Journal of Computational Mathematics
Vol.32, No.5, 2014, 491–506.

http://www.global-sci.org/jcm
doi:10.4208/jcm.1403-m4404

ERROR ESTIMATION FOR NUMERICAL METHODS USING THE ULTRA WEAK VARIATIONAL FORMULATION IN MODEL OF NEAR FIELD SCATTERING PROBLEM*

Tian Luan
School of Mathematics and Statistics, Beihua University, Jilin 132013, China
Email: luantian@163.com
Fuming Ma Minghui Liu
Institute of Mathematics, Jilin University, Changchun 130012, China
Email: mfm@jlu.edu.com lmh09@mails.jlu.edu.cn

Abstract

In this paper, we investigate the use of ultra weak variational formulation to solve a wave scattering problem in near field optics. In order to capture the sub-scale features of waves, we utilize evanescent wave functions together with plane wave functions to approximate the local properties of the field. We analyze the global convergence and give an error estimation of the method. Numerical examples are also presented to demonstrate the effectiveness of the strategy.

Mathematics subject classification: 65N12, 65N55.
Key words: Helmholtz equation, Ultra weak variational formulation, Plane wave function, Evanescent wave function, Absorbing boundary condition.

1. Introduction

Near filed optics could provide an effective approach to break the diffraction limit in conventional far-field optics [1, 2], so it has developed dramatically and been applied in diverse aspects in recent years, such as nondestructive imaging of biological samples, nanotechnology, near-field optical microscopy [3]. In order to theoretically understand the physical mechanism of this fascinating feature, it is desirable to accurately solve the underlying scattering problem.

In this paper, we focus on a typical scattering problem in the near field optics that models the total internal reflection microscopy (TIRM). More specifically, we consider a sample deposited on a homogeneous substrate and illustrated from below (transmission geometry). When the incident angle is greater than a critical value, the total internal reflection happens. Then the evanescent wave appears at the other side of the interface which is used as illumination to encode the sub-wavelength structure of the scattering object. This phenomenon is formulated mathematically by Helmholtz equation, which models time-harmonic electromagnetic wave scattering for the case of TM (transverse magnetic) polarization. However, at medium and high frequency, resolution requirements and so-called pollution effect entail an excessive computational efforts and prevent standard finite element method from effective use. Thus, numerically simulating the wave propagation is a challenging task. But the wave-based methods offer a possible way to deal with this problem. The main idea is to use special solutions of the underlying partial differential equation in each element to build the discrete space, thus a priori information about the

* Received June 17, 2013 / Revised version received December 11, 2013 / Accepted March 25, 2014 / Published online August 22, 2014 /

solution is directly incorporated in the approximation space. Possible techniques include least squares methods (LSM) [4–6], the partition of unity method (PUM) [7, 8], the discontinuous Galerkin method (DG) [9,10], and the ultra weak variational formulation (UWVF) [11,13,14]. It is the last of these techniques that will be considered for this work.

The ultra weak variational formulation is originated from the domain decomposition technique and was proposed by Cessenat and Després. In [11], Cessenat and Després studied an obstacle scattering problem in homogeneous background medium by virtue of this method using plane wave functions. They proved an error estimate for the method showing that the solution of the UWVF converges to an appropriate impedance trace of the true solution on the boundary of the domain. Inspired by their ideas, we apply UWVF to solve our problem. By introducing a coupling parameter, an ultra weak variational formulation suitable to TIRM scattering model problem is derived. And in order to capture the sub-scale features of the wave field, evanescent wave functions which are also the solution of Helmholtz equation are introduced to enrich the plane wave functions . We analyze the error estimation and give a hp-version convergence result even away from the boundary. The argument is fundamentally different from that of [11].

The remainder of this paper is organized as follows. In Section 2, we give a ultra weak variational formulation appropriate to our model problem. Then we introduce the discrete problem including construction of the approximation space. In Section 3, we analyze the error estimate of our approach. First, via an auxiliary sesquilinear, we show that our ultra weak variational formulation has sufficient coercivity to provide an error estimate. Next, an basic estimate is given by means of the duality techniques. Finally, we derive convergence result from the best approximation error and a knowledge of the approximation properties of bases. In Section 4, numerical experiments are presented to demonstrate the validity of the method. In Section 5, we conclude.

2. Continuous and Discrete Problem

2.1. Formulation of the Model Problem

In this subsection, we introduce the model problem to be studied and give some related notations used later.

The point in the plane is denoted by $\boldsymbol{x} = (x, y) \in \mathbb{R}^2$. The whole space $\mathbb{R}^2$ is divided by the substrate $\Gamma_0 = \{\boldsymbol{x} | y = 0\}$ into $\mathbb{R}^2_+ = \{y > 0\}$ and $\mathbb{R}^2_- = \{y < 0\}$. The corresponding refractive indexes are n_+ and n_- $(n_+ < n_-)$ respectively. A sample S with refractive n_s is deposited on the substrate Γ_0 and illustrated from below by time harmonic plane wave $u^i = \exp(\mathrm{i}\alpha x + \mathrm{i}\eta y)$ at an angle θ greater than a critical value, where $\alpha = k_0 n_- \sin\theta$ and $\eta = k_0 n_- \cos\theta$, and k_0 is the free-space wave number(see Fig. 2.1 for geometry of the model). Throughout we assume nonmagnetic materials and TM polarization.

When there is no sample, the field in $\mathbb{R}^2$ denoted by u^{ref} is called reference field. According to the Maxwell electromagnetic theory, u^{ref} is the solution of the following equation

$$\Delta u^{ref} + k_0^2 m^2(\boldsymbol{x}) u^{ref} = 0, \quad \boldsymbol{x} \in \mathbb{R}^2, \tag{2.1}$$

where refractive is defined by

$$m(\boldsymbol{x}) = \begin{cases} n_+, & \boldsymbol{x} \in \mathbb{R}^2_+, \\ n_-, & \boldsymbol{x} \in \mathbb{R}^2_-. \end{cases}$$

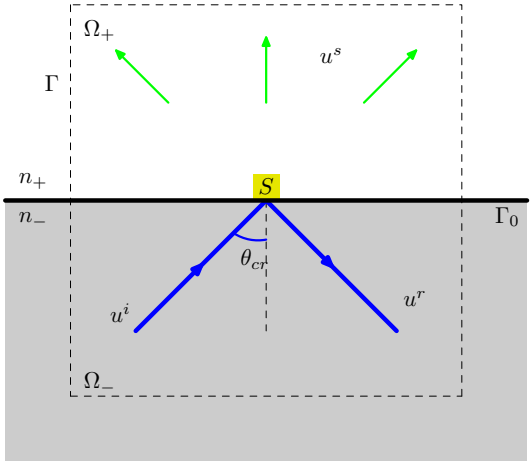

Fig. 2.1. Geometry of the model.

It has been shown in [12]that

$$u^{ref} = \begin{cases} u^t, & \boldsymbol{x} \in \mathbb{R}^2_+, \\ u^i + u^r, & \boldsymbol{x} \in \mathbb{R}^2_-, \end{cases} \tag{2.2}$$

where u^t and u^r are transmitted and reflected waves respectively. More precisely,

$$u^r = \frac{\eta - \gamma}{\eta + \gamma} \exp(\mathrm{i}\alpha x - \mathrm{i}\eta y), \qquad u^t = \frac{2\eta}{\eta + \gamma} \exp(\mathrm{i}\alpha x + \mathrm{i}\gamma y), \tag{2.3}$$

with

$$\gamma(\alpha) = \begin{cases} \sqrt{k_0^2 n_+^2 - \alpha^2}, & \text{for } k_0 n_+ > |\alpha|, \\ \mathrm{i}\sqrt{\alpha^2 - k_0^2 n_+^2}, & \text{for } k_0 n_+ < |\alpha|. \end{cases} \tag{2.4}$$

It is easily seen that when the incident angle is greater than the critical value, i.e., $k_0 n_+ < |\alpha|$, $\gamma(\alpha)$ is purely imaginary. Therefore, the transmitted wave becomes an evanescent wave, which propagates along the substrate surface but exponentially decays in the y direction.

