - t$ plot of the numerical solution. $N = 250$, P^2 elements with periodic boundary conditions on [-25,25]. Central flux ($\alpha_1 = \beta_1 = \beta_2 = 0$) is used.

Chapter 3

Superconvergence analysis of UWDG method on linear Schrödinger equation

In this chapter, we study the superconvergence properties of the UWDG method on solving the following linear Schrödinger equation:

$$
\begin{aligned}
&iu_t + u_{xx} = 0, \qquad (x,t) \in I \times (0, T_e], \\
&u(x,0) = u_0(x),
\end{aligned}
\tag{3.1}
$$

where $I = [a, b]$ and periodic boundary condition. We consider solving the equation with scale invariant flux parameters. Such choice include all commonly used fluxes, e.g. alternating, central, DDG and interior penalty DG (IPDG) fluxes.

We study the superconvergence property in two types. One type is the superconvergence of cell averages, numerical fluxes, solution at special points and superconvergence towards the projection $P_h^\star$ in Chapter 2. Depending on the flux choices and the evenness of oddness of the polynomial degree k, we obtain $2k$ or $(2k-1)$-th order superconvergence rate for cell averages and numerical flux of the function, as well as $(2k-1)$ or $(2k-2)$-th order for numerical flux of derivative. The proof relies on the correction function techniques for second order derivatives applied to [10] for DDG methods for diffusion problems. We also prove the UWDG solution is superconvergent with a rate of $k+3$ to the projection $P_h^\star$ we introduced in Chapter 2 if $k \geq 3$. At interior points whose locations are determined by roots of certain polynomials associated with the flux parameters, we show that the function values and the first and second order derivatives of the DG solution are superconvergent with order $k+2, k+1, k$, respectively. Compared with [10] for solving diffusion problems, Schrödinger equation poses unique challenges for superconvergence proof because of the lack of the dissipation mechanism from the equation. One major highlight of our proof is that we introduce specially chosen test functions in the error equation and show the superconvergence of the second derivative and jump across the cell interfaces of the difference between numerical solution and projected exact solution. This technique was originally proposed in [19] and is essential to elevate the convergence order for our analysis.

Another type of superconvergence is by postprocessing the UWDG solution such that the postprocessed solution is convergent faster than original solution. We introduce a dual problem and prove $(2k)$-th order negative norm estimate. The order is one order less than that in hyperbolic equations, due to the ultra-weak formulation which has boundary term of the product of derivatives and function values. With the negative norm estimates and divided difference estimates, we prove the $(2k)$-th order superconvergence rate for the postprocessed solution.

The rest of this chapter is organized as follows. In Section 3.1, we recall the UWDG scheme for linear Schrödinger equations and define some new notations. In Section 3.2, we restate the projection results in Section 2.2 under scale invariant parameters and introduce another related projection. Section 3.3 contains the superconvergence results of the UWDG solution in various quantities. In Section 3.5, we provide numerical tests verifying theoretical results. Some technical proof is provided in the Appendix.

The major contents of this chapter has been published in [15].

3.1 Numerical scheme

In this chapter, the semi-discrete UWDG scheme for solving linear Schrödinger equation is defined as follows: solve for the unique function $u_h = u_h(t) \in V_h^k$, $k \geq 1$, $t \in (0, T_e]$, such that

$$a_j(u_h, v_h) = 0, \quad \forall j \in \mathbb{Z}_N \tag{3.2}$$

holds for all $v_h \in V_h^k$, where a_j is defined in (2.6) and the numerical fluxes are defined in (2.14).

Some commonly used fluxes take the following choices of parameters.

- central flux, $\alpha_1 = \beta_1 = \beta_2 = 0$;
- alternating flux, $\alpha_1 = \pm\frac{1}{2}, \beta_1 = \beta_2 = 0$;
- IPDG like flux, $\alpha_1 = \beta_2 = 0, \beta_1 = \tilde{\beta_1} h^{-1}$;
- DDG like flux, $\alpha_1 = \tilde{\alpha_1}, \beta_2 = 0, \beta_1 = \tilde{\beta_1} h^{-1}$;
- more generally, any scale invariant flux, $\alpha_1 = \tilde{\alpha_1}, \beta_1 = \tilde{\beta_1} h^{-1}, \beta_2 = \tilde{\beta_2} h$;

where $\tilde{\alpha_1}, \tilde{\beta_1}, \tilde{\beta_2}$ are prescribed constants independent of mesh size. In this chapter, we will only consider scale invariant flux choices.

Compared with discretization for diffusion equations, we don't have any extra diffusion term in (2.10) to help with the estimates. Therefore, superconvergence error estimates are more challenging compared with [10].

To facilitate the discussion, we introduce notations that will be used in this chapter. Similar to [12], we define operator D^{-1} for any integrable function v on I_j by

$$D^{-1}v(x) = \frac{2}{h_j}\int_{x_{j-\frac{1}{2}}}^{x} v(x)dx = \int_{-1}^{\xi} \hat{v}(\xi)d\xi, \quad x \in I_j. \tag{3.3}$$

Using the property of Legendre polynomials, we have

$$D^{-1}L_{j,k} = \frac{1}{2k+1}\left(L_{j,k+1} - L_{j,k-1}\right), \quad k \geq 1. \tag{3.4a}$$

$$D^{-2}L_{j,k} = \frac{1}{2k+1}\left(\frac{1}{2k+3}(L_{j,k+2} - L_{j,k}) - \frac{1}{2k-1}(L_{j,k} - L_{j,k-2})\right), \quad k \geq 2, \tag{3.4b}$$

where $D^{-2} = D^{-1} \circ D^{-1}$.

3.2 Projections

Under scale invariant flux parameter assumption, we have more concise results for $P_h^\star$. To facilitate the superconvergence proof at special points, we introduce another projection operator $P_h^\dagger$ in this section.

To shorten the notation, from here on we use two notations C_m and $C_{m,n}$ to denote mesh independent constants. C_m may depend on $|u|_{W^{k+1+m,\infty}(I)}$ for assumptions A0/A1, and $\|u\|_{W^{k+3+m,\infty}(I)}$ for assumption A3, $\|u\|_{W^{k+4+m,\infty}(I)}$ for assumption A2. $C_{m,n}$ may depend on $|u|_{W^{k+1+m+2n,\infty}(I)}$ for assumptions A0/A1, on $\|u\|_{W^{k+3+m+3n,\infty}(I)}$ for assumption A3 and on $\|u\|_{W^{k+4+m+4n,\infty}(I)}$ for assumption A2.

The definition of $P_h^\star$ is given in (2.15). When scale-invariant flux parameters are used, we have the following Lemma.

Lemma 3.2.1 ($P_h^\star$ under scale invariant flux parameters)**.** *Suppose any of the assumptions A0/A1/A2/A3 holds, u satisfies the condition in Definition 2.2.1, and scale-invariant flux parameters are used. We have the following estimates*

$$\left|\acute{u}_{j,m} - u_{j,m}\right| \leq C_0 h^{k+1}, m = k-1, k, \quad \|u - P_h^\star u\|_{L^\nu(\mathcal{I}_N)} \leq C_0 h^{k+1}, \quad \nu = 2, \infty. \quad (3.5)$$

In addition, if $h_j = h_{j+1}$,

$$\left|\acute{u}_{j,m} - u_{j,m} - (\acute{u}_{j+1,m} - u_{j+1,m})\right| \leq C_1 h^{k+2}, \quad m = k-1, k. \quad (3.6)$$

Proof. When scale invariant parameters are used, when assumption A0 is satisfied, (3.5) is a direct result of (2.29) in Lemma 2.2.3. On uniform mesh, under assumption A1/A2/A3, the matrices Q_1, Q_2 and eigenvalues λ_1, λ_2 are constants independent of h. Thus, (3.5) is a direct result of (2.30) and (2.31) in Lemma 2.2.3.

To prove the estimates for $\acute{u}_{j,m} - u_{j,m} - (\acute{u}_{j+1,m} - u_{j+1,m}), m = k-1, k$, we denote

$$\mathcal{U}_j = \begin{bmatrix} \acute{u}_{j,k-1} - u_{j,k-1} \\ \acute{u}_{j,k} - u_{j,k} \end{bmatrix} - \begin{bmatrix} \acute{u}_{j+1,k-1} - u_{j+1,k-1} \\ \acute{u}_{j+1,k} - u_{j+1,k} \end{bmatrix}.$$

Plug (2.26), (2.27) in above formula. With the use of (33), $\mathcal{U}_j$ can be estimated in the same way as U_j in the proof of Lemma 2.2.3, and then (3.6) is obtained. We omit the proof for brevity. □

With the optimal estimates of $P_h^\star u$, we proved the optimal L^2 error estimate of the DG scheme in Theorem 2.3.1, which is restated in a more concise version below.

Theorem 3.2.2 (Theorem 2.3.1 under scale-invariant flux parameters)**.** *Suppose any of the* *assumptions A0/A1/A2/A3 holds, let the exact solution u of* (3.1) *be sufficiently smooth, satisfying periodic boundary condition and u_h be the UWDG solution in* (3.2)*, then*

$$\|P_h^\star u - u_h\| \le C_2 h^{k+1}, \quad \|u - u_h\| \le C_2 h^{k+1}. \tag{3.7}$$

Next, we introduce a local projection $P_h^\dagger$ as a variant of $P_h^\star$ and study its approximation properties, especially the superconvergence property at a special set of points. Such superconvergence estimates will help us reveal the superconvergence of UWDG solution at special points. Similar ideas have been employed in [9] for proving the superconvergence at the so-called generalized Radau points when using upwind-biased flux for hyperbolic equations.

Definition 3.2.1. *For DG scheme with flux choice* (2.14)*, we define a local projection operator $P_h^\dagger$ for any periodic function $u \in W^{1,\infty}(I)$ to be the unique polynomial $P_h^\dagger u \in V_h^k$ (when $k \ge 1$) satisfying*

$$\int_{I_j} P_h^\dagger u \, v_h dx = \int_{I_j} u \, v_h dx, \quad \forall v_h \in P_c^{k-2}(I_j), \tag{3.8a}$$

$$G \begin{bmatrix} P_h^\dagger u \\ (P_h^\dagger u)_x \end{bmatrix}\Bigg|_{x_{j+\frac{1}{2}}}^{-} + H \begin{bmatrix} P_h^\dagger u \\ (P_h^\dagger u)_x \end{bmatrix}\Bigg|_{x_{j-\frac{1}{2}}}^{+} = G \begin{bmatrix} u \\ u_x \end{bmatrix}\Bigg|_{x_{j+\frac{1}{2}}} + H \begin{bmatrix} u \\ u_x \end{bmatrix}\Bigg|_{x_{j-\frac{1}{2}}} \tag{3.8b}$$

for all $j \in \mathbb{Z}_N$. When $k = 1$, only condition (3.8b) *is needed.*

Projection $P_h^\dagger$ is always a local projection. Denote $P_h^\dagger u|_{I_j} = \sum_{m=0}^{k} \grave{u}_{j,m} L_{j,m}$, by (3.8a), $\grave{u}_{j,m} = u_{j,m}, m \le k-2$. The similarities in definition imply that $P_h^\star$ and $P_h^\dagger$ are very close to each other, as shown in the following lemma.

Lemma 3.2.3. *For periodic function $u \in W^{1,\infty}(I)$, if assumption A0 is satisfied, $P_h^\star u = P_h^\dagger u$. If any of the assumptions A1/A2/A3 is satisfied, $P_h^\dagger$ exists and is uniquely defined if $(-1)^{k+1}\frac{\Gamma_j}{\Lambda_j} \neq 1$ for all $j \in \mathbb{Z}_N$. Then,*

$$\|u - P_h^\dagger u\|_{L^\nu(I_j)} \leq Ch^{k+1}|u|_{W^{k+1,\nu}(I_j)}, \quad \nu = 2, \infty. \tag{3.9}$$

If any of the assumptions A1/A2/A3 is satisfied, we have

$$\|P_h^\star u - P_h^\dagger u\|_{L^\nu(\mathcal{I}_\mathcal{N})} \leq C_1 h^{k+2}, \quad \nu = 2, \infty. \tag{3.10}$$

Proof. When assumption A0 is satisfied, due to (2.20), $P_h^\star = P_h^\dagger$. The rest of the proof is given in Appendix. □

To analyze the superconvergence property at special points, we need to investigate the expansion of the projection error of $P_h^\dagger$ on every cell I_j, if A0/A1/A2/A3,

$$(u - P_h^\dagger u)|_{I_j} = [L_{j,k-1}, L_{j,k}] \begin{bmatrix} u_{j,k-1} - \grave{u}_{j,k-1} \\ u_{j,k} - \grave{u}_{j,k} \end{bmatrix} + \sum_{m=k+1}^{\infty} u_{j,m} L_{j,m} = \sum_{m=k+1}^{\infty} u_{j,m} R_{j,m}, \tag{3.11}$$

where $u_{j,m}$ is defined in (2.24) and

$$R_{j,m} = L_{j,m} - [L_{j,k-1}, L_{j,k}]\mathcal{M}_m. \tag{3.12}$$

We write out the explicit expression of the leading term in expansions

$$R_{j,k+1} = L_{j,k+1} + bL_{j,k} + cL_{j,k-1}, \tag{3.13}$$

where

$$\begin{aligned} b &= -\frac{2\alpha_1 \frac{2k+1}{h_j}}{\Gamma_j + (-1)^k \Lambda_j}, \\ c &= -\frac{\beta_1 - \frac{2(k+1)^2}{h_j}(\alpha_1^2 + \beta_1\beta_2 + \frac{1}{4}) - (-1)^{k+1}\frac{2(k+1)}{h_j}(\alpha_1^2 + \beta_1\beta_2 - \frac{1}{4})}{\Gamma_j + (-1)^k \Lambda_j} \\ &\quad - \frac{\frac{\beta_2}{h_j{}^2} k(k+2)(k+1)^2}{\Gamma_j + (-1)^k \Lambda_j} \end{aligned}$$

to determine the location of superconvergent points.

For $s = 0, 1, 2$, denote D_j^s as the roots of $\frac{d^s}{dx^s} R_{j,k+1}$, $D^s = \bigcup_{j=1}^N D_j^s$, then it follows from (3.11) and (2.25) that, for $x \in D_j^s$,

$$\partial^s (u - P_h^\dagger u)(x) = \sum_{m=k+2}^{\infty} u_{j,m} \frac{d^s}{dx^s} R_{j,m} \le C h^{k+\frac{3}{2}-s} |u|_{W^{k+2,s}(I_j)}, \tag{3.14}$$

indicating superconvergence at those points. We state such superconvergence results in Theorem 3.3.6.

Since the expression of b, c depends on h_j, on nonuniform mesh, $D_j^s, s = 0, 1, 2$ have nodes with the different relative locations on each cell. For simplicity, below we discuss the locations of D_j^0, D_j^1, D_j^2 for special flux choices on uniform mesh.

- Alternating fluxes: $b = \pm\frac{2k+1}{k}, c = -\frac{(k+1)^2}{k}$.
- Central flux: if k is even, then $b = 0, c = -\frac{(k+1)(k+2)}{k(k-1)}$; if k is odd, then $b = 0, c = -1$.

- IPDG fluxes: if k is even, then $b=0, c=-\frac{(k+1)(k+2)-2\tilde{\beta}_1}{k(k-1)-2\tilde{\beta}_1}$; if k is odd, then $b=0, c=-1$.

For central and IPDG fluxes, if k is odd, $R_{j,k+1}=L_{j,k+1}-L_{j,k-1}$, $\frac{d}{dx}R_{j,k+1}=\frac{4k+2}{h_j}L_{j,k}$, $\frac{d^2}{dx^2}R_{j,k+1}=\frac{8k+4}{h_j^2}L'_{j,k}$, implying that D_j^0, D_j^1, D_j^2 are Lobatto points of order $k+1$, Gauss points of order k and Lobatto points of order $k+1$ excluding end points, respectively, on interval I_j. Therefore, $card(D_j^0)=k+1, card(D_j^1)=k, card(D_j^2)=k-1$.

Cao *et al.* proved there exists $k+1$ superconvergence points (Radau points) when using upwind flux for linear hyperbolic problem in [12], $k+1$ superconvergence points (Lobatto points) using special flux parameter in DDG method in [10] and $k+1$ or k superconvergence points, depending on parameters, for using upwind-biased flux for linear hyperbolic problem in [9]. Analyzing the number and location of superconvergent points for our scheme is more challenging. We shall only provide lower bound estimates for the number of superconvergence points. For general parameters choices, when $k \geq 2$, $R_{j,k+1} \perp P_c^{k-2}(I_j)$, by Theorem 3.3 and Corollary 3.4 in [67], we can easily show $R_{j,k+1}$ has at least $k-1$ simple zeros, i.e., $card(D_j^0) \geq k-1$. By the same approach, we can show when $k \geq 3$, $card(D_j^1) \geq k-2$, and when $k \geq 4$, $card(D_j^2) \geq k-3$. For small k values, D_j^1, D_j^2 can possibly be empty sets.

3.3 Superconvergence properties

In this section, we study superconvergence of the numerical solution. We investigate the superconvergence of UWDG fluxes, cell averages, towards a particular projection and at some special points. This analysis is done by decomposing the error into

$$e=u-u_h=\epsilon_h+\zeta_h,\ \epsilon_h=u-u_I,\ \zeta_h=u_I-u_h \tag{3.15}$$

for some $u_I \in V_h^k$. For error analysis of DG schemes, u_I is usually taken as a projection of u. While for our purpose of superconvergence analysis, u_I needs to be carefully designed as illustrated in Section 3.3.2. Before that, we prove some intermediate superconvergence results in Section 3.3.1 without specifying u_I. Then, the choice of u_I is made in Section 3.3.2 and the main results are obtained.

3.3.1 Some intermediate superconvergence results

This subsection will collect superconvergence results of $\|(\zeta_h)_{xx}\|_{L^2(I)}, (\frac{1}{N}\sum_{j=1}^N |[\zeta_h]|^2_{j+\frac{1}{2}})^{\frac{1}{2}}$, $(\frac{1}{N}\sum_{j=1}^N |[(\zeta_h)_x]|^2_{j+\frac{1}{2}})^{\frac{1}{2}}$ without specifying u_I. The main idea is to choose special test functions in error equation, similar to the techniques used in [19] for hyperbolic problems. This is an essential step to elevate the superconvergence order in Theorem 3.3.4 when k is even.

Lemma 3.3.1. *For $k \geq 2$, let u be the exact solution to* (3.1) *and u_h be the DG solution in* (3.2)*. ϵ_h, ζ_h are defined in* (3.15)*. We choose s_h to be a function in V_h^k, such that $\int_I s_h v_h dx = a(\epsilon_h, v_h)$, $\forall v_h \in V_h^k$. Then, when any of the assumptions A0/A1 is satisfied,*

$$\|(\zeta_h)_{xx}\| \leq C\|s_h + (\zeta_h)_t\|, \tag{3.16}$$

$$(\frac{1}{N}\sum_{j=1}^N |[\zeta_h]|^2_{j+\frac{1}{2}})^{\frac{1}{2}} \leq Ch^2\|s_h + (\zeta_h)_t\|, \tag{3.17}$$

$$(\frac{1}{N}\sum_{j=1}^N |[(\zeta_h)_x]|^2_{j+\frac{1}{2}})^{\frac{1}{2}} \leq Ch\|s_h + (\zeta_h)_t\|. \tag{3.18}$$

Proof. The proof is given in Appendix. □

3.3.2 Correction functions and the main results

In this section, we shall present the main superconvergence results. The proof depends on Lemma 3.3.1 and the correction function technique introduced by Cao *et al.* in [12, 10], which is essential for superconvergence. We let $u_I = P_h^\star u$ when $k = 2$, and $u_I = P_h^\star u - w$, when $k \geq 3$, where $w \in V_h^k$ is a specially designed correction function defined below.

Similar to [10], we start the construction by defining $w_q, 1 \leq q \leq \lfloor \frac{k-1}{2} \rfloor$. For $k \geq 3$, we denote $w_0 = u - P_h^\star u$ and define a series of functions $w_q \in V_h^k$, as follows

$$\int_{I_j} w_q (v_h)_{xx} dx = -i \int_{I_j} (w_{q-1})_t v_h dx, \qquad \forall v_h \in P_c^k(I_j) \setminus P_c^1(I_j), \tag{3.19a}$$

$$\widehat{w_q} = 0, \qquad \text{at } x_{j+\frac{1}{2}}, \tag{3.19b}$$

$$\widetilde{(w_q)_x} = 0, \qquad \text{at } x_{j+\frac{1}{2}}, \tag{3.19c}$$

for all $j \in \mathbb{Z}_N$. (3.19b) and (3.19c) is equivalent to

$$G \begin{bmatrix} w_q \\ (w_q)_x \end{bmatrix} \Bigg|_{x_{j+\frac{1}{2}}}^{-} + H \begin{bmatrix} w_q \\ (w_q)_x \end{bmatrix} \Bigg|_{x_{j+\frac{1}{2}}}^{+} = 0. \tag{3.20}$$

w_q exists and is unique when any of the assumptions A0/A1/A2/A3 is satisfied for the same reason as the existence and uniqueness of $P_h^\star$.

With the construction of w_q, we define

$$w(x,t) = \sum_{q=1}^{\lfloor \frac{k-1}{2} \rfloor} w_q(x,t), \tag{3.21}$$

then

$$
\begin{aligned}
a_j(\epsilon_h, v_h) &= a_j(u - P_h^\star u, v_h) + \sum_{q=1}^{\lfloor \frac{k-1}{2} \rfloor} a_j(w_q, v_h) \\
&= \int_{I_j} (w_0)_t v_h dx + \sum_{q=1}^{\lfloor \frac{k-1}{2} \rfloor} \left(\int_{I_j} (w_q)_t v_h dx - i \int_{I_j} w_q (v_h)_{xx} dx \right) \\
&= \int_{I_j} (w_0)_t v_h dx + \sum_{q=1}^{\lfloor \frac{k-1}{2} \rfloor} \int_{I_j} (w_q - w_{q-1})_t v_h dx \\
&= \int_{I_j} (w_{\lfloor \frac{k-1}{2} \rfloor})_t v_h dx, \quad \forall v_h \in V_h^k(I).
\end{aligned}
\tag{3.22}
$$

The approximation property of w_q and $a_j(\epsilon_h, v_h)$ are presented in the following Lemma.

Lemma 3.3.2. *For $k \geq 3$, suppose u satisfies the condition in Theorem 3.2.2. For $w_q, 1 \leq q \leq \lfloor \frac{k-1}{2} \rfloor$, $q + r \leq \lfloor \frac{k-1}{2} \rfloor + 1$, we have*

$$
\begin{aligned}
\partial_t^r w_q|_{I_j} = \sum_{m=k-1-2q}^{k} \partial_t^r c_{j,m}^q L_{j,m}, \quad \partial_t^r c_{j,k-1-2q}^q = C h_j^{2q} \partial_t^{q+r} (u_{j,k-1} - \acute{u}_{j,k-1}), \\
\left| \partial_t^r c_{j,m}^q \right| \leq C_{2r,q} h^{k+1+2q},
\end{aligned}
\tag{3.23}
$$

and then

$$
\|\partial_t^r w_q\| \leq C_{2r,q} h^{k+1+2q}. \tag{3.24}
$$

For any $v_h \in V_h^k$,

$$
|a(\epsilon_h, v_h)| \leq C_{2,\lfloor \frac{k-1}{2} \rfloor} h^{k+1+2\lfloor \frac{k-1}{2} \rfloor} \|v_h\|. \tag{3.25}
$$

Proof. The proof is given in Appendix. □

Lemma 3.3.3. *For $k \geq 2$, suppose u satisfies the condition in Theorem 3.2.2. If the* *parameters satisfy any of the assumptions A0/A1 and* $u_h|_{t=0} = u_I|_{t=0}$*, we have*

$$\|(\zeta_h)_{xx}\| \le C_{4+2\lfloor\frac{k-1}{2}\rfloor} h^{k+1+2\lfloor\frac{k-1}{2}\rfloor}. \tag{3.26}$$

$$\Big(\frac{1}{N}\sum_{j=1}^{N} |[(\zeta_h)]|^2_{j+\frac{1}{2}}\Big)^{\frac{1}{2}} \le C_{4+2\lfloor\frac{k-1}{2}\rfloor} h^{k+3+2\lfloor\frac{k-1}{2}\rfloor}. \tag{3.27}$$

$$\Big(\frac{1}{N}\sum_{j=1}^{N} |[(\zeta_h)_x]|^2_{j+\frac{1}{2}}\Big)^{\frac{1}{2}} \le C_{4+2\lfloor\frac{k-1}{2}\rfloor} h^{k+2+2\lfloor\frac{k-1}{2}\rfloor}. \tag{3.28}$$

Proof. This Lemma is a direct result of Lemma 3.3.1, once the estiamtes of $\|s_h + (\zeta_h)_t\|$ is acquired. When $k = 2$, $w = 0$. $a(\epsilon_h, v_h) = \int_I (\epsilon_h)_t v_h dx$ from the definition of $P_h^\star$. That is, $s_h = (\epsilon_h)_t$ in the condition of Lemma 3.3.1. To bound $\|(\zeta_h)_t\|$, we take the time derivative of the error equation and obtain

$$a(e_t, v_h) = a((\epsilon_h)_t, v_h) + a((\zeta_h)_t, v_h) = 0.$$

Let $v_h = \overline{(\zeta_h)_t}$, since $a((\zeta_h)_t, \overline{(\zeta_h)_t}) + \overline{a((\zeta_h)_t, \overline{(\zeta_h)_t})} = \frac{d}{dt}\|(\zeta_h)_t\|^2$, by the property of $P_h^\star u$, we have

$$\frac{d}{dt}\|(\zeta_h)_t\|^2 = -a((\epsilon_h)_t, \overline{(\zeta_h)_t}) - \overline{a((\epsilon_h)_t, \overline{(\zeta_h)_t})} \le 2\|(\epsilon_h)_{tt}\|\|(\zeta_h)_t\|,$$

which implies $\frac{d}{dt}\|(\zeta_h)_t\| \le \|(\epsilon_h)_{tt}\|$. To estimate $\|(\zeta_h)_t|_{t=0}\|$, we let $t = 0$ in the error equation. Since $\zeta_h|_{t=0} = (u_h - u_I)|_{t=0} = 0$, we have

$$a(\epsilon_h, v_h) + \int_I (\zeta_h)_t|_{t=0} v_h dx = 0.$$

Let $v_h = \overline{(\zeta_h)_t}|_{t=0}$, then

$$\|(\zeta_h)_t|_{t=0}\|^2 \le \|(\epsilon_h)_t\|\|(\zeta_h)_t|_{t=0}\|.$$

Therefore,

$$\|(\zeta_h)_t\| \le \|(\epsilon_h)_t\| + t\|(\epsilon_h)_{tt}\|.$$

By Lemma 3.3.1, estimates in (3.5) and the inequality above, we can get (3.26)-(3.28).

For $k \ge 3$, by (3.22), we have $a(\epsilon_h, v_h) = \int_I (w_{\lfloor \frac{k-1}{2} \rfloor})_t v_h dx$, that is, $s_h = (w_{\lfloor \frac{k-1}{2} \rfloor})_t$ in the condition of Lemma 3.3.1. Then, following the same lines of proof as above, by replacing ϵ_h with $w_{\lfloor \frac{k-1}{2} \rfloor}$ and using Lemma 3.3.2, we are done. $\square$

Now we are ready to state the following estimates of $\|\zeta_h\|$.

Theorem 3.3.4. *For $k \ge 2$, suppose u satisfies the condition in Theorem 3.2.2. Assume $u_h|_{t=0} = u_I|_{t=0}$, then $\forall t \in (0, T_e]$,*

$$\|\zeta_h\| \le \begin{cases} C_{2,\frac{k-1}{2}} h^{2k} & \text{if } k \text{ is odd and } A0/A1/A2/A3, \\ (C_{k+2}h^{4k} + \sum_{I_j \subset I^{NU}} C_k h^{4k-1})^{\frac{1}{2}} & \text{if } k \text{ is even and } A0/A1, \\ C_{2,\frac{k-2}{2}} h^{2k-1}, & \text{if } k \text{ is even and } A2/A3, \end{cases} \tag{3.29}$$

where I^{NU} is the collection of cells in which the length of I_j is different with at least one of its neighbors.

Proof. From error equation, $a(e,\overline{\zeta_h}) = a(\epsilon_h,\overline{\zeta_h}) + a(\zeta_h,\overline{\zeta_h}) = 0$, which gives us

$$\frac{d}{dt}\|\zeta_h\|^2 = -a(\epsilon_h,\overline{\zeta_h}) - \overline{a(\epsilon_h,\overline{\zeta_h})} \leq \begin{cases} 2\|(\epsilon_h)_t\|\|\zeta_h\|, & k=2, \\ 2\|(w_{\lfloor\frac{k-1}{2}\rfloor})_t\|\|\zeta_h\|, & k\geq 3. \end{cases} \tag{3.30}$$

By (3.5), (3.24) and Gronwall's inequality, we have

$$\|\zeta_h\| \leq C_{2,\lfloor\frac{k-1}{2}\rfloor} t h^{k+1+2\lfloor\frac{k-1}{2}\rfloor}, \quad \forall t\in(0,T_e].$$

Therefore, when k is odd, or k is even and parameters satisfy any of the assumptions A2/A3, the proof is complete.

When k is even and parameters satisfy any of the assumptions A0/A1, we make use of Lemma 3.3.3 to show the improved estimates. We let $l = \lfloor\frac{k-1}{2}\rfloor = \frac{k-2}{2}$, then

$$\begin{aligned} a(\epsilon_h,\overline{\zeta_h}) &= \int_I (w_l)_t\overline{\zeta_h}dx = \sum_{j=1}^{N}\sum_{m=1}^{k}\partial_t c_{j,m}^l\int_{I_j} L_{j,m}\overline{\zeta_h}dx \\ &= \sum_{j=1}^{N}\partial_t c_{j,1}^l\int_{I_j} L_{j,1}\overline{\zeta_h}dx + \sum_{j=1}^{N}\sum_{m=2}^{k}\partial_t c_{j,m}^l\int_{I_j} L_{j,m}\overline{\zeta_h}dx \doteq \mathcal{A}_1+\mathcal{A}_2, \end{aligned}$$

where we denote the first term in the summation by $\mathcal{A}_1$, and the other term in summation as $\mathcal{A}_2$. Note that $D^{-1}L_{j,m}\perp P^0, m\geq 1$ in the inner product sense, thus $D^{-2}L_{j,m}(\pm 1) =$

$0, m \geq 2$. By integration by parts, we get

$$
\begin{aligned}
\mathcal{A}_2 &= \sum_{j=1}^{N} \sum_{m=2}^{k} \frac{h_j^2}{4} \partial_t c_{j,m}^l \int_{I_j} D^{-2} L_{j,m} \overline{(\zeta_h)}_{xx} dx \\
&\leq Ch^{-1} \sum_{j=1}^{N} |\frac{h_j^2}{4} \partial_t c_{j,m}^l|^2 + h \sum_{j=1}^{N} \sum_{m=2}^{k} (\int_{I_j} D^{-2} L_{j,m} \overline{(\zeta_h)}_{xx} dx)^2 \leq C_k h^{4k} + Ch^2 \|(\zeta_h)_{xx}\|^2 \\
&\leq C_{k+2} h^{4k}
\end{aligned}
$$

where we have used (3.23) in the first inequality, and (3.26) in the third inequality.

To estimate $\mathcal{A}_1$, we take the first and second antiderivative of $L_{j,1} = \xi$ as $\frac{h_j}{2}(\frac{\xi^2}{2} - \frac{1}{6})$, $(\frac{h_j}{2})^2 \frac{\xi^3 - \xi}{6}$ and apply integration by parts twice,

$$
\begin{aligned}
\mathcal{A}_1 &= \sum_{j=1}^{N} \frac{h_j}{2} \partial_t c_{j,1}^l \left((\frac{\xi^2}{2} - \frac{1}{6}) \bar{\zeta}_h |_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} - \int_{I_j} (\frac{\xi^2}{2} - \frac{1}{6}) (\bar{\zeta}_h)_x dx \right) \\
&= \sum_{j=1}^{N} \frac{h_j}{2} \partial_t c_{j,1}^l \left(\frac{1}{3} (\bar{\zeta}_h|_{j+\frac{1}{2}}^{-} - \bar{\zeta}_h|_{j-\frac{1}{2}}^{+}) + \frac{h_j}{2} \int_{I_j} \frac{\xi^3 - \xi}{6} (\bar{\zeta}_h)_{xx} dx \right) \\
&= \sum_{j=1}^{N} \frac{h_j}{2} \partial_t c_{j,1}^l \left(-\frac{1}{3} [\bar{\zeta}_h]_{j+\frac{1}{2}} + \frac{h_j}{2} \int_{I_j} \frac{\xi^3 - \xi}{6} (\bar{\zeta}_h)_{xx} dx \right) \\
&+ \sum_{j=1}^{N} \left(\left(\frac{h_j}{2} \right) \partial_t c_{j,1}^l - \left(\frac{h_{j+1}}{2} \right) \partial_t c_{j+1,1}^l \right) \frac{1}{3} \bar{\zeta}_h|_{j+\frac{1}{2}}^{+},
\end{aligned}
$$

where we have used the periodicity in the last equality. Therefore,

$$
\begin{aligned}
|\mathcal{A}_1| &\leq \frac{1}{2} h \sum_{j=1}^{N} \left(\left(\frac{h_j}{2} \right) |\partial_t c_{j,1}^l| \right)^2 + h^{-1} \sum_{j=1}^{N} \frac{1}{9} |[\bar{\zeta}_h]|_{j+\frac{1}{2}}^2 + Ch^2 \sum_{j=1}^{N} \|(\zeta_h)_{xx}\|_{L^2(I_j)}^2 \\
&+ \frac{1}{18} h^{-1} \sum_{j=1}^{N} \left| \left(\left(\frac{h_j}{2} \right) \partial_t c_{j,1}^l - \left(\frac{h_{j+1}}{2} \right) \partial_t c_{j+1,1}^l \right) \right|^2 + h \sum_{j=1}^{N} \|\zeta_h\|_{L^2(\partial I_j)}^2 \\
&\leq C_{k+2} h^{4k} + C\|\zeta_h\|^2 + Ch^{-1} \sum_{j=1}^{N} \left| \left(\left(\frac{h_j}{2} \right) \partial_t c_{j,1}^l - \left(\frac{h_{j+1}}{2} \right) \partial_t c_{j+1,1}^l \right) \right|^2
\end{aligned}
$$

where we used (3.23), inverse inequality, (3.26) and (3.27) in the last inequality.

