

A high-order multiscale discontinuous Galerkin method for two-dimensional Schrödinger equation in quantum transport

Bo Dong^{*} and Wei Wang[†]

Abstract

We develop and analyze a high-order multiscale discontinuous Galerkin (DG) method for two-dimensional stationary Schrödinger equations in quantum transport. The solution of the problem under consideration has oscillations mainly in one direction, so we include oscillatory non-polynomial basis functions in that direction and use polynomial basis in the other direction to approximate the solution. We prove that the resulting method converges with an optimal order when the mesh size is sufficiently small. Numerically we observe that the method converges on coarse meshes and achieves optimal higher-order convergence when the mesh size is refined to the scale of the wave length. Numerical results show that the method can capture highly oscillating solutions of Schrödinger equations more effectively than standard DG methods with polynomial basis.

Key words: discontinuous Galerkin method, multiscale method, two-dimensional Schrödinger equation

1 Introduction

We develop, analyze, and numerically test a new multiscale discontinuous Galerkin (DG) methods for the following two-dimensional second-order

^{*}Email: bdong@umassd.edu. Department of Mathematics, University of Massachusetts Dartmouth, North Dartmouth, MA 02747. Research supported by NSF grant DMS-1818998.

[†]E-mail: weiwang1@fiu.edu. Department of Mathematics & Statistics, Florida International University, Miami, FL, 33199.

equation

$$-\varepsilon^2 \Delta u - f(x, y)u = 0, \quad (1.1)$$

where $\varepsilon > 0$ is a small parameter and $f(x, y) > 0$ is a real-valued smooth function. This type of equation has the application to the stationary Schrödinger equation in the modeling of quantum transport in nanoscale semiconductors [14, 5, 15, 19],

$$-\frac{\hbar^2}{2m_e} \Delta \varphi + V(x, y)\varphi = E\varphi, \quad (1.2)$$

where $\hbar$ is the reduced Plank constant, m_e is the effective mass, $V(x, y)$ is the total electrostatic potential in the device, E is the injection energy, and the solution $\varphi(x, y)$ is the complex-valued wave function. By defining $\varepsilon = \frac{\hbar}{\sqrt{2m_e}}$, $f(x, y) = E - V(x, y)$, the Schrödinger equation (1.2) is in the form of model equation (1.1), in which ε describes the microscopic/macroscopic scale ratio. In the case of constant f , a particular solution to the Schrödinger equation (1.2) is given by a single plane wave, $u = \exp(\frac{i}{\varepsilon} \xi \cdot \mathbf{x})$, for any given wave vector ξ with $|\xi| = \sqrt{f}$ and direction vector $\mathbf{x} \in \mathbb{R}^2$. It is clear that the solution u features oscillations with frequency $1/\varepsilon$. The highly oscillatory nature of the solutions poses a huge challenge to the traditional numerical methods because they require extremely refined meshes and thus tremendous amount of computer memory and CPU time to capture such high frequency oscillations. Thus, we are interested in developing multiscale methods which incorporate the multiscale information of the problem into the finite element spaces to solve for highly oscillatory solutions on coarse meshes.

Multiscale DG methods with non-polynomial basis functions have been developed and studied in the literature, including [1, 20, 12, 21, 7, 13, 19, 18, 22]. Previous work of multiscale methods in solving stationary Schrödinger equations (mainly for one dimension) include multiscale finite element methods [16, 17, 5, 4, 15, 3] and multiscale discontinuous Galerkin method [19, 9, 10] etc. In [19], Wang and Shu first proposed a third-order multiscale DG method with an exponential space E^2 based on WKB approximation for one-dimensional Schrödinger equation. In [10], we extended the method to arbitrarily higher order multiscale DG method with the exponential space E^k for any positive integer k . In [14], a local DG method was developed for two-dimensional Schrödinger equation by using the E^2 space in one direction and the polynomial space in the other direction. Numerically the method was observed to have the third-order convergence, but error analysis was not available.

It is challenging to construct high order multiscale DG methods for general Schrödinger equations in two-dimensional space. In the one-dimensional case, the solution of Schrödinger equation oscillates only in one direction and WKB asymptotics provides a good approximation to the solution. In the general two-dimensional setting, Schrödinger equation can have solutions that oscillate in any direction, and WKB approximation is not available for multidimensional case. Therefore, it is difficult to incorporate the oscillatory nature of the solution into the two-dimensional basis functions. In this work, we consider the case that the solution of Schrödinger equation mainly oscillates in one direction due to boundary conditions and extend our one-dimensional high-order multiscale DG methods in [10] to two-dimensional space. This will serve as a stepping stone for the development of efficient multiscale finite element methods for more general two-dimensional Schrödinger equations.

In this paper, we take the space of approximate solutions as tensor product of functions in each dimension. In the direction that the solution has stronger oscillation, we use the one-dimensional approximation space E^k in [10] which incorporates the fine scale information of the problem in exponential basis functions. In the other direction that has less or no oscillation, we use standard polynomial basis. The numerical traces in our DG formulation have penalty terms with complex coefficients. When the penalty parameters are taken to be zero, the numerical traces are the same as the minimal dissipation local discontinuous Galerkin (MD-LDG) method in [8]. When they are nonzero, our method is different from local DG methods and the pure imaginary penalty terms in the numerical traces allow us to estimate the jumps of errors at element interfaces using an energy argument. Numerical results in [9] show that penalty terms also help reduce resonance errors. To get estimates on errors in the interior of elements, we consider a dual problem and prove that the method has optimal high-order convergence when the mesh size is sufficiently small using a duality argument. Numerical experiments show that our method converges on coarse meshes where standard DG methods do not. When the meshes are refined, the errors of our method are several magnitudes smaller than standard DG methods using the same number of basis functions. In the application of solving Schrödinger equation from quantum transport, our method can accurately capture oscillatory solution on coarse meshes where standard DG methods produce spurious waves.

The rest of the paper is organized as follows. In Section 2, we introduce

our model problem and define the multiscale DG method. In Section 3, we state and prove the error estimates. Numerical results are shown in Section 4. Finally, we conclude in Section 5.

2 Multiscale DG method: The Methodology

We consider the following 2D problem

$$\begin{cases} -\varepsilon^2 \Delta u - fu = 0, & \text{in } \Omega, \\ u = u_D & \text{on } \Gamma_D \subseteq \partial\Omega, \\ \frac{\partial u}{\partial \mathbf{n}} - i\omega u = 2i\omega g & \text{on } \Gamma_N = \partial\Omega \setminus \Gamma_D, \end{cases} \quad (2.1)$$

where Ω is a rectangular domain, $f = f(x, y) > 0$, ω is a positive number, i is the imaginary unit, and u_D and g are given boundary functions. This type of problems have applications in modeling the 2D quantum transport phenomena in semiconductor devices, such as resonant tunneling diode [6], nanoscale metal–oxide–semiconductor field-effect transistor (MOSFET) [16, 17, 4], and quantum directional coupler [14].

To define our multiscale DG method for the model problem (2.1), we rewrite it as a system of first order equations

$$\mathbf{q} - \varepsilon \nabla u = 0, \quad (2.2a)$$

$$-\varepsilon \nabla \cdot \mathbf{q} - fu = 0 \quad (2.2b)$$

with the boundary conditions

$$u = u_D \quad \text{on } \Gamma_D, \quad (2.2c)$$

$$\mathbf{q} \cdot \mathbf{n} - i\varepsilon\omega u = 2i\varepsilon\omega g \quad \text{on } \Gamma_N. \quad (2.2d)$$

We partition the domain using rectangular mesh, letting $\Omega_h = \{K_{ij} = I_i \times J_j | I_i = (x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}), J_j = (y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}), i = 1, \dots, N_x, j = 1, \dots, N_y\}$, $\partial\Omega_h := \{\partial K_{ij} | K_{ij} \in \Omega_h\}$ be the collection of the boundaries of all elements, $\mathcal{E}_h$ be the set of all faces, and $\mathcal{E}_h^i := \mathcal{E}_h \setminus \partial\Omega$ be the set of all interior faces. We use the notation $x_i = \frac{1}{2}(x_{i-\frac{1}{2}} + x_{i+\frac{1}{2}})$, $y_j = \frac{1}{2}(y_{j-\frac{1}{2}} + y_{j+\frac{1}{2}})$, $\Delta x_i = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$, $\Delta y_j = y_{j+\frac{1}{2}} - y_{j-\frac{1}{2}}$, $\Delta x = \max_{i=1, \dots, N_x} \Delta x_i$, $\Delta y = \max_{j=1, \dots, N_y} \Delta y_j$, and $h = \max\{\Delta x, \Delta y\}$.