When the sample appears, it gives rise to the emergence of the scattered field u^s. Let the total field u defined by

$$u = u^{ref} + u^s. \tag{2.5}$$

Then according to the electromagnetic theory of Maxwell, it satisfies the following equation

$$\Delta u + k_0^2 n^2(\boldsymbol{x}) u = 0, \quad \boldsymbol{x} \in \mathbb{R}^2, \tag{2.6}$$

where $k_0 n(\boldsymbol{x})$ is the wave number and $n(\boldsymbol{x})$ is refractive defined by

$$n(\boldsymbol{x}) = \begin{cases} n_+, & \boldsymbol{x} \in \mathbb{R}^2_+ \setminus S, \\ n_s\ , & \boldsymbol{x} \in S, \\ n_-, & \boldsymbol{x} \in \mathbb{R}^2_-. \end{cases}$$

From (2.1), (2.5) and (2.6), it is deduced that the scattered field u^s satisfies

$$\Delta u^s + k_0^2 n(\boldsymbol{x}) u^s = k_0^2 \big(m^2(\boldsymbol{x}) - n^2(\boldsymbol{x})\big) u^{ref}, \quad \boldsymbol{x} \in \mathbb{R}^2. \tag{2.7}$$

Next we have to introduce an artificial boundary Γ with unit out normal $\boldsymbol{\nu}$ to transform the problem physically unbounded to the one on a bounded domain Ω, i.e., $\Gamma = \partial\Omega$. To restrict spurious numerical reflections from the boundary Γ, we impose absorbing boundary condition on it

$$\partial_{\boldsymbol{\nu}} u^s - \mathrm{i}k_0 n(\boldsymbol{x}) u^s = 0, \quad \boldsymbol{x} \in \Gamma,$$

which is stable and simple to implement. Here and in the sequel we denote the normal derivative operator in the direction of the vector $\boldsymbol{\nu}$ by $\partial_{\boldsymbol{\nu}}$.

Thus the total field u satisfies

$$\big(\partial_{\boldsymbol{\nu}} - \mathrm{i}k_0 n(\boldsymbol{x})\big) u = \big(\partial_{\boldsymbol{\nu}} - \mathrm{i}k_0 n(\boldsymbol{x})\big) u^{ref}, \quad \boldsymbol{x} \in \Gamma.$$

So the near field scattering problem we interested can be formulated by the following boundary value problem: given the incident field u^i, to find the scattered field u^s satisfying

$$\begin{cases} \Delta u^s + k_0^2 n(\boldsymbol{x}) u^s = k_0^2 (m^2(\boldsymbol{x}) - n^2(\boldsymbol{x})) u^{ref}, & \boldsymbol{x} \in \Omega, \\ \big(\partial_{\boldsymbol{\nu}} - \mathrm{i}k_0 n(\boldsymbol{x})\big) u^s = 0, & \boldsymbol{x} \in \Gamma. \end{cases} \tag{2.8}$$

Alternatively, we could also seek the scattered field u^s from the total field u which satisfies

$$\begin{cases} \Delta u + k_0^2 n^2(\boldsymbol{x}) u = 0, & \boldsymbol{x} \in \Omega, \\ \big(\partial_{\boldsymbol{\nu}} - \mathrm{i}k_0 n(\boldsymbol{x})\big) u = g, & \boldsymbol{x} \in \Gamma, \end{cases} \tag{2.9}$$

where $g = \big(\partial_{\boldsymbol{\nu}} - \mathrm{i}k_0 n(\boldsymbol{x})\big) u^{ref}$.

2.2. Ultra Weak Variational Formulation

For this part, we first define the spaces and operators needed, then give the ultra weak variational formulation to our problem.

Let $\mathcal{T}_h = \{\Omega_k\}_{k=1}^N$ be a mesh of the domain Ω. We assume that the mesh is chosen such that refractive $n(\boldsymbol{x})$ is constant n_k on every Ω_k, i.e., $n_k = n(\boldsymbol{x})|_{\Omega_k}$. The unit outer normal of $\partial\Omega_k$ is denoted by $\tilde{\boldsymbol{\nu}}_k$. For two elements Ω_k and Ω_j, define $\Gamma_{k,j} = \partial\Omega_k \cap \partial\Omega_j$ with unit outer normal $\boldsymbol{\nu}_{k,j}$ which points from Ω_k to Ω_j if $\Gamma_{k,j} \neq \emptyset$. A face of $\partial\Omega_k$ on the exterior boundary Γ is denoted by $\Gamma_k = \Gamma \cap \partial\Omega_k$ with unit outer normal $\boldsymbol{\nu}_k$ if $\partial\Omega_k \cap \Gamma \neq \emptyset$. Let h_k be the diameter of Ω_k, and the mesh width is $h = \max_k h_k$.

To guarantee the continuity of the field across the interface between elements, we have to introduce a coupling parameter given as

$$\sigma(\boldsymbol{x}) = \begin{cases} k_0 \dfrac{n_k + n_j}{2}, & \boldsymbol{x} \in \Gamma_{k,j}, \\ k_0 n_k \ , & \boldsymbol{x} \in \Gamma_k, \end{cases}$$

which is the mean value of wave number on the skeleton of the mesh. From the definition of σ, it is known that $\sigma > 0$ and $\sigma(\boldsymbol{x}) = k_0 n(x),\ \boldsymbol{x} \in \Gamma$. For convenience, we denote $\sigma_{k,j} = \sigma|_{\Gamma_{k,j}}$ and $\sigma_k = \sigma|_{\Gamma_k},\ k = 1, 2, \cdots, N$.

The main function space in which we set our problem is a Hilbert space X, defined by

$$X = \prod_{k=1}^{N} L^2(\partial\Omega_k).$$

with the weighted scalar product

$$(\tilde{x},\tilde{y})=\sum_k\int_{\partial\Omega_k}\frac{1}{\sigma}\tilde{x}_k\overline{\tilde{y}_k}\,\mathrm{d}s,\quad \tilde{x},\ \tilde{y}\in X,$$

and the induced norm $\|\tilde{x}\|_X$. Here and in the following we denote $\tilde{x}_k=\tilde{x}|_{\partial\Omega_k}$, for $\tilde{x}\in X$.

Space H is defined by

$$H=\prod_{k=1}^{N}H_k,$$

with

$$H_k=\big\{v_k\in H^1(\Omega_k)|(\Delta+k_0^2n_k^2)v_k=0\ \textit{and}\ (\partial_{\boldsymbol{\nu}_k}+\mathrm{i}\sigma)v_k|_{\partial\Omega_k}\in L^2(\partial\Omega_k)\big\}.$$

In order to give a simple form of the formulation and prepare for discussions later , we need following operators which have the similar definitions as those in [11].

Definition 2.1. *Extension mapping $\mathcal{E}:\ X\to H$ is defined by*

$$\mathcal{E}(\tilde{z})=v,\quad k=1,2,\cdots,N,$$

where $v|_{\Omega_k}=v_k$ is the unique solution of the boundary value problem

$$\begin{cases}(\Delta+k_0^2n_k^2)v_k=0\ , & \text{in}\quad \Omega_k,\\ (\partial_{\tilde{\nu}_k}+\mathrm{i}\sigma)v_k=\tilde{z}_k, & \text{on}\quad \partial\Omega_k.\end{cases}$$

Definition 2.2. *Operator $\mathcal{F}=(\mathcal{F}_k)\in\mathcal{L}(X)$ mapping the outgoing trace to the incoming one is defined by*

$$\mathcal{F}_k(\tilde{z}_k)=\big((-\partial_{\tilde{\nu}_k}+\mathrm{i}\sigma)\mathcal{E}(\tilde{z})|_{\Omega_k}\big)|_{\partial\Omega_k},\quad k=1,2,\cdots,N.$$

Definition 2.3. *Linear operator $\Pi\in\mathcal{L}(X)$ is defined by*

$$\Pi(\tilde{z})=\begin{cases}\tilde{z}|_{\Gamma_{j,k}}, & \textit{on}\ \Gamma_{k,j},\\ 0\quad , & \textit{on}\ \Gamma_k.\end{cases}\qquad k,\ j=1,2,\cdots,N.$$

Then by Green theorem, we can directly derive following result, which states that boundary value problem (2.9) is equivalent to a variational problem over element interfaces. Since the argument is similar to Theorem 1.3 of [11], we only present the result but omit the details.