We estimate the last term in $\mathcal{A}_1$ by the estimation of the difference of $u_{j,m}$ in neighboring cells, similar to that in Proposition 3.1 of [6]. If $h_j \neq h_{j+1}$, then

$$\frac{h_j}{2}\partial_t c^l_{j,1} - \frac{h_{j+1}}{2}\partial_t c^l_{j+1,1} \leq C_k h^{2k}.$$

If $h_j = h_{j+1}$, by (3.23) and (3.6),

$$\begin{aligned}\frac{h_j}{2}\partial_t c^l_{j,1} - \frac{h_{j+1}}{2}\partial_t c^l_{j+1,1} &= C\Big(\frac{h_j}{2}\Big)^{2l+1}\partial_t^{l+1}(u_{j,k-1} - \acute{u}_{j,k-1} - (u_{j+1,k-1} - \acute{u}_{j+1,k-1})) \\ &\leq C_{k+1}h^{2k+1}.\end{aligned}$$

Therefore, we have

$$|\mathcal{A}_1| \leq C_{k+2}h^{4k} + C\|\zeta_h\|^2 + \sum_{I_j \subset I^{NU}} C_k h^{4k-1}.$$

Combine with the estimates for $\mathcal{A}_2$, we have

$$\frac{d}{dt}\|\zeta_h\|^2 \leq C_{k+2}h^{4k} + C\|\zeta_h\|^2 + \sum_{I_j \subset I^{NU}} C_k h^{4k-1}.$$

By Gronwall's inequality and the numerical initial condition, we obtain

$$\|\zeta_h\| \leq (C_{k+2}h^{4k} + \sum_{I_j \subset I^{NU}} C_k h^{4k-1})^{\frac{1}{2}}.$$

The proof is now complete. $\square$

Above theorem states that for $k \geq 2$, when k is odd or k is even and any of the assumptions A0/A1 is satisfied, $\|\zeta_h\|$ has the desired $2k$-th order convergence rate. When k is even and any of the assumptions A2/A3 is satisfied, we are unable to improve the order because that a middle step, Lemma 3.3.3, is proved only under assumption A0/A1. However, numerical result shows the superconvergence results of $\|\zeta_h\|$ and the quantities in following two theorems still hold when k is even and any of the assumptions A2/A3 is satisfied. There is room for improving the proof under such assumption.

Our main superconvergence results are listed in the following two theorems.

Theorem 3.3.5 (Superconvergence of numerical fluxes and cell averages). *Let*

$$\begin{aligned} E_f &= \Big(\frac{1}{N}\sum_{j=1}^{N}(u-\widehat{u_h})|^2_{j+\frac{1}{2}}\Big)^{\frac{1}{2}}, \quad E_{fx} = \Big(\frac{1}{N}\sum_{j=1}^{N}(u_x-\widetilde{(u_h)_x})|^2_{j+\frac{1}{2}}\Big)^{\frac{1}{2}}, \\ E_c &= \Big(\frac{1}{N}\sum_{j=1}^{N}\Big|\frac{1}{h_j}\int_{I_j} u-u_h dx\Big|^2\Big)^{\frac{1}{2}} \end{aligned} \tag{3.31}$$

be the errors in the two numerical fluxes and the cell averages, respectively. For $k \geq 2$, suppose u satisfies the condition in Theorem 3.2.2. Assume $u_h|_{t=0} = u_I|_{t=0}$, then $\forall t \in (0, T_e]$

- *if k is odd, parameters satisfy any of the assumptions A0/A1/A2/A3, we have*

$$E_f \leq C_{2,\frac{k-1}{2}} h^{2k}, \quad E_{fx} \leq C_{2,\frac{k-1}{2}} h^{2k-1}, \quad E_c \leq C_{2,\frac{k-1}{2}} h^{2k}, \tag{3.32}$$

- *if k is even, parameters satisfy any of the assumptions A0/A1, we have*

$$E_f \le (C_{k+2}h^{4k} + \sum_{I_j \subset I^{NU}} C_k h^{4k-1})^{\frac{1}{2}},\ E_{f_x} \le (C_{k+2}h^{4k} + \sum_{I_j \subset I^{NU}} C_k h^{4k-1})^{\frac{1}{2}} h^{-1}, \tag{3.33}$$

$$E_c \le (C_{k+2}h^{4k} + \sum_{I_j \subset I^{NU}} C_k h^{4k-1})^{\frac{1}{2}}, \tag{3.34}$$

where I^{NU} is the collection of cells in which the length of I_j is different with at least one of its neighbors.

- *if k is even and parameters satisfy assumption A2/A3, we have*

$$E_f \le C_{2,\frac{k-2}{2}} h^{2k-1}, \quad E_{f_x} \le C_{2,\frac{k-2}{2}} h^{2k-2}, \quad E_c \le C_{2,\frac{k-2}{2}} h^{2k-1}. \tag{3.35}$$

Proof. We first prove the estimates for E_f. By (3.19b) and the definition of $P_h^\star$, $\widehat{\epsilon}_h(x_{j+\frac{1}{2}}) = \widehat{u - u_I}(x_{j+\frac{1}{2}}) = 0$, then

$$(u - \widehat{u_h})|_{j+\frac{1}{2}} = (\widehat{\zeta_h})|_{j+\frac{1}{2}} = \big(\{\zeta_h\} - \alpha_1[\zeta_h] + \beta_2[(\zeta_h)_x]\big)|_{j+\frac{1}{2}}.$$

Therefore, by inverse inequality and the fact $\beta_2 = \tilde{\beta}_2 h$,

$$E_f \le C\Big(\frac{1}{N}\|\zeta_h\|^2_{L^2(\partial\mathcal{I}_N)}\Big)^{\frac{1}{2}} \le C\|\zeta_h\|,$$

and the desired estimates for E_f is obtained by (3.29). The estimates for E_{f_x} can be obtained following same lines.

Next, we prove the estimates for E_c. If k is odd, then $\int_{I_j} w_q dx = 0, 1 \leq q \leq \frac{k-3}{2}$, by (3.23) and orthogonality of Legendre polynomials. Thus,

$$\int_{I_j} u - u_h dx = \int_{I_j} u - P_h^\star u + \sum_{q=1}^{\lfloor \frac{k-1}{2} \rfloor} w_q + \zeta_h dx = \int_{I_j} w_{\lfloor \frac{k-1}{2} \rfloor} dx + \int_{I_j} \zeta_h dx.$$

Thus,

$$\Big| \frac{1}{h_j} \int_{I_j} u - u_h dx \Big|^2 \leq \frac{2}{h_j} (\|\zeta_h\|^2_{L^2(I_j)} + \|w_{\frac{k-1}{2}}\|^2_{L^2(I_j)}).$$

If k is even, then $\int_{I_j} w_q dx = 0, 1 \leq q \leq \frac{k-2}{2}$, by (3.23) and orthogonality of Legendre polynomials. Thus, by similar step, we have

$$\int_{I_j} u - u_h dx = \int_{I_j} \zeta_h dx, \quad \Big| \frac{1}{h_j} \int_{I_j} u - u_h dx \Big|^2 \leq \frac{2}{h_j} \|\zeta_h\|^2_{L^2(I_j)}.$$

Therefore,

$$E_c \leq C(\|\zeta_h\|^2 + \|w_{\frac{k-1}{2}}\|^2)^{1/2} \text{ if } k \text{ is odd, } E_c \leq C\|\zeta_h\| \text{ if } k \text{ is even,}$$

and the desired estimate for E_c is obtained by (3.29) and (3.24). □

Theorem 3.3.6 (Superconvergence towards projections and at special points). *Suppose u satisfies the condition in Theorem 3.2.2. Assume $u_h|_{t=0} = P_h^\star u_0$, then $\forall t \in (0, T_e]$,*

$$\|u_h - P_h^\star u\| \leq \begin{cases} (C_4 h^{4k} + \sum_{I_j \subset I^{NU}} C_2 h^{4k-1})^{\frac{1}{2}} & k = 2, \text{ if } A0 \text{ or } A1 \\ C_{2,1}(1+t)h^{k+3} & k \geq 3, \end{cases} \tag{3.36}$$

where I^{NU} is the collection of cells in which the length of I_j is different with at least one of *its neighbors.*

Assume $D^s, s = 0, 1, 2$ defined in (3.14) are not empty sets. Let

$$E_u = \Big(\frac{1}{|D^0|}\sum_{x\in D^0}|(u-u_h)(x)|^2\Big)^{\frac{1}{2}}, \quad E_{u_x} = \Big(\frac{1}{|D^1|}\sum_{x\in D^1}|(u-u_h)_x(x)|^2\Big)^{\frac{1}{2}},$$
$$E_{u_{xx}} = \Big(\frac{1}{|D^2|}\sum_{x\in D^2}|(u-u_h)_{xx}(x)|^2\Big)^{\frac{1}{2}} \tag{3.37}$$

be the average point value error for the numerical solution, the derivative of solution and the second order derivative of solution at corresponding sets of points. Then

- *if $k = 2$ and any of the assumptions A0/A1 is satisfied, we have*

$$E_u \le (C_4h^{4k} + \sum_{I_j\subset I^{NU}} C_2h^{4k-1})^{\frac{1}{2}}, \quad E_{u_x} \le h^{-1}(C_4h^{4k} + \sum_{I_j\subset I^{NU}} C_2h^{4k-1})^{\frac{1}{2}},$$
$$E_{u_{xx}} \le h^{-2}(C_4h^{4k} + \sum_{I_j\subset I^{NU}} C_2h^{4k-1})^{\frac{1}{2}}. \tag{3.38}$$

- *if $k \ge 3$ and any of the assumptions A0/A1/A2/A3 is satisfied, we have*

$$E_u \le C_{2,1}h^{k+2}, \quad E_{u_x} \le C_{2,1}h^{k+1}, \quad E_{u_{xx}} \le C_{2,1}h^k. \tag{3.39}$$

Proof. When $k = 2$, we have $u_h - P_h^\star u = -\zeta_h$. If any of the assumptions A0/A1 is satisfied, by (3.29), we have

$$\|u_h - P_h^\star u\| \le (C_4h^{4k} + \sum_{I_j\subset I^{NU}} C_2h^{4k-1})^{\frac{1}{2}}.$$

When $k \ge 3$, to relax the regularity requirement, we follow the same steps in Lemma

3.3.2, and change the definition of u_I to $u_I = P_h^\star u - w_1$. Then $\epsilon_h = u - u_I, \zeta_h = u_I - u_h$ and we obtain

$$|a(\epsilon_h, v_h)| \leq C_{2,1} h^{k+3} \|v_h\|, \quad \forall v_h \in V_h^k.$$

By the estimates above, (3.24) and the error equation, we obtain

$$\frac{d}{dt}\|\zeta_h\|^2 \leq 2\|(w_1)_t\|\|\zeta_h\| \leq C_{2,1} h^{k+3}.$$

By Gronwall's inequality,

$$\|\zeta_h\| \leq C_{2,1} t h^{k+3} + \|(\zeta_h)|_{t=0}\| = C_{2,1} t h^{k+3} + \|w_1|_{t=0}\| \leq C_{2,1}(1+t)h^{k+3}, \quad \forall t \in (0, T_e],$$

where $u_h|_{t=0} = P_h^\star u_0$ is used in the first equality. Since $u_h - P_h^\star u = -\zeta_h - w_1$, it follows that $\forall t \in (0, T_e]$,

$$\|u_h - P_h^\star u\| \leq 2(\|\zeta_h\| + \|w_1\|) \leq C_{2,1}(1+t)h^{k+3}.$$

Then the proof for (3.36) is complete.

If any of the assumptions A0/A1/A2/A3 is satisfied, then

$$\begin{aligned} E_u &\leq \Big(\frac{1}{|D^0|} \sum_{x \in D^0} |(u - P_h^\dagger u)(x)|^2 + |(P_h^\star u - u_h)(x)|^2 + |(P_h^\star u - P_h^\dagger u)(x)|^2 \Big)^{\frac{1}{2}} \\ &\leq C h^{k+2} |u|_{W^{k+2,2}(I)} + C\|P_h^\star u - u_h\| + C\|P_h^\star u - P_h^\dagger u\|, \end{aligned}$$

where (3.14), inverse inequality, and (3.10) are used in the last inequality. Then the estimates for E_u is proven by Lemma 3.2.3 and (3.36). The estimates for E_{u_x} and $E_{u_{xx}}$ can be proven following the same lines.

Remark 3.3.1. *If the initial discretization is taken as $u_h|_{t=0} = u_I|_{t=0}$, the theorem above still holds. However, the regularity requirement will be higher.*

$\square$

3.4 Superconvergence of postprocessed solution

In this section, we analyze the superconvergence property of the postprocessed DG solutions for linear Schrödinger equation (3.1) on uniform mesh by using negative Sobolev norm estimates. The postprocessor was originally introduced in [7, 55] for finite difference and finite element methods, and later applied to DG methods in [23]. The postprocessed solution is computed by the convolution of numerical solution u_h and a kernel function $K_h^{\nu,l}(x) = \frac{1}{h^d} K^{\nu,l}(\frac{x}{h})$, where d is the number of spatial dimensions, and l is the index of H^{-l} norm we're trying to estimate later. The convolution kernel has three main properties. First, it has compact support, making post processing computationally advantageous. Second, it preserves polynomials of degree up to $\nu - 1$ by convolution, thus the convergence rate is not deteriorated. Third, the kernel $K^{\nu,l}$ is a linear combination of B-splines, which allows us to express the derivatives of kernel by difference quotients (see section 4.1 in [23]).

We give the formula for the convolution kernel when the DG scheme uses approximation space V_h^k:

$$K^{2(k+1),k+1}(x) = \sum_{\gamma=-k}^{k} k_\gamma^{2(k+1),k+1} \psi^{(k+1)}(x - \gamma),$$

where $\psi^{(k+1)}$ are the B-spline bases and the computation of coefficients $k_\gamma^{2(k+1),k+1}$ can be found in [65]. Then we can define the postprocessed DG solution as

$$u^* = \int_{-\infty}^{\infty} K_h^{2(k+1),k+1}(y-x)u_h(y)dy. \tag{3.40}$$

u^* is an "averaged" version of u_h such that it is closer as an approximation to the exact solution u. Lastly, we define divided difference as

$$d_h v(x) = \frac{1}{h}(v(x+\frac{1}{2}h) - v(x-\frac{1}{2}h)).$$

Now we are ready to state an approximation result showing the smoothness of u and negative Sobolev norm of divided difference lead to a bound on $u - u^*$.

Theorem 3.4.1 (Bramble and Schatz [7]). *Suppose u^* is defined in* (3.40) *and $K_h^{2(k+1),k+1} = \frac{1}{h}K^{2(k+1),k+1}(\frac{x}{h})$, where $K^{2(k+1),k+1}$ is a kernel function as defined above. Let u be the exact solution of linear Schrödinger equation* (3.1) *satisfying periodic boundary condition, $u \in H^{2k+2}(I)$. Then for arbitrary time $t \in (0, T_e]$, h sufficiently small, we have*

$$\|u - u^*\| \le Ch^{2k+2}|u|_{H^{2k+2}(I)} + \sum_{\alpha \le k+1} \|d_h^\alpha(u-u_h)\|_{H^{-(k+1)}(\mathcal{I}_N)}, \tag{3.41}$$

where C is independent of u and h.

The right hand side of (3.41) indicates that if $\|d_h^\alpha(u-u_h)\|_{H^{-(k+1)}(\mathcal{I}_N)}$ converges at a rate higher than $k+1$, then we have superconvergence property for the postprocessed solution. In what follows, we estimate the negative-norm term following the steps in [23]. First, we introduce a dual problem: find a function v such that $v(\cdot,t)$ is periodic function with period equal to the length of I, i.e., $b-a$ for all $t \in (0, T_e]$ and

$$\begin{aligned} iv_t - v_{xx} &= 0, \quad \text{in } \ \mathbb{R} \times (0, T_e), \\ v(x, T_e) &= \Phi(x), \quad x \in \mathbb{R}, \end{aligned} \tag{3.42}$$

where Φ is an arbitrary function in $\mathcal{C}_0^\infty(I)$. We use the notation $(\phi, \psi) := \int_I \phi\, \psi dx$ in this section. At final time T_e,

$$\begin{aligned} (u(T_e) - u_h(T_e), \Phi) &= (u, v)(T_e) - (u_h, v)(T_e) \\ &= (u, v)(0) + \int_0^{T_e} \{(u, v_t) + (u_t, v)\} dt - (u_h, v)(T_e) \\ &= (u, v)(0) - (u_h, v)(0) - \int_0^{T_e} \{((u_h)_t, v) + (u_h, v_t)\} dt \\ &= (u - u_h, v)(0) - \int_0^{T_e} \{((u_h)_t, v) + (u_h, v_t)\} dt, \end{aligned}$$

where the property $uv_t + u_t v = 0$ is used to obtain the third equality.

The DG solution u_h satisfies (3.2). Therefore, we have $\forall v_h \in V_h^k$

$$\begin{aligned} ((u_h)_t, v) &= ((u_h)_t, v - v_h) + ((u_h)_t, v_h) \\ &= ((u_h)_t, v - v_h) + iA(u_h, v_h) \\ &= ((u_h)_t, v - v_h) - iA(u_h, v - v_h) + iA(u_h, v). \end{aligned}$$

Then we obtain

$$(u(T_e) - u_h(T_e), \Phi) = \Theta_M + \Theta_N + \Theta_C,$$

where

$$\Theta_M = (u - u_h, v)(0),$$
$$\Theta_N = -\int_0^{T_e} \{((u_h)_t, v - v_h) - iA(u_h, v - v_h)\}dt, \quad \forall v_h \in V_h^k,$$
$$\Theta_C = -\int_0^{T_e} \{(u_h, v_t) + iA(u_h, v)\}dt.$$

By choosing the initial numerical discretization $u_h(0) = P_h^0 u_0$ and $v_h = P_h^0 v$, we have $\Theta_M = (u - u_h, v)(0) = (u - u_h, v - v_h)(0)$ and

$$|\Theta_M| \leq \|(u - u_h)(0)\| \cdot \|(v - v_h)(0)\| \leq Ch^{2k+2} \|u\|_{H^{k+1}(I)} \|v\|_{H^{k+1}(I)}.$$

Since v is a smooth function, we have

$$\Theta_C = -\int_0^{T_e} \{(u_h, v_t) + i(u_h, v_{xx})\}dt = 0.$$

Choose $v_h = P_h^0 v$ and from the symmetry of the operator $A(\cdot, \cdot)$, we get

$$\begin{aligned}
|\Theta_N| &= \left| \int_0^{T_e} A(u_h, v - v_h)dt \right| = \left| \int_0^{T_e} A(v - v_h, u_h)dt \right| \\
&= \left| \int_0^{T_e} \sum_{j=1}^{N} \left(\widehat{v - v_h}[(u_h)_x] - \widetilde{(v - v_h)_x}[u_h] \right) \Big|_{j+\frac{1}{2}} dt \right| \\
&= \left| \int_0^{T_e} \sum_{j=1}^{N} \left(\widehat{v - v_h}[u_x - (u_h)_x] - \widetilde{(v - v_h)_x}[u - u_h] \right) \Big|_{j+\frac{1}{2}} dt \right| \\
&\leq CT_e \max_{t \in (0, T_e]} \left(\|u - u_h\|_{L^2(\partial \mathcal{I}_N)} \|\widetilde{(v - v_h)_x}\|_{L^2(\partial \mathcal{I}_N)} \right. \\
&\quad \left. + \|(u - u_h)_x\|_{L^2(\partial \mathcal{I}_N)} \|\widehat{v - v_h}\|_{L^2(\partial \mathcal{I}_N)} \right).
\end{aligned}$$

By (3.7),

$$\begin{aligned}\|u-u_h\|_{L^2(\partial\mathcal{I}_N)} &= \|u-P_h^\star u\|_{L^2(\partial\mathcal{I}_N)} + \|P_h^\star u-u_h\|_{L^2(\partial\mathcal{I}_N)} \\ &\le C_0 h^{k+\frac{1}{2}} + Ch^{-\frac{1}{2}}\|P_h^\star u-u_h\| \le C_2 h^{k+\frac{1}{2}},\end{aligned}$$

where we have used Lemma 2.2.3 and Theorem 3.2.2. Similarly, we have $\|u_x-(u_h)_x\|_{L^2(\partial\mathcal{I}_N)} \le C_2 h^{k-\frac{1}{2}}$. By the property of L^2 projection $h^{\frac{3}{2}}\|v_x-(v_h)_x\|_{L^2(\partial\mathcal{I}_N)} + h\|v_x-(v_h)_x\| + h^{\frac{1}{2}}\|v-v_h\|_{L^2(\partial\mathcal{I}_N)} + \|v-v_h\| \le Ch^{k+1}\|v\|_{H^{k+1}(I)}$. Then it is straightforward that for scale invariant fluxes,

$$\|\widehat{v-v_h}\|_{L^2(\partial\mathcal{I}_N)} \le Ch^{k+\frac{1}{2}}\|v\|_{H^{k+1}(I)}, \quad \|(\widetilde{v-v_h})_x\|_{L^2(\partial\mathcal{I}_N)} \le Ch^{k-\frac{1}{2}}\|v\|_{H^{k+1}(I)}.$$

Therefore, we have

$$|\Theta_N| \le C_2 h^{2k}\|v\|_{H^{k+1}(I)}. \tag{3.43}$$

Combine the above three estimate and the fact $\|v\|_{H^{k+1}(I)} = \|\Phi\|_{H^{k+1}(I)}$, we have

$$\|u(T_e)-u_h(T_e)\|_{H^{-(k+1)}(I)} \le C_2 h^{2k}.$$

Since we consider u_h with optimal error estimates on uniform mesh with mesh size h, then the divided difference $d_h^\alpha u$ satisfies the linear Schrödinger but with initial data $d_h^\alpha u_0, \alpha \le k+1$ on shifted mesh. Similarly, $d_h^\alpha u_h$ also satisfies the DG scheme (3.2) but with shifted mesh and initial numerical discretization $d_h^\alpha u_h = P_h^0 d_h^\alpha u_0$. Then by the same proof for $u-u_h$ above,

$$\|d_h^\alpha(u-u_h)(T_e)\|_{H^{-(k+1)}(I)} \le C_{2+\alpha} h^{2k}, \tag{3.44}$$

where we used Taylor expansion to estimate $d_h^\alpha u$ to obtain the last inequality.

The following theorem is a result of (3.44) and Theorem 3.4.1.

Theorem 3.4.2. *Let u_h be the UWDG solution of* (3.2)*, suppose the conditions in Theorem 3.4.1 and any of the assumptions A0/A1/A2/A3 is satisfied, then on a uniform mesh*

$$\|u(T_e) - u^*(T_e)\| \le C_{k+3} h^{2k}. \tag{3.45}$$

3.5 Numerical results

In this section, we provide numerical tests demonstrating superconvergence properties. In the proof, we see that the initial value of u_h matters in estimating $\|u_h - u_I\|$, thus will impact the superconvergence estimation for E_f and E_{f_x}. Therefore, in our numerical tests, we apply two types of initial discretization for u_h. For computing the postprocessed solution u^*, we use the standard L^2 projection $P_h^0 u$ as numerical initialization to demonstrate the convergence enhancement ability of postprocessor. For verifying other superconvergence quantities, we apply the initial condition $u_h|_{t=0} = u_I|_{t=0}$. In order not to deteriorate the high order convergence rates, for temporal discretization, we use explicit Runge-Kutta fourth order method with $dt = c \cdot h^{2.5}$, $c = 0.05$ when $k = 2$ and $c = 0.01$ when $k = 3, 4$.

Example 3.5.1. *We compute* (3.1) *on* $[0, 2\pi]$ *with exact solution* $u(x,t) = exp(i3(x - 3t))$ *using UWDG scheme* (3.2)*. We verify the results with several flux parameters.*

In the following tables, we show the convergence rate for quantities $E_f, E_{f_x}, E_c, E_u, E_{u_x}$, $E_{u_{xx}}$ as defined in (3.31) and (3.37) as well as

$$E^* = \|u - u^*\|, \quad E_P = \|u_h - P_h^\star u\|, \tag{3.46}$$

which represent the error after postprocessing, and the error between numerical solution and the projected exact solution $P_h^\star u$. In addition, we test the superconvergence of the intermediate quantities $\zeta_h = P_h^\star u - w - u_h$ as in Lemma 3.3.3, and introduce the following notations:

$$E_{[\zeta_h]} = (\frac{1}{N}\sum_{j=1}^{N} |[\zeta_h]|^2_{j+\frac{1}{2}})^{\frac{1}{2}}, \quad E_{[(\zeta_h)_x]} = (\frac{1}{N}\sum_{j=1}^{N} |[(\zeta_h)_x]|^2_{j+\frac{1}{2}})^{\frac{1}{2}}.$$

The numerical fluxes we tested include

1. Tables 3.1, 3.5: A0 parameters, alternating flux, $\alpha_1 = 0.5, \beta_1 = \beta_2 = 0$, with nonuniform mesh;
2. Tables 3.2, 3.6: A0 parameters, a scale invariant flux, $\alpha_1 = 0.3, \beta_1 = \frac{0.4}{h}, \beta_2 = 0.4h$, with nonuniform mesh;
3. Tables 3.3, 3.7: A1 parameters, central flux, $\alpha_1 = \beta_1 = \beta_2 = 0$, with uniform mesh;
4. Tables 3.4: A3 parameters, $\alpha_1 = 0.25, \beta_2 = 0, \beta_1 = \frac{2}{h}, \frac{5}{h}, \frac{9}{h}$ for $k = 2, 3, 4$, respectively, with uniform mesh;
5. Table 3.8: all parameters mentioned above, with uniform mesh,

where the nonuniform mesh is generated by perturbing the location of the nodes of a uniform mesh by 10% of mesh size.

We first verify the results in Theorems 3.3.5, 3.3.6 by examining Tables 3.1, 3.2, 3.3, 3.4, where the parameters satisfy assumption A0, A0, A1, A3, respectively. We observe that the scheme can achieve at least the theoretical order of convergence for the quantities in these two theorems. To be more specific, E_P shows $(k + \min(3, k))$-th order of convergence.

Table 3.1: Example 3.5.1. Error table when using alternating flux on nonuniform mesh. Ending time $T_e = 1$, $x \in [0, 2\pi]$.

	N	L^2 error	order	E_P	order	$E_{u_{xx}}$	order	E_{u_x}	order
	10	2.68E-01	-	2.53E-01	-	2.36E+00	-	8.92E-01	-
	20	2.68E-02	3.32	2.47E-02	3.36	2.78E-01	3.09	7.43E-02	3.59
P^2	40	2.00E-03	3.75	1.42E-03	4.12	6.02E-02	2.21	5.29E-03	3.81
	80	1.91E-04	3.39	9.22E-05	3.95	1.50E-02	2.00	3.83E-04	3.79
	160	2.19E-05	3.12	5.83E-06	3.98	3.78E-03	1.99	3.24E-05	3.56
	10	1.02E-02	-	7.51E-03	-	1.60E-01	-	2.50E-02	-
	20	5.65E-04	4.18	1.24E-04	5.93	2.00E-02	3.00	8.23E-04	4.93
P^3	40	2.94E-05	4.26	2.13E-06	5.86	2.43E-03	3.04	3.83E-05	4.42
	80	1.87E-06	3.98	3.02E-08	6.14	2.99E-04	3.02	2.29E-06	4.06
	160	1.18E-07	3.98	4.57E-10	6.05	3.76E-05	2.99	1.45E-07	3.98
	10	6.45E-04	-	8.76E-05	-	1.79E-02	-	8.67E-04	-
	20	2.06E-05	4.97	6.12E-07	7.16	1.25E-03	3.84	3.06E-05	4.82
P^4	40	6.83E-07	4.91	2.52E-09	7.92	7.45E-05	4.06	9.00E-07	5.09
	80	2.04E-08	5.07	1.55E-11	7.35	4.70E-06	3.99	2.78E-08	5.02
	160	6.10E-10	5.06	1.02E-13	7.24	2.85E-07	4.04	8.34E-10	5.06
	N	E_u	order	E_f	order	E_{f_x}	order	E_c	order
	10	2.83E-01	-	2.79E-01	-	9.10E-01	-	5.00E-01	-
	20	2.30E-02	3.62	2.31E-02	3.60	6.94E-02	3.71	4.48E-02	3.48
P^2	40	1.47E-03	3.97	1.46E-03	3.98	4.45E-03	3.96	2.91E-03	3.94
	80	9.12E-05	4.01	9.11E-05	4.01	2.75E-04	4.02	1.82E-04	4.00
	160	5.78E-06	3.98	5.77E-06	3.98	1.74E-05	3.98	1.15E-05	3.98
	10	6.81E-03	-	6.78E-03	-	1.98E-02	-	1.16E-02	-
	20	1.40E-04	5.61	1.36E-04	5.64	4.08E-04	5.60	2.64E-04	5.45
P^3	40	2.16E-06	6.01	2.02E-06	6.07	6.06E-06	6.07	4.03E-06	6.03
	80	3.73E-08	5.86	3.05E-08	6.05	9.13E-08	6.05	6.11E-08	6.04
	160	8.04E-10	5.53	4.59E-10	6.05	1.38E-09	6.05	9.25E-10	6.05
	10	1.09E-04	-	9.69E-05	-	2.83E-04	-	1.62E-04	-
	20	1.04E-06	6.72	5.02E-07	7.59	1.57E-06	7.49	9.69E-07	7.39
P^4	40	1.28E-08	6.34	1.76E-09	8.16	5.26E-09	8.22	3.50E-09	8.11
	80	1.94E-10	6.04	6.99E-12	7.97	2.09E-11	7.97	1.40E-11	7.97
	160	2.85E-12	6.09	2.79E-14	7.97	5.55E-13	5.23	5.07E-14	8.10

$E_u, E_{u_x}, E_{u_{xx}}$ are shown to have $(k+2)$-th, $(k+1)$-th and k-th order of convergence , respectively. Note that when $k = 2$, in Tables 3.2 and 3.4, there are situations when no superconvergence point exists. This finding shows an evidence to the assertion that D^s defined in (3.14) could be empty sets. The order of convergence for $E_f, (E_f)_x, E_c$ in all tables are $2k$. In addition, Table 3.4 shows that when k is even and assumption A3 is satisfied, the convergence order for all quantities are the same as when any of assumption

A0/A1 is satisfied, which is one order higher than the estimates in Theorems 3.3.5, 3.3.6.

Table 3.2: Example 3.5.1. Error table when using flux parameters: $\alpha_1 = 0.3, \beta_1 = \frac{0.4}{h}, \beta_2 = 0.4h$ on nonuniform mesh. Ending time $T_e = 1$, $x \in [0, 2\pi]$.

	N	L^2 error	order	E_P	order	$E_{u_{xx}}$	order	E_{u_x}	order
P^2	40	1.66E-02	-	1.28E-02	-	2.38E-01	-	DNE	-
	80	1.50E-03	3.47	7.46E-04	4.10	2.52E-02	3.24	DNE	-
	160	1.70E-04	3.14	4.91E-05	3.92	4.51E-03	2.48	DNE	-
	320	2.16E-05	2.98	3.12E-06	3.98	8.89E-04	2.34	DNE	-
	640	2.64E-06	3.03	1.96E-07	3.99	2.00E-04	2.15	DNE	-
P^3	10	2.08E-02	-	1.57E-02	-	2.00E-01	-	4.47E-02	-
	20	1.14E-03	4.19	2.75E-04	5.84	2.23E-02	3.17	1.42E-03	4.98
	40	5.91E-05	4.27	4.83E-06	5.83	2.72E-03	3.03	6.57E-05	4.43
	80	3.75E-06	3.98	6.81E-08	6.15	3.36E-04	3.02	3.91E-06	4.07
	160	2.39E-07	3.97	1.09E-09	5.96	4.23E-05	2.99	2.47E-07	3.98
P^4	10	9.72E-04	-	1.48E-04	-	1.93E-02	-	1.32E-03	-
	20	3.17E-05	4.94	1.04E-06	7.16	1.37E-03	3.82	4.63E-05	4.83
	40	1.05E-06	4.91	4.15E-09	7.97	8.18E-05	4.06	1.35E-06	5.10
	80	3.14E-08	5.07	2.51E-11	7.37	5.16E-06	3.99	4.27E-08	4.99
	160	9.39E-10	5.06	1.63E-13	7.27	3.13E-07	4.04	1.27E-09	5.07
	N	E_u	order	E_f	order	E_{f_x}	order	E_c	order
P^2	40	1.26E-02	-	1.26E-02	-	3.83E-02	-	2.49E-02	-
	80	7.43E-04	4.08	7.42E-04	4.08	2.28E-03	4.07	1.48E-03	4.07
	160	4.82E-05	3.95	4.82E-05	3.95	1.47E-04	3.95	9.63E-05	3.94
	320	3.09E-06	3.96	3.09E-06	3.96	9.39E-06	3.97	6.18E-06	3.96
	640	1.94E-07	3.99	1.94E-07	3.99	5.86E-07	4.00	3.88E-07	3.99
P^3	10	1.44E-02	-	1.41E-02	-	4.38E-02	-	2.49E-02	-
	20	3.08E-04	5.54	3.05E-04	5.53	9.17E-04	5.58	5.92E-04	5.39
	40	4.70E-06	6.04	4.56E-06	6.06	1.37E-05	6.06	9.12E-06	6.02
	80	7.56E-08	5.96	6.88E-08	6.05	2.06E-07	6.05	1.38E-07	6.04
	160	1.03E-09	6.20	1.10E-09	5.97	3.29E-09	5.97	2.21E-09	5.97
P^4	10	1.77E-04	-	1.65E-04	-	4.69E-04	-	2.73E-04	-
	20	1.56E-06	6.83	8.66E-07	7.57	2.52E-06	7.54	1.66E-06	7.36
	40	1.83E-08	6.41	3.04E-09	8.15	8.85E-09	8.15	6.02E-09	8.11
	80	2.82E-10	6.02	1.21E-11	7.98	3.76E-11	7.88	2.41E-11	7.97
	160	4.12E-12	6.10	4.41E-14	8.10	1.30E-13	8.18	8.73E-14	8.11

In Tables 3.1 and 3.2, we used nonuniform mesh in numerical test. The quantities tested have similar order of convergence compared to the order of convergence on uniform mesh. Another interesting observation is the order of convergence of E_{f_x}. Our numerical tests show that E_{f_x} converges at an order of $2k$ for all four sets of parameters, which is at least one order higher than the estimates in Theorem 3.3.5.

Table 3.3: Example 3.5.1. Error table when using central flux on uniform mesh. Ending time $T_e = 1$, $x \in [0, 2\pi]$.

