A NOVEL LEAST SQUARES METHOD FOR HELMHOLTZ EQUATIONS WITH LARGE WAVE NUMBERS

QIYA HU AND RONGLONG SONG

Abstract. In this paper we are concerned with numerical methods for non-homogeneous Helmholtz equations in inhomogeneous media. We design a least squares method for discretization of the considered Helmholtz equations. In this method, an auxiliary unknown is introduced on the common interface of any two neighboring elements and a quadratic subject functional is defined by the jumps of the traces of the solutions of local Helmholtz equations across all the common interfaces, where the local Helmholtz equations are defined on elements and are imposed Robin-type boundary conditions given by the auxiliary unknowns. A minimization problem with the subject functional is proposed to determine the auxiliary unknowns. The resulting discrete system of the auxiliary unknowns is Hermitian positive definite and so it can be solved by the PCG method. Under some assumptions we show that the generated approximate solutions possess almost the optimal error estimates with little "wave number pollution". Moreover, we construct a substructuring preconditioner for the discrete system of the auxiliary unknowns. Numerical experiments show that the proposed methods are very effective for the tested Helmholtz equations with large wave numbers.

Key words. Helmholtz equations, inhomogeneous media, large wave number, auxiliary unknowns, least squares, error estimates, preconditioner

AMS subject classifications. 65N30, 65N55.

1. Introduction

For simplicity of exposition, we only consider two dimensional problem in this paper. Let Ω be a bounded, connected and Lipschitz domain in \( \mathbb{R}^2 \). Consider the Helmholtz equations

\[
\begin{aligned}
- \Delta u - \kappa^2 u &= f & \text{in } \Omega, \\
\frac{\partial u}{\partial \mathbf{n}} + i\kappa u &= g & \text{on } \partial \Omega,
\end{aligned}
\]

where \( \mathbf{n} \) denotes the unit outward normal on the boundary \( \partial \Omega \) and \( \kappa \) is the wave number defined by \( \kappa(x) = \frac{\omega}{c(x)} > 0 \), with \( \omega > 0 \) being a constant and \( c(x) \) being a bounded and positive function defined on \( \Omega \). In applications, \( \omega \) denotes the angular frequency, which may be very large, and \( c(x) \) denotes the wave speed (the acoustic velocity), which may not be a constant function on \( \Omega \), i.e., the involved media is inhomogeneous.

Helmholtz equation is the basic model in sound propagation. It is a very important topic to design a high accuracy method for Helmholtz equations with large wave numbers.
wave numbers, such that the so called “wave number pollution” can be reduced. In recent years, many interesting methods for the discretization of Helmholtz equations with large wave numbers have been proposed, for example (but not all), the first-order system least-squares (FOSLS) method [31], the ultra weak variational formulation (UWVF) [3], the plane wave least squares (PWLS) methods [24,25,33], the plane wave discontinuous Galerkin (PWDG) methods [17,22], the method of fundamental solutions [2,5], the plane wave method with Lagrange multipliers [12], the variational theory of complex rays [36], the high order element discontinuous Galerkin method (HODG) [10,14], local discontinuous Galerkin method (LDG) [15], hybridizable discontinuous Galerkin method (HDG) [4,21] and the discontinuous Petrov-Galerkin method [7,40], the ray-based finite element method [11] and the generalized plane wave method [30]. All these methods are superior to the standard linear finite element method.

It is well known that the plane wave methods almost have no “wave number pollution” and can generate higher accuracy approximations than the other methods for solving the Helmholtz equations with large (piecewise constant) wave numbers. Unfortunately, the plane wave methods cannot be directly applied to the discretization of nonhomogeneous Helmholtz equations in inhomogeneous media. A plane wave method combined with local spectral element for nonhomogeneous Helmholtz equations in homogeneous media was proposed in [26] (see also [25]). A generalized plane wave method for homogeneous Helmholtz equations in inhomogeneous media was introduced in [30]. In the present paper, we try to design a novel discretization method for nonhomogeneous Helmholtz equations in inhomogeneous media such that the method almost has no “wave number pollution” and possesses some other nice features.

The basic ideas of the new method can be roughly described as follows. We introduce an auxiliary unknown ($q$ order polynomial) on the common interface of any two neighboring elements, and solve local discrete Helmholtz equations of $p$ order polynomials ($p \geq q + 2$) on all the elements in parallel, where each local Helmholtz equation is defined on an element and is imposed a Robin-type boundary condition given by the auxiliary unknowns. We define a minimization problem with a quadratic subject functional defined by the jumps of the traces of the solutions of the local Helmholtz equations across all the common interfaces. This minimization problem results in a Hermitian positive definite algebraic system of the auxiliary unknowns. After solving the algebraic system, we can easily obtain an approximate solution of the original Helmholtz equation.

The new method possesses the following merits: (i) it is cheap since only one unknown is introduced in each element or on each element interface (the unknowns needed to be globally solved are defined on all the interfaces and have less degrees of freedom); (ii) the method is easy to implement since the local problem on each element is directly defined by the original Helmholtz equation of the second order and the basis functions on both the elements and the element interfaces are standard; (iii) the resulting algebraic system is Hermitian positive definite, so it can be solved by the PCG method, which has stable convergence and less cost of calculation.

We show that the proposed method possesses almost the optimal error estimates with little “wave number pollution” under suitable assumptions. Besides, we construct a domain decomposition preconditioner for solving the resulting algebraic
system. Numerical results indicate that the proposed discretization method and preconditioner are very efficient for the tested Helmholtz equations with large wave numbers.

The paper is organized as follows: In Section 2, we describe the proposed least squares variational formulation for Helmholtz equations. In Section 3, we construct a substructuring preconditioner for the discrete system. The main results about error estimates are presented in Section 4. In Section 5, we give proofs of the main results in details. Finally, we report some numerical results to confirm the effectiveness of the new method in Section 6.

2. A LEAST SQUARES VARIATIONAL FORMULATION

2.1. Notations. As usual we partition Ω into elements in the sense that

$$\bar{\Omega} = \bigcup_{k=1}^{N} \bar{\Omega}_k, \quad \Omega_k \cap \Omega_j = \emptyset, \quad \text{for } k \neq j.$$  

Here each Ω_k may be curve polyhedron. We use h_k to denote the diameter of Ω_k and set h = max{h_k}. Let T_h denote the partition comprised of elements \{Ω_k\}_{k=1}^N.

As usual we assume that the partition T_h is quasi-uniform and regular.

Let γ_kj denote the common edge of two neighboring elements Ω_k and Ω_j, and set γ_k = ∂Ω_k ∩ ∂Ω when the intersection is an edge of the element Ω_k. For convenience, define γ = ∪_{k≠j} γ_kj.

Let q ≥ 1 be an integer and choose p ≥ q + 2. Throughout this paper we use the following notations:

• V_h^p(Ω_k) = \{v ∈ H^1(Ω_k) : v is a polynomial whose order does not exceed p\}.
• V_h^p(T_h) = \bigcap_{k=1}^{N} V_h^p(Ω_k).
• V_h^p(∂Ω_k) = \{v|_{∂Ω_k} : v ∈ V_h^p(Ω_k)\}.
• W(γ) = \prod_{k≠j} H^{-\frac{1}{2}}(γ_kj).
• W_h^p(γ_kj) = \{\mu ∈ H^1(γ_kj) : \mu is a polynomial whose order does not exceed q\}.
• W_h^p(γ) = \prod_{k≠j} W_h^p(γ_kj).
• W_h^p(∂Ω_k ∩ ∂Ω) = \{μ|_{∂Ω_k ∩ ∂Ω} : μ ∈ W_h^p(γ)\}.
• The jump of v across γ_kj: [v] = v_k - v_j, where v is a piecewise smooth function on T_h and v_k = v|_{Ω_k}.
• (u, v)_{Ω_k} = \int_{Ω_k} u \cdot v dx, \quad (u, v)_{∂Ω_k} = \int_{∂Ω_k} u \cdot v ds.

2.2. A continuous variational formulation. For each element Ω_k, set u|_{Ω_k} = u_k. For k > j, define λ ∈ W(γ) as

$$\lambda|_{γ_{kj}} = (\frac{\partial u_k}{\partial n_k} + i ρ u_k)|_{γ_{kj}} = (\frac{\partial u_j}{\partial n_j} - i ρ u_j)|_{γ_{kj}},$$

where ρ > 0, n_k and n_j separately denote the unit outward normal on ∂Ω_k and ∂Ω_j. It is clear that the solution u of (1.1) satisfies the local Helmholtz equation on each element Ω_k (k = 1, ..., N):

$$\begin{align*}
- Δ u_k - κ^2 u_k &= f \quad \text{in } Ω_k, \\
\frac{∂ u_k}{∂ n_k} &\pm i ρ u_k = λ \quad \text{on } ∂Ω_k \setminus ∂Ω, \\
\frac{∂ u_k}{∂ n_k} + i κ u_k &= g \quad \text{on } ∂Ω_k \cap ∂Ω. 
\end{align*}$$

(2.1)
For each element \( \Omega_k \), define the local sesquilinear form
\[
a^{(k)}(v, w) = \langle \nabla v, \nabla w \rangle_{\Omega_k} - \langle \kappa^2 v, w \rangle_{\partial \Omega_k \setminus \partial \Omega} + i\langle \kappa v, w \rangle_{\partial \Omega_k \cap \partial \Omega}, \quad v, w \in H^1(\Omega_k)
\]
and the local functional
\[
L^{(k)}(\varpi) = \langle f, \varpi \rangle_{\Omega_k} + \langle g, \varpi \rangle_{\partial \Omega_k \cap \partial \Omega}, \quad v \in H^1(\Omega_k).
\]
It is easy to see that the variational formulation of (2.1) is: to find \( u_k(\lambda) \in H^1(\Omega_k) \) such that
\[
a^{(k)}(u_k(\lambda), v) = L^{(k)}(v) + \langle \lambda, v \rangle_{\partial \Omega_k \setminus \partial \Omega}, \quad \forall v \in H^1(\Omega_k).
\]
(2.2)

We define the quadratic functional
\[
J(\mu) = \sum_{\gamma_{kj}} \int_{\gamma_{kj}} |u_k(\mu) - u_j(\mu)|^2 ds, \quad \mu \in W(\gamma)
\]
and consider the following minimization problem: find \( \lambda \in W(\gamma) \) such that
\[
J(\lambda) = \min_{\mu \in W(\gamma)} J(\mu).
\]
(2.4)

It is clear that \( u \) is the solution of (1.1) if and only if \( J(\lambda) = 0 \), which means that \( \lambda \) is the solution of the minimization problem (2.4).

In order to give the variational problem of (2.4), we write the solution of (2.1) as \( u_k(\lambda) = u_k^{(1)}(\lambda) + u_k^{(2)}(\lambda) \), which respectively satisfy
\[
a^{(k)}(u_k^{(1)}(\lambda), v) = \langle \lambda, v \rangle_{\partial \Omega_k \setminus \partial \Omega}, \quad \forall v \in H^1(\Omega_k)
\]
and
\[
a^{(k)}(u_k^{(2)}(\lambda), v) = L^{(k)}(v), \quad \forall v \in H^1(\Omega_k).
\]

Then \( J(\mu) \) can be written as
\[
J(\mu) = \sum_{\gamma_{kj}} \int_{\gamma_{kj}} |(u_k^{(1)}(\mu) - u_j^{(1)}(\mu)) + (u_k^{(2)}(\mu) - u_j^{(2)}(\mu))|^2 ds.
\]

Define the sesquilinear form
\[
s(\lambda, \mu) = \sum_{\gamma_{kj}} \int_{\gamma_{kj}} (u_k^{(1)}(\lambda) - u_j^{(1)}(\lambda)) \cdot (u_k^{(1)}(\mu) - u_j^{(1)}(\mu)) ds, \quad \lambda, \mu \in W(\gamma)
\]
and the functional
\[
l(\mu) = -\sum_{\gamma_{kj}} \int_{\gamma_{kj}} (u_k^{(2)}(\mu) - u_j^{(2)}(\mu)) \cdot (u_k^{(1)}(\mu) - u_j^{(1)}(\mu)) ds, \quad \mu \in W(\gamma).
\]

Therefore the variational problem of the minimization problem (2.4) can be expressed as follows: find \( \lambda \in W(\gamma) \) such that
\[
s(\lambda, \mu) = l(\mu), \quad \forall \mu \in W(\gamma).
\]
(2.5)
2.3. The discrete variational formulation. Let $\lambda_h \in W^q_h(\gamma)$. For each element $\Omega_k$, define $u_{h,k}(\lambda_h) \in V^p_h(\Omega_k)$ by
\[ a^{(k)}(u_{h,k}(\lambda_h), v_h) = L^{(k)}(v_h) + \langle \lambda_h, v_h \rangle_{\partial \Omega_k \setminus \partial \Omega}, \quad \forall v_h \in V^p_h(\Omega_k). \] (2.6)
It is easy to see that the above problem is uniquely solvable.

As in the continuous situation, we decompose $u_{h,k}$ into $u_{h,k} = u_{h,k}^{(1)}(\lambda_h) + u_{h,k}^{(2)}$, which are respectively defined by
\[ a^{(k)}(u_{h,k}^{(1)}(\lambda_h), v_h) = \langle \lambda_h, v_h \rangle_{\partial \Omega_k \setminus \partial \Omega}, \quad \forall v_h \in V^p_h(\Omega_k) \]
and
\[ a^{(k)}(u_{h,k}^{(2)}, v_h) = L^{(k)}(v_h), \quad \forall v_h \in V^p_h(\Omega_k). \]

Define the discrete sesquilinear form
\[ s_h(\lambda_h, \mu_h) = \sum_{\gamma_{kj}} \int_{\gamma_{kj}} (u_{h,k}^{(1)}(\lambda_h) - u_{h,j}^{(1)}(\lambda_j)) \cdot \frac{(u_{h,k}^{(1)}(\mu_h) - u_{h,j}^{(1)}(\mu_j))}{ds, \quad \lambda_h, \mu_h \in W^q_h(\gamma) \] and the functional
\[ l_h(\mu_h) = - \sum_{\gamma_{kj}} \int_{\gamma_{kj}} (u_{h,k}^{(2)} - u_{h,j}^{(2)}) \cdot \frac{(u_{h,k}^{(1)}(\mu_h) - u_{h,j}^{(1)}(\mu_j))}{ds, \quad \mu_h \in W^q_h(\gamma). \]

Therefore the discrete variational problem of (2.5) can be written as follows: find $\lambda_h \in W^q_h(\gamma)$ such that
\[ s_h(\lambda_h, \mu_h) = l_h(\mu_h), \quad \forall \mu_h \in W^q_h(\gamma). \] (2.7)

After $\lambda_h$ is solved from (2.7), we can easily compute $u_{h,k}$ in parallel by (2.6) for every $\Omega_k$. Define $u_h \in V^q_h(\Omega_h)$ by $u_h|_{\Omega_k} = u_{h,k}(\lambda_h)$ ($k = 1, \ldots, N$). Then $u_h$ should be an approximate solution of $u$. We would like to emphasize the discrete system (2.7) has relatively less degrees of freedom, so it is cheaper to be solved.

Let $S$ be the stiffness matrix associated with the sesquilinear form $s_h(\cdot, \cdot)$, and let $b$ denote the vector associated with $l_h(\cdot)$. Then the discretization problem (2.7) leads to the algebraic system
\[ SX = b. \] (2.8)

From the definition of the sesquilinear form $s_h(\cdot, \cdot)$, we know that the matrix $S$ is Hermitian positive definite, so the system (2.8) can be solved by the preconditioned CG method with a positive definite preconditioner (see the next section).

Remark 2.1. As in the traditional Lagrange multiplier method, we can derive another discrete system of $\lambda_h$ by the constraints (for all element interfaces $\gamma_{kj}$)
\[ \langle u_{h,k} - u_{h,j}, \mu \rangle_{\gamma_{kj}} = 0, \quad \forall \mu \in W^q_h(\gamma). \]

However, the coefficient matrix of the resulting system is not positive definite yet, which makes the solution of the system to be more difficult.

3. A domain decomposition preconditioner

In this section, we are devoted to the construction of a preconditioner $K$ for $S$. Since $S$ is Hermitian and positive definite, we can construct a (positive definite) substructuring preconditioner absorbing some ideas in the BDDC method first introduced in [5]. As we will see, the preconditioner designed in this section has essential differences from the one defined in the standard BDDC method.
For convenience, we will define the preconditioner in operator form. To this end, let $S : W_h^q(\gamma) \rightarrow W_h^q(\gamma)$ denote the discrete operator corresponding to the stiffness matrix $S$, i.e.,

$$(S\lambda_h, \mu_h) = s_h(\lambda_h, \mu_h), \quad \forall \lambda_h, \mu_h \in W_h^q(\gamma).$$

As usual we coarsen the partition as follows: let $\Omega$ be decomposed into a union of $D_1, D_2, \ldots, D_n$ such that $D_r$ is just a union of several elements $\Omega_k \in T_h$ and satisfies

$$\Omega = \bigcup_{r=1}^{n_0} D_r, \quad D_r \cap D_l = \emptyset \quad \text{for} \ r \neq l.$$ 

Let $d$ denote the size of the subdomains $D_1, D_2, \ldots, D_n$, and let $T_d$ denote the partition comprised of the subdomains $\{D_r\}_{r=1}^{n_0}$.