Next, we introduce the finite element spaces and the weak formulation of the approximate solutions, as well as the numerical traces that appear in the

weak formulation. For the multiscale finite element spaces in two-dimensional space, since the solution oscillates mainly in x -direction due to the boundary condition, we use the non-polynomial basis E^k [10] in x -direction and the polynomial basis P^k in y -direction. We define

$$M^k := \{v \in L^2(\Omega_h) \mid v|_{K_{ij}} \in E^k|_{I_i} \otimes P^k|_{I_j}, \forall K_{ij} \in \Omega_h\},$$

and

$$\mathbf{M}^k := \{\mathbf{w} = (w_1, w_2) \mid w_i \in M^k, i = 1, 2\},$$

where

$$E^k|_{I_i} = \begin{cases} \text{span}\{e^{i\omega(x-x_i)}, e^{-i\omega(x-x_i)}\}, & \text{if } k = 1, \\ \text{span}\{e^{i\omega(x-x_i)}, e^{-i\omega(x-x_i)}, 1, x, \dots, x^{k-2}\}, & \text{if } k \geq 2, \end{cases}$$

$$\text{and } P^k|_{J_j} = \text{span}\{1, \frac{y-y_j}{\Delta y_j}, \dots, (\frac{y-y_j}{\Delta y_j})^k\}.$$

Similar to our previous work in one dimension [10], the constant ω in the exponent of the basis of E^k is from the boundary condition in Eq. (2.1), and it is related to the wave frequency of the model problem. We remark that if the solution mainly oscillates in y -direction, the space is defined similarly with E^k in y -direction and P^k in x -direction.

The weak formulation of our DG methods for Eq. (2.2) is to find approximate solutions $(u_h, \mathbf{q}_h) \in M^k \times \mathbf{M}^k$ such that

$$(\mathbf{q}_h, \mathbf{w})_{\Omega_h} + (\varepsilon u_h, \nabla \cdot \mathbf{w})_{\Omega_h} - \langle \varepsilon \widehat{u}_h, \mathbf{w} \cdot \mathbf{n} \rangle_{\partial \Omega_h} = 0 \quad (2.3a)$$

$$(\varepsilon \mathbf{q}_h, \nabla v)_{\Omega_h} - \langle \varepsilon \widehat{\mathbf{q}}_h, v \mathbf{n} \rangle_{\partial \Omega_h} - (f u_h, v)_{\Omega_h} = 0 \quad (2.3b)$$

for any test functions $v \in M^k$ and $\mathbf{w} \in \mathbf{M}^k$. Here, we have used the notation

$$\begin{aligned} (\varphi, v)_{\Omega_h} &= \sum_{i,j} \int_{K_{i,j}} \varphi \bar{v} \, dx dy, & (\boldsymbol{\psi}, \mathbf{w})_{\Omega_h} &= \sum_{i,j} \int_{K_{i,j}} \boldsymbol{\psi} \cdot \bar{\mathbf{w}} \, dx dy, \\ \langle \varphi, v \rangle_{\partial \Omega_h} &= \sum_{i,j} \int_{\partial K_{i,j}} \varphi \bar{v} \, ds, & \langle \boldsymbol{\psi}, \mathbf{w} \rangle_{\partial \Omega_h} &= \sum_{i,j} \int_{\partial K_{i,j}} \boldsymbol{\psi} \cdot \bar{\mathbf{w}} \, ds, \end{aligned}$$

where $\bar{v}$ is the complex conjugate of any function v and $\mathbf{n}$ is the unit outward normal vector.

The choice of numerical traces plays a key role for the definition of DG methods, and different numerical traces lead to different DG methods [2]. To

define our numerical traces, let us introduce some notation. At any interior element interface $e \in \mathcal{E}_h^i$ shared by two elements K_1 and K_2 , the average and jump of a scalar function v are given by

$$\{v\} = \frac{1}{2}(v_1 + v_2), \quad \llbracket v \mathbf{n} \rrbracket = v_1 \mathbf{n}_1 + v_2 \mathbf{n}_2,$$

where $v_i = v|_{\partial K_i}$ and $\mathbf{n}_i$ is the unit normal vector on e pointing exterior to K_i for $i = 1, 2$. Similarly, for a vector-valued function $\mathbf{w}$, we set

$$\{\mathbf{w}\} = \frac{1}{2}(\mathbf{w}_1 + \mathbf{w}_2), \quad \llbracket \mathbf{w} \cdot \mathbf{n} \rrbracket = \mathbf{w}_1 \cdot \mathbf{n}_1 + \mathbf{w}_2 \cdot \mathbf{n}_2.$$

At any boundary edge $e \in \partial\Omega$, we simply let

$$\llbracket v \mathbf{n} \rrbracket = v \mathbf{n}, \quad \llbracket \mathbf{w} \cdot \mathbf{n} \rrbracket = \mathbf{w} \cdot \mathbf{n}.$$

In our schemes, we define the numerical traces as follows. At an interior element interface $e \in \mathcal{E}_h^i$,

$$\widehat{u}_h = u_h^- - \mathbf{i} \beta \llbracket \mathbf{q}_h \cdot \mathbf{n} \rrbracket, \quad (2.4a)$$

$$\widehat{\mathbf{q}}_h = \mathbf{q}_h^+ + \mathbf{i} \alpha \llbracket u_h \mathbf{n} \rrbracket, \quad (2.4b)$$

where

$$\begin{aligned} v^- &= \{v\} + \boldsymbol{\rho}_0 \cdot \llbracket v \mathbf{n} \rrbracket, & v^+ &= \{v\} - \boldsymbol{\rho}_0 \cdot \llbracket v \mathbf{n} \rrbracket, \\ \mathbf{w}^- &= \{\mathbf{w}\} + \boldsymbol{\rho}_0 \llbracket \mathbf{w} \cdot \mathbf{n} \rrbracket, & \mathbf{w}^+ &= \{\mathbf{w}\} - \boldsymbol{\rho}_0 \llbracket \mathbf{w} \cdot \mathbf{n} \rrbracket, \end{aligned}$$

where $\boldsymbol{\rho}_0$ is any function on $\mathcal{E}_h$ such that for any point $e \in \partial K \cap \mathcal{E}_h$,

$$\boldsymbol{\rho}_0 \cdot \mathbf{n}_K(e) = \frac{1}{2} \text{sign}(\mathbf{w}_0 \cdot \mathbf{n}_K(e)),$$

where $\mathbf{w}_0$ is any nonzero piecewise constant vector in $\mathbf{H}(\text{div}, \Omega)$. In analysis, we assume that the penalty parameters β and α are positive constants, which allows us to carry out error analysis in a way similar to [11, 9]. In implementation, these penalty parameters can be taken as zero, and the numerical traces will be the same as those for the standard MD-LDG method in [8]. We remark that we can recover the scheme in [14] by taking $k = 2$ and zero penalty parameters in our method.

On the domain boundary Γ_D , we define the numerical traces by

$$\langle \widehat{u}_h, v \rangle_{\Gamma_D} = \langle u_D, v \rangle_{\Gamma_D}, \quad (2.5a)$$

$$\widehat{\mathbf{q}}_h = \mathbf{q}_h + \mathbf{i} \theta (u_h - \widehat{u}_h) \mathbf{n} \quad \text{on } \Gamma_D, \quad (2.5b)$$

where the penalty parameter $\theta > 0$ on $\Gamma_D^+ := \{e \in \Gamma_D : \mathbf{w}_0 \cdot \mathbf{n}_K(e) > 0\}$ for MD-LDG in [8]. On the boundary Γ_N , we let

$$\langle \widehat{u}_h, v \rangle_{\Gamma_N} = \langle u_h + \mathbf{i} \frac{\gamma}{\varepsilon \omega} \mathbf{R}_h \cdot \mathbf{n}, v \rangle_{\Gamma_N}, \quad (2.5c)$$

$$\langle \widehat{\mathbf{q}}_h, \mathbf{w} \rangle_{\Gamma_N} = \langle \mathbf{q}_h - (1 - \gamma) \mathbf{R}_h, \mathbf{w} \rangle_{\Gamma_N}, \quad (2.5d)$$

where γ can be any real constant in $(0, 1)$ and $\mathbf{R}_h := \mathbf{q}_h - \mathbf{i} \varepsilon \omega (u_h + 2g) \mathbf{n}$. Note that the numerical traces on Γ_N are defined so that they satisfy the following equation,

$$\langle \widehat{\mathbf{q}}_h \cdot \mathbf{n} - \mathbf{i} \varepsilon \omega \widehat{u}_h, v \rangle_{\Gamma_N} = \langle 2\mathbf{i} \varepsilon \omega g, v \rangle_{\Gamma_N} \quad \text{for any } v \in V_h,$$

which is similar to the boundary condition (2.2d) for the exact solutions.

3 Error estimates

In the rest of the paper, we use the notation $\|\cdot\|_{s,D}$ for $H^s(D)$ -norm. We drop the first subindex if $s = 0$, and the second subindex if $D = \Omega$ or Ω_h . To carry out error analysis, we start with introducing some projections.