	N	L^2 error	order	E_P	order	$E_{u_{xx}}$	order	E_{u_x}	order
P^2	40	4.20E-03	-	3.21E-03	-	4.86E-01	-	3.39E-02	-
	80	4.31E-04	3.29	2.23E-04	3.85	1.33E-01	1.87	4.49E-03	2.92
	160	4.92E-05	3.13	1.43E-05	3.96	3.41E-02	1.97	5.69E-04	2.98
	320	5.99E-06	3.04	9.01E-07	3.99	8.57E-03	1.99	7.14E-05	2.99
	640	7.44E-07	3.01	5.60E-08	4.01	2.15E-03	2.00	8.94E-06	3.00
P^3	20	3.18E-04	-	7.32E-05	-	4.31E-02	-	3.34E-03	-
	40	1.71E-05	4.21	1.02E-06	6.16	5.49E-03	2.97	2.04E-04	4.03
	80	1.03E-06	4.05	1.55E-08	6.04	6.89E-04	2.99	1.27E-05	4.01
	160	6.41E-08	4.01	2.41E-10	6.01	8.62E-05	3.00	7.91E-07	4.00
	320	4.00E-09	4.00	3.76E-12	6.00	1.08E-05	3.00	4.94E-08	4.00
P^4	10	5.04E-04	-	7.75E-05	-	1.14E-01	-	1.32E-02	-
	20	2.10E-05	4.58	4.91E-07	7.30	1.11E-02	3.36	6.17E-04	4.42
	40	7.32E-07	4.84	2.65E-09	7.53	8.01E-04	3.79	2.17E-05	4.83
	80	2.36E-08	4.96	1.60E-11	7.37	5.21E-05	3.94	6.76E-07	5.01
	160	7.42E-10	4.99	1.13E-13	7.15	3.29E-06	3.99	2.19E-08	4.95
	N	E_u	order	E_f	order	E_{f_x}	order	E_c	order
P^2	40	3.24E-03	-	3.21E-03	-	9.58E-03	-	6.36E-03	-
	80	2.25E-04	3.85	2.23E-04	3.85	6.86E-04	3.80	4.44E-04	3.84
	160	1.45E-05	3.96	1.43E-05	3.96	3.90E-05	4.14	2.86E-05	3.96
	320	9.10E-07	3.99	9.01E-07	3.99	3.00E-06	3.70	1.80E-06	3.99
	640	5.66E-08	4.01	5.60E-08	4.01	1.51E-07	4.31	1.12E-07	4.01
P^3	20	1.88E-04	-	7.28E-05	-	2.16E-04	-	1.41E-04	-
	40	5.17E-06	5.19	1.02E-06	6.16	3.07E-06	6.14	2.03E-06	6.12
	80	1.63E-07	4.99	1.54E-08	6.04	4.63E-08	6.05	3.10E-08	6.03
	160	5.04E-09	5.01	2.39E-10	6.01	7.18E-10	6.01	4.81E-10	6.01
	320	1.57E-10	5.01	3.73E-12	6.00	1.12E-11	6.00	7.51E-12	6.00
P^4	10	2.21E-04	-	7.63E-05	-	2.15E-04	-	1.27E-04	-
	20	5.70E-06	5.28	4.52E-07	7.40	1.26E-06	7.42	8.66E-07	7.20
	40	1.05E-07	5.76	2.05E-09	7.78	6.17E-09	7.67	4.04E-09	7.74
	80	1.72E-09	5.93	8.31E-12	7.94	2.57E-11	7.91	1.66E-11	7.93
	160	2.72E-11	5.98	3.27E-14	7.99	9.96E-14	8.01	6.55E-14	7.98

Next, we test the order of convergence for quantities in Lemma 3.3.3. In Tables 3.5 and 3.6, we observe clean convergence order of $2k-1$, $2k+1$, $2k$ for $\|(\zeta_h)_{xx}\|$, $E_{[\zeta_h]}$, $E_{[(\zeta_h)_x]}$ when k is even and $2k$, $2k+2$, $2k+1$ for these three quantities when k is odd. In Table 3.7, the order of convergence has some fluctuation, but the quantities are shown to have the same order of convergence as those in Tables 3.5 and 3.6. These convergence rates are consistent with the results in Lemma 3.3.3.

Table 3.4: Example 3.5.1. Error table when using flux parameters: $\alpha_1 = 0.25, \beta_1 = \frac{2}{h}, \frac{5}{h}, \frac{9}{h}, \beta_2 = 0$ on uniform mesh. Ending time $T_e = 1$, $x \in [0, 2\pi]$.

	N	L^2 error	order	E_P	order	$E_{u_{xx}}$	order	E_{u_x}	order
P^2	80	1.41E-03	-	8.17E-05	-	DNE	-	1.15E-02	-
	160	1.65E-04	3.09	4.74E-06	4.11	DNE	-	1.34E-03	3.11
	320	2.03E-05	3.02	2.92E-07	4.02	DNE	-	1.65E-04	3.02
	640	2.53E-06	3.01	1.80E-08	4.03	DNE	-	2.05E-05	3.01
	1280	3.16E-07	3.00	1.22E-09	3.88	DNE	-	2.55E-06	3.01
P^3	20	8.27E-04	-	4.58E-05	-	2.98E-01	-	6.63E-03	-
	40	3.92E-05	4.40	5.20E-07	6.46	3.11E-02	3.26	3.60E-04	4.20
	80	2.29E-06	4.10	7.54E-09	6.11	3.72E-03	3.06	2.18E-05	4.05
	160	1.40E-07	4.03	1.16E-10	6.03	4.60E-04	3.02	1.35E-06	4.01
	320	8.74E-09	4.01	1.80E-12	6.01	5.74E-05	3.00	8.43E-08	4.00
P^4	20	5.10E-04	-	2.08E-04	-	3.76E-01	-	3.96E-03	-
	40	8.28E-06	5.95	2.38E-07	9.77	1.24E-02	4.92	6.76E-05	5.87
	80	1.87E-07	5.47	1.04E-09	7.84	5.64E-04	4.47	1.55E-06	5.45
	160	5.44E-09	5.10	7.11E-12	7.19	3.29E-05	4.10	4.53E-08	5.09
	N	E_u	order	E_f	order	E_{f_x}	order	E_c	order
P^2	40	3.24E-03	-	3.21E-03	-	9.58E-03	-	6.36E-03	-
	80	2.25E-04	3.85	2.23E-04	3.85	6.86E-04	3.80	4.44E-04	3.84
	160	1.45E-05	3.96	1.43E-05	3.96	3.90E-05	4.14	2.86E-05	3.96
	320	9.10E-07	3.99	9.01E-07	3.99	3.00E-06	3.70	1.80E-06	3.99
	640	5.66E-08	4.01	5.60E-08	4.01	1.51E-07	4.31	1.12E-07	4.01
P^3	20	1.88E-04	-	7.28E-05	-	2.16E-04	-	1.41E-04	-
	40	5.17E-06	5.19	1.02E-06	6.16	3.07E-06	6.14	2.03E-06	6.12
	80	1.63E-07	4.99	1.54E-08	6.04	4.63E-08	6.05	3.10E-08	6.03
	160	5.04E-09	5.01	2.39E-10	6.01	7.18E-10	6.01	4.81E-10	6.01
	320	1.57E-10	5.01	3.73E-12	6.00	1.12E-11	6.00	7.51E-12	6.00
P^4	10	2.21E-04	-	7.63E-05	-	2.15E-04	-	1.27E-04	-
	20	5.70E-06	5.28	4.52E-07	7.40	1.26E-06	7.42	8.66E-07	7.20
	40	1.05E-07	5.76	2.05E-09	7.78	6.17E-09	7.67	4.04E-09	7.74
	80	1.72E-09	5.93	8.31E-12	7.94	2.57E-11	7.91	1.66E-11	7.93
	160	2.72E-11	5.98	3.27E-14	7.99	9.96E-14	8.01	6.55E-14	7.98

Lastly, we test the order of convergence for E^* on uniform mesh for the four sets of parameters. Table 3.8 shows that E^* has a convergence rate of at least $2k$, and can go up to $2k + 2$. Similar higher order of convergence behaviors exists in the literature [23, 65].

Table 3.5: Example 3.5.1. Error table for intermediate quantities when using alternating flux on nonuniform mesh. Ending time $T_e = 1$, $x \in [0, 2\pi]$.

	N	$\|\zeta_h\|$ error	order	$\|(\zeta_h)_{xx}\|$	order	$E_{[\zeta_h]}$	order	$E_{[(\zeta_h)_x]}$	order
P^2	10	3.96E-01	-	3.02E+00	-	4.51E-02	-	3.37E-01	-
	20	3.28E-02	3.60	2.23E-01	3.76	1.57E-03	4.84	1.82E-02	4.21
	40	2.08E-03	3.98	1.42E-02	3.98	4.21E-05	5.22	9.42E-04	4.28
	80	1.29E-04	4.01	7.95E-04	4.15	1.17E-06	5.16	5.50E-05	4.10
	160	8.17E-06	3.98	9.54E-05	3.06	3.83E-08	4.94	3.55E-06	3.96
P^3	10	9.54E-03	-	8.83E-02	-	7.52E-05	-	1.31E-03	-
	20	1.93E-04	5.63	1.76E-03	5.65	1.75E-07	8.75	6.16E-06	7.74
	40	2.86E-06	6.08	2.59E-05	6.09	3.42E-10	9.00	2.45E-08	7.97
	80	4.31E-08	6.05	3.91E-07	6.05	1.45E-12	7.88	2.71E-10	6.50
	160	6.87E-10	5.97	6.19E-09	5.98	6.76E-15	7.74	2.59E-12	6.71
P^4	10	1.35E-04	-	1.41E-03	-	4.97E-07	-	2.81E-05	-
	20	7.10E-07	7.57	8.60E-06	7.36	1.20E-09	8.69	9.88E-08	8.15
	40	2.50E-09	8.15	4.24E-08	7.66	2.56E-12	8.88	2.90E-10	8.41
	80	9.90E-12	7.98	2.55E-10	7.38	2.98E-15	9.74	7.48E-13	8.60
	160	3.58E-14	8.11	2.22E-12	6.85	8.90E-18	8.39	5.46E-15	7.10

Table 3.6: Example 3.5.1. Error table for intermediate quantities when using flux parameters: $\alpha_1 = 0.3, \beta_1 = \frac{0.4}{h}, \beta_2 = 0.4h$ on nonuniform mesh. Ending time $T_e = 1$, $x \in [0, 2\pi]$.

	N	$\|\zeta_h\|$ error	order	$\|(\zeta_h)_{xx}\|$	order	$E_{[\zeta_h]}$	order	$E_{[(\zeta_h)_x]}$	order
P^2	40	1.46E-02	-	2.67E-01	-	2.55E-03	-	2.13E-02	-
	80	9.35E-04	3.97	2.57E-02	3.38	7.74E-05	5.04	1.25E-03	4.09
	160	5.96E-05	3.97	2.86E-03	3.17	2.52E-06	4.94	7.56E-05	4.05
	320	3.76E-06	3.99	3.30E-04	3.11	7.74E-08	5.02	4.76E-06	3.99
	640	2.38E-07	3.98	4.25E-05	2.96	2.57E-09	4.91	3.19E-07	3.90
P^3	10	2.02E-02	-	1.78E-01	-	5.73E-04	-	1.15E-03	-
	20	4.31E-04	5.55	3.90E-03	5.52	1.09E-06	9.04	5.48E-06	7.71
	40	6.46E-06	6.06	5.83E-05	6.06	3.27E-09	8.38	2.99E-08	7.52
	80	9.72E-08	6.05	8.76E-07	6.06	9.55E-12	8.42	1.84E-10	7.34
	160	1.55E-09	5.97	1.40E-08	5.97	4.28E-14	7.80	1.56E-12	6.89
P^4	10	2.27E-04	-	2.23E-03	-	1.03E-06	-	2.06E-06	-
	20	1.22E-06	7.54	1.30E-05	7.42	4.13E-09	7.96	1.66E-08	6.96
	40	4.30E-09	8.15	5.93E-08	7.78	1.01E-11	8.67	6.95E-11	7.90
	80	1.71E-11	7.98	4.66E-10	6.99	1.82E-14	9.12	2.33E-13	8.22
	160	6.17E-14	8.11	3.19E-12	7.19	3.61E-17	8.98	8.34E-16	8.13

Table 3.7: Example 3.5.1. Error table for intermediate quantities when using central flux on uniform mesh. Ending time $T_e = 1$, $x \in [0, 2\pi]$.

	N	$\|\zeta_h\|$ error	order	$\|(\zeta_h)_{xx}\|$	order	$E_{[\zeta_h]}$	order	$E_{[(\zeta_h)_x]}$	order
P^2	40	4.53E-03	-	3.84E-02	-	3.04E-05	-	2.79E-04	-
	80	3.15E-04	3.85	3.03E-03	3.66	1.16E-06	4.71	8.31E-06	5.07
	160	2.02E-05	3.96	1.29E-04	4.55	1.79E-07	2.70	1.31E-06	2.66
	320	1.27E-06	3.99	1.59E-05	3.03	7.14E-09	4.65	5.08E-08	4.69
	640	7.92E-08	4.01	5.73E-07	4.79	2.90E-10	4.62	2.11E-09	4.59
P^3	20	1.03E-04	-	9.27E-04	-	4.27E-08	-	5.25E-06	-
	40	1.44E-06	6.16	1.29E-05	6.17	1.75E-10	7.93	4.32E-08	6.93
	80	2.18E-08	6.04	1.99E-07	6.02	4.23E-13	8.69	1.64E-10	8.04
	160	3.38E-10	6.01	3.05E-09	6.03	2.91E-16	10.50	3.24E-14	12.31
	320	5.28E-12	6.00	4.76E-11	6.00	1.28E-18	7.83	1.41E-15	4.52
P^4	10	1.06E-04	-	1.05E-03	-	7.67E-07	-	2.26E-05	-
	20	6.37E-07	7.37	7.12E-06	7.21	3.03E-09	7.99	7.70E-08	8.20
	40	2.88E-09	7.79	3.20E-08	7.80	5.78E-12	9.03	6.36E-11	10.24
	80	1.17E-11	7.94	1.84E-10	7.44	9.33E-15	9.28	5.24E-13	6.92
	160	4.63E-14	7.98	2.45E-12	6.23	3.04E-17	8.26	8.82E-16	9.21

Table 3.8: Example 3.5.1. Postprocessing error table for the four sets of parameters. Ending time $T_e = 1$, uniform mesh on $x \in [0, 2\pi]$. The first row below labels the parameters by $(\tilde{\alpha_1}, \tilde{\beta_1}, \tilde{\beta_2})$.

Fluxes		(0.5,0,0)		(0, 0, 0)		(0.3, 0.4, 0.4)		(0.25, {2, 5, 9}, 0)	
	N	E^*	order	E^*	order	E^*	order	E^*	order
P^2	10	1.00E+00	-	2.81E-01	-	1.00E+00	-	1.53E-01	-
	20	2.84E-01	1.81	3.71E-02	2.92	1.20E-01	3.06	8.05E-02	0.93
	40	2.11E-02	3.75	3.23E-03	3.52	9.63E-03	3.64	2.68E-03	4.91
	80	1.37E-03	3.94	2.24E-04	3.85	7.55E-04	3.67	1.20E-04	4.49
	160	8.69E-05	3.98	1.44E-05	3.96	5.13E-05	3.88	6.99E-06	4.10
P^3	10	1.00E+00	-	1.00E+00	-	1.00E+00	-	1.00E+00	-
	20	6.04E-02	4.05	6.29E-02	3.99	6.05E-02	4.05	7.02E-02	3.83
	40	5.39E-04	6.81	6.05E-04	6.70	5.26E-04	6.85	5.46E-04	7.01
	80	3.28E-06	7.36	5.04E-06	6.91	2.82E-06	7.54	2.91E-06	7.55
	160	3.14E-08	6.70	6.49E-08	6.28	2.04E-08	7.11	1.79E-08	7.34
P^4	10	1.00E+00	-	1.00E+00	-	1.00E+00	-	1.00E+00	-
	20	4.54E-02	4.46	4.54E-02	4.46	4.54E-02	4.46	4.54E-02	4.46
	40	1.32E-04	8.42	1.32E-04	8.42	1.32E-04	8.42	1.36E-04	8.39
	80	1.70E-07	9.60	1.70E-07	9.60	1.70E-07	9.60	1.66E-07	9.67
	160	1.79E-10	9.89	1.80E-10	9.89	1.79E-10	9.89	1.75E-10	9.89

Chapter 4

Sparse grid central DG methods for linear hyperbolic systems

In this chapter, we develop sparse grid central discontinuous Galerkin (CDG) method for the following time-dependent linear hyperbolic system with variable coefficients

$$\frac{\partial \mathbf{u}}{\partial t} + \sum_{i=1}^{d} \frac{\partial (A_i(t,\mathbf{x})\mathbf{u})}{\partial x_i} = \mathbf{0}, \quad \mathbf{x} \in \Omega, \tag{4.1}$$

subject to appropriate initial and boundary conditions. In the expression above, $d \geq 2$ is the spatial dimension of the problem, $\mathbf{u}(t,\mathbf{x}) = (u^1(t,\mathbf{x}), \cdots, u^m(t,\mathbf{x}))^T$ is the unknown function, $A_i(t,\mathbf{x}) \in \mathbb{R}^{m\times m}, i = 1,\ldots,d$ are the given smooth variable coefficients. We assume $\Omega = [0,1]^d$ in this chapter, but the discussion can be easily generalized to arbitrary box-shaped domains. The model (4.1) arises in many contexts [46], such as simulations of acoustic, elastic waves, and Maxwell's equations in free space. The scheme we develop in this chapter can also apply to the case when $A_i(t,\mathbf{x})$ is defined through another set of equations that can be nonlinearly coupled with $\mathbf{u}$, such as the models in kinetic plasma waves and incompressible flows.

Similar to [38], in this chapter, we restrict our attention to smooth solutions of (4.1). It is known that for non-smooth solutions, adaptivity should be invoked to capture discontinuity like structures. This can be achieved using the idea in [39] and is left for our future work. Based on the sparse grid DG scheme constructed in [38], the goal of the present chapter is to design and analyze the sparse grid CDG method. Motivated by the Riemann-solver-free property and large CFL number allowance of CDG methods, we develop sparse grid CDG method to compute hyperbolic systems efficiently. We investigate stability, convergence rate and CFL condition of the resulting scheme. A novelty of this work is the design of the scheme for non-periodic problems, where a new hierarchical representation of the solution is presented, which results in a sparse finite element space that can be defined on the dual mesh. L^2 projection results are studied for this space, which helps the convergence proof of the schemes for initial-boundary value problems.

The rest of this chapter is organized as follows: in Section 4.1, we construct the sparse grid CDG formulations for periodic and non-periodic problems, and perform numerical study of the CFL conditions. In Section 4.2, we prove L^2 stability and error estimates for scalar equations. The numerical performance is validated in Section 4.3 by several benchmark tests, including scalar transport equations, acoustic and elastic waves.

The contents of this chapter has been published in [75].

4.1 Numerical Scheme

In this section, we define and discuss the properties of the proposed sparse grid CDG methods. For convenience of notations, we rewrite (4.1) in a component-wise form as

$$\frac{\partial u^l}{\partial t} + \nabla \cdot (A^l(t, \mathbf{x})\mathbf{u}) = 0, \quad l = 1, \cdots, m, \quad \mathbf{x} \in \Omega, \tag{4.2}$$

where $A^l(t,\mathbf{x}) = (A_1^l(t,\mathbf{x}),\cdots,A_d^l(t,\mathbf{x}))^T \in \mathbb{R}^{d\times m}$ denotes a collection of the l-th row of each matrix A_i. The problem is solved with given initial value $\mathbf{u}(0,\mathbf{x}) = \mathbf{u}_0(\mathbf{x})$, and periodic or Dirichlet type boundary conditions.

We proceed as follows. First, we introduce the scheme for periodic problems. In this setting, the finite element space on the primal and dual mesh can be defined in similar ways. Then, we discuss the implementation details and perform numerical study of the CFL conditions. Finally, we consider the more complicated non-periodic problems, for which a new sparse finite element space will be introduced on the dual mesh.

4.1.1 Periodic problems

To define the sparse finite element space, we first review the hierarchical decomposition of piecewise polynomial space in one dimension [76]. Consider a general interval $[a,b]$, we define the n-th level mesh $\Omega_n([a,b])$ to be a uniform partition of 2^n cells with length $h_n = 2^{-n}(b-a)$ and $I_n^j = [a+jh_n, a+(j+1)h_n]$, $j=0,\dots,2^n-1$, for any $n\ge 0$. Let

$$V_n^k([a,b]) := \{v : v \in P^k(I_n^j),\ \forall\, j=0,\dots,2^n-1\}$$

be the usual piecewise polynomials of degree at most k on Ω_n. Then, we have the nested structure

$$V_0^k([a,b]) \subset V_1^k([a,b]) \subset V_2^k([a,b]) \subset V_3^k([a,b]) \subset \cdots$$

Similar to [76], we can now define the multiwavelet subspace $W_n^k([a,b])$, $n=1,2,\dots$ as the orthogonal complement of $V_{n-1}^k([a,b])$ in $V_n^k([a,b])$ with respect to the L^2 inner product on $[a,b]$, i.e.,

$$V^k_{n-1}([a,b]) \oplus W^k_n([a,b]) = V^k_n([a,b]), \quad W^k_n([a,b]) \perp V^k_{n-1}([a,b]).$$

For notational convenience, we let $W^k_0([a,b]) := V^k_0([a,b])$, which is the standard piecewise polynomial space of degree k on $[a,b]$. This gives the hierarchical decomposition $V^k_n([a,b])$ on Ω_n as $V^k_n([a,b]) = \bigoplus_{0\le l\le n} W^k_l([a,b])$.

For a d dimensional domain $[a,b]^d$, we recall some basic notations about multi-indices. For a multi-index $\alpha = (\alpha_1, \cdots, \alpha_d) \in \mathbb{N}^d_0$, where $\mathbb{N}_0$ denotes the set of nonnegative integers, the l^1 and l^∞ norms are defined as

$$|\boldsymbol{\alpha}|_1 := \sum_{i=1}^d \alpha_i, \qquad |\boldsymbol{\alpha}|_\infty := \max_{1\le i\le d} \alpha_i.$$

The component-wise arithmetic operations and relational operations are defined as

$$\boldsymbol{\alpha}\cdot\boldsymbol{\beta} := (\alpha_1\beta_1, \ldots, \alpha_d\beta_d), \qquad c\cdot\boldsymbol{\alpha} := (c\alpha_1, \ldots, c\alpha_d), \qquad 2^{\boldsymbol{\alpha}} := (2^{\alpha_1}, \ldots, 2^{\alpha_d}),$$

$$\boldsymbol{\alpha} \le \boldsymbol{\beta} \Leftrightarrow \alpha_i \le \beta_i, \forall i, \quad \boldsymbol{\alpha} < \boldsymbol{\beta} \Leftrightarrow \boldsymbol{\alpha} \le \boldsymbol{\beta} \text{ and } \boldsymbol{\alpha} \ne \boldsymbol{\beta}.$$

By making use of the multi-index notation, we denote by $\mathbf{l} = (l_1, \cdots, l_d) \in \mathbb{N}^d_0$ the mesh level in a multivariate sense. We define the tensor-product mesh grid $\Omega_{\mathbf{l}}([a,b]^d) = \Omega_{l_1}([a,b]) \otimes \cdots \otimes \Omega_{l_d}([a,b])$ and the corresponding mesh size $h_{\mathbf{l}} = (h_{l_1}, \cdots, h_{l_d})$. Based on the grid $\Omega_{\mathbf{l}}$, we denote by $I^{\mathbf{j}}_{\mathbf{l}} = \{\mathbf{x} : x_i \in I^{j_i}_{l_i}, i = 1, \cdots, d\}$ as an elementary cell, and

$$\mathbf{V}^k_{\mathbf{l}}([a,b]^d) := \{v : v(\mathbf{x}) \in Q^k(I^{\mathbf{j}}_{\mathbf{l}}),\ \mathbf{0} \le \mathbf{j} \le 2^{\mathbf{l}} - \mathbf{1}\} = V^k_{l_1,x_1}([a,b]) \times \cdots \times V^k_{l_d,x_d}([a,b])$$

as the standard tensor-product piecewise polynomial space on this mesh, where $Q^k(I_{\mathbf{l}}^{\mathbf{j}})$ denotes the collection of polynomials of degree up to k in each dimension on cell $I_{\mathbf{l}}^{\mathbf{j}}$. If $\mathbf{l} = (N, \cdots, N)$, the grid and space will be further denoted by $\Omega_N([a,b]^d)$ and $\mathbf{V}_N^k([a,b]^d)$, respectively.

Based on a tensor-product construction, the multi-dimensional increment space can be defined as

$$\mathbf{W}_{\mathbf{l}}^k([a,b]^d) = W_{l_1,x_1}^k([a,b]) \times \cdots \times W_{l_d,x_d}^k([a,b]).$$

Therefore, we have $\mathbf{V}_N^k([a,b]^d) = \bigoplus_{\substack{|\mathbf{l}|_\infty \leq N \\ \mathbf{l} \in \mathbb{N}_0^d}} \mathbf{W}_{\mathbf{l}}^k([a,b]^d)$. The sparse finite element approximation space we consider, is defined by

$$\hat{\mathbf{V}}_N^k([a,b]^d) := \bigoplus_{\substack{|\mathbf{l}|_1 \leq N \\ \mathbf{l} \in \mathbb{N}_0^d}} \mathbf{W}_{\mathbf{l}}^k([a,b]^d).$$

This is a subset of $\mathbf{V}_N^k([a,b]^d)$, and its number of degrees of freedom scales as $O((k+1)^d 2^N N^{d-1})$ [76], which is significantly less than that of $\mathbf{V}_N^k([a,b]^d)$ with exponential dependence on Nd. This is the key to computational savings in high dimensions.

The standard CDG schemes [48, 50] is characterized by numerical approximations on two sets of overlapping grids: primal and dual meshes. Now, we are ready to incorporate the sparse finite element space defined above into the CDG framework. For the domain under consideration $\Omega = [0,1]^d$, we let $\Omega_{N,P} := \Omega_N([0,1]^d)$ be the primal mesh and $\Omega_{N,D}$, which is the periodic extension of $\Omega_N([-h_N/2, 1-h_N/2]^d)$ restricted to $[0,1]^d$, be the dual mesh. Similarly, we let $\hat{\mathbf{V}}_{N,P}^k := \hat{\mathbf{V}}_N^k([0,1]^d)$ and $\hat{\mathbf{V}}_{N,D}^k$ to be the periodic extension of $\hat{\mathbf{V}}_N^k([-h_N/2, 1-h_N/2]^d)$ restricted to $[0,1]^d$. Here and below, the subscripts P and D represent the quantities defined on the primal and dual mesh, respectively.

The approximation properties for the sparse finite element space have been established in previous work [76, 38]. By using a lemma in [38], we can have estimates for L^2 projection operator onto the spaces $\hat{\mathbf{V}}^k_{N,P}, \hat{\mathbf{V}}^k_{N,D}$.

To facilitate the discussion, below we introduce some notations about norms and semi-norms. Let $G = P, D$, on primal or dual mesh $\Omega_{N,G}$, we use $\|\cdot\|_{H^s(\Omega_{N,G})}$ to denote the standard broken Sobolev norm, i.e. $\|v\|^2_{H^s(\Omega_{N,G})} = \sum_{\mathbf{0}\le\mathbf{j}\le 2^{\mathbf{N}}-\mathbf{1}} \|v\|^2_{H^s(I^{\mathbf{j}}_{N,G})}$, where $\|v\|_{H^s(I^{\mathbf{j}}_{N,G})}$ is the standard Sobolev norm on $I^{\mathbf{j}}_{N,G}$, (and $s = 0$ is used to denote the L^2 norm). Similarly, we use $|\cdot|_{H^s(\Omega_{N,G})}$ to denote the broken Sobolev semi-norm, and $\|\cdot\|_{H^s(\Omega_{\mathbf{l},G})}, |\cdot|_{H^s(\Omega_{\mathbf{l},G})}$ to denote the broken Sobolev norm and semi-norm that are supported on a general grid $\Omega_{\mathbf{l},G}$. For any set $L = \{i_1, \dots i_r\} \subset \{1, \dots d\}$, we define L^c to be the complement set of L in $\{1, \dots d\}$. For a non-negative integer α and set L, we define the semi-norm on any domain denoted by Ω'

$$|v|_{H^{\alpha,L}(\Omega')} := \left\| \left(\frac{\partial^\alpha}{\partial x^\alpha_{i_1}} \cdots \frac{\partial^\alpha}{\partial x^\alpha_{i_r}} \right) v \right\|_{L^2(\Omega')},$$

and

$$|v|_{\mathcal{H}^{q+1}(\Omega')} := \max_{1\le r\le d} \left(\max_{\substack{L\subset\{1,2,\cdots,d\}\\ |L|=r}} |v|_{H^{q+1,L}(\Omega')} \right),$$

which is the norm for the mixed derivative of v of at most degree $q+1$ in each direction. In this chapter, we use the notation $A \lesssim B$ to represent $A \le \text{constant} \times B$, where the constant is independent of N and the mesh level considered. The following results are obtained from Lemma 3.2 in [38].

Lemma 4.1.1 (L^2 projection estimate)**.** *Let $\mathbf{P}_P, \mathbf{P}_D$ be L^2 projections onto the spaces $\hat{\mathbf{V}}^k_{N,P}, \hat{\mathbf{V}}^k_{N,D}$, respectively, then for $k \ge 1$, $1 \le q \le \min\{p, k\}$, and $v \in \mathcal{H}^{p+1}(\Omega)$, which is* *periodic on* Ω, $N \geq 1$, $d \geq 2$, *we have for* $G = P, D$,

$$|\mathbf{P}_G v - v|_{H^s(\Omega_{N,G})} \lesssim \begin{cases} N^d 2^{-N(q+1)} |v|_{\mathcal{H}^{q+1}(\Omega)} & s = 0, \\ 2^{-Nq} |v|_{\mathcal{H}^{q+1}(\Omega)} & s = 1. \end{cases} \tag{4.3}$$

This lemma shows that the L^2 norm and H^1 semi-norm of the projection error scale like $O(N^d 2^{-N(k+1)})$ and $O(2^{-Nk})$ with respect to N when the function v has bounded mixed derivatives up to enough degrees. This lemma will be used in Theorem 4.2.2 to establish convergence of the scheme.

Now, we are ready to formulate the sparse grid CDG scheme. Below we review some standard notations about jumps and averages of piecewise functions. With $G = P$ or D, let $T_{h,G}$ be the collection of all elementary cell $I^{\mathbf{j}}_{N,G}$, $\Gamma_{N,G} := \bigcup_{T \in \Omega_{N,G}} \partial T$ be the union of the interfaces for all the elements in $\Omega_{N,G}$ (here we have taken into account the periodic boundary condition when defining $\Gamma_{N,G}$) and $S(\Gamma_G) := \Pi_{T \in \Omega_{N,G}} L^2(\partial T)$ be the set of L^2 functions defined on $\Gamma_{N,G}$. For any $q \in S(\Gamma_{N,G})$ and $\mathbf{q} \in [S(\Gamma_{N,G})]^d$, we define their averages $\{q\}, \{\mathbf{q}\}$ and jumps $[q], [\mathbf{q}]$ on the interior edges as follows. Suppose e is an interior edge shared by elements T_+ and T_-, either on primal or dual mesh, we define the unit normal vectors $\boldsymbol{n}^+$ and $\boldsymbol{n}^-$ on e pointing exterior of T_+ and T_-, respectively, then

$$\begin{aligned} [q] &= q^- \boldsymbol{n}^- + q^+ \boldsymbol{n}^+, \quad \{q\} = \frac{1}{2}(q^- + q^+), \\ [\mathbf{q}] &= \mathbf{q}^- \cdot \boldsymbol{n}^- + \mathbf{q}^+ \cdot \boldsymbol{n}^+, \quad \{\mathbf{q}\} = \frac{1}{2}(\mathbf{q}^- + \mathbf{q}^+). \end{aligned}$$

The semi-discrete sparse grid CDG scheme for (4.2), based on the weak formulation introduced in [48, 50], is defined as follows: we find $u_h^l \in \hat{\mathbf{V}}^k_{N,P}$ and $v_h^l \in \hat{\mathbf{V}}^k_{N,D}$, such that

$\forall\, l = 1, \cdots, m$

$$\int_\Omega (u_h^l)_t\, \varphi_h \, d\mathbf{x} = \frac{1}{\tau_{\max}} \int_\Omega (v_h^l - u_h^l)\, \varphi_h \, d\mathbf{x} + \int_\Omega A^l(t, \mathbf{x}) \mathbf{v}_h \cdot \nabla \varphi_h \, d\mathbf{x} - \sum_{e \in \Gamma_{N,P}} \int_e A^l(t, \mathbf{x}) \mathbf{v}_h \cdot [\varphi_h] \, ds, \tag{4.4}$$

$$\int_\Omega (v_h^l)_t\, \psi_h \, d\mathbf{x} = \frac{1}{\tau_{\max}} \int_\Omega (u_h^l - v_h^l)\, \psi_h \, d\mathbf{x} + \int_\Omega A^l(t, \mathbf{x}) \mathbf{u}_h \cdot \nabla \psi_h \, d\mathbf{x} - \sum_{e \in \Gamma_{N,D}} \int_e A^l(t, \mathbf{x}) \mathbf{u}_h \cdot [\psi_h] \, ds, \tag{4.5}$$

for any $\varphi_h \in \hat{\mathbf{V}}_{N,P}^k$ and $\psi_h \in \hat{\mathbf{V}}_{N,D}^k$, where $\mathbf{u}_h = (u_h^1, \cdots, u_h^m), \mathbf{v}_h = (v_h^1, \cdots, v_h^m)$ and $\tau_{\max}$ is an upper bound for the time step due to the CFL restriction (see Section 4.1.3 for detailed discussions).

4.1.2 Discussions on implementations

Here, we briefly discuss some details about the implementation of the scheme. We perform the computation by using orthonormal multiwavelet bases constructed by Alpert [4]. In 1D, the bases of $W_l^k([0,1])$ are denoted by

$$v_{p,l}^j(x), \quad p = 1, \cdots, k+1, \quad j = 0, \cdots, 2^{l-1} - 1$$

and they satisfy $\int_a^b v_{p,l}^j(x) v_{p',l'}^{j'}(x) dx = \delta_{pp'} \delta_{jj'} \delta_{ll'}$. Figures 4.1a and 4.2a provide illustrations of the basis functions for $k = 0, 1$ and $l = 0, 1, 2$. The bases in $W_{\mathbf{l}}^k$ in multi-dimensions are defined by tensor products

$$v_{\mathbf{s}} = v_{\mathbf{p},\mathbf{l}}^{\mathbf{j}} := \prod_{i=1}^d v_{p_i, l_i}^{j_i}(x_i), \quad p_i = 1, \cdots, k+1, \quad j_i = 0, \cdots, \max(0, 2^{l_i - 1} - 1),$$

where we have used the notation $\mathbf{s} = (\mathbf{l}, \mathbf{j}, \mathbf{p})$ and $s_i = (l_i, j_i, p_i)$ to denote the multi-index for the bases.