In order to explain our ideas, we investigate basis functions associated with two neighboring subdomains $D_r$ and $D_l$, which have the non-empty common part $\partial D_r \cap \partial D_l$. Let $e$ and $e'$ be two fine edges that satisfy $e \in D_r \setminus (\partial D_r \cap \partial D_l)$ and $e' \in D_l \setminus (\partial D_r \cap \partial D_l)$, and let $\mu_e$ and $\mu_{e'}$ denote two basis functions on $e$ and $e'$ respectively. It can be checked that, if $e$ and $e'$ close to $\partial D_r \cap \partial D_l$, then $\mu_e$ and $\mu_{e'}$ still have coupling, i.e., $s_h(\mu_e, \mu_{e'}) \neq 0$. This means that, if the interface is defined in the standard manner, namely, is defined as the union of all $\partial D_r \cap \partial D_l$, the degrees of freedom in subdomain interiors cannot be eliminated independently. According to this observation, in the current situation an interface should be defined as a union of some elements instead of a union of some edges.

For each $D_r$, let $D_r^b \subset D_r$ be a union of the elements that touch the right and the lower boundary of $\partial D_r \setminus \partial \Omega$ (refer to Figure 1). We define an interface as

$$\Gamma = \bigcup_{r=1}^{n_0} D_r^b.$$ 

Of course, the definition of such an interface is not unique (see [27] and [35] for similar definitions of interfaces).

In the following we describe various subspaces and the corresponding solvers, which are needed in the construction of the desired preconditioner.

At first we define a subspace associated with each $D_r$. Set $D_r^0 = D_r \setminus D_r^b$ (see Figure 1), and define the subspace for each subdomain $D_r^0$

$$W_h^q(D_r^0) = \{ \mu \in W_h^q(\gamma) : \text{supp} \mu \subset D_r^0 \}, \quad r = 1, 2, \ldots, n_0.$$ 

The local solver on the local space $W_h^q(D_r^0)$ is defined in the standard manner. Let $S_{r}^0 : W_h^q(D_r^0) \rightarrow W_h^q(D_r^0)$ be the discrete operator which is the restriction of $S$ on $W_h^q(D_r^0)$.

For the definition of solvers associated with the interface, we need to give a decomposition of the interface $\Gamma$. Let $V_d$ denote the set of all the nodes corresponding to the coarse partition $T_d$. For a coarse node $V \in V_d$, let $D_V$ denote the top left corner element that touch the vertex $V$ (see Figure 2).

Let $D_{rl}$ denote the union of the elements that touch the intersection $\partial D_r \cap \partial D_l$ from the left side (or the upper side) but do not touch the lower (or the right) endpoints of $\partial D_r \cap \partial D_l$ (see Figure 3).

It is easy to see that the interface can be decomposed into

$$\Gamma = (\bigcup_{rl} D_{rl}) \bigcup \bigcup_{V \in V_d} D_V.$$
Next we define local interface spaces. Set
\[
\tilde{D}_{rl} = D_{rl} \cup D^0_r \cup D^0_l.
\]
and define the discrete $s_h(\cdot, \cdot)$-harmonic extension spaces

$$W^q_h(\tilde{D}_{rl}) = \{ \mu \in W^q_h(\gamma) : supp \mu \subset \tilde{D}_{rl}; s_h(\mu, w) = 0, \forall w \in W^q_h(D_r^0) \cup W^q_h(D_l^0) \}.$$  

Notice that the basis functions of these local spaces are unknown, a standard way is to transform the corresponding local interface problem into a residual equation, which is defined on the natural restriction space of the global space $W^q_h(\gamma)$ on the subdomain $\tilde{D}_{rl}$. However, solution of the residual equation is expensive since the restriction space contains many more basis functions than each local space $W^q_h(D_r^0)$, which is defined on a smaller subdomain $D_r^0$ than $\tilde{D}_{rl}$.

In order to decrease the cost of calculation, we need to reduce the sizes of the subdomains $\tilde{D}_{rl}$ and define discrete $s_h(\cdot, \cdot)$-harmonic on the reduced subdomains. We reduce $\tilde{D}_{rl}$ to $\tilde{D}_{rl}^{half}$ such that the resulting subdomains have almost the same size with $D_r$ (see Figure 4).

![Figure 4](image.png)

**Figure 4.** The rectangle ABCD denotes a subdomain $D_{rl}$ and the rectangle EFGH denotes a subdomain $\tilde{D}_{rl}^{half}$

Define the local spaces

$$W^q_h(\tilde{D}_{rl}^{half}) = \{ \mu \in W^q_h(\gamma) : supp \mu \subset \tilde{D}_{rl}^{half} \}.$$  

For $\mu \in W^q_h(\tilde{D}_{rl})$, let $\mu_{rl}^{half} \in W^q_h(\tilde{D}_{rl}^{half})$ denote the discrete $s_h(\cdot, \cdot)$-harmonic extension of $\mu|_{D_{rl}}$ into the complement domain $\tilde{D}_{rl}^{half}\backslash D_{rl}$.

Define the discrete operator $K^0_{rl} : W^q_h(\tilde{D}_{rl}) \to W^q_h(\tilde{D}_{rl})$ by

$$(K^0_{rl}\mu, w) = s_h(\mu_{rl}^{half}, w_{rl}^{half}), \mu \in W^q_h(\tilde{D}_{rl}), \forall w \in W^q_h(\tilde{D}_{rl}).$$

Notice that the action of $(K^0_{rl})^{-1}$ is implemented by solving a residual equation defined on the “half” space $W^q_h(\tilde{D}_{rl}^{half})$ (see Algorithm 3.1), so $K^0_{rl}$ can be regarded as an “inexact” local interface solver based on the “compressed” harmonic extension $\mu_{rl}^{half}$ (refer to [27]).

Finally we construct a coarse space $W^q_h(\gamma)$ by some local energy minimizations.

For a coarse node $V \in V_d$, let $\phi_V^{(m)}$ be a basis function in the subspace

$$W^q_h(D_V) = \{ \mu|_{D_V} : \mu \in W^q_h(\gamma) \}.$$  

Since the function $\phi_V^{(m)}$ is well defined only on the fine edges of $D_V$, we need to extend $\phi_V^{(m)}$ in a suitable manner such that $\phi_V^{(m)}$ has definitions on all the fine edges of $T_h$. The desired coarse space will be spanned by the extensions of all $\phi_V^{(m)}$. 
Let \( \tilde{\phi}_V^{(m)} \) be the initial extension of \( \phi_V^{(m)} \) such that \( \tilde{\phi}_V^{(m)} \) is \( s_h(\cdot, \cdot) \)-harmonic on each subspace \( W^q_h(D^0_r) \) and vanishes on all the fine edges in \( \Gamma \setminus D_V \). In order to define further extension of \( \tilde{\phi}_V^{(m)} \), let \( \Gamma_V \) denote a union of the coarse edges that touch the vertex \( V \). For each \( \Gamma_{rl} \in \Gamma_V \), let \( \Phi^{(m)}_{\Gamma_{rl}} \in W^q_h(\bar{D}_{rl}^{\text{half}}) \) be the solution of the minimization problem

\[
\min_{\Phi \in W^q_h(\bar{D}_{rl}^{\text{half}})} \left\{ s_h^{(r)}(\tilde{\phi}_V^{(m)} + \Psi, \tilde{\phi}_V^{(m)} + \Psi) + s_h^{(l)}(\tilde{\phi}_V^{(m)} + \Psi, \tilde{\phi}_V^{(m)} + \Psi) \right\},
\]

where \( s_h^{(r)}(\cdot, \cdot) \) denotes the restriction of \( s_h(\cdot, \cdot) \) on the fine edges on \( D_r \). Then \( \Phi^{(m)}_{\Gamma_{rl}} \in W^q_h(\bar{D}_{rl}^{\text{half}}) \) can be obtained by solving the local equation

\[
\sum_{k=r,l} s_h^{(k)}(\Phi^{(m)}_{\Gamma_{rl}}, v) = -\sum_{k=r,l} s_h^{(k)}(\tilde{\phi}_V^{(m)}, v), \quad \forall v \in W^q_h(\bar{D}_{rl}^{\text{half}}).
\]

Define

\[
\Phi^{(m)}_V = \tilde{\phi}_V^{(m)} + \sum_{\Gamma_{rl} \in \Gamma_V} R^t_{rl} \Phi^{(m)}_{\Gamma_{rl}},
\]

where \( R^t_{rl} \) denotes the zero extension operators from \( W^q_h(\bar{D}_{rl}) \) into \( W^q_h(\gamma) \). The coarse space \( W^q_d(\gamma) \) is spanned by all the basis functions \( \Phi^{(m)}_V \), namely,

\[
W^q_d(\gamma) = \text{span}\{\Phi^{(m)}_V\}.
\]

Let the coarse solver \( S_d : W^q_d(\gamma) \rightarrow W^q_d(\gamma) \) be the discrete operator which is the restriction of \( S \) on \( W^q_d(\gamma) \) as usual.

Now we can define the preconditioner \( K : W^q_h(\gamma) \rightarrow W^q_h(\gamma) \) as

\[
K^{-1} = \sum_r (S^0_r)^{-1} Q_r + \sum_{\Gamma_{rl}} (K^0_{rl})^{-1} Q_{rl} + S^{-1}_d Q_d,
\]

where \( Q_r, Q_{rl} \) and \( Q_d \) denote the \( L^2 \) projectors into \( W^q_h(D^0_r) \), \( W^q_h(\bar{D}_{rl}) \) and \( W^q_d(\gamma) \), respectively.

The action of the preconditioner \( K^{-1} \) can be described by the following algorithm.

**Algorithm 3.1.** For \( \xi \in W^q_d(\gamma) \), the solution \( \lambda_\xi = K^{-1} \xi \in W^q_h(\gamma) \) can be obtained as follows:

**Step 1.** Computing \( \lambda^0_r \in W^q_h(D^0_r) \) in parallel by

\[
s_h^{(r)}(\lambda^0_r, \mu_h) = (\xi, \mu_h), \quad \forall \mu_h \in W^q_h(D^0_r), \ r = 1, 2, \ldots, n_0.
\]

**Step 2.** Computing \( \lambda_{rl} \in W^q_h(\bar{D}_{rl}^{\text{half}}) \) in parallel by

\[
\sum_{k=r,l} s_h^{(k)}(\lambda_{rl}, \mu_h) = (\xi, \mu_h) - \sum_{k=r,l} s_h^{(k)}(\lambda^0_r, \mu_h), \quad \forall \mu_h \in W^q_h(\bar{D}_{rl}^{\text{half}}).
\]

**Step 3.** Computing \( \lambda_d \in W^q_d(\gamma) \) by

\[
s_h(\lambda_d, \mu_h) = (\xi, \mu_h) - \sum_r s_h^{(r)}(\lambda^0_r, \mu_h), \quad \forall \mu_h \in W^q_d(\gamma).
\]

**Step 4.** Set \( \phi = \sum \lambda_{rl} + \lambda_d \) and compute harmonic extensions \( \lambda^H_r \in W^q_h(D_r) \) for all \( r \) in parallel, such that \( \lambda^H_r = \phi \) on \( D^0_r \) and satisfies

\[
s_h^{(r)}(\lambda^H_r, \mu_h) = 0, \quad \forall \mu_h \in W^q_h(D^0_r), \ r = 1, 2, \ldots, n_0.
\]
Step 5. Computing
\[ \lambda_{\xi} = \sum_{r} \lambda_{r}^{0} + \sum_{r} \lambda_{r}^{H}. \]

**Remark 3.1.** The minimization problem (3.1) is different from that in the BDDC method. In the BDDC method, each minimization problem which determines coarse basis functions is defined on one subdomain, so the solutions of the two minimization problems associated with two neighboring subdomains have different values on their common interface. In order to define coarse basis functions, in the BDDC method one has to compute some average of the values of the two solutions on the common interface. Since the minimization problem (3.1) is defined on the subdomain \( \tilde{D}_{rl} \), the solution of this minimization problem has a unique value on the interface \( D_{rl} \) and the coarse basis functions can be directly obtained by (3.3). We found that, if minimization problems are defined as in the BDDC method, then the resulting preconditioner is unstable.

**Remark 3.2.** Notice that the variational problem (3.2) and the variational problem in Step 2 of Algorithm 3.1 correspond to the same stiffness matrix (with different right hands only). Thus the computation for the coarse basis functions by solving every subproblem (3.2) in parallel only increases a little cost by using LU decomposition made in Step 2 for each local stiffness matrix (when Step 2 is implemented in the direct method).

### 4. Main results

As pointed out in Section 1, the proposed method is also practical for the case with variable wave number, but there are some additional difficulties in the analysis for general variable wave numbers. Because of this, as in the most existing papers, we only consider the case with \( c(x) = 1 \) (so \( \kappa = \omega \)) for the purpose of analysis. Before presenting the main results, we give several assumptions.

**Assumption 1.** \( \Omega \) is a strictly star-shaped domain with an analytic boundary; \( p \geq 1 + c_{0} \log \omega \) with a constant \( c_{0} \).

The above assumption can be essentially found in [32]. The condition that \( \Omega \) is a strictly star-shaped domain with an analytic boundary appeared in the many existing works (see, for example, [10]).

**Assumption 2.** The mesh size \( h \) satisfies the condition: \( \omega h \leq C_{0} \) with a constant \( C_{0} \), where \( C_{0} \) may be mildly small.

The above assumption is weaker than that required in the most existing works. The following assumption has no restriction to the proposed method.

**Assumption 3.** The parameter \( \rho \) in the variational formula is not large: \( \rho \leq C_{0} \min\{1, \omega^{2} h\} \) for a mildly small constant \( C_{0} \).

Now we list the main results, which will be proved in the next section. Firstly, we give a result about local \( \inf - \sup \) condition.

**Theorem 4.1.** For any \( \mu \in W_{h}^{q}(\partial \Omega_{k} \setminus \partial \Omega) \), there exits a function \( v \in V_{h}^{p}(\partial \Omega_{k}) \) such that
\[
\langle \mu, v \rangle_{\partial \Omega_{k} \setminus \partial \Omega} \geq C p^{-\frac{1}{2}} \|\mu\|_{0, \partial \Omega_{k} \setminus \partial \Omega} \| v \|_{0, \partial \Omega_{k}}, \quad \text{when } q \geq 1 \text{ and } p \geq q + 2, \\
\langle \mu, v \rangle_{\partial \Omega_{k} \setminus \partial \Omega} \geq C \|\mu\|_{0, \partial \Omega_{k} \setminus \partial \Omega} \| v \|_{0, \partial \Omega_{k}}, \quad \text{when } q \geq 2 \text{ and } p = 2q,
\]
where \( C \) is a constant independent of \( \omega, h, p \) and \( q \).
Then we give a result on the coerciveness of the sesquilinear form $s_h(\cdot, \cdot)$, which implies that the discrete problem (2.7) is well posed.

**Theorem 4.2.** Let Assumption 1-Assumption 3 be satisfied. Then, for any $\mu_h \in W_q^\omega(\gamma)$, we have

$$s_h(\mu_h, \mu_h) \geq C\omega^{-2}h^2p^{-2} \sum_{\gamma_k} \|\mu_h\|_{0, \gamma_k}^2, \quad \text{when } q \geq 1 \text{ and } p \geq q + 2,$$

$$s_h(\mu_h, \mu_h) \geq C\omega^{-2}h^2p^{-1} \sum_{\gamma_k} \|\mu_h\|_{0, \gamma_k}^2, \quad \text{when } q \geq 2 \text{ and } p = 2q,$$

(4.2)

where $C$ is a constant independent of $\omega, h, p$ and $q$.

Finally, we give error estimates of the approximation $u_h$. For ease of notation, we define ($r \geq 0$)

$$|v|_{r+1, \Omega} = \left(\sum_{k=1}^{N} |v|_{r+1, \gamma_k}^2\right)^{\frac{1}{2}} \quad \text{and} \quad \|\mu\|_{r-\frac{1}{2}, \gamma} = \left(\sum_{k=1}^{N} \|\mu\|_{r-\frac{1}{2}, \partial\gamma_k \setminus \partial\gamma_0}^2\right)^{\frac{1}{2}}.$$

**Theorem 4.3.** Suppose that $q \geq 1$ and $p \geq q + 2$. Let Assumption 1-Assumption 3 be satisfied. Assume that the analytical solution $u$ of the Helmholtz problem (1.1) belongs to $H^{r+1}(\Omega)$ with $1 \leq r \leq q$ ($r \in \mathbb{N}$). Then the approximate solution $u_h$ defined in Subsection 2.3 have the error estimates

$$|u - u_h|_{1, \Omega} \leq Ch^{-1}(p^{-r}|u|_{r+1, \Omega} + q^{-r}\|\lambda\|_{r-\frac{1}{2}, \gamma})$$

and

$$\|u - u_h\|_{0, \Omega} \leq Ch^{-1}(p^{-r}|u|_{r+1, \Omega} + q^{-r}\|\lambda\|_{r-\frac{1}{2}, \gamma}),$$

where $C$ is a constant independent of $\omega, h, p$ and $q$.