3.1 Projection

We first define a local projection Π for scalar-valued functions, which is the tensor product of L^2 -projections in each direction. For any element $K_{ij} = I_i \times J_j \in \Omega_h$,

$$\Pi = \Pi_x \otimes \Pi_y : H^s(K_{ij}) \rightarrow E^k|_{I_i} \otimes P^k|_{J_j},$$

where Π_x and Π_y are the L^2 -orthogonal projections onto $E^k|_{I_i}$ and $P^k|_{J_j}$, respectively. Similarly, we can define a L^2 -projection for vector-valued functions. For any $\boldsymbol{\rho} = (\rho_1, \rho_2) \in [H^s(K_{ij})]^2$, we let

$$\mathbf{\Pi} \boldsymbol{\rho} = (\Pi \rho_1, \Pi \rho_2).$$

It is easy to see that the projection Π and $\mathbf{\Pi}$ have the following orthogonality properties: for any $\xi \in H^s(K_{ij})$ and $\boldsymbol{\rho} \in [H^s(K_{ij})]^2$,

$$(\xi - \Pi \xi, v)_{K_{ij}} = 0 \quad \text{for any } v \in E^k|_{I_i} \otimes P^k|_{J_j}, \quad (3.1a)$$

$$(\boldsymbol{\rho} - \mathbf{\Pi} \boldsymbol{\rho}, \mathbf{w})_{K_{ij}} = 0 \quad \text{for any } \mathbf{w} \in [E^k|_{I_i} \otimes P^k|_{J_j}]^2. \quad (3.1b)$$

For the L^2 -projections Π and $\mathbf{\Pi}$, we have the following approximation properties.

Lemma 3.1. *Let $\xi \in H^{s+1}(K)$ and $\boldsymbol{\rho} \in [H^{s+1}(K)]^2$, $s \geq 0$, for any $K \in \Omega_h$. If h is sufficiently small, we have*

$$\begin{aligned}\|\xi - \Pi\xi\|_K &\leq Ch^{\min\{s,k\}+1}\|\xi\|_{s+1,I}, \\ \|\xi - \Pi\xi\|_e &\leq Ch^{\min\{s,k\}+\frac{1}{2}}\|\xi\|_{s+1,I}, \quad \forall e \in \partial K, \\ \|\boldsymbol{\rho} - \Pi\boldsymbol{\rho}\|_K &\leq Ch^{\min\{s,k\}+1}\|\boldsymbol{\rho}\|_{s+1,I}, \\ \|\boldsymbol{\rho} - \Pi\boldsymbol{\rho}\|_e &\leq Ch^{\min\{s,k\}+\frac{1}{2}}\|\boldsymbol{\rho}\|_{s+1,I}, \quad \forall e \in \partial K,\end{aligned}$$

where C is independent of h .

The lemma above can be proved in the same way as the approximation estimates of Π_x in Lemma 3.2 in [10], which is based on Taylor expansions of basis functions for sufficiently small h . To avoid redundancy, we omit the details here.

3.2 Error Equations

For ease of presentation, we use the following notations. We let

$$\begin{aligned}e_u &= u - u_h, & \delta_u &= u - \Pi u, & \eta_u &= \Pi u - u_h, \\ e_{\mathbf{q}} &= \mathbf{q} - \mathbf{q}_h, & \delta_{\mathbf{q}} &= \mathbf{q} - \Pi \mathbf{q}, & \eta_{\mathbf{q}} &= \Pi \mathbf{q} - \mathbf{q}_h.\end{aligned}$$

Then

$$e_u = \delta_u + \eta_u, \quad e_{\mathbf{q}} = \delta_{\mathbf{q}} + \eta_{\mathbf{q}}.$$

It is easy to see that the exact solution $(u, \mathbf{q})$ and the approximate solution $(u_h, \mathbf{q}_h)$ both satisfy the weak formulation (2.3). Thus, using the orthogonality properties of the projections in (3.1), we get the following error equations

$$(\eta_{\mathbf{q}}, \mathbf{w}) + \varepsilon(\eta_u, \nabla \cdot \mathbf{w}) - \varepsilon\langle \widehat{e}_u, \mathbf{w} \cdot \mathbf{n} \rangle = 0, \quad (3.2a)$$

$$\varepsilon(\eta_{\mathbf{q}}, \nabla v) - \varepsilon\langle \widehat{e}_{\mathbf{q}}, v \mathbf{n} \rangle - (fe_u, v) = 0. \quad (3.2b)$$

Here, for simplicity, we have dropped the subindex if an integral is over Ω_h or $\partial\Omega_h$.

We also have the following lemma on the errors of numerical traces at interior edges and boundary edges.

Lemma 3.2. *On any interior face, we have*

$$\widehat{e}_\omega = \widehat{\delta}_\omega + \widehat{\eta}_\omega \quad \text{for } \omega = u, \mathbf{q},$$

where

$$\begin{aligned}\widehat{\delta}_u &= \delta_u^- - \mathbf{i}\beta\llbracket\delta_{\mathbf{q}} \cdot \mathbf{n}\rrbracket, & \widehat{\eta}_u &= \eta_u^- - \mathbf{i}\beta\llbracket\eta_{\mathbf{q}} \cdot \mathbf{n}\rrbracket, \\ \widehat{\delta}_{\mathbf{q}} &= \delta_{\mathbf{q}}^+ + \mathbf{i}\alpha\llbracket\delta_u \mathbf{n}\rrbracket, & \widehat{\eta}_{\mathbf{q}} &= \eta_{\mathbf{q}}^+ + \mathbf{i}\alpha\llbracket\eta_u \mathbf{n}\rrbracket.\end{aligned}$$

On the boundary, we have

$$\widehat{e}_u = \begin{cases} 0 & \text{on } \Gamma_D \\ e_u - \mathbf{i}\frac{\gamma}{\varepsilon\omega}e_{\mathbf{R}} \cdot \mathbf{n} & \text{on } \Gamma_N \end{cases} \quad \text{and} \quad \widehat{e}_{\mathbf{q}} = \begin{cases} e_{\mathbf{q}} + \mathbf{i}\theta e_u \mathbf{n} & \text{on } \Gamma_D \\ e_{\mathbf{q}} - (1 - \gamma)e_{\mathbf{R}} & \text{on } \Gamma_N \end{cases},$$

where $e_{\mathbf{R}} = e_{\mathbf{q}} - \mathbf{i}\varepsilon\omega e_u \mathbf{n}$.

The proof of Lemma 3.2 is just straight forward calculations by using the definition of numerical traces in (2.4) and (2.5).

3.3 Error estimates

Now let us state our main results and we postpone the proofs to the next subsection. We first have the following theorem on the estimates of the projections of the errors at the element interfaces.

Theorem 3.3. *Suppose that $(u, \mathbf{q})$ is the exact solution to the problem (2.2) and $(u_h, \mathbf{q}_h)$ is the solution of the multiscale DG method defined by (2.3)-(2.5). When $\alpha > 0$, $\beta > 0$, $\theta > 0$, $0 < \gamma < 1$, $0 < C_0 < f < C_1$ for some constant C_0 and C_1 , and h is sufficiently small, we have*

$$\|\llbracket\eta_u \mathbf{n}\rrbracket\|_{\mathcal{E}_h}^2 + \|\llbracket\eta_{\mathbf{q}} \cdot \mathbf{n}\rrbracket\|_{\mathcal{E}_h^i \cup \Gamma_N}^2 \leq Ch^{k+1}\|\eta_u\| + Ch^{2k+1}(\|u\|_{k+1}^2 + \|q\|_{k+1}^2),$$

where C is independent of h but may depend on ε .

We would like to point out that the conditions on the penalty parameters $\alpha, \beta, \theta, \gamma$ in Theorem 3.3 are for the purpose of error analysis. Our numerical experiments in Section 4 show that the method still works well when $\alpha = \beta = \gamma = 0$. We would also like to mention that the error estimate above is only for h small enough. But our numerical results show that the method has at least the second-order convergence on coarse meshes and higher order convergence on fine meshes.

For the projections of errors in the interior of the domain, we have the following optimal convergence result.

Theorem 3.4. *Under the assumption of Theorem 3.3, we have*

$$\|\eta_u\| \leq Ch^{k+1}(\|u\|_{k+1} + \|\mathbf{q}\|_{k+1}),$$

where C is independent of h but may depend on ε .

Using Theorem 3.4, the approximation property of the projection Π in Lemma 3.1, and triangle inequality, we get the following error estimate.

Theorem 3.5. *Under the assumption of Theorem 3.3, where h is sufficiently small, we have*

$$\|u - u_h\| \leq Ch^{k+1}(\|u\|_{k+1} + \|\mathbf{q}\|_{k+1}),$$

where C is independent of h but may depend on ε .

3.4 Proofs

In this subsection, we prove the error estimates in Theorem 3.3 and Theorem 3.4. We first use an energy argument to get a Gårding type equality on the projection of errors in Lemma 3.6. Since the second-order equation in (2.1) is not coercive, we will not be able to estimate the energy in the interior from the Gårding's equality. But the jump coefficients in $\hat{u}_h$ and $\hat{q}_h$ in (2.4) are chosen to be purely imaginary, so we can separate the real and imaginary parts of the equality and estimate the jump terms at cell interfaces from the imaginary part to prove Theorem 3.3. This approach has been used in [11, 7] for Helmholtz equation. Next, using properties of the two-dimensional projections and a duality argument, we estimate the projection of the errors in the interior of the domain for the proof of Theorem 3.4.

Let us start with showing the following lemma using an energy argument.