As for temporal schemes, we can use the total variation diminishing Runge-Kutta (TVD-RK) methods [73] to solve the ordinary differential equations for the coefficients resulting from the discretization. To calculate the right-hand-side of (4.4)-(4.5), the fast matrix-vector product by LU split or LU decomposition algorithms [69, 70, 64] can be applied, by which one can decompose all calculations into one dimensional operations. Below, we briefly describe the LU decomposition algorithm for the calculation of the following matrix-vector product which appears at the right-hand-side of (4.4)-(4.5)

$$b_{\mathbf{j}} = \sum_{\mathbf{s}:|\mathbf{l}|_1 \leq N} f_{\mathbf{s}} t^1_{s_1,j_1} \cdots t^d_{s_d,j_d},$$

where $f_{\mathbf{s}}$ can be the coefficient of the basis in sparse grid space and $t^i_{s_i,j_i}$, $i = 1, \cdots, d$, are the corresponding one-dimensional transform of coefficients from basis v_{s_i} to basis v_{j_i} in the i-th dimension in our scheme. Note that we have $n = 2^N(k+1)$ one-dimensional bases in each dimension, and we use v_{s_i} to denote the s_i-th basis. The bases are ordered according to grid increment. Using Algorithm 1 in [70], we should calculate all the one-dimensional transform along each direction associated with a block lower triangular matrix, and then calculate all the one-dimensional transforms having a block upper triangular structure. The fast matrix-vector product $f_{\mathbf{s}} \to b_{\mathbf{j}}$ on sparse grid with LU decomposition can be proceeded as follows.

1. Calculate (block) LU decomposition $t^i_{s,j} = \sum_{m=1}^{n} (Pl)^i_{s,m} (uQ)^i_{m,j}$, $s, j = 1, \cdots, n$, for $i = 1, \cdots, d$, where P^i, Q^i are the permutation matrices, l^i, u^i are lower and upper triangular matrices.

2. Compute the transform with a (block) lower triangular matrix for $i = 1, \cdots, d$,

$$b_{s_1,\cdots,s_{i-1},s_i',s_{i+1},\cdots,s_d} \leftarrow \sum_{s_i:l_1+\cdots+l_d\leq N} f_{\mathbf{s}}(Pl)^i_{s_i,s_i'}.$$

3. Compute the transform with a (block) upper triangular matrix for $i = 1, \cdots, d$,

$$b_{\mathbf{s}} \leftarrow \sum_{s_i':l_1+\cdots+l_{i-1}+l_i'+l_{i+1}+\cdots+l_d\leq N} b_{s_1,\cdots,s_{i-1},s_i',s_{i+1},\cdots,s_d} (uQ)^i_{s_i',s_i}.$$

Note that in step 1, the LU decomposition pivots only from rows or columns in the same mesh level to maintain the hierarchical structure. This pivoting can be successfully done in the sparse grid CDG scheme, but not in the sparse grid DG scheme, for which additional splitting of the flux terms are deemed necessary for variable coefficient case.

For the integrals involving variable-coefficient, we use Gaussian quadrature to compute these terms. Since these integrals are multi-dimensional integrations, we use the so-called unidirectional principle to separate the integration into multiplication of one-dimensional integrals. For example, if $\phi(x) = \phi_1(x_1)\cdots\phi_d(x_d)$ is separable,

$$\int_\Omega \phi(x) = \int_{[a,b]} \phi_1(x_1) \cdots \int_{[a,b]} \phi_d(x_d).$$

When the variable coefficient $A_i(t, x)$ is separable, we can use unidirectional principle directly. If it is not separable, we can find $A_i^h(t, x)$ as the L^2 projection of $A_i(t, x)$ onto the sparse grid finite element space, and then use $A_i^h(t, x)$ to compute the integrals.

4.1.3 Discussions on CFL conditions

It is well known that the CDG schemes allow larger CFL numbers than the standard DG methods except for piecewise constant approximations [50, 60]. Here, we perform a numerical study of the CFL conditions of DG [27], CDG [51], sparse grid DG [38], and the sparse

Table 4.1: CFL numbers of the DG method, CDG method, sparse grid DG method and sparse grid CDG method with piecewise degree k polynomials, Runge-Kutta method of order ν for Example 4.3.1 with d=2. The CFL numbers of the sparse grid DG/CDG methods are measured with regard to the most refined mesh h_N.

	DG			CDG			sparse grid DG			sparse grid CDG		
k	1	2	3	1	2	3	1	2	3	1	2	3
$\nu = 2$	0.33	–	–	0.48	–	–	0.66	–	–	0.87	–	–
$\nu = 3$	0.40	0.20	0.13	0.66	0.36	0.24	0.81	0.41	0.25	1.17	0.65	0.44
$\nu = 4$	0.46	0.23	0.14	0.90	0.52	0.35	0.92	0.46	0.28	1.58	0.94	0.62

grid CDG schemes. We only consider the two-dimensional case solving constant coefficient equation $u_t + u_{x_1} + u_{x_2} = 0$ for now. The results are listed in Table 4.1. The CFL number of DG method is obtained from Table 2.2 in [27]. The rest of the table is computed by eigenvalue analysis of the discretization matrix, and by requiring the amplification of the eigenvalues to be bounded by 1 in magnitude. We observe that the sparse grid DG method has CFL number that is about two times the CFL number of the standard DG method. The sparse grid CDG method offers the largest CFL conditions among all four methods. Here, as a side note, we find that the CFL number for two-dimensional CDG method is larger than the CFL number for one-dimensional CDG method in [51]. This table shows that one advantage of the sparse grid CDG method is the ability to take large time steps for time evolution problems. In general, further numerical results suggest that for equation $u_t + c_1 u_{x_1} + c_2 u_{x_2} = 0$, the CFL number for sparse grid DG and sparse grid CDG method will change with the value of the coefficients c_1, c_2. Results in higher dimensions are yet to be studied. A preliminary calculation shows that for equation $u_t + u_{x_1} + u_{x_2} + u_{x_3} = 0$ the CFL conditions for CDG, sparse grid DG and sparse grid CDG methods in 3D are all higher than those for the 2D case in Table 4.1. The sparse grid CDG method still possesses the largest CFL number among all four methods. Those interesting issues will be investigated in our future work.

4.1.4 Non-periodic problems

Here, we consider non-periodic problems, where equation (4.1) or (4.2) is supplemented by Dirichlet boundary condition on the inflow edges. In this case, we can no longer use periodicity to define the finite element space on the dual mesh, and a new grid hierarchy needs to be introduced.

Recall that for standard CDG methods with non-periodic boundary condition on the domain $[0,1]$, the finite element space on dual mesh with cell size $h_n = 1/2^n$ is represented by

$$V_{n,D}^k = \{v : v \in P^k(I_{n,D}^j), \forall\, j = 0, \ldots, 2^n\}, \tag{4.6}$$

where the mesh is partitioned as

$$I_{n,D}^0 = [0, \frac{h_n}{2}], \quad I_{n,D}^j = [(j-\frac{1}{2})h_n, (j+\frac{1}{2})h_n],\ j = 1, \ldots, 2^n - 1, \quad I_{n,D}^{2^n} = [1-\frac{h_n}{2}, 1],$$

which consists of $2^n - 1$ cells of size h_n, and two cells at the left and right ends of size $h_n/2$. It is easy to see that this space does not have nested structures, i.e. $V_{n-1,D}^k \not\subset V_{n,D}^k$. Therefore, we need a new hierarchy to define the increment polynomial spaces.

For a fixed refined mesh level N, we define the following grid $\Omega_{l,N,D}$ on level l, $l = 0 \ldots N$, by a collection of cells as

$$I_{l,N,D}^0 = [0, h_l - \frac{h_N}{2}], \quad I_{l,N,D}^{2^l} = [1 - \frac{h_N}{2}, 1],$$
$$I_{l,N,D}^j = [jh_l - \frac{h_N}{2}, (j+1)h_l - \frac{h_N}{2}],\ j = 1, \ldots, 2^l - 1,$$

which consists of $2^l - 1$ cells of size h_l, and a cell at the left end of size $h_l - \frac{h_N}{2}$, and a cell at the right end of size $\frac{h_N}{2}$. This grid structure is naturally nested, and therefore $V_{l,N,D}^k$

which consists of piecewise polynomials of degree k defined on $\Omega_{l,N,D}$ are also nested, and $V^k_{N,N,D} = V^k_{N,D}$ as defined in (4.6).

Then the definitions of sparse finite element space in Section 4.1.1 can be naturally extended here. We let $W^k_{l,N,D}$, $l = 1, \dots N$ be a complement set of $V^k_{l-1,N,D}$ in $V^k_{l,N,D}$, i.e.

$$V^k_{l-1,N,D} \oplus W^k_{l,N,D} = V^k_{l,N,D}.$$

However, we no longer require $W^k_{l,N,D}$ to be L^2 orthogonal to $V^k_{l-1,N,D}$, because such definition will be difficult to implement in practice. Instead, we define $W^k_{l,N,D}$ to be a span of basis functions that are shifted basis functions of W^k_l space defined in Section 4.1.1, namely,

$$W^k_{l,N,D} = W^k_l([-\frac{h_N}{2}, 1 - \frac{h_N}{2}])\big|_{[0,1]}, \quad l \geq 1.$$

By denoting $W^k_{0,N,D} = V^k_{0,N,D}$, we have decomposed $V^k_{N,D} = \bigoplus_{0 \leq l \leq N} W^k_{l,N,D}$. Illustration of basis functions by such definitions for $k = 0, 1$ and $l = 0, 1, 2$ can be found in Figures 4.1b and 4.2b. The dimension of $W^k_{0,N,D}$ is $2(k+1)$, while the dimensions of $W^k_{l,N,D}$, $l = 1, \dots N$ are $2^{l-1}(k+1)$.

Finally, the sparse finite element space on the dual mesh of domain $[0,1]^d$ is defined as

$$\hat{\hat{\mathbf{V}}}^k_{N,D} := \bigoplus_{\substack{|\mathbf{l}|_1 \leq N \\ \mathbf{l} \in \mathbb{N}_0^d}} \mathbf{W}^k_{\mathbf{l},N,D},$$

where $\mathbf{W}^k_{\mathbf{l},N,D} = W^k_{l_1,N,D,x_1} \times \cdots \times W^k_{l_d,N,D,x_d}$. This is a subset of the full grid space $\mathbf{V}^k_{N,D} = \bigoplus_{\substack{|\mathbf{l}|_\infty \leq N \\ \mathbf{l} \in \mathbb{N}_0^d}} \mathbf{W}^k_{\mathbf{l},N,D}$, and its number of degrees of freedom scales as $O(2^{d-1}(k+1)^d 2^N N^{d-1})$ (the proof is similar to Lemma 2.3 in [76]), which is larger than that of $\hat{\mathbf{V}}^k_{N,P}$, but still significantly less than that of $\mathbf{V}_{N,D}^k$ with exponential dependence on Nd.

We will now investigate the approximation property of the space $\hat{\hat{\mathbf{V}}}_{N,D}^k$. We can obtain the following result, which essentially states that the L^2 projection onto this newly constructed space has the same order of accuracy as $\mathbf{P}_P, \mathbf{P}_D$ in Lemma 4.1.1.

Lemma 4.1.2 (L^2 projection estimate onto $\hat{\hat{\mathbf{V}}}_{N,D}^k$)**.** *Let $\tilde{\mathbf{P}}_D$ be the L^2 projection onto the space $\hat{\hat{\mathbf{V}}}_{N,D}^k$, then for $k \geq 1$, $1 \leq q \leq \min\{p, k\}$, and $v \in \mathcal{H}^{p+1}(\Omega)$, $N \geq 1$, $d \geq 2$, we have*

$$|\tilde{\mathbf{P}}_D v - v|_{H^s(\Omega_{N,D})} \lesssim \begin{cases} N^d 2^{-N(q+1)} |v|_{\mathcal{H}^{q+1}(\Omega)} & s = 0, \\ 2^{-Nq} |v|_{\mathcal{H}^{q+1}(\Omega)} & s = 1. \end{cases} \tag{4.7}$$

Proof. The proof follows same procedure as Appendix A in [38]. We will mainly highlight the difference in the proof (see Steps 1 and 2 below). The main difference lies in the fact that all the hierarchical spaces (and associated projections) have dependence not only on l, but also on the finest mesh level N.

Step 1: Decomposition of $\tilde{\mathbf{P}}_D$ into tensor products of one-dimensional increment projections. We denote $P_{l,N,D}^k$ as the standard L^2 projection operator from $L^2([0,1])$ to $V_{l,N,D}^k$, and the induced increment projection

$$Q_{l,N,D}^k := \begin{cases} P_{l,N,D}^k - P_{l-1,N,D}^k, & \text{if } l = 1, \dots N, \\ P_{0,N,D}^k, & \text{if } l = 0, \end{cases}$$

and further denote

$$\tilde{\mathbf{P}}^k_{N,D} := \sum_{\substack{|\mathbf{l}|_1 \le N \\ \mathbf{l} \in \mathbb{N}_0^d}} Q^k_{l_1,N,D,x_1} \otimes \cdots \otimes Q^k_{l_d,N,D,x_d},$$

where the last subindex of $Q^k_{l_i,N,D,x_i}$ indicates that the increment operator is defined in x_i-direction. We can verify that $\tilde{\mathbf{P}}_D = \tilde{\mathbf{P}}^k_{N,D}$. In fact, for any v, it's clear that $\tilde{\mathbf{P}}^k_{N,D} v \in \hat{\hat{\mathbf{V}}}^k_{N,D}$. Therefore, we only need

$$\int_\Omega (\tilde{\mathbf{P}}^k_{N,D} v - v) w \, d\mathbf{x} = 0, \qquad \forall\, w \in \hat{\hat{\mathbf{V}}}^k_{N,D}. \tag{4.8}$$

It suffices to show (4.8) for $v \in C^\infty(\Omega)$ which is a dense subset of $L^2(\Omega)$. In fact, we have

$$v = \mathbf{P}^k_{N,D} v + v - \mathbf{P}^k_{N,D} v,$$

where $\mathbf{P}^k_{N,D} = P^k_{N,N,D,x_1} \otimes \cdots \otimes P_{N,N,D,x_d}$ is the L^2 projection onto the full grid space $\mathbf{V}^k_{N,D}$. Therefore,

$$\begin{aligned}
\int_\Omega (\tilde{\mathbf{P}}^k_{N,D} v - v) w d\mathbf{x} &= \int_\Omega (\tilde{\mathbf{P}}^k_{N,D} v - \mathbf{P}^k_{N,D} v) w d\mathbf{x} + \int_\Omega (v - \mathbf{P}^k_{N,D} v) w d\mathbf{x} \\
&= - \int_\Omega \Big(\sum_{\substack{|\mathbf{l}|_\infty \le N, |\mathbf{l}|_1 > N \\ \mathbf{l} \in \mathbb{N}_0^d}} Q^k_{l_1,N,D,x_1} \otimes \cdots \otimes Q^k_{l_d,N,D,x_d} v \Big)\, w \, d\mathbf{x}.
\end{aligned}$$

The last term in the first row of the equality above vanishes because $w \in \hat{\hat{\mathbf{V}}}^k_{N,D} \subset \mathbf{V}^k_{N,D}$.

In addition, for any $l \geq 1$, $\phi \in L^2([0,1]), \varphi \in V^k_{l-1,N,D}$

$$\int_{[0,1]} Q^k_{l,N,D}\phi\,\varphi dx = \int_{[0,1]} (I - P^k_{l-1,N,D})\phi\,\varphi dx - \int_{[0,1]} (I - P^k_{l,N,D})\phi\,\varphi dx = 0,$$

Therefore, by properties of the tensor product projections

$$\int_\Omega (\tilde{\mathbf{P}}^k_{N,D} v - v) w d\mathbf{x} = 0, \quad \forall w \in \hat{\tilde{\mathbf{V}}}^k_{N,D},$$

and the proof for $\tilde{\mathbf{P}}_D = \tilde{\mathbf{P}}^k_{N,D}$ is complete.

Step 2: Estimation of the increment projections. For a function $v \in H^{p+1}([0,1])$, we have the convergence property of the L^2 projection $P^k_{l,N,D}$ as follows: for any integer q with $1 \leq q \leq \min\{p, k\}$, $s = 0,\ 1$,

$$|P^k_{l,N,D} v - v|_{H^s(I^j_{l,N,D})} \leq c_{k,s,q} (h^j_{l,N})^{(q+1-s)} |v|_{H^{q+1}(I^j_{l,N,D})}, \quad j = 1, \cdots, 2^l - 1,$$

where the mesh size $h^j_{l,N} = \begin{cases} h_l - h_N/2, & j = 0 \\ h_l, & j = 1, \cdots, 2^l - 1, \\ h_N/2, & j = 2^l. \end{cases}$

The estimation above directly applies for $Q^k_{0,N,D} = P^k_{0,N,D}$. For $l \geq 1$, by simple algebra, we have

$$
\begin{aligned}
|Q^k_{l,N,D} v|_{H^s(I^j_{l,N,D})} &\le \tilde{c}_{k,s,q} 2^{-l(q+1-s)} |v|_{H^{q+1}(I^{\lfloor j/2 \rfloor}_{l-1,N,D})}, \quad j = 2, \cdots, 2^l - 1, \\
|Q^k_{l,N,D} v|_{H^s(I^j_{l,N,D})} &\le c_{k,s,q} (h_l)^{(q+1-s)} |v|_{H^{q+1}(I^j_{l,N,D})} \\
&\quad + c_{k,s,q} (h_{l-1} - h_N/2)^{(q+1-s)} |v|_{H^{q+1}(I^0_{l-1,N,D})}, \quad j = 0, 1, \\
&< \tilde{c}_{k,s,q} 2^{-l(q+1-s)} |v|_{H^{q+1}(I^0_{l-1,N,D})}, \\
|Q^k_{l,N,D} v|_{H^s(I^{2^l}_{l,N,D})} &= 0,
\end{aligned}
$$

with $\tilde{c}_{k,s,q} = c_{k,s,q}(1 + 2^{q+1-s})$.

The rest of the proof is then very similar to Appendix A in [38], and is omitted. □

We now provide a numerical validation of Lemma 4.1.2 by considering the error of projection $\tilde{\mathbf{P}}_D$ for a smooth function

$$
u(\mathbf{x}) = \exp\left(\prod_{i=1}^{d} x_i\right), \quad \mathbf{x} \in [0,1]^d. \tag{4.9}
$$

In Table 4.2, we report the L^2 errors and the associated orders of accuracy for $k = 1, 2, 3$, $d = 2, 3$. It is clear that the predicted order of accuracy is achieved.

With the aid of this space, the semi-discrete scheme can now be defined similarly as in (4.4)-(4.5) by using the space on the dual mesh as $\hat{\tilde{\mathbf{V}}}^k_{N,D}$, and replacing the numerical values on the boundary of the domain by corresponding functions in the Dirichlet boundary conditions.

We now comment on the implementation of this algorithm. As can be seen from Figures

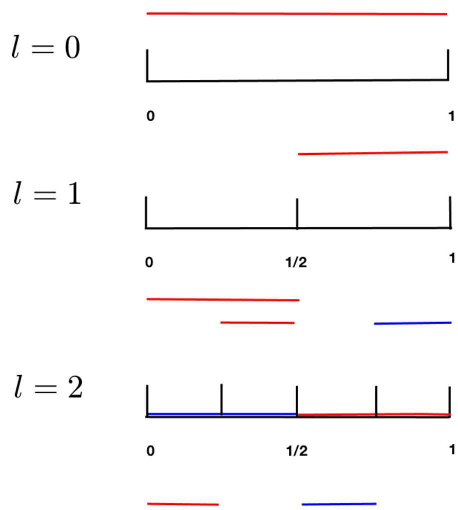
Table 4.2: L^2 errors and orders of accuracy for L^2 projection operator $\tilde{\mathbf{P}}_D$ of (4.9) onto $\hat{\tilde{\mathbf{V}}}_{N,D}^k$ when $d = 2$ and $d = 3$. N is the number of mesh levels, k is the polynomial order, d is the dimension. L^2 order is calculated with respect to h_N.

		L^2 error	order	L^2 error	order	L^2 error	order
N	h_N	$k=1$		$k=2$		$k=3$	
				$d=2$			
3	1/8	8.93E-04	–	9.14E-06	–	6.40E-08	–
4	1/16	2.61E-04	1.77	1.29E-06	2.82	4.45E-09	3.85
5	1/32	7.34E-05	1.83	1.77E-07	2.87	3.01E-10	3.89
6	1/64	2.00E-05	1.88	2.37E-08	2.90	1.98E-11	3.93
7	1/128	5.35E-06	1.90	3.11E-09	2.93	1.29E-12	3.94
				$d=3$			
3	1/8	6.19E-04	–	4.93E-06	–	3.18E-08	–
4	1/16	1.90E-04	1.70	7.45E-07	2.73	2.36E-09	3.75
5	1/32	5.71E-05	1.73	1.10E-07	2.76	1.69E-10	3.80
6	1/64	1.67E-05	1.77	1.58E-08	2.80	1.18E-11	3.84
7	1/128	4.80E-06	1.80	2.24E-09	2.82	9.35E-13	3.66

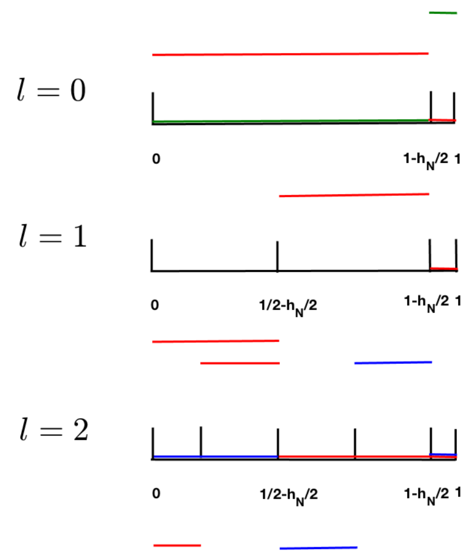
4.1b and 4.2b, there are two types of basis functions in 1D for the dual space.

- Type 1 bases (for $l \geq 0$), which are the shifted and truncated multiwavelet bases.
- Type 2 bases (for $l = 0$), which are the Legendre polynomials of degree up to k on $[1 - \frac{h_N}{2}, 1]$.

Clearly, Type 1 bases are orthogonal to Type 2 bases, because their support do not overlap. Type 2 bases are orthogonal to each other due to the definition of Legendre polynomials. However, Type 1 bases are no longer orthogonal to each other, due to the domain shift and truncation. However, only the left-most element on each level are changed. For other bases in that level, they will still retain orthogonality. The bases on left-most element in all level are orthogonal to other bases, but not to each other, i.e., the bases defined on left-most element in different levels are not orthogonal. This implies that although the mass matrix is not identity here, it will have block structures and be sparse.



(a) Primal mesh. Number of bases for $l = 0, 1, 2$ are $1, 1, 2$.



(b) Dual mesh. Number of bases for $l = 0, 1, 2$ are $2, 1, 2$.

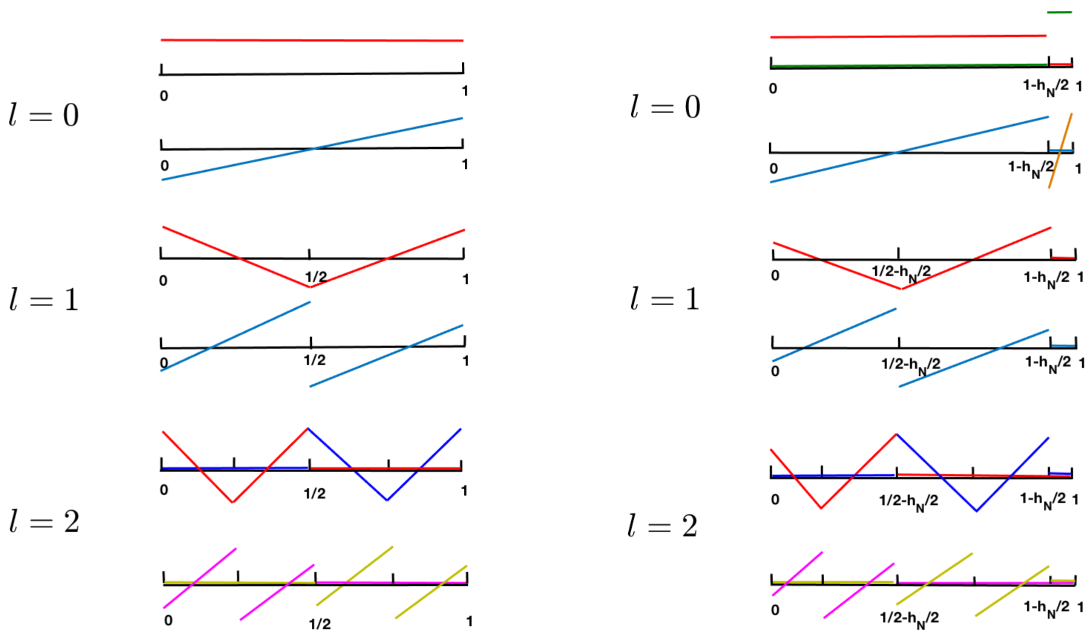
Figure 4.1: Illustration of one-dimensional bases on different levels for $k = 0$: non-periodic problems. Different colors represent different bases.

4.2 Stability and convergence

In this section, we prove L^2 stability and error estimates for the sparse grid CDG scheme for the scalar equation. We consider both periodic and non-periodic boundary conditions. For periodic problems, (4.2) reduces to

$$\frac{\partial u}{\partial t} + \nabla \cdot (\mathbf{A} u) = 0, \quad \mathbf{x} \in \Omega, \tag{4.10}$$

where $\mathbf{A} = (A_1(t, \mathbf{x}), \cdots, A_d(t, \mathbf{x}))$, and $\|\mathbf{A}\|_{L^\infty(\Omega)} < \infty, \|\nabla \cdot \mathbf{A}\|_{L^\infty(\Omega)} < \infty$. We assume $A_i \neq 0$ to avoid the discussion of different boundary conditions for degenerating coefficients. However, there is no difficulty to extend the proof below to degenerating case. For non-



(a) Primal mesh. Number of bases for $l = 0, 1, 2$ are $2, 2, 4$.

(b) Dual mesh. Number of bases for $l = 0, 1, 2$ are $4, 2, 4$.

Figure 4.2: Illustration of one-dimensional bases on different levels for $k = 1$: non-periodic problems. Different colors represent different bases.

periodic problems, the following inflow boundary conditions are prescribed,

$$u(t, \mathbf{x})|_{\partial\Omega_{x_i^{in}}} = g_i(t, \cdots, x_{i-1}, x_{i+1}, \cdots, x_d)$$

where

$$\partial\Omega_{x_i^{in}} := \begin{cases} \{\mathbf{x} \in \Omega | x_i = 0\}, & \text{if } A_i(t, \mathbf{x}) > 0, \\ \{\mathbf{x} \in \Omega | x_i = 1\}, & \text{if } A_i(t, \mathbf{x}) < 0. \end{cases}$$

Correspondingly, we denote the outflow edges by

$$\partial\Omega_{x_i^{out}} := \begin{cases} \{\mathbf{x} \in \Omega | x_i = 1\}, & \text{if } A_i(t, \mathbf{x}) > 0, \\ \{\mathbf{x} \in \Omega | x_i = 0\}, & \text{if } A_i(t, \mathbf{x}) < 0. \end{cases}$$

The scheme for periodic case reduces to: to find $u_h \in \hat{\mathbf{V}}^k_{N,P}$ and $v_h \in \hat{\mathbf{V}}^k_{N,D}$, such that

$$\int_\Omega (u_h)_t \, \varphi_h \, d\mathbf{x} = \frac{1}{\tau_{\max}} \int_\Omega (v_h - u_h) \, \varphi_h \, d\mathbf{x} + \int_\Omega v_h \mathbf{A} \cdot \nabla \varphi_h \, d\mathbf{x} - \sum_{e \in \Gamma_{N,P}} \int_e v_h \mathbf{A} \cdot [\varphi_h] \, ds, \tag{4.11}$$

$$\int_\Omega (v_h)_t \, \psi_h \, d\mathbf{x} = \frac{1}{\tau_{\max}} \int_\Omega (u_h - v_h) \, \psi_h \, d\mathbf{x} + \int_\Omega u_h \mathbf{A} \cdot \nabla \psi_h \, d\mathbf{x} - \sum_{e \in \Gamma_{N,D}} \int_e u_h \mathbf{A} \cdot [\psi_h] \, ds, \tag{4.12}$$

for any $\varphi_h \in \hat{\mathbf{V}}^k_{N,P}$ and $\psi_h \in \hat{\mathbf{V}}^k_{N,D}$. For non-periodic problems, we require $v_h, \psi_h \in \hat{\hat{\mathbf{V}}}^k_{N,D}$, and enforce $u_h|_{\partial\Omega_{x_i^{in}}} = v_h|_{\partial\Omega_{x_i^{in}}} = g_i$ on the boundary interface.

We can prove that the schemes retain similar stability properties as the standard CDG schemes.

Theorem 4.2.1 (L^2 Stability)**.** *With periodic boundary condition, the numerical solutions u_h and v_h of the sparse grid CDG scheme* (4.11)-(4.12) *for the equation* (4.10) *satisfy the following L^2 stability condition*

$$\|u_h\|^2_{L^2(\Omega_{N,P})} + \|v_h\|^2_{L^2(\Omega_{N,D})} \lesssim \|u_h(0,\mathbf{x})\|^2_{L^2(\Omega_{N,P})} + \|v_h(0,\mathbf{x})\|^2_{L^2(\Omega_{N,D})}. \tag{4.13}$$

For non-periodic boundary condition, the corresponding numerical solutions satisfy

$$\begin{aligned} \|u_h\|^2_{L^2(\Omega_{N,P})} + \|v_h\|^2_{L^2(\Omega_{N,D})} \lesssim{}& \|u_h(0,\mathbf{x})\|^2_{L^2(\Omega_{N,P})} + \|v_h(0,\mathbf{x})\|^2_{L^2(\Omega_{N,D})} \\ &+ \int_0^T \sum_{i=1}^d \int_{\partial\Omega_{x_i^{in}}} |A_i| g_i^2 \, ds \, dt \quad \textit{if } \tau_{\max} \lesssim \frac{h_N}{\|\mathbf{A}\|_1}. \end{aligned} \tag{4.14}$$

Proof. For periodic boundary condition, let $\varphi_h = u_h$ in (4.11) and $\psi_h = v_h$ in (4.12), summing the two equalities up, we have

$$\begin{aligned}
&\frac{1}{2}\frac{d}{dt}\int_\Omega((u_h)^2+(v_h)^2)d\mathbf{x}\\
=&\frac{1}{\tau_{\max}}\int_\Omega v_h\,u_h-u_h\,u_h+u_hv_h-v_hv_h\,d\mathbf{x}+\int_\Omega v_h\mathbf{A}\cdot\nabla u_h\,d\mathbf{x}-\sum_{e\in\Gamma_{N,P}}\int_e v_h\mathbf{A}\cdot[u_h]\,ds\\
&+\int_\Omega u_h\mathbf{A}\cdot\nabla v_h\,d\mathbf{x}-\sum_{e\in\Gamma_{N,D}}\int_e u_h\mathbf{A}\cdot[v_h]\,ds\\
=&-\frac{1}{\tau_{\max}}\int_\Omega(u_h-v_h)^2d\mathbf{x}+\int_\Omega\mathbf{A}\cdot\nabla(u_hv_h)d\mathbf{x}-\sum_{e\in\Gamma_{N,P}}\int_e v_h\mathbf{A}\cdot[u_h]\,ds\\
&-\sum_{e\in\Gamma_{N,D}}\int_e u_h\mathbf{A}\cdot[v_h]\,ds.
\end{aligned}$$

Apply divergence theorem, and by periodicity, we have

$$\int_\Omega\mathbf{A}\cdot\nabla(u_hv_h)d\mathbf{x}-\sum_{e\in\Gamma_{N,P}}\int_e\mathbf{A}v_h\cdot[u_h]\,ds-\sum_{e\in\Gamma_{N,D}}\int_e\mathbf{A}u_h\cdot[v_h]\,ds=-\int_\Omega\nabla\cdot\mathbf{A}u_hv_hd\mathbf{x}.$$

By the simple inequality $ab\leq\frac{1}{2}(a^2+b^2)$,

$$\frac{1}{2}\frac{d}{dt}\int_\Omega\big((u_h)^2+(v_h)^2\big)d\mathbf{x}\leq-\frac{1}{\tau_{\max}}\int_\Omega(u_h-v_h)^2d\mathbf{x}+\frac{1}{2}\|\nabla\cdot\mathbf{A}\|_{L^\infty(\Omega)}\int_\Omega((u_h)^2+(v_h)^2)d\mathbf{x}.$$

and the proof for the periodic case is complete by using Gronwall's inequality.

For non-periodic boundary condition, we follow the same lines and plug in the corresponding boundary condition,

$$\begin{aligned}
&\frac{1}{2}\frac{d}{dt}\int_\Omega((u_h)^2+(v_h)^2)d\mathbf{x}\\
=&-\frac{1}{\tau_{\max}}\int_\Omega(u_h-v_h)^2d\mathbf{x}-\int_\Omega\nabla\cdot\mathbf{A}u_hv_hd\mathbf{x}+\int_{\partial\Omega}\mathbf{A}\cdot\mathbf{n}u_hv_hds
\end{aligned}$$

$$
\begin{aligned}
&- \sum_{i=1}^{d} \left(\int_{\partial\Omega_{x_i^{in}}} \mathbf{A}\cdot\mathbf{n} g_i (u_h + v_h) ds + 2 \int_{\partial\Omega_{x_i^{out}}} \mathbf{A}\cdot\mathbf{n} u_h v_h ds \right) \\
= &- \frac{1}{\tau_{\max}} \int_\Omega (u_h - v_h)^2 d\mathbf{x} - \int_\Omega \nabla\cdot\mathbf{A} u_h v_h d\mathbf{x} \\
&+ \sum_{i=1}^{d} \left(\int_{\partial\Omega_{x_i^{in}}} |\mathbf{A}\cdot\mathbf{n}| (-u_h v_h + g_i(u_h + v_h)) ds - \int_{\partial\Omega_{x_i^{out}}} |\mathbf{A}\cdot\mathbf{n}| u_h v_h ds \right) \\
\le &- \frac{1}{\tau_{\max}} \int_\Omega (u_h - v_h)^2 d\mathbf{x} + \frac{1}{2} \|\nabla\cdot\mathbf{A}\|_{L^\infty(\Omega)} \int_\Omega ((u_h)^2 + (v_h)^2) d\mathbf{x} \\
&+ \sum_{i=1}^{d} \left(\int_{\partial\Omega_{x_i^{in}}} |\mathbf{A}\cdot\mathbf{n}| (g_i^2 + \frac{1}{2}(u_h - v_h)^2) ds + \frac{1}{2} \int_{\partial\Omega_{x_i^{out}}} |\mathbf{A}\cdot\mathbf{n}| (u_h - v_h)^2 ds \right) \\
\le &- \frac{1}{\tau_{\max}} \int_\Omega (u_h - v_h)^2 d\mathbf{x} + \frac{1}{2} \|\nabla\cdot\mathbf{A}\|_{L^\infty(\Omega)} \int_\Omega ((u_h)^2 + (v_h)^2) d\mathbf{x} \\
&+ \sum_{i=1}^{d} \left(\int_{\partial\Omega_{x_i^{in}}} |\mathbf{A}\cdot\mathbf{n}| g_i^2 ds + \int_{\partial\Omega_{x_i^{in}} \cup \partial\Omega_{x_i^{out}}} |\mathbf{A}\cdot\mathbf{n}| \frac{1}{2} (u_h - v_h)^2 ds \right) \\
= &- \frac{1}{\tau_{\max}} \int_\Omega (u_h - v_h)^2 d\mathbf{x} + \frac{1}{2} \|\nabla\cdot\mathbf{A}\|_{L^\infty(\Omega)} \int_\Omega ((u_h)^2 + (v_h)^2) d\mathbf{x} \\
&+ \sum_{i=1}^{d} \left(\int_{\partial\Omega_{x_i^{in}}} |A_i| g_i^2 ds + \int_{\partial\Omega_{x_i^{in}} \cup \partial\Omega_{x_i^{out}}} |A_i| \frac{1}{2} (u_h - v_h)^2 ds \right).
\end{aligned}
$$

by noticing $\mathbf{A}\cdot\mathbf{n}|_{\partial\Omega_{x_i^{in}}} < 0$ and $\mathbf{A}\cdot\mathbf{n}|_{\partial\Omega_{x_i^{out}}} > 0$.