**Remark 4.1.** It can be seen from Theorem 4.3 that the proposed discretization method possesses quasi-optimal convergence, which almost has no “wave number pollution” and is comparable to the convergence of the plane wave method (comparing Theorem 3.15 in [22]).

5. **Proof of the main results**

In this section, we give detailed proofs of the theorems stated in Section 4. As we shall see, the proofs are very technical, so a series of auxiliary results are needed to be built.

For ease of notation, we use the shorthand notation $x \preceq y$ and $y \succeq x$ for the inequality $x \leq Cy$ and $y \geq Cx$.

5.1. **Analysis on the local inf − sup condition.** Set $J = [0, 1]$ and let $P_p$ stands for the space of all polynomials on $J$ with orders $\leq p$. Firstly, we give a space decomposition of $P_p$ on $J$

$$P_p = P_p^1 + P_p^\prime, \quad \text{with } P_p^1 \perp P_p^\prime.$$

(5.1)

The specific definition of $P_p^1$ and $P_p^\prime$ will be given next.

Let $P_1$ and $P_p^\prime$ denote the linear part and high-order part of $P_p$, respectively. Then the two bases of $P_1$ are $\{\phi_1 = x, \phi_2 = 1 - x\}$. Let $\{\psi_k\}_{k=1}^{p-1}$ denote the basis
functions of the subspace $\mathcal{P}_{p'}$. Define

$$
\phi_1^* = \phi_1 - \sum_{k=1}^{p-1} \alpha_k \psi_k, \text{ satisfying } \langle \phi_1^*, \psi_k \rangle_J = 0, \quad k = 1, 2, \ldots, p - 1
$$

and

$$
\phi_2^* = \phi_2 - \sum_{k=1}^{p-1} \beta_k \psi_k, \text{ satisfying } \langle \phi_2^*, \psi_k \rangle_J = 0, \quad k = 1, 2, \ldots, p - 1.
$$

Apparently we can get

$$(\alpha_1, \alpha_2, \ldots, \alpha_{p-1})^t = A^{-1} b_1 \text{ and } (\beta_1, \beta_2, \ldots, \beta_{p-1})^t = A^{-1} b_2,$$

where $A = (\langle \psi_k, \psi_j \rangle_J)_{(p-1) \times (p-1)}$ and $b_1 = (\langle \phi_1, \psi_k \rangle_J)_{(p-1) \times 1}$, $b_2 = (\langle \phi_2, \psi_k \rangle_J)_{(p-1) \times 1}$, which means $\phi_1^*, \phi_2^*$ are uniquely determined. Let $\mathcal{P}_1^* = \text{span} \{ \phi_1^*, \phi_2^* \}$, which satisfies the space decomposition (5.1).

Next we give a set of orthogonal basis functions of $\mathcal{P}_{p'} = \text{span} \{ \psi_1, \psi_2, \ldots, \psi_{p-1} \}$. We choose a set of Jacobi polynomials $\{G_k\}$ (see [34]) and for convenience, we let their first coefficient be 1

$$
G_k = (-1)^{k-1} \frac{(k+3)!}{(2k+2)!} x^{-2} (1-x)^{-2} \frac{d^{k-1}}{dx^{k-1}} (x^{k+1} (1-x)^{k+1}), \quad (5.2)
$$

which satisfy

$$
\int_0^1 x^2 (1-x)^2 G_k G_j dx = \begin{cases} 0 & k \neq j, \\ \frac{(k-1)!^2 (k+3)!}{(2k+2)! (2k+3)!} & k = j. \end{cases}
$$

We also have the recursion relations

$$
\begin{cases}
G_1 = 1, \quad G_2 = x - \frac{1}{2}, \\
G_k = (x - \frac{1}{2}) G_{k-1} - \frac{(k-2)(k+2)}{4(2k-1)(2k+1)} G_{k-2} \quad k \geq 3.
\end{cases}
$$

Let $\psi_k = \frac{(2k+2)!}{(k-1)!(k+1)!} x (1-x) G_k$. It’s obviously $\psi_k(0) = \psi_k(1) = 0$ and

$$
\int_0^1 \psi_k \psi_j dx = \begin{cases} 0 & k \neq j, \\ \frac{k(k+1)(k+2)(k+3)}{2k+3} & k = j. \end{cases} \quad (5.3)
$$

Furthermore $\{\psi_k\}_{k=1}^{p-1}$ satisfy the recursion relations

$$
\begin{cases}
\psi_1 = 12x(1-x), \quad \psi_2 = 120x(1-x) (x - \frac{1}{2}), \\
\psi_k = \frac{2(2k+1)}{k-1} \psi_{k-1} - \frac{k+2}{k-1} \psi_{k-2} \quad k \geq 3.
\end{cases} \quad (5.4)
$$

The functions $\{\psi_k\}_{k=1}^{p-1}$ constitute a set of orthogonal bases of $\mathcal{P}_{p'}$.

**Lemma 5.1.** For $\{\phi_1, \phi_2\}$ and $\{\psi_k\}_{k=1}^{p-1}$ defined as before, we have

$$
\langle \phi_1, \psi_k \rangle_J = 1 \text{ and } \langle \phi_2, \psi_k \rangle_J = (-1)^{k-1}, \quad k = 1, 2, \ldots, p - 1.
$$
Proof. Using mathematical induction and the recursion relations (5.4), we can easily prove
\[ \langle \phi_1, \psi_k \rangle_J = \int_0^1 x \psi_k dx = 1 \]
and
\[ \langle \phi_2, \psi_k \rangle_J = \int_0^1 (1 - x) \psi_k dx = (-1)^{k-1}. \]
\[ \square \]

Lemma 5.2. For any integer \( m \geq 1 \), we have
\[ \sum_{k=1}^{m} \frac{2k + 3}{k(k + 1)(k + 2)(k + 3)} = \frac{1}{3} - \frac{1}{(m + 1)(m + 3)} \] (5.5)
and
\[ \sum_{k=1}^{m} \frac{(-1)^{k-1}(2k + 3)}{k(k + 1)(k + 2)(k + 3)} = \frac{1}{6} + \frac{(-1)^{m-1}}{(m + 1)(m + 2)(m + 3)}. \] (5.6)

Proof. It is easy to verify that
\[ \sum_{k=1}^{m} \frac{2k + 3}{k(k + 1)(k + 2)(k + 3)} = \sum_{k=1}^{m} \left( \frac{1}{k(k + 2)} - \frac{1}{(k + 1)(k + 3)} \right) = \frac{1}{3} - \frac{1}{(m + 1)(m + 3)} \]
and
\[ \sum_{k=1}^{m} \frac{(-1)^{k-1}(2k + 3)}{k(k + 1)(k + 2)(k + 3)} = \sum_{k=1}^{m} \left( \frac{(-1)^{k-1}}{k(k + 1)(k + 2)} + \frac{(-1)^{k-1}}{(k + 1)(k + 2)(k + 3)} \right) = \frac{1}{6} + \frac{(-1)^{m-1}}{(m + 1)(m + 2)(m + 3)}. \]
\[ \square \]

Lemma 5.3. For the two bases \( \{\phi_1^*, \phi_2^*\} \) of \( P_1^* \), we have
\[ \langle \phi_1^*, \phi_1^* \rangle_J = \langle \phi_2^*, \phi_2^* \rangle_J = \frac{1}{p(p + 2)}, \quad \langle \phi_1^*, \phi_2^* \rangle_J = \frac{(-1)^{p-1}}{p(p + 1)(p + 2)} \] (5.7)
and
\[ ||a_1\phi_1^*||_{0,J}^2 + ||a_2\phi_2^*||_{0,J}^2 \leq \frac{p+1}{p}||a_1\phi_1^* + a_2\phi_2^*||_{0,J}^2, \quad \forall a_1, a_2 \in \mathbb{R}. \] (5.8)

Proof. Using the definition of \( \phi_k^* \), (5.3), (5.5) and Lemma 5.1, we deduce that
\[ \langle \phi_1^*, \phi_1^* \rangle_J = \langle \phi_1, \phi_1 \rangle_J - b_1^t A^{-1} b_1 = \langle \phi_1, \phi_1 \rangle_J - \sum_{k=1}^{p-1} \langle \phi_1, \psi_k \rangle_J \langle \psi_k, \psi_k \rangle_J = \frac{1}{3} - \sum_{k=1}^{p-1} \frac{2k + 3}{k(k + 1)(k + 2)(k + 3)} = \frac{1}{p(p + 2)}. \]
Similarly, we have
\[ \langle \phi_2^*, \phi_2^* \rangle_J = \frac{1}{p(p + 2)}. \]
On the other hand, using Lemma 5.1, 5.3 and 5.6, we get

\[
\langle \phi_1^*, \phi_2^* \rangle_J = \langle \phi_1, \phi_2 \rangle_J - b_1^t A_1^{-1} b_2 = \langle \phi_1, \phi_2 \rangle_J - \sum_{k=1}^{p-1} \langle \phi_k, \psi_k \rangle_J \langle \phi_2, \psi_k \rangle_J
\]

\[
= \frac{1}{6} - \sum_{k=1}^{p-1} (\frac{1}{k(k+1)(k+2)(k+3)}) = (\frac{1}{p(p+1)(p+2)}).
\]

This gives the second equality of (5.7). The equality (5.8) can be easily obtained from (5.7).

**Proof of Theorem 4.1** For an element \( \Omega_k \), let \( n_k \) denote the number of the edges of \( \Omega_k \) and write its boundary as \( \partial \Omega_k = \bigcup_{J=1}^{n_k} J \), where \( J \) is the \( j \)th edge of \( \Omega_k \). If \( J \in \partial \Omega \), we set \( \mu|_{\partial \Omega} = 0 \). Then we only need to prove: for any \( \mu \in W^p_h(\partial \Omega_k) \), there exits \( v \in V^p_h(\partial \Omega_k) \), such that

\[
\langle \mu, v \rangle_{\partial \Omega_k} \geq C_{p,q} \|\mu\|_{0, \partial \Omega_k} \|v\|_{0, \partial \Omega_k},
\]

where \( C_{p,q} \) is a positive constant which only may depend on \( p \) and \( q \). Since \( \Omega_k \) is regular, we can simply set \( J = [0,1] \) by the scaling transformation.

Using the space decomposition (5.4), we can let

\[
\mu|_{J_k} = \sum_{k=1}^{p-1} \xi_k \psi_k + a_1 \phi_1^* + a_2 \phi_2^*, \quad a_1, a_2, \xi_k \in \mathbb{R}
\]

(5.9)

where \( \{ \phi_1^*, \phi_2^* \} \) are two basis functions of \( \mathcal{P}^*_1 \) and \( \{ \psi_k \} \) denote the orthogonal basis functions of \( \mathcal{P}_{q'}/ \) as defined before. Then we choose

\[
v|_{J_k} = \sum_{k=1}^{p-1} \xi_k \psi_k + \sum_{k=q}^{p-1} \langle \phi_1^* + a_2 \phi_2^*, \psi_k \rangle_J \psi_k.
\]

(5.10)

It is clear that \( v|_{J_k}(0) = v|_{J_k}(1) = 0 \). Then we have \( v \in V^p_h(\partial \Omega_k) \).

Using the orthogonality between the basis functions, we have

\[
\langle \mu, v \rangle_{J_k} = \sum_{k=1}^{p-1} \xi_k^2 \langle \psi_k, \psi_k \rangle_J + \sum_{k=q}^{p-1} \langle \phi_1^* + a_2 \phi_2^*, \psi_k \rangle_J^2 = ||v||_{0, J_k}^2.
\]

It follows that

\[
\langle \mu, v \rangle_{\partial \Omega_k} = \sum_{J=1}^{n_k} \langle \mu, v \rangle_{J} = \sum_{J=1}^{n_k} ||v||_{0, J}^2 = ||v||_{0, \partial \Omega_k}^2.
\]

Thus, we only need to prove: there exists \( C_{p,q} \), such that

\[
||v||_{0, \partial \Omega_k} \geq C_{p,q} \|\mu\|_{0, \partial \Omega_k} \quad \text{or} \quad ||v||_{0, J} \geq C_{p,q} \|\mu\|_{0, J}.
\]

To do this, we use (5.3), (5.7) and Lemma 5.1 which gives

\[
||v||_{0, J_k}^2 = \sum_{k=1}^{p-1} \xi_k^2 \langle \psi_k, \psi_k \rangle_J + \sum_{k=q}^{p-1} \frac{(2k + 3)(a_1 + (-1)^{k-1}a_2)^2}{k(k+1)(k+2)(k+3)}
\]

\[
= \sum_{k=1}^{p-1} \xi_k^2 \langle \psi_k, \psi_k \rangle_J + (\frac{1}{q(q+2)} - \frac{1}{p(p+2)}) (a_1^2 + a_2^2)
\]

\[
+ (\frac{1}{q(q+2)(q+2)} - \frac{(-1)^{p-1}}{p(p+1)(p+2)}) 2a_1 a_2.
\]
Then, using \[5.7\] again, we have
\[
\|\mu\|_{0,J_j}^2 = \sum_{k=1}^{q-1} \epsilon_k^2 (\psi_k, \bar{\psi}_k)_j + a_1^2(\phi_1^*, \phi_1^*)_j + a_2^2(\phi_2^*, \phi_2^*)_j + 2a_1a_2(\phi_1^*, \phi_2^*)_j
\]
\[
= \sum_{k=1}^{q-1} \epsilon_k^2 (\psi_k, \bar{\psi}_k)_j + \frac{1}{q(q+2)} (a_1^2 + a_2^2) + \frac{(-1)^{q-1}}{q(q+1)(q+2)} 2a_1a_2.
\]
So we choose
\[
C_{p,q}^2 = \begin{cases} 
1 - \frac{(p+1)(q+2)}{p(p+1)} & p + q = \text{even}, \\
1 - \frac{(q+1)(q+2)}{p(p+1)} & p + q = \text{odd},
\end{cases}
\]
satisfying
\[
\|u\|_{0,J_j}^2 \geq C_{p,q}^2 \|\lambda\|_{0,J_j}^2.
\]
When \(q \geq 1\) and \(p \geq q + 2\), we have \(\|u\|_{0,J_j}^2 \geq p^{-1}\|\mu\|_{0,J_j}^2\), namely,
\[
(\mu, v)_j \geq p^{-\frac{1}{2}} \|\mu\|_{0,J_j} \|v\|_{0,J_j}.
\]
When \(q \geq 2\) and \(p = 2q\), we have \(\|u\|_{0,J_j}^2 \geq \|\mu\|_{0,J_j} \|v\|_{0,J_j}\), namely,
\[
(\mu, v)_j \geq \|\mu\|_{0,J_j} \|v\|_{0,J_j}.
\]
It concludes the proof of the local \(\inf - \sup\) condition given by \[4.1\]. \(\square\)

5.2. Analysis on the coerciveness. Let \(\tilde{V}_h^p(\mathcal{T}_h) \subset H^1(\Omega)\) denote the usual finite element space of \(p\) order associated with the partition \(\mathcal{T}_h\). For a function \(v \in \tilde{V}_h^p(\mathcal{T}_h)\), we need to construct a correction function \(\hat{v}\) in \(\mathcal{T}_h^p(\mathcal{T}_h)\), which should satisfy the estimates stated in Lemma \[5.4\] given later.

For each element \(\Omega_k\), we set \(v|_{\Omega_k} = v_k\). We try to define a correction function \(\hat{v}_k\) of \(v_k\) for each \(\Omega_k\) and then define the desired function \(\hat{v}\) such that \(\hat{v}|_{\Omega_k} = \hat{v}_k\). To this end, we use the standard decomposition
\[
v_k = v_k^0 + v_k^\partial,
\]
where \(v_k^0|_{\partial\Omega_k} = v_k|_{\partial\Omega_k}\) and \(v_k^\partial \in \tilde{V}_h^p(\Omega_k)\) is the discrete harmonic extension of \(v_k|_{\partial\Omega_k}\) into \(\Omega_k\). It is easy to see that \(v_k^0|_{\partial\Omega_k} = v_k|_{\partial\Omega_k} - v_k^\partial|_{\partial\Omega_k} = 0\), which can be naturally extended into \(\Omega\). However, in general we have \(v_k^\partial|_{\gamma_{kj}} = v_j^\partial|_{\gamma_{kj}}\), where \(\gamma_{kj} = \partial\Omega_k \cap \partial\Omega_j\) is an element edge.

Since we require that the desired function \(\hat{v} \in H^1(\Omega)\), we need to define a correction \(\tilde{v}_k^\partial\) of \(v_k^\partial\) in a special manner such that \(\tilde{v}_k^\partial|_{\gamma_{kj}} = v_j^\partial|_{\gamma_{kj}}\). After it is done, we naturally define
\[
\hat{v}_k = v_k^0 + \tilde{v}_k^\partial,
\]
where \(\tilde{v}_k^\partial \in \tilde{V}_h^p(\Omega_k)\) is the discrete harmonic extension of \(\tilde{v}_k^\partial|_{\partial\Omega_k}\) into \(\Omega_k\).

Next we give a definition of \(\tilde{v}_k^\partial|_{\partial\Omega_k}\). Let \(e\) denote an edge of \(\partial\Omega_k\). When \(e = \partial\Omega_k \cap \partial\Omega\), we simply define \(\tilde{v}_k^\partial|_e = v_k^\partial|_e\). If \(e = \partial\Omega_k \cap \partial\Omega_j\), we define \(\tilde{v}_k^\partial|_e\) in the following way.