Lemma 3.6. *We have that*

$$L = R_1 + R_2,$$

where

$$\begin{aligned}
L &= \|\eta_q\|^2 - (|\eta_u|^2, f) + \mathbf{i}\beta\varepsilon\|\llbracket\eta_q \cdot \mathbf{n}\rrbracket\|_{\mathcal{E}_h^i}^2 + \mathbf{i}\alpha\varepsilon\|\llbracket\eta_u \mathbf{n}\rrbracket\|_{\mathcal{E}_h^i}^2 + \mathbf{i}\theta\varepsilon\|\eta_u\|_{\Gamma_D}^2 \\
&\quad + \mathbf{i}\frac{\gamma}{\omega}\|\eta_q \cdot \mathbf{n}\|_{\Gamma_N}^2 + \mathbf{i}(1-\gamma)\varepsilon^2\omega\|\eta_u\|_{\Gamma_N}^2 \\
R_1 &= (\eta_u, f\delta_u) \\
R_2 &= \varepsilon\langle\delta_u^- - \mathbf{i}\beta\llbracket\delta_q \cdot \mathbf{n}\rrbracket, \llbracket\eta_q \cdot \mathbf{n}\rrbracket\rangle_{\mathcal{E}_h^i} + \varepsilon\langle\llbracket\eta_u \mathbf{n}\rrbracket, \delta_q^+ + \mathbf{i}\alpha\llbracket\delta_u \mathbf{n}\rrbracket\rangle_{\mathcal{E}_h^i} \\
&\quad + \varepsilon\langle\eta_u, \delta_q \cdot \mathbf{n}\rangle_{\Gamma_D} - \mathbf{i}\theta\varepsilon\langle\eta_u, \delta_u\rangle_{\Gamma_D} + \varepsilon\langle(1-\gamma)\delta_u - \mathbf{i}\frac{\gamma}{\varepsilon\omega}\delta_q \cdot \mathbf{n}, \eta_q \cdot \mathbf{n}\rangle_{\Gamma_N} \\
&\quad + \varepsilon\langle\eta_u, \gamma\delta_q \cdot \mathbf{n} + \mathbf{i}(1-\gamma)\varepsilon\omega\delta_u\rangle_{\Gamma_N}.
\end{aligned}$$

Proof. Taking the complex conjugate of the error equation (3.2b), we have

$$\varepsilon(\nabla v, \eta_q) - \varepsilon\langle v\mathbf{n}, \widehat{e}_q \rangle - (v, fe_u) = 0. \quad (3.3)$$

Taking $\mathbf{w} = \eta_q$ in (3.2a) and $v = \eta_u$ in (3.3) and adding these two equations together, we get

$$\begin{aligned}
0 &= (\eta_q, \eta_q) + \varepsilon(\eta_u, \nabla \cdot \eta_q) - \varepsilon\langle \widehat{e}_u, \eta_q \cdot \mathbf{n} \rangle \\
&\quad + \varepsilon(\nabla \eta_u, \eta_q) - \langle \eta_u, \widehat{e}_q \cdot \mathbf{n} \rangle - (\eta_u, fe_u).
\end{aligned}$$

Using integration by parts, we have

$$0 = \|\eta_q\|^2 + \Theta - (\eta_u, f\delta_u) - (|\eta_u|^2, f), \quad (3.4)$$

where

$$\Theta = \varepsilon\langle \eta_u, \eta_q \cdot \mathbf{n} \rangle - \varepsilon\langle \widehat{e}_u, \eta_q \cdot \mathbf{n} \rangle - \varepsilon\langle \eta_u, \widehat{e}_q \cdot \mathbf{n} \rangle.$$

Using Lemma 3.2 to rewrite $\widehat{e}_u$ and $\widehat{e}_q$ in Θ , we have

$$\Theta = \Theta|_{\partial\Omega_h \setminus \partial\Omega} + \Theta|_{\partial\Omega} = A_1 + A_2 + A_3 + A_4,$$

where

$$\begin{aligned}
A_1 &= \varepsilon\langle \eta_u, \eta_q \cdot \mathbf{n} \rangle_{\partial\Omega_h \setminus \partial\Omega} - \varepsilon\langle \eta_u^-, \llbracket\eta_q \cdot \mathbf{n}\rrbracket \rangle_{\mathcal{E}_h^i} - \varepsilon\langle \llbracket\eta_u \mathbf{n}\rrbracket, \eta_q^+ \rangle_{\mathcal{E}_h^i} \\
A_2 &= -\varepsilon\langle \delta_u^-, \llbracket\eta_q \cdot \mathbf{n}\rrbracket \rangle_{\mathcal{E}_h^i} - \varepsilon\langle \llbracket\eta_u \mathbf{n}\rrbracket, \delta_q^+ \rangle_{\mathcal{E}_h^i} \\
A_3 &= \mathbf{i}\beta\varepsilon\langle \llbracket\delta_q \cdot \mathbf{n}\rrbracket + \llbracket\eta_q \cdot \mathbf{n}\rrbracket, \llbracket\eta_q \cdot \mathbf{n}\rrbracket \rangle_{\mathcal{E}_h^i} + \mathbf{i}\alpha\varepsilon\langle \llbracket\eta_u \mathbf{n}\rrbracket, \llbracket\delta_u \mathbf{n}\rrbracket + \llbracket\eta_u \mathbf{n}\rrbracket \rangle_{\mathcal{E}_h^i} \\
A_4 &= \varepsilon\langle \eta_u, \eta_q \cdot \mathbf{n} \rangle_{\partial\Omega} - \varepsilon\langle \widehat{e}_u, \eta_q \cdot \mathbf{n} \rangle_{\partial\Omega} - \varepsilon\langle \eta_u, \widehat{e}_q \cdot \mathbf{n} \rangle_{\partial\Omega}.
\end{aligned}$$

It is easy to check that

$$A_1 = 0$$

and

$$A_3 = \mathbf{i}\beta\varepsilon\langle\llbracket\delta_{\mathbf{q}}\cdot\mathbf{n}\rrbracket, \llbracket\eta_{\mathbf{q}}\cdot\mathbf{n}\rrbracket\rangle_{\mathcal{E}_h^i} + \mathbf{i}\beta\varepsilon|\llbracket\eta_{\mathbf{q}}\cdot\mathbf{n}\rrbracket|_{\mathcal{E}_h^i}^2 + \mathbf{i}\alpha\varepsilon\langle\llbracket\eta_u\mathbf{n}\rrbracket, \llbracket\delta_u\mathbf{n}\rrbracket\rangle_{\mathcal{E}_h^i} + \mathbf{i}\alpha\varepsilon|\llbracket\eta_u\mathbf{n}\rrbracket|_{\mathcal{E}_h^i}^2.$$

Using Lemma 3.2, we get

$$\begin{aligned} A_4 = & \varepsilon\langle\eta_u, \eta_{\mathbf{q}}\cdot\mathbf{n}\rangle_{\partial\Omega} - \varepsilon\langle e_u - \mathbf{i}\frac{\gamma}{\varepsilon\omega}e_{\mathbf{R}}, \eta_{\mathbf{q}}\cdot\mathbf{n}\rangle_{\Gamma_N} \\ & - \varepsilon\langle\eta_u, (e_{\mathbf{q}} + \mathbf{i}\theta e_u\mathbf{n})\cdot\mathbf{n}\rangle_{\Gamma_D} - \varepsilon\langle\eta_u, (e_{\mathbf{q}} - (1-\gamma)e_{\mathbf{R}})\cdot\mathbf{n}\rangle_{\Gamma_N} \end{aligned}$$

Since $e_w = \delta_w + \eta_w$ for $w = u, \mathbf{q}$,

$$\begin{aligned} A_4 = & -\varepsilon\langle\delta_u - \mathbf{i}\frac{\gamma}{\varepsilon\omega}e_{\mathbf{R}}, \eta_{\mathbf{q}}\cdot\mathbf{n}\rangle_{\Gamma_N} - \varepsilon\langle\eta_u, \delta_{\mathbf{q}}\cdot\mathbf{n} + \mathbf{i}\theta(\delta_u + \eta_u)\rangle_{\Gamma_D} \\ & - \varepsilon\langle\eta_u, (\delta_{\mathbf{q}} + \eta_{\mathbf{q}})\cdot\mathbf{n} - (1-\gamma)e_{\mathbf{R}}\rangle_{\Gamma_N}. \end{aligned}$$

Note that

$$e_{\mathbf{R}} = e_{\mathbf{q}}\cdot\mathbf{n} - \mathbf{i}\varepsilon\omega e_u\mathbf{n} = (\delta_{\mathbf{q}} + \eta_{\mathbf{q}})\cdot\mathbf{n} - \mathbf{i}\varepsilon\omega(\delta_u + \eta_u)\mathbf{n}.$$

After some calculations we get

$$\begin{aligned} A_4 = & -\varepsilon\langle(1-\gamma)\delta_u - \mathbf{i}\frac{\gamma}{\varepsilon\omega}\delta_{\mathbf{q}}\cdot\mathbf{n}, \eta_{\mathbf{q}}\cdot\mathbf{n}\rangle_{\Gamma_N} + \mathbf{i}\frac{\gamma}{\omega}\|\eta_{\mathbf{q}}\cdot\mathbf{n}\|_{\Gamma_N}^2 \\ & - \varepsilon\langle\eta_u, \delta_{\mathbf{q}}\cdot\mathbf{n}\rangle_{\Gamma_D} - \varepsilon\langle\eta_u, \mathbf{i}\theta\delta_u\rangle_{\Gamma_D} + \mathbf{i}\varepsilon\theta\|\eta_u\|_{\Gamma_D}^2 \\ & - \varepsilon\langle\eta_u, \gamma\delta_{\mathbf{q}}\cdot\mathbf{n} + \mathbf{i}(1-\gamma)\varepsilon\omega\delta_u\rangle_{\Gamma_N} + \mathbf{i}(1-\gamma)\varepsilon^2\omega\|\eta_u\|_{\Gamma_N}^2 \end{aligned}$$

Plugging A_2, A_3 and A_4 into Θ in (3.4), we get the conclusion. $\square$

Next we prove Theorem 3.3.