Let $T_{N,D}^i := \{T \in \Omega_{N,D} | T \cap \partial\Omega_{x_i} \neq \emptyset\}$ denote the cells on dual mesh adjacent to the boundary in the i-th direction. By inverse inequality, we have $\|u_h - v_h\|^2_{L^2(\partial\Omega_{x_i})} \lesssim h_N^{-1} \|u_h - v_h\|^2_{L^2(T_{N,D}^i)} \le h_N^{-1} \|u_h - v_h\|^2_{L^2(\Omega)}$. Therefore, if $\tau_{\max} \lesssim \frac{h_N}{\|\mathbf{A}\|_1}$,

$$
\frac{1}{2}\frac{d}{dt} \int_\Omega ((u_h)^2 + (v_h)^2) d\mathbf{x} \le \frac{1}{2} \|\nabla\cdot\mathbf{A}\|_{L^\infty(\Omega)} \int_\Omega ((u_h)^2 + (v_h)^2) d\mathbf{x} + \sum_{i=1}^{d} \int_{\partial\Omega_{x_i^{in}}} |A_i| g_i^2 ds,
$$

and the proof for the non-periodic case is complete by using Gronwall's inequality. $\square$

Now we are ready to prove L^2 error estimate of the sparse grid CDG scheme.

Theorem 4.2.2 (L^2 error estimate)**.** *Let u be the exact solution to (4.10) and u_h, v_h be the numerical solution to the semidiscrete scheme (4.11) and (4.12) with initial discretization $u_h(0, \mathbf{x}) = \mathbf{P}_P u_0, v_h(0, \mathbf{x}) = \mathbf{P}_D u_0$ for periodic boundary condition or $u_h(0, \mathbf{x}) = \mathbf{P}_P u_0, v_h(0, \mathbf{x}) = \tilde{\mathbf{P}}_D u_0$ for non-periodic boundary condition. If $\tau_{\max} \lesssim h_N$, then for $k \geq 1$, $u_0 \in \mathcal{H}^{p+1}(\Omega), 1 \leq q \leq \min\{p, k\}, N \geq 1, d \geq 2$, we have for all $t \geq 0$*

$$\|u - u_h\|_{L^2(\Omega_{N,P})} + \|u - v_h\|_{L^2(\Omega_{N,D})} \lesssim N^d 2^{-Nq} |u|_{\mathcal{H}^{q+1}(\Omega)}. \tag{4.15}$$

Proof. For periodic problems, we first introduce the standard notation of bilinear form

$$\begin{aligned}
B(u_h, v_h; \varphi_h, \psi_h) = &\int_\Omega (u_h)_t \, \varphi_h \, d\mathbf{x} - \frac{1}{\tau_{\max}} \int_\Omega (v_h - u_h) \, \varphi_h \, d\mathbf{x} - \int_\Omega v_h \mathbf{A} \cdot \nabla \varphi_h \, d\mathbf{x} \\
&+ \sum_{e \in \Gamma_P} \int_e v_h \mathbf{A} \cdot [\varphi_h] \, ds + \int_\Omega (v_h)_t \, \psi_h \, d\mathbf{x} - \frac{1}{\tau_{\max}} \int_\Omega (u_h - v_h) \, \psi_h \, d\mathbf{x} \\
&- \int_\Omega u_h \mathbf{A} \cdot \nabla \psi_h \, d\mathbf{x} + \sum_{e \in \Gamma_D} \int_e u_h \mathbf{A} \cdot [\psi_h] \, ds.
\end{aligned}$$

By Galerkin orthogonality, we have the error equation

$$B(u - u_h, u - v_h; \varphi_h, \psi_h) = 0, \quad \forall \varphi_h \in \hat{\mathbf{V}}_{N,P}^k, \psi_h \in \hat{\mathbf{V}}_{N,D}^k. \tag{4.16}$$

We take

$$\begin{aligned}
\varphi_h &= \mathbf{P}_P u - u_h, \quad \psi_h = \mathbf{P}_D u - u_h, \\
\varphi^e &= \mathbf{P}_P u - u, \qquad \psi^e = \mathbf{P}_D u - u,
\end{aligned}$$

then the error equation (4.16) becomes

$$B(\varphi_h, \psi_h; \varphi_h, \psi_h) = B(\varphi^e, \psi^e; \varphi_h, \psi_h). \tag{4.17}$$

From Theorem 4.2.1, we get

$$\frac{1}{2}\frac{d}{dt}\int_\Omega (\varphi_h^2 + \psi_h^2) d\mathbf{x} \le B(\varphi^e, \psi^e; \varphi_h, \psi_h) + \frac{1}{2}\|\nabla \cdot \mathbf{A}\|_{L^\infty(\Omega)} \int_\Omega (\varphi_h^2 + \psi_h^2) d\mathbf{x}. \tag{4.18}$$

We write the bilinear form on the right-hand side as a sum of three terms

$$B(\varphi^e, \psi^e; \varphi_h, \psi_h) = B^1 + B^2 + B^3, \tag{4.19}$$

where

$$\begin{aligned}
B^1 &= \int_\Omega (\varphi^e)_t\, \varphi_h \, d\mathbf{x} - \frac{1}{\tau_{\max}} \int_\Omega (\psi^e - \varphi^e)\, \varphi_h \, d\mathbf{x} + \int_\Omega (\psi^e)_t\, \psi_h \, d\mathbf{x} - \frac{1}{\tau_{\max}} \int_\Omega (\varphi^e - \psi^e)\, \psi_h \, d\mathbf{x}, \\
B^2 &= -\int_\Omega \psi^e \mathbf{A} \cdot \nabla \varphi_h \, d\mathbf{x} - \int_\Omega \varphi^e \mathbf{A} \cdot \nabla \psi_h \, d\mathbf{x}, \\
B^3 &= \sum_{e \in \Gamma_{N,P}} \int_e \psi^e \mathbf{A} \cdot [\varphi_h] \, ds + \sum_{e \in \Gamma_{N,D}} \int_e \varphi^e \mathbf{A} \cdot [\psi_h] \, ds.
\end{aligned}$$

By Cauchy-Schwartz inequality, Lemma 4.1.1 and $\tau_{\max} \lesssim h_N$, we have

$$B^1 \lesssim \int_\Omega (\varphi_h^2 + \psi_h^2) d\mathbf{x} + N^{2d} 2^{-2Nq} |u|^2_{\mathcal{H}^{q+1}(\Omega)}. \tag{4.20}$$

To estimate B^2, B^3, we use the following inverse inequalities $\forall w_h \in \hat{\mathbf{V}}^k_{N,G}$, for $G = P, D$,

$$|w_h|_{\mathcal{H}^1(\Omega_{N,G})} \lesssim h_N^{-1} \|w_h\|_{L^2(\Omega_{N,G})}, \quad \|w_h\|_{\Gamma_{N,G}} \lesssim h_N^{-\frac{1}{2}} \|w_h\|_{L^2(\Omega_{N,G})}$$

and trace inequality,

$$\|\phi\|^2_{L^2(\partial T)} \lesssim {h_N}^{-1}\|\phi\|^2_{L^2(T)} + h_N|\phi|_{H^1(T)}, \quad \forall \phi \in H^1(T), T \in \Omega_{N,G}.$$

Then we have

$$B^2 \lesssim \int_\Omega (\varphi_h^2 + \psi_h^2) d\mathbf{x} + N^{2d} 2^{-2Nq} \, |u|^2_{\mathcal{H}^{q+1}(\Omega)} \tag{4.21}$$

and

$$B^3 \lesssim \int_\Omega (\varphi_h^2 + \psi_h^2) d\mathbf{x} + N^{2d} 2^{-2Nq} \, |u|^2_{\mathcal{H}^{q+1}(\Omega)} \,. \tag{4.22}$$

Combining (4.20), (4.21), (4.22) with (4.18), we obtain

$$\frac{d}{dt}\int_\Omega (\varphi_h^2 + \psi_h^2) d\mathbf{x} \lesssim \int_\Omega (\varphi_h^2 + \psi_h^2) d\mathbf{x} + N^{2d} 2^{-2Nq} \, |u|^2_{\mathcal{H}^{q+1}(\Omega)} \,.$$

Together with the estimates for initial discretization and by Gronwall's inequality, the proof is complete. For non-periodic problems, the argument is very similar as long as the stability result holds. The proof is omitted for brevity. □

This theorem proves L^2 error of the scheme is $O(N^d 2^{-Nk})$ or $O(|\log h_N|^d \, h_N^k)$ when the exact solution has enough smoothness in the mixed derivative norms.

4.3 Numerical results

In this section, we present several numerical tests to validate the performance of the proposed sparse grid CDG schemes. Unless otherwise stated, we use the third-order TVD-RK temporal discretization [73] and choose the time step $\Delta t = \frac{c}{\sum_{i=1}^{d} \frac{c_i}{h_N}}$, with $c = 0.1$ for $k = 1$, 2, where

c_i is the maximum wave propagation speed in x_i-direction. To guarantee that the spatial error dominates for $k = 3$, we take $\Delta t = O(h_N^{4/3})$. $\tau_{\max}$ is taken as $\frac{1}{2k+1} h_N$ which is always smaller than the maximum time step allowed based on the CFL number in Table 4.1. For periodic problems, we only provide L^2 errors on the primal mesh, because the results on the dual mesh are similar. For non-periodic problems, the L^2 errors are the L^2 average of the errors on the primal and dual meshes.

4.3.1 Scalar case

In this subsection, we consider the scalar case, i.e. $m = 1$.

Example 4.3.1 (Linear advection with constant coefficients)**.** *We consider*

$$\begin{cases} u_t + \displaystyle\sum_{i=1}^{d} u_{x_i} = 0, \quad \mathbf{x} \in [0,1]^d, \\ u(0,\mathbf{x}) = \sin\left(2\pi \displaystyle\sum_{i=1}^{d} x_i\right), \end{cases} \tag{4.23}$$

with periodic or Dirichlet boundary conditions on the inflow edges corresponding to the given exact solution.

The exact solution is a smooth function,

$$u(t,\mathbf{x}) = \sin\left(2\pi\left(\sum_{i=1}^{d} x_i - d\,t\right)\right).$$

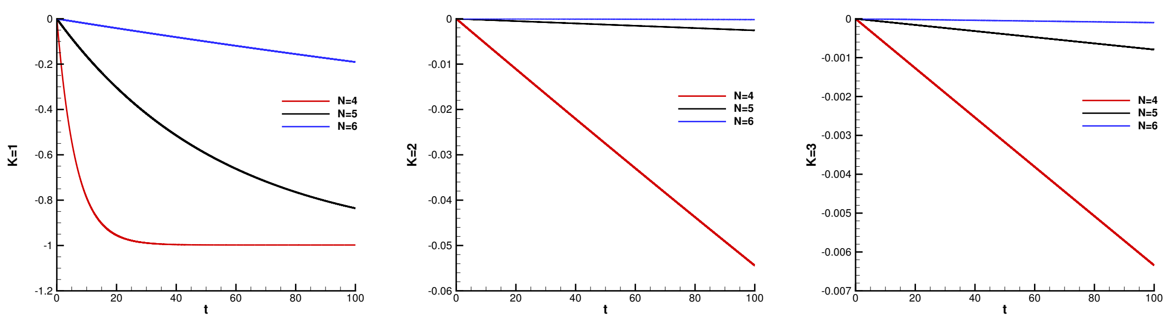
In the simulation, we compute the numerical solutions up to two periods in time, meaning that we let final time $T = 1$ for $d = 2$, $T = 2/3$ for $d = 3$, and $T = 0.5$ for $d = 4$.

We first test the scheme with periodic boundary condition. In Table 4.3, we report the

L^2 errors and orders of accuracy for $k = 1, 2, 3$ and up to dimension four. As for accuracy, we observe about half order reduction from the optimal $(k+1)$-th order for high-dimensional computations ($d = 4$). The order is slightly better for lower dimensions. The convergence order is similar to the performance of the sparse grid DG scheme in [38]. In Figure 4.3, we plot the time evolution of the error of L^2 norm of numerical solutions u_h and v_h, which is given by

$$\int_\Omega \left((u_h(t,\mathbf{x}))^2 + (v_h(t,\mathbf{x}))^2\right) d\mathbf{x} - \int_\Omega \left((u_h(0,\mathbf{x}))^2 + (v_h(0,\mathbf{x}))^2\right) d\mathbf{x}$$

for two-dimensional case for $t = 0$ to $t = 100$. From Theorem 4.2.1, such errors are proportional to the difference between u_h and v_h. We can clearly see that the higher order accurate scheme performs way better in conservation of L^2 norm due to its higher order accuracy.



(a) k=1 (b) k=2 (c) k=3

Figure 4.3: Example 4.3.1. The time evolution of the error of L^2 norm of numerical solutions u_h and v_h of the sparse grid CDG method with $d = 2$. (a) k=1, (b) k=2, (c) k=3. $N = 4, 5, 6$.

Then, we test the scheme with Dirichlet boundary condition prescribed at the inflow edge according to the exact solution. The results are listed in Table 4.4. The accuracy order is similar to the periodic case.

Finally, we use this example to compare the performance of the DG, CDG, sparse grid

Table 4.3: L^2 errors and orders of accuracy for Example 4.3.1 at $T = 1$ when $d = 2$, $T = 2/3$ when $d = 3$, and $T = 0.5$ when $d = 4$. N denotes mesh level, h_N is the size of the smallest mesh in each direction, k is the polynomial order, d is the dimension. L^2 order is calculated with respect to h_N.

		L^2 error	order	L^2 error	order	L^2 error	order
N	h_N	$k = 1$		$k = 2$		$k = 3$	
		$d = 2$					
3	1/8	3.14E-01	–	1.20E-02	–	5.84E-04	–
4	1/16	6.99E-02	2.17	2.23E-03	2.43	8.50E-05	2.78
5	1/32	1.34E-02	2.38	4.87E-04	2.20	3.84E-06	4.47
6	1/64	3.43E-03	1.97	5.97E-05	3.03	3.89E-07	3.30
7	1/128	9.21E-04	1.90	9.33E-06	2.68	1.80E-08	4.43
		$d = 3$					
3	1/8	6.77E-01	–	5.27E-02	–	2.13E-03	–
4	1/16	3.56E-01	0.93	1.10E-02	2.26	2.62E-04	3.02
5	1/32	1.05E-01	1.76	1.82E-03	2.60	2.85E-05	3.20
6	1/64	2.54E-02	2.05	5.22E-04	1.80	2.01E-06	3.83
7	1/128	7.45E-03	1.77	6.89E-05	2.92	2.01E-07	3.32
		$d = 4$					
3	1/8	7.13E-01	–	1.26E-01	–	4.41E-03	–
4	1/16	6.48E-01	0.14	3.39E-02	1.89	7.56E-04	2.54
5	1/32	3.80E-01	0.77	6.91E-03	2.29	9.82E-05	2.94
6	1/64	1.37E-01	1.47	1.39E-03	2.31	9.44E-06	3.38
7	1/128	3.81E-02	1.85	3.56E-04	1.97	8.16E-07	3.53

Table 4.4: L^2 errors and orders of accuracy for Example 4.3.1 with Dirichlet boundary condition on the inflow edges at $T = 1$ when $d = 2$ and $T = 2/3$ when $d = 3$. N denotes mesh level, h_N is the size of the smallest mesh on the primal mesh in each direction, k is the polynomial order, d is the dimension. L^2 order is calculated with respect to h_N.

		L^2 error	order	L^2 error	order	L^2 error	order
N	h_N	$k = 1$		$k = 2$		$k = 3$	
		$d = 2$					
3	1/8	2.66E-01	–	1.66E-02	–	8.21E-04	–
4	1/16	7.47E-02	1.83	3.33E-03	2.32	8.80E-05	3.22
5	1/32	1.94E-02	1.95	5.97E-04	2.48	4.79E-06	4.20
6	1/64	5.44E-03	1.83	8.60E-05	2.80	4.50E-07	3.41
7	1/128	1.49E-03	1.87	1.35E-05	2.67	2.20E-08	4.35
		$d = 3$					
3	1/8	6.15E-01	–	5.34E-02	–	2.67E-03	–
4	1/16	2.86E-01	1.10	1.40E-02	1.93	2.87E-04	3.22
5	1/32	1.14E-01	1.33	2.57E-03	2.45	3.21E-05	3.16
6	1/64	3.23E-02	1.82	5.82E-04	2.14	2.60E-06	3.63
7	1/128	1.03E-02	1.65	9.81E-05	2.57	2.86E-07	3.18

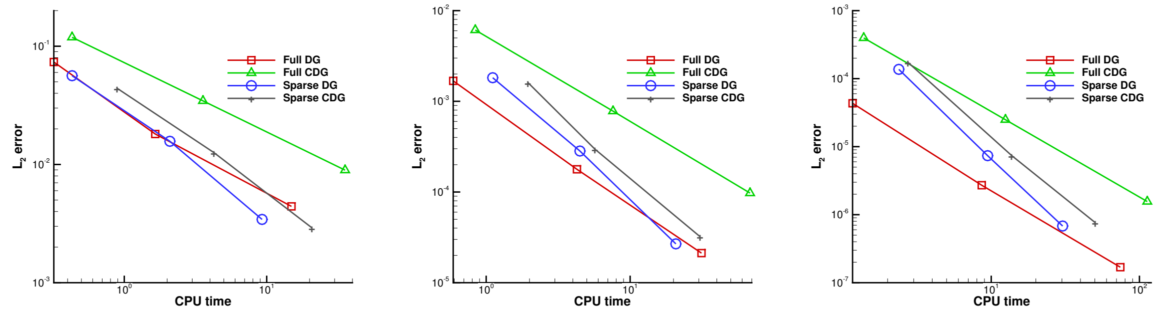
DG and sparse grid CDG methods. We use the following non-separable initial condition

$$u(0,\mathbf{x}) = \exp\left(\sin\left(2\pi\sum_{i=1}^{d} x_i\right)\right), \quad \mathbf{x} \in [0,1]^d, \tag{4.24}$$

where $d = 2$. When $k = 1, 2, 3$, Runge-Kutta methods of order $\nu = 2, 3, 4$, respectively, are used for time discretization. We take the time step according to the CFL numbers listed in Table 4.1. We plot the comparison of the methods measuring L^2 errors vs. CPU times in Figure 4.4. The computations in this example are implemented by an OpenMP code using computational resources from the Institute for Cyber-Enabled Research in Michigan State University. We can see that the sparse grid CDG method outperforms the CDG method, and the sparse grid DG method outperforms the DG method particularly when the mesh level N is more refined. When the mesh level increases from N to $N + 1$, the CPU cost for sparse grid method grows with the rate of about 4 to 5, while the factor is about 8 to 10 for full grid calculations, respectively, for this 2D case. This shows the advantage of the sparse grid approach. When comparing the sparse grid CDG method with the sparse grid DG method, it seems that for this example, the sparse grid DG method is more efficient. It will be interesting to compare the results for fully nonlinear problems in higher dimensions, for which the CDG method is more advantageous, and this is currently under investigation.

Example 4.3.2 (Solid body rotation)**.** *We consider solid-body-rotation problems, which are in the form of* (4.1) *with periodic boundary conditions and*

- $d = 2,\ A_1(t,\mathbf{x}) = -x_2 + \frac{1}{2},\ A_2(t,\mathbf{x}) = x_1 - \frac{1}{2},$
- $d = 3,\ A_1(t,\mathbf{x}) = -\frac{\sqrt{2}}{2}\left(x_2 - \frac{1}{2}\right),\ A_2(t,\mathbf{x}) = \frac{\sqrt{2}}{2}\left(x_1 - \frac{1}{2}\right) + \frac{\sqrt{2}}{2}\left(x_3 - \frac{1}{2}\right),\ A_3(t,\mathbf{x}) = -\frac{\sqrt{2}}{2}\left(x_2 - \frac{1}{2}\right).$



(a) k=1 (b) k=2 (c) k=3

Figure 4.4: L^2 errors and associated CPU times of DG, CDG, sparse grid DG and sparse grid CDG methods for Example 4.1 with initial condition (4.24) at $T = 1$ for d=2. (a) k=1, (b) k=2, (c) k=3.

Such benchmark tests are commonly used in the literature to assess performance of transport schemes. Here, the initial profile traverses along circular trajectories centered at $(1/2, 1/2)$ for $d = 2$ and about the axis $\{x_1 = x_3\} \cap \{x_2 = 1/2\}$ for $d = 3$ without deformation, and it goes back to the initial state after 2π in time. The initial conditions are set to be the following smooth cosine bells (with C^5 smoothness),

$$u(0, \mathbf{x}) = \begin{cases} b^{d-1} \cos^6\left(\frac{\pi r}{2b}\right), & \text{if} \quad r \leq b, \\ 0, & \text{otherwise}, \end{cases} \tag{4.25}$$

where $b = 0.23$ when $d = 2$ and $b = 0.45$ when $d = 3$, and $r = |\mathbf{x} - \mathbf{x}_c|$ denotes the distance between $\mathbf{x}$ and the center of the cosine bell with $\mathbf{x}_c = (0.75, 0.5)$ for $d = 2$ and $\mathbf{x}_c = (0.5, 0.55, 0.5)$ for $d = 3$.

In Table 4.5, we summarize the convergence study of the numerical solutions computed by the sparse CDG method, including the L^2 errors and orders of accuracy. For this variable coefficients equation, we observe at least k-th order convergence for all cases. The order is slightly lower than the corresponding ones in Example 4.3.1.

Table 4.5: L^2 errors and orders of accuracy for Example 4.3.2 at $T = 2\pi$. N denotes mesh level, h_N is the size of the smallest mesh in each direction, k is the polynomial order, d is the dimension. L^2 order is calculated with respect to h_N.

		L^2 error	order	L^2 error	order	L^2 error	order
N	h_N	$k=1$		$k=2$		$k=3$	
		$d=2$					
5	1/32	1.53E-02	–	5.81E-03	–	1.34E-03	–
6	1/64	1.02E-02	0.58	1.50E-03	1.95	9.64E-05	3.80
7	1/128	4.66E-03	1.13	1.46E-04	3.36	1.16E-05	3.05
8	1/256	1.42E-03	1.71	2.34E-05	2.64	1.10E-06	3.40
		$d=3$					
5	1/32	4.83E-03	–	6.25E-04	–	7.35E-05	–
6	1/64	1.87E-03	1.37	1.20E-04	2.38	9.18E-06	3.00
7	1/128	7.46E-04	1.33	3.39E-05	1.82	1.36E-06	2.75
8	1/256	2.55E-04	1.55	8.11E-06	2.06	1.94E-07	2.81

Example 4.3.3 (Deformational flow)**.** *We consider the two-dimensional deformational flow with velocity field*

$$A_1(t, \mathbf{x}) = \sin^2(\pi x_1)\sin(2\pi x_2)g(t),\ A_2(t, \mathbf{x}) = -\sin^2(\pi x_2)\sin(2\pi x_1)g(t),$$

where $g(t) = \cos(\pi t/T)$ *with* $T = 1.5$, *with periodic boundary condition.*

We still adopt the cosine bell (4.25) as the initial condition for this test, but with $\mathbf{x}_c = (0.65, 0.5)$ and $b = 0.35$. Note that the deformational test is more challenging than the solid body rotation due to the space and time dependent flow field. In particular, along the direction of the flow, the cosine bell deforms into a crescent shape at $t = T/2$, then goes back to its initial state at $t = T$ as the flow reverses. In the simulations, we compute the solution up to $t = T$. The convergence study is summarized in Table 4.6. Similar orders are observed compared with Example 4.3.2. In Figure 4.5, we plot the contour plots of the numerical solutions on the primal mesh at $t = T/2$ when the shape of the bell is greatly deformed, and $t = T$ when the solution is recovered into its initial state. It is observed that the sparse CDG scheme with higher degree k can better resolve the highly deformed solution structure.

Table 4.6: L^2 errors and orders of accuracy for Example 4.3.3 at $T = 1.5$. N denotes mesh level, h_N is the size of the smallest mesh in each direction, k is the polynomial order, d is the dimension. L^2 order is calculated with respect to h_N. $d = 2$.

N	h_N	L^2 error	order	L^2 error	order	L^2 error	order
		$k=1$		$k=2$		$k=3$	
5	1/32	1.73E-02	–	4.37E-03	–	1.14E-03	–
6	1/64	8.06E-03	1.10	1.17E-03	1.90	2.44E-04	2.22
7	1/128	3.29E-03	1.29	2.04E-04	2.52	2.05E-05	3.57
8	1/256	1.08E-03	1.61	2.78E-05	2.88	2.75E-06	2.90

4.3.2 System case

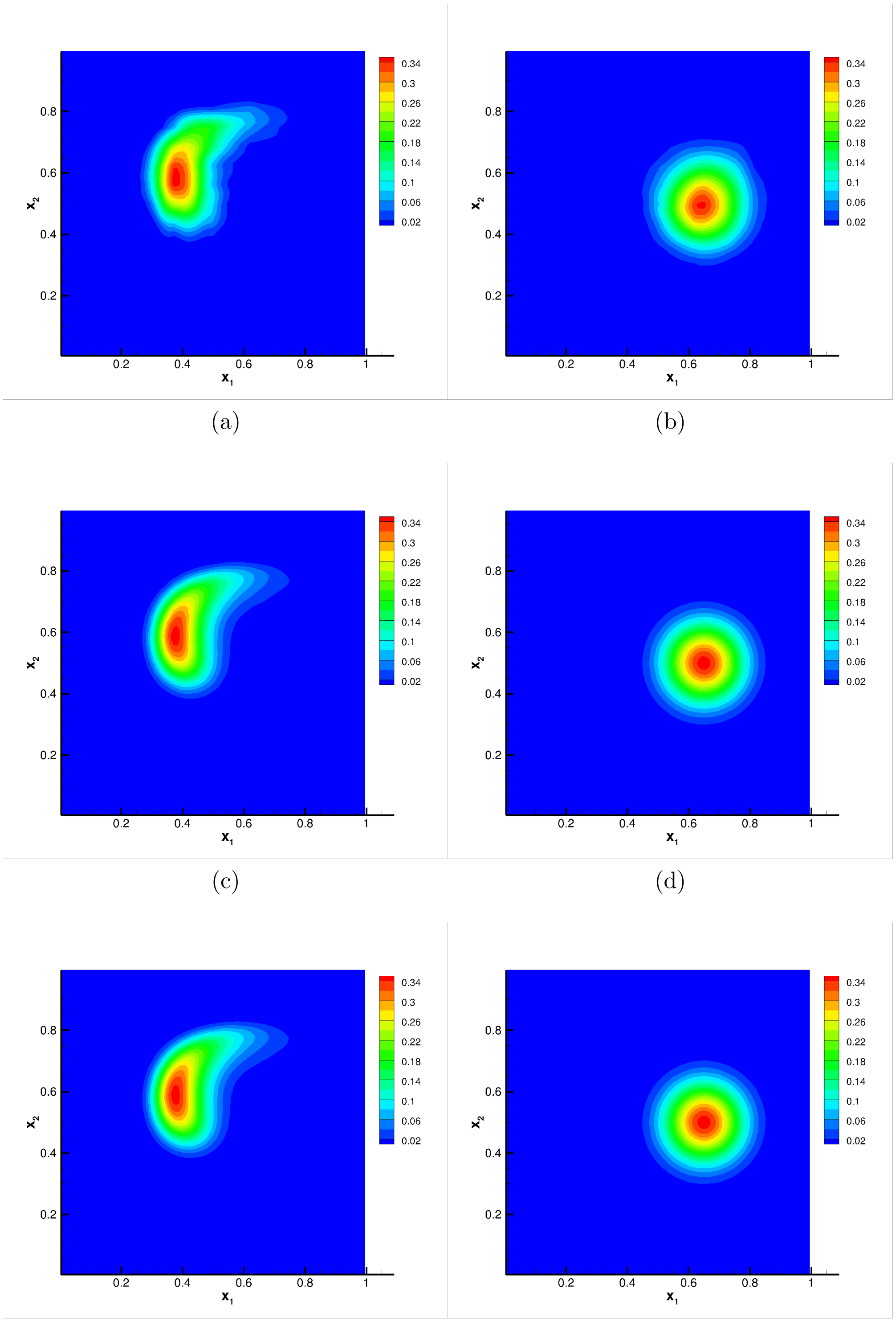
In this subsection, we consider system case, which means $m > 1$ in equation (4.1) or (4.2).

Example 4.3.4 (Acoustic wave equation with constant wave speed)**.** *We consider*

$$\begin{cases} u_t = \nabla \cdot \mathbf{v}, \quad \mathbf{x} \in [0,1]^2, \\ \mathbf{v}_t = \nabla u, \\ u(0,\mathbf{x}) = u_0(\mathbf{x}), \quad \mathbf{v}(0,\mathbf{x}) = \mathbf{v}_0(\mathbf{x}). \end{cases} \tag{4.26}$$

with periodic boundary conditions. The initial conditions $u_0(\mathbf{x})$ and $\mathbf{v}_0(\mathbf{x})$ are chosen according to the following two types of exact solutions: the standing wave

$$\begin{bmatrix} u(t,\mathbf{x}) \\ v_1(t,\mathbf{x}) \\ v_2(t,\mathbf{x}) \end{bmatrix} = \begin{bmatrix} -\sqrt{2}\sin(2\sqrt{2}\pi t)\sin(2\pi x_1)\sin(2\pi x_2) \\ \cos(2\sqrt{2}\pi t)\cos(2\pi x_1)\sin(2\pi x_2) \\ \cos(2\sqrt{2}\pi t)\sin(2\pi x_1)\cos(2\pi x_2) \end{bmatrix},$$



(a) (b) (c) (d) (e) (f)

Figure 4.5: Example 4.3.3. Deformational flow test. The contour plots of the numerical solutions on primal mesh at $t = T/2$ (a, c, e) and $t = T$ (b, d, f). $k = 1$ (a, b), $k = 2$ (c, d), and $k = 3$ (e, f). $N = 7$.

and the traveling wave

$$\begin{bmatrix} u(t,\mathbf{x}) \\ v_1(t,\mathbf{x}) \\ v_2(t,\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \sqrt{2}\sin(2\sqrt{2}\pi t + 2\pi x_1)\cos(2\pi x_2) \\ \sin(2\sqrt{2}\pi t + 2\pi x_1)\cos(2\pi x_2) \\ \cos(2\sqrt{2}\pi t + 2\pi x_1)\sin(2\pi x_2) \end{bmatrix}.$$

We compute the solution until $T = 1$. Similar to the scalar case, we present the L^2 errors and orders of accuracy for $\mathbf{u}(t,\mathbf{x}) = \begin{bmatrix} u(t,\mathbf{x}), & v_1(t,\mathbf{x}), & v_2(t,\mathbf{x}) \end{bmatrix}^T$ in Table 4.7. From the table, we still observe at least $(k+1/2)$-th order for the solution.

Table 4.7: L^2 errors and orders of accuracy for Example 4.3.4 at $T = 1$. N denotes mesh level, h_N is the size of the smallest mesh in each direction, k is the polynomial order, d is the dimension. L^2 order is calculated with respect to h_N. $d = 2$.