As in the beginning of Subsection \[5.1\], we can define the spaces \(\mathcal{P}_1^*\) and \(\mathcal{P}_p^*\) on the edge \(e\) by the standard scaling technique. Then we have the decomposition
\[
v_k^\partial|_e = v_k^\partial + v_k^\partial|_e,
\]
where $v^0_i \in P_i^1$ and $v^0_{i0} \in P_i^{p_r}$. Let $\{\phi_1^{sc}, \phi_2^{sc}\}$ denote the two basis functions of $P_i^1$. Let $v_1$ and $v_2$ denote the two endpoints of the edge $e$. It is easy to see that $v^0_i$ can be written as

$$v^0_{k1} = v_k(v_1)\phi_1^{sc} + v_k(v_2)\phi_2^{sc}.$$ 

Set

$$\Lambda_{v_i} = \{r, \Omega_r \text{ contains } v_i \text{ as one of its vertices} \} \quad (i = 1, 2),$$

and let $n_{v_i}$ denote the number of all the elements that contain $v_i$ as their common vertex, namely, the dimension of set $\Lambda_{v_i}$. For $e = \partial\Omega_k \cap \partial\Omega_j$, define

$$\hat{v}^0_{k1} = \frac{1}{n_{v_1}} \sum_{r \in \Lambda_{v_1}} v_r(v_1)\phi_1^{sc} + \frac{1}{n_{v_2}} \sum_{r \in \Lambda_{v_2}} v_r(v_2)\phi_2^{sc} \quad (5.11)$$

and

$$\hat{v}^0_{k0} = \frac{1}{2}(v^0_{k0} + v^0_{j0}). \quad (5.12)$$

Now we define $\hat{v}^0_k = \hat{v}^0_{k1} + \hat{v}^0_{k0}$ for each $e \subset \partial\Omega_k$, and let $\hat{v}^0_k \in V^p_h(\Omega_k)$ be the discrete harmonic extension of $\hat{v}^0_k|_{\partial\Omega_k}$. From the definition of $\hat{v}^0_k$, we know that $\hat{v}^0_k|_{\gamma_{kj}} = \hat{v}^0_j|_{\gamma_{kj}}$. Thus we can define $\hat{v}_k = \hat{v}^0_k + \hat{v}^0_j$. It is clear that $\hat{v}_k|_{\gamma_{kj}} = \hat{v}_j|_{\gamma_{kj}}$.

Finally we define $\hat{v}$ by $\hat{v}|_{\partial\Omega_k} = \hat{v}_k$ and we have $\hat{v} \in V^p_h(\mathcal{T}_h)$.

**Lemma 5.4.** For $v \in V^p_h(\mathcal{T}_h)$, let $\hat{v} \in V^p_h(\mathcal{T}_h)$ be defined above. Then we have

$$|v - \hat{v}|_{1,\Omega} + h^{-1/2} \langle v - \hat{v} \rangle_{0,\Omega} \lesssim h^{-1/2} p^2 \left( \sum_{\gamma_{kj}} \|v|_{1,\Omega} \right)^{1/2}. \quad (5.13)$$

Proof. Notice that $v_k = v^0_k + v^o_k$ and $\hat{v}_k = v^0_k + \hat{v}^o_k$, and using the stability of the discrete harmonic extension, we deduce that

$$|v - \hat{v}|_{1,\Omega} = \sum_{k=1}^N |v_k - \hat{v}_k|_{1,\Omega_k} = \sum_{k=1}^N |v^0_k - \hat{v}^0_k|_{1,\Omega_k} \quad (5.14)$$

and

$$\|v - \hat{v}\|_{0,\Omega} = \sum_{k=1}^N \|v_k - \hat{v}_k\|_{0,\Omega_k} = \sum_{k=1}^N \|v^0_k - \hat{v}^0_k\|_{0,\Omega_k} \quad (5.15)$$

and

$$\|v - \hat{v}\|_{0,\Omega} \lesssim \sum_{k=1}^N \|v^0_k - \hat{v}^0_k\|_{0,\Omega_k} \lesssim h^{-1} p \sum_{k=1}^N \|v^0_k - \hat{v}^0_k\|_{0,\partial\Omega_k} \lesssim h p \sum_{k=1}^N \|v^0_k - \hat{v}^0_k\|_{0,\partial\Omega_k}.$$ 

It suffices to give an estimate of $\sum_{k=1}^N \|v^0_k - \hat{v}^0_k\|_{0,\partial\Omega_k}$.

Let $e$ be an edge on $\partial\Omega_k$. When $e = \partial\Omega_k \cap \partial\Omega$, we have $(v^0_k - \hat{v}^0_k)|e = 0$, which implies that $\|v^0_k - \hat{v}^0_k\|_{0,e} = 0$.

If $e = \partial\Omega_k \cap \partial\Omega_j = \gamma_{kj}$, we have $(v^0_k - \hat{v}^0_k)|_{\gamma_{kj}} = (v^0_k - \hat{v}^0_k) + (v^0_k - \hat{v}^0_k)$. So we get

$$\|v^0_k - \hat{v}^0_k\|_{0,\gamma_{kj}} \leq \|v^0_k - \hat{v}^0_k\|_{0,\gamma_{kj}} + \|v^0_k - \hat{v}^0_k\|_{0,\gamma_{kj}}. \quad (5.16)$$
Let $v_1$ and $v_2$ denote two endpoints of $\gamma_{kj}$, and let $\Lambda_{v_i}$ ($i = 1, 2$) be the sets defined before (5.11). It follows, from (5.11) and (5.12), that
\[ v^0_{k1} - \tilde{v}^0_{k1} = \frac{1}{n_{v1}} \sum_{r \in \Lambda_{v1}} (v_k - v_r)(v_1) \phi_1^{e} + \frac{1}{n_{v2}} \sum_{r \in \Lambda_{v2}} (v_k - v_r)(v_2) \phi_2^{e} \] (5.17)
and
\[ v^0_{k0} - \tilde{v}^0_{k0} = \frac{1}{2} (v^0_{k0} - v^0_{j0}). \]

Notice that $(v^0_k - v^0_j)|_{\gamma_{kj}} = (v^0_{k1} - v^0_{j1}) + (v^0_{k0} - v^0_{j0})$, where $(v^0_{k1} - v^0_{j1}) \in \mathcal{P}_1$ and $(v^0_{k0} - v^0_{j0}) \in \mathcal{P}_0$. It is easy to see that
\[ ||v^0_k - \tilde{v}^0_k||_{0, \gamma_{kj}} = \frac{1}{2} ||v^0_k - v^0_j||_{0, \gamma_{kj}} \leq \frac{1}{2} ||v_k - v_j||_{0, \gamma_{kj}} = \frac{1}{2} ||v_k - v_j||_{0, \gamma_{kj}}. \] (5.18)

Clearly, we have $k \in \Lambda_{v_i}$ ($i = 1, 2$). Define
\[ \Lambda_{v_i}^{k, 1} = \{ r, r \in \Lambda_{v_i}, r \neq k, \text{ and } \gamma_{kr} = \partial \Omega_k \cap \partial \Omega_r \text{ is an edge} \} \quad (i = 1, 2) \]
and
\[ \Lambda_{v_i}^{k, 2} = \{ r, r \in \Lambda_{v_i}, r \neq k, \text{ and } \partial \Omega_r \cap \partial \Omega_j \text{ is the vertex } v_i \} \quad (i = 1, 2). \]

It is clear that $\Lambda_{v_i} = \{ k \} \cup \Lambda_{v_i}^{k, 1} \cup \Lambda_{v_i}^{k, 2}$ and $\Lambda_{v_i}^{k, 1} \cap \Lambda_{v_i}^{k, 2} = \emptyset$. Then, for the first item on the right side of (5.17), we have
\[ \frac{1}{n_{v1}} \sum_{r \in \Lambda_{v1}} (v_k - v_r)(v_1) \phi_1^{e} = \frac{1}{n_{v1}} \sum_{r \in \Lambda_{v1}^{k, 1}} (v_k - v_r)(v_1) \phi_1^{e} + \frac{1}{n_{v1}} \sum_{r \in \Lambda_{v1}^{k, 2}} (v_k - v_r)(v_1) \phi_1^{e}. \] (5.19)

We first give an estimate of $|||(v_k - v_r)(v_1) \phi_1^{e}|||_{0, e}$ for $r \in \Lambda_{v_1}^{k, 1}$. Let $\{ \phi_1^{\gamma_{kr}}, \phi_2^{\gamma_{kr}} \}$ denote the two basis functions of $\mathcal{P}_1$ associated with the edge $\gamma_{kr}$. We already assume that the partition $T_h$ is quasi-uniform, which yields
\[ ||\phi_1^{e}||_{0, e} \lesssim ||\phi_1^{\gamma_{kr}}||_{0, \gamma_{kr}} \text{ and } ||\phi_2^{e}||_{0, e} \lesssim ||\phi_2^{\gamma_{kr}}||_{0, \gamma_{kr}}. \]

This, together with Lemma [5.3] leads to
\[ ||(v_k - v_r)(v_1) \phi_1^{e}||_{0, e} \lesssim ||v_k - v_r||_{0, \gamma_{kr}} \lesssim ||v_k - v_j||_{0, \gamma_{kj}} \]
\[ = ||v_k - v_r||_{0, \gamma_{kr}} = ||v||_{0, \gamma_{kr}}, \quad (r \in \Lambda_{v_1}^{k, 1}). \] (5.20)

Next we estimate $|||(v_k - v_r)(v_1) \phi_1^{e}|||_{0, e}$ for $r \in \Lambda_{v_1}^{k, 2}$. Without loss of generality, we assume that there exists some element $\Omega_l$ such that $\partial \Omega_k \cap \partial \Omega_l = \gamma_{kl}$ and $\partial \Omega_l \cap \partial \Omega_r = \gamma_{lr}$ are two (different) edges that have the common vertex $v_1$. Then, by the triangle inequality, we have
\[ ||(v_k - v_r)(v_1) \phi_1^{e}||_{0, e} \leq ||(v_k - v_l)(v_1) \phi_1^{e}||_{0, e} + ||(v_l - v_r)(v_1) \phi_1^{e}||_{0, e}. \]

We can estimate the two terms on the right side of the above inequality like (5.20), and we obtain
\[ ||(v_k - v_r)(v_1) \phi_1^{e}||_{0, e} \lesssim ||v||_{0, \gamma_{kl}} + ||v||_{0, \gamma_{lr}}, \quad (r \in \Lambda_{v_1}^{k, 2}). \] (5.21)
Let $E_{v_1}$ denote the set of all the edges that have $v_1$ as their common endpoint. Substituting (5.20) and (5.21) into (5.19), yields

$$\left\| \frac{1}{n_{v_1}} \sum_{r \in H_{v_1}} (v_k - v_r)(v_1) \phi_1^e \right\|_{0,e} \lesssim \sum_{e \in E_{v_1}} \|v\|_{0,e}. \quad (5.22)$$

In an analogous way with (5.22), we can verify that

$$\left\| \frac{1}{n_{v_2}} \sum_{r \in H_{v_2}} (v_k - v_r)(v_2) \phi_2^e \right\|_{0,e} \lesssim \sum_{e \in E_{v_2}} \|v\|_{0,e}. \quad (5.23)$$

Here $E_{v_2}$ denotes the set of all the edges that have $v_2$ as their common endpoint. Plugging this and (5.22) in (5.17), leads to

$$\left\| v_k^\alpha - \tilde{v}_k^\alpha \right\|_{0,\gamma_{k_j}} \lesssim \sum_{e \in E_{v_1}} \|v\|_{0,e} + \sum_{e \in E_{v_2}} \|v\|_{0,e},$$

which, together with (5.18) and (5.16), gives

$$\left\| v_k^\alpha - \tilde{v}_k^\alpha \right\|_{0,\gamma_{k_j}} \lesssim \sum_{e \in E_{v_1}} \|v\|_{0,e} + \sum_{e \in E_{v_2}} \|v\|_{0,e}. \quad (5.24)$$

Hence, we get

$$\sum_{k=1}^{N} \left\| v_k^\alpha - \tilde{v}_k^\alpha \right\|^2_{0,\partial \Omega_h} = \sum_{\gamma_{k_j}} \left\| v_k^\alpha - \tilde{v}_k^\alpha \right\|^2_{0,\gamma_{k_j}} \lesssim \sum_{\gamma_{k_j}} \|v\|^2_{0,\gamma_{k_j}}. \quad (5.25)$$

Finally, substituting this inequality into (5.14) and (5.15), we obtain

$$|v - \tilde{v}|^2_{1,\Omega} \lesssim h^{-1}p \sum_{\gamma_{k_j}} \|v\|^2_{0,\gamma_{k_j}} \quad \text{and} \quad |v - \tilde{v}|^2_{0,\Omega} \lesssim hp \sum_{\gamma_{k_j}} \|v\|^2_{0,\gamma_{k_j}}. \quad (5.26)$$

The estimate (5.13) is a direct consequence of the above inequalities. □

Consider the dual problem with Robin-type boundary condition

$$\begin{cases}
-\Delta \phi - \omega^2 \phi = \tilde{f} & \text{in } \Omega, \\
\frac{\partial \phi}{\partial n} - i\omega \phi = 0 & \text{on } \partial \Omega.
\end{cases} \quad (5.23)$$

The weak form of (5.23) is: to find $\phi \in H^1(\Omega)$ such that

$$(\nabla \phi, \nabla \overline{\psi})_\Omega - \omega^2(\phi, \overline{\psi})_\Omega - i\omega(\phi, \overline{\psi})_{\partial \Omega} = (\tilde{f}, \overline{\psi})_\Omega, \quad \forall \psi \in H^1(\Omega). \quad (5.24)$$

The $p$-FEM discretization of (5.24) reads: to find $\phi_h \in \hat{V}_h^p(\mathcal{T}_h)$ such that

$$(\nabla \phi_h, \nabla \overline{\psi}_h)_\Omega - \omega^2(\phi_h, \overline{\psi}_h)_\Omega - i\omega(\phi_h, \overline{\psi}_h)_{\partial \Omega} = (\tilde{f}, \overline{\psi}_h)_\Omega, \quad \forall \psi_h \in \hat{V}_h^p(\mathcal{T}_h). \quad (5.25)$$

Some error estimates of this approximation were built in [32].

**Lemma 5.5.** [32] Cox 5.10 | Let $\tilde{f} \in L^2(\Omega)$ and let $\phi$ and $\phi_h$ be the solution defined above. Assume that $\frac{\omega h}{p} \leq C_0$ and let Assumption 1 be satisfied. Then we have

$$\left\| \nabla (\phi - \phi_h) \right\|_{0,\Omega} + \omega \left\| (\phi - \phi_h) \right\|_{0,\Omega} \lesssim h^{-1} \left\| \tilde{f} \right\|_{0,\Omega}. \quad (5.26)$$

By Lemma 5.4 and Lemma 5.5 we can derive another auxiliary result.
Lemma 5.6. Let Assumption 1-Assumption 3 be satisfied, and let \( v \in V^p_h(\mathcal{T}_h) \) satisfy

\[
a^{(k)}(v, w) = \langle \lambda_h, w \rangle_{\partial \Omega_k \setminus \partial \Omega} \quad (k = 1, 2, \ldots, N), \quad \forall w \in V^p_h(\mathcal{T}_h). \tag{5.26}
\]

Then

\[
||v||^2_{0, \Omega} \lesssim h^{-1} \sum_{\gamma_{kj}} ||v||^2_{0, \gamma_{kj}}. \tag{5.27}
\]

Proof. In the dual problem (5.23), we choose \( \tilde{f} = v \). Then (5.23) becomes

\[
\begin{cases}
-\Delta \phi - \omega^2 \phi = v & \text{in } \Omega, \\
\frac{\partial \phi}{\partial n} - i\omega \phi = 0 & \text{on } \partial \Omega. \tag{5.28}
\end{cases}
\]

Let \( \phi \in H^1(\Omega) \) and \( \phi_h \in \tilde{V}^p_h(\mathcal{T}_h) \) denote its weak solution and \( p \) finite element solution, which are defined respectively by (5.24) and (5.25) with \( \tilde{f} = v \).