Proof. Taking the imaginary part of L in Lemma 3.6, we get

$$\begin{aligned} & \beta\varepsilon\|\llbracket\eta_{\mathbf{q}}\cdot\mathbf{n}\rrbracket\|_{\mathcal{E}_h^i}^2 + \alpha\varepsilon\|\llbracket\eta_u\mathbf{n}\rrbracket\|_{\mathcal{E}_h^i}^2 + \theta\varepsilon\|\eta_u\|_{\Gamma_D}^2 + \frac{\gamma}{\omega}\|\eta_{\mathbf{q}}\cdot\mathbf{n}\|_{\Gamma_N}^2 + (1-\gamma)\varepsilon^2\omega\|\eta_u\|_{\Gamma_N}^2 \\ & \leq |R_1| + |R_2|. \end{aligned}$$

Let us estimate $|R_1|$ and $|R_2|$. Using Cauchy inequality, we get

$$|R_1| = |(\eta_u, f\delta_u)| \leq \|f\|_{\infty}\|\delta_u\|\|\eta_u\| \leq C\|\delta_u\|\|\eta_u\| \quad (3.5)$$

and

$$\begin{aligned}
|R_2| &\leq C\varepsilon \|[\![\eta_{\mathbf{q}} \cdot \mathbf{n}]\!]\|_{\mathcal{E}_h^i} (\|\delta_u\|_{\partial\Omega_h \setminus \partial\Omega} + \beta \|\delta_{\mathbf{q}}\|_{\partial\Omega_h \setminus \partial\Omega}) \\
&\quad + C\varepsilon \|[\![\eta_u \mathbf{n}]\!]\|_{\mathcal{E}_h^i} (\|\delta_{\mathbf{q}}\|_{\partial\Omega_h \setminus \partial\Omega} + \alpha \|\delta_u\|_{\partial\Omega_h \setminus \partial\Omega}) \\
&\quad + C\varepsilon \|[\![\eta_u]\!]\|_{\Gamma_D} (\|\delta_{\mathbf{q}}\|_{\Gamma_D} + \varepsilon \theta \|\delta_u\|_{\Gamma_D}) \\
&\quad + C\varepsilon \|\eta_{\mathbf{q}} \cdot \mathbf{n}\|_{\Gamma_N} ((1 - \gamma) \|\delta_u\|_{\Gamma_N} + \frac{\gamma}{\varepsilon \omega} \|\delta_{\mathbf{q}}\|_{\Gamma_N}) \\
&\quad + C\varepsilon \|\eta_u\|_{\Gamma_N} (\gamma \|\delta_{\mathbf{q}}\|_{\Gamma_N} + (1 - \gamma) \varepsilon \omega \|\delta_u\|_{\Gamma_N}).
\end{aligned}$$

For $\alpha > 0, \beta > 0, \theta > 0, 0 < \gamma < 1$ and $0 < C_0 < f < C_1$, we have

$$|R_2| \leq C\varepsilon (\|[\![\eta_{\mathbf{q}} \cdot \mathbf{n}]\!]\|_{\mathcal{E}_h^i \cup \Gamma_N} + \|[\![\eta_u \mathbf{n}]\!]\|_{\mathcal{E}_h}) (\|\delta_u\|_{\partial\Omega_h} + \|\delta_{\mathbf{q}}\|_{\partial\Omega_h}). \quad (3.6)$$

The conclusion follows by using (3.5)-(3.6) and Lemma 3.1. $\square$

To prove Theorem 3.4, we consider the following dual problem.

$$\begin{cases} \psi - \varepsilon \nabla \varphi = 0, & \text{in } \Omega, \\ -\varepsilon \nabla \cdot \psi - f \varphi = \xi, & \text{in } \Omega, \\ \varphi = 0 & \text{on } \Gamma_D, \\ \psi \cdot \mathbf{n} + i \varepsilon \omega \varphi = 0 & \text{on } \Gamma_N, \end{cases} \quad (3.7)$$

Assume that we have the following regularity for the dual problem

$$\|\psi\|_1 + \|\varphi\|_1 \leq C \|\xi\|, \quad (3.8)$$

where C depends on ε .

We first prove the following Lemma using duality argument.

Lemma 3.7.

$$(\eta_u, \xi) = S_1 + S_2,$$

where

$$\begin{aligned}
S_1 &= -\varepsilon \langle \eta_u, \delta_{\psi} \cdot \mathbf{n} \rangle + \varepsilon \langle \eta_{\mathbf{q}}, \delta_{\varphi} \mathbf{n} \rangle - \varepsilon \langle \widehat{e}_{\mathbf{q}}, \delta_{\varphi} \mathbf{n} \rangle + \varepsilon \langle \widehat{e}_u, \delta_{\psi} \cdot \mathbf{n} \rangle, \\
S_2 &= ((f - \mathbb{P}^0 f) \delta_u, \Pi \varphi) + ((f - \mathbb{P}^0 f) \eta_u, \delta_{\varphi}).
\end{aligned}$$

Proof. Using the dual problem (3.7), we have

$$\begin{aligned}
(\eta_u, \xi) &= (\eta_u, -\varepsilon \nabla \cdot \psi - f \varphi) \\
&= -\varepsilon (\eta_u, \nabla \cdot \delta_{\psi}) - \varepsilon (\eta_u, \nabla \cdot \Pi \psi) - (\eta_u, f \phi). \quad (3.9)
\end{aligned}$$

Let us rewrite the first two terms on the right hand side of (3.9). Using integration by parts and the orthogonality property of the projection Π (3.1b), we have

$$-\varepsilon(\eta_u, \nabla \cdot \delta_\psi) = -\varepsilon\langle \eta_u, \delta_\psi \cdot \mathbf{n} \rangle + \varepsilon(\nabla \eta_u, \delta_\psi) = -\varepsilon\langle \eta_u, \delta_\psi \cdot \mathbf{n} \rangle. \quad (3.10)$$

Taking $\mathbf{w} = \Pi\psi$ in the error equation (3.2a) and using the orthogonality property (3.1b) and the dual problem, we get

$$\begin{aligned} -\varepsilon(\eta_u, \nabla \cdot \Pi\psi) &= (\eta_q, \Pi\psi) - \varepsilon\langle \widehat{e}_u, \Pi\psi \cdot \mathbf{n} \rangle \\ &= (\eta_q, \psi) - \varepsilon\langle \widehat{e}_u, \Pi\psi \cdot \mathbf{n} \rangle \\ &= (\eta_q, \varepsilon \nabla \varphi) - \varepsilon\langle \widehat{e}_u, \Pi\psi \cdot \mathbf{n} \rangle \\ &= (\eta_q, \varepsilon \nabla \delta_\varphi) + (\eta_q, \varepsilon \nabla \Pi\varphi) - \varepsilon\langle \widehat{e}_u, \Pi\psi \cdot \mathbf{n} \rangle. \end{aligned}$$

Then using integration by parts, the orthogonality property of the projection Π (3.1a), and the error equation (3.2b) with $v = \Pi\varphi$, we obtain that

$$\begin{aligned} -\varepsilon(\eta_u, \nabla \Pi\psi) &= \langle \eta_q, \varepsilon \delta_\varphi \mathbf{n} \rangle - (\nabla \cdot \eta_q, \varepsilon \delta_\varphi) + (\eta_q, \varepsilon \nabla \Pi\varphi) - \varepsilon\langle \widehat{e}_u, \Pi\psi \cdot \mathbf{n} \rangle \\ &= \langle \eta_q, \varepsilon \delta_\varphi \mathbf{n} \rangle + \varepsilon\langle \widehat{e}_q, \Pi\varphi \mathbf{n} \rangle + (fe_u, \Pi\varphi) - \varepsilon\langle \widehat{e}_u, \Pi\psi \cdot \mathbf{n} \rangle. \end{aligned} \quad (3.11)$$

Plugging (3.10) and (3.11) into (3.9), we get

$$(\eta_u, \xi) = S_1 + S_2,$$

where

$$\begin{aligned} S_1 &= -\varepsilon\langle \eta_u, \delta_\psi \cdot \mathbf{n} \rangle + \langle \eta_q, \varepsilon \delta_\varphi \mathbf{n} \rangle + \varepsilon\langle \widehat{e}_q, \Pi\varphi \mathbf{n} \rangle - \varepsilon\langle \widehat{e}_u, \Pi\psi \cdot \mathbf{n} \rangle \\ S_2 &= (fe_u, \Pi\varphi) - (\eta_u, f\varphi). \end{aligned}$$

Since $\widehat{e}_q$ and $\widehat{e}_u$ are single-valued, using Lemma 3.2 and the boundary condition of the dual problem, it is easy to see that

$$-\varepsilon\langle \widehat{e}_q, \varphi \mathbf{n} \rangle + \varepsilon\langle \widehat{e}_u, \psi \cdot \mathbf{n} \rangle = 0.$$

So we get

$$S_1 = -\varepsilon\langle \eta_u, \delta_\psi \cdot \mathbf{n} \rangle + \varepsilon\langle \eta_q, \delta_\varphi \mathbf{n} \rangle - \varepsilon\langle \widehat{e}_q, \delta_\varphi \mathbf{n} \rangle + \varepsilon\langle \widehat{e}_u, \delta_\psi \cdot \mathbf{n} \rangle.$$

Using $e_u = \delta_u + \eta_u$, we can rewrite S_2 as

$$S_2 = (f\delta_u, \Pi\varphi) - (\eta_u, f\delta_\varphi).$$

This completes the proof. $\square$

Now let us prove Theorem 3.4.