		L^2 error	order	L^2 error	order	L^2 error	order
N	h_N	$k=1$		$k=2$		$k=3$	
		standing wave					
3	1/8	3.56E-01	–	1.05E-02	–	5.37E-04	–
4	1/16	7.93E-02	2.17	1.84E-03	2.51	4.31E-05	3.64
5	1/32	1.50E-02	2.40	3.18E-04	2.53	3.39E-06	3.67
6	1/64	3.72E-03	2.01	4.95E-05	2.68	2.77E-07	3.61
7	1/128	1.01E-03	1.88	7.60E-06	2.70	2.03E-08	3.77
		traveling wave					
3	1/8	3.97E-01	–	1.85E-02	–	7.75E-04	–
4	1/16	8.58E-02	2.21	3.36E-03	2.46	6.76E-05	3.52
5	1/32	1.97E-02	2.12	6.07E-04	2.47	5.68E-06	3.57
6	1/64	5.36E-03	1.88	9.66E-05	2.65	4.44E-07	3.68
7	1/128	1.50E-03	1.84	1.45E-05	2.74	3.39E-08	3.71

Example 4.3.5 (Two-dimensional homogeneous isotropic elastic wave [44]). *The 2D elastic wave equation in homogeneous and isotropic medium in velocity-stress formulation without external source, is a linear hyperbolic system of the form*

$$\mathbf{u}_t + A_1\mathbf{u}_{x_1} + A_2\mathbf{u}_{x_2} = 0, \tag{4.27}$$

where $\mathbf{u} = \begin{bmatrix} \sigma_{xx}, & \sigma_{yy}, & \sigma_{xy}, & v, & w \end{bmatrix}^T$, σ_{xx}, σ_{yy} *represents the normal stress and* σ_{xy} *represents the shear stress and* v, w *are the velocity in* x *and* y *directions.*

$$A_1 = -\begin{bmatrix} 0 & 0 & 0 & \lambda+2\mu & 0 \\ 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & \mu \\ \frac{1}{\rho} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\rho} & 0 & 0 \end{bmatrix}, \quad A_2 = -\begin{bmatrix} 0 & 0 & 0 & 0 & \lambda \\ 0 & 0 & 0 & 0 & \lambda+2\mu \\ 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & \frac{1}{\rho} & 0 & 0 \\ 0 & \frac{1}{\rho} & 0 & 0 & 0 \end{bmatrix},$$

where λ *and* μ *are the Lamé constants and* ρ *is the mass density of material. Eigenvalues of* A_1 *and* A_2 *are* $-c_p, -c_s, 0, c_s, c_p$*, which give us the wave speed* $c_p = \sqrt{\frac{\lambda+2\mu}{\rho}}$ *and* $c_s = \sqrt{\frac{\mu}{\rho}}$ *for P-wave and S-wave respectively. We consider the homogeneous material parameters* $\lambda = 2, \mu = 1, \rho = 1$*, then* $c_p = 2, c_s = 1$*. On domain* $\Omega = [0,1]^2$*, we take the solutions consisting of a plane P-wave traveling along diagonal direction* $\mathbf{n} = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ *and a plane S-wave traveling in the opposite direction, i.e.,*

$$\mathbf{u}(t, \mathbf{x}) = \mathbf{R}_s e^{\sin(\mathbf{k}\cdot\mathbf{x}+kc_s t)} + \mathbf{R}_p e^{\sin(\mathbf{k}\cdot\mathbf{x}-kc_p t)},$$

where $\mathbf{R}_s = [-\mu, \mu, 0, -\frac{\sqrt{2}}{2}c_s, \frac{\sqrt{2}}{2}c_s]^T, \mathbf{R}_p = [\lambda+\mu, \lambda+\mu, \mu, -\frac{\sqrt{2}}{2}c_p, -\frac{\sqrt{2}}{2}c_p]^T$ *and* $\mathbf{k} = k\mathbf{n}, k = 2\sqrt{2}\pi$*. Periodic boundary condition is applied and the initial condition is chosen as* $u(0, \mathbf{x})$*.*

We compute the solution until $T = 1$. The L^2 errors and orders of accuracy for $\mathbf{u}(t, \mathbf{x})$ are shown in Table 4.8. We observe that the convergence order is close to $k+1$.

Example 4.3.6 (Three-dimensional isotropic elastic wave [31])**.** *We extend the previous*

Table 4.8: L^2 errors and orders of accuracy for Example 4.3.5 at $T = 1$. N denotes mesh level, h_N is the size of the smallest mesh in each direction, k is the polynomial order, dimension $d = 2$. L^2 order is calculated with respect to h_N.

		L^2 error	order	L^2 error	order	L^2 error	order
N	h_N	$k=1$		$k=2$		$k=3$	
4	1/16	1.09E+00	–	2.72E-01	–	5.71E-02	–
5	1/32	7.47E-01	0.55	6.48E-02	2.07	6.19E-03	3.21
6	1/64	2.41E-01	1.63	9.65E-03	2.75	4.77E-04	3.70
7	1/128	7.14E-02	1.76	1.12E-03	3.11	2.55E-05	4.23

example to 3D and obtain the following linear hyperbolic system

$$\mathbf{u}_t + A_1\mathbf{u}_{x_1} + A_2\mathbf{u}_{x_2} + A_3\mathbf{u}_{x_3} = 0, \tag{4.28}$$

where $\mathbf{u} = \begin{bmatrix}\sigma_{xx}, & \sigma_{yy}, & \sigma_{zz}, & \sigma_{xy}, & \sigma_{yz}, & \sigma_{xz}, & u, & v, & w\end{bmatrix}^T$, *$\sigma$ is the stress tensor and u, v, w are the velocities in each spatial direction.*

$$A_1 = -\begin{bmatrix} 0&0&0&0&0&0&\lambda+2\mu&0&0\\ 0&0&0&0&0&0&\lambda&0&0\\ 0&0&0&0&0&0&\lambda&0&0\\ 0&0&0&0&0&0&0&\mu&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&\mu\\ \frac{1}{\rho}&0&0&0&0&0&0&0&0\\ 0&0&0&\frac{1}{\rho}&0&0&0&0&0\\ 0&0&0&0&0&\frac{1}{\rho}&0&0&0 \end{bmatrix}, A_2 = -\begin{bmatrix} 0&0&0&0&0&0&0&\lambda&0\\ 0&0&0&0&0&0&0&\lambda+2\mu&0\\ 0&0&0&0&0&0&0&\lambda&0\\ 0&0&0&0&0&0&\mu&0&0\\ 0&0&0&0&0&0&0&0&\mu\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&\frac{1}{\rho}&0&0&0&0&0\\ 0&\frac{1}{\rho}&0&0&0&0&0&0&0\\ 0&0&0&0&\frac{1}{\rho}&0&0&0&0 \end{bmatrix},$$

$$A_3 = -\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda+2\mu \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\rho} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\rho} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\rho} & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

where λ, μ and ρ take the same values as the previous example. Hence, we have the same values for c_p and c_s. Eigenvalues of A_1, A_2 and A_3 are $-c_p, -c_s, -c_s, 0, 0, 0, c_s, c_s, c_p$, which describe the wave speed for P-wave and S-wave (with different polarizations). On domain $\Omega = [0,1]^3$, we take the solutions consisting of a plane S-wave traveling along diagonal direction $\mathbf{n} = (-\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}})$ and a plane P-wave traveling in the opposite direction, i.e.,

$$\mathbf{u}(t,\mathbf{x}) = \mathbf{R}_s \sin(\mathbf{k}\cdot\mathbf{x} - kc_s t) + \mathbf{R}_p \sin(\mathbf{k}\cdot\mathbf{x} + kc_p t),$$

where

$$\mathbf{R}_s = [-\frac{2}{3}\mu, \frac{2}{3}\mu, 0, 0, \frac{1}{3}\mu, -\frac{1}{3}\mu, -\frac{1}{\sqrt{3}}c_s, \frac{1}{\sqrt{3}}c_s, 0]^T,$$
$$\mathbf{R}_p = [\lambda + \frac{2}{3}\mu, \lambda + \frac{2}{3}\mu, \lambda + \frac{2}{3}\mu, \frac{2}{3}\mu, \frac{2}{3}\mu, \frac{2}{3}\mu, -\frac{1}{\sqrt{3}}c_p, -\frac{1}{\sqrt{3}}c_p, -\frac{1}{\sqrt{3}}c_p]^T$$

and $\mathbf{k} = k\mathbf{n}, k = -2\sqrt{3}\pi$. Similarly, we consider periodic boundary condition and $u_0(\mathbf{x}) = u(0,\mathbf{x})$ as initial condition.

We present the numerical results at $T = 1$. In Table 4.9, we get at least $(k + 1/2)$-th order of accuracy for the solution $\mathbf{u}(t, \mathbf{x})$.

Table 4.9: L^2 errors and orders of accuracy for Example 4.3.6 at $T = 1$. N denotes mesh level, h_N is the size of the smallest mesh in each direction, k is the polynomial order, d is the dimension. L^2 order is calculated with respect to h_N. $d = 3$.

		L^2 error	order	L^2 error	order	L^2 error	order
N	h_N	$k = 1$		$k = 2$		$k = 3$	
4	1/16	2.49E+00	–	4.93E-02	–	8.91E-04	–
5	1/32	7.70E-01	1.69	8.17E-03	2.59	8.66E-05	3.36
6	1/64	1.76E-01	2.13	1.59E-03	2.36	7.12E-06	3.60
7	1/128	4.27E-02	2.04	2.79E-04	2.51	5.42E-07	3.72

APPENDIX

Detailed discussions on the choice of the T matrix as in (51) or (52)

We discuss what parameters result in $|b_1 \pm b_2| = 0$, under the assumption that α_1 has no dependence on h, $\beta_1 = \tilde{\beta}_1 h^{p_1}, \beta_2 = \tilde{\beta}_2 h^{p_2}, \tilde{\beta}_1, \tilde{\beta}_2$ are nonzero constants that do not depend on h.

$$\begin{aligned} b_1 - b_2 &= (-\beta_1 + \frac{k(k-1)}{2h})(1 - \beta_2 \frac{2k(k-1)}{h}) + \frac{k(k-1)}{h} 2\alpha_1^2 \\ &= (-\tilde{\beta}_1 h^{p_1} + \frac{k(k-1)}{2} h^{-1})(1 - 2k(k-1)\tilde{\beta}_2 h^{p_2 - 1}) + k(k-1)2\alpha_1^2 h^{-1}, \end{aligned}$$

$$\begin{aligned} b_1 + b_2 &= (-\beta_1 + \frac{k(k+1)}{2h})(1 - \beta_2 \frac{2k(k+1)}{h}) + \frac{k(k+1)}{h} 2\alpha_1^2 \\ &= (-\tilde{\beta}_1 h^{p_1} + \frac{k(k+1)}{2} h^{-1})(1 - 2k(k+1)\tilde{\beta}_2 h^{p_2 - 1}) + k(k+1)2\alpha_1^2 h^{-1}. \end{aligned}$$

If $b_1 - b_2 = 0, \forall h < h_0$, then

- $\alpha_1 \neq 0$, then $p_1 = -1,\ p_2 = 1$ and $\tilde{\beta}_1, \tilde{\beta}_2$ satisfies

$$(-\tilde{\beta}_1 + \frac{k(k-1)}{2})(1 - 2k(k-1)\tilde{\beta}_2) + k(k-1)2\alpha_1^2 = 0. \tag{29}$$

Similarly, for $b_1 + b_2 = 0, \forall h < h_0$, then

- $\alpha_1 \neq 0$, $p_1 = -1,\ p_2 = 1$ and $\tilde{\beta}_1, \tilde{\beta}_2$ satisfies

$$(-\tilde{\beta}_1 + \frac{k(k+1)}{2})(1 - 2k(k+1)\tilde{\beta}_2) + k(k+1)2\alpha_1^2 = 0. \tag{30}$$

Detailed discussions on assumption A2

Parameter choices for $|\Gamma| = |\Lambda|$ imply

$$\begin{aligned}
\Gamma \pm \Lambda &= \beta_1 + \frac{k^2(k^2-1)}{h^2}\beta_2 + \frac{k(k\pm 1)}{h}(-2\alpha_1^2 - 2\beta_1\beta_2) + \frac{-k^2 \pm k}{2h} \\
&= (\beta_1 - \frac{k(k\mp 1)}{2h})(1 - 2\beta_2\frac{k(k\pm 1)}{h}) - \frac{k(k\pm 1)}{h}2\alpha_1^2 = 0,
\end{aligned}$$

which indicates

- if $\alpha_1 \neq 0$, then $b_1 \pm b_2$ can be greatly simplified as follows.
  - If $\Gamma + \Lambda = 0$, then k is odd, and

$$\begin{aligned}
b_1 + b_2 &= \frac{k}{h}\left(1 - \beta_2\frac{2k(k+1)}{h}\right), \\
b_1 - b_2 &= -\frac{2}{k+1}\left(\beta_1 - \frac{k(k-1)}{2h}\right), \\
\Lambda &= -\frac{1}{k+1}\left(\beta_1 - \frac{k^2}{h} + \frac{k^2(k^2-1)}{h^2}\beta_2\right).
\end{aligned}$$

  - If $\Gamma - \Lambda = 0$, then k is even, and

$$\begin{aligned}
b_1 + b_2 &= \frac{2}{k-1}\left(\beta_1 - \frac{k(k+1)}{2h}\right), \\
b_1 - b_2 &= -\frac{k}{h}\left(1 - \beta_2\frac{2k(k-1)}{h}\right), \\
\Lambda &= -\frac{1}{k-1}\left(\beta_1 - \frac{k^2}{h} + \frac{k^2(k^2-1)}{h^2}\beta_2\right), \quad k > 1.
\end{aligned}$$

- If $\alpha_1 = 0$, then

$$\beta_1 = \frac{k(k \pm 1)}{2h}, \text{or } \beta_2 = \frac{h}{2k(k \pm 1)}. \tag{31}$$

More estimates related to Legendre coefficients

We provide estimates of the Legendre coefficients, especially their difference in neighboring cells of equal size.

If $u \in W^{k+2+n,\infty}(I)$, then expand $\hat{u}_j(\xi)$ at $\xi = -1$ in (2.24) by Taylor series, we have for $m \geq k+1$, $\exists z \in [-1, 1]$, s.t.

$$\begin{aligned} u_{j,m} &= C \int_{-1}^{1} \frac{d}{d\xi^{k+1}} \Big(\sum_{s=0}^{n} \frac{d}{d\xi^s} \hat{u}_j(-1) \frac{(\xi+1)^s}{s!} \\ &+ \frac{d}{d\xi^{n+1}} \hat{u}_j(z) \frac{(\xi+1)^{n+1}}{(n+1)!} \Big) \frac{d}{d\xi^{m-k-1}} (\xi^2 - 1)^m d\xi, \\ &= \sum_{s=0}^{n} \theta_s h_j^{k+1+s} u^{(k+1+s)}(x_{j-\frac{1}{2}}) + O(h_j^{k+2+n} |u|_{W^{k+2+n,\infty}(I_j)}), \end{aligned} \tag{32}$$

where θ_s are constants independent of u and h_j.

When $h_j = h_{j+1}$, we use Taylor expansion again, and compute the difference of two $u_{j,m}$ from neighboring cells

$$u_{j,m} - u_{j+1,m} = \sum_{s=1}^{n} \mu_s h_j^{k+1+s} u^{(k+1+s)}(x_{j-\frac{1}{2}}) + O(h_j^{k+2+n} |u|_{W^{k+2+n,\infty}(I_j \cup I_{j+1})}). \tag{33}$$

Then we obtain the estimates

$$|u_{j,m} - u_{j+1,m} + \sum_{s=1}^{n} \mu_s h_j^{k+1+s} u^{(k+1+s)}(x_{j-\frac{1}{2}})| \leq C h^{k+2+n} |u|_{W^{k+2+n,\infty}(I_j \cup I_{j+1})}, \tag{34}$$

where μ_s are constants independent of u and h_j.

Two convolution-like operators

To facilitate the analysis in Chapter 2 and 3, we define two operators on a periodic functions u in $L^2(I)$:

$$\boxplus u(x) = \sum_{l=0}^{N-1} (-1)^l \frac{-N+2l}{2} u(x + L\frac{l}{N}), \quad N \text{ is odd}, \tag{35a}$$

$$\boxtimes_\lambda u(x) = \frac{1}{1-\lambda^N} \sum_{l=0}^{N-1} \lambda^l u(x + L\frac{l}{N}), \quad |\lambda| = 1, \tag{35b}$$

where $L = b - a$ is the size of I and N is odd in (35a).

Expand u by Fourier series, i.e., $u(x) = \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{2\pi i n x/L}$, we have

$$\begin{aligned}
\boxplus u(x) &= \sum_{l=0}^{N-1} (-1)^l \frac{-N+2l}{2} \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{in(\frac{2\pi}{L}x + 2\pi\frac{l}{N})} \\
&= \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{\frac{2\pi}{L}inx} \sum_{l=0}^{N-1} \frac{-N+2l}{2} (-e^{i2\pi\frac{n}{N}})^l \\
&= \sum_{n=-\infty}^{\infty} \frac{-2e^{2\pi i\frac{n}{N}}}{(1+e^{2\pi i\frac{n}{N}})^2} \hat{f}(n) e^{\frac{2\pi}{L}inx}, \\
\boxtimes_\lambda u(x) &= \frac{1}{1-\lambda^N} \sum_{l=0}^{N-1} \lambda^l \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{in(\frac{2\pi}{L}x + 2\pi\frac{l}{N})} \\
&= \frac{1}{1-\lambda^N} \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{\frac{2\pi}{L}inx} \sum_{l=0}^{N-1} (\lambda e^{i2\pi\frac{n}{N}})^l \\
&= \sum_{n=-\infty}^{\infty} \frac{\hat{f}(n)}{1-\lambda e^{2\pi i\frac{n}{N}}} e^{\frac{2\pi}{L}inx}.
\end{aligned}$$

In addition, we can apply the operator on the same function recursively, we have

$$\begin{aligned}
\boxtimes_{\lambda_1}^{\nu_1} \cdots \boxtimes_{\lambda_n}^{\nu_n} u(x) &= \sum_{n=-\infty}^{\infty} \frac{1}{(1-\lambda_1 e^{2\pi i \frac{n}{N}})^{\nu_1}} \cdots \frac{1}{(1-\lambda_n e^{2\pi i \frac{n}{N}})^{\nu_n}} \hat{f}(n) e^{i\frac{2\pi}{L} inx}, \\
(\boxtimes_{\lambda})^{\nu} u(x) &= \sum_{n=-\infty}^{\infty} \frac{\hat{f}(n)}{(1-\lambda e^{2\pi i \frac{n}{N}})^{\nu}} e^{\frac{2\pi}{L} inx}, \\
(\boxplus)^{\nu} u(x) &= \sum_{n=-\infty}^{\infty} \Big(\frac{-2e^{2\pi i \frac{n}{N}}}{(1+e^{2\pi i \frac{n}{N}})^2} \Big)^{\nu} \hat{f}(n) e^{\frac{2\pi}{L} inx}.
\end{aligned}$$

Next, we estimate the two operators. Assuming $u \in W^{3,1}(I)$, then the Fourier coefficient $\hat{f}(n)$ satisfies:

$$\left|\hat{f}(n)\right| \leq C \frac{|u|_{W^{k+4,1}(I)}}{1+|n|^3}. \tag{36}$$

Since N is odd, then $\omega^n = e^{2\pi i \frac{n}{N}} \neq -1, \forall n$. Hence, $\boxplus u(x)$ are well defined. We estimate $\boxplus u(x)$ by splitting it into blocks of size N as

$$\boxplus u(x) = \sum_{l=-\infty}^{\infty} S_l, \quad \text{where} \quad S_l = \sum_{n=lN-\frac{N-1}{2}}^{lN+\frac{N-1}{2}} \frac{-2e^{2\pi i \frac{n}{N}}}{(1+e^{2\pi i \frac{n}{N}})^2} \hat{f}(n) e^{\frac{2\pi}{L} inx}.$$

Let's estimate S_{k+1}^0 first. Denote $W_1(n) = \frac{-2e^{2\pi i \frac{n}{N}}}{(1+e^{2\pi i \frac{n}{N}})^2} e^{\frac{2\pi}{L} inx}$. For $|n| \leq [\frac{3N}{8}]$, $|W_1(n)| = \frac{2}{|1+\omega^n|^2} \leq \frac{2}{|1+e^{i3\pi/4}|^2} = \frac{2}{2-\sqrt{2}}$. For other n, $|W_1(n)| \leq |W_1(\frac{N-1}{2})| = \frac{2}{|1+\omega^{(N-1)/2}|^2} \leq CN^2$

from Taylor expansions.

$$
\begin{aligned}
|S_0| &\leq \sum_{n=-[\frac{3N}{8}]}^{[\frac{3N}{8}]} \left|\hat{f}(n)W_1(n)\right| + \sum_{-\frac{N-1}{2}}^{n=-[\frac{3N}{8}]-1} \left|\hat{f}(n)W_1(n)\right| + \sum_{n=[\frac{3N}{8}]+1}^{\frac{N-1}{2}} \left|\hat{f}(n)W_1(n)\right| \\
&\leq \frac{2}{2-\sqrt{2}} \sum_{n=-[\frac{3N}{8}]}^{[\frac{3N}{8}]} \left|\hat{f}(n)\right| + CN^2 \sum_{-\frac{N-1}{2}}^{n=-[\frac{3N}{8}]-1} \left|\hat{f}(n)\right| + CN^2 \sum_{n=[\frac{3N}{8}]+1}^{\frac{N-1}{2}} \left|\hat{f}(n)\right| \\
&\leq \frac{2}{2-\sqrt{2}} \sum_{n=-[\frac{3N}{8}]}^{[\frac{3N}{8}]} \left|\hat{f}(n)\right| + CN^2 \frac{1}{1+(\frac{3N}{8})^3}(\frac{N}{4}+2)|u|_{W^{3,1}(I)} \\
&\leq C\left(\sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} \frac{1}{1+|n|^3} + \frac{1}{(\frac{3}{8})^3}\right)|u|_{W^{3,1}(I)}.
\end{aligned}
$$

Then, in a similar way,

$$
|S_l| \leq C\left(\sum_{n=lN-\frac{N-1}{2}}^{lN+\frac{N-1}{2}} \frac{1}{1+|n|^3} + \frac{1}{(|l|+\frac{3}{8})^3}\right)|u|_{W^{3,1}(I)}.
$$

Therefore,

$$
|\boxplus u(x)| \leq C\left(\sum_{n=-\infty}^{\infty} \frac{1}{1+|n|^3} + \sum_{l=-\infty}^{\infty} \frac{1}{(|l|+\frac{3}{8})^3}\right)|u|_{W^{3,1}(I)} \leq C|u|_{W^{3,1}(I)}. \tag{37}
$$

Similar to the estimation for $\boxplus u(x)$ above, we split $\boxtimes$ into blocks of size N,

$$
\boxtimes_\lambda u(x) = \sum_{l=-\infty}^{\infty} \mathcal{S}_l, \quad \text{where} \quad \mathcal{S}_l = \sum_{n=lN}^{(l+1)N-1} \hat{f}(n)W_2(n),\ W_2(n) = \frac{e^{\frac{2\pi}{L}inx}}{1-\lambda e^{2\pi i\frac{n}{N}}}.
$$

$W_2(n)$ is singular when λ is close to any n-th root of unity. Assuming $\left|1-\lambda^N\right| =$

$O(h^{\delta'})$ and $|\lambda - 1| = O(h^{\delta/2})$ with $0 \le \delta/2 \le 1$. We can write $\lambda = e^{\pm i\theta}$ and assume $\theta \in (0, \pi)$ without loss of generality. First, we establish a relation between δ and δ'. Since $|\lambda| = |\lambda^N| = 1$, we have $\delta, \delta' \ge 0$. Because $1 - \lambda^N = (e^{2\pi in/N})^N - (e^{i\theta})^N = (e^{2\pi in/N} - e^{i\theta})(\sum_{l=0}^{N-1}(e^{2\pi in/N})^{N-1-l}(e^{i\theta})^l)$, thus $\left|1 - \lambda^N\right| \le N\left|\omega^n - e^{i\theta}\right|, \forall n$. With the assumption $\left|1 - \lambda^N\right| \sim Ch^{\delta'}$, we get $\left|\omega^n - e^{i\theta}\right| \ge Ch^{\delta'+1}$. Particularly, when $n = 0$, we have $|1 - \lambda| \ge Ch^{\delta'+1}$, hence $\delta/2 \le \delta' + 1$.

In addition, $|W_2(n)| = |\lambda - \omega^n|^{-1} \le Ch^{-(\delta'+1)}$. With the assumption that $0 \le \delta/2 \le 1$, there $\exists\, n_0 \sim O(h^{\delta/2-1})$ s.t. $2\pi\frac{n_0}{N} \le \theta < 2\pi\frac{n_0+1}{N}$. Let $n_1 = \lfloor n_0/2 \rfloor, n_2 = 2n_0 - n_1$, then for $n_1 \le n \le n_2$, $\left|\hat{f}(n)\right| \le C\frac{1}{1+n_1^2}|u|_{W^{2,1}(I)}$. For other n, $|w_2(n)| \le |w_2(n_1)| \le \frac{1}{2|\sin(\pi n_1/N - \theta/2)|} \le Ch^{-\delta/2}$. Thus,

$$
\begin{aligned}
|\mathcal{S}_0| &\le Ch^{-\delta/2}\left(\sum_{n=0}^{n_1-1} + \sum_{n=n_2+1}^{N-1}\left|\hat{f}(n)\right|\right) + Ch^{-(\delta'+1)}\sum_{n=n_1}^{n_2}\left|\hat{f}(n)\right| \\
&\le C\left(h^{-\delta/2}\sum_{n=0}^{N-1}\frac{1}{1+|n|^2} + h^{-(\delta'+1)}(n_2 - n_1 + 1)\frac{1}{1+n_1^2}\right)|u|_{W^{2,1}(I)} \\
&\le C\left(h^{-\delta/2}\sum_{n=0}^{N-1}\frac{1}{1+|n|^2} + h^{-(\delta'+1)}h^{\delta/2-1}h^{2-\delta}\right)|u|_{W^{2,1}(I)} \\
&\le C\left(h^{-\delta/2}\sum_{n=0}^{N-1}\frac{1}{1+|n|^2} + h^{-\delta'-\delta/2}\right)|u|_{W^{2,1}(I)}.
\end{aligned}
$$

Using similar approaches, for $l \ne 0$,

$$
|\mathcal{S}_l| \le C\left(h^{-\delta/2}\sum_{n=lN}^{(l+1)N-1}\frac{1}{1+|n|^2} + h^{-\delta'+\delta/2}\frac{1}{|n_1/N + l|^2}\right)|u|_{W^{2,1}(I)}.
$$

Summing up, we reach the estimation

$$
\begin{aligned}
|\boxtimes_\lambda u(x)| &\leq C\left(h^{-\delta/2}\sum_{n=-\infty}^{\infty}\frac{1}{1+|n|^2}+h^{-\delta'-\delta/2}+h^{-\delta'+\delta/2}\sum_{l\in\mathbb{N},l\neq 0}\frac{1}{|l|^2}\right)|u|_{W^{2,1}(I)} \\
&\leq Ch^{-\delta'-\delta/2}|u|_{W^{2,1}(I)}.
\end{aligned}
\tag{38}
$$

Corollary .0.7. *When $\lambda_i, i \leq n$ is a complex number with $|\lambda_i| = 1$, independent of h, above estimates yields following results:*

$$
\left|\boxtimes_{\lambda_1}^{\nu_1}\cdots\boxtimes_{\lambda_n}^{\nu_n}u(x)\right| \leq C|u|_{W^{1+\sum_{i=1}^n \nu_i,1}(I)}, \quad |(\boxplus u(x))^\nu| \leq C|u|_{W^{1+2\nu,1}(I)}. \tag{39}
$$

Estimates for $\mathcal{M}_{j,m}$

Let's recall the definition

$$
\mathcal{M}_{j,m} = (A_j + B_j)^{-1}(GL_{j,m}^- + HL_{j,m}^+), \quad \forall m \in \mathbb{Z}^+, \forall j \in \mathbb{Z}_N,
$$

where $G, H, A_j, B_j, L_{j,m}^-, L_{j,m}^+, \Gamma_j, \Lambda_j$ are defined in Table 2.1.

$$
A_j + B_j = G[L_{j,k-1}^-, L_{j,k}^-] + H[L_{j,k-1}^+, L_{j,k}^+] = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{h_j}\end{bmatrix} M_+ + \begin{bmatrix}\alpha_1 & -\beta_2 \\ -\beta_1 & -\alpha_1\end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{h_j}\end{bmatrix} M_-,
$$

where

$$
\begin{aligned}
M_{\pm} &= \begin{bmatrix} 1 & 0 \\ 0 & h_j \end{bmatrix} [L^{-}_{j,k-1} \pm L^{+}_{j,k-1}, L^{-}_{j,k} \pm L^{+}_{j,k}] \\
&= \begin{bmatrix} 1 \pm (-1)^{k-1} & 1 \pm (-1)^{k} \\ k(k-1)(1 \pm (-1)^{k}) & k(k+1)(1 \pm (-1)^{k+1}) \end{bmatrix}.
\end{aligned}
$$

Therefore,

$$
(A_j + B_j)^{-1} = \frac{1}{D_1} M_{-}^{-1} \begin{bmatrix} -\alpha_1 & \beta_2 - \frac{h_j}{2k(k+(-1)^k)} \\ \beta_1 h_j - \frac{k(k-(-1)^k)}{2} & \alpha_1 h_j \end{bmatrix},
$$

where $D_1 = \frac{(-1)^k h_j}{2k(k+(-1)^k)}((-1)^k \Gamma_j + \Lambda_j)$.

In what follows, we estimate $\mathcal{M}_{j,m}$ when scale-invariant flux parameters are used in (40), and when $\alpha_1^2 + \beta_1 \beta_2 = \frac{1}{4}$ in (41).

- Scale-invariant flux parameters.

D_1 is bounded by definitions of Γ_j, Λ_j and mesh regularity condition. Then

$$(A_j + B_j)^{-1} G \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{h_j} \end{bmatrix} =$$

$$\frac{1}{D_1} M_-^{-1} \begin{bmatrix} -\alpha_1 & \tilde{\beta}_2 h h_j^{-1} - \frac{1}{2k(k+(-1)^k)} \\ \tilde{\beta}_1 h^{-1} h_j - \frac{k(k-(-1)^k)}{2} & \alpha_1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} + \alpha_1 & -\tilde{\beta}_2 h h_j^{-1} \\ -\tilde{\beta}_1 h^{-1} h_j & \frac{1}{2} - \alpha_1 \end{bmatrix},$$

$$(A_j + B_j)^{-1} H \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{h_j} \end{bmatrix} =$$

$$\frac{1}{D_1} M_-^{-1} \begin{bmatrix} -\alpha_1 & \tilde{\beta}_2 h h_j^{-1} - \frac{1}{2k(k+(-1)^k)} \\ \tilde{\beta}_1 h^{-1} h_j - \frac{k(k-(-1)^k)}{2} & \alpha_1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} - \alpha_1 & \tilde{\beta}_2 h h_j^{-1} \\ \tilde{\beta}_1 h^{-1} h_j & \frac{1}{2} + \alpha_1 \end{bmatrix}$$

and

$$\begin{aligned} \mathcal{M}_{j,m} =& (A_j + B_j)^{-1} G \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{h_j} \end{bmatrix} \begin{bmatrix} 1 \\ m(m+1) \end{bmatrix} \\ &+ (-1)^m (A_j + B_j)^{-1} H \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{h_j} \end{bmatrix} \begin{bmatrix} 1 \\ -m(m+1) \end{bmatrix}. \end{aligned}$$

If the mesh is uniform, the three formulas above are independent of mesh size h. For nonuniform mesh, by mesh regularity condition, $\exists \sigma_1, \sigma_2, s.t., \sigma_1 \le h^{-1} h_j \le \sigma_2$, therefore,

$$\left\| (A_j + B_j)^{-1} G \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{h_j} \end{bmatrix} \right\|_\infty \le C, \ \left\| (A_j + B_j)^{-1} H \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{h_j} \end{bmatrix} \right\|_\infty \le C, \ \|\mathcal{M}_{j,m}\|_\infty \le C. \tag{40}$$

- $\alpha_1^2 + \beta_1\beta_2 = \frac{1}{4}$.

  $\Lambda_j = 0$. $D_1 = \frac{h_j\Gamma_j}{2k(k+(-1)^k)}$ and above formulas simplifies to

$$(A_j + B_j)^{-1}G \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{h_j} \end{bmatrix} =$$

$$\frac{1}{2D_1}M_-^{-1} \begin{bmatrix} -\alpha_1 - \frac{1}{2} + \frac{\beta_1 h_j}{k(k+(-1)^k)} & \beta_2 h_j^{-1} - \frac{1}{2k(k+(-1)^k)} \\ \beta_1 h_j - \frac{k(k-(-1)^k)}{2} - \alpha_1 k(k-(-1)^k) & \alpha_1 - \frac{1}{2} + \beta_2 h_j^{-1} k(k-(-1)^k) \end{bmatrix},$$

$$(A_j + B_j)^{-1}H \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{h_j} \end{bmatrix} =$$

$$\frac{1}{2D_1}M_-^{-1} \begin{bmatrix} \alpha_1 - \frac{1}{2} - \frac{\beta_1 h_j}{k(k+(-1)^k)} & \beta_2 h_j^{-1} - \frac{1}{2k(k+(-1)^k)} \\ \beta_1 h_j - \frac{k(k-(-1)^k)}{2} + \alpha_1 k(k-(-1)^k) & \alpha_1 + \frac{1}{2} - \beta_2 h_j^{-1} k(k-(-1)^k) \end{bmatrix}.$$

Thus, we have the following estimation

$$\|\mathcal{M}_{j,m}\|_\infty \leq \left(1 + \frac{\max\left(|\beta_1|, \frac{|\frac{1}{2}-\alpha_1|}{h}, \frac{|\frac{1}{2}+\alpha_1|}{h}, \frac{|\beta_2|}{h^2}\right)}{|\Gamma_j|}\right). \tag{41}$$

Proof of Lemma 2.2.2

By Definition 2.2.1, $P_h^\star u|_{I_j} = \sum_{m=0}^{k-2} u_{j,m}L_{j,m} + \acute{u}_{j,k-1}L_{j,k-1} + \acute{u}_{j,k}L_{j,k}$. We solve the two coefficients $\acute{u}_{j,k-1}, \acute{u}_{j,k}$ on every cell I_j according to definition (2.16).

If assumption A0 is satisfied, it has been shown in Lemma 2.2.1 that (2.16) is equivalent to (2.20). Substitute u and u_x by (2.23), we obtain the following equation

$$(A_j + B_j) \begin{bmatrix} \acute{u}_{j,k-1} \\ \acute{u}_{j,k} \end{bmatrix} = (A_j + B_j) \begin{bmatrix} u_{j,k-1} \\ u_{j,k} \end{bmatrix} + \sum_{m=k+1}^{\infty} u_{j,m}(GL^-_{j,m} + HL^+_{j,m}), \tag{42}$$

the existence and uniqueness of the system above is ensured by assumption A0, that is, $\det(A_j + B_j) = 2(-1)^k \Gamma_j \neq 0$. And (2.26) is proven by multiplying $(A_j + B_j)^{-1}$ on both sides of above equality.