Using (5.28) and Green’s formula, we obtain

\[
||v||^2_{0, \Omega} = \sum_{k=1}^{N} (v, v)_{\Omega_k} = \sum_{k=1}^{N} ((v, -\Delta \phi)_{\Omega_k} - (v, \omega^2 \phi)_{\Omega_k})
\]

\[
= \sum_{k=1}^{N} ((\nabla v, \nabla \phi)_{\Omega_k} - (v, \nabla \phi \cdot n)_{\partial \Omega_k} - \omega^2 (v, \phi)_{\Omega_k})
\]

\[
= \sum_{k=1}^{N} (\nabla v, \nabla \phi_h)_{\Omega_k} - \sum_{k=1}^{N} \omega^2 (v, \phi_h)_{\Omega_k} - \sum_{\gamma_{kj}} \langle [v], \nabla \phi_h \cdot n \rangle_{\gamma_{kj}} - (v, i\omega \phi)_{\partial \Omega}
\]

\[
= \sum_{k=1}^{N} (\nabla v, \nabla \phi_h)_{\Omega_k} - \sum_{k=1}^{N} \omega^2 (v, \phi_h)_{\Omega_k} - \sum_{\gamma_{kj}} \langle [v], \nabla \phi_h \cdot n \rangle_{\gamma_{kj}} + i\omega (v, \phi)_{\partial \Omega}
\]

\[
+ \sum_{k=1}^{N} (\nabla v, \nabla (\phi - \phi_h))_{\Omega_k} - \sum_{k=1}^{N} \omega^2 (v, \phi - \phi_h)_{\Omega_k}. \tag{5.29}
\]

Letting \( w = \phi_h \) in (5.26) and summing the resulting equality over \( k \), and using the fact that \( \phi_h \) is continuous across the inner edges, gives

\[
\sum_{k=1}^{N} a^{(k)}(v, \phi_h) = \sum_{k=1}^{N} \langle \lambda_h, \phi_h \rangle_{\partial \Omega_k \setminus \partial \Omega} = 0,
\]

which implies that

\[
\sum_{k=1}^{N} (\nabla v, \nabla \phi_h)_{\Omega_k} - \sum_{k=1}^{N} \omega^2 (v, \phi_h)_{\Omega_k} = -i\rho \sum_{\gamma_{kj}} \langle [v], \phi_h \rangle_{\gamma_{kj}} - i\omega (v, \phi_h)_{\partial \Omega}.
\]
This, together with (5.29), leads to

\[ ||v||^2_{0, \Omega} = -i \rho \sum_{\gamma_k} \langle [v], \bar{\phi}_h \rangle_{\gamma_k} - i \omega \langle v, \bar{\phi}_h \rangle_{\partial \Omega} - \sum_{\gamma_k} \langle [v], \nabla \bar{\phi} \cdot \mathbf{n} \rangle_{\gamma_k} + i \omega \langle v, \bar{\phi} \rangle_{\partial \Omega} \]

\[ + \sum_{k=1}^{N} (\nabla v, \nabla (\phi - \phi_h))_{\Omega_k} - \sum_{k=1}^{N} \omega^2 (v, \bar{\phi} - \phi_h)_{\Omega_k} \]

\[ = \sum_{k=1}^{N} (\nabla v, \nabla (\phi - \phi_h))_{\Omega_k} - \sum_{k=1}^{N} \omega^2 (v, \bar{\phi} - \phi_h)_{\Omega_k} + i \omega \langle v, \bar{\phi} - \phi_h \rangle_{\partial \Omega} \]

\[ - \sum_{\gamma_k} \langle [v], \nabla \bar{\phi} \cdot \mathbf{n} \rangle_{\gamma_k} - i \rho \sum_{\gamma_k} \langle [v], \bar{\phi}_h \rangle_{\gamma_k}. \]  

(5.30)

For \( v \in V_h^p(\mathcal{T}_h) \), we construct \( \tilde{v} \in \tilde{V}_h^p(\mathcal{T}_h) \) as in Lemma 5.4. For ease of notation, set

\[ R = \sum_{k=1}^{N} (\nabla (v - \tilde{v}), \nabla (\phi - \phi_h))_{\Omega_k} - \sum_{k=1}^{N} \omega^2 (v - \tilde{v}, \bar{\phi} - \phi_h)_{\Omega_k} + i \omega (v - \tilde{v}, \bar{\phi} - \phi_h)_{\partial \Omega}. \]

Then (5.30) can be written as

\[ ||v||^2_{0, \Omega} = R + \sum_{k=1}^{N} (\nabla \tilde{v}, \nabla (\phi - \phi_h))_{\Omega_k} - \sum_{k=1}^{N} \omega^2 (\tilde{v}, \bar{\phi} - \phi_h)_{\Omega_k} + i \omega (\tilde{v}, \bar{\phi} - \phi_h)_{\partial \Omega} \]

\[ - \sum_{\gamma_k} \langle [v], \nabla \bar{\phi} \cdot \mathbf{n} \rangle_{\gamma_k} - i \rho \sum_{\gamma_k} \langle [v], \bar{\phi}_h \rangle_{\gamma_k}. \]  

(5.31)

Choosing \( \psi = \tilde{v} \) in (5.24) and \( \psi_h = \tilde{v} \) in (5.25), we get the difference

\[ (\nabla (\phi - \phi_h), \nabla \tilde{v})_{\Omega} - \omega^2 (\phi - \phi_h, \tilde{v})_{\Omega} - i \omega (\phi - \phi_h, \tilde{v})_{\partial \Omega} = 0. \]  

(5.32)

The complex conjugation of the above equality becomes

\[ (\nabla \tilde{v}, \nabla (\phi - \phi_h))_{\Omega} - \omega^2 (\bar{v}, \bar{\phi} - \phi_h)_{\Omega} + i \omega (v, \bar{\phi} - \phi_h)_{\partial \Omega} = 0. \]  

(5.33)

Substituting (5.33) into (5.31), we obtain

\[ ||v||^2_{0, \Omega} = R - \sum_{\gamma_k} \langle [v], \nabla \bar{\phi} \cdot \mathbf{n} \rangle_{\gamma_k} - i \rho \sum_{\gamma_k} \langle [v], \bar{\phi}_h \rangle_{\gamma_k}. \]  

(5.34)

Let \( M = \sum_{\gamma_k} ||v||^2_{0, \gamma_k} \). Using Cauchy-Schwarz inequality to the sums on the right side of (5.34), yields

\[ ||v||^2_{0, \Omega} \leq |R| + \sum_{\gamma_k} ||v||_{\partial \gamma_k} ||\nabla \phi \cdot \mathbf{n}||_{\partial \gamma_k} + \rho \sum_{\gamma_k} ||v||_{\partial \gamma_k} ||\phi_h||_{\partial \gamma_k} \]

\[ \leq |R| + \left( \sum_{\gamma_k} ||\nabla \phi \cdot \mathbf{n}||^2_{0, \gamma_k} \right)^{1/2} M^{1/2} + \rho \left( \sum_{\gamma_k} ||\phi_h||^2_{0, \gamma_k} \right)^{1/2} M^{1/2}. \]  

(5.35)
We need to estimate $|R|$. It is easy to see that
\[
|R| \leq \sum_{k=1}^{N} \|\nabla (v - \tilde{v})\|_{0, \Omega_k} \|\nabla (\phi - \phi_h)\|_{0, \Omega_k} + \sum_{k=1}^{N} \omega^2 \|v - \tilde{v}\|_{0, \Omega_k} \|\phi - \phi_h\|_{0, \Omega_k} \\
+ \omega \|v - \tilde{v}\|_{0, \partial \Omega} \|\phi - \phi_h\|_{0, \partial \Omega} \\
\leq \|\nabla (v - \tilde{v})\|_{0, \Omega} \|\nabla (\phi - \phi_h)\|_{0, \Omega} + \omega^2 \|v - \tilde{v}\|_{0, \Omega} \|\phi - \phi_h\|_{0, \Omega} \\
+ \omega \|v - \tilde{v}\|_{0, \partial \Omega} \|\phi - \phi_h\|_{0, \partial \Omega}.
\]

(5.36)

It follows by Lemma 5.5 that
\[
\|\nabla (\phi - \phi_h)\|_{0, \Omega} + \omega \|\phi - \phi_h\|_{0, \Omega} \leq hp^{-1} \|v\|_{0, \Omega}.
\]

Then, by the $\varepsilon$-inequality ($\varepsilon = \omega^{-\frac{1}{2}} < 1$), we have
\[
\|\phi - \phi_h\|_{0, \partial \Omega} \leq \omega^{-\frac{1}{2}} \|\nabla (\phi - \phi_h)\|_{0, \Omega} + \omega^\frac{1}{2} \|\phi - \phi_h\|_{0, \Omega} \leq \omega^{-\frac{1}{2}} hp^{-1} \|v\|_{0, \Omega}.
\]

Moreover, from Lemma 5.5, we have
\[
\|\nabla (v - \tilde{v})\|_{0, \Omega} + h^{-1} \|v - \tilde{v}\|_{0, \Omega} \leq h^{-\frac{1}{2}} p^2 M^\frac{1}{2}.
\]

Furthermore, by the trace inequality, we get
\[
|v - \tilde{v}|^2_{0, \partial \Omega} = \sum_{k=1}^{N} \|v - \tilde{v}\|^2_{0, \partial \Omega_k \cap \partial \Omega} \leq \sum_{k=1}^{N} (h\|\nabla (v - \tilde{v})\|^2_{0, \Omega_k} + h^{-1} \|v - \tilde{v}\|^2_{0, \Omega_k}) \\
= h\|\nabla (v - \tilde{v})\|^2_{0, \Omega} + h^{-1} \|v - \tilde{v}\|^2_{0, \Omega} \leq pM.
\]

Hence, substituting the above estimates into (5.36), we obtain
\[
|R| \leq (h^{-\frac{1}{2}} p^{-\frac{3}{2}} + \omega h^{-\frac{1}{2}} p^{-\frac{3}{2}} + \omega^2 hp^{-\frac{3}{2}}) \cdot M^\frac{1}{2} \|v\|_{0, \Omega}.
\]

(5.37)

On the other hand, from the stability estimates of the dual problem (5.28), we have
\[
\omega^{-1} |\phi|_{2, \Omega} + |\phi|_{1, \Omega} + \omega \|\phi\|_{0, \Omega} \lesssim \|v\|_{0, \Omega}.
\]

Then we can get
\[
\sum_{\gamma_{kj}} \|\nabla \phi \cdot n\|^2_{0, \gamma_{kj}} \leq \sum_{k=1}^{N} \|\nabla \phi \cdot n\|^2_{0, \partial \Omega_k} \lesssim \sum_{k=1}^{N} (h \|\phi\|^2_{2, \Omega_k} + h^{-1} \|\phi\|^2_{2, \Omega_k}) \\
= h \|\phi\|^2_{2, \Omega} + h^{-1} \|\phi\|^2_{2, \Omega} \lesssim h^{-1} \|v\|^2_{0, \Omega}.
\]

(5.38)

and
\[
\sum_{\gamma_{kj}} \|\phi_h\|^2_{0, \gamma_{kj}} \lesssim \sum_{k=1}^{N} \|\phi_h\|^2_{0, \partial \Omega_k} \lesssim \sum_{k=1}^{N} (\omega^{-1} h \|\phi\|^2_{2, \Omega_k} + \omega h^{-1} \|\phi\|^2_{0, \Omega_k}) \\
= \omega^{-1} h \|\phi\|^2_{2, \Omega} + \omega h^{-1} \|\phi\|^2_{0, \Omega} \lesssim \omega^{-1} h^{-1} \|v\|^2_{0, \Omega}.
\]

(5.39)

Substituting the inequalities (5.37), (5.38) and (5.39) into (5.35) and using the Assumption 3, yields
\[
\|v\|^2_{0, \Omega} \lesssim h^{-\frac{1}{2}} \cdot M^\frac{1}{2} \|v\|_{0, \Omega}.
\]

Finally, we obtain the desired inequality (5.27). \(\square\)

**Remark 5.1.** The above lemma can be viewed as a variant of Poincare inequality for discontinuous functions in $V_h^0(T_h)$. Comparing Lemma 3.7 of [22], we believe that the inequality (5.27) is sharp. The proof of this lemma depends on the estimate given in Lemma 5.5 (i.e., [32] Cor 5.10).
Lemma 5.7. Let Assumption 3 be satisfied. Assume that $v \in V_h^p(\mathcal{T}_h)$ satisfies the condition

$$a^{(k)}(v, \overline{w}) = \langle \lambda_h, \overline{w} \rangle_{\partial \Omega_h \setminus \partial \Omega} \quad (k = 1, 2, \cdots, N), \quad \forall w \in V_h^p(\mathcal{T}_h). \tag{5.40}$$

Then the following estimate holds

$$\sum_{\gamma_k} ||\lambda_h||_{0, \gamma_k}^2 \lesssim h^{-1}p^1 \delta ||\nabla v||^2_{0, \Omega} + h^{-1}p^\delta \omega^2 ||v||^2_{0, \Omega}.$$ 

where $\delta = 1$ when $p \geq q + 2$ or $\delta = 0$ if $p = 2q$ ($q \geq 2$).

Proof. By $\inf - \sup$ condition given in Theorem 4.1, there exists $\psi \in V_h^p(\partial \Omega_k)$, which is discrete harmonic in $\Omega_k$, such that

$$||\lambda_h||_{0, \partial \Omega_k}^2 \lesssim p^\delta \frac{\langle \lambda_h, \overline{\psi} \rangle_{\partial \Omega_k}^2}{||\psi||_{0, \partial \Omega_k}^2},$$

where $\delta = 1$ when $p \geq q + 2$ or $\delta = 0$ if $p = 2q$ ($q \geq 2$). Then, using (5.40) and Assumption 3, yields

$$||\lambda_h||_{0, \partial \Omega_k}^2 \lesssim p^\delta \frac{\langle \lambda_h, \overline{\psi} \rangle_{\partial \Omega_k}^2}{||\psi||_{0, \partial \Omega_k}^2} \lesssim p^\delta \frac{||\nabla \psi||_{0, \Omega_k}^2 ||\nabla \psi||_{0, \Omega_k}^2 + \omega^2 ||\psi||_{0, \Omega_k}^2 ||\psi||_{0, \Omega_k}^2 + \omega ||\psi||_{0, \Omega_k}^2 ||\psi||_{0, \Omega_k}^2}{||\psi||_{0, \partial \Omega_k}^2}.$$ 

(5.41)

Since $\psi$ is discrete harmonic in $\Omega_k$, we have

$$||\nabla \psi||_{0, \Omega_k}^2 \lesssim ||\psi||_{2, \partial \Omega_k}^2 \lesssim h^{-1}p ||\psi||_{0, \partial \Omega_k}^2,$$

$$||\psi||_{0, \Omega_k}^2 \lesssim h^2 ||\psi||_{2, \partial \Omega_k}^2 + h ||\psi||_{0, \partial \Omega_k} \lesssim hp ||\psi||_{0, \partial \Omega_k}^2.$$ 

This, together with (5.41), leads to

$$||\lambda_h||_{0, \partial \Omega_k}^2 \lesssim p^\delta (||\nabla \psi||_{0, \Omega_k}^2 h^{-1}p + \omega^2 ||\psi||_{0, \Omega_k}^2 hp + h||\nabla \psi||_{0, \Omega_k}^2 + \omega^2 h^{-1} ||\psi||_{0, \Omega_k}^2)$$

$$= p^\delta (h^{-1}p + hp + \omega h^{-1}) ||\nabla \psi||_{0, \Omega_k}^2 + \omega^2 h^{-1} ||\psi||_{0, \Omega_k}^2 \lesssim h^{-1}p^1 \delta ||\nabla \psi||_{0, \Omega_k}^2 + h^{-1}p^\delta \omega^2 ||\psi||_{0, \Omega_k}^2.$$ 

Summing up the above inequality over $k$, gives

$$\sum_{\gamma_k} ||\lambda_h||_{0, \gamma_k}^2 \lesssim h^{-1}p^1 ||\nabla \psi||_{0, \Omega}^2 + h^{-1}p^\delta \omega^2 ||\psi||_{0, \Omega}^2.$$

By Lemma 5.6 and Lemma 5.7, we can prove a crucial auxiliary result given below. As we will see, this auxiliary result plays a key role in the proof of Theorem 4.2 and Theorem 4.3.

Lemma 5.8. Assume that $q \geq 1$ and $p \geq q + 2$. Let Assumption 1-Assumption 3 be satisfied, and let $v \in V_h^p(\mathcal{T}_h)$ satisfy

$$a^{(k)}(v, \overline{w}) = \langle \lambda_h, \overline{w} \rangle_{\partial \Omega_h \setminus \partial \Omega} \quad (k = 1, 2, \cdots, N), \quad \forall w \in V_h^p(\mathcal{T}_h). \tag{5.42}$$

Then the following estimate holds

$$||\nabla v||_{0, \Omega}^2 + \omega^2 ||v||_{0, \Omega}^2 \lesssim \omega^2 h^{-1} \sum_{\gamma_k} |||v|||_{0, \gamma_k}^2.$$ 

(5.43)
Proof. Let \( M = \sum_{\gamma_{kj}} ||v||^2_{0,\gamma_{kj}} \). It follows by Lemma 5.6 that
\[
||v||^2_{0,\Omega} \lesssim h^{-1} M. \tag{5.44}
\]
If \( ||\nabla v||^2_{0,\Omega} \leq \omega^2 ||v||^2_{0,\Omega} \), by the above inequality we immediately get
\[
||\nabla v||^2_{0,\Omega} \lesssim \omega^2 h^{-1} M.
\]
This, together with \( \omega = 0 \), gives \( \omega = 0 \).

In the following we consider the case that the inequality \( ||\nabla v||^2_{0,\Omega} \leq \omega^2 ||v||^2_{0,\Omega} \) does not hold, namely, \( \omega^2 ||v||^2_{0,\Omega} \leq ||\nabla v||^2_{0,\Omega} \).