Theorem 2.1. *Let $u\in H^1(\Omega)$ be the solution of* (2.9) *and satisfies the regularity hypothesis $\partial_{\tilde{\nu}_k}u_k\in L^2(\partial\Omega_k)$ with $u_k=u|_{\Omega_k}$, $k=1,2,\cdots,N$. Define $\tilde{x}\in X$, such that $\tilde{x}_k=(\partial_{\tilde{\nu}_k}+\mathrm{i}\sigma)u_k|_{\partial\Omega_k}$. Then $\tilde{x}$ satisfies following ultra weak variational formulation*

$$a(\tilde{x},\tilde{y})=(\tilde{b},\tilde{y})_X,\quad \textit{for all}\ \tilde{y}\in X,\tag{2.10}$$

where

$$a(\tilde{x},\tilde{y})=\sum_k\int_{\partial\Omega_k}\frac{1}{\sigma}\tilde{x}_k\ \overline{\tilde{y}_k}\ \mathrm{d}s+\sum_k\sum_{j\neq k}\int_{\Gamma_{k,j}}\frac{1}{\sigma}\tilde{x}_j\ \overline{\mathcal{F}_k(\tilde{y}_k)}\ \mathrm{d}s.\tag{2.11}$$

Here $\tilde{b}\in X$ is defined, via the Riesz representation theorem, by

$$(\tilde{b},\tilde{y})_X=-\sum_k\int_{\Gamma_k}\frac{1}{\sigma}g\ \overline{\mathcal{F}_k(\tilde{y}_k)}\ \mathrm{d}s,\quad \textit{for all}\ \tilde{y}\in X.$$

Conversely, if $\tilde{x}$ satisfies (2.10) *then u solves problem* (2.9) *defined by*

$$u|_{\Omega_k}=u_k,\ (\Delta+k_0^2n_k^2)u_k=0,\quad (\partial_{\tilde{\nu}_k}+\mathrm{i}\sigma)u_k=\tilde{x}_k,\quad k=1,2,\cdots,N.\tag{2.12}$$

By the operators defined above, $a(\cdot,\cdot)$ can be rewritten as

$$a(\tilde{x},\tilde{y}) = (\tilde{x},\tilde{y})_X - (\Pi\tilde{x},\mathcal{F}\tilde{y})_X, \quad \text{for } \tilde{x},\ \tilde{y} \in X. \tag{2.13}$$

After denoting the dual operator of $\mathcal{F}$ by $\mathcal{F}^*$, we define $\mathcal{A} = \mathcal{F}^*\Pi$. Then it is easily known that the ultra weak variational formulation (2.10) is equivalent to the problem: find $\tilde{x} \in X$, such that

$$(\mathcal{I} - \mathcal{A})\tilde{x} = \tilde{b}, \tag{2.14}$$

where $\mathcal{I}$ is the unit operator.

2.3. Discrete Problem

For this subsection, we move to a consideration of the discrete system, including the construction of the approximation space.

Let $X_k \in L^2(\partial\Omega_k)$ denote a finite dimensional space and set $X_h = \prod_{k=1}^N X_k$. Substituting X_h for X, we obtain the discrete form of problem (2.10): seek $\tilde{x}_h \in X_h$ such that

$$a(\tilde{x}_h,\tilde{y}_h) = (\tilde{b},\tilde{y}_h)_X, \quad \text{for all } \tilde{y}_h \in X_h, \tag{2.15}$$

or equivalently

$$(\mathcal{I} - \mathcal{P}_h\mathcal{A})\tilde{x}_h = \mathcal{P}_h\tilde{b},$$

where $\mathcal{P}_h$ is the orthogonal projector from X onto X_h.

Once the unknown discrete impedance $\tilde{x}_h$ are calculated, the full solution can be approximated through solving local problems element by element.

Next, we formulate in detail the formation of the discrete space. For each k, we choose a finite number of nonzero functions $v_{k,l},\ l = 1,2,\cdots,P$ ($P \in \mathbb{N}$). On one hand they are supported in Ω_k. And on the other hand they are independent solutions of the homogeneous Helmholtz equation in Ω_k. Specifically, we use functions $v_{k,l}$ satisfying

$$\begin{cases} (\Delta + k_0^2 n_k^2)v_{k,l}|_{\Omega_k} = 0, & \text{and } v_{k,l} \neq 0, \\ v_{k,l}|_{\Omega_j} = 0, & \text{if } k \neq j. \end{cases}$$

And we define their impedance $\tilde{z}_{k,l}$ as

$$\begin{cases} \tilde{z}_{k,l}|_{\partial\Omega_k} = (\partial_{\tilde{\nu}_k} + \mathrm{i}\sigma)v_{k,l}|_{\partial\Omega_k}, \\ \tilde{z}_{k,l}|_{\partial\Omega_j} = 0, \quad \text{if } k \neq j. \end{cases} \tag{2.16}$$

Then the discrete space X_h is the one spanned by $\tilde{z}_{k,l}$, i.e.,

$$X_h = Span\big\{\tilde{z}_{k,l} | k = 1,2,\cdots,N,\ l = 1,2,\cdots,P\big\}.$$

In each element Ω_k the total field u satisfy the homogeneous Helmholtz equation

$$\Delta u + k_0^2 n_k^2 u = 0, \quad \text{in } \Omega_k, \tag{2.17}$$

so it could be approximated to any tolerance by plane waves [15]. Thus we take plane wave functions for $v_{k,l}$. More precisely, for $p \leq P,\ p \in \mathbb{N}$,

$$v_{k,l} = \exp \mathrm{i}k_0 n_k \boldsymbol{d}_{k,l} \cdot (\boldsymbol{x} - \boldsymbol{x}_k), \quad l = 1,2,\cdots,p, \tag{2.18}$$

where $\boldsymbol{x}_k = (x_k, y_k)$ is a point in Ω_k and $\boldsymbol{d}_{k,l}$ are the directions of plane waves satisfying $|\boldsymbol{d}_{k,l}| = 1$.

To capture the small-scale feature of the field, we also utilize evanescent wave functions of the form

$$v_{k,p+1} = \exp\big(\mathrm{i}\beta(y - y_k) + \mathrm{i}\alpha(x - x_k)\big), \tag{2.19a}$$

$$v_{k,p+2} = \exp\big(\mathrm{i}\beta(y - y_k) - \mathrm{i}\alpha(x - x_k)\big), \tag{2.19b}$$

where β on Ω_k is given by $\mathrm{i}\sqrt{\alpha^2 - k_0^2 n_+^2}$ with $|\alpha| > k_0 n_+$. Practically, they do not active in every element but only in parts of elements in $\mathbb{R}^2_+$, since the evanescent waves emerging in our model problem exponentially decay away from interface Γ_0.

3. Error Estimation and Convergence Analysis

The purpose for this section is to estimate the error and analyze the convergence rate of the method. A basic error estimate is proved, and the convergence result follows from a knowledge of the approximation properties of the bases.