If any of the assumption A1/A2/A3 is satisfied, (2.16) can be written as

$$G \sum_{m=0}^{k} \acute{u}_{j,m} L^-_m + H \sum_{m=0}^{k} \acute{u}_{j+1,m} L^+_m = G \sum_{m=0}^{\infty} u_{j,m} L^-_m + H \sum_{m=0}^{\infty} u_{j+1,m} L^+_m,$$

where we used (2.23) and $G + H = I_2$ in above equality. Since $\acute{u}_{j,m} = u_{j,m}$ when $m \leq k-2$,

$$A \begin{bmatrix} \acute{u}_{j,k-1} \\ \acute{u}_{j,k} \end{bmatrix} + B \begin{bmatrix} \acute{u}_{j+1,k-1} \\ \acute{u}_{j+1,k} \end{bmatrix} = \sum_{m=k-1}^{\infty} u_{j,m} G L^-_m + u_{j+1,m} H L^+_m.$$

In order to solve for $\acute{u}_{j,k-1}, \acute{u}_{j,k}$, we group above coupled equations for all j in a $2N \times 2N$ linear system as follows,

$$M \begin{bmatrix} \acute{u}_{1,k-1} \\ \acute{u}_{1,k} \\ \cdots \\ \acute{u}_{N-1,k-1} \\ \acute{u}_{N-1,k} \\ \acute{u}_{N,k-1} \\ \acute{u}_{N,k} \end{bmatrix} = \begin{bmatrix} \eta_1 \\ \theta_1 \\ \cdots \\ \eta_{N-1} \\ \theta_{N-1} \\ \eta_N \\ \theta_N \end{bmatrix}, \tag{43}$$

where

$$\begin{bmatrix} \eta_j \\ \\ \theta_j \end{bmatrix} = G \sum_{m=k-1}^{\infty} u_{j,m} L_m^- + H \sum_{m=k-1}^{\infty} u_{j+1,m} L_m^+,$$

and $M = circ(A, B, 0_2, \cdots, 0_2)$, denoting a $2N \times 2N$ block-circulant matrix with first two rows as $(A, B, 0_2, \cdots, 0_2)$, with 0_2 as a 2×2 zero matrix. We can calculate that

$$\det A = \det B = \frac{-2k}{h}(\alpha_1^2 + \beta_1\beta_2 - \frac{1}{4}) \neq 0. \tag{44}$$

It is clear that the existence and uniqueness of $P_h^\star$ is equivalent to $\det M \neq 0$. By a direct computation, $\det M = \det A^N \det(I_2 - Q^N)$, where I_2 denotes the 2×2 identity matrix, and

$$Q = -A^{-1}B = \frac{(-1)^{k+1}}{\Lambda} \begin{bmatrix} c_1 + c_2 & b_1 + b_2 \\ \\ b_1 - b_2 & c_1 - c_2 \end{bmatrix},$$

with

$$c_1 = \beta_1 + \frac{k^2(k^2-1)}{h^2}\beta_2 - \frac{2k^2}{h}(\alpha_1^2 + \beta_1\beta_2 + \frac{1}{4}) := \Gamma, \tag{45}$$

$$c_2 = \frac{k}{h}(2\alpha_1), \tag{46}$$

$$b_1 = -\beta_1 - \frac{k^2(k^2+1)}{h^2}\beta_2 + \frac{2k^2}{h}(\alpha_1^2 + \beta_1\beta_2 + \frac{1}{4}), \tag{47}$$

$$b_2 = -\frac{2k^3}{h^2}\beta_2 + \frac{2k}{h}(\alpha_1^2 + \beta_1\beta_2 + \frac{1}{4}). \tag{48}$$

The eigenvalues of Q are

$$\lambda_1 = \frac{(-1)^{(k+1)}}{\Lambda}(\Gamma + \sqrt{\Gamma^2 - \Lambda^2}), \quad \lambda_2 = \frac{(-1)^{(k+1)}}{\Lambda}(\Gamma - \sqrt{\Gamma^2 - \Lambda^2}). \tag{49}$$

Since $\det Q = \det B / \det A = 1$, we have the relations $\lambda_1 \lambda_2 = 1$ and

$$b_1^2 - b_2^2 = \Gamma^2 - \Lambda^2 - c_2^2. \tag{50}$$

Below we discuss the existence and uniqueness of $P_h^\star$ based on the types of eigenvalues of Q.

A1 If $|\Gamma| > |\Lambda|$, then $\lambda_{1,2}$ are real and different. Therefore, we can perform eigenvalue decomposition of Q,

$$Q = TDT^{-1},$$

where

$$D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix},$$

and

$$T = \begin{bmatrix} 1 & -\frac{b_1+b_2}{c_2+\sqrt{\Gamma^2-\Lambda^2}} \\ \frac{b_1-b_2}{c_2+\sqrt{\Gamma^2-\Lambda^2}} & 1 \end{bmatrix}, T^{-1} = \frac{1}{\det T} \begin{bmatrix} 1 & \frac{b_1+b_2}{c_2+\sqrt{\Gamma^2-\Lambda^2}} \\ -\frac{b_1-b_2}{c_2+\sqrt{\Gamma^2-\Lambda^2}} & 1 \end{bmatrix}, \tag{51}$$

where $\det T = \frac{2\sqrt{\Gamma^2-\Lambda^2}}{c_2+\sqrt{\Gamma^2-\Lambda^2}}$, except for the case when $(b_1 - b_2)(b_1 + b_2) = 0$ and $c_2 < 0$, where

$$T = \begin{bmatrix} 1 & -\frac{b_1+b_2}{2c_2} \\ \frac{b_1-b_2}{2c_2} & 1 \end{bmatrix}, T^{-1} = \begin{bmatrix} 1 & \frac{b_1+b_2}{2c_2} \\ -\frac{b_1-b_2}{2c_2} & 1 \end{bmatrix}. \tag{52}$$

In both situations, we have

$$\det M = \det A^N \det(I_2 - \begin{bmatrix} \lambda_1^N & 0 \\ 0 & \lambda_2^N \end{bmatrix}) = \det A^N \det(\begin{bmatrix} 1-\lambda_1^N & 0 \\ 0 & 1-\lambda_2^N \end{bmatrix}).$$

$\det M \neq 0$ if and only if $(\lambda_1)^N \neq 1$ and $(\lambda_2)^N \neq 1$, which is true since $|\lambda_1| \neq 1$ and $|\lambda_2| \neq 1$.

A2 If $|\Gamma| = |\Lambda|$, then $\lambda_1 = \lambda_2 = (-1)^{k+1}\frac{\Gamma}{\Lambda}$ and we have two repeated eigenvalues. Perform Jordan decomposition:

$$\begin{bmatrix} c_1+c_2 & b_1+b_2 \\ b_1-b_2 & c_1-c_2 \end{bmatrix} = \mathcal{T} \begin{bmatrix} c_1 & 1 \\ 0 & c_1 \end{bmatrix} \mathcal{T}^{-1},$$

and

$$\mathcal{T} = \begin{bmatrix} c_2 & 1 \\ b_1-b_2 & 0 \end{bmatrix}, \qquad \text{if}\, b_1 \neq b_2, \tag{53}$$

$$\mathcal{T} = \begin{bmatrix} 2b_1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \text{if}\, b_1 = b_2.$$

We define

$$\mathcal{J} = \frac{(-1)^{k+1}}{\Lambda} \begin{bmatrix} c_1 & 1 \\ 0 & c_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \frac{(-1)^{k+1}}{\Lambda} \\ 0 & \lambda_1 \end{bmatrix}, \quad Q = \mathcal{T}\mathcal{J}\mathcal{T}^{-1},$$

then

$$
\begin{aligned}
Q^j &= \mathcal{T}\mathcal{J}^j\mathcal{T}^{-1}, \quad \mathcal{J}^j = \begin{bmatrix} \lambda_1^j & \kappa_j \\ 0 & \lambda_1^j \end{bmatrix}, \\
I_2 - Q^N &= \mathcal{T}\begin{bmatrix} 2 & -\kappa_N \\ 0 & 2 \end{bmatrix}\mathcal{T}^{-1},
\end{aligned}
$$

where $\kappa_j = \frac{(-1)^{(k+1)j}}{\Lambda^j} j\Gamma^{j-1}$.

In both situations, $\det M \neq 0$ if and only if $(\lambda_1)^N \neq 1$, meaning that we require N to be odd and further, if k is odd, we require $\Gamma = -\Lambda$; if k is even, we require $\Gamma = \Lambda$. In both cases, $\lambda_1 = \lambda_2 = -1$.

<u>A3</u> If $|\Gamma| < |\Lambda|$, then $\lambda_{1,2}$ are complex, $|\lambda_{1,2}| = 1$, $\lambda_1 = \overline{\lambda_2}$, still Q is diagonalizable, and similar to A1, $\det M \neq 0$ turns to $(\lambda_1)^N \neq 1$ and $(\lambda_2)^N = \overline{(\lambda_1)^N} \neq 1$, i.e. we require

$$
(-1)^{(k+1)N}\left(\frac{\Gamma}{\Lambda} + \sqrt{\left(\frac{\Gamma}{\Lambda}\right)^2 - 1}\right)^N \neq 1.
$$

Summarize above results, we proved the existence and uniqueness for $P_h^\star$ when any of the assumptions A0/A1/A2/A3 is satisfied.

In order to obtain the exact formula of $\acute{u}_{j,k-1}$ and $\acute{u}_{j,k}$, we analyze the inverse of the matrix M. It is known that the inverse of a nonsingular circulant matrix is also circulant, so is a block-circulant matrix. In particular,

$$
M^{-1} = circ(r_0, r_1, \cdots, r_{N-1}) \otimes A^{-1}
$$

where $\otimes$ means Kronecker product for block matrices and r_l is a 2×2 matrix defined as,

$$r_j = Q^j (I_2 - Q^N)^{-1}, \quad j = 0, \cdots, N-1. \tag{54}$$

Therefore, if any of the assumptions A1/A2/A3 is satisfied,

$$\begin{aligned}
&\begin{bmatrix} \acute{u}_{j,k-1} \\ \acute{u}_{j,k} \end{bmatrix} \\
&= \sum_{l=0}^{N-1} r_l A^{-1} \Big(A \begin{bmatrix} u_{j+l,k-1} \\ u_{j+l,k} \end{bmatrix} + B \begin{bmatrix} u_{j+l+1,k-1} \\ u_{j+l+1,k} \end{bmatrix} + \sum_{m=k+1}^{\infty} u_{j+l,m} G L_m^- + u_{j+l+1,m} H L_m^+ \Big), \\
&= \sum_{l=0}^{N-1} r_l \Big(\begin{bmatrix} u_{j+l,k-1} \\ u_{j+l,k} \end{bmatrix} - Q \begin{bmatrix} u_{j+l+1,k-1} \\ u_{j+l+1,k} \end{bmatrix} + \sum_{m=k+1}^{\infty} u_{j+l,m} [L_{k-1}^-, L_k^-]^{-1} L_m^- \\
&- u_{j+l+1,m} Q [L_{k-1}^+, L_k^+]^{-1} L_m^+ \Big) \\
&= \begin{bmatrix} u_{j,k-1} \\ u_{j,k} \end{bmatrix} + \sum_{m=k+1}^{\infty} \Big(\sum_{l=1}^{N-1} u_{j+l,m} V_{2,m} + u_{j,m} r_0 [L_{k-1}^-, L_k^-]^{-1} L_m^- \\
&- u_{j+N,m} r_N [L_{k-1}^-, L_k^-]^{-1} L_m^- \Big) \\
&= \begin{bmatrix} u_{j,k-1} \\ u_{j,k} \end{bmatrix} + \sum_{m=k+1}^{\infty} \big(u_{j,m} V_{1,m} + \sum_{l=0}^{N-1} u_{j+l,m} r_l V_{2,m} \big),
\end{aligned}$$

where $r_N = Q^N (I_2 - Q^N)^{-1} = r_0 - I_2$ is used in the third equality. And (2.27) is proven.

Proof of Lemma 2.2.3

Denote

$$U_j = \begin{bmatrix} \acute{u}_{j,k-1} - u_{j,k-1} \\ \acute{u}_{j,k} - u_{j,k} \end{bmatrix},$$

When assumption A0 is satisfied (2.28) is a direct result of (2.26) and (41).

If any of the assumptions A1/A2/A3 is satisfied, we have

$$U_j = \begin{bmatrix} \acute{u}_{j,k-1} - u_{j,k-1} \\ \acute{u}_{j,k} - u_{j,k} \end{bmatrix} = \sum_{m=k+1}^{\infty} (u_{j,m} V_{1,m} + \sum_{l=0}^{N-1} u_{j+l,m} r_l V_{2,m}).$$

In order to estimate U_j, we first compute r_l to get its detailed dependence on l. If A1/A3, Q is diagonalizable, then

$$\begin{aligned} r_l = TD^l(I_2 - D^N)^{-1}T^{-1} &= \frac{\lambda_1^l}{1-\lambda_1^N} T \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} T^{-1} + \frac{\lambda_2^l}{1-\lambda_2^N} T \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} T^{-1} \\ &= \frac{\lambda_1^l}{1-\lambda_1^N} Q_1 + \frac{\lambda_2^l}{1-\lambda_2^N} (I_2 - Q_1), \end{aligned} \tag{55}$$

where

$$Q_1 = \frac{1}{2\sqrt{\Gamma^2 - \Lambda^2}} \begin{bmatrix} c_2 + \sqrt{\Gamma^2 - \Lambda^2} & b_1 + b_2 \\ b_1 - b_2 & -c_2 + \sqrt{\Gamma^2 - \Lambda^2} \end{bmatrix}, \tag{56}$$

when T is given by (51), and

$$Q_1 = \frac{1}{2c_2}\begin{bmatrix} 2c_2 & b_1+b_2 \\ b_1-b_2 & 0 \end{bmatrix}, \tag{57}$$

when T is given by (52).

If assumption A2 is satisfied,

$$r_l = Q^l(I_2 - Q^N)^{-1} = \frac{(-1)^l}{2} I_2 + (-1)^l \frac{-N+2l}{4\Gamma} Q_2, \tag{58}$$

where

$$Q_2 = \mathcal{T}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\mathcal{T}^{-1} = \begin{bmatrix} c_2 & b_1+b_2 \\ b_1-b_2 & -c_2 \end{bmatrix}. \tag{59}$$

When assumption A1 is satisfied, eigenvalues λ_1, λ_2 are real.

$$\sum_{l=0}^{N-1} \left|\frac{\lambda_{1,2}^l}{1-\lambda_{1,2}^N}\right| = \frac{1}{1-|\lambda_{1,2}|}\frac{1-|\lambda_{1,2}|^N}{|1-\lambda_{1,2}^N|}.$$

Without loss of generality, we assume $|\lambda_1| < 1 < |\lambda_2|$, then

$$\sum_{l=0}^{N-1} \left|\frac{\lambda_1^l}{1-\lambda_1^N}\right| \le \frac{1}{1-|\lambda_1|} = \frac{|\lambda_2|}{|\lambda_2|-1},$$
$$\sum_{l=0}^{N-1} \left|\frac{\lambda_2^l}{1-\lambda_2^N}\right| \le \frac{1}{|\lambda_2|-1}.$$

And thus

$$\sum_{l=0}^{N-1} \|r_l\|_\infty \le \frac{\|Q_1\|_\infty}{1-|\lambda_1|} + \frac{\|I_2-Q_1\|_\infty}{1-|\lambda_2|}. \tag{60}$$

Then we have

$$
\begin{aligned}
\|U_j\|_\infty &\le C(1+\sum_{l=0}^{N-1}\|r_l\|_\infty)|u|_{W^{k+1,\infty}(I)} \\
&\le Ch^{k+1}|u|_{W^{k+1,\infty}(I)}(1+\frac{\|Q_1\|_\infty}{1-|\lambda_1|}+\frac{\|I_2-Q_1\|_\infty}{1-|\lambda_2|}),
\end{aligned}
$$

where (2.25) and the fact that $V_{1,m}, V_{2,m}, \forall m \ge 0$ are constant matrices independent of h are used in above inequalities.

When assumption A3 is satisfied, eigenvalues λ_1, λ_2 are complex, with $|\lambda_{1,2}| = 1$ and above estimation does not apply. We perform more detailed computation and use Fourier analysis to bound U_j by utilizing the smoothness and periodicity of u. If $u \in W^{k+2+n,\infty}(I)$,

$$
\begin{aligned}
U_j &= \sum_{l=0}^{N-1} r_l \sum_{m=k+1}^{\infty} \big(u_{l+j,m} - \sum_{s=0}^{n}\theta_s h^{k+1+s} u^{(k+1+s)}(x_{j+l-\frac{1}{2}})) V_{2,m}\big) \\
&+ \sum_{l=0}^{N-1} r_l \sum_{m=k+1}^{\infty}\sum_{s=0}^{n}\theta_s h^{k+1+s} u^{(k+1+s)}(x_{j+l-\frac{1}{2}}) V_{2,m} + \sum_{m=k+1}^{\infty} u_{j,m} V_{1,m} \\
&= O(h^{k+2+n}|u|_{W^{k+2+n,\infty}(I_j)}) + \sum_{m=k+1}^{\infty} u_{j,m} V_{1,m} \\
&+ \sum_{m=k+1}^{\infty}\sum_{s=0}^{n}\theta_s h^{k+1+s} \sum_{l=0}^{N-1}(\frac{\lambda_1^l}{1-\lambda_1^N} Q_1 + \frac{\lambda_2^l}{1-\lambda_2^N}(I_2-Q_1)) u^{(k+1+s)}(x_{j+l-\frac{1}{2}}) V_{2,m} \\
&= O(h^{k+2+n}|u|_{W^{k+2+n,\infty}(I_j)}) + \sum_{m=k+1}^{\infty} u_{j,m} V_{1,m} \\
&+ \sum_{m=k+1}^{\infty}\sum_{s=0}^{n}\theta_s h^{k+1+s} (Q_1 \boxtimes_{\lambda_1} + (I_2-Q_1)\boxtimes_{\lambda_2}) u^{(k+1+s)}(x_{j-\frac{1}{2}}) V_{2,m} \\
&\le Ch^{k+1}\|u\|_{W^{k+3,\infty}(I)}(1+h^{-\delta'-\delta/2}(\|Q_1\|_\infty + \|I_2-Q_1\|_\infty)),
\end{aligned}
$$

where (32) is used in the first equality and (38) is used in the last inequality.

When assumption A2 is satisfied, by similar computation, if $u \in W^{k+2+n,\infty}(I)$,

$$
\begin{aligned}
U_j &= O(h^{k+2+n}|u|_{W^{k+2+n,\infty}(I_j)}) + \sum_{m=k+1}^{\infty} u_{j,m} V_{1,m} \\
&+ \sum_{m=k+1}^{\infty} \sum_{s=0}^{n} \frac{1}{2}\theta_s \sum_{l=0}^{N-1} \big((-1)^l + (-1)^l \frac{-N+2l}{2\Gamma} Q_2\big) h^{k+1+s} u^{(k+1+s)}(x_{j+l-\frac{1}{2}}) V_{2,m} \\
&\le Ch^{k+1}|u|_{W^{k+1,\infty}(I_j)} + \left| \sum_{m=k+1}^{\infty} \sum_{s=0}^{n} \theta_s h^{k+1+s} \frac{Q_2}{2\Gamma} \boxplus u^{(k+1+s)}(x_{j-\frac{1}{2}}) V_{2,m} \right| \\
&+ \left| \sum_{m=k+1}^{\infty} \sum_{s=0}^{n} \frac{1}{2}\theta_s h^{k+1+s} \big(u^{(k+1+s)}(x_{j+N-\frac{3}{2}}) + \sum_{l'=0}^{\frac{N-1}{2}} u^{(k+1+s)}(x_{j+2l'-\frac{1}{2}}) \right. \\
&\left. - u^{(k+1+s)}(x_{j+2l'+\frac{1}{2}})\big) V_{2,m} \right| \\
&\le Ch^{k+1}\|u\|_{W^{k+4,\infty}(I)}(1 + \frac{\|Q_2\|_\infty}{|\Gamma|}),
\end{aligned}
$$

where (34) and (39) are used in the last inequality.

The estimates of U_j for assumptions A1, A2 and A3 are finished, and (2.29), (2.30) and (2.31) are direct results of the estimation of $\|U_j\|_\infty$.

Proof of Lemma 3.2.3

Proof. Since $P_h^\star u = P_h^\dagger u$ when A0, the formula for $\grave{u}_{j,k-1}, \grave{u}_{j,k}$ is the same as (2.26). That is,

$$
\begin{bmatrix} \grave{u}_{j,k-1} \\ \grave{u}_{j,k} \end{bmatrix} = \begin{bmatrix} u_{j,k-1} \\ u_{j,k} \end{bmatrix} + \sum_{m=k+1}^{\infty} u_{j,m} \mathcal{M}_m. \tag{61}
$$

Under assumption A1/A2/A3, above formula is well-defined if and only if $\mathcal{M}_m$ is not singular. By the analysis of $\mathcal{M}_{j,m}$ in Appendix, the existence and uniqueness condition is $\det(A+B) = 2((-1)^k\Gamma + \Lambda) \neq 0$. Thus, by (40) and (2.25), (3.9) is proven.

If any of the assumptions A1/A2/A3 is satisfied, then the difference of two projections can be written as

$$Wu|_{I_j} = P_h^\star u|_{I_j} - P_h^\dagger u|_{I_j} = (\acute{u}_{j,k-1} - \grave{u}_{j,k-1})L_{j,k-1} + (\acute{u}_{j,k} - \grave{u}_{j,k})L_{j,k}.$$

The properties of $P_h^\star u$ and $P_h^\dagger u$ yield the following coupled system

$$A \begin{bmatrix} \acute{u}_{j,k-1} - \grave{u}_{j,k-1} \\ \acute{u}_{j,k} - \grave{u}_{j,k} \end{bmatrix} + B \begin{bmatrix} \acute{u}_{j+1,k-1} - \grave{u}_{j+1,k-1} \\ \acute{u}_{j+1,k} - \grave{u}_{j+1,k} \end{bmatrix} = \begin{bmatrix} \tau_j \\ \iota_j \end{bmatrix}, \quad \forall j \in \mathbb{Z}_N,$$

$$\begin{aligned}
\begin{bmatrix} \tau_j \\ \iota_j \end{bmatrix} &= \begin{bmatrix} u \\ u_x \end{bmatrix}\Bigg|_{x_{j+\frac{1}{2}}} - G \begin{bmatrix} P_h^\dagger u \\ (P_h^\dagger u)_x \end{bmatrix}\Bigg|^-_{x_{j+\frac{1}{2}}} - H \begin{bmatrix} P_h^\dagger u \\ (P_h^\dagger u)_x \end{bmatrix}\Bigg|^+_{x_{j+\frac{1}{2}}} \\
&= G \begin{bmatrix} (u - P_h^\dagger u)|^-_{x_{j+\frac{1}{2}}} - (u - P_h^\dagger u)|^-_{x_{j+\frac{3}{2}}} \\ (u - P_h^\dagger u)_x|^-_{x_{j+\frac{1}{2}}} - (u - P_h^\dagger u)_x|^-_{x_{j+\frac{3}{2}}} \end{bmatrix},
\end{aligned}$$

where the second equality was obtained by the definition of $P_h^\dagger u$ (3.8b).

Gather the relations above for all j results in a large $2N \times 2N$ linear system with block circulant matrix M, defined in (43), as coefficient matrix, then the solution is

$$\begin{bmatrix} \acute{u}_{j,k-1} - \grave{u}_{j,k-1} \\ \acute{u}_{j,k} - \grave{u}_{j,k} \end{bmatrix} = \sum_{l=0}^{N-1} r_l A^{-1} \begin{bmatrix} \tau_{l+j} \\ \iota_{l+j} \end{bmatrix}, \quad j \in \mathbb{Z}_N,$$

where by periodicity, when $l + j > N$, $\tau_{l+j} = \tau_{l+j-N}, \iota_{l+j} = \iota_{l+j-N}$.

On uniform mesh, by the definition of $R_{j,m}$ in (3.12), $R_{j,m}(1)$ and $(R_{j,m})_x(1)$ are independent of j, we denote the corresponding values as $R_m(1)$ and $(R_m)_x(1)$ and let

$R_m^- = [R_m(1), (R_m)_x(1)]^T$. By (3.11), we have

$$\begin{bmatrix} (u-P_h^{\dagger}u)|_{x_{j+\frac{1}{2}}}^- - (u-P_h^{\dagger}u)|_{x_{j+\frac{3}{2}}}^- \\ (u-P_h^{\dagger}u)_x|_{x_{j+\frac{1}{2}}}^- - (u-P_h^{\dagger}u)_x|_{x_{j+\frac{3}{2}}}^- \end{bmatrix} = \sum_{m=k+1}^{\infty} (u_{j,m} - u_{j+1,m}) R_m^-$$

and

$$\begin{bmatrix} \acute{u}_{j,k-1} - \grave{u}_{j,k-1} \\ \acute{u}_{j,k} - \grave{u}_{j,k} \end{bmatrix} = \sum_{l=0}^{N-1} r_l \Big(\sum_{m=k+1}^{\infty} (u_{l+j,m} - u_{l+j+1,m}) A^{-1} G R_m^- \Big) \doteq \mathcal{S}_j, \quad j \in \mathbb{Z}_N. \quad (62)$$

Using (33), we can estimate $\mathcal{S}_j$ by the same lines as the estimation of U_j in the proof of Lemma 2.2.3 in Appendix, and (3.10) is proven. □

Proof of Lemma 3.3.1

Proof. By error equation, the symmetry of $A(\cdot,\cdot)$ and the definition of s_h, we have

$$0 = a(e, v_h) = a(\epsilon_h, v_h) + a(\zeta_h, v_h) = \int_I s_h v_h dx + \int_I (\zeta_h)_t v_h dx - iA(v_h, \zeta_h), \quad \forall v_h \in V_h^k. \tag{63}$$

Now, we are going to choose three special test functions to extract superconvergence properties (3.16)-(3.18) about ζ_h. We first prove (3.16). In order to have $A(v_h, \zeta_h) = \|(\zeta_h)_{xx}\|^2$, we choose a function $v_1 \in V_h^k$, such that $\forall j \in \mathbb{Z}_N, v_1|_{I_j} = \alpha_{j,k-1} L_{j,k-1} + \alpha_{j,k} L_{j,k} + \overline{(\zeta_h)_{xx}}$, $\int_{I_j} v_1 (\zeta_h)_{xx} dx = \|(\zeta_h)_{xx}\|^2_{L^2(I_j)}$, $\hat{v}_1|_{j+\frac{1}{2}} = 0$ and $\widetilde{(v_1)_x}|_{j+\frac{1}{2}} = 0$.

When the assumption A0 holds, the definition of v_1 yields the following local system for each pair of $\alpha_{j,k-1}$ and $\alpha_{j,k}$,

$$(A_j + B_j) \begin{bmatrix} \alpha_{j,k-1} \\ \alpha_{j,k} \end{bmatrix} = -G \begin{bmatrix} \overline{(\zeta_h)^-_{xx}} \\ \overline{(\zeta_h)^-_{xxx}} \end{bmatrix} \Big|_{j+\frac{1}{2}} - H \begin{bmatrix} \overline{(\zeta_h)^+_{xx}} \\ \overline{(\zeta_h)^+_{xxx}} \end{bmatrix} \Big|_{j-\frac{1}{2}}, \quad \forall j \in \mathbb{Z}_N.$$

$$\begin{aligned} \begin{bmatrix} \alpha_{j,k-1} \\ \alpha_{j,k} \end{bmatrix} =& -(A_j + B_j)^{-1} G \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{h_j} \end{bmatrix} \begin{bmatrix} \overline{(\zeta_h)^-_{xx}} \\ h_j \overline{(\zeta_h)^-_{xxx}} \end{bmatrix} \Big|_{j+\frac{1}{2}} \\ & -(A_j + B_j)^{-1} H \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{h_j} \end{bmatrix} \begin{bmatrix} \overline{(\zeta_h)^+_{xx}} \\ h_j \overline{(\zeta_h)^+_{xxx}} \end{bmatrix} \Big|_{j-\frac{1}{2}}, \end{aligned} \tag{64}$$

thus v_1 is nontrivial and uniquely defined under assumption A0. By orthogonality of Legendre polynomials, it follows that

$$\begin{aligned} \|v_1\|^2_{L^2(I_j)} &= |\alpha_{j,k-1}|^2 \int_{I_j} L^2_{j,k-1} dx + |\alpha_{j,k}|^2 \int_{I_j} L^2_{j,k} dx + \|(\zeta_h)_{xx}\|^2_{L^2(I_j)} \\ &\le C(h_j \|(\zeta_h)_{xx}\|^2_{L^2(\partial I_j)} + h_j^3 \|(\zeta_h)_{xxx}\|^2_{L^2(\partial I_j)} + \|(\zeta_h)_{xx}\|^2_{L^2(I_j)}) \\ &\le C \|(\zeta_h)_{xx}\|^2_{L^2(I_j)}, \end{aligned}$$

where (40), trace inequalities and inverse inequalities are used in above inequality.

Let $v_h = v_1$, then (63) becomes

$$0 = \int_I s_h v_1 dx + \int_I (\zeta_h)_t v_1 dx - i\|(\zeta_h)_{xx}\|^2.$$

Hence $\|(\zeta_h)_{xx}\|^2 \le \|s_h + (\zeta_h)_t\| \cdot \|v_1\| \le C\|s_h + (\zeta_h)_t\| \cdot \|(\zeta_h)_{xx}\|$. Therefore, (3.16) is proven when assumption A0 is satisfied.

Similarly, in order to have $A(v_h, \zeta_h) = -\sum_{j=1}^N |[\zeta_h]|^2_{j+\frac{1}{2}}$, we define $v_2 \in V_h^k$, such that

$\forall j \in \mathbb{Z}_N, v_2|_{I_j} = \alpha_{j,k-1}L_{j,k-1} + \alpha_{j,k}L_{j,k}$, $\int_{I_j} v_2(\zeta_h)_{xx}dx = 0$, $\hat{v}_2|_{j+\frac{1}{2}} = 0$ and $\widetilde{(v_2)_x}|_{j+\frac{1}{2}} = \overline{[\zeta_h]}|_{j+\frac{1}{2}}$. When assumption A0 is satisfied, this definition yields the following local system for each pair of $\alpha_{j,k-1}$ and $\alpha_{j,k}$,

$$(A_j + B_j)\begin{bmatrix}\alpha_{j,k-1}\\ \alpha_{j,k}\end{bmatrix} = G\begin{bmatrix}0\\ \overline{[\zeta_h]}\end{bmatrix}\Bigg|_{j+\frac{1}{2}} + H\begin{bmatrix}0\\ \overline{[\zeta_h]}\end{bmatrix}\Bigg|_{j-\frac{1}{2}}, \quad \forall j \in \mathbb{Z}_N.$$

By same algebra as above, we have

$$\begin{bmatrix}\alpha_{j,k-1}\\ \alpha_{j,k}\end{bmatrix} = (A_j+B_j)^{-1}G\begin{bmatrix}1 & 0\\ 0 & \frac{1}{h_j}\end{bmatrix}\begin{bmatrix}0\\ h_j\overline{[\zeta_h]}\end{bmatrix}\Bigg|_{j+\frac{1}{2}} + (A_j+B_j)^{-1}H\begin{bmatrix}1 & 0\\ 0 & \frac{1}{h_j}\end{bmatrix}\begin{bmatrix}0\\ h_j\overline{[\zeta_h]}\end{bmatrix}\Bigg|_{j-\frac{1}{2}}.$$

By (40), it follows directly that

$$\|v_2\|^2_{L^2(I_j)} \le Ch_j^3(|[\zeta_h]|^2_{j+\frac{1}{2}} + |[\zeta_h]|^2_{j-\frac{1}{2}}).$$

Plug v_2 in (63), we obtain

$$\sum_{j=1}^{N} |[\zeta_h]|^2_{j+\frac{1}{2}} = i\int_I s_h v_2 dx + i\int_I (\zeta_h)_t v_2 dx \le \|s_h + (\zeta_h)_t\|\|v_2\|.$$

Therefore, (3.17) is proven when assumption A0 is satisfied.

Finally, in order to have $A(v_h, \zeta_h) = \sum_{j=1}^N |[(\zeta_h)_x]|^2_{j+\frac{1}{2}}$, we choose $v_3 \in V_h^k$, such that $\forall j \in \mathbb{Z}_N, v_3|_{I_j} = \alpha_{j,k-1}L_{j,k-1}+\alpha_{j,k}L_{j,k}$ such that $\int_{I_j} v_3(\zeta_h)_{xx}dx = 0$, $\hat{v}_3|_{j+\frac{1}{2}} = \overline{[(\zeta_h)_x]}|_{j+\frac{1}{2}}$ and $\widetilde{(v_3)_x}|_{j+\frac{1}{2}} = 0$. Follow the same lines as the estimates for v_2, we end up with the estimates

$$\|v_3\|^2_{L^2(I_j)} \le Ch_j(|[(\zeta_h)_x]|^2_{j+\frac{1}{2}} + |[(\zeta_h)_x]|^2_{j-\frac{1}{2}}).$$

Plug v_3 in (63), we obtain (3.18) when assumption A0 is satisfied.