Choosing \( w = v \) in \( \omega \) and summing the resulting equality over \( k \), gives
\[
\sum_{k=1}^{N} \left( ||\nabla v||^2_{0,\Omega} - \omega^2 ||v||^2_{0,\Omega} \right) = \sum_{k=1}^{N} \langle \lambda_k, v \rangle_{\Omega} - \sum_{\gamma_{kj}} \langle \lambda_k, v \rangle_{\gamma_{kj}}.
\]
Considering the module of the above equality and using Cauchy-Schwarz inequality, yields
\[
||\nabla v||^2_{0,\Omega} - \omega^2 ||v||^2_{0,\Omega} \leq \sum_{\gamma_{kj}} ||\lambda_k||_{0,\gamma_{kj}} ||v||_{0,\gamma_{kj}} \leq \left( \sum_{\gamma_{kj}} ||\lambda_k||^2_{0,\gamma_{kj}} \right)^{\frac{1}{2}} M^{\frac{1}{2}}. \tag{5.45}
\]
Combining \( \omega \) and \( \omega \), we get
\[
||\nabla v||^2_{0,\Omega} \leq \omega^2 ||v||^2_{0,\Omega} + \left( \sum_{\gamma_{kj}} ||\lambda_k||^2_{0,\gamma_{kj}} \right)^{\frac{1}{2}} M^{\frac{1}{2}} \lesssim \omega^2 h^{-1} M + \left( \sum_{\gamma_{kj}} ||\lambda_k||^2_{0,\gamma_{kj}} \right)^{\frac{1}{2}} M^{\frac{1}{2}}. \tag{5.46}
\]
It follows by Lemma 5.7 that
\[
\sum_{\gamma_{kj}} ||\lambda_k||^2_{0,\gamma_{kj}} \lesssim h^{-1} p + \delta ||\nabla v||^2_{0,\Omega} + h^{-1} p + \omega^2 ||v||^2_{0,\Omega} \lesssim h^{-1} p + \delta ||\nabla v||^2_{0,\Omega}. \tag{5.47}
\]
Here we have used the assumption \( \omega^2 ||v||^2_{0,\Omega} \leq ||\nabla v||^2_{0,\Omega} \) in the derivation of the second inequality. Substituting \( \omega \) into \( \omega \), yields
\[
||\nabla v||^2_{0,\Omega} \lesssim \omega^2 h^{-1} M + h^{-1} p + \frac{1}{2} ||\nabla v||_{0,\Omega} M^{\frac{1}{2}}, \quad \delta = 0 \text{ or } 1.
\]
The above inequality implies that
\[
||\nabla v||^2_{0,\Omega} \lesssim \omega^2 h^{-1} M.
\]
In summary, we obtain the desired result \( \omega = 0 \). \qed

**Proof of Theorem 4.2** From the definition of \( u_{h,\delta}^{(1)}(\lambda_h) \) in Subsection 2.3, we know that
\[
a^{(k)}(u_{h,\delta}^{(1)}(\lambda_h), \overline{w}_h) = \langle \lambda_h, \overline{w}_h \rangle_{\partial \Omega' \setminus \partial \Omega}, \quad k = 1, \cdots, N; \quad \forall w_h \in V_h^p(T_h).
\]
Namely, \( u_{h,\delta}^{(1)}(\lambda_h) \) satisfies \( \omega \). It follows by Lemma 5.7 that
\[
\sum_{\gamma_{kj}} ||\lambda_k||^2_{0,\gamma_{kj}} \lesssim h^{-1} p + \delta ||\nabla u_{h,\delta}^{(1)}(\lambda_h)||^2_{0,\Omega} + h^{-1} p + \omega^2 ||u_{h,\delta}^{(1)}(\lambda_h)||^2_{0,\Omega}. \tag{5.48}
\]
Thus, together with (5.48), leads to
\[ \sum_{\gamma_k} \|\lambda h\|_{0,\gamma_k}^2 \lesssim \omega^2 h^{-2} \|u_h^{(1)}(\lambda_h)\|_{0,\gamma_k}^2. \]

This, together with (5.48), leads to
\[ \sum_{\gamma_k} \|\lambda h\|_{0,\gamma_k}^2 \lesssim \omega^2 h^{-2} p^{1+\delta} \sum_{\gamma_k} \|u_h^{(1)}(\lambda_h)\|_{0,\gamma_k}^2. \]

Thus
\[ s_h(\lambda_h, \lambda_h) = \sum_{\gamma_k} \|u_h^{(1)}(\lambda_h)\|_{0,\gamma_k}^2 \geq \omega^{-2} h^2 p^{-(1+\delta)} \sum_{\gamma_k} \|\lambda h\|_{0,\gamma_k}^2, \]

where \( \delta = 1 \) when \( p \geq q + 2 \) (\( q \geq 1 \)) or \( \delta = 0 \) if \( p = 2q \) (\( q \geq 2 \)).

5.3. **Analysis on the error estimates.** In order to prove Theorem 4.3, we need more auxiliary results.

**Lemma 5.9.** Let Assumption 2 and Assumption 3 be satisfied. For one element \( \Omega_k \), assume that \( v \) satisfies
\[ -\omega^2 (v, 1)_{\Omega_k} \pm i \rho (v, 1)_{\partial \Omega_k \setminus \partial \Omega} + i \omega (v, 1)_{\partial \Omega_k \cap \partial \Omega} = 0 \]  
(5.49)

and set
\[ \|\nabla v\|_{0,\Omega_k}^2 - \omega^2 \|v\|_{0,\Omega_k}^2 \pm i \rho \|v\|_{0,\partial \Omega_k \setminus \partial \Omega} + i \omega \|v\|_{0,\partial \Omega_k \cap \partial \Omega} = L(v). \]  
(5.50)

Then
\[ \|\nabla v\|_{0,\Omega_k}^2 \leq C|L(v)| \quad \text{and} \quad \omega^2 \|v\|_{0,\Omega_k}^2 \leq C|L(v)|. \]

Proof. We first assume that \( \partial \Omega_k \cap \partial \Omega \neq \emptyset \). Considering the imaginary part of (5.50), we have
\[ \omega \|v\|_{0,\partial \Omega_k \cap \partial \Omega} \leq |L(v)| + \rho \|v\|_{0,\partial \Omega_k \setminus \partial \Omega} \lesssim \frac{1}{\omega} |L(v)| + \rho h \|\nabla v\|_{0,\Omega_k}^2 + \rho \omega^{-1} \|v\|_{0,\Omega_k}^2. \]  
(5.51)

By Poincaré inequality and (5.51), yields
\[ \omega^2 \|v\|_{0,\Omega_k}^2 \lesssim \omega^2 h^2 \|\nabla v\|_{0,\Omega_k}^2 + \omega^2 h \|v\|_{0,\partial \Omega_k \cap \partial \Omega}^2 \lesssim \omega^2 h^2 \|\nabla v\|_{0,\Omega_k}^2 + \omega h |L(v)| + \rho h^2 \|v\|_{0,\Omega_k}^2 + \rho \omega \|v\|_{0,\Omega_k}^2, \]

which implies that
\[ (1 - \rho C \omega^{-1}) \omega^2 \|v\|_{0,\Omega_k}^2 \lesssim (\omega^2 h^2 + \rho \omega h^2) \|\nabla v\|_{0,\Omega_k}^2 + \omega h |L(v)|. \]

Then, from Assumption 3, we have
\[ \omega^2 \|v\|_{0,\Omega_k}^2 \lesssim \omega^2 h^2 \|\nabla v\|_{0,\Omega_k}^2 + \omega h |L(v)|. \]  
(5.52)

On the other hand, considering the real part of (5.50) and using (5.52), we deduce that
\[ \|\nabla v\|_{0,\Omega_k}^2 \leq \omega^2 \|v\|_{0,\Omega_k}^2 + |L(v)| \lesssim \omega^2 h^2 \|\nabla v\|_{0,\Omega_k}^2 + \omega h |L(v)| + |L(v)|, \]

which gives
\[ (1 - C \omega^2 h^2) \|\nabla v\|_{0,\Omega_k}^2 \lesssim |L(v)|. \]

This, together with (5.52), leads to
\[ (1 - C \omega^2 h^2) \omega^2 \|v\|_{0,\Omega_k}^2 \lesssim |L(v)|. \]
In the following we assume that $\partial \Omega_k \cap \partial \Omega = \emptyset$. It follows by (5.49) that
\[
\omega^2 |\Omega_k| |\gamma_h(v)| = \rho |\langle \hat{v}, 1 \rangle_{\partial \Omega_k \setminus \partial \Omega}| \leq \rho |\partial \Omega_k|^2 \|v\|_{0, \partial \Omega_k}
\]
where $\gamma_h(v) = \frac{1}{|\Omega_k|} \int_{\Omega_k} v \, dx$. Thus
\[
\omega^4 |\gamma_h(v)|^2 \leq \rho^2 |\partial \Omega_k| \cdot |\Omega_k|^{-2} \|v\|_{0, \partial \Omega_k}^2.
\]
This, together with the trace inequality (or $\varepsilon$-inequality), leads to
\[
\omega^4 |\gamma_h(v)|^2 \leq \rho^2 |\partial \Omega_k| \cdot |\Omega_k|^{-1} \|v\|_{0, \partial \Omega_k}^2 \leq \rho^2 h^{-1} (h \|\nabla v\|_{0, \Omega_k}^2 + h^{-1} \|v\|_{0, \Omega_k}^2) = \rho^2 \|\nabla v\|_{0, \Omega_k}^2 + \rho^2 h^{-2} \|v\|_{0, \Omega_k}^2.
\] (5.53)
Using Friedrichs’ inequality and (5.53), we deduce that
\[
\omega^4 \|v\|_{0, \Omega_k}^2 \leq \omega^4 |v - \gamma_h(v)|_{0, \Omega_k}^2 + \omega^4 |\gamma_h(v)|_{0, \Omega_k}^2 \leq \omega^4 h^2 \|\nabla v\|_{0, \Omega_k}^2 + \rho^2 \|\nabla v\|_{0, \Omega_k}^2 + \rho^2 h^{-2} \|v\|_{0, \Omega_k}^2.
\]
So we get
\[
(1 - \rho^2 C \omega^{-4} h^{-2}) \|v\|_{0, \Omega_k}^2 \leq (h^2 + \rho^2 \omega^{-4}) \|\nabla v\|_{0, \Omega_k}^2.
\]
Thus, by Assumption 3, we have
\[
\|v\|_{0, \Omega_k}^2 \lesssim h^2 \|\nabla v\|_{0, \Omega_k}^2.
\]
In addition, considering the real part of (5.50) and the above inequality, we have
\[
\|\nabla v\|_{0, \Omega_k}^2 \leq \omega^2 \|v\|_{0, \Omega_k}^2 + |L| \lesssim \omega^2 h^2 \|\nabla v\|_{0, \Omega_k}^2 + |L(v)|.
\]
Therefore we obtain
\[
(1 - C \omega^2 h^2) \|\nabla v\|_{0, \Omega_k}^2 \lesssim |L(v)|
\]
and
\[
(1 - C \omega^2 h^2) \omega^2 \|v\|_{0, \Omega_k}^2 \lesssim |L(v)|.
\]
In summary, under Assumption 2 we have
\[
\|\nabla v\|_{0, \Omega_k}^2 \lesssim |L(v)| \quad \text{and} \quad \omega^2 \|v\|_{0, \Omega_k}^2 \lesssim |L(v)|.
\]

In the following we need to use an auxiliary function $\hat{u}_h(\lambda)$ for $\lambda \in W(\gamma)$. For each element $\Omega_k$, let $\hat{u}_{h,k}(\lambda) \in V_h^p(\Omega_k)$ be determined by the variational problem
\[
a^{(k)}(\hat{u}_{h,k}(\lambda), \tau_h) = L^{(k)}(\tau_h) + \langle \lambda, \tau_h \rangle_{\partial \Omega_k \setminus \partial \Omega}, \quad \forall \tau_h \in V_h^p(\Omega_k).
\] (5.54)
Then define $\hat{u}_h(\lambda) \in V_h^p(\Omega)$ such that $\hat{u}_h(\lambda)|_{\Omega_k} = \hat{u}_{h,k}(\lambda)$ ($k = 1, \cdots, N$).

**Lemma 5.10.** Assume that $u \in H^{r+1}(\Omega)$ with $1 \leq r \leq p$. Let Assumption 2 and Assumption 3 be satisfied. Then
\[
|u(\lambda) - \hat{u}_h(\lambda)|_{1, \Omega} \leq C \omega^{-1} h^r p^{-r} |u|_{r+1, \Omega}
\]
and
\[
\|u(\lambda) - \hat{u}_h(\lambda)\|_{0, \Omega} \leq C \omega^{-1} h^r p^{-r} |u|_{r+1, \Omega}.
\]
Proof. Let \( \varepsilon_u = u(\lambda) - \hat{u}_h(\lambda) \) and \( \gamma_{\Omega_h}(\varepsilon_u) = \frac{1}{|\Omega_h|} \int_{\Omega_h} \varepsilon_u dx \). Choosing \( v = v_h \) in (2.2) and taking the difference between (2.2) and (5.54), we get
\[
a^{(k)}(\varepsilon_u, \tau_h) = 0 \quad \forall v_h \in V_h^p(\Omega_h), \ k = 1, 2, \ldots, N. \quad (5.55)
\]

Let \( Q_h^p : L^2(\Omega_h) \rightarrow V_h^p(\Omega_h) \) denote the standard \( L^2 \) projection operator. Using the approximation of the projection operator (refer to [20]), we have
\[
||\nabla(I - Q_h^p)u(\lambda)||_{0,\Omega_h} \lesssim h^{r+1} p^{-(r+1)} |u(\lambda)|_{r+1,\Omega_h}
\]
and
\[
||(I - Q_h^p)u(\lambda)||_{0,\Omega_h} \lesssim h^{r+1} p^{-(r+1)} |u(\lambda)|_{r+1,\Omega_h}.
\]

It is clear that
\[
\varepsilon_u = (I - Q_h^p)u(\lambda) + Q_h^p u(\lambda) - \hat{u}_h(\lambda).
\]
By choosing \( v_h = Q_h^p u(\lambda) - \hat{u}_h(\lambda) \) in (5.55), we get
\[
L(\varepsilon_u) = ||\nabla \varepsilon_u||_{0,\Omega_h}^2 - \omega^2 ||\varepsilon_u||_{0,\Omega_h}^2 + i \omega ||\varepsilon_u||_{0,\Omega_h} \gamma_{\Omega_h} = (\nabla \varepsilon_u, \nabla(I - Q_h^p)u(\lambda))_{\Omega_h} - \omega^2 (\varepsilon_u, (I - Q_h^p)u(\lambda))_{\Omega_h}
\]
\[
\quad \pm i \omega (\varepsilon_u, (I - Q_h^p)u(\lambda))_{\Omega_h} \gamma_{\Omega_h} + i \omega (\varepsilon_u, (I - Q_h^p)u(\lambda))_{\Omega_h} \gamma_{\Omega_h}.
\]

On the other hand, set \( v_h = 1 \) in (5.55), we have
\[
-\omega^2 (\varepsilon_u, 1)_{\Omega_h} \pm i \rho (\varepsilon_u, 1)_{\Omega_h} \gamma_{\Omega_h} + i \omega (\varepsilon_u, 1)_{\Omega_h} \gamma_{\Omega_h} = 0.
\]

Then \( \varepsilon_u \) satisfies the conditions in Lemma 5.9 so we obtain
\[
||\nabla \varepsilon_u||_{0,\Omega_h}^2 \lesssim |L(\varepsilon_u)| \quad \text{and} \quad \omega^2 ||\varepsilon_u||_{0,\Omega_h}^2 \lesssim |L(\varepsilon_u)|.
\]

It suffices to estimate \(|L(\varepsilon_u)|\). By the above definition of \( L \), we can deduce that
\[
|L(\varepsilon_u)| \lesssim ||\nabla \varepsilon_u||_{0,\Omega_h} ||\nabla(I - Q_h^p)u||_{0,\Omega_h} + \omega ||\varepsilon_u||_{0,\Omega_h} ||(I - Q_h^p)u||_{0,\Omega_h}
\]
\[
+ \omega ||\varepsilon_u||_{0,\Omega_h} ||(I - Q_h^p)u||_{0,\Omega_h} 
\]
\[
\lesssim ||\nabla \varepsilon_u||_{0,\Omega_h} ||\nabla(I - Q_h^p)u||_{0,\Omega_h} + \omega ||\varepsilon_u||_{0,\Omega_h} ||(I - Q_h^p)u||_{0,\Omega_h}
\]
\[
\lesssim \omega \left( h^\frac{r}{2} ||\varepsilon_u||_{0,\Omega_h} + h^{-\frac{r}{2}} ||\varepsilon_u||_{0,\Omega_h} \gamma_{\Omega_h} \right) \left( h^\frac{r}{2} p^{1 - \frac{r}{2}} ||\nabla(I - Q_h^p)u||_{0,\Omega_h} + h^{-\frac{r}{2}} p^2 ||(I - Q_h^p)u||_{0,\Omega_h} \right)
\]
\[
\lesssim \left( ||\nabla \varepsilon_u||_{0,\Omega_h} \gamma_{\Omega_h} + \omega ||\varepsilon_u||_{0,\Omega_h} \right) h^{r-1} p^{r-1} |u(\lambda)|_{r+1,\Omega_h}.
\]

Consequently, we have
\[
||\nabla \varepsilon_u||_{0,\Omega_h}^2 \lesssim \left( ||\nabla \varepsilon_u||_{0,\Omega_h} \gamma_{\Omega_h} + \omega ||\varepsilon_u||_{0,\Omega_h} \right) h^{r-1} p^{r-1} |u(\lambda)|_{r+1,\Omega_h}
\]
and
\[
\omega^2 ||\varepsilon_u||_{0,\Omega_h}^2 \lesssim \left( ||\nabla \varepsilon_u||_{0,\Omega_h} \gamma_{\Omega_h} + \omega ||\varepsilon_u||_{0,\Omega_h} \right) h^{r-1} p^{r-1} |u(\lambda)|_{r+1,\Omega_h}.
\]

Therefore, we get
\[
||\nabla \varepsilon_u||_{0,\Omega_h} \lesssim h^{r} p^{r-1} |u(\lambda)|_{r+1,\Omega_h} \quad \text{and} \quad \omega ||\varepsilon_u||_{0,\Omega_h} \lesssim h^{r} p^{r-1} |u(\lambda)|_{r+1,\Omega_h}. \quad (5.56)
\]

Then the estimates in this lemma can be obtained by summing (5.56) over \( k \). \( \square \)

Let \( Q_h^p : L^2(\gamma) \rightarrow W_h^p(\gamma) \) denote the \( L^2 \) projector, and let \( u_h(\lambda) \in V_h^p(\Omega_h) \) be defined as in (2.6), by choosing \( \lambda_h = Q_h^p \lambda \).