For $s \in \mathbb{R}$, $H^s(E)$ denotes the usual Sobolev space on domain E with norm $||\cdot||_{s,E}$. For any positive integer k, we will also consider on $H^k(E)$ the semi-norm $|.|_{k,E}$. And let $\langle\cdot,\cdot\rangle_{\partial E}$ be the $L^2(\partial E)$ inner product or duality pairing as appropriate. By $[u]_{\Gamma_{k,j}}$ and $[\partial_{\boldsymbol{\nu}} u]_{\Gamma_{k,j}}$ we denote the jump of u and its normal derivative $\partial_{\boldsymbol{\nu}} u$ across the interface $\Gamma_{k,j}$ respectively, i.e.,

$$[u]_{\Gamma_{k,j}} = u|_{\Gamma_{k,j}} - u|_{\Gamma_{j,k}}, \quad [\partial_{\boldsymbol{\nu}} u]_{\Gamma_{k,j}} = \partial_{\boldsymbol{\nu}_{k,j}} u_k - \partial_{\boldsymbol{\nu}_{k,j}} u_j.$$

We start by proving that $a(\cdot,\cdot)$ has sufficient coercivity to provide an error estimate.

Lemma 3.1. *For $\tilde{x} \in X$, let $u = \mathcal{E}(\tilde{x})$, then it holds*

$$\begin{aligned}\Re\big(a(\tilde{x},\tilde{x})\big) = &\sum_{k<j}\Big(\sigma_{k,j}||[u]_{\Gamma_{k,j}}||^2_{0,\Gamma_{k,j}} + \frac{1}{\sigma_{k,j}}||[\partial_{\boldsymbol{\nu}} u]_{\Gamma_{k,j}}||^2_{0,\Gamma_{k,j}}\Big)\\ &+ \sum_j\Big(\sigma_j||u||^2_{0,\Gamma_j} + \frac{1}{\sigma_j}||\partial_{\boldsymbol{\nu}_j} u||^2_{0,\Gamma_j}\Big).\end{aligned} \tag{3.1}$$

Here $\Re(\cdot)$ denotes the real part of the corresponding expression.

Proof. We define the following auxiliary form

$$a_0(\tilde{x},\tilde{x}) = a(\tilde{x},\tilde{x}) - \sum_j \int_{\Gamma_j}\Big(\sigma_j|u|^2 + \frac{1}{\sigma_j}|\partial_{\boldsymbol{\nu}_j} u|^2\Big)\,\mathrm{d}s. \tag{3.2}$$

Then obviously

$$a(\tilde{x},\tilde{x}) = a_0(\tilde{x},\tilde{x}) + \sum_j \int_{\Gamma_j}\Big(\sigma_j|u|^2 + \frac{1}{\sigma_j}|\partial_{\boldsymbol{\nu}_j} u|^2\Big)\,\mathrm{d}s.$$

Since $\tilde{x}_k = (\partial_{\boldsymbol{\nu}_k} + \mathrm{i}\sigma)u_k|_{\partial\Omega_k}$, we can now formulate a_0 as follows by rewriting the definition in terms of a sum over faces in the grid.

$$\begin{aligned}a_0(\tilde{x},\tilde{x}) = \sum_{k<j}\int_{\Gamma_{k,j}}\Big\{&\mathrm{i}(u_k\partial_{\boldsymbol{\nu}_{k,j}}\bar{u}_k - \bar{u}_k\partial_{\boldsymbol{\nu}_{k,j}}u_k) + \mathrm{i}(u_j\partial_{\boldsymbol{\nu}_{j,k}}\bar{u}_j - \bar{u}_j\partial_{\boldsymbol{\nu}_{j,k}}u_j)\\ &+ \mathrm{i}(u_j\partial_{\boldsymbol{\nu}_{k,j}}\bar{u}_k + \bar{u}_j\partial_{\boldsymbol{\nu}_{k,j}}u_k) + \mathrm{i}(u_k\partial_{\boldsymbol{\nu}_{j,k}}\bar{u}_j + \bar{u}_k\partial_{\boldsymbol{\nu}_{j,k}}u_j)\\ &+ \sigma_{k,j}|[u]_{\Gamma_{k,j}}|^2 + \frac{1}{\sigma_{k,j}}|[\partial_{\boldsymbol{\nu}} u]_{\Gamma_{k,j}}|^2\Big\}\,\mathrm{d}s + \sum_j\int_{\Gamma_j}\mathrm{i}\big(u_j\partial_{\boldsymbol{\nu}_j}\bar{u}_j - \bar{u}_j\partial_{\boldsymbol{\nu}_j}u_j\big)\,\mathrm{d}s.\end{aligned} \tag{3.3}$$

And it is noticed that

$$\Re\Big(\mathrm{i}(u_j\partial_{\boldsymbol{\nu}_{k,j}}\bar{u}_k+\bar{u}_j\partial_{\boldsymbol{\nu}_{k,j}}u_k)+\mathrm{i}(u_k\partial_{\boldsymbol{\nu}_{j,k}}\bar{u}_j+\bar{u}_k\partial_{\boldsymbol{\nu}_{j,k}}u_j)\Big)=0.$$

Using this in (3.3) we obtain

$$\begin{aligned}\Re\big(a_0(\tilde{x},\tilde{x})\big)=&\sum_k\int_{\partial\Omega_k}\mathrm{i}\big(u_k\partial_{\boldsymbol{\nu}_k}\bar{u}_k-\bar{u}_k\partial_{\boldsymbol{\nu}_k}u_k\big)\,\mathrm{d}s\\&+\sum_{k<j}\int_{\Gamma_{k,j}}\Big(\sigma_{k,j}|[u]_{\Gamma_{k,j}}|^2+\frac{1}{\sigma_{k,j}}|[\partial_{\boldsymbol{\nu}}u]_{\Gamma_{k,j}}|^2\Big)\,\mathrm{d}s.\end{aligned}$$

By Green's second identity and the Helmholtz equation, it results

$$\int_{\partial\Omega_k}\Big(u_k\partial_{\boldsymbol{\nu}_k}\bar{u}_k-\bar{u}_k\partial_{\boldsymbol{\nu}_k}u_k\Big)=0,\quad k=1,2,\cdots,N.$$

Thus

$$\Re\big(a_0(\tilde{x},\tilde{x})\big)=\sum_{k<j}\int_{\Gamma_{k,j}}\Big(\sigma_{k,j}|[u]_{\Gamma_{k,j}}|^2+\frac{1}{\sigma_{k,j}}|[\partial_{\boldsymbol{\nu}}u]_{\Gamma_{k,j}}|^2\Big)\,\mathrm{d}s.$$

Next, inspired by the work of [4], we prove a basic error estimate by the duality technique, which states that the interior error of the numerical solution is controlled by the weighted sum of the the errors on the skeleton of the mesh.

Lemma 3.2. *Let $\tilde{x}\in X$ be the solution of (2.10), and $\tilde{x}_h\in X_h$ is the solution of (2.15), then for $u=\mathcal{E}(\tilde{x})$, $u_h=\mathcal{E}(\tilde{x}_h)$, it holds*

$$\begin{aligned}\|u-u_h\|_{0,\Omega}\le Ch^{-1/2}\Bigg(&\sum_{k<j}\Bigg(\sigma_{k,j}\|[u-u_h]_{\Gamma_{k,j}}\|^2_{0,\Gamma_{k,j}}\\&+\frac{1}{\sigma_{k,j}}\|[\partial_{\boldsymbol{\nu}}(u-u_h)]_{\Gamma_{k,j}}\|^2_{0,\Gamma_{k,j}}\Bigg)+\sum_j\sigma_j||u-u_h\|^2_{0,\Gamma_j}\Bigg)^{1/2}.\end{aligned}\tag{3.4}$$

The constant C is independent of u and h.