Proof. Taking $\xi = \eta_u$ in Lemma 3.7, we get

$$\|\eta_u\|^2 \leq |S_1| + |S_2|. \quad (3.12)$$

Now let us estimate $|S_1|$ and $|S_2|$. We first estimate S_2 . By Cauchy inequality and Lemma 3.1,

$$\begin{aligned} |S_2| &\leq \|f\|_\infty \|\delta_u\| \|\Pi\varphi\| + \|f\|_\infty \|\eta_u\| \|\delta_\varphi\| \\ &\leq Ch^{k+1} \|u\|_{k+1} \|\varphi\| + Ch \|\varphi\|_1 \|\eta_u\|. \end{aligned}$$

Using the regularity property, (3.8), we have

$$|S_2| \leq Ch^{k+1} \|u\|_{k+1} \|\eta_u\| + Ch \|\eta_u\|^2. \quad (3.13)$$

Next we estimate S_1 . Using the expressions of $\widehat{e}_u$ and $\widehat{e}_q$ in Lemma 3.2, after some calculations we get

$$\begin{aligned} S_1 &= \varepsilon \langle \delta_u^- - \mathbf{i}\beta[\delta_q \cdot \mathbf{n}] - \mathbf{i}\beta[\eta_q \cdot \mathbf{n}], [\delta_\psi \cdot \mathbf{n}] \rangle_{\mathcal{E}_h^i} - \varepsilon \langle [\eta_u \mathbf{n}], \delta_\psi^+ \rangle_{\mathcal{E}_h^i} \\ &\quad - \varepsilon \langle \delta_q^+ + \mathbf{i}\alpha[\delta_u \mathbf{n}] + \mathbf{i}\alpha[\eta_u \mathbf{n}], [\delta_\varphi \mathbf{n}] \rangle_{\mathcal{E}_h^i} + \varepsilon \langle [\eta_q \cdot \mathbf{n}], \delta_\varphi^- \rangle_{\mathcal{E}_h^i} \\ &\quad - \varepsilon \langle \eta_u, \delta_\psi \cdot \mathbf{n} \rangle_{\Gamma_D} - \varepsilon \langle \delta_q - \theta(\delta_u + \eta_u) \mathbf{n}, \delta_\varphi \mathbf{n} \rangle_{\Gamma_D} \\ &\quad + \varepsilon \langle (1 - \gamma)\delta_u - \gamma\eta_u - \mathbf{i} \frac{\gamma}{\varepsilon\omega} (\delta_q + \eta_q) \cdot \mathbf{n}, \delta_\psi \cdot \mathbf{n} \rangle_{\Gamma_N} \\ &\quad - \varepsilon \langle (\gamma\delta_q - (1 - \gamma)\eta_q + (1 - \gamma)\mathbf{i}\varepsilon\omega(\delta_u + \eta_u \mathbf{n}), \delta_\varphi \mathbf{n} \rangle_{\Gamma_N}. \end{aligned}$$

So

$$\begin{aligned} |S_1| &\leq C(\|\delta_u\|_{\partial\Omega_h} + \|\delta_q\|_{\partial\Omega_h})(\|\delta_\varphi\|_{\partial\Omega_h} + \|\delta_\psi\|_{\partial\Omega_h}) \\ &\quad + (\|[\eta_u \mathbf{n}]\|_{\mathcal{E}_h} + \|[\eta_q \cdot \mathbf{n}]\|_{\mathcal{E}_h^i \cup \Gamma_N})(\|\delta_\varphi\|_{\partial\Omega_h} + \|\delta_\psi\|_{\partial\Omega_h}). \end{aligned}$$

Using Lemma 3.1 and Theorem 3.3, we get

$$|S_1| \leq Ch^{k+1}(\|\varphi\|_1 + \|\psi\|_1)(\|u\|_{k+1} + \|\mathbf{q}\|_{k+1}) + Ch^{1/2}(\|\varphi\|_1 + \|\psi\|_1)\|\eta_u\|.$$

By (3.8), we have

$$|S_1| \leq Ch^{k+1} \|\eta_u\| (\|u\|_{k+1} + \|\mathbf{q}\|_{k+1}) + Ch^{1/2} \|\eta_u\|^2. \quad (3.14)$$

Combining (3.12)-(3.14), for h sufficiently small, we have

$$\|\eta_u\|^2 \leq Ch^{k+1} \|\eta_u\| (\|u\|_{k+1} + \|\mathbf{q}\|_{k+1}),$$

which implies that

$$\|\eta_u\| \leq Ch^{k+1} (\|u\|_{k+1} + \|\mathbf{q}\|_{k+1}).$$

This completes the proof. $\square$

4 Numerical results

In this section, we perform several 2D numerical tests using the proposed multiscale DG methods. Since the solutions have oscillations mainly in x -direction, we apply the multiscale DG methods with the multiscale spaces $M^1 = E^1 \otimes P^1$, $M^2 = E^2 \otimes P^2$ and $M^3 = E^3 \otimes P^3$.

In the first example, we show that the multiscale DG methods are able to capture the oscillatory solution exactly if the solution is in the multiscale finite element spaces. In the second example, we show the optimal order of convergence of the multiscale DG for two levels of ε . In the third example, we apply the method in the simulation of a two-dimensional Schrödinger problem. In the second and third examples, we also compare multiscale DG methods using M^k with traditional DG methods using Q^k , where Q^k is the space of piecewise polynomials up to degree k in each variable and has the same number of basis functions as M^k . Numerical results show that our multiscale DG methods provide more accurate approximations than traditional DG methods for these oscillatory solutions.

We would like to remark that we have tested multiscale DG methods for different choices of penalty parameters, although our analysis requires all penalty parameters to be positive and is inconclusive for zero penalty parameters. When the penalty parameters α and β are positive, the resulting linear system involves a full matrix and is computationally expensive to solve on fine meshes. When α and β are zeros, the matrix is banded and we are able to solve on more refined meshes. In all our numerical tests, we take $\theta = 1/h$ on Γ_D^+ and zero elsewhere, which is the same as the standard MD-LDG method in [8]. From the numerical experiments, we observe similar magnitudes of errors and the same orders of convergence when choosing different penalty parameters in our method.

4.1 Constant f , $\omega = \frac{\sqrt{f}}{\varepsilon}$

Example 4.1. In the first example, we consider the simple case of Eq. (2.1) with constant function $f(x) = 10$. The boundary conditions are given as

$$\begin{cases} u_x(0, y) + i\omega u(0, y) = 2i\omega \\ u_x(1, y) - i\omega u(1, y) = 0 \\ u(x, 0) = e^{i\frac{\sqrt{10}}{\varepsilon}x} \\ u(x, 1) = e^{i\frac{\sqrt{10}}{\varepsilon}x} \end{cases},$$

Table 4.1: Example 4.1: L^2 -errors by multiscale DG for $f(x) = 10$.

	$N_x \times N_y$	$\alpha = \beta = 0, \gamma = 0$			$\alpha = \beta = 1, \gamma = 0.5$		
		M^1	M^2	M^3	M^1	M^2	M^3
$\varepsilon = 0.1$	4×4	4.79E-13	4.41E-13	7.47E-13	5.44E-13	5.62E-13	5.29E-13
	8×8	4.65E-13	3.17E-13	1.21E-13	5.11E-13	4.79E-13	6.59E-12
$\varepsilon = 0.03$	4×4	9.30E-12	1.98E-11	2.08E-11	7.63E-12	1.14E-11	1.07E-11
	8×8	1.58E-11	1.98E-11	7.11E-11	1.39E-11	1.48E-11	1.32E-11

where $\omega = \sqrt{10}/\varepsilon = \sqrt{f}/\varepsilon$. It has the exact solution $u = e^{i\frac{\sqrt{10}}{\varepsilon}x}$. In this example, the exact solution lies in the multiscale finite element spaces. Thus the proposed multiscale DG methods with these spaces are able to compute the solution exactly with round-off errors. The L^2 -errors of the multiscale DG methods with multiscale spaces M^1 , M^2 and M^3 are shown in Table 4.1 for two different levels of ε , i.e., $\varepsilon = 0.1$ and a smaller $\varepsilon = 0.03$. We tested both the multiscale DG with zero penalty parameters $\alpha = \beta = \gamma = 0$ and positive penalty parameters $\alpha = \beta = 1$, $\gamma = 0.5$. It is clear to see nearly round-off errors in double precision for all results. We remark that in our computations, the integrals involving exponential functions are integrated numerically by quadrature rules which produces small errors.