**Lemma 5.11.** Suppose that \( q \geq 1, p \geq q+2 \) and \( \lambda \in H^{r-\frac{1}{4}}(\gamma) \) with \( 1 \leq r \leq q+1 \). Let Assumption 1-Assumption 3 be satisfied. Then
\[
||\hat{u}_h(\lambda) - u_h(Q_h^p \lambda)||_{1,\Omega_h} \leq Ch^{r-q} \gamma_{\frac{1}{4}} \quad \text{and} \quad ||\hat{u}_h(\lambda) - u_h(Q_h^p \lambda)||_{0,\Omega_h} \leq C \omega^{-1} h^{r-q} \gamma_{\frac{1}{4}} \gamma_{\frac{1}{4}}.
Proof. Set \( \tilde{u}_h = \tilde{u}_h(\lambda) - u_h(Q^p_h(\lambda)) \) and \( \varepsilon_\lambda = \lambda - Q^p_h(\lambda) \). From (2.6) and (5.54), we deduce that

\[
a^{(k)}(\tilde{u}_h, v_h) = \langle \varepsilon_\lambda, v_h \rangle_{\partial \Omega_h \setminus \partial \Omega} \quad \forall \; v_h \in V^p_h(\Omega_k), \; k = 1, 2, \ldots, N. \tag{5.57}
\]

In particular, choosing \( v_h = 1 \) in (5.57), we have

\[
-\omega^2(\tilde{u}_h, 1)_{\partial \Omega} \pm i\rho(\tilde{u}_h, 1)_{\partial \Omega_h \setminus \partial \Omega} + i\rho(\tilde{u}_h, 1)_{\partial \Omega_h \cap \partial \Omega} = \int_{\partial \Omega_h \setminus \partial \Omega} \varepsilon_\lambda ds = 0.
\]

On the other hand, set \( v_h = \tilde{u}_h \) in (5.57), then we get

\[
L(\tilde{u}_h) = \|\nabla \tilde{u}_h\|^2_{0, \Omega} - \omega^2\|\tilde{u}_h\|^2_{0, \Omega} \pm i\rho\|\tilde{u}_h\|^2_{0, \partial \Omega_h \setminus \partial \Omega} + i\rho\|\tilde{u}_h\|^2_{0, \partial \Omega_h \cap \partial \Omega}
\]

\[
= \langle \varepsilon_\lambda, \tilde{u}_h \rangle_{\partial \Omega_h \setminus \partial \Omega} = \langle \varepsilon_\lambda, \tilde{u}_h - \gamma_{\Omega_h}(\tilde{u}_h) \rangle_{\partial \Omega_h \setminus \partial \Omega}.
\]

Here we have used the fact that \( \gamma_{\Omega_h}(\tilde{u}_h) = \frac{1}{|\Omega_h|} \int_{\Omega_h} \tilde{u}_h \, dx \) is a constant. It is easy to see that

\[
|L(\tilde{u}_h)| \leq \|\varepsilon_\lambda\|_{-\frac{1}{2}, \partial \Omega_h \setminus \partial \Omega} \|\tilde{u}_h - \gamma_{\Omega_h}(\tilde{u}_h)\|_{\frac{1}{2}, \partial \Omega_h \setminus \partial \Omega} \|\nabla \tilde{u}_h\|_{0, \Omega}.
\]

Then, by applying Lemma 5.9 to \( \tilde{u}_h \), we get

\[
\|\nabla \tilde{u}_h\|^2_{0, \Omega} \leq \|\varepsilon_\lambda\|_{-\frac{1}{2}, \partial \Omega_h \setminus \partial \Omega} \|\nabla \tilde{u}_h\|_{0, \Omega} \quad \text{and} \quad \omega^2\|\tilde{u}_h\|^2_{0, \Omega} \leq \|\varepsilon_\lambda\|_{-\frac{1}{2}, \partial \Omega_h \setminus \partial \Omega} \|\nabla \tilde{u}_h\|_{0, \Omega}.
\]

Thus we obtain

\[
\|\nabla \tilde{u}_h\|^2_{0, \Omega} \leq \|\varepsilon_\lambda\|_{-\frac{1}{2}, \partial \Omega_h \setminus \partial \Omega} \quad \text{and} \quad \omega\|\tilde{u}_h\|^2_{0, \Omega} \leq \|\varepsilon_\lambda\|_{-\frac{1}{2}, \partial \Omega_h \setminus \partial \Omega}. \tag{5.58}
\]

By using the approximation of the projection operators, we have

\[
\|\varepsilon_\lambda\|_{-\frac{1}{2}, \partial \Omega_h \setminus \partial \Omega} \leq h^r q^{-r} \|\lambda\|_{-r, -\frac{1}{2}, \partial \Omega_h \setminus \partial \Omega}.
\]

Substituting this into (5.58), yields

\[
\|\nabla \tilde{u}_h\|^2_{1, \Omega} \leq h^r q^{-r} \|\lambda\|_{-r, -\frac{1}{2}, \partial \Omega_h \setminus \partial \Omega} \quad \text{and} \quad \omega\|\tilde{u}_h\|^2_{0, \Omega} \leq h^r q^{-r} \|\lambda\|_{-r, -\frac{1}{2}, \partial \Omega_h \setminus \partial \Omega}.
\]

By summing the above inequalities over \( k \), we obtain the desired estimates. \( \square \)

**Lemma 5.12.** Suppose that \( q \geq 1 \), \( p \geq q + 2 \) and \( u \in H^{r+1}(\Omega) \) with \( 1 \leq r \leq p \). Let Assumption 1-Assumption 3 be satisfied. Then

\[
|u_h(Q^p_h(\lambda)) - u_h(\lambda)|_{1, \Omega} \leq C h^{r-1}(p^{-r}|u|_{r+1, \Omega} + q^{-r}||\lambda||_{r, -\frac{1}{2}, \gamma})
\]

and

\[
||u_h(Q^p_h(\lambda)) - u_h(\lambda)||_{0, \Omega} \leq C \omega^{-1} h^{r-1}(p^{-r}|u|_{r+1, \Omega} + q^{-r}||\lambda||_{r, -\frac{1}{2}, \gamma}).
\]

Proof. Set \( \varepsilon_h = u_h(Q^p_h(\lambda)) - u_h(\lambda) \) and let \( \varepsilon_\lambda = Q^p_h(\lambda) - \lambda_h \). Notice that the function \( u_h(Q^p_h(\lambda)) \) is defined by (2.6) with \( \lambda_h = Q^p_h(\lambda) \). Then, it follows by (2.6) that

\[
a^{(k)}(\varepsilon_h, v_h) = \langle \varepsilon_\lambda, v_h \rangle_{\partial \Omega_h \setminus \partial \Omega} \quad \forall \; v_h \in V^p_h(\Omega_k), \; k = 1, 2, \ldots, N. \tag{5.59}
\]

Setting \( v = \varepsilon_h \) in Lemma 5.8, we get

\[
\|\nabla \varepsilon_h\|^2_{0, \Omega} + \omega^2||\varepsilon_h||^2_{0, \Omega} \leq \omega^2 h^{-1} \sum_{\gamma_{kj}} ||\varepsilon_h||^2_{0, \gamma_{kj}}. \tag{5.60}
\]

By the definition of \( \lambda_h \), which corresponds to the minimal energy, we deduce that

\[
\sum_{\gamma_{kj}} ||\varepsilon_h||^2_{0, \gamma_{kj}} \leq \sum_{\gamma_{kj}} ||u_h(Q^p_h(\lambda))||^2_{0, \gamma_{kj}} + \sum_{\gamma_{kj}} ||u_h(\lambda)||^2_{0, \gamma_{kj}} \leq 2 \sum_{\gamma_{kj}} ||u_h(Q^p_h(\lambda))||^2_{0, \gamma_{kj}} + 2 \sum_{\gamma_{kj}} ||u_h(Q^p_h(\lambda)) - \lambda(\lambda)||^2_{0, \gamma_{kj}}. \tag{5.61}
\]
Here we have used the fact that the function $u(\lambda)$ has the zero jump across each edge $\gamma_{k_j}$. Using $\varepsilon$-inequality, yields
\[
|||u_h(Q_h^r \lambda) - u(\lambda)|||_{0,\gamma_{k_j}} \lesssim h|||\nabla(u_h(Q_h^r \lambda) - u(\lambda))|||_{0,\Omega_h} + h^{-1}|||u_h(Q_h^r \lambda) - u(\lambda)|||^2_{0,\Omega_h} \\
\lesssim h(|||\nabla(u_h(Q_h^r \lambda) - \hat{u}_h(\lambda))|||_{0,\Omega} + |||\nabla(\hat{u}_h(\lambda) - u(\lambda))|||_{0,\Omega}) \\
+ h^{-1}(|||u_h(Q_h^r \lambda) - \hat{u}_h(\lambda)|||_{0,\Omega} + |||\hat{u}_h(\lambda) - u(\lambda)|||^2_{0,\Omega}).
\]
Substituting this inequality into (5.61) and using Lemma 5.10 and Lemma 5.11 yields
\[
\sum_{\gamma_{k_j}} |||\varepsilon_h|||^2_{0,\gamma_{k_j}} \lesssim (h + h^{-1}\omega^{-2})(r^{-2r}|u|_{r+1,\Omega}^2 + h^{2r}q^{-2r}\|\lambda\|^2_{r-\frac{1}{2}}).
\]
This, together with (5.60), leads to
\[
|||\nabla\varepsilon_h|||_{0,\Omega}^2 + \omega^2|||\varepsilon_h|||_{0,\Omega}^2 \lesssim (\omega^2 + h^{-2})h^{2r}(r^{-2r}|u|_{r+1,\Omega}^2 + q^{-2r}\|\lambda\|^2_{r-\frac{1}{2}}).
\]
Now we immediately obtain the estimates in this lemma.

**Proof of Theorem 4.3** By the triangle inequality, we have
\[
||u(\lambda) - u_h(\lambda_h)||_{0,\Omega} \\
\leq ||u(\lambda) - \hat{u}_h(\lambda)||_{0,\Omega} + ||\hat{u}_h(\lambda) - u_h(Q_h^r \lambda)||_{0,\Omega} + ||u_h(Q_h^r \lambda) - u_h(\lambda_h)||_{0,\Omega}
\]
and
\[
||u(\lambda) - u_h(\lambda_h)||_{1,\Omega} \\
\leq ||u(\lambda) - \hat{u}_h(\lambda)||_{1,\Omega} + ||\hat{u}_h(\lambda) - u_h(Q_h^r \lambda)||_{1,\Omega} + ||u_h(Q_h^r \lambda) - u_h(\lambda_h)||_{1,\Omega}.
\]
Then, by the estimates given in Lemma 5.10, 5.11, and 5.12 we obtain
\[
||u(\lambda) - u_h(\lambda_h)||_{0,\Omega} \lesssim \omega^{-1}h^{r-1}(\rho^{-r}|u|_{r+1,\Omega} + q^{-r}\|\lambda\|_{r-\frac{1}{2}})
\]
and
\[
||u(\lambda) - u_h(\lambda_h)||_{1,\Omega} \lesssim h^{r-1}(\rho^{-r}|u|_{r+1,\Omega} + q^{-r}\|\lambda\|_{r-\frac{1}{2}}).
\]

**Remark 5.2.** Since we have used the results in 32 to prove Theorem 4.2 and Theorem 4.3 (see Remark 5.1 and the explanations given above Lemma 5.8), we have to use Assumption 1 in this paper. But, as we will see in the next section, this assumption seems unnecessary. We believe that the main results are also valid without this assumption.

6. Numerical Experiments

In this section we report some numerical results to illustrate that the proposed least squares method and domain decomposition preconditioner are efficient for Helmholtz equations with large wave numbers.

In the discretization method described in Section 2, the parameter $\rho$ can be relatively arbitrarily positive number. We find that the different choices of $\rho$ does not affect the accuracy of the resulting approximations provided that the value of $\rho$ is less than 1. In this section, we simply choose $\rho = 10^{-5}$ for numerical experiments.

For the considered example, the domain $\Omega$ is a rectangle so we adopt a uniform partition $\mathcal{T}_h$ for the domain $\Omega$ as follows: $\Omega$ is divided into some small rectangles with a same size $h$, where $h$ denotes the length of the longest edge of the elements.
To measure the accuracy of the numerical solution $u_h$, we introduce the following relative $L^2$ error:

$$
\text{Err.} = \frac{||u_{ex} - u_h||^2_{L^2(\Omega)}}{||u_{ex}||^2_{L^2(\Omega)}}.
$$

For a discretization method, when the value of $\omega h$ is fixed but $\omega$ increases ($h$ decreases), the relative $L^2$ error $\text{Err.}$ may obviously increase (if the number of basis functions on each element does not increase). This phenomenon is called “wave number pollution”. The efficiency of a discretization method for Helmholtz equations can be characterized by the degree of wave number pollution. For convenience, we define a positive parameter $\delta$ to measure the degree of wave number pollution as follows: the parameter $\delta$ is the minimal positive number such that, when $\omega$ increases and $h$ decreases to keep the value of $\omega^{1+\delta} h$ being a constant, the relative $L^2$ error $\text{Err.}$ does not increase. If $\delta = 0$, the discretization method has no “wave number pollution”. For the standard linear finite element method, the existing results imply that $\frac{1}{2} < \delta < 1$. A discretization method is ideal means that $\delta \to 0^+$. For concrete examples, it is difficult to exactly calculate such parameter $\delta$. Because of this, we want to give a similar definition of $\delta$, which can be explicitly calculated.

When $\omega$ increases from $\omega_1$ to $\omega_2$, the mesh size $h$ decreases from $h_1$ to $h_2$. We fix the value $\omega h$, i.e., $\omega_2 h_2 = \omega_1 h_1$. Let $\text{Err}_1$ and $\text{Err}_2$ denote the relative $L^2$ errors with $\omega = \omega_1$ ($h = h_1$) and $\omega = \omega_2$ ($h = h_2$), respectively. Define $\delta > 0$ by

$$
\frac{\omega_2^{1+\delta} h_2}{\omega_1^{1+\delta} h_1} = \frac{\text{Err}_2}{\text{Err}_1}.
$$

It is easy to see that the parameter $\delta$ can be expressed as

$$
\delta = \frac{\ln(\text{Err}_2/\text{Err}_1)}{\ln(\omega_2/\omega_1)} + \left(\frac{\ln(h_1/h_2)}{\ln(\omega_2/\omega_1)} - 1\right) = \frac{\ln(\text{Err}_2/\text{Err}_1)}{\ln(\omega_2/\omega_1)} \quad (\text{since } \omega_2 h_2 = \omega_1 h_1).
$$

For a given $\omega$, we define the error order with respect to $h$ in the standard manner, namely,

$$
\text{order} = \frac{\ln(\text{Err}_2/\text{Err}_1)}{\ln(h_2/h_1)},
$$

where $\text{Err}_1$ and $\text{Err}_2$ denote the relative $L^2$ errors corresponding to $h = h_1$ and $h = h_2$ ($p, q$ are fixed), respectively.