Proof. Let v satisfy the auxiliary problem

$$\begin{cases}\Delta v+k_0^2n^2(\boldsymbol{x})v=\phi, & \boldsymbol{x}\in\Omega,\\ v=0, & \boldsymbol{x}\in\Gamma.\end{cases}\tag{3.5}$$

where ϕ is an arbitrary function in $L^2(\Omega)$. Using integration by parts, and the fact that u_h satisfy the Helmholtz equation on each element, we can write

$$\begin{aligned}(u-u_h,\phi)_\Omega&=\sum_j\big(u-u_h,\Delta v+k_0^2n_j^2v\big)_{\Omega_j}\\&=\sum_j\Big(\big(\Delta(u-u_h)+k_0^2n_j^2(u-u_h),v\big)_{\Omega_j}+\big\langle(u-u_h),\partial_{\tilde{\nu}_j}v\big\rangle_{\partial\Omega_j}-\big\langle\partial_{\tilde{\nu}_j}(u-u_h),v\big\rangle_{\partial\Omega_j}\Big)\\&=\sum_j\Big(\big\langle(u-u_h),\partial_{\tilde{\nu}_j}v\big\rangle_{\partial\Omega_j}-\big\langle\partial_{\tilde{\nu}_j}(u-u_h),v\big\rangle_{\partial\Omega_j}\Big)\\&=\sum_{k<j}\Big(\big\langle[u-u_h]_{\Gamma_{k,j}},\partial_{\boldsymbol{\nu}_{k,j}}v\big\rangle_{\Gamma_{k,j}}-\big\langle[\partial_{\boldsymbol{\nu}}(u-u_h)]_{\Gamma_{k,j}},v\big\rangle_{\Gamma_{k,j}}\Big)\\&\quad+\sum_j\Big(\big\langle u-u_h,\partial_{\boldsymbol{\nu}_j}v\big\rangle_{\Gamma_j}-\big\langle\partial_{\boldsymbol{\nu}_j}(u-u_h),v\big\rangle_{\Gamma_j}\Big).\end{aligned}$$

By the boundary condition of v, we rearrange the sum on Γ to obtain

$$\begin{aligned}&(u-u_h,\phi)_\Omega\\ =&\sum_{k<j}\Big(\big\langle[u-u_h]_{\Gamma_{k,j}},\partial_{\boldsymbol{\nu}_{k,j}}v\big\rangle_{\Gamma_{k,j}}-\big\langle[\partial_{\boldsymbol{\nu}}(u-u_h)]_{\Gamma_{k,j}},v\big\rangle_{\Gamma_{k,j}}\Big)+\sum_j\big\langle(u-u_h),\partial_{\boldsymbol{\nu}_j}v\big\rangle_{\Gamma_j}.\end{aligned}$$

Hence using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}|(u-u_h,\phi)_\Omega|\le&\sum_{k<j}\Big(\|[u-u_h]_{\Gamma_{k,j}}\|_{0,\Gamma_{k,j}}\|\partial_{\boldsymbol{\nu}_{k,j}}v\|_{0,\Gamma_{k,j}}+\|[\partial_{\boldsymbol{\nu}}(u-u_h)]_{\Gamma_{k,j}}\|_{0,\Gamma_{k,j}}\|v\|_{0,\Gamma_{k,j}}\Big)\\&+\sum_j\|u-u_h\|_{0,\Gamma_j}\|\partial_{\boldsymbol{\nu}_j}v\|_{0,\Gamma_j}.\end{aligned}$$

In terms of Cauchy-Schwarz inequality again, it follows

$$\begin{aligned}&|(u-u_h,\phi)_\Omega|\\ \le&\Big\{\sum_{k<j}\big(\sigma_{k,j}\|[u-u_h]_{\Gamma_{k,j}}\|^2_{0,\Gamma_{k,j}}+\frac{1}{\sigma_{k,j}}\|[\partial_{\boldsymbol{\nu}}(u-u_h)]_{\Gamma_{k,j}}\|^2_{0,\Gamma_{k,j}}\big)+\sum_j\sigma_j\|u-u_h\|^2_{0,\Gamma_j}\Big\}^{1/2}\\&\cdot\Big\{\sum_{k<j}\big(\frac{1}{\sigma_{k,j}}\|\partial_{\boldsymbol{\nu}_{k,j}}v\|^2_{0,\Gamma_{k,j}}+\sigma_{k,j}\|v\|^2_{0,\Gamma_{k,j}}\big)+\sum_j\frac{1}{\sigma_j}\|\partial_{\boldsymbol{\nu}_j}v\|^2_{0,\Gamma_j}\Big\}^{1/2}.\end{aligned}$$

Thus we have the bound

$$\begin{aligned}&|(u-u_h,\phi)_\Omega|\\ \le&\Big\{\sum_{k<j}\big(\sigma_{k,j}\|[u-u_h]_{\Gamma_{k,j}}\|^2_{0,\Gamma_{k,j}}+\frac{1}{\sigma_{k,j}}\|[\partial_{\boldsymbol{\nu}}(u-u_h)]_{\Gamma_{k,j}}\|^2_{0,\Gamma_{k,j}}\big)\\&+\sum_j\sigma_j\|u-u_h\|^2_{0,\Gamma_j}\Big\}^{1/2}|||v|||,\end{aligned}\tag{3.6}$$

where

$$|||v|||=\Big\{\sum_{k<j}\big(\frac{1}{\sigma_{k,j}}\|\partial_{\boldsymbol{\nu}_{k,j}}v\|^2_{0,\Gamma_{k,j}}+\sigma_{k,j}\|v\|^2_{0,\Gamma_{k,j}}\big)+\sum_j\frac{1}{\sigma_j}\|\partial_{\boldsymbol{\nu}_j}v\|^2_{0,\Gamma_j}\Big\}^{1/2}.$$

We estimate $|||v|||$ using the trace estimate [16]

$$\|v\|^2_{0,\partial\Omega_k}\le C\|v\|_{0,\Omega_k}\big(h_k^{-1}\|v\|_{0,\Omega_k}+|v|_{1,\Omega_k}\big),\tag{3.7}$$

where constant $C>0$ is independent of v and h_k. Then by means of the trace theorem and the regularity of v [16], we obtain

$$\begin{aligned}|||v|||\le&C\Big\{\sum_j\big(h_j|v|^2_{2,\Omega_j}+h_j^{-1}||\nabla v||^2_{0,\Omega_j}+h_j||\nabla v||^2_{0,\Omega_j}+h_j^{-1}||v||^2_{0,\Omega_j}\big)\Big\}^{1/2}\\ \le&C\Big(\sum_j h_j^{-1}||v||^2_{2,\Omega_j}\Big)^{1/2}\le C\Big(\sum_j h_j^{-1}||\phi||^2_{0,\Omega_j}\Big)^{1/2}\le Ch^{-1/2}||\phi||_{0,\Omega},\end{aligned}$$

where constant $C>0$ is independent of v and h . We combine this with (3.6) to obtain

$$\begin{aligned}|(u-u_h,\phi)_\Omega|\le&Ch^{-1/2}\Big\{\sum_{k<j}\big(\sigma_{k,j}\|[u-u_h]_{\Gamma_{k,j}}\|^2_{0,\Gamma_{k,j}}\\&+\frac{1}{\sigma_{k,j}}\|[\partial_{\boldsymbol{\nu}}(u-u_h)]_{\Gamma_{k,j}}\|^2_{0,\Gamma_{k,j}}\big)+\sum_j\sigma_j\|(u-u_h)\|^2_{0,\Gamma_j}\Big\}^{1/2}||\phi||_{0,\Omega}.\end{aligned}\tag{3.8}$$

Finally, it results

$$\|u-u_h\|_{0,\Omega} \le Ch^{-1/2}\Big\{\sum_{k<j}\big(\sigma_{k,j}\|[u-u_h]_{\Gamma_{k,j}}\|^2_{0,\Gamma_{k,j}} + \frac{1}{\sigma_{k,j}}\|[\partial_{\boldsymbol{\nu}}(u-u_h)]_{\Gamma_{k,j}}\|^2_{0,\Gamma_{k,j}}\big) + \sum_j \sigma_j\|u-u_h\|^2_{0,\Gamma_j}\Big\}^{1/2}. \tag{3.9}$$

This completes the proof of the lemma. □

To estimate the error, we quote following result which was proved for the case $\sigma = 1$ in [11], but the proofs carry over directly to the current situation.