4.2 Constant f , $\omega \neq \frac{\sqrt{f}}{\varepsilon}$

Example 4.2. In this example, we consider a constant $f(x) = 10 + \varepsilon^2\pi^2$. The boundary conditions are given as

$$\begin{cases} u_x(0, y) + i\omega u(0, y) = 2i\omega \sin(\pi y) \\ u_x(1, y) - i\omega u(1, y) = 0 \\ u(x, 0) = 0 \\ u(x, 1) = 0 \end{cases},$$

where $\omega = \sqrt{10}/\varepsilon$. It has the exact solution $u = e^{i\frac{\sqrt{10}}{\varepsilon}x} \sin(\pi y)$ which contains a highly oscillatory wave function in the x -direction and a sine function in the y -direction. The multiscale finite element spaces are able to capture the oscillatory wave function in the x -direction.

We first perform the experiments for the multiscale DG with different magnitudes of penalty parameters, i.e., $\alpha = \beta = 0$, $\alpha = \beta = 0.1$, and $\alpha = \beta = 1$. Table 4.2 and 4.3 list L^2 -errors and orders by multiscale DG for

$\varepsilon = 0.1$ and $\varepsilon = 0.03$, respectively. We observe similar error magnitudes and orders of convergence for different levels of penalty parameters.

Next, we compare our multiscale DG method with $\alpha = \beta = \gamma = 0$ and its traditional polynomial-based counterpart, the MD-LDG method, in the same order of finite elements spaces. The number of elements we use for the multiscale DG is one-fourth of those for the MD-LDG, because the multiscale finite element spaces are designed to effectively capture the oscillatory solution even on coarse meshes.

Tables 4.4, 4.5, 4.6 and Tables 4.7, 4.8, 4.9 show the comparisons between multiscale DG methods with multiscale spaces M^1 , M^2 and M^3 and traditional polynomial DG methods with Q^1 , Q^2 and Q^3 for $\varepsilon = 0.1$ and a smaller $\varepsilon = 0.03$, respectively. From 4.4, 4.5, 4.6, we can see that when $\varepsilon = 0.1$, $h < \varepsilon$, both multiscale DG methods with multiscale spaces M^1 , M^2 and M^3 and traditional polynomial DG methods with Q^1 , Q^2 and Q^3 achieve the optimal $(k+1)$ -th order of convergence. However, the errors of multiscale DG methods are several magnitudes smaller than those of traditional polynomial DG methods even if the mesh size of multiscale DG in x -direction is much larger. When using a smaller ε , $\varepsilon = 0.03$, in Tables 4.7-4.9 we can still see $(k+1)$ -th order of convergence for multiscale DG methods at all mesh levels. However, standard polynomial DG methods can not approximate the solutions well until the mesh is refined to $h < \varepsilon$. For example, polynomial DG methods with Q^1 and Q^2 do not have any order of convergence until mesh is refined to 64×64 . We remark that due to the limit space in single processor, we are not able to compute polynomial DG methods Q^3 with 64×64 in Table 4.6 and 4.9.

4.3 Applications to 2D Schrödinger Equation

Example 4.3. In this example, we apply the multiscale method to a 2D stationary Schrödinger problem on the domain $[0, 1] \times [0, 1]$,

$$\begin{aligned} -\varepsilon^2 \Delta \phi + V(x, y) \phi &= E \phi, \\ \phi_x(0, y) + i\omega \phi(0, y) &= 2i\omega, \\ \phi_x(1, y) - i\omega \phi(1, y) &= 0, \\ \phi(x, 0) &= 0, \\ \phi(x, 1) &= 0. \end{aligned}$$

We let $\varepsilon = 0.03$, the energy $E = 1$, the external potential $V(x, y) = \frac{1}{2} \sin x \cos y - 1$, and $\omega = \sqrt{2}/\varepsilon$.

Table 4.2: Example 4.2: L^2 -errors and orders of accuracy by multiscale DG for $\varepsilon = 0.1$.

		$\alpha = \beta = 0, \gamma = 0$		$\alpha = \beta = 0.1, \gamma = 0.5$		$\alpha = \beta = 1, \gamma = 0.5$	
	$N_x \times N_y$	error	order	error	order	error	order
M^1	4×4	2.95E-02	–	2.64E-02	–	1.84E-02	–
	8×8	6.70E-03	2.14	6.43E-03	2.04	4.34E-03	2.08
	16×16	1.66E-03	2.01	1.61E-03	2.00	1.07E-03	2.02
M^2	4×4	1.77E-03	–	1.70E-03	–	1.66E-03	–
	8×8	2.20E-04	3.01	2.12E-04	3.01	2.51E-04	2.72
	16×16	2.62E-05	3.07	2.61E-05	3.02	3.16E-05	2.99
M^3	4×4	9.47E-05	–	8.35E-05	–	6.10E-05	–
	8×8	5.07E-06	4.22	4.94E-06	4.08	3.53E-06	4.11

Table 4.3: Example 4.2: L^2 -errors and orders of accuracy by multiscale DG for $\varepsilon = 0.03$.

		$\alpha = \beta = 0, \gamma = 0$		$\alpha = \beta = 0.1, \gamma = 0.5$		$\alpha = \beta = 1, \gamma = 0.5$	
	$N_x \times N_y$	error	order	error	order	error	order
M^1	4×4	3.03E-02	–	2.44E-02	–	1.73E-02	–
	8×8	7.46E-03	2.02	6.44E-03	1.92	4.25E-03	2.03
	16×16	1.67E-03	2.16	1.61E-03	2.00	1.05E-03	2.01
M^2	4×4	2.05E-03	–	1.62E-03	–	1.61E-03	–
	8×8	3.06E-04	2.74	2.15E-04	2.91	2.36E-04	2.77
	16×16	2.67E-05	3.52	2.62E-05	3.04	2.97E-05	2.99
M^3	4×4	8.37E-05	–	8.00E-05	–	6.10E-05	–
	8×8	5.19E-06	4.01	4.96E-06	4.01	3.52E-06	4.12

Table 4.4: Example 4.2: Comparison between multiscale DG M^1 and polynomial MD-LDG Q^1 for $\varepsilon = 0.1$.

multiscale M^1			polynomial Q^1		
$N_x \times N_y$	error	order	$N_x \times N_y$	error	order
			8×8	7.55E-01	
4×16	1.83E-03	–	16×16	2.55E-01	1.56
8×32	4.42E-04	2.05	32×32	4.53E-02	2.49
16×64	1.05E-04	2.08	64×64	1.09E-02	2.06

Table 4.5: Example 4.2: Comparison between multiscale DG M^2 and polynomial MD-LDG Q^2 for $\varepsilon = 0.1$.

multiscale M^2			polynomial Q^2		
$N_x \times N_y$	error	order	$N_x \times N_y$	error	order
			8×8	2.49E-01	
4×16	9.07E-05	—	16×16	2.93E-02	3.09
8×32	3.31E-06	4.78	32×32	3.38E-03	3.12
16×64	4.08E-07	3.02	64×64	4.18E-04	3.02

Table 4.6: Example 4.2: Comparison between multiscale DG M^3 and polynomial ND-LDG Q^3 for $\varepsilon = 0.1$.

multiscale M^3			polynomial Q^3		
$N_x \times N_y$	error	order	$N_x \times N_y$	error	order
			8×8	5.50E-02	
4×16	5.80E-07	—	16×16	3.23E-03	4.09
8×32	2.46E-08	4.56	32×32	2.02E-04	4.00
16×64	1.32E-09	4.22	64×64	—	—

Table 4.7: Example 4.2: Comparison between multiscale DG M^1 and polynomial MD-LDG Q^1 for $\varepsilon = 0.03$.

multiscale M^1			polynomial Q^1		
$N_x \times N_y$	error	order	$N_x \times N_y$	error	order
			8×8	7.06E-01	
4×16	2.21E-03		16×16	7.15E-01	-0.02
8×32	4.29E-04	2.37	32×32	6.98E-01	0.03
16×64	1.05E-04	2.03	64×64	3.16E-01	1.14

Table 4.8: Example 4.2: Comparison between multiscale DG M^2 and polynomial MD-LDG Q^2 for $\varepsilon = 0.03$.

multiscale M^2			polynomial Q^2		
$N_x \times N_y$	error	order	$N_x \times N_y$	error	order
			8×8	7.22E-01	
4×16	5.19E-05	—	16×16	7.03E-01	0.04
8×32	3.36E-06	3.95	32×32	4.86E-01	0.53
16×64	4.18E-07	3.01	64×64	1.59E-02	4.93

Table 4.9: Example 4.2: Comparison between multiscale DG M^3 and polynomial MD-LDG Q^3 for $\varepsilon = 0.03$.

multiscale M^3			polynomial Q^3		
$N_x \times N_y$	error	order	$N_x \times N_y$	error	order
			8×8	7.16E-01	
4×16	4.37E-07	—	16×16	7.32E-01	-0.03
8×32	5.41E-08	3.01	32×32	3.16E-02	4.54
16×64	1.40E-09	5.27	64×64	—	—

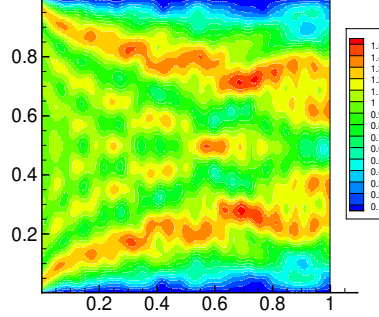


Figure 4.1: Example 4.3: reference solution.