### 6.1. Wave propagation in a duct with rigid walls.

In this subsection, we try to give some comparisons between the proposed method and the plane wave least squares method. To this end, we consider the following model Helmholtz equation for the acoustic pressure $u$ (see [28])

$$
\begin{cases}
-\Delta u - \omega^2 u = 0 & \text{in } \Omega, \\
\frac{\partial u}{\partial n} + i\omega u = g & \text{on } \partial \Omega,
\end{cases}
$$

(6.1)

where $\Omega = [0, 2] \times [0, 1]$, and $g = (\frac{\partial u}{\partial n} + i\omega) u_{ex}$. The analytic solution $u_{ex}$ of the problem can be obtained in the closed form as

$$
u_{ex}(x, y) = \cos(k\pi y)(A_1 e^{-i\omega x} + A_2 e^{i\omega x})
$$

with $\omega_x = \sqrt{\omega^2 - (k\pi)^2}$, and the coefficients $A_1$ and $A_2$ satisfying the equation

$$
\begin{pmatrix}
\omega_x \\
\omega_x - \omega
\end{pmatrix}
\begin{pmatrix}
A_1 \\
A_2
\end{pmatrix} = 
\begin{pmatrix}
-i \\
0
\end{pmatrix}.
$$

(6.2)
Let “PWLS” and “NLS” denote the plane wave least squares method (see Subsection 2.3 in [27]) and the novel least squares method proposed in this paper, respectively. Besides, let $\hat{p}$ be the number of plane wave basis functions on every element. For convenience, we use “dof.” to denote the number of degrees of freedom in the resulting algebraic systems (which mean the system (2.8) for the NLS method).

In the following two tables, we compare the required numbers of degrees of freedom to achieve almost the same accuracies of the approximate solutions generated by the two methods.

**Table 1**

| $\frac{1}{h}$ | PWLS, $\hat{p} = 12$ | NLS, $(q, p) = (3, 5)$ |
|----------------|-------------------|-------------------|
| $\frac{1}{2}$  | 18816             | 1.85e-4           |
| $\frac{1}{3}$  | 31104             | 2.75e-5           |
| $\frac{1}{4}$  | 46464             | 7.83e-6           |
| $\frac{1}{5}$  | 64896             | 2.81e-6           |

**Table 2**

| $\frac{1}{h}$ | PWLS, $\hat{p} = 12$ | NLS, $(q, p) = (3, 5)$ |
|----------------|-------------------|-------------------|
| $\frac{1}{3}$  | 55296             | 1.29e-4           |
| $\frac{1}{4}$  | 75264             | 3.43e-4           |
| $\frac{1}{5}$  | 98304             | 5.73e-4           |
| $\frac{1}{7}$  | 124416            | 7.85e-4           |

It can be seen from the above data that, for the new least squares method, less degrees of freedom in the algebraic system are enough to achieve almost the same accuracies (with the same choices of $\omega$ and $h$). For the proposed method, a little extra cost is needed when solving all the local problems defined on the elements (in parallel). Besides, the system (2.8) has more complex structure than the one in the PWLS method and so its preconditioner is more difficult to construct. In summary, the proposed method is at least comparable to the plane wave method even if the wave number is a constant (otherwise, the plane wave method may be unpractical).

6.2. **An example with variable wave numbers.** In this subsection, we consider the following Helmholtz equations with variable wave numbers

\[
\begin{cases}
-\Delta u - \kappa^2 u = f & \text{in } \Omega, \\
\frac{\partial u}{\partial n} + i\kappa u = g & \text{on } \partial\Omega,
\end{cases}
\]

where $\Omega = [0, 1] \times [0, 1]$ and $\kappa = \frac{\omega}{c(x, y)}$. We define the velocity field $c(x)$ as a smooth converging lens with a Gaussian profile at the center $(r_1, r_2) = (1/2, 1/2)$ (refer to
\begin{equation}
  c(x, y) = \frac{4}{3} \left(1 - \frac{1}{8} \exp \left(-32((x - r_1)^2 + (y - r_2)^2)\right)\right).
\end{equation}

The analytic solution of the problem is given by
\begin{equation}
  u_{ex}(x, y) = c(x, y) \exp(i\omega xy).
\end{equation}

For this example, the standard plane wave methods are unpractical. In Table 3 and Table 4, we list the accuracies of the approximate solutions generated by the proposed least squares method, where the algebraic systems are solved in the exact manner.

**Table 3**

| $\omega h = 1$, $(q, p) = (2, 4)$ | $\omega h = 2$, $(q, p) = (3, 5)$ |
|------------------|------------------|
| $\omega$ | $h$ | $\Delta$ | $\delta h$ | $h$ | $\delta$ |
| 64 | $\frac{1}{7}$ | 3.0794e-5 | | 128 | $\frac{1}{7}$ | 2.9133e-5 | 0.0354 |
| 256 | $\frac{1}{7}$ | 3.2405e-5 | 0.0493 | 256 | $\frac{1}{7}$ | 2.9522e-5 | 0.0191 |
| 512 | $\frac{1}{7}$ | 3.3182e-5 | 0.0342 | 512 | $\frac{1}{7}$ | 2.9811e-5 | 0.0141 |

**Table 4**

| $(q, p) = (2, 4)$ | $(q, p) = (3, 5)$ |
|------------------|------------------|
| $\omega$ | $h$ | Err. | order | $h$ | Err. | order |
| 64 | $\frac{1}{7}$ | 4.4838e-4 | 3.8640 | 64 | $\frac{1}{7}$ | 2.0161e-6 | 3.9330 |
| 64 | $\frac{1}{7}$ | 3.0794e-5 | 3.8640 | 64 | $\frac{1}{7}$ | 2.0161e-6 | 3.9330 |
| 64 | $\frac{1}{7}$ | 1.2821e-7 | 3.9750 | 64 | $\frac{1}{7}$ | 2.1740e-8 | 5.0871 |
| 64 | $\frac{1}{7}$ | 8.0676e-9 | 3.9902 | 64 | $\frac{1}{7}$ | 6.7097e-10 | 5.0180 |

From the above two tables, we can see that the approximate solutions generated by the proposed method indeed have high accuracies and have little “wave number pollution”.

Since the resulting stiffness matrix is Hermitian positive definite, we can solve the system by the CG method and the PCG method with the preconditioner constructed in Section 3. As usual we choose $d \approx \sqrt{h}$ as the subdomain size in this preconditioner to guarantee the loading balance. The stopping criterion in the iterative algorithms is that the relative $L^2$-norm of the residual of the iterative approximation satisfies $\epsilon < 1.0e-6$.

Moreover, let $N_{iter}^{CG}$ represent the iteration count for solving the algebraic system by CG method and $N_{iter}^{PCC}$ represent the iteration count for solving the algebraic system by PCG method with the DD preconditioner. When the wave number $\omega$ increases (and the mesh size $h$ decreases), the iteration count $N_{iter}$ (represent $N_{iter}^{CG}$ or $N_{iter}^{PCC}$) also increases. In order to describe the growth rate of the iteration count $N_{iter}$ with respect to the wave number $\omega$, we introduce a new notation $\rho_{iter}$. Let
\(\omega_1\) and \(\omega_2\) be two wave numbers, and let \(N^{(1)}_{\text{iter}}\) and \(N^{(2)}_{\text{iter}}\) denote the corresponding iteration counts, respectively. Then we define the positive number \(\rho_{\text{iter}}\) by

\[
\frac{N^{(2)}_{\text{iter}}}{N^{(1)}_{\text{iter}}} = \left(\frac{\omega_2}{\omega_1}\right)^{\rho_{\text{iter}}}.
\]

For example, when \(\rho_{\text{iter}} = 1\), the growth is linear; if \(\rho_{\text{iter}} \to 0^+\), then the preconditioner possesses the optimal convergence. For a preconditioner, the positive number \(\rho_{\text{iter}}\) defined above is called “relative growth rate” of the iteration count. Of course, we hope that the relative growth rate \(\rho_{\text{iter}}\) is small.

In Table 5 and Table 6, we compare the iteration counts and its relative growth rate for the CG method and PCG method with the DD preconditioner constructed in Section 3.

### Table 5

Effectiveness of the preconditioner: the case with \(\omega h \approx 1\) and \((q,p) = (2, 4)\)

| \(\omega\) | \(h\) | \(d\) | \(N_{\text{iter}}^{\text{CG}}\) | \(\rho_{\text{iter}}^{\text{CG}}\) | \(N_{\text{iter}}^{\text{PCG}}\) | \(\rho_{\text{iter}}^{\text{PCG}}\) | Err. |
|---|---|---|---|---|---|---|---|
| 20\(\pi\) | \(\frac{1}{51}\) | \(\frac{1}{5}\) | 1556 | 105 | 2504 | 0.6864 | 139 | 0.4047 | 3.8198e-5 |
| 40\(\pi\) | \(\frac{1}{121}\) | \(\frac{1}{11}\) | 5158 | 1.0426 | 9643 | 0.9027 | 191 | 0.3941 | 4.2254e-5 |
| 80\(\pi\) | \(\frac{1}{241}\) | \(\frac{1}{12}\) | 967 | 0.6838 | 1714 | 0.8258 | 251 | 0.3999 | 4.6951e-5 |

### Table 6

Effectiveness of the preconditioner: the case with \(\omega h \approx 2\) and \((q,p) = (3, 5)\)

| \(\omega\) | \(h\) | \(d\) | \(N_{\text{iter}}^{\text{CG}}\) | \(\rho_{\text{iter}}^{\text{CG}}\) | \(N_{\text{iter}}^{\text{PCG}}\) | \(\rho_{\text{iter}}^{\text{PCG}}\) | Err. |
|---|---|---|---|---|---|---|---|
| 20\(\pi\) | \(\frac{1}{51}\) | \(\frac{1}{5}\) | 451 | 77 | 602 | 0.4166 | 104 | 0.4337 | 2.7665e-5 |
| 40\(\pi\) | \(\frac{1}{121}\) | \(\frac{1}{11}\) | 967 | 0.6838 | 1714 | 0.8258 | 251 | 0.3999 | 3.6951e-5 |
| 80\(\pi\) | \(\frac{1}{241}\) | \(\frac{1}{12}\) | 967 | 0.6838 | 1714 | 0.8258 | 251 | 0.3999 | 3.6951e-5 |

The above data indicate that the proposed preconditioner is efficient and the iteration counts of the corresponding PCG method has small relative growth rate when the wave number increases.

### References

[1] C. Alves and C. Chen, A new method of fundamental solutions applied to nonhomogeneous elliptic problems, Advances in Computational Mathematics, 23(2005), 125-142
[2] C. Alves and S. Valtchev, Numerical comparison of two meshfree methods for acoustic wave scattering, Engineering Analysis with Boundary Elements, 29(2005), 371-382
[3] O. Cessenat and B. Després, Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problem, SIAM J. Numer. Anal., 35(1998), 255-299.
[4] H. Chen, P. Lu and X. Xu, A hybridizable discontinuous Galerkin method for the Helmholtz equation with high wave number, SIAM J. Numer. Anal. , 51(2013),2166-2188
[5] E. Deckers, O. Atak, L. Coox, R. Damico, H. Devriendt, S. Jonckheere, K. Koo, B. Phuymers, D. Vandepitte and W. Desmet, The wave based method: An overview of 15 years of research, Wave Motion, 51(2014), 550-565
[6] W. Desmet, A wave based prediction technique for coupled vibro-acoustic analysis. Kuleuven, division PMA, Ph.D. Thesis 98D12, 1998.
[7] L. Demkowicz, J. Gopalakrishnan, I. Muga and J. Zitelli, Wavenumber explicit analysis of a DPG method for the multidimensional Helmholtz equation, Comput. Methods Appl. Mech. Engrg., 213 (2012), 126–138.

[8] C. Dohrmann, A preconditioner for substructuring based on constrained energy minimization. SIAM Journal on Scientific Computing, 2003, 25(1):246-258.

[9] B. Engquist and L. Ying, Sweeping preconditioner for the Helmholtz equation: moving perfectly matched layers. Multiscale modeling simulation, 9(2): 686-710, 2011.

[10] Y. Du and H. Wu, Preasymptotic error analysis of higher order FEM and CIP-FEM for Helmholtz equation with high wave number. SIAM J. Numer. Anal., 53(2015), No. 2, pp. 782-804.

[11] J. Fang, J. Qian, L. Zepeda-Náñez and H. Zhao, Learning Dominant Wave Directions For Plane Wave Methods For High-Frequency Helmholtz Equations, arXiv 1608.08871v1, 2016.

[12] C. Farhat, I. Harari, and U. Hetmaniuk, A discontinuous Galerkin method with Lagrange multipliers for the solution of Helmholtz problems in the mid-frequency regime, Comput. Methods Appl. Mech. Engrg., 192 (2003), 1389-1419.

[13] X. Feng and O. A. Karakashian, Two-level additive Schwarz methods for a discontinuous Galerkin approximation of second order elliptic problems, SIAM J. Numer. Anal., 39 (2001), 1343–1365.

[14] X. Feng and H. Wu, hp-discontinuous Galerkin methods for the Helmholtz equation with large wave number. Math. Comp., 80 (2011), 1997–2024.

[15] X. Feng and Y. Xing, Absolutely stable local discontinuous Galerkin methods for the Helmholtz equation with large wave number, Math. Comp., 82 (2013), pp. 1269-1290.

[16] G. Gabard, Discontinuous Galerkin methods with plane waves for time-harmonic problems, Journal of Computational Physics, 225 (2007), 1961-1984.

[17] C. J. Gittelson, R. Hiptmair, and I. Perugia, Plane wave discontinuous Galerkin methods: Analysis of the h-version, M2AN Math. Model. Numer. Anal., 43 (2009), 297-331.

[18] B. Guo and W. Sun, The optimal convergence of the hp version of the finite element method with quasi-uniform meshes, SIAM J. Numer. Anal., 45 (2007), 698-730.

[19] U. Hetmaniuk, Stability estimates for a class of Helmholtz problems. Commun. Math. Sci., 5 (2007), 665–678.

[20] E. Herbert and C. Waluga, hp analysis of a hybrid DG method for Stokes flow, Ima Journal of Numerical Analysis, 33.2(2013): 687-721.

[21] R. Griesmaier and P. Monk, Error analysis for a hybridizable discontinuous Galerkin method for the Helmholtz equation, J. Sci. Comput., 49 (2011), pp. 291-310.

[22] R. Hiptmair, A. Moiola, and I. Perugia, Plane wave discontinuous Galerkin methods for the 2D Helmholtz equation: analysis of the p-version. SIAM J. Numer. Anal., 49 (2011), No.1, 264–284.

[23] C. Howarth, P. Childs and A. Moiola, Implementation of an interior point source in the ultra weak variational formulation through source extraction, J. Comput. and Appl. Math., 271(2014), 295-306.

[24] Q. Hu and L. Yuan, A weighted variational formulation based on plane wave basis for discretization of Helmholtz equations, Int. J. Numer. Anal. Model., 11 (2014), 587–607.

[25] Q. Hu and L. Yuan, A plane wave least-squares method for time-harmonic Maxwell’s equations in absorbing media, SIAM J. Sci. Comput., 36 (2014), A1911–A1936.

[26] Q. Hu and L. Yuan, A Plane wave method combined with local spectral elements for nonhomogeneous Helmholtz equation and time-harmonic Maxwell equations, Adv. Comput. Math., 44(2018), pp. 245-275.

[27] Q. Hu and H. Zhang, Substructuring preconditioners for the systems arising from plane wave discretization of Helmholtz equations, SIAM J. Sci. Comput., 38(2016), pp. A2232-A2261.

[28] T. Huttunen, P. Gamallo and R. Astley, Comparison of two wave element methods for the Helmholtz problem, Commun. Numer. Meth. Engng., 25(2009), 35-52.

[29] T. Huttunen, M. Malinen and P. Monk, Solving Maxwell’s equations using the ultra weak variational formulation, J. Comput. Phys., 223 (2007), 731–758.

[30] L. Imbert-Gerard and P. Monk, Numerical simulation of wave propagation in inhomogeneous media using Generalized Plane Waves, [arXiv:1511.08251] [math.NA], 2015.

[31] B. Lee, T. Manteuffel, S. McCormick AND J. Ruge, A first-order system least-squares for the Helmholtz equations, SIAM J. SCI. COMPUT., 21(2000), No. 5, 1927-1949.
[32] J. M. Melenk and S. Sauter, Wavenumber explicit convergence analysis for Galerkin discretizations of the Helmholtz equation, SIAM J. Numer. Anal., 49(2011), pp. 1210-1243.
[33] P. Monk and D. Wang, A least-squares method for the helmholtz equation, Comput. Methods Appl. Mech. Engrg., 175 (1999), 121–136.
[34] Szegő, G. Orthogonal Polynomials (fourth edition). Volume 23. AMS Coll. Publications, (1975)
[35] J. Peng, J. Wang and S. Shu, Adaptive BDDC algorithms for the system arising from plane wave discretization of Helmholtz equations, Int. J. Numer. Methods Eng. 116(2018): 683-707.
[36] H. Riou, P. Ladevèze, B. Sourcis, The multiscale VTCR approach applied to acoustics problems, J. Comput. Acous., 16(2008), No. 4, 487-505.
[37] B. Riviére, Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations: Theory and Implementation, SIAM, Philadelphia, 2008.
[38] E. Turkel, D. Gordon, R. Gordon and S. Tsynkov. Compact 2D and 3D sixth order schemes for the Helmholtz equation with variable wave number. J. Comput. Phys., 232(1)(2013), 272–287.
[39] H. Wu, Pre-asymptotic error analysis of CIP-FEM and FEM for the Helmholtz with high wave number. Part I: linear version equation, IMA J. Numer. Anal., (2013), 1-23.
[40] L. Yuan and Q. Hu, A plane wave discontinuous Petro-Galerkin method for Helmholtz equation and time-harmonic Maxwell’s equations with complex wave number, J. Numerical Methods and Computer Applications, 36(2015), 185-196