Lemma 3.3. *Let $\tilde{x} \in X$ be the solution of* (2.10)*, and $\tilde{x}_h \in X_h$ is the solution of* (2.15)*. Then we have*

$$\|(\mathcal{I}-\mathcal{A})(\tilde{x}-\tilde{x}_h)\|_X \le 2\|(\mathcal{I}-\mathcal{P}_h)\tilde{x}\|_X. \tag{3.10}$$

Combining above results, we could conclude following estimate showing that the error in the computational domain can be bounded by the best approximation error.

Lemma 3.4. *Let $\tilde{x} \in X$ and $\tilde{x}_h \in X_h$ be the solution of* (2.10) *and* (2.15)*, respectively. Then for $u = \mathcal{E}(\tilde{x})$ and $u_h = \mathcal{E}(\tilde{x}_h)$, it holds*

$$\|u-u_h\|_{0,\Omega} \le Ch^{-1/2}\|(\mathcal{I}-\mathcal{P}_h)\tilde{x}\|_X, \tag{3.11}$$

where the constant C is independent of u and h.

Proof. From Lemma 3.1 and Lemma 3.2, we have

$$\begin{aligned}\|u-u_h\|^2_{0,\Omega} &\le h^{-1}\Big\{\sum_{k<j}\big(\sigma_{k,j}\|[u-u_h]_{\Gamma_{k,j}}\|^2_{0,\Gamma_{k,j}} + \frac{1}{\sigma_{k,j}}\|[\partial_{\boldsymbol{\nu}}(u-u_h)]_{\Gamma_{k,j}}\|^2_{0,\Gamma_{k,j}}\big) \\ &\qquad + \sum_j \sigma_j\|u-u_h\|^2_{0,\Gamma_j}\Big\} \\ &\le Ch^{-1}\Re\big(a(\tilde{x}-\tilde{x}_h,\tilde{x}-\tilde{x}_h)\big).\end{aligned} \tag{3.12}$$

It yields by subtracting (2.10) from (2.15)

$$a(\tilde{x}-\tilde{x}_h,\tilde{y}_h) = 0, \quad \text{for all } \tilde{y}_h \in X_h.$$

Hence

$$\begin{aligned}a(\tilde{x}-\tilde{x}_h,\tilde{x}-\tilde{x}_h) &= a\big(\tilde{x}-\tilde{x}_h,(\mathcal{I}-\mathcal{P}_h)\tilde{x}\big) + a\big(\tilde{x}-\tilde{x}_h,\mathcal{P}_h\tilde{x}-\tilde{x}_h\big) \\ &= a\big(\tilde{x}-\tilde{x}_h,(\mathcal{I}-\mathcal{P}_h)\tilde{x}\big).\end{aligned} \tag{3.13}$$

In addition, it holds

$$|a(\tilde{x},\tilde{y})| = |\big((\mathcal{I}-\mathcal{A})\tilde{x},\tilde{y}\big)| \le \|(\mathcal{I}-\mathcal{A})\tilde{x}\|_X\|\tilde{y}\|_X, \quad \textit{for } \tilde{x},\ \tilde{y} \in X.$$

By Lemma 3.3, it results

$$|a(\tilde{x}-\tilde{x}_h,\tilde{x}-\tilde{x}_h)| = |a\big(\tilde{x}-\tilde{x}_h,(\mathcal{I}-\mathcal{P}_h)\tilde{x}\big)| \le 2\|(\mathcal{I}-\mathcal{P}_h)\tilde{x}\|^2_X.$$

Thus

$$\begin{aligned}\|u-u_h\|_{0,\Omega}^2 &\le Ch^{-1}\Re\big(a(\tilde{x}-\tilde{x}_h,\tilde{x}-\tilde{x}_h)\big)\\ &\le Ch^{-1}|a(\tilde{x}-\tilde{x}_h,\tilde{x}-\tilde{x}_h)| \le Ch^{-1}\|(\mathcal{I}-\mathcal{P}_h)\tilde{x}\|_X^2.\end{aligned}$$

This completes the proof of the lemma. □

Next we analyze the convergence rate of the method under the following additional assumptions:

- Ω is convex;
- each element Ω_k of $\mathcal{T}_h$ is a convex Lipschitz domain;
- there is a constant $\rho \in (0,1)$ such that each element $\Omega_k \in \mathcal{T}_h$ contains a ball of radius ρh_k;
- there is a constant $\mu \in (0,1)$ such that for each $\Omega_k \in \mathcal{T}_h$, $h_k \ge \mu h$.

Then we cite the following technical result proved in [15] with the help of Vekua theory and approximation estimates for harmonic polynomials.

Lemma 3.5. *Let $D \in \mathbb{R}^2$ be a bounded, convex Lipschitz domain of diameter h_D which contains a ball of radius rh_D with $0 < r \le 1/2$. Let $u \in H^{s+1}(D)$, $s \ge 0$, be the solution of $\Delta u + \omega^2 u = 0$ with constant $\omega > 0$. Fix $q \ge 1$, $q \in \mathbb{N}$, set $p = 2q+1$ and let the directions $\{\boldsymbol{d}_l = (\cos\theta_l, \sin\theta_l)\}_{l=1}^p$ satisfy the condition that there exists $\delta \in (0,1]$ such that the minimum angles between two different directions is greater than or equal to $2\pi\delta/p$. Then, for $q \ge 2s+1$ there exist $\alpha_1, \alpha_2, \cdots, \alpha_p \in \mathbb{C}$, such that for every $0 \le j \le s+1$, it holds*

$$\begin{aligned}&\|u - \sum_{l=1}^{p} \alpha_l \exp(\mathrm{i}\omega \boldsymbol{x}\cdot\boldsymbol{d}_l)\|_{j,\omega,D}\\ \le\ & C\big[1+(\omega h_D)^{q+j-s+8}\big]\exp\big((7/4-3/4r)\omega h_D\big)h^{s+1-j}\\ &\cdot\Big[\Big(\frac{2\log(q+2)}{q+2}\Big)^{s+1-j} + 4\exp(-5)r\delta^4(q+1)^{-q/2}\Big]\|u\|_{s+1,\omega,D},\end{aligned} \tag{3.14}$$

where the constant $C > 0$ depends only on j, s and the shape of D, but not on h_D, ω, p, δ and u. Here $\|\cdot\|_{j,\omega,D}$ is the ω-weighted Sobolev norm.

It is noticed that $\log(q+2)/q+2$ asymptotically behaves like $\log p/p$ for increasing q, whereas the term $(q+1)^{-q/2}$ decays faster. Therefore, the estimate, for large p, can be expressed as

$$\Big\|u - \sum_{l=1}^{p} \alpha_l \exp(\mathrm{i}\omega \boldsymbol{x}\cdot\boldsymbol{d}_l)\Big\|_{j,\omega,D} \le Ch_D^{s+1-j}\Big(\frac{\log(p)}{p}\Big)^{s+1-j}\|u\|_{s+1,\omega,D}. \tag{3.15}$$

We indicate that, thanks to the assumptions for the domain Ω and the mesh $\mathcal{T}_h$, we can apply the above lemma to the element $\Omega_k \in \mathcal{T}_h$ to get the following convergence result of our method.

Theorem 3.1. *Let $u \in \prod_j H^{s+1}(\Omega_j)$ and $\tilde{x}_h \in X_h$ be the solution of (2.9) and (2.10), respectively. Then for $u_h = \mathcal{E}(\tilde{x}_h)$ and large p, it holds*

$$\|u-u_h\|_{0,\Omega} \le Ch^{s-1}\Big(\frac{\log(p)}{p}\Big)^{s-1/2}\|u\|_{s+1}, \tag{3.16}$$

where $\|u\|_{s+1}^2 = \sum_{j=1}^N \|u\|_{s+1,k_0n_j,\Omega_j}^2$, *and the constant* C *is independent of* u, p *and* h, *but relies on* σ.