Under assumption A1, we need to compute $\sum_{j=1}^{N}(|\alpha_{j,k-1}|^2+|\alpha_{j,k}|^2)$ to estimate $\|v_1\|^2$. The definition of v_1 yields the following coupled system

$$A\begin{bmatrix}\alpha_{j,k-1}\\ \alpha_{j,k}\end{bmatrix}+B\begin{bmatrix}\alpha_{j+1,k-1}\\ \alpha_{j+1,k}\end{bmatrix}=-G\begin{bmatrix}\overline{(\zeta_h)_{xx}^-}\\ \overline{(\zeta_h)_{xxx}^-}\end{bmatrix}\Bigg|_{j+\frac{1}{2}}-H\begin{bmatrix}\overline{(\zeta_h)_{xx}^+}\\ \overline{(\zeta_h)_{xxx}^+}\end{bmatrix}\Bigg|_{j+\frac{1}{2}},\quad j\in\mathbb{Z}_N. \quad (65)$$

Write it in matrix form

$$M\boldsymbol{\alpha}=\boldsymbol{b},\quad \boldsymbol{\alpha}=[\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_N]^T,$$

where M is defined in (43) and

$$\boldsymbol{\alpha}_j=[\alpha_{j,k-1},\alpha_{j,k}],\boldsymbol{b}=[\boldsymbol{b}_1,\cdots,\boldsymbol{b}_N]^T,\boldsymbol{b}_j=-G\begin{bmatrix}\overline{(\zeta_h)_{xx}^-}\\ \overline{(\zeta_h)_{xxx}^-}\end{bmatrix}\Bigg|_{j+\frac{1}{2}}-H\begin{bmatrix}\overline{(\zeta_h)_{xx}^+}\\ \overline{(\zeta_h)_{xxx}^+}\end{bmatrix}\Bigg|_{j+\frac{1}{2}}.$$

Left multiply A^{-1} in (65), we get an equivalent system

$$M'\boldsymbol{\alpha}=\boldsymbol{b}',\quad M'=circ(I_2,A^{-1}B,0_2,\cdots,0_2),\boldsymbol{b}'=[\boldsymbol{b}'_1,\cdots,\boldsymbol{b}'_N]^T,\boldsymbol{b}'_j=A^{-1}\boldsymbol{b}_j,$$

and $(M')^{-1}=circ(r_0,\cdots,r_{N-1})$. By Theorem 5.6.4 in [29] and similar to the proof in Lemma 3.1 in [10],

$$M'=(\mathcal{F}_N^*\otimes I_2)\boldsymbol{\Omega}(\mathcal{F}_N\otimes I_2),$$

where $\mathcal{F}_N$ is the discrete Fourier transform matrix defined by $(\mathcal{F}_N)_{ij}=\frac{1}{\sqrt{N}}\overline{\omega}^{(i-1)(j-1)}$, $\omega=e^{i\frac{2\pi}{N}}$. $\mathcal{F}_N$ is symmetric and unitary and

$$\boldsymbol{\Omega}=\mathrm{diag}(I_2+A^{-1}B,I_2+\omega A^{-1}B,\cdots,I_2+\omega^{N-1}A^{-1}B).$$

The assumption $\frac{|\Gamma|}{|\Lambda|} > 1$ in A1 ensures that the eigenvalues of $Q = -A^{-1}B$ are not 1, thus $I_2 + \omega^j A^{-1}B, \forall j$, is nonsingular and $\mathbf{\Omega}$ is invertible. Then

$$|\rho((M')^{-1})| = \|(M')^{-1}\|_2 \leq \|\mathcal{F}_N^* \otimes I_2\|_2 \|\mathbf{\Omega}\|_2 \|\mathcal{F}_N \otimes I_2\|_2 \leq C. \tag{66}$$

Therefore,

$$\begin{aligned}\sum_{j=1}^{N}(|\alpha_{j,k-1}|^2 + |\alpha_{j,k}|^2) = \boldsymbol{\alpha}^T\boldsymbol{\alpha} = (\boldsymbol{b}')^T(M')^{-T}(M')^{-1}(\boldsymbol{b}')^T \\ \leq \|(M')^{-1}\|_2^2\|\boldsymbol{b}'\|_2^2 \leq C\sum_{j=1}^{N}\|\boldsymbol{b}'_j\|_2^2.\end{aligned}$$

Since $A^{-1}G\begin{bmatrix}1 & 0\\ 0 & \frac{1}{h}\end{bmatrix}, A^{-1}H\begin{bmatrix}1 & 0\\ 0 & \frac{1}{h}\end{bmatrix}$ are constant matrices, we have

$$\begin{aligned}\|\boldsymbol{b}'_j\|_2^2 &\leq C\left(\left\|\begin{bmatrix}\overline{(\zeta_h)_{xx}^-}\\ \overline{h(\zeta_h)_{xxx}^-}\end{bmatrix}\Big|_{j+\frac{1}{2}}\right\|_2 + \left\|\begin{bmatrix}\overline{(\zeta_h)_{xx}^+}\\ \overline{h(\zeta_h)_{xxx}^+}\end{bmatrix}\Big|_{j+\frac{1}{2}}\right\|_2\right)\\ &\leq C(\|(\zeta_h)_{xx}\|^2_{L^2(\partial I_j)} + \|(\zeta_h)_{xx}\|^2_{L^2(\partial I_{j+1})}\\ &\quad + h^2\|(\zeta_h)_{xxx}\|^2_{L^2(\partial I_j)} + h^2\|(\zeta_h)_{xxx}\|^2_{L^2(\partial I_{j+1})})\\ &\leq C(\|(\zeta_h)_{xx}\|^2_{L^2(\partial I_j)} + \|(\zeta_h)_{xx}\|^2_{L^2(\partial I_{j+1})}),\end{aligned}$$

where inverse inequality is used to obtain the last inequality. Finally, we obtain the estimate

$$\|v_1\|^2 = \sum_{j=1}^{N}|\alpha_{j,k-1}|^2\|L_{j,k-1}\|^2_{L^2(I_j)} + \sum_{j=1}^{N}|\alpha_{j,k}|^2\|L_{j,k}\|^2_{L^2(I_j)} + \|(\zeta_h)_{xx}\|^2$$

$$
\begin{aligned}
&\le \|(\zeta_h)_{xx}\|^2 + Ch\sum_{j=1}^{N}(|\alpha_{j,k-1}|^2 + |\alpha_{j,k}|^2) \\
&\le \|(\zeta_h)_{xx}\|^2 + Ch\sum_{j=1}^{N}(\|(\zeta_h)_{xx}\|^2_{L^2(\partial I_j)} + \|(\zeta_h)_{xx}\|^2_{L^2(\partial I_{j+1})}) \\
&\le \|(\zeta_h)_{xx}\|^2 + Ch\|(\zeta_h)_{xx}\|^2_{L^2(\partial \mathcal{I}_N)} \le C\|(\zeta_h)_{xx}\|^2,
\end{aligned}
$$

where inverse inequality is used to obtain the last inequality. Then the estimates for (3.16) hold true. (3.17) and (3.18) can be proven by the same procedure when assumption A1 is satisfied, and the steps are omitted for brevity. $\square$

Remark .0.1. *When assumption A2 or A3 is satisfied, the eigenvalues of Q are -1 or two complex number with magnitude 1, then a constant bound for $\rho((M')^{-1})$ as in* (66) *is not possible. Therefore, we cannot obtain similar results for assumption A2 and A3.*

Proof for Lemma 3.3.2

Proof. Since $w_q \in V_h^k$, we have

$$
w_q|_{I_j} = \sum_{m=0}^{k} c_{j,m}^q L_{j,m}. \tag{67}
$$

Let $v_h = D^{-2}L_{j,m}, m \le k-2$ in (3.19a), we obtain

$$
c_{j,m}^q = -i\frac{2m+1}{h_j}\frac{h_j^2}{4}\int_{I_j} \partial_t w_{q-1} D^{-2}L_{j,m}dx. \tag{68}
$$

Since $D^{-2}L_{j,m} \in P_c^{m+2}(I_j)$, by the property $u - P_h^\star u \perp V_h^{k-2}$ in the L^2 inner product sense, we have

$$c^1_{j,m} = \begin{cases} -i\frac{2m+1}{h_j}\frac{h_j^2}{4}\int_{I_j}\partial_t(u-P^\star_h u)D^{-2}L_{j,m}dx = 0, & m \le k-4, \\ -i\dfrac{2m+1}{h_j}\dfrac{h_j^2}{4}\displaystyle\int_{I_j}\partial_t((u_{j,k-1}-\acute{u}_{j,k-1})L_{j,k-1} & \\ \qquad\qquad +(u_{j,k}-\acute{u}_{j,k})L_{j,k})D^{-2}L_{j,m}dx, & m = k-3, k-2. \end{cases} \tag{69}$$

By induction using (67), (68), (69), for $0 \le m \le k-2-2q$, $c^q_{j,m} = 0$.

Furthermore, the first nonzero coefficient can be written in a simpler form related to $u_{j,k-1}$ by induction.

When $q = 1$, we compute $c^1_{j,k-3}$ by (69) and the definition of w_0. That is

$$\begin{aligned} c^1_{j,k-3} &= -i\frac{2(k-3)+1}{h_j}\left(\frac{h_j}{2}\right)^2\partial_t(u_{j,k-1}-\acute{u}_{j,k-1})\int_{I_j}D^{-2}L_{j,k-3}L_{j,k-1}dx \\ &= Ch_j^2\partial_t(u_{j,k-1}-\acute{u}_{j,k-1}). \end{aligned}$$

Suppose $c^{q-1}_{k+1-2q} = Ch_j^{2q-2}\partial_t^{q-1}(u_{j,k-1}-\acute{u}_{j,k-1})$, then

$$\begin{aligned} c^q_{j,k-1-2q} &= -i\frac{2(k-1-2q)+1}{h_j}\left(\frac{h_j}{2}\right)^2\partial_t c^{q-1}_{j,k+1-2q}\int_{I_j}D^{-2}L_{j,k-1-2q}L_{j,k+1-2q}dx \\ &= Ch_j{}^{2q}\partial_t^q(u_{j,k-1}-\acute{u}_{j,k-1}). \end{aligned}$$

The induction is completed and the second formula in (3.23) is proven when $r = 0$.

Next, we begin estimating the coefficient $c^q_{j,m}$. By Holder's inequality and (68), we have the estimates for $c^q_{j,m}, k-1-2q \le m \le k-2$,

$$\left|c^q_{j,m}\right| \le Ch^{2-\frac{1}{s}}\|\partial_t w_{q-1}\|_{L^s(I_j)}.$$

To estimate the coefficients $c^q_{j,k-1}, c^q_{j,k}$, we need to discuss them under different assumptions. If assumption A0 is satisfied, meaning (3.19b) and (3.19c) can be decoupled and therefore w_q is locally defined by (3.19). By (3.20) and following the same algebra of solving the k-th and $(k+1)$-th coefficients in (42),

$$\begin{bmatrix} c^q_{j,k-1} \\ c^q_{j,k} \end{bmatrix} = -\sum_{m=0}^{k-2} \mathcal{M}_{j,m} c^q_{j,m}.$$

By (40), for all $j \in \mathbb{Z}_N$,

$$\left|c^q_{j,k-1}\right|^2 + \left|c^q_{j,k}\right|^2 \le C \sum_{m=k-2q-3}^{k-2} \left|c^q_{j,m}\right|^2 \le Ch^3 \|\partial_t w_{q-1}\|^2_{L^2(I_j)}.$$

$$\max(\left|c^q_{j,k-1}\right|, \left|c^q_{j,k}\right|) \le C \max_{k-2q-3 \le m \le k-2} \left|c^q_{j,m}\right| \le Ch^2 \|\partial_t w_{q-1}\|_{L^\infty(I_j)}.$$

If any of the assumptions A1/A2/A3 is satisfied, (3.20) defines a coupled system. From the same lines for obtaining (2.27) in Appendix, the solution for $c^q_{j,k-1}, c^q_{j,k}$ is

$$\begin{aligned} \begin{bmatrix} c^q_{j,k-1} \\ c^q_{j,k} \end{bmatrix} &= -\sum_{m=k-1-2q}^{k-2} \sum_{l=0}^{N-1} r_l A^{-1}(GL^-_m c^q_{j+l,m} + HL^+_m c^q_{j+l+1,m}) \\ &= -\sum_{m=k-1-2q}^{k-2} \Big(c^q_{j,m} V_{1,m} + \sum_{l=0}^{N-1} c^q_{j+l,m} r_l V_{2,m} \Big). \end{aligned} \tag{70}$$

Under assumption A1 and scale invariant flux assumption, (60) implies $\sum_{l=0}^{N-1} \|r_l\|_\infty \le C$. We have the estimate for $c^q_{j,m}, m = k-1, k$, that is

$$\left\| \begin{bmatrix} c^q_{j,k-1} \\ c^q_{j,k} \end{bmatrix} \right\|_\infty \le C(1 + \sum_{l=0}^{N-1} \|r_l\|_\infty) \max |c^q_{j+l,m}| \le Ch^2 \|\partial_t w_{q-1}\|_{L^\infty(\mathcal{I}_N)}.$$

Under assumption A2 or A3, $\sum_{l=0}^{N-1} \|r_l\|_\infty$ is unbounded. Thus we use Fourier analysis to bound the coefficients utilizing the smoothness and periodicity by similar steps of estimating U_j in the proof of Lemma 2.2.3. In the rest of the proof, we make use of two operators $\boxtimes$ and $\boxplus$, which are defined in (35b) and (35a).

When assumption A3 is satisfied, $Q = -A^{-1}B$ has two imaginary eigenvalues λ_1, λ_2 with $|\lambda_1| = |\lambda_2| = 1$. $r_l = \frac{\lambda_1^l}{1-\lambda_1^N} Q_1 + \frac{\lambda_2^l}{1-\lambda_2^N}(I_2 - Q_1)$, where Q_1 is a constant matrix independent of h, and defined in (56) and (57). We perform more detailed computation of the coefficients. In (69), plug in (2.27), for $m = k-3, k-2$, when $u_t \in W^{k+2+n,\infty}(I)$,

$$
\begin{aligned}
c_{j,m}^1 &= i\frac{2m+1}{h}\frac{h^2}{4}\int_{I_j} [L_{j,k-1}, L_{j,k}]\partial_t \sum_{p=k+1}^{\infty} \Big(u_{j,p}V_{1,p} + \sum_{l=0}^{N-1} u_{j+l,p} r_l V_{2,p}\Big) D^{-2} L_{j,m} dx \\
&= i\frac{2m+1}{2}\frac{h^2}{4} \sum_{p=k+1}^{\infty} \partial_t \Big(u_{j,p} F_{p,m}^1 + \sum_{l=0}^{N-1} u_{j+l,p} r_l F_{p,m}^2\Big) \\
&= i\frac{2m+1}{8} \sum_{p=k+1}^{\infty} \Big(\sum_{s=0}^{n} \mu_s h^{k+3+s} u_t^{(k+1+s)}(x_{j-\frac{1}{2}}) F_{p,m}^1 + O(h^{k+n+3}|u_t|_{W^{k+2+n,\infty}(I)}) \\
&+ \sum_{l=0}^{N-1} \big(\frac{\lambda_1^l}{1-\lambda_1^N} Q_1 + \frac{\lambda_2^l}{1-\lambda_2^N}(I_2 - Q_1)\big) \sum_{s=0}^{n} \mu_s h^{k+3+s} u_t^{(k+1+s)}(x_{j+l-\frac{1}{2}}) F_{p,m}^2\Big) \\
&= i\frac{2m+1}{8} \sum_{p=k+1}^{\infty} \sum_{s=0}^{n} \mu_s h^{k+3+s} \big(u_t^{(k+1+s)}(x_{j-\frac{1}{2}}) F_{p,m}^1 \\
&+ (Q_1 \boxtimes_{\lambda_1} + (I_2 - Q_1)\boxtimes_{\lambda_2}) u_t^{(k+1+s)}(x_{j-\frac{1}{2}}) F_{p,m}^2\big) + O(h^{k+3+n}|u_t|_{W^{k+2+n,\infty}(I)}),
\end{aligned}
$$

where $F_{p,m}^\nu = \frac{2}{h}\int_{I_j} [L_{j,k-1}, L_{j,k}] V_{\nu,p} D^{-2} L_{j,m} dx, \nu = 1, 2$, are constants independent of h and (32) is used in the third equality.

Plug the formula above into (70), by similar computation, we have

$$
\begin{bmatrix} c_{j,k-1}^1 \\ c_{j,k}^1 \end{bmatrix} = -i\frac{2m+1}{8} h^2 \sum_{m=k-3}^{k-2} \sum_{p=k+1}^{\infty} \sum_{s=0}^{n} \mu_s h^{k+1+s} \big(u_t^{(k+1+s)}(x_{j-\frac{1}{2}}) F_{p,m}^1 V_{1,m}
$$

$$+ (Q_1 \boxtimes_{\lambda_1} + (I_2 - Q_1)\boxtimes_{\lambda_2}) u_t^{(k+1+s)}(x_{j-\frac{1}{2}})(F^2_{p,m}V_{1,m} + F^1_{p,m}V_{2,m})$$
$$+ (Q_1 \boxtimes_{\lambda_1} + (I_2 - Q_1)\boxtimes_{\lambda_2})^2 u_t^{(k+1+s)}(x_{j-\frac{1}{2}})F^2_{p,m}V_{2,m})$$
$$+ O(h^{k+2+n}|u_t|_{W^{k+2+n,\infty}(I)}).$$

By (39), we have

$$(Q_1 \boxtimes_{\lambda_1} + (I_2 - Q_1)\boxtimes_{\lambda_2})^{\nu} u_t^{(k+1+s)}(x_{j-\frac{1}{2}}) \le C|u_t|_{W^{k+2+s+\nu,1}(I)} \le C|u|_{W^{k+4+s+\nu,1}(I)}.$$

Therefore,

$$|c^1_{j,m}| \le C_2 h^{k+3}, \quad m = k-3, k-2, \ \text{ and } |c^1_{j,m}| \le C_3 h^{k+3}, \quad m = k-1, k.$$

By induction and similar computation, we can obtain the formula for $c^q_{j,m}$. For brevity, we omit the computation and directly show the estimates

$$|c^q_{j,m}| \le C_{3q} h^{k+1+2q}, \quad k-1-2q \le m \le k.$$

When assumption A2 is satisfied, $Q = -A^{-1}B$ has two repeated eigenvalues. $r_l = \frac{(-1)^l}{2} I_2 + (-1)^l \frac{-N+2l}{4\Gamma} Q_2$, where Q_2/Γ is a constant matrix. For $m = k-3, k-2$, when $u_t \in W^{k+2+n,\infty}(I)$, we compute $c^1_{j,m}$ by the same procedure as previous case and obtain

$$c^1_{j,m} = i\frac{2m+1}{8}h^2 \sum_{p=k+1}^{\infty} \sum_{s=0}^{n} \mu_s h^{k+1+s} \big(u_t^{(k+1+s)}(x_{j-\frac{1}{2}})F^1_{p,m}$$
$$+ \frac{1}{2}(\boxtimes_{-1} + \frac{Q_2}{\Gamma}\boxplus) u_t^{(k+1+s)}(x_{j-\frac{1}{2}})F^2_{p,m}\big) + O(h^{k+3+n}|u_t|_{W^{k+2+n,\infty}(I)}).$$

Plug formula above into (70), we have

$$\begin{bmatrix} c^1_{j,k-1} \\ c^1_{j,k} \end{bmatrix} = -i\frac{2m+1}{8}h^2 \sum_{m=k-3}^{k-2}\sum_{p=k+1}^{\infty}\sum_{s=0}^{n}\mu_s h^{k+1+s}\big(u_t^{(k+1+s)}(x_{j-\frac{1}{2}})F^1_{p,m}V_{1,m}$$
$$+\frac{1}{2}(\boxtimes_{-1}+\frac{Q_2}{\Gamma}\boxplus)u_t^{(k+1+s)}(x_{j-\frac{1}{2}})(F^2_{p,m}V_{1,m}+F^1_{p,m}V_{2,m})$$
$$+\frac{1}{4}(\boxtimes_{-1}+\frac{Q_2}{\Gamma}\boxplus)^2u_t^{(k+1+s)}(x_{j-\frac{1}{2}})F^2_{p,m}V_{2,m}\big)+O(h^{k+2+n}|u_t|_{W^{k+2+n,\infty}(I)}).$$

By (39), we have

$$(\boxtimes_{-1}+\frac{Q_2}{\Gamma}\boxplus)^{\nu}u_t^{(k+1+s)}(x_{j-\frac{1}{2}})\le C|u_t|_{W^{k+2+s+2\nu,1}(I)}\le C|u|_{W^{k+4+s+2\nu,1}(I)}$$

and

$$|c^1_{j,m}|\le C_2h^{k+3},\quad m=k-3,k-2,\ \text{ and } |c^1_{j,m}|\le C_4h^{k+3},\quad m=k-1,k.$$

By induction and similar computation, we can obtain the formula for $c^q_{j,m}$. For brevity, we omit the computation and directly show the estimates

$$|c^q_{j,m}|\le C_{4q}h^{k+1+2q},\quad k-1-2q\le m\le k.$$

All the analysis above works when we change definition of w_q to $\partial_t^r w_q$ (and change $(w_{q-1})_t$ to $\partial_t^{r+1}w_{q-1}$ accordingly) in (3.19). Summarize the estimates for $c^q_{j,m}$ under all three assumptions, for $1\le q\le\lfloor\frac{k-1}{2}\rfloor$, we have

$$|\partial_t^r c^q_{j,m}|\le C_{2r,q}h^{k+1+2q},\quad \|\partial_t^r w_q\|\le C(\sum_{j=1}^{N}\sum_{m=k-2q-1}^{k}|\partial_t^r c^q_{j,m}|^2h_j)^{\frac{1}{2}}\le C_{2r,q}h^{k+1+2q}.$$

Then (3.23), (3.24) is proven. And (3.25) is a direct result of above estimate and (3.22). $\square$

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] S. Adjerid, K. D. Devine, J. E. Flaherty, and L. Krivodonova. A posteriori error estimation for discontinuous Galerkin solutions of hyperbolic problems. *Computer Methods in Applied Mechanics and Engineering*, 191(11-12):1097–1112, 2002.

[2] S. Adjerid and T. C. Massey. Superconvergence of discontinuous Galerkin solutions for a nonlinear scalar hyperbolic problem. *Computer Methods in Applied Mechanics and Engineering*, 195(25-28):3331–3346, 2006.

[3] M. Ainsworth, P. Monk, and W. Muniz. Dispersive and dissipative properties of discontinuous Galerkin finite element methods for the second-order wave equation. *Journal of Scientific Computing*, 27(1):5–40, 2006.

[4] B. Alpert. A class of bases in Lˆ2 for the sparse representation of integral operators. *SIAM Journal on Mathematical Analysis*, 24(1):246–262, 1993.

[5] U. M. Ascher, S. J. Ruuth, and R. J. Spiteri. Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations. *Applied Numerical Mathematics*, 25(2-3):151–167, 1997.

[6] J. Bona, H. Chen, O. Karakashian, and Y. Xing. Conservative, discontinuous Galerkin–methods for the generalized Korteweg–de Vries equation. *Mathematics of Computation*, 82(283):1401–1432, 2013.

[7] J. H. Bramble and A. H. Schatz. Higher order local accuracy by averaging in the finite element method. *Mathematics of Computation*, 31(137):94–111, 1977.

[8] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:147–269, 2004.

[9] W. Cao, D. Li, Y. Yang, and Z. Zhang. Superconvergence of discontinuous Galerkin methods based on upwind-biased fluxes for 1d linear hyperbolic equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 51(2):467–486, 2017.

[10] W. Cao, H. Liu, and Z. Zhang. Superconvergence of the direct discontinuous Galerkin method for convection-diffusion equations. *Numerical Methods for Partial Differential Equations*, 33(1):290–317, 2017.

[11] W. Cao, C.-W. Shu, Y. Yang, and Z. Zhang. Superconvergence of discontinuous Galerkin method for scalar nonlinear hyperbolic equations. *SIAM Journal on Numerical Analysis*, 56(2):732–765, 2018.

[12] W. Cao, Z. Zhang, and Q. Zou. Superconvergence of discontinuous Galerkin methods for linear hyperbolic equations. *SIAM Journal on Numerical Analysis*, 52(5):2555–2573, 2014.

[13] O. Cessenat and B. Despres. Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional helmholtz problem. *SIAM Journal on Numerical Analysis*, 35(1):255–299, 1998.

[14] Q. Chang, E. Jia, and W. Sun. Difference schemes for solving the generalized nonlinear Schrödinger equation. *Journal of Computational Physics*, 148(2):397–415, 1999.

[15] A. Chen, Y. Cheng, Y. Liu, and M. Zhang. Superconvergence of ultra-weak discontinuous Galerkin methods for the linear Schrödinger equation in one dimension. *arXiv e-prints*, page arXiv:1905.08161, May 2019.

[16] A. Chen, F. Li, and Y. Cheng. An ultra-weak discontinuous Galerkin method for Schrödinger equation in one dimension. *Journal of Scientific Computing*, 78(2):772–815, Feb 2019.

[17] Y. Cheng, C.-S. Chou, F. Li, and Y. Xing. L^2 stable discontinuous Galerkin methods for one-dimensional two-way wave equations. *Mathematics of Computation*, 86(303):121–155, 2017.

[18] Y. Cheng and C.-W. Shu. A discontinuous Galerkin finite element method for time dependent partial differential equations with higher order derivatives. *Mathematics of Computation*, 77(262):699–730, 2008.

[19] Y. Cheng and C.-W. Shu. Superconvergence of discontinuous Galerkin and local discontinuous Galerkin schemes for linear hyperbolic and convection-diffusion equations in one space dimension. *SIAM Journal on Numerical Analysis*, 47(6):4044–4072, 2010.

[20] P. G. Ciarlet. *The finite element method for elliptic problems*. North-Holland Publishing Co., Amsterdam-New York-Oxford, 1978. Studies in Mathematics and its Applications, Vol. 4.

[21] 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. *Mathematics of Computation*, 54(190):545–581, 1990.

[22] B. Cockburn, S.-Y. Lin, and C.-W. Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws iii: one-dimensional systems. *Journal of Computational Physics*, 84(1):90–113, 1989.

[23] B. Cockburn, M. Luskin, C.-W. Shu, and E. Süli. Enhanced accuracy by post-processing for finite element methods for hyperbolic equations. *Mathematics of Computation*, 72(242):577–606, 2003.

[24] B. Cockburn and C.-W. Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. ii. general framework. *Mathematics of Computation*, 52(186):411–435, 1989.

[25] B. Cockburn and C.-W. Shu. The local discontinuous Galerkin method for time-dependent convection-diffusion systems. *SIAM Journal on Numerical Analysis*, 35(6):2440–2463, 1998.

[26] B. Cockburn and C.-W. Shu. The Runge-Kutta discontinuous Galerkin method for conservation laws v: multidimensional systems. *Journal of Computational Physics*, 141(2):199–224, 1998.

[27] B. Cockburn and C.-W. Shu. Runge-Kutta discontinuous Galerkin methods for convection-dominated problems. *Journal of Scientific Computing*, 16:173–261, 2001.

[28] I. Dag. A quadratic b-spline finite element method for solving nonlinear Schrödinger equation. *Computer Methods in Applied Mechanics and Engineering*, 174(1-2):247–258, 1999.

[29] P. J. Davis. *Circulant matrices*. Wiley, New York, 1979.

[30] J. Douglas and T. Dupont. Interior penalty procedures for elliptic and parabolic Galerkin methods. *Computing Methods in Applied Sciences*, pages 207–216, 1976.

[31] M. Dumbser and M. Käser. An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes—II. The three-dimensional isotropic case. *Geophysical Journal International*, 167(1):319–336, 2006.

[32] V. Gradinaru. Fourier transform on sparse grids: code design and the time dependent Schrödinger equation. *Computing*, 80(1):1–22, 2007.

[33] M. Griebel. A parallelizable and vectorizable multi-level algorithm on sparse grids. In W. Hackbusch, editor, *Parallel Algorithms for Partial Differential Equations, Notes on Numerical Fluid Mechanics*, volume 31, pages 241–251, Braunschweig, 1991. Vieweg-Verlag.

[34] M. Griebel. Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences. *Computing*, 61(2):151–179, 1998.

[35] M. Griebel and J. Hamaekers. Sparse grids for the Schrödinger equation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 41(02):215–247, 2007.

[36] M. Griebel and G. Zumbusch. Adaptive sparse grids for hyperbolic conservation laws. In *Hyperbolic Problems: Theory, Numerics, Applications*, pages 411–422. Springer, 1999.

[37] M. J. Grote, A. Schneebeli, and D. Schötzau. Discontinuous Galerkin finite element method for the wave equation. *SIAM Journal on Numerical Analysis*, 44(6):2408–2431, 2006.

[38] W. Guo and Y. Cheng. A sparse grid discontinuous galerkin method for high-dimensional transport equations and its application to kinetic simulations. *SIAM Journal on Scientific Computing*, 38(6):A3381–A3409, 2016.

[39] W. Guo and Y. Cheng. An adaptive multiresolution discontinuous Galerkin method for time-dependent transport equations in multidimensions. *SIAM Journal on Scientific Computing*, 39(6):A2962–A2992, 2017.

[40] P. Hemker. Sparse-grid finite-volume multigrid for 3D-problems. *Advances in Computational Mathematics*, 4(1):83–110, 1995.

[41] L. Ji, Y. Xu, and J. K. Ryan. Negative-order norm estimates for nonlinear hyperbolic conservation laws. *Journal of Scientific Computing*, 54(2):531–548, Feb 2013.

[42] O. Karakashian and C. Makridakis. A space-time finite element method for the nonlinear Schrödinger equation: the discontinuous Galerkin method. *Mathematics of Computation*, 67(222):479–499, 1998.

[43] O. Karakashian and C. Makridakis. A space-time finite element method for the nonlinear Schrödinger equation: the continuous Galerkin method. *SIAM Journal on Numerical Analysis*, 36(6):1779–1807, 1999.

[44] M. Käser and M. Dumbser. An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes—I. The two-dimensional isotropic case with external source terms. *Geophysical Journal International*, 166(2):855–877, 2006.

[45] A. Kurganov and E. Tadmor. New high-resolution central schemes for nonlinear conservation laws and convection–diffusion equations. *Journal of Computational Physics*, 160(1):241–282, 2000.

[46] R. J. LeVeque. *Finite volume methods for hyperbolic problems*, volume 31. Cambridge university press, 2002.

[47] X. Liang, A. Q. M. Khaliq, and Y. Xing. Fourth order exponential time differencing method with local discontinuous Galerkin approximation for coupled nonlinear Schrödinger equations. *Communications in Computational Physics*, 17(2):510–541, 2015.

[48] Y. Liu. *Central schemes and central discontinuous Galerkin methods on overlapping cells*. PhD thesis, Georgia Institute of Technology, 2005.

[49] Y. Liu. Central schemes on overlapping cells. *Journal of Computational Physics*, 209(1):82–104, 2005.

[50] Y. Liu, C.-W. Shu, E. Tadmor, and M. Zhang. Central discontinuous Galerkin methods on overlapping cells with a nonoscillatory hierarchical reconstruction. *SIAM Journal on Numerical Analysis*, 45(6):2442–2467, 2007.

[51] Y. Liu, C.-W. Shu, E. Tadmor, and M. Zhang. L2 stability analysis of the central discontinuous Galerkin method and a comparison between the central and regular discontinuous Galerkin methods. *ESAIM: Mathematical Modelling and Numerical Analysis*, 42(4):593–607, 2008.

[52] W. Lu, Y. Huang, and H. Liu. Mass preserving discontinuous Galerkin methods for Schrödinger equations. *Journal of Computational Physics*, 282:210–226, 2015.

[53] X. Ma and N. Zabaras. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *Journal of Computational Physics*, 228(8):3084–3113, 2009.

[54] X. Meng and J. K. Ryan. Discontinuous Galerkin methods for nonlinear scalar hyperbolic conservation laws: divided difference estimates and accuracy enhancement. *Numerische Mathematik*, 136(1):27–73, May 2017.

[55] M. S. Mock and P. D. Lax. The computation of discontinuous solutions of linear hyperbolic equations. *Communications on Pure and Applied Mathematics*, 31(4):423–430, 1978.

[56] H. Nessyahu and E. Tadmor. Non-oscillatory central differencing for hyperbolic conservation laws. *Journal of Computational Physics*, 87(2):408–463, 1990.

[57] F. Nobile, R. Tempone, and C. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, 2008.

[58] D. Pathria and J. L. Morris. Pseudo-spectral solution of nonlinear Schrödinger equations. *Journal of Computational Physics*, 87(1):108–125, 1990.

[59] W. H. Reed and T. Hill. Triangular mesh methods for the neutron transport equation. Technical report, Los Alamos Scientific Lab., N. Mex.(USA), 1973.

[60] M. A. Reyna and F. Li. Operator bounds and time step conditions for the DG and central DG methods. *Journal of Scientific Computing*, 62(2):532–554, 2015.

[61] B. Riviere and M. F. Wheeler. Discontinuous finite element methods for acoustic and elastic wave problems. *Contemporary Mathematics*, 329:271–282, 2003.

[62] B. Rivière, M. F. Wheeler, and V. Girault. Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems. part i. *Computational Geosciences*, 3(3):337–360, 1999.

[63] B. Rivière, M. F. Wheeler, and V. Girault. A priori error estimates for finite element methods based on discontinuous approximation spaces for elliptic problems. *SIAM Journal on Numerical Analysis*, 39(3):902–931, 2001.

[64] Z. Rong, J. Shen, and H. Yu. A nodal sparse grid spectral element method for multi-dimensional elliptic partial differential equations. *International Journal of Numerical Analysis and Modeling*, 14(4-5):762–783, 2017.

[65] J. Ryan, C.-W. Shu, and H. Atkins. Extension of a post processing technique for the discontinuous Galerkin method for hyperbolic equations with application to an aeroacoustic problem. *SIAM Journal on Scientific Computing*, 26(3):821–843, 2005.

[66] C. Schwab, E. Süli, and R. Todor. Sparse finite element approximation of high-dimensional transport-dominated diffusion problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 42(05):777–819, 2008.

[67] J. Shen, T. Tang, and L.-L. Wang. *Spectral methods: algorithms, analysis and applications*, volume 41. Springer Science & Business Media, 2011.

[68] J. Shen and L.-L. Wang. Sparse spectral approximations of high-dimensional problems based on hyperbolic cross. *SIAM Journal on Numerical Analysis*, 48(3):1087–1109, 2010.

[69] J. Shen and H. Yu. Efficient spectral sparse grid methods and applications to high-dimensional elliptic problems. *SIAM Journal of Scientific Computing*, 32(6):3228–3250, 2010.

[70] J. Shen and H. Yu. Efficient spectral sparse grid methods and applications to high-dimensional elliptic equations II. unbounded domains. *SIAM Journal of Scientific Computing*, 34(2):A1141–A1164, 2012.

[71] Q. Sheng, A. Khaliq, and E. Al-Said. Solving the generalized nonlinear Schrödinger equation via quartic spline approximation. *Journal of Computational Physics*, 166(2):400–417, 2001.

[72] C.-W. Shu. Discontinuous Galerkin methods for time-dependent convection dominated problems: Basics, recent developments and comparison with other methods. In A. C. G.R. Barrenechea, F. Brezzi and E. Georgoulis, editors, *Building Bridges: Connections*

and Challenges in Modern Approaches to Numerical Partial Differential Equations, volume 114, pages 369–397. Springer, Switzerland, 2016.

[73] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics*, 77:439–471, 1988.

[74] T. R. Taha and M. I. Ablowitz. Analytical and numerical aspects of certain nonlinear evolution equations. ii. Numerical, nonlinear Schrödinger equation. *Journal of Computational Physics*, 55(2):203–230, 1984.

[75] Z. Tao, A. Chen, M. Zhang, and Y. Cheng. Sparse grid central discontinuous galerkin method for linear hyperbolic systems in high dimensions. *SIAM Journal on Scientific Computing*, 41(3):A1626–A1651, 2019.

[76] Z. Wang, Q. Tang, W. Guo, and Y. Cheng. Sparse grid discontinuous Galerkin methods for high-dimensional elliptic equations. *Journal of Computational Physics*, 314:244—263, 2016.

[77] M. F. Wheeler. An elliptic collocation-finite element method with interior penalties. *SIAM Journal on Numerical Analysis*, 15(1):152–161, 1978.

[78] D. Xiu. Efficient collocational approach for parametric uncertainty analysis. *Communications in Computational Physics*, 2(2):293–309, 2007.

[79] D. Xiu and J. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM Journal of Scientific Computing*, 27(3):1118–1139, 2005.

[80] Y. Xu and C.-W. Shu. Local discontinuous Galerkin methods for nonlinear Schrödinger equations. *Journal of Computational Physics*, 205(1):72–97, 2005.

[81] Y. Xu and C.-W. Shu. Optimal error estimates of the semidiscrete local discontinuous Galerkin methods for high order wave equations. *SIAM Journal on Numerical Analysis*, 50(1):79–104, 2012.

[82] Y. Yang and C.-W. Shu. Analysis of optimal superconvergence of discontinuous Galerkin method for linear hyperbolic equations. *SIAM Journal on Numerical Analysis*, 50(6):3110–3133, 2012.

[83] C. Zenger. Sparse grids. In W. Hackbusch, editor, *Parallel Algorithms for Partial Differential Equations, Notes on Numerical Fluid Mechanics*, volume 31, pages 241–251, Braunschweig, 1991. Vieweg-Verlag.

[84] L. Zhou, Y. Xu, Z. Zhang, and W. Cao. Superconvergence of local discontinuous Galerkin method for one-dimensional linear schrödinger equations. *Journal of Scientific Computing*, 73(2-3):1290–1315, 2017.