We do not have the formula of the exact solution. The reference solutions are computed by polynomial DG P^2 with mesh 80×80 in Figure 4.1. On Figure 4.2, we compare the numerical results by multiscale DG M^1 and polynomial DG Q^1 on the same coarse mesh 32×32 . We can see that for the same degree of freedom, multiscale DG M^1 is able to capture the shape and height of the waves pretty well. However, polynomial DG produce very spurious waves. This is because standard polynomial DG needs $h < \varepsilon$ in order to resolve the small oscillations in the solution. When the mesh is refined, standard polynomial DG will be able to approximate the solution well. Next, we compare the numerical results by multiscale DG M^2 with polynomial DG Q^2 on coarse mesh 16×32 ; see Figure 4.3. Again, we observe that multiscale DG M^2 is able to capture the oscillatory waves very well, but polynomial DG Q^2 with the same coarse mesh produces spurious waves. Therefore, the multiscale DG is able to approximate the solution on coarse mesh and uses less degree of freedom than standard polynomial DG. Thus it is more efficient and accurate than the standard DG methods for solving the problems involving small scales.

5 Concluding remarks

In this paper, we extend our previous work on the high order multiscale discontinuous Galerkin method for one-dimensional stationary Schrödinger equations to two-dimensional space. The solution under consideration has frequency change mainly in one direction, so we use oscillatory non-polynomial

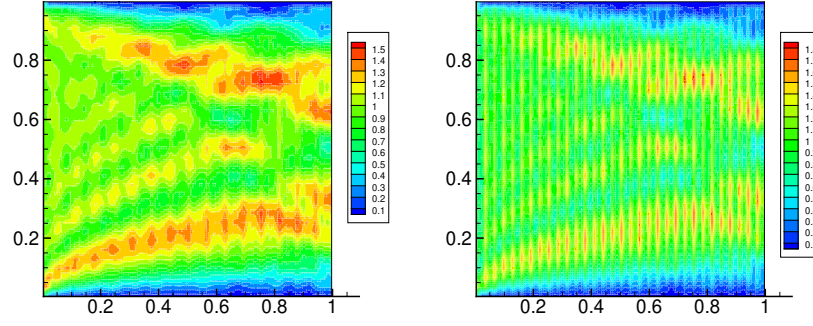


Figure 4.2: Example 4.3: Numerical solutions 32×32 . Left: multiscale DG M^1 . Right: polynomial DG Q^1 .

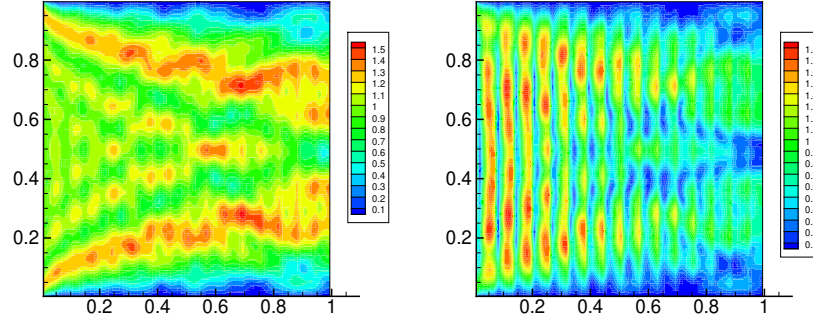


Figure 4.3: Example 4.3: Numerical solutions 16×32 . Left: multiscale DG M^2 . Right: polynomial DG Q^2 .

basis functions in that direction and polynomial basis in the other direction. We prove that the resulting DG method converges with an optimal order when the mesh size h is small enough. Numerically we observe that the method converges on coarse meshes and has an optimal convergence order when the mesh size is refined to the scale of the wave length. Numerically, we show the accuracy and convergence order of second order, third order and fourth order multiscale discontinuous Galerkin (DG) methods and compare them with traditional polynomial DG methods. We also demonstrate their ability to capture highly oscillating solutions of the Schrödinger equation in the application of the stationary Schrödinger equations. In the future work, we would like to investigate the explicit dependence on the parameter ε in the error estimates. We would also like to develop the multiscale discontinuous Galerkin method for more general two-dimensional problems with oscillatory solutions in both directions.

References

- [1] J. Aarnes and B.-O. Heimsund, *Multiscale discontinuous Galerkin methods for elliptic problems with multiple scales*, Multiscale methods in science and engineering 1–20, Lect. Notes Comput. Sci. Eng., 44. Springer, Berlin, 2005.
- [2] D.N. Arnold, F. Brezzi, B. Cockburn and L.D. Marini, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, SIAM J. Numer. Anal., **39**, 2002, 1749–1779.
- [3] A. Arnold, N Ben Abdallah and C. Negulescu, WKB-based schemes for the oscillatory 1D Schrödinger equation in the semiclassical limit, SIAM J. Numer. Anal., 49, 2011, 1436–1460.
- [4] N. Ben Abdallah, M. Mouis and C. Negulescu, *An accelerated algorithm for 2D simulations of the quantum ballistic transport in nanoscale MOS-FETs*, J. Comput. Phys., **225**, 2007, 74–99.
- [5] N. Ben Abdallah and O. Pinaud, *Multiscale simulation of transport in an open quantum system: Resonances and WKB interpolation*, J. Comput. Phys., **213**, 2006, 288–310.

- [6] N. Ben Abdallah and H. Wu, *A Generalized Stationary Algorithm for Resonant Tunneling: Multi-Mode Approximation and High Dimension*, Commun. Comput. Phys. **10**, 2011, 882-898.
- [7] A. Buffa and P. Monk, *Error estimates for the Ultra Weak Variational Formulation of the Helmholtz equation*, ESAIM: M2AN Math. Model. Numer. Anal. **42**, 2008, 925–940.
- [8] B. Cockburn and B. Dong, *An analysis of the minimal dissipation local discontinuous Galerkin method for convection-diffusion problems*, J. Sci. Comput., **32**, 2007, 233–262.
- [9] B. Dong, C.-W. Shu and W. Wang, *A new multiscale discontinuous Galerkin method for the one-dimensional stationary Schrodinger equation*, J. Sci. Comput., **66**, 2016, 321–345.
- [10] B. Dong and W. Wang, *High-order multiscale discontinuous Galerkin methods for the one-dimensional stationary Schrodinger equation*, J. Comput. and Applied Mathematics, **380**, 2020, 1–11.
- [11] X. Feng and H. Wu, *Discontinuous Galerkin methods for the Helmholtz equation with large wave number*, SIAM J. Numer. Anal., **47**, 2009, 2872–2896.
- [12] G. Gabard, *Discontinuous Galerkin methods with plane waves for time-harmonic problems*, J. Comput. Phys., **225**, 2007, 1961–1984.
- [13] C. Gittelsohn, R. Hiptmair, and I. Perugia, *Plane wave discontinuous Galerkin methods: Analysis of the h-version*, ESAIM: M2AN Math. Model. Numer. Anal., **43**, 2009, 297–331.
- [14] L. Guo and Y. Xu, *Local Discontinuous Galerkin Methods for the 2D Simulation of Quantum Transport Phenomena on Quantum Directional Coupler*, Commun. Comput. Phys., **15**, 2014, 1012–1028.
- [15] C. Negulescu, *Numerical analysis of a multiscale finite element scheme for the resolution of the stationary Schrödinger equation*, Numer. Math., **108**, 2008, 625–652.
- [16] C. Negulescu, N. Ben Abdallah, E. Polizzi and M. Mouis, *Simulation schemes in 2D nanoscale MOSFETs: A WKB based method*, J. Comput. Electron., **3**, 2004, 397–400.

- [17] E. Polizzi and N. Ben Abdallah, *Subband decomposition approach for the simulation of quantum electron transport in nanostructures*, J. Comput. Phys., **202**, 2005, 150–180.
- [18] W. Wang, J. Guzmán and C.-W. Shu, *The multiscale discontinuous Galerkin method for solving a class of second order elliptic problems with rough coefficients*, Int. J. Numer. Anal. Model, **8**, 2011, 28–47.
- [19] W. Wang and C.-W. Shu, *The WKB local discontinuous Galerkin method for the simulation of Schrödinger equation in a resonant tunneling diode*, J. Sci. Comput., **40**, 2009, 360–374.
- [20] L. Yuan and C.-W. Shu, *Discontinuous Galerkin method based on non-polynomial approximation spaces*, J. Comput. Phys., **218**, 2006, 295–323.
- [21] L. Yuan and C.-W. Shu, *Discontinuous Galerkin method for a class of elliptic multi-scale problems*, Int. J. Numer. Meth. Fluids, **56**, 2008, 1017–1032.
- [22] Y. Zhang, W. Wang, J. Guzmán and C.-W. Shu, *Multi-scale discontinuous Galerkin method for solving elliptic problems with curvilinear unidirectional rough coefficients*, J. Sci. Comput., **61**, 2014, 42–60.