Proof. By Lemma 3.5, there is a function $u_a \in V_h$ such that in each element Ω_k it holds

$$\|u-u_a\|_{j,\omega,\Omega_k} \;\le\; Ch^{s+1-j}\Big(\frac{\log(p)}{p}\Big)^{s+1-j}\|u\|_{s+1,k_0n_k,\Omega_k}. \tag{3.17}$$

with $0 \le j \le s+1$.

Then define $\tilde{x}_a$ by $\tilde{x}_a|_{\partial\Omega_j} = \big((\partial_{\tilde{\nu}_j} + \mathrm{i}\sigma)u_a|_{\Omega_j}\big)|_{\partial\Omega_j}$, $j = 1, 2, \cdots, N$. We use the trace estimate (3.7) again and combine it with above estimate to obtain

$$\begin{aligned}
\|x-x_a\|_X^2 &= \sum_j \int_{\partial\Omega_j} \frac{1}{\sigma}|(\partial_{\tilde{\nu}_j} + \mathrm{i}\sigma)(u-u_a)|^2 \, ds \\
&\le \sum_j \left(\frac{1}{\sigma}\|\partial_{\tilde{\nu}_j}(u-u_a)\|_{0,\partial\Omega_j}^2 + \|u-u_a\|_{0,\partial\Omega_j}^2\right) \\
&\le \sum_k Ch_k^{2s-1}\Big(\frac{\log(p)}{p}\Big)^{2s-1}\|u\|_{s+1,k_0n_k,\Omega_k} \\
&\le Ch^{2s-1}\Big(\frac{\log(p)}{p}\Big)^{2s-1}\|u\|_{s+1}^2.
\end{aligned} \tag{3.18}$$

Thus, it holds

$$\|(\mathcal{I}-\mathcal{P}_h)x)\|_X \le \|x-x_a\|_X \le Ch^{s-1/2}\Big(\frac{\log(p)}{p}\Big)^{s-1/2}\|u\|_{s+1}.$$

Finally in terms of Lemma 3.4, it results

$$\|u-u_h\|_{0,\Omega} \le Ch^{s-1}\Big(\frac{\log(p)}{p}\Big)^{s-1/2}\|u\|_{s+1}.$$

This completes the proof of the theorem. □

Unfortunately, we can not indicate the dependence of the convergence rate on the number of evanescent wave functions. This is due to the fact that their approximation properties are not available at present as far as we know. But notice that they are used only in parts of the elements, thus the convergence result is still reasonable.

4. Numerical Results

In this part, we demonstrate the numerical results of our method on case of high-frequency. All simulations are performed under MATLAB2010b.

For simplicity, the computational domain is taken as a square $\Omega = [-l, l] \times [-l, l]$ and the mesh consists of uniformly small squares with side length l/n, $n \in \mathbb{N}$. It is easily seen that the parameter n determine the fineness of the mesh.

In our experiment, we set $l = 1/2$, and the free space wave number $k_0 = 1$, $n_+ = 45$, $n_- = 75$, $n_s = 60$. The incident field is plane wave $u^i = \exp(\mathrm{i}\alpha x + \mathrm{i}\eta y)$ with incident angle $\theta = \pi/4$.

In each element Ω_k, we choose p propagation directions $\boldsymbol{d}_{k,l}$ of plane waves as follows

$$\boldsymbol{d}_{k,r} = (\cos\theta_r, \sin\theta_r), \quad \theta_r = 2\pi(r-1)/p, \quad r = 1, 2, \cdots, p.$$

And we use two evanescent wave functions of the form (2.19) to enrich the plane wave functions, which propagate along the substrate in positive and negative x directions respectively. The point $\boldsymbol{x}_k = (x_k, y_k)$ is taken as the center of the element Ω_k, $k = 1, 2, \cdots, N$.

To test the accuracy and convergence of the method, we calculate the reference field u_{ph}^{ref} caused only by the two-layered background medium. Then the boundary value problem to be solved is following

$$\begin{cases} \Delta u + k_0^2 m^2(\boldsymbol{x}) u = 0, & \boldsymbol{x} \in \Omega, \\ \big(\partial_{\boldsymbol{\nu}} - \mathrm{i}k_0 m(\boldsymbol{x})\big) u = g, & \boldsymbol{x} \in \Gamma, \end{cases}$$

where $g = \big(\partial_{\boldsymbol{\nu}} - \mathrm{i}k_0 m(\boldsymbol{x})\big) u^{ref}$. Since its exact solution is just u^{ref} given by (2.2), there is no modeling error for this case. Our numerical error is only caused by the algorithm. We calculate errors between the real solution u^{ref} and the numerical one u_{ph}^{ref} with L^2 norm on Ω i.e.

$$error = \Big(\sum_{k=1}^{N} ||u^{ref} - u_{ph}^{ref}||_{0,\Omega_k}^2 \Big)^{1/2},$$

and on each Ω_k we evaluate the integral by the numerical quadrature formula.

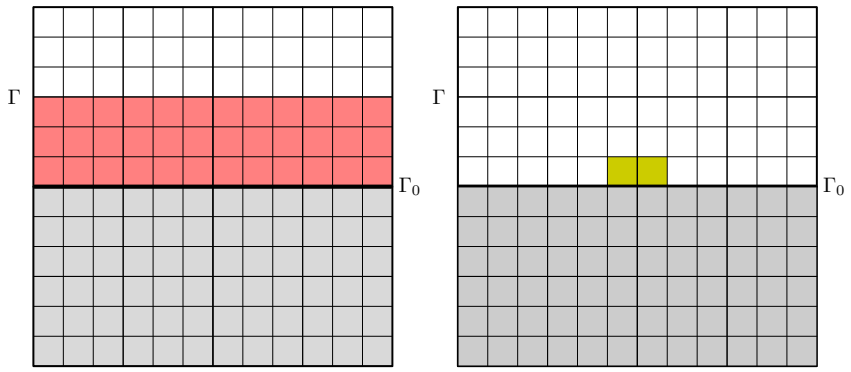

Fig. 4.1. Left plot: Real part of reference field u_h^{ref}. Right plot: Real part of sacttering field u_{ph}^s.

We first focus on the case of mesh refinement. Five meshes are used with $n = 2$ for the coarse mesh and $n = 16$ for the finest one. Since evanescent wave functions exponentially decay as a function of depth in y direction, they are used only in the elements in $\Omega_e = [-l, l] \times [0, l/2]$ (see the red domain in the left plot of Fig. 4.1). We present the L^2 errors on Ω for $p = 10$ in Table 4.1. It is seen that errors go down with mesh refinement but at a slow rate.

Table 4.1: Errors for mesh refinement for $p = 10$.

n	2	4	8	12	16
error	5.47e-001	1.05e-001	3.08e-002	2.52e-003	4.54e-004

Next, we move to a consideration of convergence as the number of bases increase on a fixed mesh. We vary p from 5 to 50 and keep evanescent wave functions unchanged. Errors with respect to the number of plane wave functions for $n = 3$ and $n = 6$ are presented respectively

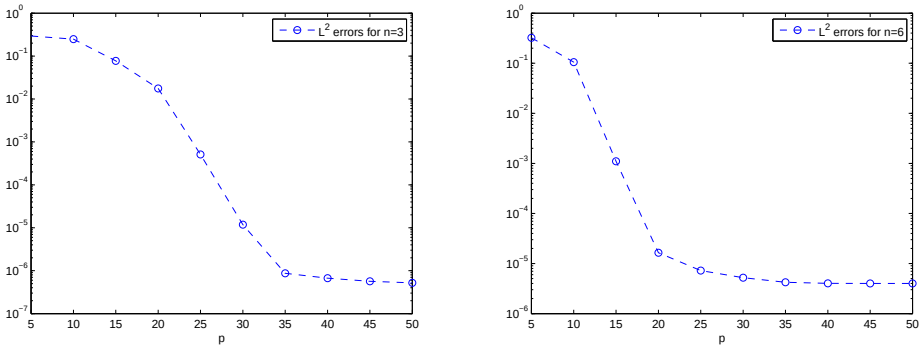

Fig. 4.2. Left plot: L^2 errors versus p for $n = 3$. Right plot: L^2 errors versus p for $n = 6$.

in Fig. 4.2. It can be observed that the error decreases rapidly when p increase. And high accuracy can be achieved at a low density of degree of freedom (eg. $p = 30$ for both cases).

To investigate the effectiveness of evanescent wave functions, we compare the errors computed on the zone Ω_e where they are active. Table 4.2 lists the corresponding numerical data, in which *err* represents the errors counted by both plane wave and evanescent wave bases and *Err* stands for the errors calculated only with plane waves. By comparison, we conclude that a better accuracy is attained by use of evanescent wave functions.

Table 4.2: Comparison of the errors on Ω_e for $n = 3$.

p	10	15	20	25	30
err	6.472e-002	9.628e-003	1.163e-004	7.806e-005	8.750e-007
Err	2.918e-001	6.684e-002	2.314e-003	4.353e-004	5.333e-006

Furthermore, we carry out preliminary tests about the influence of the number of evanescent wave functions on the convergence rate. The evanescent wave functions used are of the form

$$\begin{aligned} v_{k,p+2l-1} &= \exp\big(\mathrm{i}\beta_l(y-y_k) + \mathrm{i}\alpha_l(x-x_k)\big) \\ v_{k,p+2l} &= \exp\big(\mathrm{i}\beta_l(y-y_k) - \mathrm{i}\alpha_l(x-x_k)\big), \quad l = 1,2,3,4,5, \end{aligned} \tag{4.1}$$

where β_l on Ω_k is given by $\mathrm{i}\sqrt{\alpha_l^2 - k_0^2 n_+^2}$ with $\alpha_l = lk_0 n_- \sin\theta$.

We still focus on the L^2 errors on the region of Ω_e with the case $n = 3$. Since the approximate space is spanned by two kinds of wave functions, we add two evanescent wave functions each time while let the number of plane wave functions $p = 10$ remains invariant. The numerical data is listed in the following table. From the results, we observe that the errors go down with the increase of the the number of evanescent wave functions gradually. Thus, we speculate that the convergence rate may be influenced by the number of evanescent wave functions.

Table 4.3: Errors against l for $p = 10$ on Ω_e.

l	1	2	3	4	5
error	6.472e-002	3.824e-002	8.177e-003	5.643e-003	9.047e-004

Finally, we simulate the situation that a sample $S = [-l/6, l/6] \times [0, l/6]$ appears in the middle of the substrate (see the yellow domain in the right plot of Fig. 4.1). Due to its presence , the interaction between the reference field and the sample produces scattering waves. When

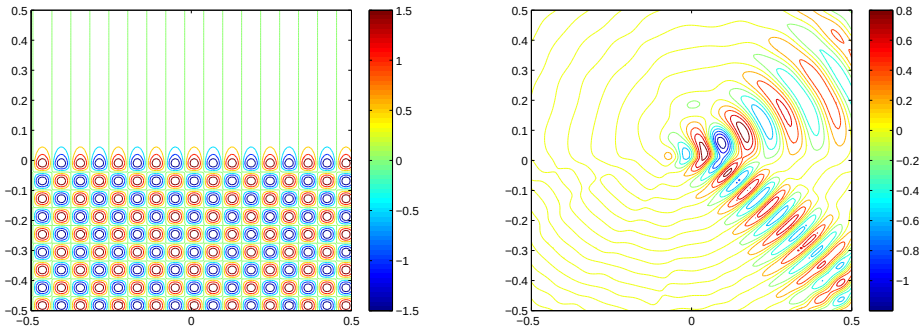

Fig. 4.3. Left plot: Contours of reference field u_{ph}^{ref}. Right plot: Contours of scattering field u_{ph}^{s}.

the total field u_{ph} is calculated, the scattering field u_{ph}^{s} can be obtained by $u_{ph}^{s} = u_{ph} - u_{ph}^{ref}$. We depict the contours of the real part of u_{ph}^{ref} and u_{ph}^{s} in Fig. 4.3 (see the left plot and the right one respectively). Physically, the sample will give rise to refraction, thus it should be an enforcement of the field in both directions of reflection and refraction. Obviously, our numerical results are in good accordance with this phenomenon, which testify the validity of our method again.

It is clear from our results that using evanescent wave functions to enrich plane wave functions is an effective way to improve accuracy.

5. Conclusions

For this work, we demonstrate an efficient and accurate numerical method for a scattering problem in near field optics. Using evanescent wave and plane wave functions to approximate the local properties of the field, we can get a fast global convergence. Furthermore, since the priori information about the solution is directly incorporated in the trial space, it can deal well with the case with high-frequency. The superiority of our method is mainly due to the good approximation properties of bases functions and the fact that various integrals can be evaluated in a closed form.

Acknowledgments. The authors would like to thank the reviewers and Dr.Zheng Enxi for many valuable suggestions. This work is supported by the National Natural Science Foundation of China (Grant No. 11371172, 51178001), Science and technology research project of the education department of Jilin Province (Grant No. 2014213).

References

[1] E. Betzig, J.K. Trautman, Near-field optics: microscopy, spectroscopy, and surface modification beyond the diffraction limit, *Science*, **257** (1992), 189-195.

[2] D. Courjon, C. Bainier, Near field microscopy and near field optics, *Rep. Pro. Phys.*, **57** (1994), 989-1028.

[3] P. Carney, J. Schotland, Near-field tomography, in *Inside out: Inverse Problems and Application*, G. Uhlmann (ed.), Cambridge U. Press, England, 2003, pp. 133-168.

[4] P. Monk, Q.D. Wang, A leaset-squares method for the Helmholtz equation, *Comput. Meths. Appl. Mech. Engrg.*, **175** (1999), 121-136.

[5] A.H. Barnett , T. Betcke, An exponentially convergent nonpolynomial finite element method for time-harmonic scattering from polygons, *J. Sci. Comput.*, **32** (2011), 1417-1441.
[6] E.X. Zheng, F.M. Ma and D.Y. Zhang, A least-squares non-polynomial finite element method for solving the polygonal-line grating problem, *J. Math. Anal. Appl.*, **397** (2013), 550-560.
[7] J. Melenk, I. Babuška, The partition of unity finite element method: basic theory and applications, *Comput. Meth. Appl. Mech. Eng.*, **139** (1996), 289-314.
[8] I. Babuska, J.M. Melenk, The partition of unity method, *Int. J. Numer. Meths. Eng.*, **40** (1997), 727-758.
[9] C.J. Gittelson, R. Hiptmair and I. Perugia, Plane wave discontinous Galerkin methods: analysis of the h-version, *Math. Model. Numer. Anal.*, **43** (2009), 297-332.
[10] R. Hiptmair, A. Moilor and L. Perugia, Plane wave discontinous Galerkin method for the 2D Helmholtz equation: analysis of the p-version, *J. Numer. Anual.*, **49** (2011), 264-284.
[11] O. Cessenat, B. Després, Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problems, *J. Numer. Anual.*, **35** (1998), 255-299.
[12] P.J. Li, Numerical simulations of global approach for photon scanning tunneling microscopy: coupling of finite-element and boundary integral methods, *J. Opt. Soc. Am. A*, **25** (2008), 1929-1936.
[13] T. Huttunen, P. Monk and J.P. Kaipio, Computational aspects of the ultra weak variational formulation, *J. Comput. Phys.*, **182** (2002), 27-46.
[14] T. Huttunen, J.P. Kaipio and P. Monk, The perfectly matched layer for the ultra weak variational formulation of the 3D Helmholtz equation, *Int. J. Numer. Meth. Eng.*, **61** (2004), 1072-1092.
[15] A. Moiola, R. Hiptmair and I. Perugia, Plane wave approximation of homogeneous Helmholtz solutions, *Z. Angew. Math. Phys.*, **62** (2011), 809-837.
[16] S.C. Brenner, L.R. Scott, The Mathematical Theory of Finite Element Methods, Springer Verlag, New York, 2002.